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

0707

TESTING THE NORMALITY OF THE LATENT VARIABLES IN THE POLYCHORIC CORRELATION MODEL

ALMEIDA, C. and M. MOUCHART

This file can be downloaded from
http://www.stat.ucl.ac.be/ISpub

Testing the normality of the latent variables in the polychoric correlation model*

Carlos ALMEIDA † and Michel MOUCHART ‡

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

April 17, 2007

Abstract

This paper develops a specification of the model defining the polychoric correlations, where the manifest ordinal variables are assumedly generated as a discretization of latent continuous ones. Taking advantage of the fact that in this model, the marginal distributions of the latent variables are not identified, we use the tool of the copula concept; some identification issues are analysed. In the second part, we develop a Bayesian encompassing specification test for testing the Gaussianity of the underlying copula and consider the discretization model as a case of partial observability.

Keywords: Bayesian encompassing, partial observability, nonparametric specification test, discretization model, Dirichlet priors, polychoric correlations, ordinal variables.

AMS 2000 subject classification: 62B05, 62F03, 62F15

*Financial support from the IAP research network Nr P5/24 of the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs) is gratefully acknowledged. Without implications, discussions with J.-M. Rolin and comments by C. Genest and by participants of seminars at the Institute of Statistics, U.C.L. (Louvain la Neuve, Belgium) and at Department of Statistics, P.U.C. (Santiago, Chile) have been really appreciated.

†tel. + 32 10 47 91 74, fax: + 32 10 47 30 32, e-mail: almeida@stat.ucl.ac.be

‡tel. + 32 10 47 43 18, fax: + 32 10 47 30 32, e-mail: mouchart@stat.ucl.ac.be (Corresponding author)

1 Introduction

A frequently used approach for the statistical analysis of ordinal data consists in modelling the data as a discretization of an underlying latent variable. The idea is, for instance, at the root of logit models for ordinal data (see *e.g.* Agresti (1984, chapter 6). When dealing with a (column) vector of I ordinal variables $X = (X_1, \ldots, X_I)^\top$ with $X_i \in \{1, \ldots, r_i\}$, it is accordingly natural to associate to each coordinate X_i a latent variable ξ_i and a vector of thresholds $-\infty = \alpha_0^{(i)} < \alpha_1^{(i)} < \ldots < \alpha_{r_i}^{(i)} = \infty$ with the interpretation:

$$X_i \leq k \Leftrightarrow \xi_i \leq \alpha_k^{(i)}. \tag{1}$$

This approach has suggested measuring the association among ordinal variables by the association among the corresponding latent variables. Thus, early in the twentieth century (Pearson (1900), Pearson and Pearson (1922), see also the bibliography in Goodman (1981)), a *polychoric correlation*, ρ_P, has been proposed, as the Pearson's correlation (corr) among the corresponding latent variables:

$$\rho_P(X_i, X_j) = \text{corr}(\xi_i, \xi_j). \tag{2}$$

In the eighties, the practitioners of covariance structure models using packages such as LISREL or EQS widened the scope of these models, originally conceived for continuous variables, by using, for ordinal variables, a matrix of polychoric correlations the same way they used Pearson's correlations for continuous variables (see *e.g.* Muthén (1983); Muthén (1984), Jöreskog, Sörbom, du Toit and du Toit (2002) among others).

Recently, the association among continuous random variables has been approached through the decomposition of the joint distribution of a random vector into the set of marginal distributions of each coordinate and a *copula*, *i.e.* a multivariate distribution with margins uniform on $[0, 1]$. The idea is that the copula concentrates the properties of association within the random vector independently of the specification of each coordinate (see *e.g.* Nelsen (1999)).

One object of this paper is to provide an analysis of several identification problems raised by the model leading to the polychoric correlations. We shall show that the use of a copula approach enhances the understanding of the identifying restrictions on the parameters, in a parametric approach, and of the form of the distribution of the latent variables, in a nonparametric approach. We shall next pay particular attention to the role of the normality assumption on the latent variables and the meaning of testing the normality hypothesis. Finally we propose a specification test using the Bayesian encompassing principle in the context of partial observability.

The Bayesian specification test for models involving partial observability has been exposed in Almeida and Mouchart (2005, 2007); for the case of total observability, see Florens, Richard and Rolin (2003) and for the general setup Florens, Mouchart and Rolin (1990) and Florens and Mouchart (1993). The main idea is to compare the parametric specification against a nonparametric alternative by using the inference over the nonparametric parameter in the alternative hypotheses both directly, in the alternative

model, and indirectly, in an extension of the parametric model through the so called *Bayesian Pseudo-True Value*; the parametric model is accordingly a reduction by sufficiency of this extended model. If these two inferences are "near", the simpler (parametric) model is preferred.

Because, in nonparametric models, we are dealing with undominated families of distributions (and therefore with undefined densities) we opt for a σ-algebraic notation, details of which are given in Appendix A.

This paper is organised as follows. Next section provides a general view of the model and an analysis of a first identification problem under arbitrary distribution specification. Section 3 focus attention on identification problems raised by a normality assumption. Section 4 proposes a copula approach of the discretization model and Appendix C provides a short overview on copulas. Section 5 develops a test of the normality assumption based on a Bayesian version of the encompassing principle, reminded briefly in Appendix D. Section 6 analyses, by means of a simulation experiment, some properties of the test just developed and Section 7 proposes an application of the test for a case of meta-analysis of clinical trial data. The paper is concluded by some remarks. Appendix B gives a formal proof of a Theorem stated in Section 3.

2 A General Specification

2.1 Discretization of the latent variable

Let X be a vector of I categorical variables X_i each with range $1, \ldots, r_i$:

$$X = (X_1, \ldots, X_I)^\top \in \prod_{1 \le i \le I} \{1, \ldots, r_i\} \equiv R_X, \quad d = \text{card } R_X = \prod_{1 \le i \le I} r_i,$$

where $R_X \subset \boldsymbol{N}^I$ denotes the range of X. A disjunctive coding is constructed by defining: $Z_{\underline{k}} = \boldsymbol{I}_{\{X=\underline{k}\}}$ for each $\underline{k} = (k_1, \ldots, k_I) \in R_X$; more specifically:

$$Z = (Z_{\underline{k}} : \underline{k} \in R_X) \in \left\{ (z_{\underline{k}} : \underline{k} \in R_X) : z_{\underline{k}} \in \{0,1\}, \sum_{\underline{k} \in R_X} z_{\underline{k}} = 1 \right\} \equiv R_Z, \tag{3}$$

with $R_Z \subset \{0,1\}^d$. This coding is subjected to a generalised Bernoulli distribution:

$$Z \mid \tau \ \sim \ GBe_{(d)}(\tau), \text{ or equivalently: } P(\ Z = z \mid \tau\ ) = \prod_{\underline{k} \in R_X} \tau_{\underline{k}}^{z_{\underline{k}}} \tag{4}$$

with $\tau = (\tau_{\underline{k}} : \underline{k} \in R_X) \in T \subset \mathcal{S}_{d-1}$ and $\tau_{\underline{k}} = E[\ Z_{\underline{k}} \ \mid \tau\ ] \in [0,1]$, where $\mathcal{S}_{d-1}$ is the $(d-1)$-dimensional Simplex, *i.e.* $\mathcal{S}_{d-1} = \{u \in \boldsymbol{R}_+^d : \sum u_i = 1\}$.

Under an *i.i.d.* sampling of (4), a minimum sufficient statistic is the sum of the data:

$$N = \sum_{1 \le i \le n} Z_{(i)} \quad \text{with} \quad N = (N_{\underline{k}}; \underline{k} \in R_X), \quad N_{\underline{k}} \in \{0, \ldots, n\}, \quad \sum_{\underline{k} \in R_X} N_{\underline{k}} = n. \tag{5}$$

The data N of I ordinal variables may be viewed as an I-dimensional contingency table distributed as a multinomial distribution:

$$N \mid \tau \ \sim \ MN_{(d)}(n, \tau). \tag{6}$$

The labeling of possible values of X_i is arbitrary. The only relevant feature is the number r_i, of different labels.

The ordered property of the ordinal variables X_i is recovered by positing a continuous latent random variable ξ_i and an ordered vector of thresholds $\alpha^{(i)} = (\alpha_1^{(i)}, \ldots, \alpha_{r_i - 1}^{(i)})$ (with the convention $\alpha_0^{(i)} \equiv -\infty$ and $\alpha_{r_i}^{(i)} \equiv \infty$) with the interpretation given in (1). Therefore the statistical model, bearing on the manifest vector X, is characterized by the array

$$\alpha = \{\alpha^{(i)} : i = 1, \ldots, I\} \tag{7}$$

and the joint distribution of the latent vector ξ, say ψ. The array α operates a decomposition of $\boldsymbol{R}^I$ into $\prod_{i=1}^I r_i = d$ cubes:

$$c_{\underline{k}} = c_{k_1, \ldots, k_I} = \prod_{i=1}^{I} (\alpha_{k_i - 1}^{(i)}, \alpha_{k_i}^{(i)}] \qquad \underline{k} \in R_X. \tag{8}$$

Note that, $c_{\underline{k}}$ is a function of the parameter α. The statistical model may accordingly be described as follows:

$$P(\ X = \underline{k} \mid \omega\ ) = \psi(c_{\underline{k}}), \text{ with } \omega = (\psi, \alpha) \in \Omega, \tag{9}$$

where ψ is the multivariate probability distribution of the ideally measured variables ξ and α gathers the thresholds as given in (7).

2.2 A first identification problem

The correspondence between the parametrization of the saturated model and that of (9) is given by

$$\psi(c_{\underline{k}}) = \omega_{X\underline{k}}, \tag{10}$$

with $(\omega_{X\underline{k}} : \underline{k} \in R_X) = \omega_X \in \Omega_X = \mathcal{S}_{d-1}$. The parametrization $\omega_X = (\omega_{X\underline{k}} : \underline{k} \in R_X)$ is clearly identified because the $\omega_{X\underline{k}}$'s represent cell probabilities of an I-dimensional contingency table. The correspondence (10) reveals a first identification problem.

Let indeed G be the group of continuous strictly increasing functions $g : \boldsymbol{R} \to \boldsymbol{R}$, $G_{(I)}$ be the group of coordinate-wise transformations defined as: $\underline{g} = (g_1, \ldots, g_I)$ with $g_i \in G$ and define the

corresponding transformation of ω:

$$\begin{aligned} \alpha_{\underline{g}} &= \{g_k(\alpha_k^{(i)}) : k = 1, \dots r_i - 1,\ i = 1, \dots, I\}, \\ \psi_{\underline{g}} &= \psi \circ \underline{g}^{-1}, \end{aligned} \tag{11}$$

then,

$$P(\bullet \mid \psi, \alpha) = P(\bullet \mid \psi_{\underline{g}}, \alpha_{\underline{g}}). \tag{12}$$

From (12), we obtain the following theorem

Theorem 2.1. *In the threshold model* (1)*, let* $\psi_i,\ i = 1, \dots, I$ *be the marginal distributions of* ψ*. Then the* ψ_i*'s are not identified*

As a consequence, a reasonable measure of association for data generated by the discretization model (1) and (9) should not depend on the marginal distributions of the latent variables ξ_i. A natural measure of association might be the Spearman's rho, which is the Pearson correlation among the corresponding latent variables transformed by their own distribution function:

$$\rho_S(X_i, X_j) = \text{corr}(\psi_i(\xi_i), \psi_j(\xi_j)). \tag{13}$$

By the probability integral transform theorem, $\psi_j(\xi_j)$ follows a uniform distribution on $[0, 1]$. Furthermore, in case of a finite sample on observable variables U and V, Corr $(F_U(U), F_V(V))$ may be estimated, non parametrically, as the rank correlation obtained by plugging in (13) the empirical marginal distributions, for details see *e.g.* Kruskal (1958). As the marginal distribution functions of the latent variables, ψ_k, are not identified and can therefore not be meaningfully estimated, an operational alternative to the unestimable $\text{corr}(\psi_i(\xi_i), \psi_j(\xi_j))$ could be $\text{corr}(\xi_i, \xi_j)$. This is precisely the polychoric correlation.

Definition 1. The *matrix of polychoric correlations* for the I-dimensional vector X of ordinal variables is defined as the $I \times I$ correlation matrix of the corresponding continuous latent variables $\{\xi_i : i = 1, \dots, I\}$.

$$R = (\rho_{ij}) \qquad \text{where } \rho_{ij} = \text{corr}(\xi_i, \xi_j). \tag{14}$$

This concept is clearly not invariant under strictly increasing (non linear) transformations of ξ_i, in spite of the identification problem raised in (12) and Theorem 2.1. In next section we consider the identification of R under a normality assumption for the joint distribution of the latent vector $(\xi_1, \dots, \xi_I)$.

3 Identifiability under Normality

The use of polychoric correlations is often grounded on the hypothesis $\xi \sim N(\mu, \Sigma)$, or, in the notation developed in the above section, $\psi = N(\mu, \Sigma)$. In order to preserve normality, we consider the subset of affine transformations included in $G_{(I)}$; these are the transformations such as:

$$x \in \mathbb{R}^I \mapsto g(x) = Bx + c, \tag{15}$$

with $B = Diag\{b_1, \ldots, b_I\}, b_i > 0$ and $c \in \mathbb{R}^I$. This subgroup is denoted by $G_{(LI)}$ where L refers to the "Linear" feature of these transformations and I is the dimension of the space.

Following the same development as above, the array α is transformed into $\alpha_{B,c} = \{b_i a_k^{(i)} + c_i : k = 1, \ldots r_i - 1,\ i = 1, \ldots, I\}$. Then $(N(\mu, \Sigma), \alpha)$ and $(N(B\mu + c, B^\top \Sigma B), \alpha_{B,c})$, are observationally equivalent. This identification problem is typically solved by fixing $b_i = [Var(\xi_i)]^{-\frac{1}{2}}$ and $c_i = -b_i E(\xi_i)$ for $i = 1, \ldots, I$ and therefore fixing the marginals. Under $\xi_i \sim N(0, 1)$, the parameter in (9) is reduced to:

$$\gamma = (R, \alpha) \in \Gamma, \tag{16}$$

where R is a correlation matrix, α is an array giving the thresholds and Γ denotes the parameter space. The model (9) now becomes:

$$P(\ X = \underline{k} \mid \gamma\ ) = \Phi_R(c_{\underline{k}}) = \gamma_{X\underline{k}}, \tag{17}$$

with $(\gamma_{X\underline{k}} = \underline{k} \in R_X) = \gamma_X \in \Gamma_X \subset \Omega_X$, and where Φ_R is the multivariate normal distribution with zero mean, unit variance and R correlation matrix.

The dimension of Γ is equal to:

$$\text{Dim}\ \Gamma = \frac{I(I-1)}{2} + \sum_{1 \le i \le I} (r_i - 1), \tag{18}$$

where $\text{Dim}\ \mathcal{C}$ stands for the dimension of the smallest affine space containing $\mathcal{C}$. If we assume that $\min\{I, r_1, \ldots, r_I\} \ge 2$, we have that:

$$\text{Dim}\ \Gamma = \frac{I(I-1)}{2} + \sum_{1 \le i \le I} (r_i - 1) \le \left(\prod_{1 \le i \le I} r_i \right) - 1 = \text{Dim}\ \Omega_X \tag{19}$$

with the equality if and only if $I = r_1 = r_2 = 2$.

As ω_X, the parameter of the saturated model, is obviously identified and a smooth function of γ, condition (19) says that a necessary condition of identification of γ is always satisfied. A complete characterization of the identifiability of this model is given by the following result.

Theorem 3.1. *Under the normality hypothesis, if* $\min\{I, r_1, \ldots, r_I\} \ge 2$*, then* $\gamma = (R, \alpha)$ *in* (16) *is identified, or equivalently, the mapping* $\gamma \mapsto P(\bullet \mid \gamma)$*, defined in* (17)*, is one-to-one, provided that the polychoric correlations matrix* R *is not singular. In such a case there is a bijection between* Γ *and* Γ_X*.* ■

The proof is given in the Appendix B. We also have the following corollary:

Corollary 3.2. *Under the hypothesis of Theorem 3.1, γ is just identified if $I = r_1 = r_2 = 2$; otherwise, the model is overidentified.* i. e. *The multinomial model (17) is not saturated* ■

Therefore, once $\min\{I, r_1, \ldots, r_I\} > 2$, the normality assumption implies

restrictions on ω_X, that may be used for testing purposes. Two remarks are in order. Firstly a global test of normality becomes rapidly unmanageable when d is increasing. The procedures programmed in several packages, such as LISREL or EQS, only test for bivariate normalities, even though alternative procedures are also available, as for instance in Muthén and Hofacker (1988). Secondly, the null hypothesis actually tested by these procedures contains not only the normal distributions but also the other distributions implying the same restrictions on ω_X. When interpreting the results of such a test, the difficulty is to make these restrictions explicit: equation (19) only gives information on the dimension of the parameter spaces Ω_X and Γ but does not provide an explicit bijection between Γ and a subset of Ω_X representing the null hypothesis. This leaves open the possibility that another parametric specification could imply the same restrictions on Ω_X as the normal specification.

4 A copula approach to the discretization model

A brief summary on copulas is given in Appendix C. In a nonparametric specification, and using Sklar's Theorem, the discretization model (1) and (9) can be parametrized, instead of (9), as follows:

$$\omega = (\{\psi_i : i = 1, \ldots, I\}, C, \alpha), \tag{20}$$

where C represents the unique copula such that:

$$\psi(x_1, \ldots, x_I) = C(\psi_1(x_1), \ldots, \psi_I(x_I)). \tag{21}$$

As the marginal distributions $\{\psi_i : i = 1, \ldots, I\}$ are not identified, the thresholds α are more suitably defined on the support of the marginal distributions of the copula C rather than on the support of $\{\psi_i : i = 1, \ldots, I\}$. More specifically, we reparametrize (20) into:

$$\pi_k^{(i)} = \psi_i(\alpha_k^{(i)}) \in [0, 1], \qquad \alpha_k^{(i)} = {\psi_i}^{-1}(\pi_k^{(i)}). \tag{22}$$

Therefore (1) becomes

$$\forall\, \underline{k} \in R_X \quad \{X \leq \underline{k}\} \;=\; \{\psi_i(\xi_i) \leq \pi_{k_i}^{(i)}, i = 1, \ldots, I\}, \tag{23}$$

and the statistical model (9) is rewritten:

$$P(\, X \leq \underline{k} \mid \omega) = C(\pi_{k_i}^{(i)} : 1 \leq i \leq I) = \omega_{X\underline{k}}, \quad \underline{k} \in R_X. \tag{24}$$

Consequently, the non-identifiability of the margin ψ_i (*i.e.* $X \perp\!\!\!\perp (\psi_i : i = 1, \ldots, I) \mid C, \pi$ in Bayesian

terms) leads to the sufficient reparametrization of (20) as:

$$\omega_C = (C, \pi) \tag{25}$$

where the subscript C stands for "Copula" and (7) is re-parametrized, by means of (22), into:

$$\pi = \{\pi_k^{(i)} : k = 1, \ldots, r_i - 1,\ i = 1, \ldots, I\}. \tag{26}$$

Thus, for all i, $\pi_0^{(i)} = 0$ and $\pi_{r_i}^{(i)} = 1$.

The thresholds $(\pi_k^{(i)} : k = 1, \ldots, r_i)$ may be also viewed as the distribution function of the manifest variable, namely the probability that the ordinal variable takes a value equal or inferior to k, and from (22), are such that the $\alpha_k^{(i)}$'s correspond to the $\pi_k^{(i)}$-quantiles of the unidentified marginal distribution ψ_i. Furthermore, the threshold parameters $\pi_k^{(i)}$ are defined independently of the copulas. Therefore the $\pi_k^{(i)}$'s may be unbiasedly and consistently estimated by the sample proportions. In contrast, the $\alpha_k^{(i)}$'s may be consistently estimated, but not unbiasedly (except in very particular cases), only relatively to an arbitrary specification of ψ_i.

When a family of copulas is finitely parameterized, the necessary condition of identification (19) becomes:

$$\text{Dim } \Theta_{\mathcal{C}} \leq \prod_{i=1}^{I} r_i - 1 - \sum_{i=1}^{I} (r_i - 1), \tag{27}$$

where $\Theta_{\mathcal{C}}$ is the parameter space of a given family $\mathcal{C}$ of I-dimensional copulas defining the model. In the case of a Gaussian copula relative to a multivariate normal distribution with correlation matrix R, say C_R^G, the model (1)-(9) becomes:

$$P(\ X \leq \underline{k}\ \mid \omega_{NCO}) = C_R^G(\pi_{k_i}^{(i)} : 1 \leq i \leq I), \qquad \text{with } \omega_{NCO} = (R, \pi), \tag{28}$$

where ω_{NCO} is the copula parametrization of the model under normality. As the parametrization ω_{NCO} is clearly in bijection with γ, the conditions of application of Theorem 3.1 remain the same.

5 Testing Normality in the Bivariate Case

The meaning of the correlation matrix R is based on a normality assumption for the distribution of the latent vector ξ. This hypothesis is testable as long as it implies restrictions on the *identified* parameters of the saturated statistical model. Even so, a specification test of a normal hypothesis against a non parametric alternative hypothesis is a difficult task because of the non observability of the latent variables. This section illustrates the computations involved by a testing procedure based on an encompassing principle in a Bayesian framework. Some basics of the general procedure for Bayesian encompassing can be found in the Appendix D.

5.1 Bayesian Specifications of the discretization model

Sampling Models After Section 2, let us consider the case $I = 2$, namely two ordinal variables $(X_1, X_2)^\top \in R_X = \{1, \ldots, r_1\} \times \{1, \ldots, r_2\}$ considered as discretization of the latent variables $(\xi_1, \xi_2)^\top \in \boldsymbol{R}^2$; again denote $d = card(R_X) = r_1 \cdot r_2$.

At the level of these latent variables, the sampling parametric null and nonparametric alternative models are specified by:

$$\mathcal{E}^0 : \quad \xi_{(\ell)} \mid \theta \ \sim \ ind.\, N_{(2)} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix} \right], \quad \ell = 1, \ldots, n \tag{29}$$

$$\mathcal{E}^1 : \quad \xi_{(\ell)} \mid \psi \ \sim \ ind.\, \psi, \quad \ell = 1, \ldots, n \tag{30}$$

where $\theta \in [-1, 1]$ is a Euclidean parameter whereas ψ, a probability measure on $\boldsymbol{R}^2$, is a functional parameter.

Suppose now that only an n-sample of a discretization of these latent variables is observed, namely:

$$X_{(\ell)} \doteq f_\alpha(\xi_{(\ell)}) = \text{Disc}(\xi_{(\ell)}, \alpha), \quad \ell = 1, \ldots, n \tag{31}$$

where "Disc" denotes the discretization function according the array of thresholds α as defined in (1). As before, $\gamma = (\theta, \alpha)$ and $\omega = (\psi, \alpha)$.

We again consider the disjunctive coding:

$$Z_{(\ell)} = (Z_{\underline{k},\ell} : \underline{k} \in R_X), \ \ Z_{\underline{k},\ell} = \boldsymbol{I}_{\{X_{(\ell)} = \underline{k}\}} \in \{0, 1\}, \ \sum_{\underline{k} \in R_X} Z_{\underline{k},\ell} = 1. \tag{32}$$

The statistical models can be written as follows:

$$\mathcal{E}^0 : \quad Z_{(\ell)} \mid \theta, \alpha \ \sim \ ind.\, GBe_{(d)}(\gamma_X) \tag{33}$$

$$\mathcal{E}^1 : \quad Z_{(\ell)} \mid \psi, \alpha \ \sim \ ind.\, GBe_{(d)}(\omega_X) \tag{34}$$

where the identified parameters γ_X and ω_X provide the cell probabilities and are defined by:

$$\gamma_X = (\gamma_{X\underline{k}} : \underline{k} \in R_X), \ \gamma_{X\underline{k}} = P^0(Z_{\underline{k},\ell} = 1 \mid \theta, \alpha) = P^0(X_{(\ell)} = \underline{k} \mid \theta, \alpha)$$
$$\omega_X = (\omega_{X\underline{k}} : \underline{k} \in R_X), \ \omega_{X\underline{k}} = P^1(Z_{\underline{k},\ell} = 1 \mid \psi, \alpha) = P^1(X_{(\ell)} = \underline{k} \mid \psi, \alpha).$$

Observe also that:

$$\gamma_X = \Phi_\theta \circ f_\alpha^{-1} \in \Gamma_X \subset \mathcal{S}_{d-1} \tag{35}$$

$$\omega_X = \psi \circ f_\alpha^{-1} \in \Omega_X = \mathcal{S}_{d-1} \tag{36}$$

where Φ_θ denotes the normal bivariate distribution with normal standard marginals and correlation θ.

Again define the sufficient statistic:

$$N = (N_{\underline{k}} : \underline{k} \in R_X), \quad N_{\underline{k}} = \sum_{1 \leq \ell \leq n} Z_{\underline{k},\ell}, \tag{37}$$

i.e. a two-entry contingency table with ordered classes, with sampling distributions:

$$\mathcal{E}^0 : \quad N \mid \theta, \alpha \ \sim \ MN_{(d)}(n, \gamma_X) \tag{38}$$

$$\mathcal{E}^1 : \quad N \mid \psi, \alpha \ \sim \ MN_{(d)}(n, \omega_X). \tag{39}$$

This is therefore a case where the two statistical models are characterised by a same sampling process, namely a multinomial one. The specification test, at the level of manifest variables, becomes accordingly a test on the prior specification for the models reduced to the manifest variables. In other words, the sampling distributions of the structural models generating $(\xi \mid \theta)$ and $(\xi \mid \psi)$, along with their respective prior specifications, are hopefully associated with different prior specifications on the parametrization identified by the manifest variables. In the present case both γ_X and ω_X take values in the $(d-1)$-dimensional Simplex. In a sampling theory approach, the testability of the two hypothesis depends on whether the null model implies testable restriction, *i.e.* whether the parameter space of γ_X is strictly included in the parameter space of ω_X.

Prior specifications It should be mentioned that on the one side the two structural models $\mathcal{E}^0$ and $\mathcal{E}^1$ involve the structural parameters (θ, α) and (ψ, α) respectively and that these parameters, haying a contextually specific meaning, are likely to carry prior information. On the other side, the two statistical models identify the statistical parameters γ_X and ω_X; as these statistical parameters are, more or less, complex functions of structural parameters, the prior information, if substantial, could be deduced from the prior distributions on the structural parameters, whereas if poor, could be specified, with some approximations, directly of the statistical parameters.

In this Bayesian test, we specify, in the null model, a prior distribution on the finite dimensional structural parameters (θ, α) from which we deduce a distribution on γ_X. In the alternative model, a prior distribution is specified directly on the statistical parameter ω_X.

(i) In the null model. The separation between the partial observability process and the structural model suggests to also assume that:

$$\theta \perp\!\!\!\perp \pi; Q^0. \tag{40}$$

The prior distribution for the correlation in the null model is specified as:

$$\frac{\theta + 1}{2} \ \sim \ \text{Beta}(a, b). \tag{41}$$

For α, we first use the reparametrization (22):

$$\pi_k^{(i,0)} = \Phi(\alpha_k^{(i)}), \quad k = 1, \ldots, r_i,\ i = 1, 2 \tag{42}$$

where the second superindex, 0, stands for the null model and Φ is the cdf of the normal standard distribution. The thresholds on $[0, 1]$ are conveniently reparametrized into the Simplex as follows:

$$\delta_k^{(i,0)} = \pi_k^{(i,0)} - \pi_{k-1}^{(i,0)},\ \ \delta^{(i,0)} = (\delta_k^{(i,0)} : k = 1, \ldots, r_i) \in \mathcal{S}_{r_i - 1},\ \ i = 1, 2, \tag{43}$$

and the prior distribution is specified as:

$$\delta^{(1,0)} \perp\!\!\!\perp \delta^{(2,0)}, \text{ or, equivalently: } \pi^{(1,0)} \perp\!\!\!\perp \pi^{(2,0)} \tag{44}$$

$$\delta^{(i,0)} \ \sim\ \text{Di}_{r_i}(n_0^{(i,0)}\, F_0^{(i,0)}), \quad i = 1, 2 \tag{45}$$

with

$$F_0^{(i,0)} = (F_{0k}^{(i,0)} : k = 1, \ldots, r_i), \quad F_{0k}^{(i,0)} = P^0(X_i = k), \quad i = 1, 2. \tag{46}$$

The statistical null model is:

$$\begin{aligned} &N \mid \theta, \pi \ \sim\ MN_d(n, \gamma_X) \qquad \text{where } \gamma_X = h(\theta, \pi) \\ &\gamma_X \ \sim\ (M_\theta^0 \otimes M_\pi^0) \circ h^{-1}, \end{aligned} \tag{47}$$

where the function $h(\cdot, \cdot)$ evaluate the cell probabilities of the contingency tables.

(ii) In the alternative model, the prior distribution on the functional parameter is specified as follows:

$$\omega_X \ \sim\ \text{Di}_d(n_0^{(1)}\, F_0^{(1)}) \tag{48}$$

$$\psi, \alpha \mid \omega_X \ \sim\ \text{An arbitrary distribution} \tag{49}$$

where "Di$_d$" stands for the d-dimensional Dirichlet distribution, $n_0^{(1)} > 0$ and $F_0^{(1)}$ a matrix with the predictive probabilities of each cell in the alternative model, namely:

$$F_0^{(1)} = (F_{0\underline{k}}^{(1)} : \underline{k} \in R_X), \quad F_{0\underline{k}}^{(1)} = P^1(X = \underline{k}). \tag{50}$$

(iii) Compatibility. Two prior specifications, (47) and (48), share in common a same empirical meaning of the thresholds on the margins of the copula. These prior specification should consider explicitly whether some compatibility should be required. More precisely, let us denote similarly to (42):

$$\pi_k^{(i,1)} = \psi_i(\alpha_k^{(i)}), \quad k = 1, \ldots, r_i,\ i = 1, 2 \tag{51}$$

as before, the second superindex, 1, stands for the alternative model and ψ_i is the cdf of the unidentified

marginal distribution of the latent variable ξ_i. With the same reparametrization as in the null model:

$$\delta_k^{(i,1)} = \pi_k^{(i,1)} - \pi_{k-1}^{(i,1)}, \;\; \delta^{(i,1)} = (\delta_k^{(i,1)} = k = 1, \ldots, r_i) \in \mathcal{S}_{r_i - 1}, \;\; i = 1, 2, \tag{52}$$

has a meaning given by $P^1(X_i = k \mid \omega) = \delta_k^{(i,1)}$ and represents the marginal distributions corresponding to the joint distribution ω_X, namely:

$$\delta_{k_i}^{(i,1)} = \sum_{k_{\bar{i}}=1}^{r_{\bar{i}}} \omega_{X\, k_1 k_2}, \quad i = 1, 2 \tag{53}$$

with $\bar{i} = 1$ if $i = 2$ and $\bar{i} = 2$ if $i = 1$. Using properties of the finite dimensional Dirichlet distribution, and denoting the margins of the matrix $F_0^{(1)}$ in (50) by:

$$F_{0 k_i}^{(i,1)} = \sum_{k_{\bar{i}}=1}^{r_{\bar{i}}} F_{0 k_1 k_2}^{(1)}, \quad F_0^{(i,1)} = (F_{0 k_i}^{(i,1)} : k_i = 1, \ldots, r_i), \tag{54}$$

we obtain, from (48),

$$\delta^{(i,1)} \sim \mathrm{Di}_{d_i}(n_0^{(1)}\, F_0^{(i,1)}), \quad i = 1, 2. \tag{55}$$

Because $(\pi^{(i,0)} : i = 1, 2)$ and $(\pi^{(i,1)} : i = 1, 2)$ represent both the marginal distribution functions of the manifest variables X_i, we make two remarks:

1. The condition $F_0^{(i,0)} = F_0^{(i,1)}$, $i = 1, 2$ means the same marginal predictive distributions in both model;
$$F_0^{(i,\cdot)} = P^0(\, X_i = k \,) = P^1(\, X_i = k \,), \quad k = 1, \ldots, r_i, i = 1, 2. \tag{56}$$

2. If additionally $n_0^{(0,0)} = n_0^{(1,0)} = n_0^{(1)}$, the same prior distribution on $\delta^{(i,0)}$ and on $\delta^{(i,1)}$ are specified. Thus, it is plausible to specify a same prior distribution over π in both models, and suppress accordingly the superindex relative to the model. Note however that the structural meaning of the thresholds α is relative to the unidentified marginal distributions of the latent variables.

Remark. In the null model, α and π are bijectively related because the marginal distributions of ξ_1 and ξ_2 are fixed, by identification constraints. In the alternative model only (α, ψ) and (π, ψ) are bijectively related; in other words ω_X is a function of (α, ψ) whereas (π, ψ) and π is a function of ω_X. ■

5.2 Bayesian encompassing specification test

The partial observability process is defined by a function known up to a Euclidean parameter α, or equivalently π. The extension of the null model in order to include the functional parameter of the alternative model is obtained by a conditional probability $M_{\psi|\theta}$ of the extended probability $Q^{0,*}$ of

(ψ, θ, ξ, X) endowed with the following *extended Bayesian Pseudo-True Value* condition, namely:

$$\psi \perp\!\!\!\perp \xi, \pi \mid \theta; Q^{0,*}; \tag{57}$$

more motivation is given in Almeida and Mouchart (2007).

Remember that ω_X has been defined as the minimal sufficient parameter in the alternative statistical model; thus: $\psi, \alpha \perp\!\!\!\perp X \mid \omega_X; Q^1$. Furthermore, in the extended model, we also have the sufficiency of ω_X, namely $\psi, \alpha \perp\!\!\!\perp X \mid \omega_X; Q^{0,*}$ if we assume the condition $\psi \perp\!\!\!\perp \theta \mid \omega_X; Q^{0,*}$ as in Theorem 3 in Almeida and Mouchart (2007). This suggests the plausibility of the condition:

$$M^{0,*}_{\psi|\omega_X} = M^1_{\psi|\omega_X}, \tag{58}$$

and permits to make the comparison based only the identified parameters through:

$$d(N) = d^*(M^{0,*}_{\omega_X|N}, M^1_{\omega_X|N}) \tag{59}$$

where d^* is a distance or divergence.

Because the high dimensionality of ω_X may be at the origin of numerical problems, we choose $\lambda \in \mathbb{R}$, an adequate subparameter of ω_X, which takes into account the properties which we want to put in evidence. Thus, we look for a characteristic of the nonparametric specification ω_X that express how far is ω_X from the closest parameter generated by the parametric specification. Let us write $\gamma_{X\underline{k}}(\theta, \pi^{(1)}, \pi^{(2)})$ for the sampling probability of the cell $\underline{k} \in R_X$ in the parametric model. For a given value of the parameter θ, here the polychoric correlation, and the thresholds defined on the marginals of ξ scaled on the $[0, 1]$-interval, we have that:

$$\gamma_{X\underline{k}}(\theta, \pi^{(1)}, \pi^{(2)}) = P^0(X = \underline{k} \mid \theta, \pi^{(1)}, \pi^{(2)}). \tag{60}$$

The value of θ making $\gamma_{X\underline{k}}(\theta, \pi^{(1)}, \pi^{(2)})$ "closest" to ω_X is obtained through a distance, or a divergence, between two distributions $\omega_X = (\omega_{\underline{k}} : \underline{k} \in R_X)$ and $\gamma_X = (\gamma_{\underline{k}} : \underline{k} \in R_X)$ under the condition of common marginal distribution implied by ω_X, namely $(\tilde{\pi}^{(1)}(\omega_X), \tilde{\pi}^{(2)}(\omega_X))$. Several specific forms of λ may be envisaged, *viz.*

$$\lambda_0(\omega_X) = \min_\theta \left(\sum_{\underline{k} \in R_X} \omega_{X\underline{k}} \log \left(\frac{\omega_{X\underline{k}}}{\gamma_{X\underline{k}}(\theta, \tilde{\pi}^{(1)}(\omega_X), \tilde{\pi}^{(2)}(\omega_X))} \right) \right) \tag{61}$$

$$\lambda_1(\omega_X) = \min_\theta \sum_{\underline{k} \in R_X} \mid \omega_{X\underline{k}} - \gamma_{X\underline{k}}(\theta, \tilde{\pi}^{(1)}(\omega_X), \tilde{\pi}^{(2)}(\omega_X)) \mid \tag{62}$$

$$\lambda_2(\omega_X) = \min_\theta \sum_{\underline{k} \in R_X} (\omega_{X\underline{k}} - \gamma_{X\underline{k}}(\theta, \tilde{\pi}^{(1)}(\omega_X), \tilde{\pi}^{(2)}(\omega_X)))^2 \tag{63}$$

If we choose the Kullback-Leibler divergence, the test statistic is:

$$d(N) = d^*_{KL}(M^{0,*}_{\lambda_r|N}, M^1_{\lambda_r|N}), \quad r = 0, 1 \text{ or, } 2. \tag{64}$$

5.3 Posterior distributions in both models

Let us now discuss how to obtain numerically the two posterior distributions required for evaluation of (64). For the alternative model, we use the specification of the prior distribution of ω_X, as given in (48), and take advantage of its natural conjugate– property, w.r.t. the multinomial sampling:

$$\omega_X \mid N \ \sim \ \text{Di}_d(n_0\, F_0 + N). \tag{65}$$

For the null model, is taken as:

$$M_{\omega_X|\theta,\pi} = E^0[\, M^1_{\omega_X|N} \mid \theta, \pi\,]. \tag{66}$$

Therefore, the posterior distribution in the extended model is given by:

$$M^{*,0}_{\omega_X|N} = E^0[\, M^1_{\omega_X|\tilde{N}} \mid N\,] \tag{67}$$

where $\tilde{N}$ is a virtual sample from Q^0 such that $\tilde{N} \perp\!\!\!\perp N \mid \theta, \pi; Q^0$, for details see Florens, Richard and Rolin (2003).

When generating a sample from $M^{0,*}_{\lambda|N}$, we firstly need to generate a sample of the posterior distribution of the parameter (θ, π) in the null model, then we generate the virtual sample $\tilde{N}$ from the sampling distribution $P^0_{X|\theta,\pi}$. Finally a sample from $M^1_{\omega_X|\tilde{N}}$ in the alternative model is generated and the functional λ is evaluated. The generation of the posterior distribution in the null model is a parametric problem, and it is treated with a MCMC algorithm.

In order to describe an MCMC algorithm in the null model, the lowercase letters are used for densities w.r.t. a σ-finite measure. Let us also denote by $N^{(1)}$ and $N^{(2)}$ the marginal totals of the contingency table; they are equivalent to the empirical distributions of X_1 and X_2 respectively; more explicitly:

$$N^{(1)} = (N_{k_1,\bullet} : k_1 \in \{1, \ldots, r_1\}) \text{ where } N_{k_1,\bullet} = \sum_{1 \le k_2 \le r_2} N_{k_1,k_2} \tag{68}$$

$$N^{(2)} = (N_{\bullet,k_2} : k_2 \in \{1, \ldots, r_2\}) \text{ where } N_{\bullet,k_2} = \sum_{1 \le k_1 \le r_2} N_{k_1,k_2}. \tag{69}$$

In both models the sampling distributions of these marginal totals are multinomial, namely:

$$\mathcal{E}^0 : \quad N^{(i)} \mid \theta, \alpha \ \sim \ MN_{(r_i)}(n, \pi^{(i)}), \qquad i = 1, 2 \tag{70}$$

$$\mathcal{E}^1 : \quad N^{(i)} \mid \psi, \alpha \ \sim \ MN_{(r_i)}(n, \pi^{(i)}), \qquad i = 1, 2. \tag{71}$$

We then build the following accelerated Gibbs sampler:

$$m^0(\pi^{(1)} \mid \pi^{(2)}, \theta, N) \;\propto\; m^0(\pi^{(1)} \mid N^{(1)}) \, \frac{p^0(N \mid \pi^{(1)}, \pi^{(2)}, \theta)}{p^0(N^{(1)} \mid \pi^{(1)})} \tag{72}$$

$$m^0(\pi^{(2)} \mid \pi^{(1)}, \theta, N) \;\propto\; m^0(\pi^{(2)} \mid N^{(2)}) \, \frac{p^0(N \mid \pi^{(1)}, \pi^{(2)}, \theta)}{p^0(N^{(2)} \mid \pi^{(2)})} \tag{73}$$

$$m^0(\theta \mid \pi^{(1)}, \pi^{(2)}, N) \;\propto\; m^0(\theta) \, p^0(N \mid \pi^{(1)}, \pi^{(2)}, \theta \,) \tag{74}$$

Taking into account the prior independence (44) and (40), the acceleration is obtained by substituting in (72) and in (73), the usual terms $m^0(\pi^{(i)})$ by the ratio $m^0(\pi^{(i)} \mid N^{(i)}) * [p^0(N^{(i)} \mid \pi^{(i)})]^{-1}$, $i = 1, 2$. Indeed, in the Gibbs sampler, the posterior distributions $m^0(\pi^{(1)} \mid N^{(1)})$ and $m^0(\pi^{(2)} \mid N^{(2)})$ are easily simulated thanks to the specification (45) and (44), and they are close to the target distributions. For the computational implementation, we use the algorithm developed in Damien, Wakefield and Walker (1999).

Finally, the test statistic is computed by Monte-Carlo integration using the algorithm developed in Wang, Kulkarni and Verdú (2005), and is calibrated against the predictive distribution of the null model, P^0_N, by simulation.

6 Simulations

Description By means of these simulation exercises, we want to check two issues raised by Section 5. Firstly whether the suggested algorithm has a suitable numerical behaviour and secondly whether the proposed test procedure is able to discriminate hypotheses. In all simulations, we consider a same construction of the BPTV, namely that given in (66).

The simulations required in this exercise appear in four different steps:

(i) The generation of simulated contingency tables N is determined by a particular sampling procedure. In this exercise, we consider samplings from the alternative region.

(ii) The evaluation of the statistic $d(N)$ in (64) is obtained as a particular case of the general procedure sketched in Section 5. Note that this step does not depend on the way the data has been simulated. This evaluation requires, as an intermediary step, the simulation of the posterior distributions $M^0_{\omega_X|N}$ and $M^1_{\omega_X|N}$.

(iii) The estimation of the predictive distribution $P^0_{d(N)}$ is obtained through an *iid* simulation of $\tilde{N}$ from P^0_N. Each simulated $\tilde{N}$ is transformed into $d(\tilde{N})$ obtained through the same procedure as (ii).

(iv) The coverage rate, *i.e.* the percentage of cases where the test statistic (64) falls in the 0.05 right tail of the null predictive distribution. This one is used as a measure of the discriminating power of the procedure. The empirical coverage is expected to be higher than 0.05 and to increase with

the sample size. As the test statistic is calibrated by simulation, these coverage rates are random variables.

Let us be more specific on the simulations presented in steps (i) and (ii) above, the last two steps being already explicit enough.

For the first step we generate two scenarios leading to KxK contingency tables, with $K = 3, 4$, issued from repeated generation of ordinal data and for each scenario we consider two possibilities. In the first possibility (A) we consider directly the Bayesian experiment concentrated on (N, ω_X), the manifest variable and on the parametrization identified by the manifest variable, without generating first the latent variable from the alternative model. In the second possibility (B) we first generate a point ψ from the region of the alternative model characterised by a finite mixture of normal distributions from which we simulate a table N and proceeds as in the first possibility. The motivation for examining these two possibilities is to check whether they lead to different discriminating powers of the encompassing test. For each of these two possibilities we repeat two trials, in order to check the stability of the algorithm, and evaluate a coverage rate.

For the second step, we first simulate, for each scenario, and for each sample N generated from step (i), the posterior distributions $M^{0,*}_{\omega_X|N}$ and $M^{1}_{\omega_X|N}$ from which we derive $M^{0,*}_{\lambda_0|N}$ and $M^{1}_{\lambda_0|N}$ with λ_0 as defined in (61) and finally evaluate $d(N) = d^*(M^{0,*}_{\lambda_0|N}, M^{1}_{\lambda_0|N})$.

It should be emphasised that all these simulations, in the null and in the alternative experiments, are based on the predictive distributions of the latent and/or of the manifest variables, given that we simulate first the prior distribution and next the corresponding sampling distribution, under respectively the null and the alternative experiment. Therefore the results of these simulations do not concern the sampling properties but the Bayesian properties of the encompassing test.

In both scenarios, the latent null model has the same structure as in Section 5:

$$\begin{aligned} &\xi_{(\ell)} \mid \theta \ \sim \ ind.\ N_2(0, R),\ \ell = 1, \ldots, n \qquad R = \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix} \\ &\frac{\theta + 1}{2} \ \sim \ \text{Be}(1, 1), \end{aligned} \tag{75}$$

with $n = 20, 50, 100, 200, 500$ and 1000.

Each contingency table is a result from the discretization of a vector of latent variables, the threshold of which may be characterised by a point in a Simplex, as in (43). For both the null and the alternative hypotheses, these points are generated for each margin i, through a Dirichlet distribution

$$\delta^{(i)} \ \sim \ \text{Di}_K(n_0\ P_0), \qquad i = 1, 2, \quad K = 3, 4 \tag{76}$$

with $n_0 = 9, P_0 = (1, 1, 1)^\top/3$ for $K = 3$ and $n_0 = 16, P_0 = (1, 1, 1, 1)^\top/4$ for $K = 4$. The statistical null model is the same as (47).

For each scenario, corresponding to $K = 3$ or 4, the two possibilities (A) and (B) for generating

points from the alternative model specifications are given below.
Alternative model: First possibility (A)

$$\begin{aligned} N \mid \omega_X &\sim MN_d(n, \omega_X), \\ \omega_X &\sim \text{Di}_d(n_0\, Q_0), \end{aligned} \tag{77}$$

with $Q_0 = (1,1,1)\,(1,1,1)^\top/9$ and for the scenario 1, and $Q_0 = (1,1,1,1)\,(1,1,1,1)^\top/16$ and $n_0 = 16$ for the scenario 2. In this possibility, the ordered nature of the margins is not taken into account. As the encompassing test relies on the statistical model obtained after integration of the latent variables, the distribution of the structural parameters (α, ψ) conditionally on the identified parameter (ω_X) is arbitrary and therefore is not specified in this possibility.
Alternative model: Second possibility (B)

In this case, we explore a particular region of the parameter space, namely finite mixtures of normal distributions. These distributions are parametrized, see (78), by a finite number of characteristics denoted as $\tilde{\psi}$:

$$\xi \mid \tilde{\psi} \sim \sum_1^{N_C} q_i N(\mu_i, \Gamma_i) \tag{78}$$

where

$$\begin{aligned} &\tilde{\psi} = (N_C, q, \mu, \rho) \\ &q = (q_1, \ldots, q_{N_C}), \quad \mu = (\mu_1, \ldots, \mu_{N_C}), \quad \rho = (\rho_1, \ldots, \rho_{N_C}) \\ &\Gamma_i = \begin{pmatrix} 1 & \rho_i \\ \rho_i & 1 \end{pmatrix}, \end{aligned}$$

The prior distribution is specified as:

$$\begin{aligned} &N_C - 1 \sim Po(2) \\ &q \mid N_C \sim \text{Di}_{N_C}((1,\ldots,1)) \\ &\underset{1\le i\le N_C}{\perp\!\!\!\perp} \mu_i \mid N_C, q, \quad \mu_i \mid N_C, q \sim N_2(0, I) \\ &\underset{1\le i\le N_C}{\perp\!\!\!\perp} \rho_i \mid N_C, q, \mu, \quad \frac{\rho_i + 1}{2} \mid N_C, q, \mu \sim \text{Be}(1,1) \end{aligned}$$

where $Po(2)$ denotes a Poisson distribution of parameter equal to 2.

The thresholds (in $[0,1]$) defining the discretization are specified as in the null model *i.e.* (76). The test is calibrated using 500 samples simulated from the null model.

Remark that both in the null model and in the two alternatives, the sampling distribution of the corresponding statistical models is the same, namely the multinomial sampling; in the two alternative specifications, the identified parameter is saturated; the difference between the two possibilities is in the prior distributions of the identified parameter even though the support in both cases is the same Simplex

$\mathcal{S}_{d-1}$.

The coverage rates are estimated as follows. For the null hypothesis, we simulate in each case 500 contingency tables $\tilde{N}$ from Q^0 and evaluate each time the test statistic $d(\tilde{N})$ corresponding to (64). The empirical distribution of these simulation provides an estimation of the null predictive distribution $\hat{P}^0_{d(N)}$ and an estimation of the 0.95-quantile $\hat{q}^0_{0.95}$.

Next we simulate 500 contingency tables $\tilde{N}^A$ from one alternative possibility (A) and estimate a coverage through the percentage of cases where the statistic $d(\tilde{N}^A)$ is larger than the threshold $\hat{q}^0_{0.95}$.

For the possibility (B), we retrieve in each trial, the same simulation under the null hypothesis Q^0 already obtained for the possibility (A), but generate twice other 500 contingency tables from the alternative possibility (B). Thus, for given n, the two trials of possibility (A) require 4x500 simulations whereas the two trials of possibility (B) require 2x500 new simulations. Therefore each row of Tables 1 and 2 requires 6x500 simulated contingency tables and evaluations of the statistic $d(N)$.

For each trial, and each possibility, the computation of $d(N)$, see (64), is kept unchanged: only the way N is simulated is modified; moreover, the posterior distribution $M^1_{\omega_X|N}$ has always the same analytical form, as given in (65) and, for evaluating $d(N)$, $M^1_{\lambda_0|N}$ is deduced from $M^1_{\omega_X|N}$.

Results The results are summarised in the Table 1 for the scenario 1 ($K = 3$) and in the Table 2 for the scenario 2 ($K = 4$). For the first possibility (A), with $n = 20$ we observe (first row) a coverage rate 0.406 for the first trial and 0.430 for the second trial. For the second possibility () these values are 0.068 and 0.066.

	First possibility (A)		Second possibility (B)		time
n	trial 1	trial 2	trial 1	trial 2	
20	0.406	0.430	0.068	0.066	$14'$
50	0.676	0.610	0.096	0.106	$17'$
100	0.808	0.826	0.124	0.184	$19'$
200	0.946	0.912	0.226	0.262	$21'$
500	0.978	0.978	0.352	0.400	$15'$
1000	0.988	0.994	0.474	0.428	$29'$

Table 1: Coverage rates for a 3 X 3 table

	First possibility (A)		Second possibility (B)		time
n	trial 1	trial 2	trial 1	trial 2	
20	0.426	0.374	0.092	0.076	$28'$
50	0.776	0.794	0.110	0.122	$32'$
100	0.946	0.944	0.184	0.178	$33'$
200	0.990	0.988	0.258	0.298	$35'$
500	1.000	1.000	0.434	0.406	$42'$
1000	1.000	1.000	0.556	0.544	$54'$

Table 2: Coverage rates for a 4 X 4 table

In both scenarios, and with the two alternative model specifications, the coverage rates are consistently increasing with the sample size. For all sample sizes, we also observe the stability of the coverage rate in the two trials for the two specifications of the alternative model, in both scenarios.

From the two scenarios, we find that the discrimination power increases with the value of K, *i.e.* the refinement of the discretization; this is coherent with the fact that an ordinal variable with more values provides more information. The difference of the coverage rate between the two possibilities, (A) and (B), of the alternative model is interesting. Indeed, let us compare the determinant and the trace of the predictive covariance matrix of the Kx$K-1$ free cells frequencies N_{ij}, for $K = 4$. For the first possibility (A), the predictive covariance matrix, see Bernardo and Smith (1994), is given by:

$$Var^{A}(vec(N)) = n\,\frac{n_0+n}{n_0+1}(diag(vec(Q_0)) - vec(Q_0)vec(Q_0)^{\top}) \tag{79}$$

where vec transforms a matrix into a (column) vector and again $Q_0 = (1,1,1,1)\,(1,1,1,1)^{\top}/16$. For the second possibility (B), we simulate a sample of size 1000 to estimate these matrix. The results are given in Table 3.

n	20	50	100	200	500	1000
Possibility	(A)					
Det	137193.6	$1.1352e+15$	$1.7547e+23$	$6.4506e+31$	$2.8291e+43$	$2.4030e+52$
Trace	37.22426	170.6112	599.7243	2233.456	13338.69	52527.57
Possibility	(B)					
Det	19730400	$4.0336e+17$	$2.2701e+26$	$9.4263e+34$	$3.5540e+46$	$7.6234e+55$
Trace	56.61611	288.5579	1123.910	4303.456	25828.93	107384.3

Table 3: Predictive variances for the alternatives

Considering the determinant and the trace of a covariance matrix as (rough) measures of global dispersion, Table 3 shows that the possibility (B) displays more variation than possibility (A) (in particular for a sample size of 100 on, the determinant of the covariance matrix is 1000 times bigger. This difference is a likely explanation of the higher coverage rate of possibility(A) shown in Tables 1 and 2.

The computation time for simulating the contingency tables, estimating the predictive distribution $\hat{P}^0_{d(N)}$, the corresponding quantile $\hat{q}^0_{0.95}$ and locating, relatively to the threshold $\hat{q}^0_{0.95}$, the statistic test generated under the alternatives is negligible, once the values of $d(N)$'s have been obtained. But the computations of the test statistic $d(N)$ is heavier, even though the posterior distributions $M^{0,*}_{\lambda_0|N}$ and $M^{1}_{\lambda_0|N}$ are based on 100 drawings only. This is due to the fact that an MCMC algorithm is used for each simulated sample. In the last column we give a computation time obtained by averaging over the 6 series of 500 simulations corresponding to each row. The computation time increases with the sample size n, but less than proportionally. Moreover the computation times to the 4x4 tables is roughly the double of the computation time for the 3x3 tables, corresponding to a switch from 9 to 16 cells.

From this simulation exercise, it may be concluded that the proposed procedure is numerically feasible and enjoys of a reasonable discriminatory power but, as to be expected with nonparametric models, requires substantial sample sizes for being reliable.

7 Application

We now examine the working of the test so far developed on real data taken from Vandenhende (2003) and dealing with a meta-analysis of clinic trials on acute migraine. The two ordinal variables are: X_1 = the intensity of pain and X_2 = nausea presence. The observed contingency table, corresponding to $n = 801$, is given in Table 4. In this exercise, we try to evaluate two issues: How quickly the evaluations of interest, p-values and $0,95$-quantiles, tend to stabilise and how quickly the computation time increases with the number of replications.

$X_1 \setminus X_2$	1	2	3	4	Totals
1	136	13	3	2	154
2	174	49	14	2	239
3	121	80	41	3	245
4	37	40	53	33	163
Totals	468	182	111	40	801

Table 4: Data of clinical trials

Let us analyse these data under the same null models as in the simulation exercise, namely (75) and (76) with $K = 4$ and therefore the same statistical models as in (47). For the alternative model we again take a nonparametric specification with Dirichlet prior as in (77) with $n_0 = 16$ and $Q_0 = (1,1,1,1)\,(1,1,1,1)^\top/16$. We want to estimate a Bayesian p-value, *i.e.* the null predictive probability that the statistic $d(N)$, in (64) with $\lambda_r = \lambda_0(\omega_X)$ in (61), takes a value higher than the observed one.

Once the data N, in the form of Table 4, have been obtained, the Bayesian encompassing test consists in evaluating the statistic $d(N)$ and in estimating the null predictive distribution $P^0_{d(N)}$. The evaluation of $d(N)$ is relative to the two Bayesian models characterised by $Q^{0,*}$ and Q^1 whereas the null predictive distribution of $d(N)$ depends on Q^0 only, once the functional form of $d(N)$ has been fixed. In this section we want to evaluate some numerical aspects of the computations, relative to the particular sample N given in Table 4.

The evaluation of the statistic $d(N)$ requires firstly to evaluate the posterior distributions $M^{0,*}_{\lambda_0|N}$ and $M^1_{\lambda_0|N}$, respectively obtained from $M^{0,*}_{\omega_X|N}$ and $M^1_{\omega_X|N}$. The simulations of $M^{0,*}_{\omega_X|N}$ require an MCMC integration with, say B_1, replications whereas for the simulations of $M^1_{\omega_X|N}$, a Dirichlet distribution, we decide to generate $B_2 = 1000$ replications as they do not rise numerical difficulties. Once the ω_X's have been simulated by one of these posterior distributions, they are transformed into $\lambda_0(\omega_X)$ and the distributions $M^{0,*}_{\lambda_0|N}$ and $M^1_{\lambda_0|N}$ are constructed accordingly. Finally the statistic $d(N)$ is computed as a distance between $M^{0,*}_{\lambda_0|N}$ and $M^1_{\lambda_0|N}$ as in (64). In this exercise the numerical stability of the proposed procedure is examined by repeating the computations for two different values of B_1, namely 400 and 800. The estimation of the predictive distribution $P^0_{d(N)}$ raises more substantial problems because we want to derive p-values and 0.95-quantiles, *i.e.* properties of the right tail.

The results, summarised in Table 5, are organised as follows. For each two values of (B_1, B_2) equal to $(400, 1000)$ and $(800, 1000)$, we compute 10 times $d(N)$ and report its average, namely 0.6480 and

0.6329 (first rows). Next we simulate the predictive distribution of $d(N)$ under the null model by using 4000 replications, these 4000 replications are separated into 8x500, 4x1000 and 2x2000 in order to estimate and compare, for each group, the estimations of the 0.95-quantile and the p-value. The average time corresponds to the average time required for simulating B_0 replications of $d(N)$ under the null hypothesis, namely 500, 1000 and 2000. That time is essentially proportional to the simulation sample size B_0.

$\widehat{d(N)} = 0.6480$	$B_1 = 400$				$B_2 = 1000$	
	$B_0 = 500$		$B_0 = 1000$		$B_0 = 2000$	
trial	$\hat{q}^0_{0.95}$	p-value	$\hat{q}^0_{0.95}$	p-value	$\hat{q}^0_{0.95}$	p-value
1	0.8875	0.100	0.9424	0.106	0.9395	0.1035
2	0.9464	0.112	0.9347	0.101	0.9157	0.099
3	0.9494	0.106	0.8938	0.094		
4	0.8853	0.096	0.9166	0.104		
5	0.8839	0.092				
6	0.9328	0.096				
7	0.6888	0.066				
8	1.1378	0.142				
Average	0.9140	0.10125	0.9219	0.10125	0.9276	0.10125
St. Dev.	0.1230	0.0213				
Max	1.1378	0.142				
Min	0.6888	0.066				
Aver. time		$170'$		$339'$		$678'$

$\widehat{d(N)} = 0.6329$	$B_1 = 800$				$B_2 = 1000$	
	$B_0 = 500$		$B_0 = 1000$		$B_0 = 2000$	
trial	$\hat{q}^0_{0.95}$	p-value	$\hat{q}^0_{0.95}$	p-value	$\hat{q}^0_{0.95}$	p-value
1	0.9182	0.104	0.9460	0.107	0.9460	0.102
2	0.9474	0.110	0.9409	0.097	0.9123	0.1045
3	0.9556	0.102	0.8934	0.103		
4	0.8968	0.092	0.9282	0.106		
5	0.9322	0.102				
6	0.8792	0.104				
7	0.8644	0.096				
8	1.0745	0.116				
Average	0.9335	0.10325	0.9271	0.10325	0.9292	0.10325
St. Dev.	0.0653	0.00748				
Max	1.0745	0.116				
Min	0.8792	0.092				
Aver. time		$281'$		$561'$		$1122'$

Table 5: Tail properties of the predictive null distribution of $d(N)$.

The numerical results motivate the following remaks:

(i) The sizable difference the two averages of $d(N)$, namely 0.6480 and 0.6329, suggests that different values of B_1 and probably of B_2 also, possibly introduce different biases in the numerical evaluation of the Kullback-Leibler divergence underlying $d(N)$. When B_1 increases from 400 to 800, we probably have a better evaluation of $d(N)$; we then observe, for $B_0 = 500$, a decrease in the standard deviation, from 0.1230 to 0.0653 and from 0.021 to 0.007.

(ii) When B_0, the number of replications of the simulated samples, increases, the variability of the p-values corresponding to each trial stabilises through an arithmetic mean process, *i.e.* for $(B_1, B_2) = (400, 1000)$, the third column $(0.100, \ldots, 0.142)$ is less stable than the fifth column $(0.106, \ldots, 0.104)$ and less stable than the seventh column $(0.135, 0.099)$. Whereas the averaging of the evaluations of 0.95-quantile is less straightforward because of its non-linearity.

(iii) The computation time is high: The case when $B_1 = 800$ is more or less twice than the time when $B_1 = 400$.

(iv) Finally we do not reject the normality hypothesis of the latent variables at the level of 0.05 in view of the estimated values for the 0.95-quantile (around 0.9 for an $d(N)$ estimated around of 0.6) and for the p-value (around 0.10).

8 Conclusions

Polychoric correlations are frequently used for the analysis of ordinal variables when modeled as a discretization of an underlying latent variable and are typically introduced under a normality hypothesis. This paper has revisited this hypothesis of normality with two objectives: *Firstly to make explicit the object of this hypothesis and secondly to deduce a more precise interpretation of the polychoric correlations*. These two objectives have been achieved by a systematic analysis of identification. One basic object of identification is indeed to check the empirical meaning of the parameters in a structural model: An unidentified parameter may not be consistently estimated and may not be interpreted as the expectation or as the probability limit of some statistic.

The basic idea is to decompose the joint distribution of the latent variable into two variation-free components: The set of its marginals and a copula. The first result, Theorem 2.1, says that the marginals are not identified. Therefore, the normality hypothesis, underlying the polychoric correlations, only bears on the gaussianity of the copula. *As a consequence, testing the normality of the latent variables should be viewed as testing the gaussianity of the copula only*.

As the range of the manifest variable is a finite set, the sampling distribution is multinomial is saturated under a nonparametric alternative with parameter ω_X. Therefore testing the form of the distribution of the latent variables may be achieved only if the null hypothesis implies restrictions on ω_X. When testing the normality, Theorem 3.1 and equation (19) give identification conditions which make the test feasible. It should be stressed that *the implicit null hypothesis is not the hypothesis that the cop-*

ula of ξ is Gaussian but is the hypothesis that the copula of ξ implies the same parametric restriction, on the saturated model, as a Gaussian copula.

As a consequence, *the interpretation of the polychoric correlations within a normality assumption rests on an arbitrary (and untestable) choice of selecting normal marginals and on a testable hypothesis of gaussian copula*. There is a considerable literature, particularly in social sciences, concerning the measures of association among ordinal variables but in the framework of the discretization model (1) and (9), any intrinsic measure of association should be coordinate free (see Kruskal (1958) for similar conclusions) and only based on the copula of the latent variables.

The copula specification (23) *and* (24) *endows the threshold values $\pi_k^{(i)}$ with a simple interpretation of the expected value (or probability limit) of a sample proportion*. Such an interpretation does not imply that the marginal distributions of the latent variables are uniform on $[0, 1]$: it only refers to the always true (for continuous distribution) and therefore unrestrictive fact that the latent continuous variables transformed by their own distribution functions are uniformly distributed on $[0, 1]$. Furthermore, the threshold values $\pi_k^{(i)}$ are easily estimated, unbiasedly and consistently, by the sample proportions without requiring arbitrary specifications of the marginal distributions ψ_i. This is different from the array α where $a_k^{(i)}$ can be interpreted relatively to an arbitrary specification of ψ_i only.

Once the proper role of the normality hypothesis has been recognised one may envisage testing that assumption. In this paper we focus the attention on the *bivariate case* in the framework of a specification test; *i.e.* a test where the *alternative hypothesis is a general nonparametric one*.

As far as a specification test is concerned, a Bayesian encompassing test has been developed for the case of total observability in Florens, Richard and Rolin (2003). For the case of partial observability, Almeida and Mouchart (2005, 2007) consider two different situations, the second one of which encompasses the discretization model. The construction of the test is presented in the Section 5 as an application of the Theorems 1, 2 and 3 in Almeida and Mouchart (2007). *These theorems ensure a suitable meaning of a test statistic based on the manifest variables only*.

Next, we have controlled the operationality of the proposed test. We have noticed that the computation of the test statistic involves estimating the posterior distributions of the identified parameters both under the null and under the alternative model. *This posterior distribution under the null model is simulated by an MCMC algorithm which makes the numerical procedure heavier*. The simulations, in Section 6, suggest that the proposed test is feasible, in terms of computational cost and reasonably reliable in terms of coverage rate. The example of application treated in Section 7 suggests that the proposed test can be operational, even if the computational cost is high and the choice of *the simulation parameters for the calibration has to face a trade-off between computational time and precision*.

A Remarks on notations

Some explanation on the notation used in this paper might be useful. We use the capital letters M, P and Q for probability measures on the parameter space, on the sample space and on the product space respectively. The superindex makes reference to the model considered (when we compare models), and the subindex refers to the relevant random element; when there is no subindex, a complete space is referred to.

In the nonparametric specification we use undominated families of distributions, the existence of densities is not always justified; thus we find preferable to work in terms of probability measures. In particular a measure on the product space is often defined through the Markovian product, denoted $\otimes$. For example, for any A, a measurable set in the θ-space, and any B, a measurable set in the ξ-space, we define:

$$(M_\theta \otimes P_{\xi|\theta})\,(A \times B) = \int_A P_{\xi|\theta}(B)\; dM_\theta, \tag{80}$$

where $P_{\xi|\theta}$ denotes a transition of probability used as conditional probability of $(\xi \mid \theta)$. The Markovian product measure is defined using the extension theorem over the σ-field generated by the cylinders. In the dominated case, this corresponds to the measure defined by the product of densities:

$$q(\theta, \xi) \;=\; m(\theta)\, p(\xi \mid \theta). \tag{81}$$

Finally, we also use the simplified notation: $M_\psi = \int M_{\psi|\xi} dP_\xi = E[\, M_{\psi|\xi}\,]$ as a compact notation for:

$$M_\psi(A) \;=\; \int M_{\psi|\xi}(A) dP_\xi, \tag{82}$$

where A is a measurable set (of the ψ-space), $M_{\psi|\xi}$ is a transition of probability and P_ξ is a probability measure (on the ξ-space).

B Proof of Theorem 3.1

Proof. From (16), we need to prove:

$$P(\bullet \mid R, \alpha) \neq P(\bullet \mid \tilde{R}, \tilde{\alpha}) \quad \Longleftarrow \quad (R, \alpha) \neq (\tilde{R}, \tilde{\alpha}).$$

The proof is split into two parts:
(i) $P(\bullet \mid R, \alpha) \neq P(\bullet \mid \tilde{R}, \tilde{\alpha}) \quad \Longleftarrow \quad \alpha \neq \tilde{\alpha}$.
Heuristically, this implication comes from the fact that the thresholds only depend on the marginal distributions, assumed to be $N(0, 1)$ for each coordinate. A more precise argument runs as follows. The assumption $P(\bullet \mid R, \alpha) = P(\bullet \mid \tilde{R}, \tilde{\alpha})$ is by definition equivalent to:

$$\forall\, \underline{k} \in R_X \qquad P(\, X \leq \underline{k} \mid R, \alpha\,) = P(\, X \leq \underline{k} \mid \tilde{R}, \tilde{\alpha}\,).$$

Let us consider $\alpha_k^{(i)} \neq \tilde{\alpha}_k^{(i)}$ for some $i = 1, \ldots, I \quad k = 1, \ldots, r_i$, define $\bar{i} = \{1, \ldots, I\} \setminus \{i\}$ and let $\underline{k} = (r_1, \ldots, r_{i-1}, k, r_{i+1}, \ldots, r_I)$. We obtain for any value of R and $\tilde{R}$:

$$P(X \leq \underline{k} \mid R, \alpha) = P(\xi_\ell < \infty,\ \ell \in \bar{i}, \xi_i \leq \alpha_k^{(i)} \mid R) = \Phi(\alpha_k^{(i)})$$
$$\neq P(X \leq \underline{k} \mid \tilde{R}, \tilde{\alpha}) = P(\xi_\ell < \infty,\ \ell \in \bar{i}, \xi_i \leq \tilde{\alpha}_k^{(i)} \mid \tilde{R}) = \Phi(\tilde{\alpha}_k^{(i)}).$$

Therefore, by the injectivity of Φ, the standard normal distribution function,α is identified.
(ii) $P(\bullet \mid R, \alpha) \neq P(\bullet \mid \tilde{R}, \alpha) \quad \Longleftarrow \quad R \neq \tilde{R}$.
As $|R| \neq 0$, one has $|\rho_{i_1 i_2}| \neq 1\ \forall\, i_1 < i_2$. Let $\rho_{ij} \neq \tilde{\rho}_{ij}$ with $i < j$, and consider

$$\underline{k} = (r_1, , \ldots, r_{i-1}, k_1, r_{i+1}, \ldots, r_{j-1}, k_2, r_{j+1}, \ldots, r_I);$$

then:

$$P(X \leq \underline{k} \mid R, \alpha) = P(\ \xi_\ell < \infty, \ell \in \{1, \ldots, I\} \setminus \{i, j\}, \xi_i \leq \alpha_{k_1}^{(i)}, \xi_j \leq \alpha_{k_2}^{(j)} \mid R)$$
$$= \Phi_{\rho_{ij}}(\alpha_{k_1}^{(i)}, \alpha_{k_2}^{(j)}),$$

where $\Phi_{\rho_{ij}}$ is the bivariate normal distribution function with zero mean, unit variance and correlation equal to ρ_{ij}. For all $\rho_{ij} \in]-1, 1[$ and $(a, b) \in \boldsymbol{R}^2$; one has:

$$\frac{\partial}{\partial \rho_{ij}} \Phi_{\rho_{ij}}(a, b) = \varphi_{\rho_{ij}}(a, b) > 0$$

where $\varphi_{\rho_{ij}}$ is the bivariate normal density function corresponding to $\Phi_{\rho_{ij}}$. (see Johnson and Kotz (1972) or Tallis (1962)). Thus, for all $(a, b) \in \boldsymbol{R}^2$, the mapping $\rho_{ij} \mapsto \Phi_{\rho_{ij}}(a, b)$ is a strictly increasing continuous function, so it is injective. Then by the injectivity of this function:

$$P(X \leq \underline{k} \mid R, \alpha) \neq P(X \leq \underline{k} \mid \tilde{R}, \alpha)$$

So R is identified. ■

C Copula: a short overview

Once it is recognized that the marginal distributions of the latent variables are not identified in the discretization model, it is natural to decompose the characterization of the multivariate distribution of the latent variables into, on the one hand, the set of its marginal distributions and, on the other hand, a complementary aspect that would be variation-free with respect to the first one and would capture the association between the latent variables. This is exactly the idea of a copula, a precise definition and characterization of which are now given.

Definition 2. A copula C is the distribution function of a multivariate probability distribution with

uniform margins on the unit interval. ■

We need two main results about copulas, the first one is a simple application of the integral transform theorem and the other one provides its reciprocal affirmation. Hereafter we use C or F indifferently as probability measure or as distribution function when there is no ambiguity. The proof of these two theorems, along with a detailed study of copulas may be found in Nelsen (1999).

Theorem C.1. *Let $F_1, \ldots, F_I$ be I univariate distributions and C a I-dimensional copula. Then $F(x_1, \ldots, x_I) = C(F_1(x_1), \ldots, F_I(x_I))$ defines a multivariate distribution F with margins $F_1, \ldots, F_I$.* ■

Theorem C.2. *(Sklar's theorem) Let F be a I-dimensional distribution with continuous marginal distributions $F_1, \ldots, F_I$. Then F has a unique copula representation $F(x_1, \ldots, x_I) = C(F_1(x_1), \ldots, F_K(x_I))$.* ■

The last two theorems give us a bijection between the set of I-dimensional multivariate distributions and the Cartesian product of the set of I univariate distributions and the set of I-dimensional copulas. An important property of copulas is their invariance to the group of strictly increasing coordinate-wise transformations; more precisely the following result is immediate:

Theorem C.3. *If F is a multivariate distribution on $\boldsymbol{R}^I$ and $g \in G_{(I)}$, then F and $F \circ g^{-1}$ have the same copula. More specifically, if*

$$F(x_1, \ldots, x_I) = C(F_1(x_1), \ldots, F_I(x_I)),$$

then for any $g \in G_{(I)}$:

$$F \circ g^{-1}(y_1, \ldots, y_I) = C(F_1 \circ g_1^{-1}(y_1), \ldots, F_I \circ g_I^{-1}(y_I))$$

A copula of a particular interest is the following:

Definition 3. For any correlation matrix Γ the corresponding Gaussian copula is defined by:

$$C_\Gamma^G(u_1, \ldots, u_I) = \Phi_\Gamma(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_I)) \quad (u_1, \ldots, u_I) \in [0,1]^I$$

where Φ is the standard normal univariate distribution function. ■

Thus a multivariate normal distribution may be viewed as a Gaussian copula along with normally distributed margins.

D Bayesian encompassing: A short review

In line with the seminal papers Cox (1962, 1961) about testing non-nested hypotheses, the encompassing principle has been developed for comparing two experiments sharing a same sample space; the main idea is to compare the inference made on the parameter of the second model when using the first one through an embedding mechanism and when using the second model directly, for details see Mizon (1984) and Mizon and Richard (1986). In the Bayesian framework, the generalities of this procedure are exposed in Florens, Mouchart and Rolin (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 distributions 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 obtained by using a distance or discrepancy between probability measures defined on the parametric space. From a Bayesian point of view, the encompassing principle may be viewed a symmetrization on the parametric space of the comparison of experiments, in the sense of Blackwell (1951) and Le Cam (1964) originally introduced for developing the concept of sufficiency on the sample space; for details, see Florens, Mouchart and Rolin (1990, section 3.5).

Let us be more explicit and consider two statistical models on a same sample space, namely $\{P^0_{\xi|\theta} : \theta \in \Theta\}$ and $\{P^1_{\xi|\psi} : \psi \in \Psi\}$, endowed with prior distributions, M^0_θ and M^1_ψ respectively. The first model may be extended so as to incorporate also ψ, the parameter of the second model, by specifying a probability transition to represent a conditional distribution of $\psi \mid \theta, \xi$. A *Bayesian Pseudo-True Value*, BPTV, is such a transition endowed with a further condition of conditional independence, namely $\psi \perp\!\!\!\perp \xi \mid \theta$, also called *BPTV condition*, which gives to θ a property of sufficiency w.r.t. ψ within the extended model. Moreover, it permits us to interpret the first model as the marginalization by sufficiency of the extended model. In the notation of Bayesian experiments as Markovian product, the extended model can be written as:

$$Q^{0,*} = M^0_\theta \otimes P^0_{\xi|\theta} \otimes M_{\psi|\theta} \qquad \text{under the BPTV condition; } \xi \perp\!\!\!\perp \psi \mid \theta; Q^{0,*}. \tag{83}$$

Thus, the comparison of Bayesian inferences is made through comparing the posterior distributions of ψ both in the extended and in the alternative models, namely: $M^{0,*}_{\psi|\xi}$ and $M^1_{\psi|\xi}$. A statistics of test is accordingly a distance or divergence between these two posterior distributions. This tests statistic is calibrated against the predictive measure in the null model, P^0_ξ.

One possibility for the specification of the BPTV suggested in Florens and Mouchart (1993), is the sampling expectation in the first model of the posterior measure in the second one, namely:

$$M_{\psi|\theta} = \int M^1_{\psi|\xi} dP^0_{\xi|\theta} \qquad (= E^0[M^1_{\psi|\xi} \mid \theta\,]) \tag{84}$$

When considering that the BPTV $M_{\psi|\theta}$ provides an interpretation of ψ within the first model, the sug-

gestion (84) is based on the idea that one way of looking at ψ from the first model point of view, is to evaluate the sampling expectation, under Q^0, of the posterior distribution of ψ obtained in its own model, namely the second one. This is in line with Cox (1961)'s paper.

Florens, Richard and Rolin (2003) develops an operational specification test. They use as null hypothesis a parametric specification of the sampling model and, as alternative, a non parametric one; in the alternative model, a Dirichlet process is used as prior distribution. With this specification and using the BPTV specified as in (84) they show that the posterior measure $M^{0,*}_{\psi|\xi}$ is a mixture of Dirichlet processes and they use direct simulation of Dirichlet process, as developed in Rolin (1992) or in Sethuraman (1994), in order to compute the statistics of test and for its calibration against the predictive measure in the null model, P^0_ξ. For a non parametric alternative they suggest to focus the attention on the posterior distributions of a finite dimensional functional of the parameter in the alternative model, say λ.

In Almeida and Mouchart (2005, 2007) the encompassing specification test has been extended to the cases of partial observability; in the first paper when the function defining the partial observability is completely known, and the second paper when that function is known up to a Euclidean parameter only. The discretization model corresponds to the second case.

References

AGRESTI, A. (1984). *Analysis of ordinal categorical data*. New York: Wiley.

ALMEIDA, C. and MOUCHART, M. (2005). Bayesian encompassing test under partial observability. Discussion Paper 05-14, Institut de statistique, Université catholique de Louvain, Louvain-la-Neuve, http://www.stat.ucl.ac.be/ISpub/dp/2005/dp0514.pdf.

ALMEIDA, C. and MOUCHART, M. (2007). Bayesian encompassing specification test under not completely known partial observability. *Bayesian Analysis*, **2**, http://ba.stat.cmu.edu/journal/forthcoming/almeida.pdf.

BERNARDO, J.M. and SMITH, A.F. (1994). *Bayesian theory*. Chichester: Wiley & Sons.

BLACKWELL, D. (1951). Comparison of experiments. In *Proc. 2th Berkeley Sympos. Math. Statist. and Prob.*, 93–102. Berkeley, Calif.: Univ. California Press.

COX, D.R. (1961). Tests of separate families of hypotheses. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob.*, 105–123. Berkeley, Calif.: Univ. California Press.

COX, D.R. (1962). Further results on tests of separate families of hypotheses. *J. Roy. Statist. Soc. Ser. B*, **24**, 406–424.

DAMIEN, P., WAKEFIELD, J. and WALKER, S. (1999). Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. *J. Roy. Statist. Soc. Ser. B*, **61**, 331–344.

FLORENS, J.P. and MOUCHART, M. (1993). Bayesian testing and testing Bayesians, chapter 11. In G. Maddala, C. Rao and H. Vinod, editors, *Handbook in Statistics*, 303–334. Amsterdam: North-Holland.

FLORENS, J.P., MOUCHART, M. and ROLIN, J.M. (1990). *Elements of Bayesian statistics*. New York: Marcel Dekker.

FLORENS, J.P., RICHARD, J.F. and ROLIN, J.M. (2003). Bayesian encompassing specification test of a parametric model against a nonparametric alternative. *Unpublished mimeo*, revised version of Discussion Paper 96-08, Institut de statistique, Université catholique de Louvain, Louvain-la-Neuve, http://www.stat.ucl.ac.be/ISpub/dp/1996/dp9608.ps.

GOODMAN, L.A. (1981). Association models and the bivariate normal for contingency tables with ordered categories. *Biometrika*, **68**, 347–355.

JOHNSON, N.L. and KOTZ, S. (1972). *Distributions in statistics: Continuous multivariate distributions*. New York: Wiley.

JÖRESKOG, K., SÖRBOM, D., DU TOIT, S. and DU TOIT, M. (2002). *LISREL8: New statistical features*. SSI Scientific International Inc.

KRUSKAL, W.H. (1958). Ordinal measures of association. *J. Amer. Statist. Assoc.*, **53**, 814–861.

LE CAM, L. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Statist.*, **35**, 1419–1455.

MIZON, G.E. (1984). The encompassing approach in econometrics. In D.F. Hendry and K.F. Wallis, editors, *Econometrics and Quantitative Economics*, 135 – 172. Oxford: Basil Blackwell.

MIZON, G.E. and RICHARD, J.F. (1986). The encompassing principle and its application to testing nonnested hypotheses. *Econometrica*, **54**, 657–678.

MUTHÉN, B. (1984). A general structural equation model with dichotomous, ordered categorical variables, and continuous latent variables indicators. *Psychometrika*, **49**, 115–132.

MUTHÉN, B. and HOFACKER, C. (1988). Testing the assumptions underlying tetrachoric correlations. *Psychometrika*, **83**, 563 – 578.

MUTHÉN, B. (1983). Latent variable structural equation modeling with categorical data. *J. Econometrics*, **22**, 43–65.

NELSEN, R. (1999). *An introduction to copulas. Properties and applications*. Lecture Notes in Statistics. Springer-Verlag.

PEARSON, K. (1900). Mathematical contribution to the theory of evolution VII: On the correlation of characters not quantitatively measurable. *Phil. Trans. R. Soc. Lond.*, **195A**, 1 – 47.

PEARSON, K. and PEARSON, E.S. (1922). On polychoric coefficients of correlation. *Biometrika*, **52**, 127 – 156.

ROLIN, J.M. (1992). Some useful properties of the Dirichlet process. Discussion Paper 92-02, Institut de statistique, Université catholique de Louvain, Louvain-la-Neuve, http://www.stat.ucl.ac.be/ISpub/dp/1992/dp9202.ps.

SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica*, **4**, 639–650.

TALLIS, G.M. (1962). The maximum likelihood estimation of correlation from contingency tables. *Biometrics*, **18**, 342 – 353.

VANDENHENDE, F. (2003). *Copula models for the analysis of longitudinal ordinal responses in clinical trials on acute migraine*. Ph.D. thesis, Institut de Statistique, Université catholique de Louvain, Louvain-la-Neuve.

WANG, Q., KULKARNI, S.R. and VERDÚ, S. (2005). Divergence estimation of continuous distributions based on data-dependent partitions. *IEEE Transactions on Information Theory*, **51**, 3064 – 3074.