Gaussian copula-generated distributions for correlated discrete and continuous variables

A. R. de LEON[*], B. WU and N. WITHANAGE

Department of Mathematics and Statistics
University of Calgary
Calgary, Alberta T2N1N4, Canada

SUMMARY

Accounting for association between mixed discrete and continuous random variables is an important statistical problem in many practical applications. In this paper, we use copulas to construct a class of joint distributions of mixed discrete and continuous random variables. In particular, we employ the Gaussian copula to generate meta distributions for mixed discrete and continuous random variables. Examples of meta-Gaussian mixed-variable distributions are the robit-normal and probit-normal-exponential distributions, the first for modelling the distribution of mixed binary-continuous data and the second for a mixture of continuous, binary and trichotomous variables. The new class of distributions is general enough to include many models currently available. We study properties of the distributions and outline likelihood estimation; a small simulation study is used to investigate the finite-sample properties of estimates. Finally, we present an application to discriminant analysis of mixed binary and continuous data.

Keywords: conditional grouped continuous model; Gaussian copula; general location model; logit model; normal distribution; probit model; t-distribution.

1. Introduction

Many statistical applications involve the collection and analysis of multivariate data comprising a mixture of discrete and continuous variables. Examples can be found in medicine (where continuous laboratory measurements may be included with such variables as presence or absence of a certain symptom for each patient), in health studies (where data may involve a patient's choice of health care unit, his state of health, his global quality of life, along with a number of quantitative health-related variables), and in many other fields. Multivariate modeling of such data often leads to complications in practice due to a relative lack of standard models.

An obvious, but often inefficient, approach to handling mixed data is to convert one type of variable to another, and then to employ appropriate methods. Although this approach is simple enough and may work in practice, the crude coding of qualitative variables or categorization of

[*]email: adeleon@math.ucalgary.ca

quantitative variables makes it conceptually unattractive and unsatisfactory in many applications. Instead, a model-based alternative is possible by directly or indirectly specifying a joint distribution.

Factorization models directly specify the joint distribution as the product of a conditional distribution of one set of variables and a marginal distribution of the other. General location models (GLOMs) (Olkin and Tate, 1961) are based on conditional Gaussian distributions and have received much attention in the literature. They assume conditional normality of continuous components and an arbitrary distribution for discrete components. A reverse factorization entails specifying a latent continuous multivariate distribution from which discrete variables are derived by forming discrete classes. Probit-style models for categorical data generated from a multivariate normal distribution of underlying latent variables yield so-called conditional grouped continuous models (CGCMs) (de Leon, 2005; Anderson and Pemberton, 1985). A number of refinements and extensions of these models have since been studied by several authors; e.g., see de Leon and Carrière (2007) for the general mixed-data model, a hybrid of GLOM and CGCM which provides a unified treatment of these models.

While factorization provides a general route for directly specifying mixed-variable distributions, it is not without its shortcomings. Factorization models use a structural approach in classifying variables as continuous or discrete to decide the direction of conditioning. A hierarchy in the data is thus induced, with conditioning components treated as intermediate variables, and conditioned components as primary responses. As such, factorization models are not invariant to the direction of conditioning taken and the factorization adopted. Variations of GLOM and CGCM have been studied in various applications resulting in models that are not comparable, as parameters have different interpretations depending on the factorization used. It is thus possible for different factorization models to yield very different inferences, especially of associations.

Indirect approaches to specifying mixed-variable joint distributions have also been studied. One approach that has found widespread adoption in practice introduces shared or correlated random effects to incorporate correlations between variables in the resulting joint model. The basic idea in this approach is to use random effects to build in correlation between mixed variables. The approach does not resort to factorization, and thus treats variables symmetrically. Its hierarchical structure allows for considerable flexibility in accounting for different measurement levels, delineation of various associations, incorporation of covariate effects, and extension to longitudinal and clustered data settings. However, random effects models also have their drawbacks. For one, correlations may be restricted to lie within artificially narrow ranges; for another, computational difficulties may arise in high-dimensional problems (see de Leon and Carrière Chough, 2010, and McCulloch, 2007, for examples).

A recent alternative strategy involves the use of copulas, as discussed by Song (2007), among others. The approach embeds absolutely continuous univariate marginal distribution functions $F_{X_1}(\cdot)$, $\cdots, F_{X_P}(\cdot)$, into their corresponding P-dimensional distribution function $F_{X_1,\cdots,X_p}(\cdot)$ via a copula $C(\cdot)$ as

$$F_{X_1,\cdots,X_P}(x_1,\cdots,x_P) \quad = \quad C\left(F_{X_1}(x_1),\cdots,F_{X_P}(x_P)\right). \tag{1}$$

The meta distribution $F_{X_1,\cdots,X_P}(\cdot)$ (McNeil et al., 2005) is thus specified via its margins and a copula that "glues" them together. In parametric contexts, the margins need not come from the same parametric family, allowing researchers great flexibility in modelling variables of different types. The copula accounts for "dependence" between variables in a way that is separate from their marginal specifications.

The choice of an appropriate copula with which to "couple" marginal distributions depends on the suitability of the copula's dependence parameter for describing the data's dependence structure. Gaussian copulas are an important family which has been used in a variety of applications (e.g., Song, 2007). The P-dimensional Gaussian copula is defined as

$$C(u_1,\cdots,u_P) \quad = \quad \Phi_P\left(\Phi^{-1}(u_1),\cdots,\Phi^{-1}(u_P);\mathbf{R}\right), \tag{2}$$

for $u_1,\cdots,u_P \in [0,1]$, where $\Phi^{-1}(\cdot)$ is the inverse function of the standard normal distribution function $\Phi(\cdot)$, and $\Phi_P(\cdot;\mathbf{R})$ is the P-dimensional standard multivariate normal distribution function (i.e., zero means and unit variances), with correlation matrix $\mathbf{R}$. For continuous random variables $X_1 \sim F_{X_1}(\cdot), \cdots, X_P \sim F_{X_P}(\cdot)$ whose joint distribution function $F_{X_1,\cdots,X_P}(\cdot)$ is specified by Gaussian copula (2), we get

$$F_{X_1,\cdots,X_P}(x_1,\cdots,x_P) = \Phi_P\left(\Phi^{-1}(u_1),\cdots,\Phi^{-1}(u_P);\mathbf{R}\right), \qquad (3)$$

where $u_1 = F_{X_1}(x_1), \cdots, u_p = F_{X_P}(x_P)$ with $U_1 = F_{X_1}(X_1) \sim$ uniform$[0,1], \cdots$, $U_P = F_{X_P}(X_P) \sim$ uniform$[0,1]$ the probability integral transforms (PIT), and $\mathbf{R}$ is the correlation matrix of so-called normal scores $\Phi^{-1}(U_1), \cdots, \Phi^{-1}(U_P)$. The flexibility and analytical tractability of Gaussian copulas make them a handy tool in many applications. Their popularity is due to the fact that they describe dependence between variables in much the same way that Gaussian distributions do.

While not new, applications of copulas to discrete data (e.g., Nikoloulopoulos and Karlis, 2010, 2009, 2008; Song et al., 2009; Zimmer and Trivedi, 2006; Trégouët et al., 2004; Meester and MacKay, 1994) have only recently been elucidated and clarified (Genest and Nešlehová, 2007). Embrechts (2009) alludes to problems arising from the use of copulas to construct discrete distributions, and cautions that "everything that can go wrong, will go wrong." As Genest and Nešlehová (2007) show, a number of complications arise from the direct application of copula models to discrete data. One such complication, which may be merely a theoretical issue, concerns the failure of the copula to uniquely determine the distribution. Another, but a more practical one, involves the interpretability of the dependence parameters.

This paper considers a class of mixed-variable distributions generated by copulas in general, and Gaussian copulas, in particular. A latent-variable approach is adopted to sidestep complications of direct application of copula models to discrete variables. The proposed distributions treat mixed variables symmetrically and do not resort to factorization. The class is general enough to include previous models studied in the literature as special cases. We discuss in detail its properties (e.g., moments, correlations) and investigate likelihood estimation for the model. An application to discriminant analysis is used to illustrate the distributions.

2. Mixed-data meta distributions

In this section, we construct the copula-generated model for the joint distribution of mixed variables, first using a general copula function and then using the Gaussian copula.

Suppose we have discrete variables $Z_1, \cdots, Z_Q$, and continuous variables $Y_1, \cdots, Y_C$. Let Z_j have $L_j + 1$ discrete values $a_j^{(1)} < \cdots < a_j^{(L_j+1)}$, $j = 1, \cdots, Q$. Underlying Z_j is Y_j^*, a continuous latent variable whose relationship with Z_j is defined by the following threshold model:

$$Z_j = \begin{cases} z_j^{(L_j)} & \text{iff } Y_j^* > \alpha_j^{(L_j)} \\ z_j^{(\ell_j)} & \text{iff } \alpha_j^{(\ell_j)} < Y_j^* \leq \alpha_j^{(\ell_j+1)}, \quad \ell_j = 1, \cdots, L_j - 1, \\ z_j^{(0)} & \text{iff } Y_j^* \leq \alpha_j^{(1)} \end{cases}$$

where $\alpha_j^{(1)}, \cdots, \alpha_j^{(L_j)}$, are unknown cutpoints or thresholds. Without loss of generality, we assume that $z_j^{(\ell_j)} = \ell_j$, $\ell_j = 0, \cdots, L_j$.

Given marginal distributions $F_{Y_1^*}(\cdot), \cdots, F_{Y_Q^*}(\cdot)$, and $F_{Y_1}(\cdot), \cdots, F_{Y_C}(\cdot)$, for continuous and latent variables, respectively, we assume that the joint distribution of $\mathbf{Y}^* = (Y_1^*, \cdots, Y_Q^*)^\top$, and $\mathbf{Y} = (Y_1, \cdots, Y_C)^\top$, is determined by a copula $C(\cdot)$. The joint distribution of $\mathbf{Z} = (Z_1, \cdots, Z_Q)^\top$ and $\mathbf{Y}$ is then

$$P\begin{pmatrix} Z_1 = \ell_1, \cdots, Z_Q = \ell_Q, \\ Y_1 \leq y_1, \cdots, Y_C \leq y_C \end{pmatrix} = \sum_{j=1}^{Q} \sum_{\epsilon_j=0}^{1} (-1)^{Q+\sum_{j=1}^{Q}\epsilon_j} C\begin{pmatrix} u_1^{(\ell_1+\epsilon_1)}, \cdots, u_Q^{(\ell_Q+\epsilon_Q)}, \\ v_1, \cdots, v_C \end{pmatrix}, \qquad (4)$$

where $u_j^{(\ell_j+\epsilon_j)} = F_{Y_j^*}(\alpha_j^{(\ell_j+\epsilon_j)})$, $j = 1, \cdots, Q$; $v_k = F_{Y_k}(y_k)$, $k = 1, \cdots, C$. It is clear that the above expression holds in the case $Q = 2$, and it can be easily shown to hold for general $Q > 2$ by induction.

The corresponding density of $\begin{pmatrix} \mathbf{Z} \\ \mathbf{Y} \end{pmatrix}$ is given by

$$f_{\mathbf{Z},\mathbf{Y}}(\boldsymbol{\ell}, \mathbf{y}) = \sum_{j=1}^{Q} \sum_{\epsilon_j=0}^{1} (-1)^{Q+\sum_{j=1}^{Q} \epsilon_j} \frac{\partial^C}{\partial y_1 \cdots \partial y_C} C \begin{pmatrix} u_1^{(\ell_1+\epsilon_1)}, \cdots, u_Q^{(\ell_Q+\epsilon_Q)}, \\ v_1, \cdots, v_C \end{pmatrix}, \qquad (5)$$

where $\boldsymbol{\ell} = (\ell_1, \cdots, \ell_Q)^\top$ and $\mathbf{y} = (y_1, \cdots, y_C)^\top$. Although (4) and (5) similarly appear in Song et al. (2009) and Song (2007), the approach outlined above contrasts with Song et al.'s (2009) and Song's (2007) direct application of copulas to model discrete variables. Genest and Nešlehová (2007) demonstrate that copulas directly linking discrete margins are unique only on the Cartesian product of marginal ranges due to non-uniformity of PITs in the discrete case; in fact, the class of possible copulas can be quite large, especially in the binary case. Although this may seem to be of mere theoretical interest without any practical implications, it has a host of serious consequences that bear directly on dependence modeling of discrete variables. For one, common rank-based association measures like Kendall's tau and Spearman's rho may now depend on the margins (see, e.g., Nešlehová, 2007; Mesfioui and Tajar, 2005). For another, the range of their possible values may be restricted–severely in some cases–rendering interpretations of such measures problematic.

Our use of latent variables to describe $Z_1, \cdots, Z_Q$, manages to sidestep these issues. Because the copula is introduced at the latent level, the resulting copula distribution is completely unique. In addition, the copula's dependence parameter is still margin-free and can be interpreted the usual way as proxy for the associations between the mixed variables. The model relies on association measures akin to polychoric and polyserial correlations to model the dependence structure of the data, without consequent restrictions on their admissible ranges. See section 2.2 for more details.

Note that the margins $F_{Y_1^*}(\cdot), \cdots, F_{Y_Q^*}(\cdot)$, and $F_{Y_1}(\cdot), \cdots, F_{Y_C}(\cdot)$, can be any continuous distributions and $C(\cdot)$ can be any copula, thus allowing researchers great flexibility in modelling mixed-variable joint distributions. In what follows, we take a fully parametric specification of the marginal distributions, and adopt the Gaussian copula to construct the joint distribution. This approach does not resort to conditioning, resulting in a symmetrical treatment of variables while preserving the margins, an attractive feature in practice.

2.1. Meta-Gaussian distributions

We now adopt the Gaussian copula to construct a meta-Gaussian distribution for $\begin{pmatrix} \mathbf{Z} \\ \mathbf{Y} \end{pmatrix}$. The use of Gaussian copula is appealing, since it describes dependence in the same way that the multivariate normal distribution does, with the difference that it does so for arbitrary random variables.

Following Song (2007, 2000), who used the Gaussian copula to generate a class of multivariate dispersion models, Proposition 1 below lays out density (5) of $\begin{pmatrix} \mathbf{Z} \\ \mathbf{Y} \end{pmatrix}$ using the Gaussian copula.

Proposition 1. *Assuming joint distribution (4) of* $\begin{pmatrix} \mathbf{Z} \\ \mathbf{Y} \end{pmatrix}$ *is expressed in terms of the Gaussian copula with correlation matrix* $\mathbf{R}$*, then joint density (5) is given by*

$$f_{\mathbf{Z},\mathbf{Y}}(\boldsymbol{\ell}, \mathbf{y}) = \frac{\phi(t_{C-1|C})\phi(t_{C-2|C:C-1}) \cdots \phi(t_{1|C:2})}{\prod_{k=1}^{C-1} \sqrt{1 - r^2_{W_k W_C}} \prod_{k=1}^{C-2} \sqrt{1 - r^2_{W_k W_{C-1}|W_C}} \cdots \sqrt{1 - r^2_{W_1 W_2 | W_C:W_3}}}$$

$$\times \sum_{\epsilon_1=0}^{1} \cdots \sum_{\epsilon_Q=0}^{1} (-1)^{Q+\sum_{j=1}^{Q}\epsilon_j} \Phi_Q \left(s_{1|C:1}^{(\ell_1+\epsilon_1)}, \cdots, s_{Q|C:1}^{(\ell_Q+\epsilon_Q)}; \mathbf{R}_{|W_C:W_1} \right)$$
$$\times f_{Y_C}(y_C) \prod_{k=1}^{C-1} \frac{f_{Y_k}(y_k)}{\phi(t_k)}, \tag{6}$$

with $f_{Y_k}(\cdot)$ the marginal density of Y_k, $\phi(\cdot)$ the standard normal density, and

$$\begin{aligned} t_{C-k|C:C-k+1} &= \frac{t_{C-k|C:C-k+2} - r_{W_{C-k}W_{C-k+1}|W_C:W_{C-k+2}} t_{C-k+1|C:C-k+2}}{\sqrt{1 - r^2_{W_{C-k}W_{C-k+1}|W_C:W_{C-k+2}}}} \\ s_{j|C:k}^{(\ell_j+\epsilon_j)} &= \frac{s_{j|C:k+1}^{(\ell_j+\epsilon_j)} - r_{W_j^* W_k|W_C:W_{k+1}} t_{k|C:k+1}}{\sqrt{1 - r^2_{W_j^* W_k|W_C:W_{k+1}}}}, \end{aligned}$$

where $W_j^* = \Phi^{-1}\{F_{Y_j^*}(Y_j^*)\}$ and $W_k = \Phi^{-1}\{F_{Y_k}(Y_k)\}$ are the normal scores (latent in the case of W_j^*), $s_j^{(\ell_j+\epsilon_j)} = \Phi^{-1}(u_j^{(\ell_j+\epsilon_j)})$, $t_k = \Phi^{-1}(v_k)$, $r_{W_{C-k}W_{C-k+1}|W_C:W_{C-k+2}}$ is the partial correlation between W_{C-k} and W_{C-k+1}, after eliminating $W_{C-k+2}, \cdots, W_C$, $r_{W_j^* W_{C-k}|W_C:W_{C-k+1}}$ is the partial correlation between W_j^* and W_{C-k}, after eliminating $W_{C-k+1}, \cdots, W_C$, and $\mathbf{R}_{|W_C:W_{C-k+1}} = \mathbf{R}_{W_1^*:W_{C-k}|W_C:W_{C-k+1}}$ is the partial correlation matrix for $W_1^*, \cdots, W_Q^*, W_1, \cdots, W_{C-k}$, after eliminating $W_{C-k+1}, \cdots, W_C$, for $k = 1, \cdots, C-1$.

To see how (6) is obtained, put $P = Q + C$ and note that

$$\Phi_P \begin{pmatrix} s_1^{(\ell_1+\epsilon_1)}, \cdots, s_Q^{(\ell_Q+\epsilon_Q)}, & ; \mathbf{R} \\ t_1, \cdots, t_C & \end{pmatrix} = \int_{-\infty}^{t_C} \phi(t) \Phi_{P-1} \begin{pmatrix} s_{1|C}^{(\ell_1+\epsilon_1)}, \cdots, s_{Q|C}^{(\ell_Q+\epsilon_Q)}, & ; \mathbf{R}_{|W_C} \\ t_{1|C}, \cdots, t_{C-1|C} & \end{pmatrix} dt, \tag{7}$$

so that we get

$$\frac{\partial}{\partial y_C} \Phi_P \begin{pmatrix} s_1^{(\ell_1+\epsilon_1)}, \cdots, s_Q^{(\ell_Q+\epsilon_Q)}, & ; \mathbf{R} \\ t_1, \cdots, t_C & \end{pmatrix} = \Phi_{P-1} \begin{pmatrix} s_{1|C}^{(\ell_1+\epsilon_1)}, \cdots, s_{Q|C}^{(\ell_Q+\epsilon_Q)}, & ; \mathbf{R}_{|W_C} \\ t_{1|C}, \cdots, t_{C-1|C} & \end{pmatrix} f_{Y_C}(y_C),$$

since $\partial t_C / \partial y_C = f_{Y_C}(y_C)/\phi(t_C)$. Next we have

$$\begin{aligned} \frac{\partial^2}{\partial y_{C-1} \partial y_C} \Phi_P \begin{pmatrix} s_1^{(\ell_1+\epsilon_1)}, \cdots, s_Q^{(\ell_Q+\epsilon_Q)}, & ; \mathbf{R} \\ t_1, \cdots, t_C & \end{pmatrix} &= \frac{\partial}{\partial y_{C-1}} \Phi_{P-1} \begin{pmatrix} s_{1|C}^{(\ell_1+\epsilon_1)}, \cdots, s_{Q|C}^{(\ell_Q+\epsilon_Q)}, & ; \mathbf{R}_{|W_C} \\ t_{1|C}, \cdots, t_{C-1|C} & \end{pmatrix} \\ &\quad \times f_{Y_C}(y_C). \end{aligned}$$

Applying (7), we get

$$\begin{aligned} \frac{\partial}{\partial y_{C-1}} \Phi_{P-1} \begin{pmatrix} s_{1|C}^{(\ell_1+\epsilon_1)}, \cdots, s_{Q|C}^{(\ell_Q+\epsilon_Q)}, & ; \mathbf{R}_{|W_C} \\ t_{1|C}, \cdots, t_{C-1|C} & \end{pmatrix} &= \Phi_{P-2} \begin{pmatrix} s_{1|C:C-1}^{(\ell_1+\epsilon_1)}, \cdots, s_{Q|C:C-1}^{(\ell_Q+\epsilon_Q)}, & ; \mathbf{R}_{|W_C:W_{C-1}} \\ t_{1|C:C-1}, \cdots, t_{C-2|C:C-1} & \end{pmatrix} \\ &\quad \times \frac{\phi(t_{C-1|C})}{\sqrt{1 - r^2_{W_{C-1}W_C}}} \frac{f_{Y_{C-1}}(y_{C-1})}{\phi(t_{C-1})}, \end{aligned}$$

where $\partial t_{C-1|C}/\partial y_{C-1} = \phi(t_{C-1|C})\partial t_{C-1}/\partial y_{C-1}$; it is then clear that repeated application of (7) leads to (6). Note that the order of differentiation can vary; in the above, we first differentiated with respect to y_C, then y_{C-1}, and so on, for notational convenience. Note as well that for the multivariate normal distribution, among others, partial and conditional correlations coincide (Baba et al., 2004). For example, the partial correlation $r_{W_{k-1}W_k|W_C:W_{C-k+2}}$ between W_{k-1} and W_k, after eliminating $W_{C-k+2}, \cdots, W_C$, is equal to the conditional correlation between W_{k-1} and W_k, given $W_{C-k+2}, \cdots, W_C$. These correlations can be computed recursively, e.g.,

$$r_{W_{k-1}W_k|W_C:W_{C-k+2}} = \frac{r_{W_{k-1}W_k|W_C:W_{C-k+3}} - r_{W_{k-1}W_{C-k+2}|W_C:W_{C-k+3}} r_{W_kW_{C-k+2}|W_C:W_{C-k+3}}}{\sqrt{(1 - r^2_{W_{k-1}W_{C-k+2}|W_C:W_{C-k+3}})(1 - r^2_{W_kW_{C-k+2}|W_C:W_{C-k+3}})}}.$$

The choice of the Gaussian copula, while arbitrary, is convenient because of its nice marginal and conditional properties. Meta-Gaussian distributions for mixed data have been studied previously by Song (2007, 2000). They were introduced for continuous data by Kelly and Krzysztofowicz (1997) and later generalized to a class of meta-elliptical distributions by Fang et al. (2002); see also Zhang et al. (2011).

The meta-Gaussian mixed-variable distribution (6) is general enough to include previously studied models as special cases. For instance, adopting normal margins for both continuous and latent variables specializes (6) as a CGCM; see also the normal-probit models of Poon and Lee (1987) and Catalano and Ryan (1992). It specializes to Song et al.'s (2009) and Song's (2007) model in the mixed binary-continuous case and can thus be considered as an extension of the latter to mixed polychotomous and continuous variables.

Logistic margins for $Y_1^*, \cdots, Y_Q^*$, result in a discrete-variable model similar to Nikoloulopoulos and Karlis' (2008) logit copula model. de Leon and Wu (2011) recently used a (generalized) t-latent distribution to yield a robit model (Liu, 2004) for binary variables as a robust alternative to probit and logit models. Indeed, the versatility and flexibility of our approach lie in the wide range of marginal models that can be accommodated.

We next derive marginal distributions of subvectors of $\mathbf{Z}$ and $\mathbf{Y}$. As Proposition 2 below shows, (6) is "reproducible" in the sense that its margins have the same form as (6). Consider $\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix}$ and $\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}$, where $\mathbf{Z}_1$ and $\mathbf{Y}_1$ are $Q_1 \times 1$ and $C_1 \times 1$, respectively, with $Q_1 = Q - Q_2$ and $C_1 = C - C_2$. We also partition $\mathbf{R}$ accordingly as

$$\mathbf{R} = \left(\begin{array}{c|c} \mathbf{R}_{\mathbf{W}^*} & \mathbf{R}_{\mathbf{W}^*\mathbf{W}} \\ \hline & \mathbf{R}_{\mathbf{W}} \end{array}\right) = \left(\begin{array}{cc|cc} \mathbf{R}_{\mathbf{W}_1^*} & \mathbf{R}_{\mathbf{W}_1^*\mathbf{W}_2^*} & \mathbf{R}_{\mathbf{W}_1^*\mathbf{W}_1} & \mathbf{R}_{\mathbf{W}_1^*\mathbf{W}_2} \\ & \mathbf{R}_{\mathbf{W}_2^*} & \mathbf{R}_{\mathbf{W}_2^*\mathbf{W}_1} & \mathbf{R}_{\mathbf{W}_2^*\mathbf{W}_2} \\ \hline & & \mathbf{R}_{\mathbf{W}_1} & \mathbf{R}_{\mathbf{W}_1\mathbf{W}_2} \\ & & & \mathbf{R}_{\mathbf{W}_2} \end{array}\right), \tag{8}$$

where $\mathbf{W}^* = \begin{pmatrix} \mathbf{W}_1^* \\ \mathbf{W}_2^* \end{pmatrix}$, and $\mathbf{W} = \begin{pmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \end{pmatrix}$, with $\mathbf{W}_1^* = (W_1^*, \cdots, W_{Q_1}^*)^\top$, $\mathbf{W}_2^* = (W_{Q_1+1}^*, \cdots, W_Q^*)^\top$, $\mathbf{W}_1 = (W_1, \cdots, W_{C_1})^\top$, and $\mathbf{W}_2 = (W_{C_1+1}, \cdots, W_C)^\top$.

Proposition 2. *Assuming the joint distribution of* $\begin{pmatrix} \mathbf{Z} \\ \mathbf{Y} \end{pmatrix}$ *is given by (6), the marginal density of* $\begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Y}_1 \end{pmatrix}$ *is then*

$$f_{\mathbf{Z}_1,\mathbf{Y}_1}(\boldsymbol{\ell}_1, \mathbf{y}_1) = \frac{\phi(t_{C_1-1|C_1})\phi(t_{C_1-2|C_1:C_1-1})\cdots\phi(t_{1|C_1:2})}{\prod_{k=1}^{C_1-1}\sqrt{1 - r^2_{W_kW_{C_1}}}\prod_{k=1}^{C_1-2}\sqrt{1 - r^2_{W_kW_{C_1-1}|W_{C_1}}}\cdots\sqrt{1 - r^2_{W_1W_2|W_{C_1}:W_3}}}$$

$$\times \sum_{\epsilon_1=0}^{1} \cdots \sum_{\epsilon_{Q_1}=0}^{1} (-1)^{Q_1+\sum_{j=1}^{Q_1}\epsilon_j} \Phi_{Q_1}\left(s_{1|C_1:1}^{(\ell_1+\epsilon_1)}, \cdots, s_{Q_1|C_1:1}^{(\ell_{Q_1}+\epsilon_{Q_1})}; \mathbf{R}_{\mathbf{W}_1^*,\mathbf{W}_1|W_{C_1}:W_1}\right)$$
$$\times f_{Y_{C_1}}(y_{C_1}) \prod_{k=1}^{C_1-1} \frac{f_{Y_k}(y_k)}{\phi(t_k)}, \tag{9}$$

where $\boldsymbol{\ell}_1^\top = (\ell_1, \cdots, \ell_{Q_1})$*,* $\mathbf{y}_1^\top = (y_1, \cdots, y_{C_1})$*, and the rest are defined analogously to those in Proposition 1, with* $\mathbf{R}_{\mathbf{W}_1^*,\mathbf{W}_1|W_{C_1}:W_{C_1-k+1}} = \mathbf{R}_{W_1^*:W_{Q_1}^*,W_1:W_{C_1-k}|W_{C_1}:W_{C_1-k+1}}$ *containing partial correlations based on* $\begin{pmatrix} \mathbf{R}_{\mathbf{W}_1^*} & \mathbf{R}_{\mathbf{W}_1^*\mathbf{W}_1} \\ & \mathbf{R}_{\mathbf{W}_1} \end{pmatrix}$*, for* $W_1^*, \cdots, W_{Q_1}^*$*, and* $W_1, \cdots, W_{C_1-k}$*, after eliminating* $W_{C_1-k+1}, \cdots, W_{C_1}$*,* $k = 1, \cdots, C_1 - 1$*.*

Proposition 2 easily follows from the "closure" property of Gaussian copulas:

$$\begin{aligned} \lim_{u_p \uparrow 1} \Phi_P\left(\Phi^{-1}(u_1), \cdots, \Phi^{-1}(u_P); \mathbf{R}\right) &= \lim_{q_p \uparrow \infty} \int_{-\infty}^{q_1} \cdots \int_{-\infty}^{q_P} \phi_P(x_1, \cdots, x_P; \mathbf{R}) dx_1 \cdots dx_P \\ &= \Phi_{P-1}\left(\cdots, \Phi^{-1}(u_p), \Phi^{-1}(u_{p-1}) \cdots; \mathbf{R}_{-(p)}\right), \end{aligned} \tag{10}$$

where $q_p = \Phi^{-1}(u_p)$, $p = 1, \cdots, P$, and $\mathbf{R}_{-(p)}$ is $\mathbf{R}$ with the pth column and pth row deleted. Applying (10) on the distribution of $\begin{pmatrix} \mathbf{Y}^* \\ \mathbf{Y} \end{pmatrix}$, we get the distribution of $\begin{pmatrix} \mathbf{Y}_1^* \\ \mathbf{Y}_1 \end{pmatrix}$ as a $(Q_1 + C_1)$-dimensional Gaussian copula with correlation matrix $\begin{pmatrix} \mathbf{R}_{\mathbf{W}_1^*} & \mathbf{R}_{\mathbf{W}_1^*\mathbf{W}_1} \\ & \mathbf{R}_{\mathbf{W}_1} \end{pmatrix}$, where $\mathbf{Y}_1^* = (Y_1^*, \cdots, Y_{Q_1}^*)^\top$. Density (9) is now immediate from Proposition 1.

It is also possible to obtain the conditional densities. For example, the conditional density of $\begin{pmatrix} \mathbf{Z}_2 \\ \mathbf{Y}_2 \end{pmatrix}$ given $\begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Y}_1 \end{pmatrix}$ easily follows from Propositions 1 and 2 as

$$\begin{aligned} f_{\mathbf{Z}_2,\mathbf{Y}_2|\mathbf{Z}_1,\mathbf{Y}_1}(\boldsymbol{\ell}_2, \mathbf{y}_2|\boldsymbol{\ell}_1, \mathbf{y}_1) \propto\ & \frac{\phi(t_{C-1|C})\phi(t_{C-2|C:C-1})\cdots\phi(t_{1|C:2})}{\prod_{k=1}^{C-1}\sqrt{1-r_{W_kW_C}^2}\prod_{k=1}^{C-2}\sqrt{1-r_{W_kW_{C-1}|W_C}^2}\cdots\sqrt{1-r_{W_1W_2|W_C:W_3}^2}} \\ & \times \sum_{\epsilon_1=0}^{1} \cdots \sum_{\epsilon_Q=0}^{1} (-1)^{Q+\sum_{j=1}^{Q}\epsilon_j} \Phi_Q\left(s_{1|C:1}^{(\ell_1+\epsilon_1)}, \cdots, s_{Q|C:1}^{(\ell_Q+\epsilon_Q)}; \mathbf{R}_{|W_C:W_1}\right) \\ & \times \frac{f_{Y_C}(y_C)}{\phi(t_{C_1})} \prod_{k=C_1+1}^{C-1} \frac{f_{Y_k}(y_k)}{\phi(t_k)}; \end{aligned} \tag{11}$$

note that while density (6) is reproducible under marginalization, conditional density (11) clearly does not have the same form as (6). This contrasts with earlier meta-Gaussian distributions for continuous data developed by Kelly and Krzysztofowicz (1997) (see also Bodnar et al., 2010) which have reproducible marginal and conditional densities.

2.2. Associations

Copula models usually rely on rank-based association measures, such as Kendall's tau or Spearman's rho (Balakrishnan and Lai, 2009, Chapter 4) to evaluate the strength of dependence between variables, as they are invariant to monotonic transformations (e.g., normal scores). However, unlike with continuous variables for which they provide margin-free measures of the level of dependence, this no longer holds in the discrete case (e.g., Genest and Nešlehová, 2007). Denuit and Lambert (2005) and Mesfioui and Tajar (2005) adopt a "continuous-ation" approach as a possible remedy. Interpretability of the association measures is another issue since their range varies, and there is thus a need to re-scale them.

Directly applying copulas to model multivariate discrete data, as in Song et al. (2009), suffers from the same problem many multivariate discrete distributions do: correlations and dependence parameters are unnaturally constrained to ensure the propriety of the joint probabilities. See, for example, the recent case of Li and Wong (2011) and Nikoloulopoulos (2011). Using continuous latent variables to describe discrete variables and constructing the copula model at the latent level gets around this problem.

The correlation matrix $\mathbf{R}$ contains three types of correlations: correlation $r_{W_j^* W_{j'}^*}$ between latent normal scores W_j^* and $W_{j'}^*$ (based on latent variables Y_j^* and $Y_{j'}^*$), correlation $r_{W_j^* W_k}$ between a latent normal score W_j^* and normal score W_k based on continuous variable Y_k, and correlation $r_{W_k W_{k'}}$ between normal scores based on Y_k and $Y_{k'}$. Bodnar et al. (2010) show that any pair of variables (latent and otherwise) in $\begin{pmatrix} \mathbf{Y}^* \\ \mathbf{Y} \end{pmatrix}$, hence any pair in $\begin{pmatrix} \mathbf{Z} \\ \mathbf{Y} \end{pmatrix}$, are independent if and only if the corresponding correlation between their normal scores (latent and otherwise) is 0.

The correlations in $\mathbf{R}$ are called normal or dependence correlation coefficients (Bodnar et al., 2010; Klaassen and Wellner, 1997), when variables are observable, as is the case with $r_{W_k W_{k'}}$. Since $\mathbf{Y}^*$ is latent and unobservable and because $r_{W_j^* W_{j'}^*}$ and $r_{W_j^* W_k}$ are analogous to polychoric and polyserial correlations $r_{Y_j^* Y_{j'}^*}$ and $r_{Y_j^* Y_k}$, respectively, we refer to $r_{W_j^* W_{j'}^*}$ as a *polychoric normal correlation coefficient*, and to $r_{W_j^* W_k}$ as a *polyserial normal correlation coefficient*. From the nonlinearity of normal quantile transforms, it follows from Theorem 6.1 of Klaassen and Wellner (1997) that

$$r_{Y_j^* Y_{j'}^*} < |r_{W_j^* W_{j'}^*}|, \;\; r_{Y_j^* Y_k} < |r_{W_j^* W_k}|, \;\; r_{Y_k Y_{k'}} < |r_{W_k W_{k'}}|. \tag{12}$$

Given $r_{W_j^* W_{j'}^*}$, $r_{W_j^* W_k}$, and $r_{W_k W_{k'}}$, it is possible to obtain $r_{Y_j^* Y_{j'}^*}$, $r_{Y_j^* Y_k}$, and $r_{Y_k Y_{k'}}$, as

$$r_{Y_j^* Y_{j'}^*} = \psi_{jj'}^{**}(r_{W_j^* W_{j'}^*}), \;\; r_{Y_j^* Y_k} = \psi_{jk}^{*}(r_{W_j^* W_k}), \;\; r_{Y_k Y_{k'}} = \psi_{kk'}(r_{W_k W_{k'}}), \tag{13}$$

for some functions $\psi_{jj'}^{**}$, ψ_{jk}^{*}, and $\psi_{kk'}$. Kugiumtzis and Bora-Senta (2010) use piece-wise linear approximations based on truncated standard bivariate normal variables to obtain $\psi_{jj'}^{**}$, ψ_{jk}^{*}, and $\psi_{kk'}$. While easy to implement, it may occasionally yield a non-positive definite correlation matrix of $r_{Y_j^* Y_{j'}^*}$, $r_{Y_j^* Y_k}$, and $r_{Y_k Y_{k'}}$; this arises mainly due to such correlations needing to satisfy certain admissible ranges.

Because $\begin{pmatrix} \mathbf{W}^* \\ \mathbf{W} \end{pmatrix}$ is a monotone transformation of $\begin{pmatrix} \mathbf{Y}^* \\ \mathbf{Y} \end{pmatrix}$, it follows that the Kendall's tau measures are such that $\tau_{W_j^* W_{j'}^*} = \tau_{Y_j^* Y_{j'}^*}$, $\tau_{W_j^* W_k} = \tau_{Y_j^* Y_k}$, and $\tau_{W_k W_{k'}} = \tau_{Y_k Y_{k'}}$. Assuming a meta-Gaussian distribution as in (6) for $\begin{pmatrix} \mathbf{Y}^* \\ \mathbf{Y} \end{pmatrix}$, these measures are easy to calculate given $\mathbf{R}$:

$$\tau_{W_j^* W_{j'}^*} = \frac{2}{\pi}\sin^{-1}(r_{W_j^* W_{j'}^*}), \;\; \tau_{W_j^* W_k} = \frac{2}{\pi}\sin^{-1}(r_{W_j^* W_k}), \;\; \tau_{W_k W_{k'}} = \frac{2}{\pi}\sin^{-1}(r_{W_k W_{k'}}); \tag{14}$$

they can also capture the full range of possible associations in $\begin{pmatrix} \mathbf{Y}^* \\ \mathbf{Y} \end{pmatrix}$, making them quite attractive in practice. A similar approach in the meta-Gaussian case may be adapted to Spearman's correlation rho $\rho_S = 6\sin^{-1}(\sin(\pi\tau/2)/2)/\pi$.

2.3. Examples

We now give two examples illustrating the class of meta-Gaussian mixed-variable distributions in (6). The first involves a continuous, a binary, and a trichotomous variable; the second is a binary-continuous pair.

Example 1. (*Probit-normal-exponential meta-Gaussian distribution*) Let $Y_1 \sim \text{normal}(\mu, \sigma^2)$ and $Y_2 \sim \text{exponential}(\lambda)$, and let Z_1 and Z_2 be binary and trichotomous random variables, such that

$$\begin{aligned} Z_1 &= \begin{cases} 1 & \text{iff } Y_1^* > \alpha_1 \\ 0 & \text{iff } Y_1^* \leq \alpha_1 \end{cases}, \\ Z_2 &= \begin{cases} 2 & \text{iff } Y_2^* > \alpha_2^{(2)} \\ 1 & \text{iff } \alpha_2^{(1)} < Y_2^* \leq \alpha_2^{(2)} \\ 0 & \text{iff } Y_2^* \leq \alpha_2^{(1)} \end{cases}, \end{aligned}$$

where Y_1^* and Y_2^* are standard normal latent variables; in addition to the usual assumption of unit latent variances, latent means are fixed at 0 since they are not identifiable from the cutpoints. This is the model described in (6) with $C = Q = 2$, and with $L_1 = 1$ and $L_2 = 2$. Normal latent distributions for Y_1^* and Y_2^* imply that Z_1 has a probit distribution and Z_2 is distributed according to Anderson and Pemberton's (1985) grouped continuous model (GCM). Assuming their joint distribution is specified by the Gaussian copula with correlation matrix $\mathbf{R}$ as in Proposition 1, we get the joint density of $(Z_1, Z_2, Y_1, Y_2)^\top$ as

$$f_{Z_1,Z_2,Y_1,Y_2}(\ell_1, \ell_2, y_1, y_2) = \frac{\lambda e^{-\lambda y_2} \phi(t_{1|2})}{\sigma\sqrt{1 - r^2_{W_1 W_2}}} \varphi(\ell_1, \ell_2, y_1, y_2), \tag{15}$$

[Table 1 about here.]

for $y_1 \in \mathbb{R}$ and $y_2 \geq 0$, where $\varphi(\ell_1, \ell_2, y_1, y_2)$ is evaluated at $\ell_1 = 0, 1$, and $\ell_2 = 0, 1, 2$, in Table 1, with $\Phi_2(\cdot; r^*)$ the standard bivariate normal distribution function with correlation r^*, and $t_1 = (y_1 - \mu)/\sigma$,

$$t_2 = \Phi^{-1}(1 - e^{-\lambda y_2}), \qquad t_{1|2} = \frac{t_1 - r_{W_1 W_2} t_2}{\sqrt{1 - r^2_{W_1 W_2}}}, \tag{16}$$

$$s_{1|21} = \frac{s_{1|2} - r_{W_1^* W_1|W_2} t_{1|2}}{\sqrt{1 - r^2_{W_1^* W_1|W_2}}}, \qquad s_{1|2} = \frac{\alpha_1 - r_{W_1^* W_2} t_2}{\sqrt{1 - r^2_{W_1^* W_2}}}, \tag{17}$$

$$s^{(j)}_{2|21} = \frac{s^{(j)}_{2|2} - r_{W_2^* W_1|W_2} t_{1|2}}{\sqrt{1 - r^2_{W_2^* W_1|W_2}}}, \qquad s^{(j)}_{2|2} = \frac{\alpha_2^{(j)} - r_{W_2^* W_2} t_2}{\sqrt{1 - r^2_{W_2^* W_2}}}, \tag{18}$$

$$r^* = \frac{r_{W_1^* W_2^*|W_2} - r_{W_1^* W_1|W_2} r_{W_2^* W_1|W_2}}{\sqrt{(1 - r^2_{W_1^* W_1|W_2})(1 - r^2_{W_2^* W_1|W_2})}}, \qquad r_{W_j^* W_1|W_2} = \frac{r_{W_j^* W_1} - r_{W_j^* W_2} r_{W_1 W_2}}{\sqrt{(1 - r^2_{W_j^* W_2})(1 - r^2_{W_1 W_2})}}, \tag{19}$$

for $j = 1, 2$; $r_{W_1^* W_2^*|W_2}$ is defined analogously to $r_{W_j^* W_1|W_2}$. The marginal densities of $(Z_1, Y_1)^\top$ and

$(Z_2, Y_2)^\top$ are similarly obtained via Proposition 2:

$$f_{Z_1,Y_1}(\ell_1, y_1) = \begin{cases} \frac{1}{\sigma}\phi(t_1)\Phi(-s_{1|1}) & \text{iff } \ell_1 = 1 \\ \frac{1}{\sigma}\phi(t_1)\Phi(s_{1|1}) & \text{iff } \ell_1 = 0 \end{cases}, \tag{20}$$

$$f_{Z_2,Y_2}(\ell_2, y_2) = \begin{cases} \lambda e^{-\lambda y_2}\Phi(-s^{(2)}_{2|2}) & \text{iff } \ell_2 = 2 \\ \lambda e^{-\lambda y_2}\{\Phi(s^{(2)}_{2|2}) - \Phi(s^{(1)}_{2|2})\} & \text{iff } \ell_2 = 1 \\ \lambda e^{-\lambda y_2}\Phi(s^{(1)}_{2|2}) & \text{iff } \ell_2 = 0 \end{cases}, \tag{21}$$

[Figure 1 about here.]

for $y_1 \in \mathbb{R}$ and $y_2 \geq 0$, where $s_{1|1} = (\alpha_1 - r_{W_1^* W_1} t_1)/\sqrt{1 - r^2_{W_1^* W_1}}$. Note that density (20) is a bivariate CGCM.

Figure 1 displays marginal densities (20) and (21) for $Y_1 \sim$ normal$(0, 1)$ and $Y_2 \sim$ exponential(2), with $\alpha_1 = 0.6745$ (i.e., $\Phi(\alpha_1) = 0.75$), $\alpha_2^{(1)} = -1.2816$ (i.e., $\Phi(\alpha_2^{(1)}) = 0.1$), $\alpha_2^{(2)} = 0.5244$ (i.e., $\Phi(\alpha_2^{(2)}) = 0.7$), and with with $r_{W_j^* W_{j'}^*} = r_{W_j^* W_k} = r_{W_k W_{k'}} = 0.5$, for all $j \neq j', k \neq k'$. Using Kugiumtzis and Bora-Senta's (2010) piece-wise linear method, these correspond to $r_{Y_1 Y_2} = r_{Y_1^* Y_2} = r_{Y_2^* Y_2} = 0.4423$, and $r_{Y_1^* Y_1} = r_{Y_2^* Y_1} = r_{Y_1^* Y_2^*} = 0.4868$. Note that these cutpoints correspond to skewed margins for Z_1 and Z_2. Figure 2(a) shows a scatterplot of data generated from the corresponding joint density (15); horizontal and vertical lines indicate $E(Y_1) = 0$ and $E(Y_2) = 0.5$.

The corresponding conditional density of $(Z_2, Y_2)^\top$ given $Z_1 = \ell_1$ and $Y_1 = y_1$, is

$$f_{Z_2,Y_2|Z_1,Y_1}(\ell_2, y_2|\ell_1, y_1) = \frac{\lambda e^{-\lambda y_2}\phi(t_{1|2})}{\sqrt{1 - r^2_{W_1 W_2}}\phi(t_1)} \times \begin{cases} \frac{1}{\Phi(s_{1|1})}\varphi(0, \ell_2, y_1, y_2) & \text{iff } \ell_1 = 0 \\ \frac{1}{\Phi(-s_{1|1})}\varphi(1, \ell_2, y_1, y_2) & \text{iff } \ell_1 = 1 \end{cases}, \tag{22}$$

for $y_2 \geq 0$ and $\ell_2 = 0, 1, 2$. This density is shown in Figure 2(b) with the same cutpoints and correlations as those in Figure 1, and with $\ell_1 = y_1 = 0$; similar plots may be constructed for $\ell_1 = 1$ or $y_1 \neq 0$.

[Figure 2 about here.]

[Figure 3 about here.]

Example 2. (*Robit-normal meta-Gaussian distribution*) Let $Y \sim$ normal(μ, σ^2), and assume a latent variable $Y^* \sim t_\nu$ such that $Z = \mathbf{1}(Y^* > \alpha)$. This is the model described in (6) with $C = Q = 1$ and $L = 1$. We model the joint distribution of Z and Y using a meta-Gaussian distribution. From Proposition 1, we get the joint density of Z and Y as

$$f_{Z,Y}(\ell, y) = \begin{cases} \frac{1}{\sigma}\phi(t)\Phi\left(-\frac{\Phi^{-1}(u) - rt}{\sqrt{1-r^2}}\right) & \text{iff } \ell = 1 \\ \frac{1}{\sigma}\phi(t)\Phi\left(\frac{\Phi^{-1}(u) - rt}{\sqrt{1-r^2}}\right) & \text{iff } \ell = 0 \end{cases}, \tag{23}$$

where $t = (y - \mu)/\sigma$, $1 - u = P(Z = 0) = P(Y^* \leq \alpha) = F_\nu(\alpha)$ is the distribution function of t_ν evaluated at α, and $r = r_{W^* Y}$ is the biserial normal correlation coefficient between latent normal score $W^* = \Phi^{-1}\{F_\nu(Y^*)\}$ and Y. Use of t-distributions as latent distributions is discussed in Liu (2004), who introduces the term "robit" model based on t-latent distributions as robust alternatives to and approximations of logit (i.e., $\nu \approx 7$, at which t_ν approximates standard logistic) and probit (i.e., t_ν approaches normal$(0,1)$ for large ν) models. A recent application of this model in a regression context is discussed in de Leon and Wu (2011).

An analytic expression for the regression function of Y given Z (Owen, 1970) is

$$E(Y|Z=\ell) \quad = \quad \mu + \begin{cases} \frac{r\sigma}{1-u}\phi\left\{\Phi^{-1}(u)\right\} & \text{iff } \ell = 1 \\ -\frac{r\sigma}{u}\phi\left\{\Phi^{-1}(u)\right\} & \text{iff } \ell = 0 \end{cases}, \tag{24}$$

thus demonstrating the non-linearity of the dependence between Z and Y, for $r \neq 0$. Figure 3 displays both joint density (23) and its regression function (24) for $Y \sim$ normal$(0,1)$, with $\nu = 5$ and $\alpha = 0.7267$ (i.e., $F_\nu(\alpha) = 0.75$), and with $r = 0.8$. For comparison, plot (a) also shows the joint density with standard logistic (i.e., logit model for Z) and standard normal (i.e., probit model for Z) latent distributions for Y^*. Plot (b) plots regression function (24) as a function of $u = F_\nu(\alpha)$ (i.e., as a function of cutpoint α); the value $u = 0.75$, indicated by the vertical line in the plot, corresponds to the actual cutpoint $\alpha = 0.7267$.

[Table 2 about here.]

3. Likelihood estimation

Let $\begin{pmatrix} \mathbf{z}_i \\ \mathbf{y}_i \end{pmatrix}$, $i = 1, \cdots, N$, denote the observed data. Putting $\boldsymbol{\psi}$ as the vector of parameters, the loglikelihood function takes the following form:

$$\begin{aligned} \ell(\boldsymbol{\psi}) \quad = \quad & \sum_{\boldsymbol{\ell},i} \mathbf{1}(\mathbf{z}_i = \boldsymbol{\ell}) \log \left\{ \sum_{\epsilon_1=0}^{1} \cdots \sum_{\epsilon_Q=0}^{1} (-1)^{Q+\sum_{j=1}^{Q}\epsilon_j} \Phi_Q\left(s_{1|C:1}^{(\ell_1+\epsilon_1)}, \cdots, s_{Q|C:1}^{(\ell_Q+\epsilon_Q)}; \mathbf{R}_{|W_C:W_1} \right) \right\} \\ & + \sum_i \left(\sum_{k=1}^{C} \log f_{Y_k}(y_{ik}) + \sum_{k=1}^{C-1} \log \left\{ \frac{\phi(t_{i,C-k|C:C-k+1})}{\phi(t_{ik})} \right\} \right) - \frac{N}{2} \sum_{k=1}^{C-1} \log(1 - r^2_{W_k W_C}) \\ & - \frac{N}{2} \sum_{k=1}^{C-2} \log(1 - r^2_{W_k W_{C-1}|W_C}) - \cdots - \frac{N}{2} \log(1 - r^2_{W_1 W_2|W_C:W_3}). \end{aligned} \tag{25}$$

Here, $\boldsymbol{\psi}$ contains marginal parameters (including cutpoints for $Z_1, \cdots, Z_Q$) as well as the (conditional and marginal) correlations in (25). Note that the correlations in $\boldsymbol{\psi}$ are a one-to-one reparametrization of the marginal correlations in $\mathbf{R}$. To see this, consider Example 1. In addition to marginal parameters μ, σ, and λ, and cutpoints $\alpha_1, \alpha_2^{(1)}$, and $\alpha_2^{(2)}$, the likelihood function based on (15) involves marginal correlations $r_{W_1W_2}, r_{W_1^*W_2}$, and $r_{W_2^*W_2}$, and conditional correlations $r_{W_1^*W_1|W_2}, r_{W_2^*W_1|W_2}$, and r^*. Noting that

$$r_{W_1^*W_1} \quad = \quad r_{W_1^*W_2} r_{W_1W_2} + r_{W_1^*W_1|W_2} \sqrt{(1 - r^2_{W_1^*W_2})(1 - r^2_{W_1W_2})}, \tag{26}$$

$$r_{W_2^*W_1} \quad = \quad r_{W_2^*W_2} r_{W_1W_2} + r_{W_2^*W_1|W_2} \sqrt{(1 - r^2_{W_2^*W_2})(1 - r^2_{W_1W_2})}, \tag{27}$$

$$\begin{aligned} r_{W_1^*W_2^*} \quad = \quad & r_{W_1^*W_2} r_{W_2^*W_2} + r_{W_1^*W_1|W_2} r_{W_2^*W_1|W_2} \sqrt{(1 - r^2_{W_1^*W_2})(1 - r^2_{W_2^*W_2})} \\ & + r^* \sqrt{(1 - r^2_{W_1^*W_1|W_2})(1 - r^2_{W_2^*W_1|W_2})(1 - r^2_{W_1^*W_2})(1 - r^2_{W_2^*W_2})}, \end{aligned} \tag{28}$$

it is clear that $\mathbf{R}$ can be recovered from $\boldsymbol{\psi}$.

[Table 3 about here.]

The maximum likelihood estimate (MLE) $\widehat{\boldsymbol{\psi}}$ is obtained by maximizing (25) using an iterative technique such as Newton-Raphson or quasi-Newton methods. One potential problem is that the estimate $\widehat{\mathbf{R}}$ may be non-positive definite; a practical remedy is to work instead with Fisher's z-transformation (de Leon and Carrière, 2007). It can be easily verified that $\widehat{\boldsymbol{\psi}}$ is consistent and asymptotically multivariate normal with mean $\boldsymbol{\psi}$ and covariance matrix given by the inverse of the Fisher information matrix $\imath(\boldsymbol{\psi}) = E\{-h(\boldsymbol{\psi})\} = E\{s(\boldsymbol{\psi})s^\top(\boldsymbol{\psi})\}$, where $s(\boldsymbol{\psi}) = \partial\ell(\boldsymbol{\psi})/\partial\boldsymbol{\psi}$ is the score function and $h(\boldsymbol{\psi}) = \partial^2\ell(\boldsymbol{\psi})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\top$ is the Hessian matrix. Standard errors (SEs) for $\widehat{\boldsymbol{\psi}}$ are calculated from diagonals of $\{s(\widehat{\boldsymbol{\psi}})s^\top(\widehat{\boldsymbol{\psi}})\}^{-1}$ or of $-h^{-1}(\widehat{\boldsymbol{\psi}})$, provided either matrix is invertible.

[Figure 4 about here.]

To evaluate the empirical performance of maximum likelihood estimates in the probit-normal-exponential meta-Gaussian distribution in Example 1, a total of $R = 1000$ repeated samples of size $N = 100, 200$, were generated with $\mu = 0, \sigma^2 = 1$, and $\lambda = 2$, cutpoints $\alpha_1 = 0.6745, \alpha_2^{(1)} = -1.2816$, and $\alpha_2^{(2)} = 0.5244$, and with with $r_{W_j^* W_{j'}^*} = r_{W_j^* W_k} = r_{W_k W_{k'}} = 0.5$, for all $j \neq j', k \neq k'$. Estimation was implemented in R using the optimization function `optim`, with correlations re-parameterized using Fisher's z-transformation. Expressions for the score function are obtained by direct differentiation (see Appendix). Standard errors (SEs) are calculated from the inverse of the observed Fisher information, and those for the marginal correlation estimates are obtained by delta method via (26)-(28). Details are found in the Appendix.

Results are presented in Table 2. These simulations generally suggest that MLEs for the model perform well in finite samples. Using the approximate margin of error $1.96\times$(SD/parameter) based on $R = 1000$ simulation repeats, relative biases for $N = 100, 200$, clearly suggest that the maximum likelihood method yields reasonably unbiased estimates. In addition, SEs of estimates are able to capture their true variability, with relative efficiencies all uniformly close to 100%. We also find generally slight improvement in bias and efficiency of estimates for sample size $N = 200$ over those for $N = 100$. Overall, it appears that maximum likelihood estimation provides good point estimates and standard errors for the model.

A drawback of maximum likelihood estimation in this case has to do with the computation of the rectangle probabilities in (25), which involves repeated multidimensional integration. Consequently, full likelihood inference might be difficult for high-dimensional settings; see Nikoloulopoulos and Karlis (2009). A possible remedy which achieves important computational economies is the use of composite likelihood methods; see, e.g., Varin (2008) and Zhao and Joe (2005). Computational and statistical performance (i.e. bias and efficiency) of this method has been shown to range from acceptably good to excellent. A study of this alternative will be the subject of future work.

4. Application in discrimination

In this section, we adopt the probit-normal-exponential meta-Gaussian distribution to develop classification procedures for mixed binary-continuous data. Consider two mixed binary-continuous populations Π_1 and Π_2, with respective densities $f_1(\cdot;\boldsymbol{\psi}_1)$ and $f_2(\cdot;\boldsymbol{\psi}_2)$ given by the probit-normal-exponential meta-Gaussian density (15), where $\boldsymbol{\psi}_h^\top = (\mu_h, \sigma_h, \lambda_h, \alpha_{h1}, \alpha_{h2}, \mathbf{r}_h^\top)$, with $\mathbf{r}_h^\top = (r_{h,W_1W_2}, r_{h,W_1^*W_2}, r_{h,W_2^*W_2}, r_{h,W_1^*W_1|W_2}, r_{h,W_2^*W_1|W_2}, r_h^*)$, for $h = 1, 2$. With equal prior probabilities and unit misclassification costs, the optimum classification rule assigns an individual with $Z_1 = \ell_1$, $Z_2 = \ell_2$, $Y_1 = y_1$, and $Y_2 = y_2$, to Π_1 if and only if

$$\mathcal{R}_{\ell_1\ell_2} \quad : \quad \log\left\{\frac{f_1(\ell_1, \ell_2, y_1, y_2; \boldsymbol{\psi}_1)}{f_2(\ell_1, \ell_2, y_1, y_2; \boldsymbol{\psi}_2)}\right\} > 0;$$

otherwise, it is assigned to Π_2. Note that separate discriminant functions are specified for $\ell_1, \ell_2 = 0, 1$. This is similar to Krzanowski's (1975) location linear discriminant function for binary-continuous data modelled by GLOM. The plug-in estimate $\widehat{\mathcal{R}}_{\ell_1\ell_2}$ of $\mathcal{R}_{\ell_1\ell_2}$ is obtained from the MLEs of $\boldsymbol{\psi}_h$; in practice, we may assume homogeneity of Π_1 and Π_2 (i.e., $\sigma_1 = \sigma_2$). The actual error rate (AER) is given by AER $= \{P(\Pi_1|\Pi_2) + P(\Pi_2|\Pi_1)\}/2$, where $P(\Pi_h|\Pi_{h'})$ is the probability of misclassifying $\Pi_{h'}$ as Π_h, given by

$$P(\Pi_h|\Pi_{h'}) \;=\; P(\Pi_{h'} \text{ misclassified as } \Pi_h) \;=\; \sum_{\ell_1=0}^{1}\sum_{\ell_2=0}^{1}\int_{\widehat{\mathcal{R}}_{\ell_1\ell_2}} f_{h'}(\ell_1, \ell_2, y_1, y_2; \boldsymbol{\psi}_{h'})dy_1dy_2, \tag{29}$$

for $h \neq h'$. To obtain an estimate of its mean μ_{AER}, AER may be evaluated by using MLEs of $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$, and then evaluating (29) by Monte Carlo methods. An alternative is Lachenbruch's hold-out procedure, which works well for moderate-sized samples and provides a nearly unbiased estimate of μ_{AER}; although it is relatively easy to implement in R, for example, this approach is more computationally intensive than the plug-in Monte Carlo method.

We illustrate the above classification method on advanced breast cancer data (Krzanowski, 1975) involving 186 patients who underwent ablative surgery for advanced breast cancer between 1958 and 1965 at Guy's Hospital, London. In $N_1 = 99$ cases, the treatment was deemed to be 'successful' or 'intermediate' (denoted as group Π_1), while in the remaining $N_2 = 87$ cases, it was a 'failure' (denoted as group Π_2). Six continuous and five binary variables are included in the data, of which we are primarily interested in the first two binary variables, and the first (rescaled by dividing by 50) and fifth (rescaled by dividing by 1000) continuous variables. The purpose is to classify a patient into group Π_1 or group Π_2 based on the mixed variables by modelling their joint distribution via the probit-normal-exponential meta-Gaussian distribution. Histograms for the two continuous variables, displayed in Figure 4, suggest that assuming normal and exponential distributions for the variables is reasonable. Maximum likelihood estimates of parameters for groups Π_1 and Π_2 are shown in Table 3. While MLEs of marginal parameters are reasonably close for both groups, those for marginal normal correlations suggest differences in dependence structures between them. Using piece-wise linear approximations, we get $\widehat{r}_{Y_1Y_2} = -0.382$, $\widehat{r}_{Y_1^*Y_1} = -0.347$, $\widehat{r}_{Y_2^*Y_1} = -0.047$, $\widehat{r}_{Y_1^*Y_2} = 0.107$, $\widehat{r}_{Y_2^*Y_2} = 0.217$, and $\widehat{r}_{Y_1^*Y_2^*} = -0.406$, for group Π_1; and $\widehat{r}_{Y_1Y_2} = -0.195$, $\widehat{r}_{Y_1^*Y_1} = -0.379$, $\widehat{r}_{Y_2^*Y_1} = -0.013$, $\widehat{r}_{Y_1^*Y_2} = 0.012$, $\widehat{r}_{Y_2^*Y_2} = -0.05$, and $\widehat{r}_{Y_1^*Y_2^*} = 0.368$, for group Π_2.

Out of $N_1 = 99$ patients in group Π_1, 26 were misclassified by hold-out method, and out of $N_2 = 87$ in group Π_2, 38 were misclassified, thus yielding a hold-out error estimate $\widehat{\mu}_{\text{AER}} = 0.344$. With $P\widehat{(\Pi_2|\Pi_1)} = 0.263$ and $P\widehat{(\Pi_1|\Pi_2)} = 0.437$, we see that we are more likely to misclassify a group Π_2 patient. While these results are comparable with Krzanowski's (1975) analysis, our approach based on copulas offer a more flexible method to model mixed data. Rather than having to transform variables to make them fit into a mold such as GLOM (adopted by Krzanowski, 1975) or CGCM, the practitioner can arbitrarily choose the margins and separately account for dependence via an appropriate copula.

5. Conclusion

This article describes a general class of meta-Gaussian distributions that can simultaneously accommodate various types of data, particularly mixed discrete and continuous variables. The distributions can be viewed as extensions of existing mixed-data models such as the GLOM and CGCM. An attractive feature of the distributions is their use of copulas to separately model dependencies between variables, thereby preserving the variables' distinct marginal properties. They thus offer a flexible alternative to conventional approaches that generally rely on the assumption that variables, or some suitable transformations of them, follow a Gaussian distribution. The meta-Gaussian distributions can serve as platforms for extending conventional multivariate methods to the case of mixed data, as illustrated by the discriminant analysis in the paper.

Appendix. Score function for density (15).

To derive the score function $s(\boldsymbol{\psi})$ for density (15), let $\ell(\boldsymbol{\psi}) = \sum_{i=1}^{N} \ell_i(\boldsymbol{\psi})$, where $\ell_i(\boldsymbol{\psi}) = \ell_i(\boldsymbol{\psi}; \ell_{i1}, \ell_{i2}, y_{i1}, y_{i2})$ is the loglikelihood contribution of individual i. It follows that $s(\boldsymbol{\psi}) = \sum_{i=1}^{N} s_i(\boldsymbol{\psi})$, where $s_i(\boldsymbol{\psi}) = \partial \ell_i(\boldsymbol{\psi})/\partial \boldsymbol{\psi}$ is the score contribution of individual i. With $\boldsymbol{\psi}^\top = (\mu, \sigma, \lambda, \alpha_1, \alpha_2^{(1)}, \alpha_2^{(2)}, r_{W_1W_2}, r_{W_1^*W_2}, r_{W_2^*W_2}, r_{W_1^*W_1|W_2}, r_{W_2^*W_1|W_2}, r^*)$ and for $\ell_{i1} = \ell_{i2} = 0$, we have

$$\ell_i(\boldsymbol{\psi}) \;=\; \log\lambda - \lambda y_{i2} - \log\sigma + \log\phi(t_{i,1|2}) - \frac{1}{2}\log(1 - r^2_{W_1W_2}) + \log\Phi_2(s_{i,1|21}, s^{(1)}_{i,2|21}; r^*). \quad (30)$$

Using identities (11.5)-(11.7) of Balakrishnan and Lai (2009, Chapter 11), we get the elements of $s_i(\boldsymbol{\psi})$ from (30) as follows:

$$\begin{aligned}\frac{\partial \ell_i(\boldsymbol{\psi})}{\partial \mu} \;=\; & -\frac{t_{i,1|2}}{\sigma\sqrt{1 - r^2_{W_1W_2}}} + \frac{1}{\sigma\sqrt{1 - r^2_{W_1W_2}}}\left\{\frac{r_{W_1^*W_1|W_2}\phi(s_{i,1|21})\Phi\left(\frac{s^{(1)}_{i,2|21} - r^* s_{i,1|21}}{\sqrt{1-r^{*2}}}\right)}{\sqrt{1 - r^2_{W_1^*W_1|W_2}}\Phi_2(s_{i,1|21}, s^{(1)}_{i,2|21}; r^*)}\right.\\ & \left. + \frac{r_{W_2^*W_1|W_2}\phi(s^{(1)}_{i,2|21})\Phi\left(\frac{s_{i,1|21} - r^* s^{(1)}_{i,2|21}}{\sqrt{1-r^{*2}}}\right)}{\sqrt{1 - r^2_{W_2^*W_1|W_2}}\Phi_2(s_{i,1|21}, s^{(1)}_{i,2|21}; r^*)}\right\}, \end{aligned} \quad (31)$$

$$\begin{aligned}\frac{\partial \ell_i(\boldsymbol{\psi})}{\partial \sigma} \;=\; & \frac{t_{i1}t_{i,1|2}}{\sigma^2\sqrt{1 - r^2_{W_1W_2}}} - \frac{1}{\sigma} + \frac{y_{i1} - \mu}{\sigma^2\sqrt{1 - r^2_{W_1W_2}}}\left\{\frac{r_{W_1^*W_1|W_2}\phi(s_{i,1|21})\Phi\left(\frac{s^{(1)}_{i,2|21} - r^* s_{i,1|21}}{\sqrt{1-r^{*2}}}\right)}{\sqrt{1 - r^2_{W_1^*W_1|W_2}}\Phi_2(s_{i,1|21}, s^{(1)}_{i,2|21}; r^*)}\right.\\ & \left. + \frac{r_{W_2^*W_1|W_2}\phi(s^{(1)}_{i,2|21})\Phi\left(\frac{s_{i,1|21} - r^* s^{(1)}_{i,2|21}}{\sqrt{1-r^{*2}}}\right)}{\sqrt{1 - r^2_{W_2^*W_1|W_2}}\Phi_2(s_{i,1|21}, s^{(1)}_{i,2|21}; r^*)}\right\}, \end{aligned} \quad (32)$$

$$\begin{aligned}\frac{\partial \ell_i(\boldsymbol{\psi})}{\partial \lambda} \;=\; & \frac{\phi(s_{i,1|21})\Phi\left(\frac{s^{(1)}_{i,2|21} - r^* s_{i,1|21}}{\sqrt{1-r^{*2}}}\right)}{\sqrt{1 - r^2_{W_1^*W_1|W_2}}\Phi_2(s_{i,1|21}, s^{(1)}_{i,2|21}; r^*)}\left(\frac{r_{W_1^*W_1|W_2}r_{W_1W_2}}{\sqrt{1 - r^2_{W_1W_2}}} - \frac{r_{W_1^*W_2}}{\sqrt{1 - r^2_{W_1^*W_2}}}\right)\\ & - \frac{\phi(s^{(1)}_{i,2|21})\Phi\left(\frac{s_{i,1|21} - r^* s^{(1)}_{i,2|21}}{\sqrt{1-r^{*2}}}\right)}{\sqrt{1 - r^2_{W_2^*W_1|W_2}}\Phi_2(s_{i,1|21}, s^{(1)}_{i,2|21}; r^*)}\left(\frac{r_{W_2^*W_1|W_2}r_{W_1W_2}}{\sqrt{1 - r^2_{W_1W_2}}} - \frac{r_{W_2^*W_2}}{\sqrt{1 - r^2_{W_2^*W_2}}}\right)\\ & + \frac{1}{\lambda} - \frac{\lambda e^{-\lambda y_{i2}} r_{W_1W_2} t_{i,1|2}}{\phi(t_{i,2})\sqrt{1 - r^2_{W_1W_2}}}, \end{aligned} \quad (33)$$

$$\frac{\partial \ell_i(\boldsymbol{\psi})}{\partial \alpha_1} = \frac{\phi(s_{i,1|21})\Phi\left(\frac{s^{(1)}_{i,2|21} - r^* s_{i,1|21}}{\sqrt{1-r^{*2}}}\right)}{\sqrt{(1-r^2_{W_1^* W_2})(1-r^2_{W_1^* W_1|W_2})}\Phi_2(s_{i,1|21}, s^{(1)}_{i,2|21}; r^*)}, \tag{34}$$

$$\frac{\partial \ell_i(\boldsymbol{\psi})}{\partial \alpha_2^{(1)}} = \frac{\phi(s^{(1)}_{i,2|21})\Phi\left(\frac{s_{i,1|21} - r^* s^{(1)}_{i,2|21}}{\sqrt{1-r^{*2}}}\right)}{\sqrt{(1-r^2_{W_2^* W_2})(1-r^2_{W_2^* W_1|W_2})}\Phi_2(s_{i,1|21}, s^{(1)}_{i,2|21}; r^*)}, \tag{35}$$

$$\frac{\partial \ell_i(\boldsymbol{\psi})}{\partial \alpha_2^{(2)}} = 0, \tag{36}$$

$$\begin{aligned}\frac{\partial \ell_i(\boldsymbol{\psi})}{\partial r_{W_1 W_2}} = {} & \frac{r_{W_2^* W_1|W_2}\phi(s^{(1)}_{i,2|21})\Phi\left(\frac{s_{i,1|21} - r^* s^{(1)}_{i,2|21}}{\sqrt{1-r^{*2}}}\right)}{\sqrt{1-r^2_{W_2^* W_1|W_2}}\Phi_2(s_{i,1|21}, s^{(1)}_{i,2|21}; r^*)} - \frac{r_{W_1^* W_1|W_2}\phi(s_{i,1|21})\Phi\left(\frac{s^{(1)}_{i,2|21} - r^* s_{i,1|21}}{\sqrt{1-r^{*2}}}\right)}{\sqrt{1-r^2_{W_1^* W_1|W_2}}\Phi_2(s_{i,1|21}, s^{(1)}_{i,2|21}; r^*)} \\ & + \frac{r_{W_1 W_1}}{1-r^2_{W_1 W_1}} + \frac{t_{i,1|2}(t_{i,1} r_{W_1 W_2} - t_{i,2})}{(1-r^2_{W_1 W_1})\sqrt{1-r^2_{W_1 W_1}}}\end{aligned} \tag{37}$$

$$\frac{\partial \ell_i(\boldsymbol{\psi})}{\partial r_{W_1^* W_2}} = \frac{(\alpha_1 r_{W_1^* W_2} - t_{i,2})\phi(s_{i,1|21})\Phi\left(\frac{s^{(1)}_{i,2|21} - r^* s_{i,1|21}}{\sqrt{1-r^{*2}}}\right)}{(1-r^2_{W_1^* W_2})\sqrt{(1-r^2_{W_1^* W_2})(1-r^2_{W_1^* W_1|W_2})}\Phi_2(s_{i,1|21}, s^{(1)}_{i,2|21}; r^*)} \tag{38}$$

$$\frac{\partial \ell_i(\boldsymbol{\psi})}{\partial r_{W_2^* W_2}} = \frac{(\alpha_2^{(1)} r_{W_2^* W_2} - t_{i,2})\phi(s^{(1)}_{i,2|21})\Phi\left(\frac{s_{i,1|21} - r^* s^{(1)}_{i,2|21}}{\sqrt{1-r^{*2}}}\right)}{(1-r^2_{W_2^* W_2})\sqrt{(1-r^2_{W_2^* W_2})(1-r^2_{W_2^* W_1|W_2})}\Phi_2(s_{i,1|21}, s^{(1)}_{i,2|21}; r^*)} \tag{39}$$

$$\frac{\partial \ell_i(\boldsymbol{\psi})}{\partial r_{W_1^* W_1|W_2}} = \frac{(s_{i,1|2} r_{W_1^* W_1|W_2} - t_{i,1|2})\phi(s_{i,1|21})\Phi\left(\frac{s^{(1)}_{i,2|21} - r^* s_{i,1|21}}{\sqrt{1-r^{*2}}}\right)}{(1-r^2_{W_1^* W_1|W_2})\sqrt{1-r^2_{W_1^* W_1|W_2}}\Phi_2(s_{i,1|21}, s^{(1)}_{i,2|21}; r^*)} \tag{40}$$

$$\frac{\partial \ell_i(\boldsymbol{\psi})}{\partial r_{W_2^* W_1|W_2}} = \frac{(s_{i,2|2} r_{W_2^* W_1|W_2} - t_{i,1|2})\phi(s^{(1)}_{i,2|21})\Phi\left(\frac{s_{i,1|21} - r^* s^{(1)}_{i,2|21}}{\sqrt{1-r^{*2}}}\right)}{(1-r^2_{W_2^* W_1|W_2})\sqrt{1-r^2_{W_2^* W_1|W_2}}\Phi_2(s_{i,1|21}, s^{(1)}_{i,2|21}; r^*)}, \tag{41}$$

$$\frac{\partial \ell_i(\boldsymbol{\psi})}{\partial r^*} = \frac{\phi_2(s_{i,1|21}, s^{(1)}_{i,2|21}; r^*)}{\Phi_2(s_{i,1|21}, s^{(1)}_{i,2|21}; r^*)}, \tag{42}$$

where $\phi_2(\cdot; r^*)$ is the standard bivariate normal density with correlation r^*. Expressions analogous to (31)-(42) can be obtained for cases $(\ell_{i1}, \ell_{i2}) = (0,1), (0,2), (1,0), (1,1)$, and $(1,2)$.

Acknowledgement

This research is partially supported by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada. Part of this work was carried out while A. R. de Leon was visiting the

School of Statistics, University of the Philippines.

References

[1] Baba, K., Shibata, R., and Sibuya, M., 2004, Partial correlation and conditional correlation as measures of conditional independence. *Australian and New Zealand Journal of Statistics*, **46**, 657–664.

[2] Balakrishnan, N. and Lai, C-D., 2009, *Continuous Bivariate Distributions*, 2nd ed., Springer.

[3] Bodnar, O., Bodnar, T., and Gupta, A. K., 2010, Estimation and inference for dependence in multivariate data. *Journal of Multivariate Analysis*, **101**, 869–881.

[4] Catalano, P. J. and Ryan, L. M., 1992, Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association*, **87**, 651–658.

[5] de Leon, A. R. and Carrière, K. C., 2007, General mixed-data model: extension of general location and grouped continuous models. *Canadian Journal of Statistics*, **35**, 533–548.

[6] de Leon, A. R. and Carrière Chough, K., 2010, Mixed-outcome data. In *Encyclopedia of Biopharmaceutical Statistics*, 3rd ed., S. C. Chow (ed.), vol. 2, pp. 817–822.

[7] de Leon, A. R. and Wu, B., 2011, Copula-based regression models for a bivariate mixed discrete and continuous outcome. *Statistics in Medicine*, **30**, 175–185.

[8] Denuit, M. and Lambert, P., 2005, Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis*, **93**, 40–57.

[9] Embrechts, P., 2009, Copulas: A personal view. *Journal of Risk & insurance*, **76**, 639–650.

[10] Fang, H-B., Fang, K-T., and Kotz, S., 2002, The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis*, **82**, 1–16.

[11] Genest, C. and Nešlehovà, J., 2007, A primer on copulas for count data. *ASTIN Bulletin*, **37**, 475–515.

[12] Hoff, P. D., 2007, Extending the rank likelihood for semiparametric copula estimation. *Annals of Applied Statistics*, **1**, 265–283.

[13] Joe, H., 1997, *Multivariate Models and Dependence Concepts*, Chapman & Hall.

[14] Kelly, K. S. and Krzysztofowicz, R., 1997, A bivariate meta-Gaussian density for use in hydrology. *Stochastic Hydrology and Hydraulics*, **11**, 17–31.

[15] Klaassen, C. A. J. and Wellner, J. A., 1997, Efficient estimation in the bivariate normal copula model: normal margins are least favourable. *Bernoulli*, **3**, 55–77.

[16] Krzanowski, W. J., 1975, Discrimination and classification using both binary and continuous variables. *Journal of the American Statistical Association*, **70**, 782-790.

[17] Kugiumtzis, D. and Bora-Senta, E., 2010, Normal correlation coefficient of non-normal variables using piece-wise linear approximation. *Computational Statistics*, **25**, 645–662.

[18] Li, J. and Wong, W. K., 2011, Two-dimensional toxic dose and multivariate logistic regression, with application to decompression sickness. *Biostatistics*, **12**, 143–155.

[19] Liu, C., 2004, Robit regression: a simple robust alternative to logistic and probit regression. In *Applied Bayesian Modeling and Causal Inference from Incomplete Data Perspectives*, Gelman A, Meng XL (eds.), Wiley, pp. 227–238.

[20] McCulloch, C., 2007, Joint modeling of mixed outcome types using latent variables. *Statistical Methods in Medical Research*, **17**, 1–21.

[21] McNeil, A. J., Frey, R., and Embrechts, P., 2005, *Quantitative Risk Management: Concepts, Techniques, Tools*, Princeton.

[22] Meester, S. and MacKay, J., 1994, A parametric model for cluster correlated categorical data. *Biometrics*, **50**, 954-963.

[23] Mesfioui, M. and Tajar, A., 2005, On the properties of some nonparametric concordance measures in the discrete case. *Journal of Nonparametric Statistics*, **17**, 541–554.

[24] Nešlehovà, J., 2007, On rank correlation measures for non-continuous random variables. *Journal of Multivariate Analysis*, **95**, 544–567.

[25] Nikoloulopoulos, A. K., 2011, Comment on "Two-dimensional toxic dose and multivariate logistic regression, with application to decompression sickness," by J. Li and W. K. Wong. Letter to the Editor of *Biostatistics.* To appear.

[26] Nikoloulopoulos, A. K. and Karlis, D., 2010, Modeling multivariate count data using copulas. *Communications in Statistics–Simulations & Computation*, **39**, 172–187.

[27] Nikoloulopoulos, A. K. and Karlis, D., 2009, Finite normal mixture copulas for multivariate discrete data modeling. *Journal of Statistical Planning & Inference*, **139**, 3878–3890.

[28] Nikoloulopoulos, A. K. and Karlis, D., 2008, Multivariate logit copula model with an application to dental data. *Statistics in Medicine*, **27**, 6393–6406.

[29] Olkin, I. and Tate, R. F., 1961, Multivariate correlation models with mixed discrete and continuous variables. *Annals of Mathematical Statistics*, **32**, 448–465 (correction: *Annals of Mathematical Statistics*, **36**, 343–344, 1965).

[30] Owen, D. B., 1980, A table of normal integrals. *Communications in Statistics–Simulation and Computations*, **9**, 389–419.

[31] Poon, W-Y. and Lee, S-Y., 1987, Maximum likelihood estimation of multivariate polyserial and polychoric correlation coefficients. *Psychometrika*, **52**, 409–430 (correction in **53**, p. 301).

[32] Shen, C. and Weissfeld, L., 2006, A copula model for repeated measurements with non-ignorable non-monotone missing outcome. *Statistics in Medicine*, **30**, 2427–2440.

[33] Song, P. X-K., 2007, *Correlated Data Analysis: Modeling, Analytics, and Applications*, Springer.

[34] Song, P. X-K., 2000, Multivariate dispersion models generated from Gaussian copula. *Scandinavian Journal of Statistics*, **27**, 305–320.

[35] Song, P. X-K., Li, M., and Yuan, Y., 2009, Joint regression analysis of correlated data using Gaussian copulas. *Biometrics*, **65**, 60–68.

[36] Trégouët, D.-A., Ducimetière, P., Bocquet, V., Visvikis, S., Soubrier, F., and Tiret, L., 2004, A parametric copula model for analysis of familial binary data. *American Journal of Human Genetics*, **64**, 886–893.

[37] Varin, C., 2008, On composite marginal likelihoods. *Advances in Statistical Analysis*, **92**, 1–28.

[38] Zhang, R., Czado, C., and Min, A., 2011, Efficient maximum likelihood estimation of copula-based meta t-distributions. *Computational Statistics and Data Analysis*, **55**, 1196–1214.

[39] Zhao, Y. and Joe, H., 2005, Composite likelihood estimation in multivariate data analysis. *Canadian Journal of Statistics*, **33**, 335–356.

[40] Zimmer, D. and Trivedi, P., 2006, Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand. *Journal of Buisness and Economic Statistics*, **24**, 63–76.

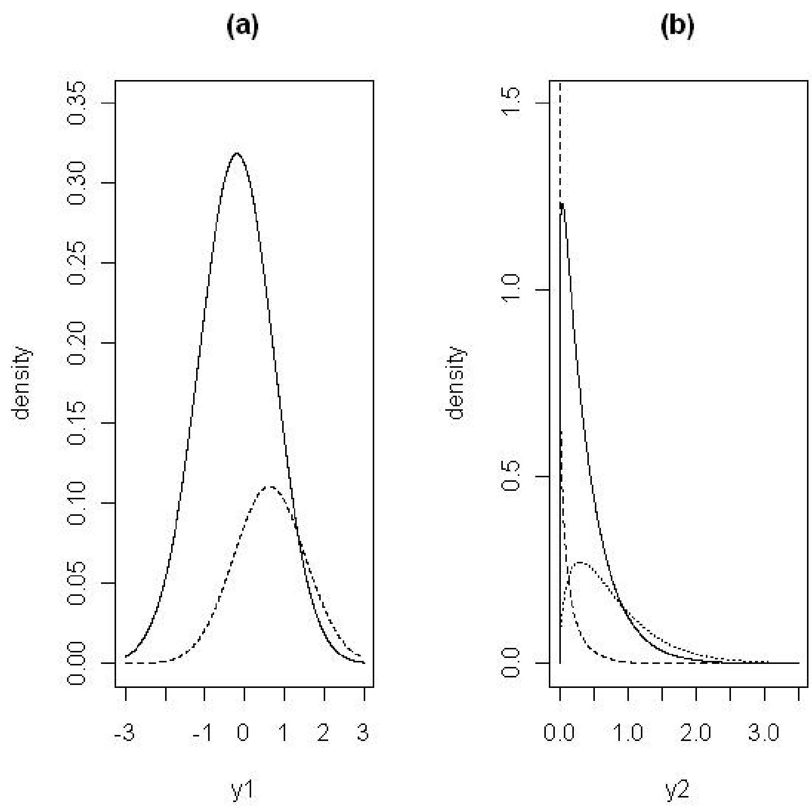

Figure 1: (a) is marginal density (20) of Z_1 and $Y_1 \sim$ normal$(0,1)$, and (b) is marginal density (21) of Z_2 and $Y_2 \sim$ exponential(2), with $\alpha_1 = .6745$, $\alpha_2^{(1)} = -1.2816$, $\alpha_2^{(2)} = .5244$, and with $r_{W_j^* W_{j'}^*} = r_{W_j^* W_k} = r_{W_k W_{k'}} = .5$, for all $j \neq j', k \neq k'$. Solid line corresponds to $\ell_1 = 0$ and dashed line to $\ell_1 = 1$ in (a); dashed line corresponds to $\ell_2 = 0$, solid line to $\ell_2 = 1$, and dotted line to $\ell_2 = 2$ in (b).

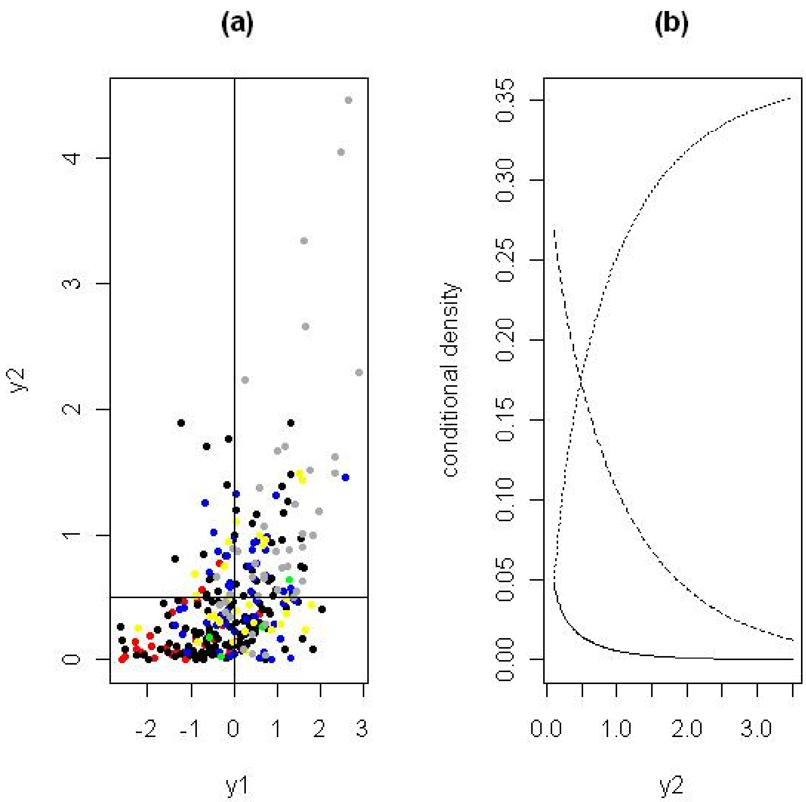

Figure 2: (a) is a scatterplot of $Y_1 \sim$ exponential(2) and $Y_2 \sim$ normal(0, 1) for $(\ell_1, \ell_2) = (0, 0)$ (red), $(0, 1)$ (black), $(0, 2)$ (blue), $(1, 0)$ (green), $(1, 1)$ (yellow), and $(1, 2)$ (gray), from joint density (15) with $N = 300$, and (b) is conditional density (22) of Z_2 and Y_2, given $\ell_1 = y_1 = 0$; both plots have $\alpha_1 = .6745$, $\alpha_2^{(1)} = -1.2816$, $\alpha_2^{(2)} = .5244$, and $r_{W_j^* W_{j'}^*} = r_{W_j^* W_k} = r_{W_k W_{k'}} = .5$, for all $j \neq j', k \neq k'$; in (a), vertical line corresponds to $E(Y_1) = 0$ and horizontal line to $E(Y_2) = .5$. Solid line in (b) corresponds to $\ell_2 = 0$, dashed line to $\ell_2 = 1$, and dotted line to $\ell_2 = 2$.

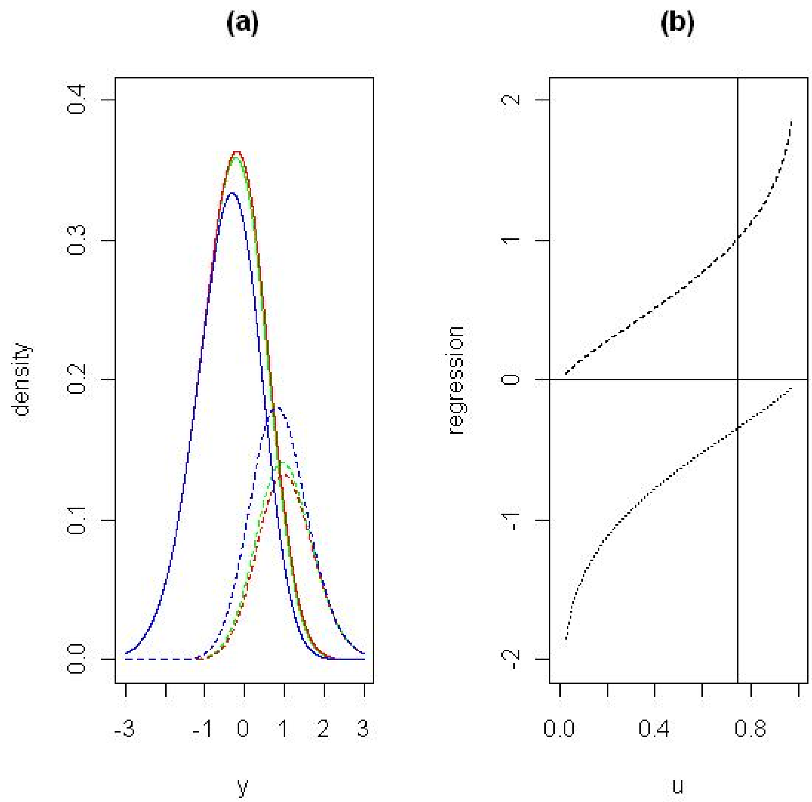

Figure 3: (a) is joint density (23) of Z and $Y \sim$ normal$(0, 1)$, and (b) is corresponding conditional mean (24) as a function of u, with $\alpha = .7267$, $\nu = 5$, and $r = .8$. Dashed line corresponds to $\ell = 0$ and dotted line to $\ell = 1$ in (a); dashed line to $\ell = 1$ and dotted line to $\ell = 0$ in (b); horizontal line in (b) is $\mu = 0$, and vertical line is $u = F_\nu(\alpha) = .75$. In plot (a), blue, red, and green plots correspond to standard logistic, standard normal, and t-latent distributions, respectively.

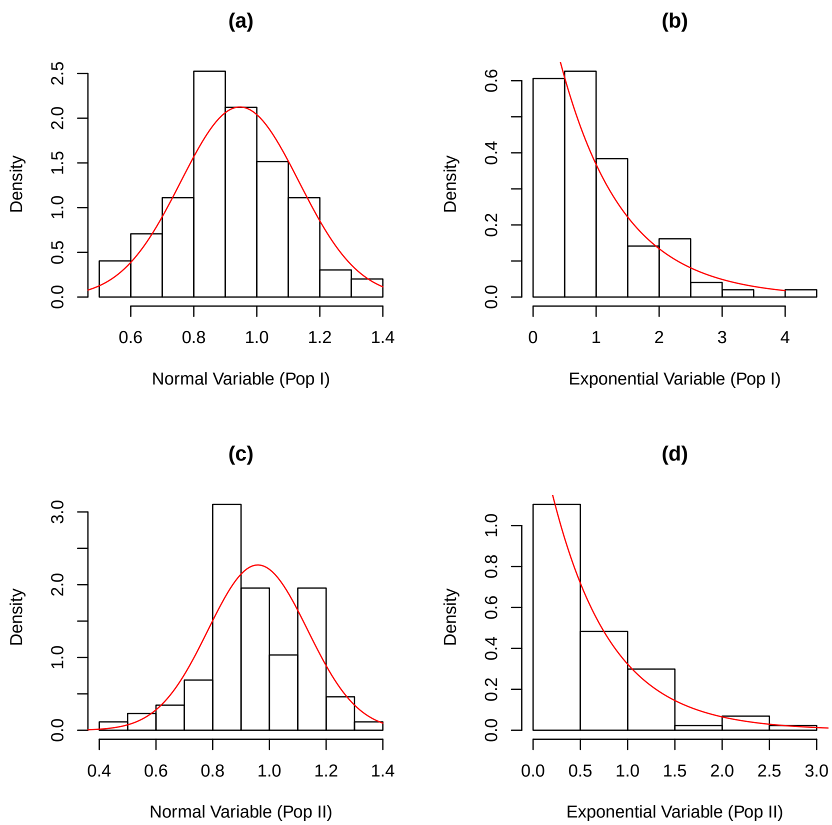

Figure 4: Histograms of continuous variables for breast cancer data. Histograms (a) and (c) correspond to the normal variable for groups Π_1 and Π_2, respectively; (b) and (d) are those of the exponential variable. Fitted normal and exponential densities are shown in red.

Table 1: $\varphi(\ell_1, \ell_2, y_1, y_2)$ evaluated at $\ell_1 = 0, 1$, and $\ell_2 = 0, 1, 2$.

	$\ell_1 = 0$	$\ell_1 = 1$
$\ell_2 = 0$	$\Phi_2(s_{1\|21}, s^{(1)}_{2\|21}; r^*)$	$\Phi_2(s_{1\|21}, s^{(2)}_{2\|21}; r^*) - \Phi_2(s_{1\|21}, s^{(1)}_{2\|21}; r^*)$
$\ell_2 = 1$	$\Phi_2(s_{1\|21}, s^{(2)}_{2\|21}; r^*) - \Phi_2(s_{1\|21}, s^{(1)}_{2\|21}; r^*)$	$\Phi(s^{(2)}_{2\|21}) - \Phi(s^{(1)}_{2\|21}) - \Phi_2(s_{1\|21}, s^{(2)}_{2\|21}; r^*)$ $+\Phi_2(s_{1\|21}, s^{(1)}_{2\|21}; r^*)$
$\ell_2 = 2$	$\Phi(s_{1\|21}) - \Phi_2(s_{1\|21}, s^{(2)}_{2\|21}; r^*)$	$1 - \Phi(s^{(2)}_{2\|21}) - \Phi(s_{1\|21})$ $+\Phi_2(s_{1\|21}, s^{(2)}_{2\|21}; r^*)$

Table 2: Average estimate (Ave Est), relative bias (RB=Bias÷Parameter), empirical standard deviation (SD), and relative efficiency (RE=Ave SE÷SD) of MLEs for probit-normal-exponential meta-Gaussian distribution (15) for $N = 100, 200$, and $R = 1000$. SEs of estimates were obtained from the observed Fisher information matrix; those for (marginal) correlations were calculated by delta method.

Parameter	N	Ave Est	RB	Ave SE	RE
$\mu = 0$	100	–.004	—	.069	.966
	200	–.002	—	.069	.996
$\sigma = 1$	100	.997	–.003	.049	1.034
	200	.999	–.001	.05	.972
$\lambda = 2$	100	2.015	.007	.142	1.01
	200	2.009	.005	.142	.992
$\alpha_1 = .674$	100	.691	.025	.096	.972
	200	.691	.024	.096	.989
$\alpha_2^{(1)} = -1.282$	100	–1.312	.023	.119	1.001
	200	–1.314	.025	.119	1.036
$\alpha_2^{(2)} = .524$	100	.535	.02	.093	1.023
	200	.528	.008	.092	.985
$r_{W_1W_2} = .5$	100	.504	.008	.05	.973
	200	.502	.004	.05	.953
$r_{W_2^*W_2^*} = .5$	100	.541	.083	.09	1.144
	200	.544	.088	.089	1.052
$r_{W_1^*W_2} = .5$	100	.512	.025	.074	1.008
	200	.514	.028	.074	.991
$r_{W_2^*W_2} = .5$	100	.504	.008	.061	1.006
	200	.511	.021	.061	.971
$r_{W_1^*W_1} = .5$	100	.514	.027	.075	.989
	200	.51	.02	.075	.974
$r_{W_2^*W_1} = .5$	100	.506	.013	.062	1.016
	200	.506	.012	.062	.966

Table 3: Estimates of parameters and fractions of hold-out misclassifications for breast cancer data. Classification was done with equal prior probabilities for the two groups and unit costs of misclassification.

Estimate	Group	
	Π_1	Π_2
$\widehat{\mu}$	.946	.959
$\widehat{\sigma}$	.188	.176
$\widehat{\lambda}$	1.009	1.6
$\widehat{\alpha}_1$	−.589	−.184
$\widehat{\alpha}_2$	.05	.21
$\widehat{r}_{W_1W_2}$	−.432	−.221
$\widehat{r}_{W_1^*W_2^*}$	−.418	.379
$\widehat{r}_{W_1^*W_2}$	.122	.013
$\widehat{r}_{W_2^*W_2}$	.245	−.056
$\widehat{r}_{W_1^*W_1}$	−.357	−.39
$\widehat{r}_{W_2^*W_1}$	−.049	−.013
Misclassified fraction	$\frac{26}{99}$	$\frac{38}{87}$