STATISTICA, anno LXXIV, n. 1, 2014

TESTING NORMALITY OF LATENT VARIABLES IN THE POLYCHORIC CORRELATION

Carlos Almeida

School of Mathematics, Yachay Tech, Yachay City of Knowledge, 100119 Urcuqui, Ecuador
Departamento de Ciencias Exactas, Universidad de las Fuerzas Armadas - ESPE, 170501, Sangolqui, Ecuador

Michel Mouchart

Institut de statistique, biostatistique et sciences actuarielles (ISBA), Université catholique de Louvain, Louvain-la-Neuve, Belgium

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 (see *e.g.* Agresti (1984, chapter 6) on logit models for ordinal data). When dealing with a vector of I ordinal variables $X = (X_1, \ldots, X_I)^\top$ with $X_i \in \{1, \ldots, r_i\}$, a natural idea consists in associating to each coordinate X_i a latent variable ξ_i and a vector $\alpha^{(i)}$ of thresholds $-\infty = \alpha_0^{(i)} < \alpha_1^{(i)} < \ldots < \alpha_{r_i}^{(i)} = \infty$ with the interpretation:

$$X_i \leq k \Leftrightarrow \xi_i \leq \alpha_k^{(i)}, \quad k \in \{1, \ldots, r_i\}. \tag{1}$$

Eq. (1) is denoted as "$X_i = \mathrm{Disc}(\xi_i, \alpha^{(i)})$", as function of ξ_i and $\alpha^{(i)}$. Early in the twentieth century Pearson (1900), Pearson and Pearson (1922), see also the bibliography in Goodman (1981), have proposed the *polychoric correlation*, *i.e.* the Pearson's correlation among the corresponding latent variables for measuring association among ordinal variables. 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, polychoric correlations the same way they used Pearson's correlations for continuous variables (see *e.g.* Muthén (1983, 1984), Jöreskog *et al.* (2002) among others). A relevant results is due to Olsson (1979), who describes for the first time a maximum likelihood algorithm for computing the polychoric correlations. Jöreskog (2002) discusses a way for testing the normality based on the chi square distance or on the log-likelihood ratio, but does not make a thorough discussion of the meaning of the underlying normality hypothesis.

This paper provides an analysis of several identification problems raised by the model leading to the polychoric correlations. A first one is the unidentifiability of the marginal distributions; 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. Next we analyse 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 a Bayesian encompassing principle in the context of partial observability.

Bayesian encompassing testing compares the parametric specification against a nonparametric alternative by using the inference on the parameter of interest defined in the nonparametric model both directly, in the alternative model, and indirectly, in an extension of the parametric model through the so called *Bayesian Pseudo-true Value* (BPTV). Under appropriate hypotheses, the parametric model becomes a reduction by sufficiency of this extended model. If these two inferences are "near", the simpler model is preferred. The Bayesian specification test with partial observability has been exposed in Almeida and Mouchart (2005, 2007a) and in Almeida (2007); for the case of total observability, see Florens *et al.* (2003) and for the general setup Florens *et al.* (1990) and Florens and Mouchart (1993).

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 2 proposes a copula approach of the discretization model and reinterprets the normality hypothesis of the latent variables as the Gaussianity of the copula. Section 3 develops a test of the normality assumption (Gaussianity of the copula) based on a Bayesian version of the encompassing principle. Section 4 assesses, by means of a simulation experiment, the computational feasibility, the numerical stability and the discriminating power. In Section 5, an application completes the paper by illustrating the working of the procedure on a meta-analysis of clinical trials on acute migraine. The final section proposes, in the form of conclusions, an evaluation of the actual achievement of the paper.

2. A GENERAL SPECIFICATION

Discretization of the latent variable. Let X be a vector of I categorical variables X_i with range $1,\ldots,r_i$:

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

where $R_X\subset \boldsymbol{N}^I$ stands for the range of X. Denoting $Z_{\underline{k}}=\mathbb{1}_{\{X=\underline{k}\}}$ for each $\underline{k}=(k_1,\ldots,k_I)\in R_X$, the model is written as:

$$P(\,X=\underline{k}\mid\tau\,)=\prod_{\underline{k}\in R_X}\tau_{\underline{k}}^{z_{\underline{k}}} \tag{2}$$

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

For an n-size sample, the data N may be viewed as an I-dimensional contingency table distributed as a multinomial distribution:

$$N\mid\tau\ \sim\ \mathrm{MN}_d(n,\tau). \tag{3}$$

In (3), 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

$$\alpha = \{\alpha^{(i)} : i = 1, \ldots, I\} \tag{4}$$

and the joint distribution of the latent vector ξ, say ψ (we follow the Bayesian tradition to use Greek letters to denote unknown parameters). The array α operates a decomposition of $\boldsymbol{R}^I$ into $\prod_{i=1}^I r_i = d$ cubes:

$$c_{\underline{k}} = c_{k_1,\ldots,k_I} = \prod_{i=1}^{I} (\alpha_{k_i-1}^{(i)}, \alpha_{k_i}^{(i)}] \qquad \underline{k} \in R_X. \tag{5}$$

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

$$P(\, X = \underline{k} \mid \omega \,) = \psi(c_{\underline{k}}), \text{ with } \omega = (\psi, \alpha) \in \Omega, \tag{6}$$

where ψ is the multivariate probability distribution of the latent variables ξ and α gathers the thresholds as given in (4).

The correspondence between the parametrization τ of the statistical model (3) and that of the structural parametrization $\omega = (\psi, \alpha)$ of (6) is given by

$$\psi(c_{\underline{k}}) = \tau_{\underline{k}}, \tag{7}$$

The parametrization τ is clearly identified because the $\tau_{\underline{k}}$'s represent cell probabilities of an I-dimensional contingency table. The statistical model (3) is saturated when $\tau \in T = \mathscr{S}_{d-1}$, *i.e.* when there is no restriction on $\mathscr{S}_{d-1}$. When the structural parametrization $\omega = (\psi, \alpha)$ has a large enough parameter space Ω as typically in a semi-parametric specification, the statistical model, implied by (3), is saturated.

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 $\{\xi_i : i = 1, \ldots, I\}$.*

$$R = (\rho_{ij}) \qquad \textit{where } \rho_{ij} = corr(\xi_i, \xi_j). \tag{8}$$

A first identification problem. The correspondence (7) reveals a first identification problem made explicit in next proposition.

PROPOSITION 2. *In the threshold model* (1), *the marginal distributions ψ_i, $i = 1, \ldots, I$ of ψ are not identified.*

Indeed, Eq. (1) is invariant under any strictly increasing transformation of both sides, transforming eventually the cubes defined in (5), without affecting the probabilities τ_k in (7) .

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)).

The concept of polychoric correlation clearly not invariant under strictly increasing (non linear) transformations of ξ_i, in view of the identification problem raised in Proposition 2.

A copula approach to the discretization model Using Sklar's Theorem (*e.g.* Nelsen (1999)), the discretization model (1) or (6) can be parametrised, as follows:

$$\omega = (\{\psi_i : i = 1,\ldots,I\}, C, \alpha), \tag{9}$$

where C represents the unique copula such that:

$$\psi(x_1,\ldots,x_I) = C(\psi_1(x_1),\ldots,\psi_I(x_I)). \tag{10}$$

with the multivariate probability distribution now taken in the form of its distribution function. As the marginal distributions functions $\{\psi_i : i = 1,\ldots,I\}$ 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 the thresholds into:

$$\pi = \{\pi_k^{(i)} : k = 1,\ldots,r_i - 1,\ i = 1,\ldots,I\}, \quad \pi_k^{(i)} = \psi_i(\alpha_k^{(i)}) \in [0,1], \quad \alpha_k^{(i)} = {\psi_i}^{-1}(\pi_k^{(i)}). \tag{11}$$

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

$$X_i \leq k \Leftrightarrow \psi_i(\xi_i) \leq \pi_k^{(i)}, \quad k \in \{1,\ldots,r_i\}, \tag{12}$$

and the statistical model (6) is rewritten:

$$\forall\, \underline{k} \in R_X, \quad P(\, X \leq \underline{k} \mid \omega) = C(\pi_{k_i}^{(i)} : 1 \leq i \leq I). \tag{13}$$

From (13), we conclude that (C,π) is a sufficient parametrization of the statistical model (*i.e.*, in Bayesian terms: $X \perp\!\!\!\perp \omega \mid C,\pi$). We according reparametrise the model (6) with:

$$\omega_C = (C,\pi). \tag{14}$$

The thresholds $(\pi_k^{(i)} : k = 1,\ldots,r_i)$ may be 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 (11), 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 and 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.

Note that τ, the parameter of the saturated model (3), is obviously identified. When a family of copulas is finitely parameterized, *i.e.* ($\mathscr{C} = \{C_\theta : \theta \in \Theta_{\mathscr{C}}\}$) where $\Theta_{\mathscr{C}}$ is a subspace of dimension $d_{\mathscr{C}}$, a necessary condition of identification is given by:

$$d_{\mathscr{C}} + \sum_{i=1}^{I}(r_i - 1) \leq d, \tag{15}$$

heuristically, the left-hand side gives the number of parameters to be estimated under a parametric specification whereas the right-hand side gives the number of identified parameters in the saturated statistical model. Next proposition provides a more specific identification condition for a parametric model.

PROPOSITION 3. *Let us consider a parametric family of copulas:* $\{C_\theta : \theta \in \Theta\}$, $h(\theta)$, *a subparameter of* (θ, π), *is identified by* X *if the application* $h(\theta) \mapsto C_\theta(\pi)$ *is one-to-one for all* $\pi = (\pi_1, \ldots, \pi_I) \in (0,1]^I$ *such that at least two values* π_i *are not equal to* 1.

Indeed, for an arbitrary $\underline{k} = (k_1, \ldots, k_I) \in R_X$ with at least two components k_i different of r_i, there are in $P(X \leq \underline{k} \mid \theta, \pi) = C_\theta(\pi_{k_1}^{(1)}, \ldots, \pi_{k_I}^{(I)})$, at least two coordinates with $\pi_{k_i}^{(i)} \neq 1$. Therefore $(h(\theta), \pi) \mapsto P(\bullet \mid \theta, \pi)$ is one-to-one.

In the case of a Gaussian copula relative to a multivariate normal distribution with correlation matrix R, say C_R^G, the model (6) becomes:

$$P(X \leq \underline{k} \mid \gamma) = C_R^G(\pi_{k_i}^{(i)} : 1 \leq i \leq I), \qquad \text{with } \gamma = (R, \pi), \tag{16}$$

where γ is the copula parametrization of the model under normality. Furthermore, the condition of the Proposition 3 is satisfied; indeed, for the bi-dimensional case, denoting by Φ_θ (resp. φ_θ) the distribution (resp. density) function of a normal bivariate distribution with normal standard margins and correlation θ, we can use the derivation of the distribution function w.r.t. the parameter in the normal case, as in Johnson and Kotz (1972) or in Tallis (1962) and consequently, if $(\pi_1, \pi_2) \in (0,1)^2$,

$$\frac{\partial C_\theta^G(\pi_1, \pi_2)}{\partial \theta} = \frac{\partial \Phi_\theta(\Phi^{-1}(\pi_1), \Phi^{-1}(\pi_2))}{\partial \theta} = \varphi_\theta(\Phi^{-1}(\pi_1), \Phi^{-1}(\pi_2)) > 0.$$

Therefore, the application $\theta \mapsto C_\theta^G(\pi_1, \pi_2)$ is uniformly monotone, and consequently injective, for all $(\pi_1, \pi_2) \in (0,1)^2$. The identifiability of the parameter specifying the Farlie-Gumbel-Morgensen (FMG) copula provides another example:

$$C_\theta(\pi_1, \pi_2) = \pi_1 \pi_2 (1 + \theta(1 - \pi_1)(1 - \pi_2)),$$

see *e.g.* Nelsen (1999). In fact, $\dfrac{\partial C_\theta(\pi_1, \pi_2)}{\partial \theta} = \pi_1 \pi_2 (1-\pi_1)(1-\pi_2) > 0$, and similarly to the Gaussian case, the application $\theta \mapsto C_\theta(\pi_1, \pi_2)$ is uniformly injective for all $(\pi_1, \pi_2) \in (0,1)^2$.

It should be clear that a nonparametric specification of ψ cannot be completely identified by the statistical model (3). Moreover, a parametric specification of ψ, or of C, may be tested against a nonparametric specification only if the parametric specification implies restriction on $\mathscr{S}_{d-1}$, the saturated parameter space of the statistical model (3); this is, in particular, the case of a strict inequality in (15).

Identifiability under Normality. The use of polychoric correlations is often grounded on the Gaussianity hypothesis $\xi \sim N(\mu, \Sigma)$, or, in the notation developed in the above section, $\psi = N(\mu, \Sigma)$. The fact that the marginal distributions of the latent variables are not identified implies that fixing them to $N(0,1)$ are as arbitrary as fixing them to $U(0,1)$. Therefore, the Normality of the latent variables is observationally equivalent to the Gaussianity of the copula underlying the specification (13).

In this case, the dimension of the Gaussian copulas space is the number of correlations; then, the necessary condition (15) becomes:

$$\frac{I(I-1)}{2} + \sum_{i=1}^{I} r_i - I \leq \prod_{i=1}^{I} r_i,$$

which is verified if $\min\{I, r_i\} \geq 2$. The inequality becomes strict when $\max\{I, r_i\} > 2$. Therefore, if there is more than two variables or at least one variable with more that two classes, the normality assumption implies restrictions on τ, that may be used for testing purposes.

Two remarks are in order. Firstly a global test of normality rapidly becomes computationally demanding 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 τ. When interpreting the results of such a test, the difficulty is to make these restrictions explicit: equation (15) only gives information on the dimension of the parameter space $\Theta_{\mathscr{C}}$. This leaves open the possibility that another parametric specification could imply the same restrictions on $\mathscr{S}_{d-1}$ as the normal specification.

3. TESTING NORMALITY IN THE BIVARIATE CASE

A specification test of a normal hypothesis against a non parametric alternative hypothesis is, in the present situation, 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.

3.1. *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 discretizations of the latent variables $(\xi_1, \xi_2)^\top \in \boldsymbol{R}^2$; again denote $d = \text{card}(R_X) = r_1 \cdot r_2$.

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

$$\mathscr{E}^0: \quad \xi_{(\ell)} \mid \theta \sim ind.\, \mathrm{N}_2\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \theta\\ \theta & 1\end{pmatrix}\right], \quad \ell = 1,\ldots,n \tag{17}$$

$$\mathscr{E}^1: \quad \xi_{(\ell)} \mid \psi \sim ind.\, \psi, \quad \ell = 1,\ldots,n \tag{18}$$

where $\theta \in (-1,1)$ is a Euclidean parameter whereas ψ, a probability measure on $\boldsymbol{R}^2$, is an unrestricted functional parameter. For convenience, the null model $\mathscr{E}^0$ incorporates from the start restricted forms of the marginal distributions, namely $\mathrm{N}(0,1)$, whereas the alternative $\mathscr{E}^1$ leaves the (unidentified) marginal distributions free.

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

$$X_{(\ell)} = \mathrm{Disc}(\xi_{(\ell)}, \alpha) \doteq f_\alpha(\xi_{(\ell)}), \quad \ell = 1,\ldots,n \tag{19}$$

where "Disc" now denotes the discretization function of the vector $\xi_{(\ell)}$ according the array of thresholds α as defined in (1). As before, $\gamma = (\theta, \alpha)$ and $\omega = (\psi, \alpha)$. The data takes the form of a two-entry contingency table N with ordered margins and sampling distributions:

$$\mathscr{E}^0: \quad N \mid \theta, \alpha \sim \mathrm{MN}_d(n, \gamma_X) \tag{20}$$

$$\mathscr{E}^1: \quad N \mid \psi, \alpha \sim \mathrm{MN}_d(n, \omega_X), \tag{21}$$

where $\gamma_X \in \Gamma_X \subset \mathscr{S}_{d-1}$ stands for the cell probabilities under the restrictions implied by the parametric model whereas $\omega_X \in \Omega_X = \mathscr{S}_{d-1}$ stands for the cell probabilities under the saturated nonparametric model. Thus, the model $\mathscr{E}^0$ may be tested against $\mathscr{E}^1$ only if Γ_X is strictly included in Ω_X.

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. Moreover, 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.

Prior specifications The two structural models $\mathscr{E}^0$ and $\mathscr{E}^1$ involve the structural parameters (θ, α) and (ψ, α) respectively and these parameters, having a contextually specific meaning, are likely to carry specific prior information. As the identified parameters γ_X and ω_X could be 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 on these identified 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 identified parameter ω_X.

(i) In the null model. The separation between the partial observability process and the structural model suggests to assume the prior independence of the parameters, namely:

$$\theta \perp\!\!\!\perp \pi; Q^0. \tag{22}$$

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

$$\frac{\theta+1}{2} \sim \text{Beta}(a,b). \tag{23}$$

The thresholds π on $[0,1]$ are conveniently reparametrized into the Simplex as follows:

$$\delta_k^{(i,0)} = \pi_k^{(i,0)} - \pi_{k-1}^{(i,0)},\ \ \delta^{(i,0)} = (\delta_k^{(i,0)} : k = 1,\ldots,r_i) \in \mathscr{S}_{r_i-1},\ \ i = 1,2, \tag{24}$$

and the prior distribution is specified as:

$$\delta^{(1,0)} \perp\!\!\!\perp \delta^{(2,0)}, \text{ or, equivalently: } \pi^{(1,0)} \perp\!\!\!\perp \pi^{(2,0)} \tag{25}$$

$$\delta^{(i,0)} \sim \text{Di}_{r_i}(n_0^{(i,0)} F_0^{(i,0)}), \quad i = 1,2 \tag{26}$$

where $\text{Di}_{r_i}(\cdot)$ stands for a Dirichlet distribution in the $(r_i - 1)$-dimensional simplex, $n_0^{(i,0)} > 0$ and $F_0^{(i,0)}$ is a discrete distribution on r_i points, namely:

$$F_0^{(i,0)} = (F_{0_k}^{(i,0)} : k = 1,\ldots,r_i), \quad F_{0_k}^{(i,0)} = P^0(X_i = k), \quad i = 1,2. \tag{27}$$

The Bayesian statistical null model is:

$$N \mid \theta, \pi \sim \text{MN}_d(n, \gamma_X) \qquad \text{where } \gamma_X = h(\theta,\pi) \tag{28}$$

$$\gamma_X \sim (M_\theta^0 \otimes M_\pi^0) \circ h^{-1}, \tag{29}$$

where the function $h(\cdot,\cdot)$ evaluates the cell probabilities of the contingency table, taking into account the Gaussianity of the latent variables ξ (parametrized by θ).
(ii) In the alternative model, the prior distribution on the functional parameter is:

$$\omega_X \sim \text{Di}_d(n_0^{(1)} F_0^{(1)}) \tag{30}$$

$$\psi, \alpha \mid \omega_X \sim \text{an arbitrary distribution} \tag{31}$$

with $n_0^{(1)} > 0$ and $F_0^{(1)}$ a matrix with the predictive probabilities of each cell in the alternative model, namely:

$$F_0^{(1)} = (F_{0_{\underline{k}}}^{(1)} : \underline{k} \in R_X), \quad F_{0_{\underline{k}}}^{(1)} = P^1(X = \underline{k}). \tag{32}$$

Note that in the null model, the prior distribution on the cell probabilities incorporating the restriction coming from the Gaussianity of the latent variables is derived from the prior specification (22), (23), (25) and (26) through (29), whereas in the alternative model the prior distribution of the cell probabilities is directly specified through (30) with the distribution supporting completely the simplex $\mathscr{S}_{d-1}$.

(iii) Compatibility. Two prior specifications, (29) and (30), 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. With the same reparametrization as in the null model, for $i = 1,2$,

$$\delta_k^{(i,1)} = \pi_k^{(i,1)} - \pi_{k-1}^{(i,1)} = P^1(X_i = k \mid \omega), \;\; \delta^{(i,1)} = (\delta_k^{(i,1)} : k = 1,\ldots,r_i) \in \mathscr{S}_{r_i-1}, \quad (33)$$

represents the marginal distributions corresponding to the joint distribution ω_X, namely:

$$\delta_{k_1}^{(1,1)} = \sum_{k_2=1}^{r_2} \omega_{X\,k_1 k_2}, \qquad \delta_{k_2}^{(2,1)} = \sum_{k_1=1}^{r_1} \omega_{X\,k_1 k_2} \quad (34)$$

Using properties of the finite dimensional Dirichlet distribution, and denoting the margins of the matrix $F_0^{(1)}$ in (32) by:

$$F_{0_{k_1}}^{(1,1)} = \sum_{k_2=1}^{r_2} F_{0_{k_1 k_2}}^{(1)}, \quad F_0^{(1,1)} = (F_{0_{k_1}}^{(1,1)} : k_1 = 1,\ldots,r_1) \quad (35)$$

$$F_{0_{k_2}}^{(2,1)} = \sum_{k_1=1}^{r_1} F_{0_{k_1 k_2}}^{(1)}, \quad F_0^{(2,1)} = (F_{0_{k_2}}^{(2,1)} : k_2 = 1,\ldots,r_2), \quad (36)$$

we obtain, from (30),

$$\delta^{(i,1)} \sim \mathrm{Di}_{r_i}(n_0^{(1)} F_0^{(i,1)}), \quad i = 1,2. \quad (37)$$

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, two features should be pointed out:

1. The condition $F_0^{(i,0)} = F_0^{(i,1)}$, $i = 1,2$ means same marginal predictive distributions in both model;

$$F_0^{(i,\cdot)} = P^0(X_i = k\,) = P^1(X_i = k\,), \quad k = 1,\ldots,r_i, i = 1,2. \quad (38)$$

2. If additionally

$$n_0^{(0,0)} = n_0^{(1,0)} = n_0^{(1)}, \quad (39)$$

same prior distributions on $\delta^{(i,0)}$ and on $\delta^{(i,1)}$ are specified.

We observe that, even under (38) and (39), in the null model specification, the marginal distributions $(\delta^{(1)}, \delta^{(2)})$ are a-priori independent (25), whereas in the alternative model this is not the case because they are derived from a Dirichlet joint distribution (30).

3.2. Bayesian encompassing specification test

The partial observability process, defined in (19) by a function known up to a Euclidean parameter α, or equivalently π, calls for an extension of the Bayesian version of the encompassing principle, described in Almeida and Mouchart (2007a,b) and in Almeida (2007). This extension is based on an extension of the null model $\mathscr{E}^0$, characterised by

a probability measure Q^0 bearing on (ξ, π, θ), into a model $\mathscr{E}^{0,*}$ incorporating ψ, the parameter of $\mathscr{E}^1$, under an *extended Bayesian Pseudo-True Value* condition, namely:

$$\psi \perp\!\!\!\perp \xi, \pi \mid \theta; Q^{0,*}; \tag{40}$$

This extension, obtained by specifying a conditional probability $M_{\psi|\theta}$ in the extended probability $Q^{0,*}$, is such that Q^0 is a marginal probability of $Q^{0,*}$, and provides two posterior distributions of ψ, namely $M^{0,*}_{\psi|X}$ in $\mathscr{E}^{0,*}$ and $M^1_{\psi|X}$ in $\mathscr{E}^1$. The encompassing principle uses a distance, or divergence, $d(X) = d^*(M^{0,*}_{\psi|X}, M^1_{\psi|X})$ as a test statistics to be calibrated against the null predictive distribution P^0_X; more details and motivation are given in Almeida and Mouchart (2007b,a) Almeida (2007), Florens *et al.* (2003), and Florens and Mouchart (1993).

As ω_X has been defined as the minimal sufficient parameter in the alternative statistical model, we have: $\psi, \alpha \perp\!\!\!\perp X \mid \omega_X; Q^1$. Furthermore, in the extended model, we also have, as shown in Theorem 3 in Almeida and Mouchart (2007a), the sufficiency of ω_X; namely, $\psi, \alpha \perp\!\!\!\perp X \mid \omega_X; Q^{0,*}$ if we assume the condition $\psi \perp\!\!\!\perp \theta \mid \omega_X; Q^{0,*}$. The arbitrary character of $M_{\psi|\omega_X}$ given in (31) suggests the plausibility of the condition: $M^{0,*}_{\psi|\omega_X} = M^1_{\psi|\omega_X}$, and permits to make the comparison based only on the two posterior distributions of the identified parameters, namely:

$$d(N) = d^*(M^{0,*}_{\omega_X|N}, M^1_{\omega_X|N}) \tag{41}$$

In line with Florens *et al.* (2003), we choose $\lambda \in \boldsymbol{R}$, an adequate subparameter of ω_X, which takes into account the properties which we want to put forward, here 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 and the thresholds defined on the marginals of ξ scaled on the $[0, 1]$-interval, namely:

$$\gamma_{X\underline{k}}(\theta, \pi^{(1)}, \pi^{(2)}) = P^0(X = \underline{k} \mid \theta, \pi^{(1)}, \pi^{(2)}). \tag{42}$$

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 $(\tilde{\pi}^{(1)}(\omega_X), \tilde{\pi}^{(2)}(\omega_X))$. As several specific forms of λ may be envisaged, we choose the following one:

$$\lambda(\omega_X) = \min_{\theta} \left(\sum_{\underline{k} \in R_X} \omega_{X\underline{k}} \log \left(\frac{\omega_{X\underline{k}}}{\gamma_{X\underline{k}}(\theta, \tilde{\pi}^{(1)}(\omega_X), \tilde{\pi}^{(2)}(\omega_X))} \right) \right), \tag{43}$$

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

$$d(N) = d^*_{KL}(M^{0,*}_{\lambda|N}, M^1_{\lambda|N}). \tag{44}$$

3.3. Posterior distributions in both models

Let us now discuss how to obtain numerically the two posterior distributions required for evaluating (44). For the alternative model, we use the specification of the prior distribution of ω_X, as given in (30), and take advantage of its natural conjugate property w.r.t. the multinomial sampling:

$$\omega_X \mid N \sim \mathrm{Di}_d(n_0 F_0 + N). \tag{45}$$

For the null extended model, we use, see Almeida and Mouchart (2007a,b) for details and motivation, the following Bayesian Pseudo-True Value:

$$M_{\omega_X|\theta,\pi} = E^0[\, M^1_{\omega_X|N} \mid \theta, \pi \,]. \tag{46}$$

Under the specification (46), the posterior distribution of ω_X in the extended model is given by:

$$M^{*,0}_{\omega_X|N} = E^0[\, M^1_{\omega_X|\tilde{N}} \mid N \,] \tag{47}$$

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 *et al.* (2003).

For a sample from $M^{0,*}_{\lambda|N}$, we generate a sample of the posterior distribution of the parameter (θ, π) in the null model, then a 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 treated with an MCMC algorithm.

In order to describe an MCMC algorithm in the null model, the lowercase letters are used for densities w.r.t. a suitable 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 marginal distributions of X_1 and X_2 respectively; more explicitly:

$$N^{(1)} = (N_{k_1,\bullet} : k_1 \in \{1,\ldots,r_1\}) \text{ where } N_{k_1,\bullet} = \sum_{1\le k_2\le r_2} N_{k_1,k_2} \tag{48}$$

$$N^{(2)} = (N_{\bullet,k_2} : k_2 \in \{1,\ldots,r_2\}) \text{ where } N_{\bullet,k_2} = \sum_{1\le k_1\le r_2} N_{k_1,k_2}. \tag{49}$$

Under the compatibility conditions (38) and (39), the sampling distributions of these marginal totals are multinomial and identical in both models, namely:

$$\mathscr{E}^0: \quad N^{(i)} \mid \theta, \alpha \sim \mathrm{MN}_{r_i}(n, \delta^{(i)}), \qquad i = 1,2 \tag{50}$$

$$\mathscr{E}^1: \quad N^{(i)} \mid \psi, \alpha \sim \mathrm{MN}_{r_i}(n, \delta^{(i)}), \qquad i = 1,2. \tag{51}$$

We then build the following accelerated Gibbs sampler:

$$m^0(\pi^{(1)} \mid \pi^{(2)}, \theta, N) \propto m^0(\pi^{(1)} \mid N^{(1)}) \, \frac{p^0(N \mid \pi^{(1)}, \pi^{(2)}, \theta)}{p^0(N^{(1)} \mid \pi^{(1)})} \tag{52}$$

$$m^0(\pi^{(2)} \mid \pi^{(1)}, \theta, N) \propto m^0(\pi^{(2)} \mid N^{(2)}) \, \frac{p^0(N \mid \pi^{(1)}, \pi^{(2)}, \theta)}{p^0(N^{(2)} \mid \pi^{(2)})} \tag{53}$$

$$m^0(\theta \mid \pi^{(1)}, \pi^{(2)}, N) \propto m^0(\theta)\, p^0(N \mid \pi^{(1)}, \pi^{(2)}, \theta\,) \tag{54}$$

For the computational implementation, we use the algorithm developed in Damien *et al.* (1999). Finally, the test statistic is computed by Monte-Carlo integration using the algorithm developed in Wang *et al.* (2005), and is calibrated against the predictive distribution of the null model, P_N^0, by simulation.

4. SIMULATIONS

Description Through these simulation exercises, we check two issues raised by Section 3. Firstly whether the suggested algorithm has a suitable numerical behaviour and secondly whether the proposed test procedure is able to discriminate hypotheses. All simulations use a same construction of the BPTV, namely that given in (46). 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 statistic $d(N)$ in (44) is evaluated as a particular case of the procedure sketched in Section 3. This step does not depend on the way the data has been simulated and requires 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 $d(\tilde{N})$ as in (ii).
(iv) The coverage rate is the percentage of cases where the test statistic (44) falls in the 0.05 right tail of the null predictive distribution. This is used as a measure of the discriminating power of the procedure: tables $\tilde{N}$ simulated from the alternative region are expected to have the statistic $d(\tilde{N})$ falling in the rejection region. More precisely, the empirical coverage should be higher than 0.05 and is expected 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 $r_1 \cdot r_2$ contingency tables, with $r_1 = r_2 = r$ and $r = 3, 4$, 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 numerical stability of the algorithm, and evaluate a coverage rate.

For the second step, we 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 $M^{0,*}_{\lambda|N}$ and $M^1_{\lambda|N}$ with λ as defined in (43) and finally evaluate $d(N) = d^*(M^{0,*}_{\lambda|N}, M^1_{\lambda|N})$.

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 3, namely:

$$\begin{aligned} &\xi_{(\ell)} \mid \theta \sim ind.\, \mathrm{N}_2(0, R),\ \ell = 1, \ldots, n \qquad R = \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix} \\ &\frac{\theta + 1}{2} \sim \mathrm{Beta}(1, 1), \end{aligned} \tag{55}$$

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 (24). For both the null and the alternative hypotheses, these points are generated for each margin i, through a Dirichlet distribution

$$\delta^{(i)} \sim \mathrm{Di}_r(n_0\, P_0), \qquad i = 1, 2, \quad r = 3, 4 \tag{56}$$

with $n_0 = 9, P_0 = (1, 1, 1)^\top/3$ for $r = 3$ and $n_0 = 16, P_0 = (1, 1, 1, 1)^\top/4$ for $r = 4$. The statistical null model is the same as (28).

For each scenario, corresponding to $r = 3$ or 4, the two possibilities (A) and (B) for generating points from the alternative model specifications are given below.

Alternative model: First possibility (A)

$$\begin{aligned} &N \mid \omega_X \sim \mathrm{MN}_d(n, \omega_X), \\ &\omega_X \sim \mathrm{Di}_d(n_0\, Q_0), \end{aligned} \tag{57}$$

with $n_0 = 9$ and $Q_0 = (1, 1, 1)(1, 1, 1)^\top/9$ for the scenario 1, and $n_0 = 16$ and $Q_0 = (1, 1, 1, 1)(1, 1, 1, 1)^\top/16$ 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 (58), by a finite number of characteristics denoted as $\tilde{\psi}$:

$$\xi \mid \tilde{\psi} \sim \sum_{1}^{N_C} q_i \mathrm{N}(\mu_i, R_i) \tag{58}$$

with $\tilde{\psi} = (N_C, q, \mu, \rho)$, $q = (q_1, \ldots, q_{N_C})$, $\mu = (\mu_1, \ldots, \mu_{N_C})$, $\rho = (\rho_1, \ldots, \rho_{N_C})$ and

$R_i = \begin{pmatrix} 1 & \rho_i \\ \rho_i & 1 \end{pmatrix}$. The prior distribution is specified as:

$$
\begin{aligned}
& N_C - 1 \sim \text{Po}(2) \\
& q \mid N_C \sim \text{Di}_{N_C}((1,\ldots,1)) \\
& \perp\!\!\!\perp_{1\le i\le N_C} \mu_i \mid N_C, q, \quad \mu_i \mid N_C, q \sim \text{N}_2(0, I) \\
& \perp\!\!\!\perp_{1\le i\le N_C} \rho_i \mid N_C, q, \mu, \quad \frac{\rho_i + 1}{2} \mid N_C, q, \mu \sim \text{Beta}(1,1)
\end{aligned}
$$

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.* (56). The test is calibrated using 500 samples simulated from the null model.

Remember 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 $\mathscr{S}_{d-1}$ without restrictions.

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 (44). The empirical distribution of these simulations 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.

For each trial, and each possibility, the computation of $d(N)$, see (44), is kept unchanged: only the way N is simulated is modified. The posterior distribution $M^1_{\omega_X|N}$ has always the same analytical form, as given in (45) and, for evaluating $d(N)$, $M^1_{\lambda|N}$ is deduced from $M^1_{\omega_X|N}$.

Results. The results are summarised in the Table 1 for the scenario 1 ($r = 3$) and in the Table 2 for the scenario 2 ($r = 4$). Note that each row of Tables 1 and 2 requires 6x500 simulated contingency tables and evaluations of the statistic $d(N)$.

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 (B) these values are 0.068 and 0.066.

These tables suggest the following remarks:
(i)] In both scenarios, and with the two alternative model specifications, the *coverage rates are consistently increasing* with the sample size.

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	0.406	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 2
Coverage rates for a 4 X 4 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	$54'$

(ii) 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.
(iii) From the two scenarios, we find that the *discrimination power* increases with the value of r, *i.e.* the refinement of the discretization; this is coherent with the fact that an ordinal variable with more values provides more information.
(iv) 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 r^2-1 free cell frequencies N_{ij}, for $r=4$. For the first possibility (A), the predictive covariance matrix, see Bernardo and Smith (1994), is given by:

$$Var^A(vec(N)) = n\,\frac{n_0+n}{n_0+1}(diag(vec(Q_0)) - vec(Q_0)vec(Q_0)^\top) \tag{59}$$

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

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, the determinant of the covariance matrix in possibility (B) is 1000 times bigger than in possibility (A). This difference is a likely explanation of the higher coverage rate of possibility(A) shown in Tables 1 and 2. In other words, simulation of possibilities (A) and (B) correspond to two dif-

TABLE 3
Predictive variances for the alternatives.

	First possibility (A)		Second possibility (B)	
n	Det	Trace	Det	Trace
20	$1.3719e+5$	37.22	$1.9730e+7$	56.62
50	$1.1352e+15$	170.61	$4.0336e+17$	288.56
100	$1.7547e+23$	599.72	$2.2701e+26$	1123.91
200	$6.4506e+31$	2233.45	$9.4263e+34$	4303.46
500	$2.8291e+43$	13338.69	$3.5540e+46$	25828.93
1000	$2.4030e+52$	52527.57	$7.6234e+55$	107384.30

ferent prior specifications associated with higher predictive variability for (B) than for (A).

(v) The computation time for simulating the contingency tables, estimating the predictive distribution $\hat{P}^0_{d(N)}$ with 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 $M^{0,*}_{\lambda|N}$ and $M^1_{\lambda|N}$ 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 of Tables 1 and 2 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.

5. 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 clinical 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. This exercise aims at evaluating 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.

Let us analyse these data under the same null models as in the simulation exercise, namely (55) and (56) with $r = 4$ and therefore the same statistical models as in (28). For the alternative model we again take a Dirichlet prior as in (57) with $n_0 = 16$ and $Q_0 = (1,1,1,1)(1,1,1,1)^\top/16$, and estimate a Bayesian p-value, *i.e.* the null predictive probability that the statistic $d(N)$, in (44), 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

TABLE 4
Data of clinical trials.

$X_1 \setminus X_2$	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

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 $M^{0,*}_{\lambda|N}$ and $M^1_{\lambda|N}$, 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 $\lambda(\omega_X)$ and the distributions $M^{0,*}_{\lambda|N}$ and $M^1_{\lambda|N}$ are constructed accordingly. Finally the statistic $d(N)$ is computed as a divergence between $M^{0,*}_{\lambda|N}$ and $M^1_{\lambda|N}$ as in (44). 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 and Table 6, 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, where B_0 is taken to be 500, 1000 and 2000. That time is essentially proportional to the simulation sample size B_0.

The numerical results motivate the following remarks:

(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 indeed observe, for $B_0 = 500$, a decrease in the standard deviation, from 0.1230 to 0.0653 for the quantiles and from 0.021 to 0.007 for the p-value.

(ii) When B_0, the number of replications of the simulated samples, increases, the variabil-

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

$\widehat{d(N)} = 0.6480$	$B_1 = 400$			$B_2 = 1000$		
	$B_0 = 500$		$B_0 = 1000$		$B_0 = 2000$	
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		$170'$		$339'$		$678'$

TABLE 6
Tail properties of the predictive null distribution of $d(N)$.

$\widehat{d(N)} = 0.6329$	$B_1 = 800$			$B_2 = 1000$		
	$B_0 = 500$		$B_0 = 1000$		$B_0 = 2000$	
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		$281'$		$561'$		$1122'$

ity 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 that of 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 a $d(N)$ estimated around of 0.6) and for the p-value (around 0.10).

6. Conclusions

In an unpublished paper, Florens *et al.* (2003) have sketched a procedure to operationalise a Bayesian test of a parametric against a non-parametric alternative and have shown that a Bayesian version of the encompassing principle provides a promising avenue. In this paper we have considered a considerably more complex ingredient, namely a problem of partial observability. Taking a test of normality of latent variables in the field of polychoric correlation as a case of study, we developed a completely operational extension of the Florens *et al.* (2003) proposal. Numerical and statistical properties of the proposed procedure have been explored through a simulation experiment and the operational feasibility has been shown through an application in the field of clinical trials.

Polychoric correlations are frequently used for the analysis of ordinal variables modelled as a discretization of an underlying latent variable, say ξ, 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 an analysis of identification.

The joint distribution of the latent variable is decomposed into two variation-free components: the set of its marginals and a copula. Proposition 2 says that the marginals are not identified. *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 which 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 Gaussianity of the copula, Proposition 3 ensures the identification of the natural parametrization and makes the test feasible. It should be stressed that *the implicit null hypothesis is not the hypothesis that the copula of ξ is Gaussian but is the hypothesis that the copula of ξ implies the same parametric restriction, on ω_X, 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*. Note that, under Gaussianity of the copula, uncorrelated variables means independent ones, and independence is invariant under coordinate-wise scale choice.

The copula specification (12) *and* (13) *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*. A Bayesian encompassing test has been developed for the case of total observability in Florens *et al.* (2003). For the case of partial observability, Almeida and Mouchart (2005, 2007a) consider two different situations, the second one of which encompasses the discretization model. The construction of the test is presented in the Section 3 as an application of the Theorems 1, 2 and 3 in Almeida and Mouchart (2007a). *These theorems ensure a suitable meaning of a test statistic based on the manifest variables only*.

Finally, we have controlled the operationality of the proposed test.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 heavy*. The simulations, in Section 4, suggest that the proposed test is feasible, in terms of computational cost and reasonably reliable in terms of coverage rate. The application treated in Section 5 suggests that the proposed test is operational, even if the computational cost is high and the choice of *the simulation parameters for the calibration has to face a natural trade-off between computational time and precision*.

ACKNOWLEDGEMENTS

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. The first author has been partially suported by the Project Prometeo of the Secretaría de Educación Superior de Ciencia, Tecnología e Innovación de la República del Ecuador.

REFERENCES

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

C. ALMEIDA (2007). *Testing specifications on partial observability models: a Bayesian encompassing approach*. Ph.D. thesis, Institut de statistique, Université catholique de Louvain, Louvain-la-Neuve.

C. ALMEIDA, M. MOUCHART (2005). *Bayesian encompassing test under partial observability*. Discussion Paper 05-14, Institut de statistique, Université catholique de Lou-

vain, Louvain-la-Neuve. http://www.stat.ucl.ac.be/ISpub/dp/2005/dp0514.pdf http://www.stat.ucl.ac.be/ISpub/dp/2005/dp0514.pdf, submitted.

C. ALMEIDA, M. MOUCHART (2007a). *Bayesian encompassing specification test under not completely known partial observability*. Bayesian Analysis, 2, pp. 303–318. http://ba.stat.cmu.edu/journal/2007/vol02/issue02/almeida.pdf http://ba.stat.cmu.edu/journal/2007/vol02/issue02/almeida.pdf.

C. ALMEIDA, M. MOUCHART (2007b). *Bayesian encompassing specification test under not completely known partial observability (summary of Almeida and Mouchart (2007a))*. In J.-M. BERNARDO, M.-J. BAYARRI, J.-O. BERGER, A.-P. DAWID, D. HECKERMAN, A.-F.-M. SMITH, M. WEST (eds.), *Bayesian Statistics 8*. Oxford:Oxford University Press, pp. 567–572.

J.-M. BERNARDO, A. F. M. SMITH (1994). *Bayesian theory*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Ltd., Chichester.

P. DAMIEN, J. WAKEFIELD, S. WALKER (1999). *Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables*. J. Roy. Statist. Soc. Ser. B, 61, no. 2, pp. 331–344.

J.-P. FLORENS, M. MOUCHART (1993). *Bayesian testing and testing Bayesians*. In *Econometrics*, North-Holland, Amsterdam, vol. 11 of *Handbook of Statist.*, pp. 303–334.

J.-P. FLORENS, M. MOUCHART, J.-M. ROLIN (1990). *Elements of Bayesian statistics*, vol. 134 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker Inc., New York.

J.-P. FLORENS, J.-F. RICHARD, J.-M. ROLIN (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 http://www.stat.ucl.ac.be/ISpub/dp/1996/dp9608.ps.

L. A. GOODMAN (1981). *Association models and the bivariate normal for contingency tables with ordered categories*. Biometrika, 68, no. 2, pp. 347–355.

N. L. JOHNSON, S. KOTZ (1972). *Distributions in statistics: continuous multivariate distributions*. John Wiley & Sons Inc., New York. Wiley Series in Probability and Mathematical Statistics.

K. JÖRESKOG (2002). *Structural equation modeling with ordinal variables sudsing LISREL*. Tech. rep., technical report. http://www.ssicentral.com/lisrel/techdocs/ordinal.pdf http://www.ssicentral.com/lisrel/techdocs/ordinal.pdf.

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

B. MUTHÉN (1983). *Latent variable structural equation modeling with categorical data*. J. Econometrics, 22, no. 1-2, pp. 43–65.

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

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

R. B. NELSEN (1999). *An introduction to copulas*, vol. 139 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

U. OLSSON (1979). *Maximul likelihood estmation of the polychoric correlation coefficient*. Psychometrika, 44, pp. 443 – 460.

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

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

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

F. VANDENHENDE (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.

Q. WANG, S. R. KULKARNI, S. VERDÚ (2005). *Divergence estimation of continuous distributions based on data-dependent partitions*. IEEE Trans. Inform. Theory, 51, no. 9, pp. 3064–3074.

SUMMARY

Testing normality of latent variables in the polychoric correlation

This paper explores the feasibility of simultaneously facing three sources of complexity in Bayesian testing, namely *(i)* testing a parametric against a non-parametric alternative *(ii)* adjusting for partial observability *(iii)* developing a test under a Bayesian encompassing principle. Testing the normality of latent variables in the polychoric correlation model is taken as a case study. This paper starts from the specification of the model defining the polychoric correlation in the framework of manifest ordinal variables viewed as discretizations of underlying latent variables. Taking advantage of the fact that in this model, the marginal distributions of the latent variables are not identified, we use the approach of copula. Some identification issues are analysed. Next, 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. The computational feasibility, the numerical stability and the discriminating power of the procedure are checked through a simulation experiment. An application completes the paper by illustrating the working of the procedure

on a meta-analysis of clinical trials on acute migraine. The final section proposes, in the form of conclusions, an evaluation of the actual achievements of the paper.

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