Published in final edited form as:

Biometrics. 2018 March ; 74(1): 185–195. doi:10.1111/biom.12711.

Inferring network structure in non-normal and mixed discrete-continuous genomic data

Anindya Bhadra,

Department of Statistics, Purdue University, 250 N. University St., West Lafayette, IN 47907

Arvind Rao, and

Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, 1400 Pressler Dr., Houston, TX 77030

Veerabhadran Baladandayuthapani

Department of Biostatistics, The University of Texas MD Anderson Cancer Center, 1400 Pressler Dr., Houston, TX 77030

Abstract

Inferring dependence structure through undirected graphs is crucial for uncovering the major modes of multivariate interaction among high-dimensional genomic markers that are potentially associated with cancer. Traditionally, conditional independence has been studied using sparse Gaussian graphical models for continuous data and sparse Ising models for discrete data. However, there are two clear situations when these approaches are inadequate. The first occurs when the data are continuous but display non-normal marginal behavior such as heavy tails or skewness, rendering an assumption of normality inappropriate. The second occurs when a part of the data is ordinal or discrete (e.g., presence or absence of a mutation) and the other part is continuous (e.g., expression levels of genes or proteins). In this case, the existing Bayesian approaches typically employ a latent variable framework for the discrete part that precludes inferring conditional independence among the data that are *actually observed*. The current article overcomes these two challenges in a unified framework using Gaussian scale mixtures. Our framework is able to handle continuous data that are not normal and data that are of mixed continuous and discrete nature, while still being able to infer a sparse conditional sign independence structure among the observed data. Extensive performance comparison in simulations with alternative techniques and an analysis of a real cancer genomics data set demonstrate the effectiveness of the proposed approach.

Keywords

Bayesian methods; Conditional sign independence; Genomic data; Graphical models; Mixed discrete and continuous data; Scale mixtures

Supplementary Material

The Supplementary Material available with this paper at the Biometrics website on Wiley Online Library contains background information on Gaussian graphical models, proofs of Proposition 1 and Lemma 1, MCMC diagnostics, additional simulations, details of GBM data, additional figures and tables referenced in Sections 4 and 5, and computer code written in MATLAB.

1 Introduction

With rapid advances in high-throughput genomic technologies using array and sequencing-based approaches, it is now possible to collect detailed high-resolution molecular information across the entire genomic landscape at various levels. The data can be genetic (e.g., mutations or single nucleotide polymorphisms), genomic (e.g., expression levels of messenger RNA and microRNA), epigenomic (e.g., DNA methylation) or proteomic (e.g., protein expression). The interrelations among these data provide key insights into the etiology of many diseases, including cancer. Statistically, the question of uncovering the major modes of multivariate interactions in genomic data can be phrased in terms of inferring a *conditional independence* graph. A unifying feature of these genomics problems is that the number of parameters far exceeds the sample size. Therefore, a multivariate sparse Gaussian graphical model is commonly applied to analyze the conditional independence structure (see, e.g., Lauritzen, 1996; Carvalho et al., 2007; Friedman et al., 2008; Meinhausen and Bühlmann, 2006; Bhadra and Mallick, 2013; Feldman et al., 2014). Given this high-dimensional setting, the purpose of the current article is to study multivariate interactions in two important situations where a Gaussian graphical model is inappropriate. These are (i) when the data are continuous, but display non-normal features such as heavy tails or skewness and (ii) when the data are of mixed discrete and continuous nature.

First, consider the case where all data are continuous but possibly non-normal. This is particularly important in genomics where the data often display features such as heavy tails. Moreover, in a multivariate setting, each marginal may display a separate characteristic. As a motivating example, in Figure 1 we plot the expression levels of two genes (AKT3 and CDK4) that are implicated in glioblastoma multiforme (GBM), which is the most aggressive form of brain cancer (TCGA, 2008). It is apparent that each marginal deviates from normality *in a different way*, especially in the tails (Kolmogorov-Smirnov test p-values 6.26e-6 and 1.49e-4, respectively). Since diseases such as cancer are often characterized by extreme changes in gene expression (Gray and Collins, 2000), capturing the tail behavior is crucial. Biological consequences of using a misspecified Gaussian model are serious, potentially resulting in an inference of wrong associations (Marko and Weil, 2012). There are some recent works in Bayesian literature for allowing for more flexible marginal behavior in the data, e.g., the alternative multivariate-*t* or Dirichlet-*t* of Finegold and Drton (2011, 2014), but, in view of Figure 1, it raises the question why one particular distribution (e.g., a *t*-distribution) would be appropriate along all the marginals. Furthermore, a *t*-distributed marginal cannot model important behavior often observed in genomics, e.g., skewness.

A second problem with genomic data is that it is heterogeneous (mixed discrete, ordinal and continuous). For example, presence or absence of mutations are modeled as binary variables; copy number aberrations as ordinal variables (gain/loss/normal); and expression levels of microRNA or messenger RNA are continuous. Characterizing the dependence among heterogeneous types of data is not well-understood, even in low dimensions. A typical Bayesian approach is to model the discrete part with latent continuous random variables and then to infer the conditional independence structure among the observed and latent continuous variables. It is unclear, however, how this latent dependence or correlation

translates to the observed data (Pitt et al., 2006). Outside of Bayesian approaches, this problem has received some recent attention, but the proposed techniques are limited to exponential family distributions (Cheng et al., 2013; Chen et al., 2015; Yang et al., 2015; Lee and Hastie, 2015).

Given these two problems, the focus of the current work is to delineate a unifying framework that can infer “conditional sign independence” in the face of data that are non-Gaussian and are of mixed discrete/continuous nature. We define two random variables ζ_1 and ζ_2 to be conditionally sign independent given ζ_3 , if the sign of ζ_1 given ζ_3 remains independent of whether ζ_2 is also known. A more precise definition is given later in Definition 1. Note that this definition has an intuitive appeal in multivariate genomic data of mixed nature. Here it might not make sense to compare the numeric values of data that are truly quantitative (e.g., gene expression) versus data that are binary $\{1, -1\}$ coded dummy variables (presence or absence of a mutation). But one might still be interested to see if positive values of the dummy variable (indicating presence of mutation) co-occurs with positive expression level of some gene (also known as up-regulation), conditional on the rest of the variables of interest. One might also want to investigate if two arbitrarily coded binary deleterious mutations are likely to co-occur, accounting for the effect of the rest of the variables.

Using a Gaussian scale mixture representation of the marginals, we show that it is possible to draw these conclusions. A key contribution of our work is that we can make statements concerning conditional sign independence among *observed* discrete and continuous random variables. This property makes our approach distinct from the literature on Bayesian copula graphical models (e.g., Pitt et al., 2006) that can only make statements conditional on some latent variables. The rest of the manuscript is organized as follows. We discuss the two main innovations of the paper, characterization of conditional sign independence in non-Gaussian and mixed discrete-continuous data in Sections 2 and 3 respectively. Simulation results and extensive performance comparison with alternative approaches are in Section 4. We analyze a cancer genomics data set in Section 5. We conclude by pointing out some directions of future investigation, including a possible E–M scheme that can be useful in non-Bayesian analysis of mixed data, in Section 6.

2 Inferring conditional sign independence in non-Gaussian continuous data using Gaussian scale mixtures

Let $\mathbf{Y}$ be the $n \times q$ data matrix, where n is the sample size and q is the number of variables. Consider the case where all q variables are continuous, but do not necessarily possess Gaussian marginals. We formulate the proposed model through a continuous, monotone, *random* transformation function of the marginals $\mathcal{F}=(f_1, \dots, f_q)$. Specifically, we assume

$$\mathcal{F}(\mathbf{Y})|\sum_{\mathbf{G}} \sim \text{MN}_{n \times q}(\mathbf{0}, \mathbf{I}_n, \sum_{\mathbf{G}}), \quad (1)$$

where $\mathcal{F}(\mathbf{Y})$ is an $n \times q$ transformed data matrix, modeled as a matrix-variate normal (Dawid, 1981), $\mathbf{0}_{n \times q}$ is an $n \times q$ mean matrix of zeros, Σ_G is the $q \times q$ column covariance matrix of q possibly correlated variables and $\mathbf{I}_n$ is an identity matrix of size n . Here $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ is a graph such that for all $u, v \in \mathbf{V}$, one has $\{u, v\} \notin \mathbf{E} \iff \sum_{\mathbf{G}}^{-1}(u, v) = 0$, where $\mathbf{V}$ corresponds to set of the transformed variables $f_1(Y_1), \dots, f_q(Y_q)$. More details on matrix normal and Bayesian approaches to Gaussian graphical models are given in Supplementary Section S.1. Two important points to note regarding this formulation are the following:

1. In a Bayesian formulation, one can further put priors on each random transformation function, thereby capturing a wide range of marginal behaviors.
2. Liu et al. (2009, 2012) showed that for continuous multivariate data, a deterministic monotone transform of the marginals aids interpretability. More specifically, Liu et al. (2009) showed if the transformation functions $f_1, \dots, f_q$ in Equation (1) are independent and monotone then conditional independence in the transformed data implies conditional independence in the original data. Liu et al. (2012) relaxed the Gaussianity assumption of Equation (1) to symmetric elliptically contoured distributions. The price one pays for the relaxed assumption is that now it is only possible to infer Kendall's rank correlation (Kendall, 1938).

However, not much is known regarding the nature of dependence in the observed data when the transformation functions are random. We start by stating the following definition.

Definition 1—Two random variables ζ_1 and ζ_2 are said to be conditionally sign independent given ζ_3 , if $\mathbb{P}(\zeta_1 < 0 | \zeta_3) = \mathbb{P}(\zeta_1 < 0 | \zeta_2, \zeta_3)$; provided these conditional probabilities exist.

Note that it is only necessary to state the definition in any one direction and the conditional sign independence in the other direction follows readily. We are now ready to state our main result for random scale transformations.

Proposition 1

- i. (Conditional sign independence). Consider in Equation (1) the scale transformation $\mathcal{F}(\mathbf{Y}) = \mathbf{Y}\mathbf{D}$, where the elements of $\mathbf{D} = \text{diag}(1/d_i)$ are independent with $0 < d_i < \infty$ almost surely with $\int dp(d_i) < \infty$ for $i = 1, \dots, q$. Under the model of Equation (1),

$$\left\{ \sum_{\mathbf{G}}^{-1} \right\}_{\gamma, \nu} = 0 \iff \mathbb{P}(Y_\gamma < 0 | Y_{-\{\gamma, \nu\}}) = \mathbb{P}(Y_\gamma < 0 | Y_{-\gamma}).$$

- ii. (Conditional uncorrelatedness). Moreover if d_i s are almost surely the same random variable τ with $(\tau^{-1}) < \infty$ then

$$\left\{ \sum_{\mathbf{G}}^{-1} \right\}_{\gamma, \nu} = 0 \iff (Y_\gamma | Y_{-\{\gamma, \nu\}}) = (Y_\gamma | Y_{-\gamma}).$$

The proof is given in supplementary Section S.2. Part (i) implies that a missing edge $\{\gamma, \nu\}$ in the graph $\mathbf{G}$ implies the sign of Y_γ is independent from that of Y_ν given the rest of the

variables. Admittedly, this result is weaker than conditional independence for Gaussian graphical models (the case where $d_i = 1$ for all i , a.s.) or, as part (ii) implies, conditional uncorrelatedness for symmetric elliptically contoured distributions (the case where $d_1 = \dots = d_q = \tau$ a.s. with $E(\tau^{-1}) < \infty$). An example of the latter is given by Finegold and Drton (2011) for the multivariate t distribution. In this case, note that if $Y \sim t_\nu(\mu, \Sigma_G)$, a multivariate- t distribution with degrees of freedom ν , location vector μ , and scale matrix Σ_G , then a scale mixture representation is $Y|\tau, \Sigma_G \sim N(\mu, \tau \Sigma_G)$, $\tau \sim \text{Inv-Gamma}(\nu/2, \nu/2)$. Since the same scale parameter τ is used for all the margins, conditional uncorrelatedness follows (also proved in Proposition 1 of Finegold and Drton, 2011).

This should not come as a surprise, however, since progressively relaxed model assumptions usually come at the cost of progressively weaker statistical conclusions that can be drawn from the model. One cannot expect the relative magnitude among the Y s to be preserved under different scaling along different marginals. However, the sign of a random variable is independent of its scaling, so long as $0 < d_i < \infty$ a.s., providing an intuitive justification of why part (i) of Proposition 1 holds.

2.1 Some examples of continuous marginals in a Gaussian scale mixture

To further motivate the proposed framework, we now give a few examples of the wide range of marginals we can capture for continuous data in order to infer conditional sign independence.

Example 1—(Power exponential family). Consider the (monotone) scale transformation $\mathcal{F}(Y) = YD = \{y_1/d_1, \dots, y_q/d_q\}$ for a $q \times q$ diagonal matrix $D = \text{diag}(1/d_i)$. Let p be a generic density and consider the Gaussian scale mixture representation

$$p(y_i) = \int_0^\infty (2\pi d_i)^{-1/2} \exp\left\{-y_i^2/(2d_i)\right\} dp(d_i). \quad (2)$$

West (1987) showed that the marginal of y_i is of the form $p(y_i) = k \exp(-|y_i|^b)$ (power-exponential family) if d_i follows a stable distribution with index $b/2$. Since the power-exponential family includes Gaussian ($b = 2$) or double-exponential ($b = 1$) as special cases, we can make provisions for such marginals.

Example 2—(Generalized hyperbolic family). If the mixing distribution in Equation (2) is generalized inverse Gaussian (GIG), the marginals are in the generalized hyperbolic (GH) family. This is due to Barndorff-Nielsen (1977) who showed if the mixing distribution is

$$p(d_i) = \frac{(\psi/\chi)^{\lambda/2}}{2K_\lambda(\sqrt{\chi\psi})} d_i^{\lambda-1} \exp\left\{-(1/2)(\chi d_i^{-1} + \psi d_i)\right\}, \quad (3)$$

then the marginal is in the GH family and can be written as

$$p(y_i) = \frac{(\psi/\chi)^\lambda}{\sqrt{2\pi} K_\lambda(\psi\chi)} \times K_{\lambda-1/2} \left(\psi \sqrt{\chi + y_i^2} \right) \times \left(\sqrt{\chi + y_i^2 + \psi} \right)^{\lambda-1/2}.$$

Here $K_\lambda(\cdot)$ is the modified Bessel function of the third kind with index λ . The domain of the parameters (ψ, χ, λ) and multivariate generalizations are given by Barndorff-Nielsen (1978). The GH family includes t marginals as a special case, if each d_i is independent inverse gamma. With the appropriate mixing density on d_i , we can model other useful marginals, e.g. normal-gamma (Griffin and Brown, 2010) or variance gamma (Kotz et al., 2001).

Example 3—(Skewed location-scale family). Consider the location-scale transformation $\mathcal{F}(\mathbf{Y}) = \{(y_1 - \mu_1)/d_1, \dots, (y_q - \mu_q)/d_q\}$, with the relation $\mu_i = \alpha_i + \beta_i d_i$ for constants α_i and β_i . In this case, Barndorff-Nielsen (1977) showed mixing over d_i with mixing distribution given by Equation (3) gives rise to marginals with asymmetric tails. *This is useful for modeling skewness.* The pure scale transformation is a special case with $\alpha_i = \beta_i = 0$.

For all the above examples, Metropolis-Hastings samplers can be implemented, enabling practical implementation. While these examples demonstrate the flexibility of the marginal behavior we can model, a fundamental question remains. Given the data, how do we decide what is an appropriate distribution of the scale parameter in a Gaussian scale mixture representation? We prove the following lemma.

Lemma 1—(i) (Polynomial tails). If the tail of the i 'th marginal $f_i(y_i)$ decays as $|y_i|^{2\lambda_i-1}$ for some $\lambda_i > 0$ as $|y_i| \rightarrow \infty$, the mixing density of d_i has tail decaying as $d_i^{\lambda_i-1}$ as $d_i \rightarrow \infty$. (ii) (Exponential tails). If the tail of the i 'th marginal $f_i(y_i)$ decays as

$$|y_i|^{2\lambda_i-1} \exp \left\{ -2(\psi_i)^{1/2} |y_i| \right\} \text{ for } \lambda_i \in \mathbb{R}, \psi_i > 0, \text{ the mixing density of } d_i \text{ has tail decaying as } d_i^{\lambda_i-1} \exp(-\psi_i d_i) \text{ as } d_i \rightarrow \infty.$$

A proof is given in Supplementary Section S.3. The above result points to the power of Gaussian scale mixture representation in which the scale can be carefully calibrated to appropriately model the corresponding marginal. In general, any heavy polynomially decaying tail can be modeled. Tails decaying at exponential rates (e.g., Laplace) can also be modeled. Lemma 1 shows that depending on each marginal, one can decide what would be an appropriate mixing density, giving a practical guide to choosing $\mathbf{D}$. For this purpose, plotting marginal q-q plots or histograms will suffice, and one need not be concerned regarding higher order interactions at this point.

Comparing the proposed method to recent alternatives, such as the “alternative multivariate t ” (Finegold and Drton, 2011), we find two main advantages. First, in our case, the univariate marginals need not all have the same distribution. Our approach includes t -distributed marginals of Finegold and Drton (2011) as a special case (if all mixing distributions on the d_i s are independent inverse gamma), but is of course, much more

flexible. Second, the alternative multivariate- t can only model symmetric tails. However, our approach can handle asymmetric tails using a location-scale mixture, thereby capturing skewness.

2.2 MCMC procedure for inferring $\mathbf{G}$

We have $\mathbf{YD} = \{\mathbf{y}_1/d_1, \dots, \mathbf{y}_q/d_q\} \sim \text{MN}_{n \times q}(0, \mathbf{I}_n, \Sigma_{\mathbf{G}})$ for a $q \times q$ diagonal matrix $\mathbf{D} = \text{diag}(1/d_j)$. Let the prior on $\Sigma_{\mathbf{G}}$ be $\Sigma_{\mathbf{G}}/\mathbf{G}, \mathbf{D} \sim \text{HIW}_{\mathbf{G}}(b, \rho \mathbf{I}_q)$. Then, integrating out $\Sigma_{\mathbf{G}}$,

$$\mathbf{YD}|\mathbf{G}, \mathbf{D} \sim \text{HMT}_{n \times q}(b, \mathbf{I}_n, \rho \mathbf{I}_q),$$

where HMT denotes a hyper-matrix t distribution (see Supplementary Section S.1 for more details). One can now use suitable mixing distributions on d_j and it is straightforward to perform MCMC to update $\mathbf{G}$ and $\mathbf{D}$, and to obtain samples from the conditional posterior of $(\Sigma_{\mathbf{G}}/\mathbf{Y}, \mathbf{G}, \mathbf{D})$. The missing edges in the inferred graph $\mathbf{G}$ points to conditional sign independence among possibly non-Gaussian continuous random variables. It is also possible to integrate out $\mathbf{D}$ to obtain the marginal of $\mathbf{Y}/\mathbf{G}$ up to a constant of proportionality, although the inferred $\mathbf{D}$ provides us knowledge of the marginal behavior through Lemma 1.

3 Inferring dependence structure across heterogeneous data types

In this section we consider the problem of network inference on mixed binary and continuous data. Let our data contain $\mathbf{Z} \in \{0, 1\}^d$ discrete and $\mathbf{Y} \in \mathbb{R}^q$ continuous variables for the same n samples (with the $d+q$ variables sharing the same dependence structure across all the n samples). A joint model for $\mathbf{X} = (\mathbf{Z}, \mathbf{Y})$ can be specified in terms of the conditionally Gaussian (CG) density of Lauritzen (1996) as follows:

$$f(\mathbf{x}) = f(\mathbf{z}, \mathbf{y}) = f(\mathbf{z})f(\mathbf{y}|\mathbf{z}) = \exp\left(g_{\mathbf{z}} + h_{\mathbf{z}}^T \mathbf{y} - \frac{1}{2} \mathbf{y}^T K_{\mathbf{z}} \mathbf{y}\right).$$

Define

$$P_{\mathbf{z}} = P(\mathbf{Z} = \mathbf{z}) = (2\pi)^{q/2} \{\det(K_{\mathbf{z}})\}^{-1/2} \exp(g_{\mathbf{z}} + h_{\mathbf{z}}^T K_{\mathbf{z}}^{-1} h_{\mathbf{z}}/2), \\ \xi_{\mathbf{z}} = E(\mathbf{Y}|\mathbf{Z} = \mathbf{z}) = K_{\mathbf{z}}^{-1} h_{\mathbf{z}}, \quad \Sigma_{\mathbf{z}} = \text{Var}(\mathbf{Y}|\mathbf{Z} = \mathbf{z}) = K_{\mathbf{z}}^{-1},$$

where the conditional distribution of $\mathbf{Y}/\mathbf{Z} = \mathbf{z}$ is $\text{N}(\xi_{\mathbf{z}}, \Sigma_{\mathbf{z}})$. It is possible to have a fairly general form for the tuple $(g_{\mathbf{z}}, h_{\mathbf{z}}, K_{\mathbf{z}})$ defining the distribution. Following Cheng et al. (2013), we consider a special case of the model

$$\log f(\mathbf{z}, \mathbf{y}) = \sum_{j=1}^d \lambda_j z_j + \sum_{\substack{j,k=1 \\ j>k}}^d \lambda_{jk} z_j z_k + \sum_{\gamma=1}^q \left(\sum_{j=1}^d \eta_j^{\gamma} z_j \right) y_{\gamma} - \frac{1}{2} \sum_{\gamma, \mu=1}^q y_{\gamma} k^{\gamma\mu} y_{\mu}. \quad (4)$$

Comparing with above, it is clear that we have

$g_{\mathbf{z}} = \sum_{j=1}^d \lambda_j z_j + \sum_{j>k} \lambda_{jk} z_j z_k; h_{\mathbf{z}}^T = \sum_{j=1}^d \eta_j^\gamma z_j$ and $K_{\mathbf{z}} = \{k^{\gamma\mu}\}$. Note also that our model is slightly simplified compared to Cheng et al. (2013), because $K_{\mathbf{z}}$ does not depend on the discrete variables, the case termed the “homogeneous model” by Lauritzen (1996). As pointed out by Cheng et al. (2013), this simplified model implies for $j, k \in \{1, \dots, d\}$ and $\gamma, \mu \in \{1, \dots, q\}$ that

$$Z_j \perp Z_k | \mathbf{X} \setminus \{Z_j, Z_k\} \iff \lambda_{jk} = 0; Z_j \perp Y_\gamma | \mathbf{X} \setminus \{Z_j, Y_\gamma\} \iff \eta_j^\gamma = 0; Y_\mu \perp Y_\gamma | \mathbf{X} \setminus \{Y_\mu, Y_\gamma\} \iff k^{\gamma\mu} = 0.$$

Thus, fitting this model allows one to infer conditional independence relationships across discrete and continuous variables. Note also that the model implies for $j = 1, \dots, d$ and $\gamma = 1, \dots, q$ the node conditional distributions

$$Z_j | \mathbf{X} \setminus Z_j \sim \text{Binomial} \left(n, \text{logit} \left(\sum_{\substack{k=1 \\ k \neq j}}^d \lambda_{jk} Z_k + \sum_{\gamma=1}^q \eta_j^\gamma Y_\gamma \right) \right), \quad (5)$$

$$Y_\gamma | \mathbf{X} \setminus Y_\gamma \sim N \left(\frac{1}{k^{\gamma\gamma}} \left(\sum_{j=1}^d \eta_j^\gamma Z_j - \sum_{\substack{\mu=1 \\ \mu \neq \gamma}}^q k^{\gamma\mu} Y_\mu \right), \frac{1}{k^{\gamma\gamma}} \right), \quad (6)$$

where $\text{logit}(\psi) = \{1 + \exp(-\psi)\}^{-1}$ for $\psi \in \mathbb{R}$. In the case of purely discrete or purely continuous data, the above conditional relationships correspond to a joint Ising distribution for discrete data and a joint multivariate Gaussian distribution for continuous data, respectively (Lauritzen, 1996). Directly maximizing the joint log likelihood in Equation (4) is known to be difficult (Lee and Hastie, 2015; Cheng et al., 2013). Thus, following the neighborhood selection approach of Meinhausen and Bühlmann (2006), existing works for pure discrete data fit penalized logistic regressions for the discrete part (e.g., Ravikumar et al., 2010) and penalized Gaussian regressions for the continuous part (e.g., Friedman et al., 2008) in high-dimensional settings to maximize the node conditional likelihoods (or pseudolikelihoods) of Equations (5–6). Building on these, Cheng et al. (2013) devised an alternating algorithm to simultaneously fit both types of regressions for mixed data. However, a rather surprising fact is that the binomial distribution can be written as a Gaussian location-scale mixture as well. We now show this allows a direct characterization of the joint density of $(\mathbf{Z}, \mathbf{Y})$ as a multivariate normal, conditional on mixing Pólya-Gamma variables for the discrete parts. To begin, note that if $U \sim \text{Binomial}(n, \text{logit}(\psi))$ then Polson et al. (2013) demonstrated the following location-scale mixture representation:

$$\left(U - \frac{n}{2}\right) | \omega \sim N(\omega\psi, \omega); \quad \omega \sim \text{PG}(n, 0),$$

where $\text{PG}(n, 0)$ denotes a Pólya-Gamma random variable, which can be expressed as an infinite weighted sum of Gamma random variables. Its density and moments are given by Polson et al. (2013) and an efficient sampler is available in the R package BayesLogit (Polson et al., 2012). Introducing latent Pólya-Gamma variables, Equations (5) and (6) become

$$\left(Z_j - \frac{n}{2}\right) | \omega_j, \mathbf{X} \setminus Z_j \sim N \left(\omega_j \left(\sum_{\substack{k=1 \\ k \neq j}}^d \lambda_{jk} Z_k + \sum_{\gamma=1}^d \eta_j^\gamma Y_\gamma \right), \omega_j \right), \quad \omega_j \stackrel{i.i.d}{\sim} \text{PG}(n, 0), \quad (7)$$

$$Y_\gamma | \mathbf{X} \setminus Y_\gamma \sim N \left(\frac{1}{k^{\gamma\gamma}} \left(\sum_{j=1}^d \eta_j^\gamma Z_j - \sum_{\substack{\mu=1 \\ \mu \neq 1}}^q k^{\gamma\mu} Y_\mu \right), \frac{1}{k^{\gamma\gamma}} \right). \quad (8)$$

One can now see from Equations (7) and (8) that all the $(d+q)$ node conditional distributions of one variable given the rest follow univariate normal distributions. By properties of multivariate normal, the joint distribution of the variables $(\mathbf{Z}, \mathbf{Y})$ given $\boldsymbol{\omega} = (\omega_1, \dots, \omega_d)$ must also correspond to a multivariate normal that will preserve these conditional means and variances (see, e.g., Khatri and Rao, 1976). Thus, define the transformed data

$$\tilde{\mathbf{X}} = (Z_1 - n/2, \dots, Z_d - n/2, Y_1, \dots, Y_q) | \boldsymbol{\omega} \sim \text{MN}_{n \times (d+q)}(\mathbf{0}, \mathbf{I}_n, \Sigma), \quad (9)$$

$$\omega_j \stackrel{i.i.d}{\sim} \text{PG}(n, 0) \quad \text{for } j=1, \dots, d. \quad (10)$$

Define $\lambda_{ji} = 1/\omega_j$. Then, the $(d+q) \times (d+q)$ symmetric Σ^{-1} is given by

$$\Sigma^{-1} = \begin{pmatrix} \lambda_{11} & \dots & -\lambda_{1d} & -\eta_1^1 & \dots & -\eta_1^q \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ -\lambda_{d1} & \dots & \lambda_{dd} & -\eta_d^1 & \dots & -\eta_d^q \\ -\eta_1^1 & \dots & -\eta_d^1 & k^{11} & \dots & k^{1q} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ -\eta_1^q & \dots & -\eta_d^q & k^{q1} & \dots & k^{qq} \end{pmatrix}.$$

The ω_j terms are independent and one can easily verify that $\int dp(\omega_j) < \infty$ when $\omega_j \sim \text{PG}(n, 0)$. Note that an inverse Wishart prior on Σ is not sensible any more because that will not induce inverse Pólya-Gamma priors on $(\lambda_{11}, \dots, \lambda_{dd})$. Thus in order to model this inverse covariance matrix, we follow the idea introduced by Wong et al. (2003), who decouple the modeling for the diagonal and off-diagonal elements. Write

$$\Omega = \Sigma^{-1} = \Theta \Gamma \Theta,$$

where Θ is a $(d+q)$ diagonal matrix with i th diagonal entry $\Theta_i = \sqrt{\Omega_{ii}}$ and Γ is related to Ω as $\Gamma_{ij} = -\Omega_{ij} / \sqrt{\Omega_{ii}\Omega_{jj}}$, i.e., the entries of Γ are the negative of the partial correlation matrix, with ones on the diagonal (Wong et al., 2003). Then, we parameterize

$$(\Theta_1^2, \dots, \Theta_d^2) = (\lambda_{11}, \dots, \lambda_{dd}) \sim 1/\text{PG}(n, 0), \quad (11)$$

$$(\Theta_{d+1}^2, \dots, \Theta_{d+q}^2) = (k^{11}, \dots, k^{qq}) \sim 1/\text{Inv-Gamma}(\alpha, \beta), \quad (12)$$

where all random variables are distributed independently and α, β are hyperparameters. We follow the same prior specification on the entries on Γ as Wong et al. (2003), which enables a sparse estimation of Γ . Thus, our parameterization differs from that of Wong et al. (2003)

only for the entries $(\Theta_1^2, \dots, \Theta_d^2)$ where they use Gamma priors, and we need to use inverted Pólya-Gamma priors. We conjecture that using the representation of Pólya-Gamma random variable as an infinite weighted sum of gamma random variables, it might be possible to characterize the induced distribution on Σ^{-1} more explicitly, although we have not pursued this. In any case, with this modification, one can employ the same MCMC sampling

procedure as in Wong et al. (2003) in order to iteratively update $(\Theta_i | \tilde{\mathbf{X}}, \Theta_{-i}, \Gamma)$ and $(\Gamma_{ij} | \tilde{\mathbf{X}}, \Theta, \Gamma_{-\{ij\}})$. Conditional independence holds according to off-diagonal zeros in inferred Γ , between the discrete-discrete, continuous-continuous or discrete-continuous random variables. Further note that we have assumed the continuous part of the data follows multivariate Gaussian distribution. An application of Proposition 1 shows that non-normal marginals can be modeled by appropriate choices of scale distributions for each marginal

$Y_1, \dots, Y_q$ and one would still be able to infer conditional sign independence. Contrast this with the framework of Cheng et al. (2013), which cannot handle non-normal marginals.

Following the well-known latent variable technique of Albert and Chib (1993) for probit models, the existing literature for Bayesian modeling of mixed data introduces a latent continuous counterpart for the observed discrete data for which posterior sampling is feasible (Pitt et al., 2006; Dobra and Lenkoski, 2011). Conditional independence is then inferred among the observed and latent continuous variables. Unfortunately, there is no direct characterization of the conditional independence relationship between the observed discrete data and their latent counterpart (Pitt et al., 2006). Our approach overcomes this through a direct scale transformation to infer dependence relationship at the level of the observed data.

4 Simulation study

We performed simulation experiments comparing the proposed method with competing approaches. We present the results for continuous non-Gaussian data and mixed discrete-continuous data in Sections 4.1 and 4.2 respectively.

4.1 Non-normal continuous data

We chose $n = 100$ and $q = 50$. We then simulated data according to the true inverse covariance matrix shown on the top left of Figure 2. The true Σ^{-1} is a symmetric banded diagonal matrix with diagonal elements equal to $v = 3$, the first sub-diagonal $= 0.25v = 0.75$ and the second subdiagonal $= -0.2v = -0.6$, the rest of the elements being zero. Thus, the true Σ^{-1} is sparse and there are both positive and negative partial correlations present. Positive definiteness of Σ^{-1} can be easily verified using the diagonal dominance property. We simulate data as $\mathbf{Y} \sim \text{MN}(\mathbf{0}, \mathbf{I}_n, \Sigma) \cdot \mathbf{D}$. Where $\mathbf{D} = \text{diag}(1/d_i)$ is a diagonal matrix with $d_i \sim \text{Exponential}(\text{mean} = 10)$ for $i = 1, \dots, 25$ and $d_i \sim \text{Inv-Gamma}(\text{shape} = 3, \text{scale} = 10)$ for $i = 26, \dots, 50$. Thus, the first 25 marginals in the observed data have double-exponential distribution while the remaining 25 have polynomially decaying t -distribution.

For this data, we compared four approaches: the proposed method based on Gaussian scale mixtures (GSM), alternative multivariate- t (Alt- t) of Finegold and Drton (2011), a sparse Bayesian Gaussian graphical model (GGM) as described in Supplementary Section S.1 and the Gaussian copula graphical model (GCGM) of Pitt et al. (2006). We implemented the first three methods in MATLAB and for GCGM we used the implementation in the R package BDgraph by Mohammadi and Wit (2015). GGM is implemented according to Supplementary Equations (S.2–S.4). For hyperparameters we used $b = 10$, $\rho = 0.5$ and prior weight $w_{uv} = 0.1$ for all edges in this example, but performed sensitivity analysis to ensure the choice of hyperparameters do not have a large effect on results. To implement Alt- t , we further put independent Inv-Gamma(2, 7) prior on all d_i . To implement GSM, we put independent Exponential(5) on the first 25 and Inv-Gamma(2, 7) on the rest. Results appear to be stable over a range of hyperparameter values. We used 50,000 MCMC iterations with a burn-in period of 20,000 iterations for all methods. MCMC diagnostics are presented in Supplementary Section S.5. Figure 2 shows the true and estimated Σ^{-1} for the first three methods (see Figure S.1 in the supplement for the estimate of GCGM). An interesting

observation is the scale next to each panel. It appears the Gaussian graphical model deals with different scaling across different marginals, for which it is a misspecified model, by heavily shrinking all entries of the resultant estimate of Σ^{-1} . On the other hand, the alternative- t , which expects polynomially decaying t marginals along all coordinates, appears to inflate the absolute values of some of the resulting estimates compared to the proposed method. Nevertheless, we remind the reader that the values of estimated Σ^{-1} are not directly comparable across the three methods, although their signs are. Table 1 reports the “sign concordance,” defined as the fraction of the elements of true Σ^{-1} signs correctly detected, by the competing methods. We report our results separately for zero, positive and negative elements; as well as the overall concordance for each method. In terms of overall concordance, the proposed approach has the best performance followed by GGM, Alt- t and GCGM.

For this data, we also tried non-Bayesian graphical lasso method, but it failed to converge after 5,000 iterations and we do not have numeric values to report. We also experimented with other sparse structures of the true Σ^{-1} . We considered structured cases, such as top left 5×5 off-diagonal block non-zero (half of them positive, the other half negative), rest off-diagonals zero; and unstructured cases, such as randomly selected 5% elements positive, 5% negative, rest 0, subject to the condition that this corresponds to a valid decomposable graph. Positive definiteness was ensured by diagonal dominance. The finding that the proposed method displays superior performance in sign detection remains robust. A larger simulation study with $n = 100$ and $q = 100$ is presented in Supplementary Section S.4.

4.2 Mixed binary and continuous data

Here we chose $n = 100$, $d = 9$ and $q = 41$. That is, we considered a total of 50 variables, the first 9 of them discrete and the remaining 41 continuous and there are 100 observations for each variable. The true inverse covariance matrix is shown in the top panel of Figure 3. The true Σ^{-1} is a symmetric banded diagonal matrix with diagonal elements equal to $v = 4$, the first sub-diagonal $= 0.2v = 0.8$ and the second subdiagonal $= -0.2v = -0.8$. In addition, we wanted to see if the method can successfully capture dependence between discrete and

continuous random variables. Thus, we set $\sum_{1:5,40:45}^{-1} = \sum_{40,45:1:5}^{-1} = -0.7$, introducing negative dependence. The mixed discrete and continuous data were then simulated according to the Equations (9–10). In order to create discrete observations, we rounded each entry of the first 9 columns to the nearest integer.

For estimation purposes, we compared the performance of GSM and GCGM. As in the previous subsection, we used native MATLAB implementation of GSM and the implementation in the package BDgraph for GCGM. To implement GSM, we used the parameterization in Equations (11–12). We simulated the required $PG(n, 0)$ random variables using the Bayeslogit package. For the hyperparameters, we used $\alpha = \beta = 1/2$ which appeared to work well in practice. As before, we used 50,000 MCMC iterations and a burn-in period of 20,000 iterations and monitored the log-likelihood to ensure convergence. The estimated Σ^{-1} by GSM is shown in the right panel of Figure 3 (see Figure S.2 in the supplement for the estimate of GCGM). The performance of GSM and GCGM in terms of capturing conditional sign dependence is reported in Table 2. Note that the alternative

multivariate- t and Gaussian graphical models are not suited for comparisons over mixed discrete-continuous data. Although GCGM of Pitt et al. (2006) can work with mixed discrete and continuous data, the interpretation of their estimated covariance matrix, which uses a latent continuous counterpart for the discrete variables, differs from ours which uses no such latent variable representation, other than the mixing Pólya-Gamma scale parameter. Nevertheless, it appears from Table 2 that GCGM does a poor job compared to GSM. Its sign concordance is lower for the zero entries as well as for the positive and negative entries. The estimate of GCGM is not as sparse as it should be, which is also apparent from Figure S.2. This finding of the behavior of GCGM is also consistent with Section 4.1, where it tends to produce a less sparse estimate compared to the other methods. Recall that both our approach (GSM) and GCGM can work with non-Gaussian distributions for the continuous data. Thus, although the data in this simulation uses normal marginals for continuous components, we experimented with non-normal marginals and the results remain quite robust.

5 Analysis of glioblastoma multiforme data

Our data consists of continuous expression levels and mutation status for 49 genes that overlap with the three critical signaling pathways - the RTK/PI3K signaling pathway, the p53 signaling pathway, and the Rb signaling pathway, which are known to be involved in migration, survival and apoptosis progression of cell cycles in GBM (Furnari et al., 2007). Of these 49 genes, 20 did not show evidence of mutation in any location. Thus, our data consists of $q = 49$ gene expressions and $d = 29$ binary mutations for $n = 103$ glioblastoma multiforme (GBM) patients. The raw data are publicly available through the Cancer Genome Atlas (TCGA) data portal (<http://tcga-data.nci.nih.gov/tcga/>). We standardize the continuous components by subtracting the mean and dividing by the standard deviation. In Figure 1, we provided an illustration of non-normal marginals in the continuous components by plotting the expression levels for AKT3 and CDK4 genes. These non-normal features are preserved under standardization. The complete list of genes whose expression levels and mutation status we consider is given in Supplementary Table S.1. More details on the GBM data set can be found in Supplementary Section S.6.

We illustrate in Figure 4 the conditional sign dependence network obtained by the proposed Gaussian scale mixture (GSM) method. Each connection represents a non-zero entry in the estimated inverse covariance matrix. Nodes with high connectivity appear closer to the center of the figure and those with lower degrees of connectivity are closer to the edges. A node with clear background and with a subscript “MUT_” denotes in the figure that the node corresponds to a binary mutation in a given gene; and a node with a solid background represents a continuous valued expression level. Several mutations show a high degree of negative association to other mutations and to expression levels of other genes. This includes the mutations in TP53 (negatively associated with mutations in MDM4, RB1, MET and to the expression level of PDGFRA), mutations in FGFR1 (negatively associated with mutations in PIK3R2, PIK3CB and positively to the expression levels of AKT1), mutations in PIK3R2 (negatively associated to mutations in FGFR1, ERBB2 and PIK3CB). Expression levels of IGF1R shows a high degree of connectivity (negatively to expression levels of PIK3CB, PTEN, CCND1).

On the other hand, some other expression levels appear isolated and do not appear to be connected to the other mutations and expressions under consideration. These include the expression levels of the MDM family (MDM2 and MDM4). It is interesting to note however that the mutations in the MDM family of genes are connected to other nodes, suggesting that this mutation acts by changing the expression levels of other genes (i.e., exhibits a *trans* effect). The influence of mutations in TP53 for GBM has been known to affect the prognosis (Shiraishi et al., 2002) and its reactivation via an MDM inhibitor has been observed (Costa et al., 2013), suggesting an interaction. Our analysis is in accordance with known pathway interactions in GBM (e.g., compare with Figure 4A of Brennan et al., 2013) and uncovers several new associations via joint analysis of binary and continuous valued data.

6 Conclusions

We proposed an approach based on Gaussian scale mixtures that is capable of handling the problem of network inference in presence of non-normal marginals and mixed discrete and continuous random variables in a unified framework. We introduced the concept of conditional sign independence and showed that it is possible to infer this based on the proposed method. By this measure, we showed by simulations that the proposed method performs better than alternatives such as copula Gaussian graphical models.

Some natural extensions of the proposed framework can be considered as future work. Prominent among them is the extension of the mixed binary/continuous framework in Section 3 to the mixed binary/ordinal/continuous case. In this case, the discrete variables would follow a multi-category logistic model instead of just two, and one may proceed using the framework of Polson et al. (2013) for multiple categories. Although for the purpose of this paper we are interested in Bayesian techniques, a scale mixture approach lends itself naturally to expectation-maximization (E–M) algorithms for maximizing likelihoods. If one is interested in estimating the inverse covariance matrix in a penalized likelihood framework, one can use our proposed framework where in the E-step instead of sampling Θ , one would substitute its conditional expectation given the rest, and simulation of $\mathbf{T}$ would be replaced by a penalized Gaussian likelihood maximization step, which is usually quite simple. For the special case of alternative multivariate- t , the E–M scheme was discussed by Finegold and Drton (2011). The current framework shows it is applicable more broadly, as long as one is able to compute the posterior expectations. This is especially promising for the case of mixed binary and continuous data, since Polson et al. (2013) provide very simple formulas for the expectation of Pólya-Gamma random variables. Thus, even in the non-Bayesian case, our proposed framework points to a possible alternative latent variable framework for implementing E–M to find the mle and it would be interesting to compare its performance to the pseudolikelihood approaches of Cheng et al. (2013) or Lee and Hastie (2015).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the Co-Editor, the AE and two anonymous referees for constructive suggestions. AB is supported by NSF grant DMS-1613063. AR is supported by a Research Scholar Grant RSG-16-005-01 from the American Cancer Society, an Institutional Research Grant from the MD Anderson Cancer Center (MDACC) and a Career Development Award from the MDACC Brain Tumor SPOR. VB is partially supported by NIH grants R01CA160736, R01CA19439, NSF grant DMS-1463233. AR and VB are also partially supported by the NIH through MDACC Support Grant P30CA016672.

References

- Albert JH, Chib S. Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*. 1993; 88:669–679.
- Barndorff-Nielsen OE. Exponentially Decreasing Distributions for the Logarithm of Particle Size. *Royal Society of London Proceedings Series A*. 1977; 353:401–419.
- Barndorff-Nielsen OE. Hyperbolic distributions and distributions on hyperbolae. *Scandinavian Journal of Statistics*. 1978; 5:151–157.
- Bhadra A, Mallick BK. Joint high-dimensional Bayesian variable and covariance selection with an application to eQTL analysis. *Biometrics*. 2013; 69:447–457. [PubMed: 23607608]
- Brennan CW, Verhaak RG, McKenna A, et al. The somatic genomic landscape of glioblastoma. *Cell*. 2013; 155:462–477. [PubMed: 24120142]
- Carvalho CM, Massam H, West M. Simulation of hyper-inverse Wishart distributions in graphical models. *Biometrika*. 2007; 94:647–659.
- Chen S, Witten DM, Shojaie A. Selection and estimation for mixed graphical models. *Biometrika*. 2015; 102:47–64. [PubMed: 27625437]
- Cheng J, Levina E, Zhu J. High-dimensional mixed graphical models. *arXiv preprint arXiv:1304.2810*. 2013
- Costa B, Bendinelli S, Gabelloni P, et al. Human glioblastoma multiforme: p53 reactivation by a novel MDM2 inhibitor. *PLoS One*. 2013; 8:e72281. [PubMed: 23977270]
- Dawid AP. Some matrix-variate distribution theory: Notational considerations and a Bayesian application. *Biometrika*. 1981; 68:265–274.
- Dobra A, Lenkoski A. Copula Gaussian graphical models and their application to modeling functional disability data. *Ann Appl Stat*. 2011; 5:969–993.
- Feldman G, Bhadra A, Kirshner S. Bayesian feature selection in high-dimensional regression in presence of correlated noise. *Stat*. 2014; 3:258–272.
- Finegold M, Drton M. Robust graphical modeling of gene networks using classical and alternative t -distributions. *Ann Appl Stat*. 2011; 5:1057–1080.
- Finegold M, Drton M. Robust Bayesian graphical modeling using dirichlet t -distributions. *Bayesian Analysis*. 2014; 9:521–550.
- Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*. 2008; 9:432–441. [PubMed: 18079126]
- Furnari FB, Fenton T, Bachoo RM, et al. Malignant astrocytic glioma: genetics, biology, and paths to treatment. *Genes & development*. 2007; 21:2683–2710. [PubMed: 17974913]
- Gray JW, Collins C. Genome changes and gene expression in human solid tumors. *Carcinogenesis*. 2000; 21:443–452. [PubMed: 10688864]
- Griffin JE, Brown PJ. Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*. 2010; 5:171–188.
- Kendall MG. A new measure of rank correlation. *Biometrika*. 1938; 30:81–93.
- Khatri C, Rao CR. Characterizations of multivariate normality. I. through independence of some statistics. *Journal of Multivariate Analysis*. 1976; 6:81–94.
- Kotz, S., Kozubowski, TJ., Podgrski, K. The Laplace distribution and generalizations. Birkhäuser; Boston: 2001.
- Lauritzen, SL. Graphical Models. Oxford University Press; Oxford: 1996.

- Lee JD, Hastie TJ. Learning the structure of mixed graphical models. *Journal of Computational and Graphical Statistics*. 2015; 24:230–253. [PubMed: 26085782]
- Liu, H., Han, F., Zhang, CH. Transelliptical graphical models. In: Bartlett, PL, Pereira, FCN, Burges, CJC, Bottou, L., Weinberger, KQ., editors. *NIPS*. 2012. p. 809-817.
- Liu H, Lafferty J, Wasserman L. The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *J Mach Learn Res*. 2009; 10:2295–2328.
- Marko NF, Weil RJ. Non-Gaussian distributions affect identification of expression patterns, functional annotation, and prospective classification in human cancer genomes. *PLoS ONE*. 2012; 7:e46935. [PubMed: 23118863]
- Meinhausen N, Bühlmann P. High-dimensional graphs and variable selection with the lasso. *Annals of Statistics*. 2006; 34:1436–1462.
- Mohammadi A, Wit EC. BDgraph: Bayesian structure learning of graphs in R. *arXiv preprint arXiv: 1501.05108*. 2015
- Pitt M, Chan D, Kohn R. Efficient Bayesian inference for Gaussian copula regression models. *Biometrika*. 2006; 93:537–554.
- Polson NG, Scott JG, Windle J. R package BayesLogit. 2012
- Polson NG, Scott JG, Windle J. Bayesian inference for logistic models using Pólya-Gamma latent variables. *J Am Stat Assoc*. 2013; 108:1339–1349.
- Ravikumar P, Wainwright MJ, Lafferty JD. High-dimensional Ising model selection using ℓ_1 -regularized logistic regression. *Ann Statist*. 2010; 38:1287–1319.
- Shiraishi S, Tada K, Nakamura H, et al. Influence of p53 mutations on prognosis of patients with glioblastoma. *Cancer*. 2002; 95:249–257. [PubMed: 12124823]
- TCGA. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 2008; 455:1061–1068. [PubMed: 18772890]
- West M. On scale mixtures of normal distributions. *Biometrika*. 1987; 74:646–648.
- Wong F, Carter CK, Kohn R. Efficient estimation of covariance selection models. *Biometrika*. 2003; 90:809–830.
- Yang Y, Ravimumar P, Allen G, Liu Z. On Graphical Models via Univariate Exponential Family Distributions. *Journal of Machine Learning Research*. 2015; 16:3813–3847. [PubMed: 27570498]

Figure 1.

An illustration of non-normal marginals in genomic data. Normal q-q plots for the marginals of AKT3 and CDK4 expression levels based on TCGA glioblastoma samples, clearly demonstrating non-Gaussian tails. These two genes have been implicated in glioblastoma by TCGA (2008).

Figure 2.

True and estimated Σ^{-1} for continuous non-normal data. Clockwise from top left: true, estimated by proposed method using Gaussian scale mixtures (GSM), by Gaussian graphical model (GGM) and by alternative multivariate- t (Alt- t). This figure appears in color in the electronic version of this article.

Figure 3.

True and estimated Σ^{-1} for mixed discrete and continuous data. Left: true, right: estimated by the proposed Gaussian scale mixture (GSM) method. This figure appears in color in the electronic version of this article.

Author Manuscript

Table 1

Sign concordance, defined as the fraction of signs of the elements of Σ^{-1} correctly recovered by the competing methods, for continuous non-normal data for $n = 100$, $q = 50$.

Method	Sign Concordance (Zero) (# True Zero = 2256)	Sign Concordance (+) (# True + = 148)	Sign Concordance(-) (# True - = 96)	%Overall Concordance
GSM	0.9796	0.8243	0.4792	0.9512
Alt-t	0.9486	0.8378	0.5833	0.928
GGM	0.9761	0.7297	0.3958	0.9392
GCGM	0.9464	0.8176	0.5104	0.922

Table 2

Sign concordance, defined as the fraction of signs of the elements of Σ^{-1} correctly recovered by the competing methods, for mixed discrete-continuous data for $n = 100$, $d = 9$ and $q = 41$.

Method	Sign Concordance (Zero) (# True Zero = 2256)	Sign Concordance (+) (# True + = 148)	Sign Concordance (-) (# True - = 96)	%Overall Concordance
GSM	0.9516	0.8919	0.625	0.9356
GCGM	0.9219	0.8108	0.5313	0.9004