I N S T I T U T D E
S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

0514

BAYESIAN ENCOMPASSING TEST UNDER PARTIAL OBSERVABILITY

ALMEIDA, C. and M. MOUCHART

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

Bayesian encompassing test under partial observability *

Carlos Almeida and Michel Mouchart †

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

April 27, 2005

Abstract

Florens, Richard and Rolin (2003) proposed a specification test of a parametric hypothesis against a nonparametric one, in the framework of a Bayesian encompassing test. Building on that work we elaborate the procedure under a condition of partial observability. The general procedure is illustrated by the case where only the sign is observable, and more generally when the available data come from a binary reduction of a vector of latent variables. This example is also used to point out some difficulties when implementing the proposed procedure.

Keywords: Bayesian encompassing, partial observability, nonparametric specification test, Dirichlet priors.

*Financial support from the contract 'Project d'Actions de Recherche Concertées' Nr 98/03–217 from the Belgian government, and from the IAP research network Nr P5/24 of the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs) is gratefully acknowledged. Comments by Jean-Marie Rolin, on a previous version, are gratefully acknowledged.

†Corresponding author: mouchart@stat.ucl.ac.be

1 Introduction

This paper focuses on testing a parametric model $\mathcal{E}_0$, parametrized by θ, against another (possibly nonparametric) alternative model $\mathcal{E}_1$, parametrized by ψ, when the actually observed data is a (possibly known) function of latent variables on which are considered both models. More precisely, we have structural models generating latent variables: $\xi = (\xi_1, \ldots, \xi_n)$; $\xi_i \in \mathbb{R}^K$, along with two complete Bayesian specifications, *i.e.* a "null" Bayesian experiment $\mathcal{E}_0$ and an "alternative" Bayesian experiment $\mathcal{E}_1$:

$$\mathcal{E}_0 : \quad \begin{cases} \xi \mid \theta \sim \ P^0_{\xi|\theta} \\ \theta \sim M^0_\theta \end{cases} \qquad Q^0 = M^0_\theta \otimes P^0_{\xi|\theta} \tag{1}$$

$$\mathcal{E}_1 : \quad \begin{cases} \xi \mid \psi \sim \ P^1_{\xi|\psi} \\ \psi \sim M^1_\psi \end{cases} \qquad Q^1 = M^1_\psi \otimes P^1_{\xi|\psi} \tag{2}$$

where $P^0_{\xi|\theta}$ and M^0_θ (resp. $P^1_{\xi|\psi}$ and M^1_ψ) are the sampling transition and the prior probability in $\mathcal{E}_0$ (resp. in $\mathcal{E}_1$).

Some words about notations may be in order. On the one hand we deal with probability measures and transition probabilities (implicitly assuming the existence of regular versions of conditional probabilities) rather than with densities because we shall deal with undominated families of distributions (because of some degeneracies and because of dealing with nonparametric problems). On the second hand, probability measures and transitions are denoted by capital letters with upper and lower indices. Upper indices mark different measures (on a same space) whereas lower indices denote random variables (often under identification with the σ-field generated by these variables); when lower indices are not present, we refer to an implicitly defined complete joint distribution. Often we combine a probability measure and a transition of probability by a Markovian product denoted $\otimes$; more explicitly, when $Q^0_{\theta,\xi}$ is defined by $M^0_\theta \otimes P^0_{\xi|\theta}$, we mean that for any measurable set A on the θ-space and B on the ξ-space the probability on the rectangle $A \times B$ is defined as:

$$Q^0_{\theta,\xi}(A \times B) = \int_{\theta \in A} P^0_{\xi|\theta}(B) \, dM^0_\theta, \tag{3}$$

and the probability measure $Q^0_{\theta,\xi}$ is obtained as the unique extension of (3) to the σ-field generated by the rectangles, based on the θ-space and on the ξ-space.

In the context of a specification test, $\mathcal{E}_0$ is parametric, *i.e.* $\theta \in \Theta$ is a Euclidean parameter whereas $\mathcal{E}_1$ is nonparametric, *i.e.* $\psi \in \Psi$ is a functional parameter. For the sake of expositions, we shall assume that $\mathcal{E}_0$ is identified

(*i.e.* that θ is a minimal sufficient parametrization of the sampling transition in $\mathcal{E}_0$). Later we shall discuss (Theorem 2.1) the identification of $\mathcal{E}_1$. Suppose, for a moment, that ψ is identified in $\mathcal{E}_1$. In order to make the comparison meaningful, there is a proper injection from Θ into Ψ, that is, for all euclidean parameter $\theta \in \Theta$ there is a corresponding unique (because of identification) functional parameter $\psi = h(\theta) \in \Psi$, but the function h is not surjective, more specifically, the range $h(\Theta)$ is a "thin" subset of Ψ.

Besides the structural experiments (1) and (2), there is an observability process that formalizes the fact that ξ is not completely observable. When the observability process is deterministic, two cases should be distinguished: either the observability is completely known and we write $X = g(\xi)$ where X is observable and g is a known function; for instance $X = (\min(\xi_1, \xi_2), \mathbf{1}_{\{\xi_1 \leq \xi_2\}})$ in the case of censored data, or the observability process depends on an unknown parameter α and we write $X = g(\xi, \alpha)$; for instance $X = disc(\xi, \alpha)$ where $disc(\xi, \alpha)$ stands for a discretization of the variable ξ according to the threshold array α, as used in the treatment of ordinal variables, for details in the specification of a discretization model, see Almeida and Mouchart (2003). In this paper, we only consider the first case.

The test is built on the encompassing principle; in the Bayesian framework, the generalities of this procedure are exposed in Florens et *al.* (1990, section 3.5.2) and Florens and Mouchart (1993) and consists in extending the Bayesian experiment $\mathcal{E}_0$ to $\mathcal{E}_0^*$, in order to include, under appropriate conditions, the parameter of the alternative Bayesian experiment $\mathcal{E}_1$. The test is completed by comparing the posterior distribution of ψ in $\mathcal{E}_1$ and in $\mathcal{E}_0^*$. Heuristically, the null hypothesis is not to be rejected if the two posterior distributions are not too different. The quantification of that difference is made using a distance or discrepancy between probability measures defined in the parametric space; for practical reasons, in nonparametric context, only finite dimensional functionals of the parameter will be compared (for a further justification, see Florens et *al.* (2003)).

The extended model $\mathcal{E}_0^*$ is built as follows:

$$\begin{aligned} \mathcal{E}_0^* : \qquad (\theta, \psi, \xi) \ \sim \ Q^{0,*} &= M_\theta^0 \otimes M_{\psi|\theta} \otimes P_{\xi|\theta}^0 = P_\xi^0 \otimes M_{\theta|\xi}^0 \otimes M_{\psi|\theta} \\ &= Q^0 \otimes M_{\psi|\theta}, \end{aligned} \tag{4}$$

where the transition $M_{\psi|\theta}$, called a Bayesian Pseudo True Value (B.P.T.V), endows ψ with an interpretation in $\mathcal{E}_0^*$ under the sufficiency condition

$$\xi \perp\!\!\!\perp \psi \mid \theta; Q^{0,*}; \tag{5}$$

this condition means that the sampling process in the extended experiment $\mathcal{E}_0^*$ is the same as in the null experiment $\mathcal{E}_0$ and, therefore, that θ is a sufficient parametrization in $\mathcal{E}_0^*$.

The order of exposition is as follows. The next section introduces the basic model along with some further notation and summarizes the main findings of Florens et *al.* (2003). Section 3 exposes the main contribution of the paper, namely the general procedure when the observability process is known. Section 4 illustrates the general procedure and the main problems to be faced for the case of the sign observation, and more generally for the case of a binary reduction of a vector of latent variables. The appendix presents the formal proofs of the theorems stated in the paper.

2 The Total Observability case

Let us shortly review the problem treated in Florens et *al.* (2003) viewed as a particular case where $X = \xi$, which is the case of total observability and the alternative model is the non-parametric alternative; thus the encompassing test in this case become a specification test. These authors use the Dirichlet process in the alternative model ($M^1_\psi = \mathcal{D}i(n_0 F_0)$) and the B.P.T.V. is specified as the sampling expectation (under $\mathcal{E}_0$) of the posterior distribution of ψ in $\mathcal{E}_1$, namely:

$$M_{\psi|\theta} = \int M^1_{\psi|\xi} dP^0_{\xi|\theta} \quad (= E^0[M^1_{\psi|\xi} \mid \theta]\,). \tag{6}$$

Then, in $\mathcal{E}^*_0$, ψ and $(\psi \mid \xi)$ are distributed according to mixtures of Dirichlet process:

$$\psi \;\sim\; M^{0,*}_\psi = \int M^1_{\psi|\xi}\; dP^0_\xi \quad (= E^0[M^1_{\psi|\xi}]\,), \tag{7}$$

$$\psi \mid \xi \;\sim\; M^{0,*}_{\psi|\xi} = \int M^1_{\psi|\tilde{\xi}}\; dP^0_{\tilde{\xi}|\xi} \quad (= E^0[M^1_{\psi|\tilde{\xi}} \mid \xi]\,), \tag{8}$$

where $\tilde{\xi} = \{\tilde{\xi}_i : 1 \le i \le n\}$ is a sample from $P^0_{\tilde{\xi}|\xi}$ generated as follows:

- θ is generated from $M^0_{\theta|\xi}$.
- $\{\tilde{\xi}_i : 1 \le i \le n\}$ an *i.i.d.* sample generated from $P^0_{\xi|\theta}$.

A first identification issue needs to be considered in the alternative model. Suppose that ψ were not minimal sufficient and suppose one may exhibit a minimal sufficient parameter ψ^*, *i.e.*:

$$\xi \perp\!\!\!\perp \psi \mid \psi^*; Q^1. \tag{9}$$

Theorem 2.1. *Using the B.P.T.V. defined in* (6), (9) *implies*

$$\psi \perp\!\!\!\perp \theta \mid \psi^*; Q^{0,*}.$$

The proof is in the Appendix. Heuristically, the B.P.T.V. would interpret only ψ^* within model $\mathcal{E}_0^*$ and it would not be natural to interpret, in $\mathcal{E}_0^*$, the unidentified aspect of ψ. From now, we assume that ψ is identified in Q^1, *i.e.* that ψ is a minimal sufficient parametrization of $P^1_{\xi|\psi}$.

When comparing $M^1_{\psi|\xi}$ and $M^{0,*}_{\psi|\xi}$, Florens et *al.* (2003) notice that the usual discrepancy functions, such as the Kullback-Leibler divergence or the Hellinger distance, raise considerable operational difficulties and opt for a direct simulation of finite dimensional functionals defined on the posterior distribution of ξ, actually the first two moments, making use of the discrete representation of the trajectories of the Dirichlet process as developed in Rolin (1992) and Sethuraman (1994).

More specifically, the simulation of the trajectories of ψ conditionally on $\xi = (\xi_1, \ldots, \xi_n)$, an n-size sample, is based on the following representations:
(i) In $\mathcal{E}_1$ from (2), the posterior distibution is:

$$\psi \mid \xi \ \sim \ M^1_{\psi|\xi} = \mathcal{D}i(n_* F_*), \tag{10}$$

where $n_* = n_0 + n$ and $F_* = \dfrac{n_0 F_0 + n F_n}{n_0 + n}$, F_n being the empirical distribution function. Therefore, any ψ_1^ξ, a realization of $(\psi \mid \xi)$, may be represented as:

$$\psi_1^\xi = (1-\gamma) \sum_{1 \le k < \infty} \alpha_k \delta_{\{\eta_k\}} + \gamma \sum_{1 \le i \le n} \beta_i \delta_{\{\xi_i\}}, \tag{11}$$

where $\delta_{\{a\}}$ denotes the Dirac mass measure at the point a and the other elements are subject to the following conditions:

$$\gamma \perp\!\!\!\perp \{\alpha_k : 1 \le k < \infty\} \perp\!\!\!\perp \{\eta_k : 1 \le k < \infty\} \perp\!\!\!\perp \{\beta_i : 1 \le i \le n\},$$

and distributed according to:

$$\begin{aligned}
&\gamma \sim \mathrm{Beta}(n, n_0),\\
&\eta_k : 1 \le k < \infty, \text{ are } i.i.d.\ F_0,\\
&\alpha_k = v_k \prod_{1 \le l \le k-1} (1 - v_l), \quad 1 \le k < \infty,\\
&\qquad \text{where } v_k : 1 \le k < \infty \text{ are } i.i.d.\ \mathrm{Beta}(1, n_0),\\
&\beta_k : 1 \le k < n \text{ are uniformly distributed on } \mathcal{S}_{n-1},
\end{aligned} \tag{12}$$

where $\mathcal{S}_{n-1}$ is the simplex of dimension $n-1$.
(ii) Similarly, any ψ_0^ξ, realization of $(\psi \mid \xi)$ in the model $\mathcal{E}_0^*$, may be represented as:

$$\psi_0^\xi = (1-\gamma) \sum_{1 \leq k < \infty} \alpha_k \delta_{\{\eta_k\}} + \gamma \sum_{1 \leq i \leq n} \beta_i \delta_{\{\tilde{\xi}_i\}} \tag{13}$$

where γ, $\{\alpha_k : 1 \leq k < \infty\}$, $\{\eta_k : 1 \leq k < \infty\}$ and $\{\beta_i : 1 \leq i \leq n\}$ are defined as in (12) and $\{\tilde{\xi}_i : 1 \leq i \leq n\}$ is generated as in (8).

In order to compare the two random probability distributions $M^1_{\psi|\xi}$ and $M^{0,*}_{\psi|\xi}$, finite dimensional characteristics ($\lambda = \ell(\psi) \in \mathbb{R}^s$) of finite approximations of the trajectories are used (equations (11) and (13)). More explicitly, Florens et *al.* (2003) propose the following procedure. For a given sample $\xi = (\xi_1, \ldots, \xi_n)$, two trajectories of ψ are simulated N times, *i.e.* $\psi^\xi_{0,i}$ (resp. $\psi^\xi_{1,i}$) $i = 1, \ldots, N$ by means of (13) (resp. (11)) where each trajectories of ψ is truncated at K points (*i.e.* in the simulations of (11) and (13): $1 \leq k \leq K$ instead of $1 \leq k < \infty$). For each $\psi^\xi_{0,i}$ (resp. $\psi^\xi_{1,i}$), finite dimensional characteristics $\lambda^\xi_{0,i} = \ell(\psi^\xi_{0,i})$ (resp. $\lambda^\xi_{0,i} = \ell(\psi^\xi_{0,i})$) are computed. As the discrepancy function $d(\xi)$ suggested in Florens et *al.* (2003) is based on densities, the density of $(\lambda \mid \xi)$ under $Q^{0,*}$ (resp. Q^1) is obtained by a kernel smoothing on the $\lambda^\xi_{0,i}$'s (resp. $\lambda^\xi_{1,i}$'s) and so $\hat{f}_0(\lambda \mid \xi)$ (resp. $\hat{f}_1(\lambda \mid \xi)$) is obtained. These authors proposed a Monte Carlo approximation of $d(\xi) = \displaystyle\int \log\left(\frac{\hat{f}_1(\lambda \mid \xi)}{\hat{f}_0(\lambda \mid \xi)}\right) \hat{f}_0(\lambda \mid \xi) d\lambda$, namely:

$$\hat{d}(\xi) = \frac{1}{N} \sum_{1 \leq i \leq N} \log \frac{\hat{f}_1(\lambda^\xi_{0,i} \mid \xi)}{\hat{f}_0(\lambda^\xi_{0,i} \mid \xi)} \tag{14}$$

It remains to calibrate $\hat{d}(\xi)$ w.r.t. P^0_ξ, the predictive measure in the null model using Monte Carlo methods. Therefore:

$$\widehat{p-value}\ (\xi) = \frac{1}{R} \sum_{1 \leq r \leq R} \mathbf{1}_{\{\hat{d}(\xi_{(r)}) \geq \hat{d}(\xi)\}} \tag{15}$$

where R is the number of the replicated samples $\xi_{(r)}$'s generated from P^0_ξ.

3 Known observability process

Extension of the models

A first case of partial observability is faced when the observational mechanism is completely known, that is the observation is a known function of the latent

variable, namely:

$$X = g(\xi), \qquad g \text{ is known.} \tag{16}$$

We first extend $\mathcal{E}_0^*$ and $\mathcal{E}_1$ so as to incorporate X. The fact that X is a deterministic function of ξ implies that:

$$X \perp\!\!\!\perp (\theta, \psi) \mid \xi; Q^{0,*} \text{ and } X \perp\!\!\!\perp \psi \mid \xi; Q^1, \tag{17}$$

$$P^0_{X|\xi} = P^1_{X|\xi} \stackrel{def}{=} P_{X|\xi} = \delta_{\{X=g(\xi)\}}, \tag{18}$$

$Q^{0,*}$ and Q^1 now denote the probability measures defining respectively the Bayesian experiments $\mathcal{E}_0^*$ and $\mathcal{E}_1$ extended in order to incorporate X. The sufficiency condition (5) implies

$$\xi \perp\!\!\!\perp \psi \mid X, \theta; Q^{0,*}. \tag{19}$$

Thus, the probability structure of the two extended experiments $\mathcal{E}_1$ and $\mathcal{E}_0^*$ can be described as follows:

$$\begin{aligned} Q^{0,*} &= M^0_\theta \otimes M_{\psi|\theta} \otimes P^0_{\xi|\theta} \otimes P_{X|\xi} \\ &= P^0_X \otimes M^{0,*}_{\psi|X} \otimes M^{0,*}_{\theta|\psi,X} \otimes P^{0,*}_{\xi|X,\theta} \\ Q^1 &= M^1_\psi \otimes P^1_{\xi|\psi} \otimes P_{X|\xi} \\ &= P^1_X \otimes M^1_{\psi|X} \otimes P^1_{\xi|\psi,X}. \end{aligned}$$

Identification issues

The partial observability typically loses some identification (unless X were a sufficient statistics, see Oulhaj and Mouchart (2003)). More specifically, in $\mathcal{E}_0^*$ we may decompose $P^{0,*}_{\xi|\theta}$ into $P^{0,*}_{X|\theta}$ and $P^{0,*}_{\xi|\theta,X}$; the minimal sufficient parameter of $P^{0,*}_{X|\theta}$ may be (strictly) smaller than the minimal sufficient parameter of $P^{0,*}_{\xi|\theta}$. Thus, the partial observability carries us to consider the reduction of models by sufficiency into the models involving minimal sufficient parameters only. Let us denote θ_X and ψ_X, the minimal sufficient parameters for the process generating X, respectively in $\mathcal{E}_0$ and $\mathcal{E}_1$, *viz.*

$$(a) \quad \theta \perp\!\!\!\perp X \mid \theta_X; Q^0, \qquad (b) \quad \psi \perp\!\!\!\perp X \mid \psi_X; Q^1. \tag{20}$$

Next theorem shows that for the B.P.T.V., the sufficiency condition (5) implies the same condition involving identified parameters only.

Theorem 3.1. *Under (20a), we have, in $\mathcal{E}_0^*$, that for any specification of the B.P.T.V. satisfying* (5)*:*

$$X \perp\!\!\!\perp (\psi, \theta) \mid \theta_X; Q^{0,*} \tag{21}$$

The proof is in Appendix. Thus, once a specification of the B.P.T.V. at the level of latent variables ξ satisfies the condition (5), the same condition is automatically satisfied the level of manifest variables X; indeed condition (21) is actually equivalent to $X \perp\!\!\!\perp \psi \mid \theta; Q^{0,*}$ along with the definition of θ_X in condition (20a).

In the context of a specifiation test, the minimal sufficiency of ψ_X in the alternative model ($P^1_{\xi|\psi} = \psi$) is trivially obtained by noticing that

$$\xi \mid \psi; Q^1 \ \sim \ \psi \ \Longrightarrow \ X \mid \psi; Q^1 \ \sim \ \psi \circ g^{-1}$$

therefore $\psi_X = \psi \circ g^{-1}$. When the prior specification of ψ in the alternative model is a Dirichlet process, $\psi \ \sim \ \mathcal{D}i(n_0 F_0)$, then

$$\text{in } Q^1\text{:} \qquad \psi_X \ \sim \ \mathcal{D}i(n_0 F_{0_X}) \text{ with } F_{0_X} = F_0 \circ g^{-1}. \tag{22}$$

The properties of the Dirichlet process imply that the support of the prior specification $\mathcal{D}i(n_0(F_0 \circ g^{-1}))$ is dense (for the weak topology) in the space of probability measures dominated by $F_0 \circ g^{-1}$, *i.e.* for all $\varepsilon > 0$, for all probability measure F dominated by $F_0 \circ g^{-1}$, and for each non null event B ($F_{0_X}(B) > 0$), the probability of the event $\{\psi : |\psi(B) - F(B)| < \varepsilon\}$ is not zero. This feature is often used to comfort the use of a Dirichlet prior specification, see *e.g.* Ferguson (1973), in spite of the fact that the trajectories are almost surely discrete.

Thus, the specification of the alternative model is reduced, in this case, to the specification of n_0, the weight of the prior information, and to the specification of F_{0_X}, the center of the Dirichlet process, where particular attention must be given to the sets of null F_{0_X}-measure .

Two alternative strategies

Let us now consider two alternative strategies for dealing with partially observable variables. A *first strategy* is to specify two models on (ξ, θ) and (ξ, ψ), to define a B.P.T.V. $M_{\psi|\theta}$ and so obtain two extended models characterized by $Q^{0,*}_{\xi,X,\psi,\theta}$ and $Q^1_{\xi,X,\psi}$ from which we construct a test statistic in the form of $d^*(M^{0,*}_{\psi_X|X}, M^1_{\psi_X|X})$. Using the same arguments as above, $M^{0,*}_{\psi_X|X}$ and $M^1_{\psi_X|X}$ are evaluated from $Q^{0,*}_{\xi,X,\psi,\theta}$ and $Q^1_{\xi,X,\psi}$ by integrating out ξ. This strategy implicitly uses a B.P.T.V. $M_{\psi_X|\theta_X}$ of the following form:

$$M_{\psi_X|\theta_X} = \int M_{\psi_X|\theta} \, dM^0_{\theta|\theta_X} \quad (= \ E^0[M_{\psi_X|\theta} \mid \theta_X]) \tag{23}$$

where $M_{\psi_X|\theta}$ is a marginal distribution from some $M_{\psi|\theta}$. When the B.P.T.V. $M_{\psi|\theta}$ has the form (6), the B.P.T.V $M_{\psi_X|\theta_X}$ may also be written as:

$$M_{\psi_X|\theta_X} = \int M^1_{\psi_X|\xi} dP^0_{\xi|\theta_X} \quad (= E^0[M^1_{\psi_X|\xi} \mid \theta_X]). \tag{24}$$

As a *second strategy*, notice that $Q^1_{\xi,\psi}$ implies a unique Q^1_{X,ψ_X} and that $Q^0_{\xi,\theta}$ implies a unique Q^0_{X,θ_X}. Thus, after integrating out ξ, one may extend Q^0_{X,θ_X} into $Q^{0,*}_{X,\theta_X,\psi_X}$ by introducing a B.P.T.V. as in (6), namely:

$$M^{(a)}_{\psi_X|\theta_X} = \int M^1_{\psi_X|X} dP^0_{X|\theta_X} \quad (= E^0[M^1_{\psi_X|X} \mid \theta_X]). \tag{25}$$

For given sampling probabilities $P^0_{\xi|\theta}$ and $P^1_{\xi|\psi}$ and given observability process $X = g(\xi)$, the answer to the question whether, or not, we have:

$$M_{\psi_X|\theta_X} = M^{(a)}_{\psi_X|\theta_X} \tag{26}$$

depends on the prior specification M^0_θ and M^1_ψ and on the B.P.T.V $M_{\psi|\theta}$ when it is not built as in (6). Condition (26) will be called a *condition of embeddability*. Next theorem gives a sufficient condition on the B.P.T.V. in order to ensure that condition of embeddability

Theorem 3.2. *Let $Q^{0,*}_{X,\psi,\theta} = Q^{0,*}$ be constructed from a prior specification $M^0_{\theta_X}$ and B.P.T.V. such that:*

$$\psi \perp\!\!\!\perp \theta \mid \psi_X; Q^{0,*} \tag{27}$$

$$\psi \perp\!\!\!\perp \theta \mid \theta_X; Q^{0,*}, \tag{28}$$

then $M_{\psi|\theta}$ has necessarily the form:

$$M_{\psi|\theta} = M_{\psi_X|\theta_X} \otimes M_{\psi|\psi_X}. \tag{29}$$

and

$$(X, \theta_X) \perp\!\!\!\perp \psi \mid \psi_X; Q^{0,*} \tag{30}$$

Furthermore, the condition of embeddability (26) *is satisfied for whatever prior specification of M^0_θ and M^1_ψ satisfying the conditions* (27) *and* (28).

See the proof in the Appendix. The conditions in the Theorem 3.2 say that in the extended model $\mathcal{E}^*_0$, the complete parameters from both models are independent conditionally on the identified parameters of either model; next section gives an example where these conditions are satisfied.

These conditions also permit to write the posterior distribution of (ψ, θ) as follows:

$$M^{0,*}_{\psi,\theta|X} = M^{0,*}_{\psi_X,\theta_X|X} \otimes M^0_{\theta|\theta_X} \otimes M_{\psi|\psi_X}. \tag{31}$$

The second conclusion (30) of the theorem implies the sufficiency, in the extended model, of the parameter identified in the alternative model (ψ_X).

Under the conditions of this theorem, the posterior distributions of $(\psi \mid X)$ in both models are given by:

$$M^{0,*}_{\psi|X} = M^{0,*}_{\psi_X|X} \otimes M_{\psi|\psi_X}, \tag{32}$$

$$M^1_{\psi|X} = M^1_{\psi_X|X} \otimes M^1_{\psi|\psi_X}. \tag{33}$$

As $(\psi \mid \psi_X)$ is not revised by X, we can suppose, as a coherence condition, that its distribution is the same in both models ($M_{\psi|\psi_X} = M^1_{\psi|\psi_X}$). Using the specification of the B.P.T.V. (25) and the conditions of the Theorem 3.2, the encompassing test statistic will therefore be based on the comparison between $M^{0,*}_{\psi_X|X}$ and $M^1_{\psi_X|X}$ only.

The main conclusion of this section is that if the observation process is completely known, the test is constructed at the level of manifest variables, then, under suitable conditions, the loss of information due to the partial observability is taken into account through the reduction to the identified parameters, independently of the encompassing procedure. In other words, this seemingly natural conclusion nevertheless requires specific conditions in order to ensure the coherence between the structural models, in terms of the latent variables, and the statistical models, in terms of the manifest variables.

Similarly to section 2, the test would be based on a finite dimensional functional λ defined on the posterior distribution of $(\psi_X \mid X)$. Therefore, the encompassing test statistics takes the form:

$$d(X) = d^*(M^{0,*}_{\lambda|X}, M^1_{\lambda|X}) \tag{34}$$

and is calibrated against P^0_X, the predictive distribution of the manifest variable under $\mathcal{E}_0$.

4 The sign observation

The Model

Let us now illustrate, by means of the simplest example, namely the observation of the sign of a latent variable, the main ideas underlying the procedure sketched in the previous section. Moreover, considering particular specifications sheds also some lights on potential difficulties.

Let us consider the following situation:

$$\mathcal{E}_0 : \quad \begin{cases} \xi \mid (\mu, \sigma^2) \ \sim \ N(\mu, \sigma^2) \\ (\mu, \sigma^2) \ \sim \ M^0_{\mu,\sigma^2} \end{cases} \qquad \mathcal{E}_1 : \quad \begin{cases} \xi \mid \psi \ \sim \ \psi \\ \psi \ \sim \ \mathcal{D}i(n_0 F_0) \end{cases} \tag{35}$$

Here, $\mathcal{E}_0$ and $\mathcal{E}_1$ are two Bayesian experiments at the level of real valued but unobservable variables and we are interested in comparing the "null" model under normality hypothesis ($\mathcal{E}_0$) against a nonparametric alternative ($\mathcal{E}_1$).

The partial observability aspect is captured by observing only the sign of the latent variable, *viz*:

$$X = \mathbf{1}_{\{\xi \geq 0\}} \tag{36}$$

The statistical model is obtained by marginalization on the observed variable X.

$$\mathcal{E}_0 : \begin{cases} X \mid (\mu, \sigma^2) \ \sim \ \mathrm{Be}(N(\frac{\mu}{\sigma})) \\ (\mu, \sigma^2) \ \sim \ M^0_{\mu,\sigma^2} \end{cases} \qquad \mathcal{E}_1 : \begin{cases} X \mid \psi \ \sim \ \mathrm{Be}(\psi([0, \infty[)) \\ \psi \ \sim \ \mathcal{D}i(n_0 F_0) \end{cases} \tag{37}$$

where $\mathrm{Be}(\cdot)$ denotes the Bernoulli distribution and N now stands for the distribution function of the standardized normal distribution. Therefore the minimal sufficient parameters, θ_X and ψ_X, are

$$\theta_X = t(\mu, \sigma) = N\left(\frac{\mu}{\sigma}\right) \qquad\qquad \psi_X = \psi([0, \infty[)$$

Using the characteristics of the Dirichlet process, the statistical models reduced on the identified parameters can be written as:

$$\mathcal{E}_0 : \begin{cases} X \mid \theta_X \ \sim \ \mathrm{Be}(\theta_X) \\ \theta_X \sim M^0_{\theta_X} = M_{\mu,\sigma^2} \circ t^{-1} \end{cases} \quad \mathcal{E}_1 : \begin{cases} X \mid \psi_X \ \sim \ \mathrm{Be}(\psi_X) \\ \psi_X \sim \mathrm{Beta}(n_0 f_0, n_0(1 - f_0)) \end{cases} \tag{38}$$

where $f_0 = F_0([0, \infty[)$ and $\mathrm{Beta}(a, b)$ is the beta distribution with density:

$$p_\beta(u \mid a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} u^{a-1}(1-u)^{b-1} \mathbf{1}_{[0,1]}(u). \tag{39}$$

(Then, $E[U] = \dfrac{a}{a+b}$ and $Var[U] = \dfrac{ab}{(a+b)^2(a+b+1)}$).

In this case, the sampling probabilities are the same under $\mathcal{E}^0$ and under $\mathcal{E}^1$; when $\mathcal{E}^1$ is reduced to the manifest variable X, ψ_X is not any more a functional parameter and the sampling model is a saturated one. This will be so as soon as the range space of $X = g(\xi)$ has finite cardinality.

Degenerate B.P.T.V.

As the sampling process is the same under $\mathcal{E}^0$ and $\mathcal{E}^1$, a seemingly natural form of the B.P.T.V might be:

$$M_{\psi_X|\theta_X} = \delta_{\{\theta_X\}}, \tag{40}$$

rather than a specification satisfying (6). In such a case, from (40), we obtain

$$M^{0,*}_{\psi_X|X} = \int \delta_{\{\theta_X\}} dM^0_{\theta_X|X} = M^0_{\theta_X|X}. \tag{41}$$

Moreover, as ψ_X and θ_X represent a same parameter of a unique sampling process (in $\mathcal{E}^0$ and $\mathcal{E}^1$) they may be identified and denoted by π and equation (41) says that $M^{0,*}_{\psi_X|X}$ may be actually written as $M^0_{\pi|X}$. Thus the test statistics may be written as:

$$d^*(M^{0,*}_{\psi_X|X}, M^1_{\psi_X|X}) = d^*(M^0_{\pi|X}, M^1_{\pi|X}). \tag{42}$$

In particular:

$$M^0_\pi = M^1_\pi \quad \Longrightarrow \quad M^0_{\pi|X} = M^1_{\pi|X}. \tag{43}$$

Thus, if coherence is invoked to force the prior distribution on π to be identical in $\mathcal{E}^0$ and in $\mathcal{E}^1$, test statistics (34) is almost surely equal to zero and the procedure simply breaks down.

Both the null and the alternative models are consistent in the sense of the weak convergence to a Dirac mass for the posterior distribution, see Ghosh and Ramamoorthy (2002). Thus when the coherence condition (43) is not fulfilled, the total variation distance between posterior distributions in both models converges nevertheless almost surely to zero when the sample size tends to infinity (see Theorem 1.3.1 Ghosh and Ramamoorthi (2002)) and again the procedure breaks down at least asymptotically.

Non degenerate B.P.T.V.

a) Construction of a non-degenerate B.P.T.V.

The degenerate specification (40) may be justified when the interpretation of a parameter only depends on the sampling process of a variable (here, X) identifying that parameter (here, ψ_X and θ_X). When the consequence (41), or (43), does not seem palatable, one may argue that the interpretation of a parameter should depend "on the world where it is embedded", *i.e.* that $\mathcal{E}^0$ and $\mathcal{E}^1$ actually represent two different structural visions of the world, at the level of the latent variables, and that (40) does by no means represent

any necessity. Thus let us now examine the situation with a non degenerate B.P.T.V.

The specification of the alternative model in (38) implies that the posterior distribution $M^1_{\psi_X|X}$ is a beta distribution, more specifically:

$$\begin{aligned} M^1_{\psi_X|X} &= \text{Beta}(n_0 f_0 + X, n_0(1-f_0) + 1 - X) \\ &= X \text{ Beta}(n_0 f_0 + 1, n_0(1-f_0)) \\ &\quad + (1-X) \text{ Beta}(n_0 f_0, n_0(1-f_0)+1). \end{aligned} \tag{44}$$

Motivated by the results of section 2, the specification of the B.P.T.V. involves the identified parameters only. Interpreting the B.P.T.V. as the bridge between the two approaches, encapsulated in $\mathcal{E}^0$ and $\mathcal{E}^1$, the specification (6) is natural and may now be written as:

$$\begin{aligned} M_{\psi_X|\theta_X} &= \int M^1_{\psi_X|X} dP^0_{X|\theta_X} \\ &= \theta_X \text{ Beta}(n_0 f_0 + 1, n_0(1-f_0)) \\ &\quad + (1-\theta_X) \text{ Beta}(n_0 f_0, n_0(1-f_0)+1) \end{aligned} \tag{45}$$

Then in the extended model $\mathcal{E}^*_0$, the prior measure of ψ_X and the posterior distribution of $\psi_X \mid X$ are mixtures of beta distributions given by:

$$\begin{aligned} M^{0,*}_{\psi_X} &= E^0[\theta_X] \text{ Beta}(n_0 f_0 + 1, n_0(1-f_0)) \\ &\quad + (1 - E^0[\theta_X]) \text{ Beta}(n_0 f_0, n_0(1-f_0)+1) \end{aligned} \tag{46}$$

$$\begin{aligned} M^{0,*}_{\psi_X|X} &= a(X) \text{ Beta}(n_0 f_0 + 1, n_0(1-f_0)) \\ &\quad + (1 - a(X)) \text{ Beta}(n_0 f_0, n_0(1-f_0)+1). \end{aligned} \tag{47}$$

with

$$a(X) = E^0[\theta_X \mid X] = \frac{\int \theta_X^{1+X}(1-\theta_X)^{1-X} dM^0_{\theta_X}}{\int \theta_X^{X}(1-\theta_X)^{1-X} dM^0_{\theta_X}} \tag{48}$$

As ψ_X is finite-dimensional, the encompassing test statistics boils down to evaluate some discrepancy function between (47) and (44), *i.e.* an integral on $[0,1]$ of a non-negative function. For instance, if the posterior distribution in the null model can be approximated by a beta distribution and the posterior expectation can be approximated by :

$$E^0[\theta_X \mid X] \approx \frac{a_0 + X}{a_0 + b_0 + 1} \tag{49}$$

then the L_1 distance, denoted d_1, may be approximated as follows:

$$d_1(M^{0,*}_{\psi_X|X}, M^1_{\psi_X|X}) = |a(X) - X| \int_0^1 |u - (1-f_0)| p_\beta(u \mid n_0 f_0, n_0(1-f_0)) du$$
$$\approx \frac{b_0 X + a_0(1-X)}{a_0 + b_0 + 1} \; E^1[|\psi_X - (1-f_0)|]. \tag{50}$$

Details are given in the Appendix.

b) Convergence

One question arises here: what happens when the sample size tends to infinity? The intuition suggests that the two models would not be discriminated because of a same sampling model and, if the sample size increases, the importance of the prior judgment must be decreasing. The following theorem comforts that intuition. For that purpose, let us write d_W for a metric which metrizes the weak convergence - for instance: Lévi (Huber,1981, Th.3.3) or Prohorov (Huber, 1981, Th.3.8) - and consider a test statistic (34) in the form:

$$d(X_1^n) = d_W(M^{0,*}_{\psi_X|X_1^n}, M^1_{\psi_X|X_1^n}). \tag{51}$$

Then we obtain the following theorem:

Theorem 4.1. *Under the condition of embeddability, given in Theorem 3.2, a test statistic* (34) *in the form* (51)*, tends to zero almost surely relative to the predictive measure of the null model, viz.*

$$P^0_{X_1^\infty}[d(X_1^n) \longrightarrow 0] = 1 \tag{52}$$

The proof is in the appendix

Some numerical computations may be helpful to visualize that convergence and are summarized in the table 1. The distance considered was the Kantorovich-Wasserstein one, the definition of which is:

$$d_W(\mu, \nu) := \int_{-\infty}^{\infty} |\, \mu(]-\infty, y]) - \nu(]-\infty, y]) \,| \; dy \tag{53}$$

This distance, valued in $[0, 1]$, is also known to metrize the weak convergence in the space of probability measures on the real line with a fixed bounded support (see the mathematical Appendix of Bickel and Freedman (1981) for details). In this example, the prior in the null model is a normal distribution with unit mean and unit variance and for the alternative model, $n_0 = 1$,

and $f_0 = F_0([0, \infty[) = 0.5$. For finite sample sizes, namely $5, 10, 20, 50$ and 100, and finite sample space, namely $\{0, 1\}^n$ it is possible to enumerate all possible values of $X_1, \ldots, X_n$ and their exact predictive probabilities (*i.e* P^0) and therefore their expectations and variances.

n	5	10	20	50	100
$\max d(X_1^n)$	0.1844	0.1143	0.0653	0.0408	0.0290
$E[d(X_1^n)]$	0.0820	0.0597	0.0426	0.0270	0.0190
$\sigma[d(X_1^n)]$	0.0314	0.0202	0.0147	0.0103	0.0078

Table 1: Characteristics of the Distribution of the Statistics Test, $\mu_0 = 1, n_0 = 1, f_0 = 0.5$

.

Table 1 summarizes those numerical results and illustrates the actual speed of convergence, to zero with respect to the predictive distribution P^0, of the encompassing test statistics.

Binary choice model

A very similar analysis can be done for the *Binary Choice Model*. Let $\eta = (\eta_1, \eta_2)$ be two latent variables representing, in some context, latent utilities. If we define $\xi = \eta_2 - \eta_1$, the partial observability process, as used *e.g.* in discrete choice models, is given by:

$$X = g(\eta_1, \eta_2) = \mathbf{1}_{\{\eta_1 < \eta_2\}} = \mathbf{1}_{\{\xi \geq 0\}} \tag{54}$$

Considering two models as above and using the same notations for the minimal sufficient parameters, we have:

$$\mathcal{E}_0 : \begin{cases} X \mid \theta_X & \sim \ \mathrm{Be}(\theta_X) \\ \theta_X & \sim \ M^0_{\theta_X} = M^0_\theta \circ t^{-1} \end{cases} \qquad \mathcal{E}_1 : \begin{cases} X \mid \psi_X & \sim \ \mathrm{Be}(\psi_X) \\ \psi_X & \sim \ \mathrm{Beta}(n_0 f_0, n_0(1 - f_0)) \end{cases} \tag{55}$$

where θ represents the parameter characterizing the joint distribution of (η_1, η_2) and $\theta_X = t(\theta) = P^0_{\xi|\theta}(\eta_1 < \eta_2)$ and $\psi_X = P^1_{\xi|\psi}(\eta_1 < \eta_2) = \psi(\eta_1 < \eta_2)$. We therefore obtain the same structure as in the sign observation model.

5 Concluding Remarks

This paper treats two issues pervading the problems of statistical modelling, particularly in social sciences and biostatistics. One issue is that the distributional assumptions are often not based on a contextual argument and

are rather justified by an approximation one; in such circumstances a basic concern is to evaluate whether a parametric approximation of a "true" distribution is fairly reliable. This is the issue of a specification test. Another issue is that, in a structural approach to model building, latent variables provide a natural support for a model firstly based on contextual knowledge and the observability problem comes into the picture at a second stage when a statistical model, bearing on manifest (or observable) variables only, is deduced from the structural model. As a matter of fact, the discretization model of a latent variable as a model for analyzing ordinal data (see *e.g.* Almeida and Mouchart (2003)) has been instrumental in stimulating the authors' interest for this approach. Thus the object of this paper has been to integrate the two issues of specification testing and of partial observability.

The contribution of this paper is twofold. Firstly, when elaborating the general framework we have singled out identification issues and alternative strategies for a complete specification of the model. Next we have elaborated the detail of the procedure until we obtain an operational solution. The generality of the proposed general procedure has been controlled by working out completely the particular case of the sign observation. This particular case has drawn our attention on several aspects that come up when applying the procedure, in particular an identification problem put to an extreme, the meaning of a parameter bound, or not, to a particular model (implying alternative strategies for specifying the Bayesian Pseudo True Value) and the problem of the asymptotic distinguishability in that particular case.

Appendix: Formal Complements

In general, unargued implications come from basic properties of conditional independence, most often, through a direct application of Theorem 2.2.10 in Florens et *al.*(1990).

5.1 Proof of Theorem 2.1

$$
\begin{aligned}
(6) \quad \Leftrightarrow M_{\psi|\theta} &= \int M^1_{\psi|\xi} dP^0_{\xi|\theta} \\
&= \int \left[M^1_{\psi^*|\xi} \otimes M^1_{\psi|\psi^*,\xi} \right] dP_{\xi|\theta} \\
(9) \quad \Rightarrow \qquad &= \int \left[M^1_{\psi^*|\xi} \otimes M^1_{\psi|\psi^*} \right] dP_{\xi|\theta} \\
&= \left[\int M^1_{\psi^*|\xi} dP_{\xi|\theta} \right] \otimes M^1_{\psi|\psi^*} \\
(6) \quad \Rightarrow \qquad &= M_{\psi^*|\theta} \otimes M^1_{\psi|\psi^*}
\end{aligned}
$$

■

5.2 Proof of Theorem 3.1

Remember that all ensuing conditional independence hold in $Q^{0,*}$.

$$
\begin{aligned}
&(5) \text{ and } (16) \quad \Rightarrow \quad X \perp\!\!\!\perp \psi \mid \theta && (56) \\
&(20a) \text{ and } (56) \Rightarrow \quad X \perp\!\!\!\perp (\theta, \psi) \mid \theta_X && (57)
\end{aligned}
$$

■

5.3 Proof of theorem 3.2

$$
\begin{aligned}
(28) \Rightarrow \quad M_{\psi|\theta} &= M_{\psi|\theta_X} && (58) \\
&= M_{\psi_X|\theta_X} \otimes M_{\psi|\psi_X,\theta_X} && (59) \\
(27) \Rightarrow \quad &= M_{\psi_X|\theta_X} \otimes M_{\psi|\psi_X}, && (60)
\end{aligned}
$$

$$
\begin{aligned}
&(5) && \Rightarrow \quad X \perp\!\!\!\perp \psi \mid \theta && (61) \\
&(61) \text{ and } (28) && \Rightarrow \quad (\theta, X) \perp\!\!\!\perp \psi \mid \theta_X && (62) \\
&(62) && \Rightarrow \quad X \perp\!\!\!\perp \psi \mid \theta_X && (63)
\end{aligned}
$$

Then, (30) follows from (63) and (27) ■

5.4 Proof of equation (50)

The L_1 distance between the two posterior distributions (44) and (47) may be written as:

$$\begin{aligned} d_1(M^{0,*}_{\psi_X|X}, M^1_{\psi_X|X}) &= |a(X) - X| \\ &\cdot \int_0^1 |p_\beta(u \mid n_0 f_0 + 1, n_0(1 - f_0)) - p_\beta(u \mid n_0 f_0, n_0(1 - f_0) + 1)| du \end{aligned} \tag{64}$$

By the beta density definition, this equation is equivalent to:

$$\begin{aligned} d_1(M^{0,*}_{\psi_X|X}, M^1_{\psi_X|X}) &= |a(X) - X| \\ &\cdot \int_0^1 \left| \frac{\Gamma(n_0 f_0 + 1)\Gamma(n_0(1 - f_0))}{\Gamma(n_0 + 1)} u^{n_0 f_0} (1 - u)^{n_0(1 - f_0) - 1} \right. \\ &\left. - \frac{\Gamma(n_0 f_0)\Gamma(n_0(1 - f_0) + 1)}{\Gamma(n_0 + 1)} u^{n_0 f_0 - 1} (1 - u)^{n_0(1 - f_0)} \right| du \end{aligned} \tag{65}$$

which can be simplified to

$$\begin{aligned} &= |a(X) - X| \\ &\quad \cdot \int_0^1 |f_0 u - (1 - f_0)(1 - u)| \, p_\beta(u \mid n_0 f_0, n_0(1 - f_0)) du \end{aligned} \tag{66}$$

$$\begin{aligned} &= |a(X) - X| \\ &\quad \cdot \int_0^1 |u - (1 - f_0)| \, p_\beta(u \mid n_0 f_0, n_0(1 - f_0)) du \end{aligned} \tag{67}$$

$$= |a(X) - X| \; E^1[|\psi_X - 1 - f_0|] \tag{68}$$

Using the approximation of the posterior expectation (49), (68) may be written as (50). ■

5.5 Proof of theorem 4.1

5.5.1 Main Argument

From (44), the posterior expectation and variance for the alternative models are:

$$E^1[\psi_X \mid X_1^n] = \frac{n_0}{n_0+n} f_0 + \frac{1}{n_0+n}\overline{X} \tag{69}$$

$$Var^1[\psi_X \mid X_1^n] = \frac{(n_0 f_0 + n\overline{X})(n_0(1-f_0)+n(1-\overline{X}))}{(n_0+n)^2(n_0+n+1)} \tag{70}$$

Then, with respect to the null model:

$$P^0[E^1[\psi_X \mid X_1^n] \longrightarrow \theta_X \mid \theta_X] = 1 \tag{71}$$

$$P^0[Var^1[\psi_X \mid X_1^n] \longrightarrow 0 \mid \theta_X] = 1 \tag{72}$$

$$\text{Therefore} \quad P^0[M^1_{\psi_X|X_1^n} \xrightarrow{w} \delta_{\{\theta_X\}} \mid \theta_X] = 1 \tag{73}$$

Similarly in the extended model (47), we show below that

$$E^{0,*}[\psi_X \mid X_1^n] = \frac{n_0}{n_0+n} f_0 + \frac{n}{n_0+n} E^0[\theta_X \mid X_1^n] \tag{74}$$

$$Var^{0,*}[\psi_X \mid X_1^n] \leq \frac{n_0 f_0 + 2nE^0[\theta_X \mid X_1^n] - nE^0[\theta_X^2 \mid X_1^n]}{(n_0+n)^2} + \frac{n^2}{(n_0+n)^2} Var^0[\theta_X \mid X_1^n] \tag{75}$$

which implies:

$$P^0[E^{0,*}[\psi_X \mid X_1^n] \longrightarrow \theta_X \mid \theta_X] = 1 \tag{76}$$

$$P^0[Var^{0,*}[\psi_X \mid X_1^n] \longrightarrow 0 \mid \theta_X] = 1 \tag{77}$$

$$\text{Therefore} \quad P^0[M^{0,*}_{\psi_X|X_1^n} \xrightarrow{w} \delta_{\{\theta_X\}} \mid \theta_X] = 1 \tag{78}$$

If d_W is a distance which metrizes the weak convergence, (73) and (78) implies

$$P^0[d_W(M^{0,*}_{\psi|X_1^n}, M^1_{\psi|X_1^n}) \longrightarrow 0 \mid \theta_X] = 1 \tag{79}$$

which implies:

$$P^0[d_W(M^{0,*}_{\psi|X_1^n}, M^1_{\psi|X_1^n}) \longrightarrow 0] = 1 \tag{80}$$

■

5.5.2 Proof of the equation (74)

We start with a usual identity:

$$\begin{aligned} E^{0,*}[\psi_X \mid X_1^n] &= E^{0,*}[E^{0,*}[\psi_X \mid X_1^n, \theta_X] \mid X_1^n] \\ &= E^0[E^{0,*}[\psi_X \mid \theta_X] \mid X_1^n] \qquad \text{from (21)} \end{aligned} \tag{81}$$

Now, from (25) and Theorem 3.2,

$$E^{0,*}[\psi_X \mid \theta_X] = \int E^1[\psi_X \mid X_1^n] dP^0_{X_1^n|\theta_X} \tag{82}$$

$$= \int \frac{n_0 f_0 + n\overline{X}}{n + n_0} dP^0_{X_1^n|\theta_X} \qquad \text{from (69)} \tag{83}$$

$$= \frac{n_0 f_0 + nE^0[\overline{X} \mid \theta_X]}{n + n_0} \tag{84}$$

$$= \frac{n_0 f_0 + n\theta_X}{n + n_0} \tag{85}$$

Then (81) becomes:

$$E^{0,*}[\psi_X \mid X_1^n] = E^{0,*}[\frac{n_0 f_0 + n\theta_X}{n + n_0} \mid X_1^n] \tag{86}$$

$$= \frac{n_0 f_0 + nE^0[\theta_X \mid X_1^n]}{n + n_0} \tag{87}$$

■

5.5.3 Proof of the equation (75)

Again, we start with a usual identity:

$$\begin{aligned} Var^{0,*}[\psi_X \mid X_1^n] &= E^{0,*}[Var^{0,*}[\psi_X \mid X_1^n, \theta_X] \mid X_1^n] \\ &\quad + Var^{0,*}[E^{0,*}[\psi_X \mid X_1^n, \theta_X] \mid X_1^n] \end{aligned} \tag{88}$$

$$\begin{aligned} &= E^{0,*}[Var^{0,*}[\psi_X \mid \theta_X] \mid X_1^n] \\ &\quad + Var^{0,*}[E^{0,*}[\psi_X \mid \theta_X] \mid X_1^n] \end{aligned} \tag{89}$$

where (89) is obtained from (21). Note that:

$$Var^{0,*}[\psi_X \mid \theta_X] = E^{0,*}[\psi_X^2 \mid \theta_X] - [E^{0,*}[\psi_X \mid \theta_X]]^2 \tag{90}$$

Using the same argument as for (82)

$$E^{0,*}[\psi_X^2 \mid \theta_X] = \int E^1[\psi_X^2 \mid X_1^n] dP^0_{X_1^n|\theta_X} \tag{91}$$

$$= \int \frac{(n_0 f_0 + n\overline{X})(n_0 f_0 + n\overline{X} + 1)}{(n+n_0)(n+n_0+1)} dP^0_{X_1^n|\theta_X} \quad \text{from (69)} \tag{92}$$

$$= \int \frac{n_0 f_0(n_0 f_0 + 1) + (2n_0 f_0 + 1)n\overline{X} + (n\overline{X})^2}{(n+n_0)(n+n_0+1)} dP^0_{X_1^n|\theta_X} \tag{93}$$

$$= \frac{n_0 f_0(n_0 f_0 + 1) + 2n(n_0 f_0 + 1)\theta_X + n(n-1)\theta_X^2}{(n+n_0)(n+n_0+1)} \tag{94}$$

From the general property of a variance, we successively derive:

$$Var^{0,*}[\psi_X \mid \theta_X] \leq \frac{n_0 + n + 1}{n_0 + n} E^{0,*}[\psi_X^2 \mid \theta_X] - [E^{0,*}[\psi_X \mid \theta_X]]^2 \tag{95}$$

$$= \frac{n_0 f_0(n_0 f_0 + 1) + 2n(n_0 f_0 + 1)\theta_X + n(n-1)\theta_X^2}{(n+n_0)^2} - \frac{(n_0 f_0 + n\theta_X)^2}{(n_0+n)^2} \tag{96}$$

$$= \frac{n_0 f_0 + 2n\theta_X - n\theta_X^2}{(n_0+n)^2} \tag{97}$$

Therefore:

$$E^{0,*}[Var^{0,*}[\psi_X \mid \theta_X] \mid X_1^n] \leq \frac{n_0 f_0 + 2nE^0[\theta_X \mid X_1^n] - nE^0[\theta_X^2 \mid X_1^n]}{(n_0+n)^2} \tag{98}$$

Now

$$Var^{0,*}[E^{0,*}[\psi_X \mid \theta_X] \mid X_1^n] = Var^{0,*}\left[\frac{n_0 f_0 + n\theta_X}{n_0+n} \mid X_1^n\right] \tag{99}$$

$$= \frac{n^2}{(n_0+n)^2} Var^0[\theta_X \mid X_1^n] \tag{100}$$

Thus:

$$Var^{0,*}[\psi_X \mid X_1^n] \leq \frac{n_0 f_0 + 2nE^0[\theta_X \mid X_1^n] - nE^0[\theta_X^2 \mid X_1^n]}{(n_0+n)^2} + \frac{n^2}{(n_0+n)^2} Var^0[\theta_X \mid X_1^n] \tag{101}$$

Clearly, the first term tends to zero as n tends to infinity and also the second term because the posterior variance of θ_X tends to zero by Doob's Theorem (see Doob (1949) and also Ghosh and Ramamoorthy (2003, Theorem 1.3.2)).

■

References

Almeida, C. and Mouchart, M. (2003a). Identification of polychoric correlations: A copula approach, Discussion Paper DP0321, Institut de statistique, Université catholique de Louvain.

Almeida, C. and Mouchart, M. (2003b). A Note on a copula approach to polychoric correlations, Discussion Paper DP0322, Institut de statistique, Université catholique de Louvain.

Bickel, P.J. and Freedman, D.A. (1981). Some Asymptotic Theory for the Bootstrap, *Annals of Statistics*, **9** , 1196 – 1217.

Doob, J.L. (1949). Application to the theory of martingales. *Colloques Internationaux du C.N.R.S.*, Paris: C.N.R.S., 22 – 28.

Ferguson, T.S (1973). A Bayesian Analysis of some nonparametric problems, *Annals of Statistics*, **1** , 209 – 230.

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

Florens, J.-P. and Mouchart, M. (1993). Bayesian Testing and Testing Bayesians, chap. 11. in Maddala, G., Rao, C. and Vinod, H. (Ed.), *Handbook in Statistics*, Amsterdam: North Holland, 303 – 333.

Florens, J.-P., Richard, J.-F. and Rolin, J.-P (2003). Bayesian Encompassing Specification Test of a parametric model against a nonparametric alternative . *unpublished Mimeo.*

Ghosh, J.K. and Ramamoorthi, R.V. (2002). *Bayesian Nonparametrics*, New York: Springer-Verlag.

Huber, P.J. (1981). *Robust Statistics*, New York: Wiley.

Mouchart, M. and Oulhaj, A. (2003). A Note on Partial Sufficiency with connection to the Identification Problem, *Metron*, **LXI**, 267 – 283.

Rolin, J.M. (1992). Some useful properties of the Dirichlet process, Discussion Paper DP9202, Institut de statistique, Université catholique de Louvain.

Sethuraman (1994). A constructive definition of the Dirichlet process. *Statistica Sinica* **11** , 639 – 650.