The Canadian Journal of Statistics
Vol. 42, No. 2, 2014, Pages 185–203
La revue canadienne de statistique

Graphical lassos for meta-elliptical distributions

Martin BILODEAU*

Department of Mathematics and Statistics, University of Montreal, P. O. Box 6128, Station Centre-ville, Montreal, Canada H3C 3J7

Key words and phrases: Graphical models; high dimensional statistics; meta-elliptical distributions; meta-Gaussian distributions; partial correlations; receiver operating characteristic; sparse graphs.

MSC 2010: Primary 62H12; secondary 62H20

Abstract: Gaussian graphical lasso is a tool for estimating sparse graphs using a Gaussian log-likelihood with an ℓ_1 penalty on the inverse covariance matrix. This paper proposes a generalization to meta-elliptical distributions. Conditional uncorrelatedness is characterized in meta-elliptical families. The proposed meta-elliptical and re-weighted Kendall graphical lassos are computed from pseudo-observations which are functions of ranks of observations. They are invariant to strictly increasing transformations of the variables and do not assume the existence of moments. Simulations of receiver operating characteristic curves show noticeable improvements (in comparison with graphical lassos designed for meta-Gaussian distributions) for distributions which are not meta-Gaussian. These improvements are realized without ill effects when the distribution is meta-Gaussian. Deterministic and random contaminations of data are used to verify the robustness of the re-weighted Kendall graphical lasso. *The Canadian Journal of Statistics* 42: 185–203; 2014 © 2014 Statistical Society of Canada

Résumé: Le lasso graphique gaussien est un estimateur de graphe épars basé sur la log vraisemblance comportant une pénalité ℓ_1 sur l'inverse de la matrice de covariance. L'auteur propose une généralisation aux distributions méta-elliptiques. La non-corrélation conditionnelle est caractérisée dans les familles méta-elliptiques. Deux lassos graphiques sont proposés : le lasso méta-elliptique et le lasso de Kendall repondéré, tous deux calculés à partir de pseudo-observations basées sur les rangs. Ils sont invariants aux transformations strictement monotones croissantes et ne présupposent l'existence d'aucun moment. Dans le cadre de simulations, ils offrent une performance (en termes de courbe ROC) comparable aux lassos spécifiques aux distributions méta-gaussiennes lorsque les données suivent cette distribution. Une amélioration notable est cependant observée lorsque les données ne suivent pas la distribution méta-gaussienne. La robustesse du lasso de Kendall repondéré est aussi illustrée au moyen de données contaminées de manière aléatoire ou déterministe. *La revue canadienne de statistique* 42: 185–203; 2014 © 2014 Société statistique du Canada

1. INTRODUCTION

Some methods have been proposed for non-Gaussian and robust Gaussian graphical models. Vogel & Fried (2011) assumed an elliptical distribution. The only elliptical distribution for which components may be independent is the Gaussian distribution. Hence, they proposed the concept of conditional uncorrelatedness, in lieu of conditional independence, to identify edges of a graph. Estimation methods proposed to estimate the scatter matrix are the sample covariance, with a kurtosis adjustment, and the robust Tyler's M-estimator. Robust Gaussian graphical modelling in Becker (2005) and Gottard & Pacillo (2010) is done by replacing the sample covariance matrix by the re-weighted minimum covariance determinant estimator. This estimator is hard to compute in

* *Author to whom correspondence may be addressed.*
E-mail: martin.bilodeau@umontreal.ca

high dimensions. Miyamura & Kano (2006) proposed the use of an alternative M-type estimator. Regardless of the estimator of the scatter or covariance matrix chosen, computations of partial correlations are done via the well-known formula involving the inverse of the estimator. Hence, they are applicable when the number of observations n is greater than the number of variables p.

Finegold & Drton (2011) used multivariate t distributions as model. The estimation method is an iterative EM (expectation-maximization) algorithm in which a graphical lasso problem is solved at each iteration. As in Vogel & Fried (2011), conditional uncorrelatedness is used to identify edges. An alternative t distribution is also investigated for which, unfortunately, conditional uncorrelatedness is no longer implied by the nullity of a certain parameter of partial correlation. According to Finegold & Drton (2011), "simple transformations of the data may be effective at minimizing the effect of outliers or contaminated data on a small scale." They further wrote that "a normal quantile transformation, in particular, appears to be effective in many cases." They did not, however, elaborate any further on this approach. For meta-Gaussian distributions, Liu, Lafferty, & Wasserman (2009) proposed to input the Gaussian scores rank correlation matrix into the Gaussian graphical lasso. Liu et al. (2012) replaced the Gaussian scores rank correlation matrix with the back transformed matrix of Spearman's or Kendall's rank correlations for more robustness and still high efficiency at meta-Gaussian distributions.

Another approach to graphical modelling is shrinkage of the sample covariance matrix; see Ledoit & Wolf (2004), Chen et al. (2010) and Schäfer & Strimmer (2005). The shrinkage estimator is usually a convex combination of the sample covariance (or correlation) matrix and a target in the form of a multiple of the identity matrix. The shrinkage factor is determined optimally to minimize the expected mean squared error. An advantage of shrinkage estimators for computing partial correlations is their non-singularity when $n \leq p$. However, they are based on the sample covariance matrix which is a poorly efficient estimate of covariance for non-Gaussian models, especially for distributions with heavy tails. Even in classical asymptotic theory (fixed p and $n \to \infty$), the efficiency of the sample covariance matrix relative to the maximum likelihood estimator, obtained by Tyler (1983), is roughly 9% for $p = 20$ when the distribution is multivariate t with $\nu = 5$ degrees of freedom. This is due to the poor robustness properties of the sample covariance matrix.

Yet another simple approach is to estimate a sparse graphical model by fitting a linear regression to each variable, using all remaining variables as predictors. The graph has the undirected edge (i, j) if the estimated coefficient of variable i on j and the estimated coefficient of variable j on i are non-zero. Meinshausen & Bühlmann (2006) proposed the lasso regression and they established that asymptotically, this consistently estimates the edges of the graph. Yuan & Lin (2007) showed that this simple approach can be viewed as an approximation to the exact maximization of the ℓ_1 penalized likelihood. Tenenhaus et al. (2010) adopted partial least squares regression. These regression methods do not consider the positive definite constraint. Graphical lasso is more efficient because of the inclusion of the positive definite constraint and the use of likelihood (Yuan & Lin, 2007).

This paper proposes two graphical lassos adapted to meta-elliptical distributions. The univariate marginal distribution functions are assumed continuous without any moment restrictions. Non-edges of an undirected graphical model are interpreted in terms of conditional uncorrelatedness assessed with a correlation measure invariant to strictly increasing transformations such as Spearman's rho or Kendall's tau. Conditional uncorrelatedness is characterized with the inverse of the linear correlation matrix. The estimator is obtained from an estimating equation similar to the sub-gradient of the Gaussian graphical lasso. It can be computed by iterating the Gaussian graphical lasso of Friedman, Hastie, & Tibshirani (2008) in a manner inspired by the fixed point algorithm of Kent & Tyler (1991) used to estimate the scatter matrix of an elliptical distribution, and without recourse to an EM algorithm. The computational burden of the two proposed graphical lassos is reduced to a feasible level by taking as an initial estimate the Kendall graphical lasso of Liu et al. (2012) and performing only one re-weighting iteration.

The paper is structured as follows. The proposed graphical lassos are introduced in Section 2 with an example of closing prices from stocks in the S & P 500 market. Elliptical distributions are reviewed in Section 3. Meta-elliptical distributions are reviewed in Section 4 and a characterization is given to identify non-edges in undirected graphs. The meta-elliptical graphical lasso and the re-weighted Kendall graphical lasso are motivated in Section 5 and their relations to the meta-Gaussian graphical lasso of Liu, Lafferty, & Wasserman (2009) and the t graphical lasso of Finegold & Drton (2011) are explained. Finally, a simulation of receiver operating characteristic curves is done in Section 6. The gain in efficiency when the distribution is meta-elliptical is achieved without ill effects when the distribution is meta-Gaussian. Simulations of the re-weighted Kendall graphical lasso also show that it can be used as a safe replacement to the Kendall graphical lasso of Liu et al. (2012). It requires roughly twice the amount of computations.

2. THE META-ELLIPTICAL AND THE RE-WEIGHTED KENDALL GRAPHICAL LASSOS

Since all lasso estimators considered in this paper are for graphical models, the qualifier 'graphical' will be omitted from now on in the expression graphical lasso. First, I briefly review the Gaussian lasso. Assume the random vector $X = (X^{(1)}, \ldots, X^{(p)})$ follows a Gaussian distribution with a positive definite correlation matrix R. Following Whittaker (1990), Cox & Wermuth (1996), or Lauritzen (1996), each graphical model is associated with an undirected graph $G = (V, E)$ with vertex set $V = \{1, \ldots, p\}$, and defined by requiring that for each non-edge $(i, j) \notin E$, the variables $X^{(i)}$ and $X^{(j)}$ are conditionally independent given all remaining variables. This conditional independence holds if and only if $\theta_{ij} = 0$, where θ_{ij} is the element in position (i, j) of $\Theta = R^{-1}$. Therefore, determining the edges of a graph is equivalent to determining the non-zero elements of Θ. For an estimate $\hat{\theta}_{ij}$, a false positive occurs when $(i, j) \notin E$ and $\hat{\theta}_{ij} \neq 0$; similarly, a true positive is when $(i, j) \in E$ and $\hat{\theta}_{ij} \neq 0$.

Consider a sample $X_l = (X_l^{(1)}, \ldots, X_l^{(p)})$ $(l = 1, \ldots, n)$ of n independent observations from the Gaussian distribution. The Gaussian lasso is the solution to the ℓ_1 penalized Gaussian log-likelihood optimization of

$$\min_{\Theta \succeq 0} - \log \det \Theta + \mathrm{tr}(S\Theta) + \lambda ||\Theta||_1 \tag{1}$$

over positive semidefinite matrices $\Theta \succeq 0$, where S is the sample covariance matrix. Here tr denotes the trace and $||\Theta||_1 = \sum_{i,j} |\theta_{ij}|$ is the ℓ_1 norm. Larger values of the regularization parameter λ lead to more θ_{ij} being estimated as zero. A slightly different problem in which the diagonal elements of Θ are not penalized is obtained by substituting $S - \lambda I$ for S in problem (1). For example, it may not be desirable to penalize diagonal elements when S is replaced by a correlation matrix. This optimization problem can be solved with the algorithms `dpglasso` of Mazumder & Hastie (2012a) and `glasso` of Friedman, Hastie, & Tibshirani (2008). However, the latter occasionally fails to converge with warm starts which may happen when computing a path of solutions over a grid of regularization parameters λ. Cross-validation using regression or likelihood approaches can be used to select λ (Friedman, Hastie, & Tibshirani, 2008).

Problem (1) is a convex optimization problem in the variable Θ (Boyd & Vandenberghe, 2004). A necessary and sufficient condition for Θ to be a solution (Witten, Friedman, & Simon (2011)) is that it satisfies

$$W - S - \lambda \Gamma(\Theta) = 0, \tag{2}$$

where $W = \Theta^{-1}$, $\Gamma(\Theta) : p \times p$ is a matrix whose (i, j) element is $\gamma_{ij} = \text{sign}(\theta_{ij})$, if $\theta_{ij} \neq 0$, and $\gamma_{ij} \in [-1, 1]$, if $\theta_{ij} = 0$. Since $\theta_{ii} > 0$, then $w_{ii} = s_{ii} + \lambda$. Hence, if a correlation matrix $\hat{R}$ is used as input for S and diagonal elements are not penalized, then the output Θ^{-1} is a positive definite correlation matrix.

In this paper, I propose the meta-elliptical lasso and a variant, the re-weighted Kendall lasso. Let g be the known density generator of the meta-elliptical distribution introduced later in Definition 1 of Section 4 and let F be the corresponding univariate distribution function given by Equation (3). A weight function defined with the density generator is $u(s) = -2g'(s)/g(s)$. The meta-elliptical lasso estimator of Θ is the matrix Θ_2 obtained as follows:

1. Compute pseudo-observations $\hat{Z}_l$ with components

   $$\hat{Z}_l^{(i)} = F^{-1}\left\{R_l^{(i)}/(n+1)\right\},$$

   where $R_l^{(i)}$ is the rank of $X_l^{(i)}$ among $X_1^{(i)}, \ldots, X_n^{(i)}$.
2. Compute the back transformed Kendall correlation matrix $\hat{R}_1 = \left(\sin(\pi\hat{\tau}_{ij}/2)\right)$, where $\hat{\tau}_{ij}$ is the Kendall correlation computed from pseudo-observations.
3. Obtain $\Theta_1 = \arg\min_{\Theta \succeq 0} -\log\det\Theta + \text{tr}(\hat{R}_1\Theta) + \lambda||\Theta||_1$.
4. Compute the covariance matrix using re-weighted pseudo-observations $\sqrt{u(s_l)}\hat{Z}_l$,

   $$\frac{1}{n}\sum_{l=1}^{n}(\sqrt{u(s_l)}\hat{Z}_l)(\sqrt{u(s_l)}\hat{Z}_l^{\mathrm{T}}) = \frac{1}{n}\sum_{l=1}^{n}u(s_l)\hat{Z}_l\hat{Z}_l^{\mathrm{T}}$$
   $$= \frac{1}{n}\sum_{l=1}^{n}u(\hat{Z}_l^{\mathrm{T}}\Theta_1\hat{Z}_l)\hat{Z}_l\hat{Z}_l^{\mathrm{T}},$$

   where $s_l = \hat{Z}_l^{\mathrm{T}}\Theta_1\hat{Z}_l$, and obtain the corresponding Pearson correlation matrix denoted $\hat{R}_2$.
5. Obtain $\Theta_2 = \arg\min_{\Theta \succeq 0} -\log\det\Theta + \text{tr}(\hat{R}_2\Theta) + \lambda||\Theta||_1$.

Some comments on this algorithm are in order. Components of pseudo-observations are elliptical scores similar to better known Gaussian scores when $F = \Phi$. Kendall correlations from pseudo-observations $\hat{Z}_l$ in Step 2 or from observations X_l are identical. The sine transformation in Step 2 is made because of the relation $\tau_{ij} = (2/\pi)\arcsin(r_{ij})$ between Kendall correlations and linear correlations for elliptical distributions; see Section 4 for details. The preliminary estimate Θ_1 is thus the Kendall lasso of Liu et al. (2012). Squared Mahalanobis distances for all observations are then computed with Θ_1 and used to re-weight pseudo-observations. The final estimate Θ_2 is obtained from the Gaussian lasso using the Pearson correlation matrix from re-weighted pseudo-observations. As for the re-weighted Kendall lasso, the only difference is that, in Step 4 above, Pearson correlations are replaced by back transformed Kendall correlations. The meta-elliptical lasso has a greater efficiency at meta-elliptical distributions (over other lassos for meta-Gaussian distributions) in terms of receiver operating characteristic curves. However, it is poorly robust due to the use of Pearson correlations. The re-weighted Kendall lasso takes aim at greater efficiencies at meta-elliptical distributions while being more robust.

2.1. Analysis of Stocks from the S & P 500 Market

As an application, consider the `stockdata` of the `R` package `huge`. It represents closing prices from all stocks in the S & P 500 for all days that the market was open between January 1, 2003 and January 1, 2008. This gives 1,258 observations for the 452 stocks that remained in the S & P 500 during the entire time period. The data have been preprocessed by calculating the log-ratio of

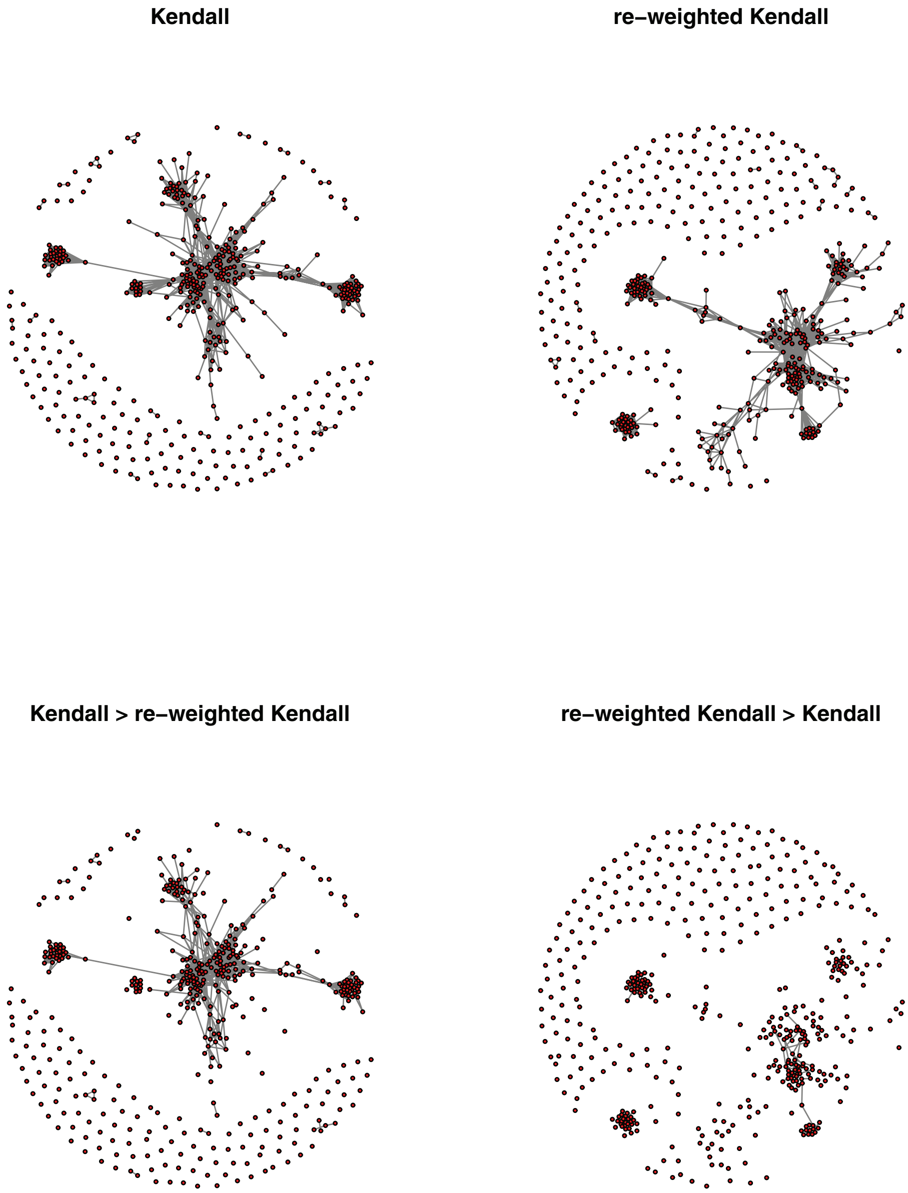

FIGURE 1: Kendall and re-weighted Kendall lassos (top) with $\lambda = 0.5$ yielded, respectively, 2,346 and 1,731 edges for sparsity levels of 2.3% and 1.7%. Kendall > re-weighted Kendall (bottom) means edges are present in Kendall's graph but not in the other, and vice versa. [Colour figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

the price at time t to price at time $t-1$. Hence, the data analysed has $n = 1{,}257$ observations and $p = 452$ variables. I compare the Kendall lasso of Liu, Lafferty, & Wasserman (2009) with the re-weighted Kendall lasso with the weight function $u(s) = (1+p)/(1+s)$ of a multivariate Cauchy distribution. The two lassos used a regularization parameter $\lambda = 0.5$ yielding 2,346 edges for the Kendall lasso and 1,731 edges for the re-weighted Kendall lasso corresponding to sparsity levels of about 2.3% and 1.7%, respectively. The two graphs on top of Figure 1 have 1,692 common edges, but 654 edges in the Kendall graph are absent from the re-weighted Kendall graph, and

39 edges in the re-weighted Kendall graph are absent from the Kendall graph. The graphs of discordant edges are the two graphs at the bottom of Figure 1. The simulations in Section 6 show that re-weighting does not damage the efficiency of the Kendall lasso when the distribution is meta-Gaussian. However, a substantial gain of efficiency is possible by re-weighting when the distribution has a Cauchy copula, or an intermediate t copula with small degrees of freedom. A sensible practice may be to fit both lassos and pay additional attention to the discordant edges.

This application was done with `glasso` version 1.7 because, and contrary to `dpglasso`, it includes thresholding of the entries of the correlation matrix. This strategy found in Mazumder & Hastie (2012b) and Witten, Friedman, & Simon (2011) reduces the amount of computations to solve Equation (2) by decomposing the graph into smaller graphs of connected components. The graphs were produced with the package `igraph` version 0.6.5-2. I used the function `cor.fk` of the `pcaPP` package which implements the algorithm of Knight (1966) to compute the Kendall correlation matrix in $O(p^2 n \log n)$ operations, rather than $O(p^2 n^2)$ for the function `cor` of the `stats` package.

Before embarking on meta-elliptical distributions as graphical models some basic properties of elliptical distributions are now reviewed.

3. ELLIPTICAL DISTRIBUTIONS

An absolutely continuous random vector $Z = (Z^{(1)}, \ldots, Z^{(p)})$, with location parameter 0, is said to be elliptically contoured with positive definite scatter matrix Σ if it admits a density of the form

$$h(z) = (\det \Sigma)^{-1/2} g\left(z^{\mathrm{T}} \Sigma^{-1} z\right).$$

A change of variables to polar coordinates establishes that h is a density if g is a non-negative function satisfying

$$\int_0^\infty r^{p-1} g(r^2)\,\mathrm{d}r = \frac{\Gamma(p/2)}{2\pi^{p/2}}.$$

Examples of elliptical distributions are the Gaussian distribution, $g(s) \propto \exp(-s/2)$, the t distribution with ν degrees of freedom, $g(s) \propto (1 + s/\nu)^{-(\nu+p)/2}$, and the power exponential family, $g(s) \propto \exp(-s^\alpha/2)$, for $\alpha > 0$.

If $\Sigma = (\sigma_{ij})$, the quantity $r_{ij} = \sigma_{ij}/(\sigma_{ii}^{1/2}\sigma_{jj}^{1/2})$ is defined as the linear correlation coefficient between variables i and j. The variables $Z^{(i)}/\sigma_{ii}^{1/2}$ are then identically distributed with a common distribution function (Fang, Fang, & Kotz, 2002)

$$F(z) = \frac{1}{2} + \frac{\pi^{(p-1)/2}}{\Gamma\left[(p-1)/2\right]} \int_0^z \int_{s^2}^\infty (t - s^2)^{(p-1)/2-1} g(t)\,\mathrm{d}t\,\mathrm{d}s, \tag{3}$$

and their joint distribution is elliptical with density

$$h(z) = (\det R)^{-1/2} g\left(z^{\mathrm{T}} R^{-1} z\right), \tag{4}$$

where $R = (r_{ij})$ is the linear correlation matrix. The cumulative distribution function corresponding to the density h in Equation (4) is denoted H. For the $N_p(0, R)$ distribution, $F = \Phi$ corresponds to a $N(0, 1)$, and for the t distribution with ν degrees of freedom, location 0, and linear correlation matrix R, denoted $t_{p,\nu}(0, R)$, $F = F_\nu$ corresponds to the univariate t distribution with ν degrees of freedom. Hult & Lindskog (2002) emphasized that the linear correlation coefficient is an extension of the usual definition in terms of variances and covariances. The linear correlation coefficient should be interpreted as a scalar measure of dependence and, as such, it should not rely on finiteness of certain moments.

3.1. Conditional Distribution of Elliptical Distributions

Of importance to the concept of conditional uncorrelatedness is the fact that conditional distributions of elliptical distributions are again elliptical. Assume $(Z^{(1)}, \ldots, Z^{(p)})$ is elliptical with the density h in Equation (4). The conditional distribution of $(Z^{(1)}, Z^{(2)})$, given all remaining variables, is elliptical with location

$$\delta = R_{12}R_{22}^{-1}\left(z^{(3)}, \ldots, z^{(p)}\right)^{\mathrm{T}} \equiv (\delta_1, \delta_2)^{\mathrm{T}} \tag{5}$$

and scatter matrix

$$R_{11.2} = R_{11} - R_{12}R_{22}^{-1}R_{21} \equiv \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \tag{6}$$

defined from the partitioned matrix

$$R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}.$$

For a fixed value of the conditioning variable, the standardized variables $(Z^{(i)} - \delta_i)/\gamma_{ii}^{1/2}$ have the same distribution function, denoted $\tilde{F}$, for $i = 1, 2$. For example, the $N_p(0, R)$ distribution has $F = \tilde{F} = \Phi$, whereas the $t_{p,\nu}(0, R)$ distribution has $F = F_\nu$ and $\tilde{F} = F_{\nu+p-2}$. The linear correlation in this conditional distribution is the partial linear correlation denoted r^{12}. Computationally, since the conditional scatter matrix is the same as for Gaussian distributions, the partial linear correlation is obtained by the same formula as the one for the partial Pearson correlation, specifically

$$r^{12} = -\frac{\theta_{12}}{\theta_{11}^{1/2}\theta_{22}^{1/2}}, \tag{7}$$

where $\Theta = (\theta_{ij})$ is the matrix R^{-1}.

The reader should be reminded that this simple formula is a direct consequence of the expression for the inverse of a partitioned matrix. For later use, since

$$R_{11.2} = \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{pmatrix}^{-1} \propto \begin{pmatrix} \theta_{22} & -\theta_{12} \\ -\theta_{21} & \theta_{11} \end{pmatrix}, \tag{8}$$

the linear correlation computed from $R_{11.2}$ is given by Equation (7). In general, $r^{ij} = -\theta_{ij}/(\theta_{ii}^{1/2}\theta_{jj}^{1/2})$ is the partial linear correlation between $Z^{(i)}$ and $Z^{(j)}$.

4. META-ELLIPTICAL DISTRIBUTIONS

A distribution function C on the unit cube $[0, 1]^p$ with uniform marginal distributions is called a copula. Sklar (1959) links an arbitrary multivariate distribution function K to a copula function via the marginal distribution functions $K_1,\ldots,K_p$. Suppose K is a multivariate distribution function with univariate marginal distribution functions $K_1,\ldots,K_p$. Then there is a copula C such that

$$K(x_1, \ldots, x_p) = C\left\{K_1(x_1), \ldots, K_p(x_p)\right\}. \tag{9}$$

If K is continuous, then the copula C is unique and is

$$C(u_1, \ldots, u_p) = K\left\{K_1^{-1}(u_1), \ldots, K_p^{-1}(u_p)\right\},$$

for $u = (u_1, \ldots, u_p)$ in $(0, 1)^p$, where $K_i^{-1}(u) = \inf\{x : K_i(x) \geq u\}$ $(i = 1, \ldots, p)$. Conversely, if C is a copula on $[0, 1]^p$ and $K_1, \ldots, K_p$ are univariate distribution functions, then the function K in Equation (9) is a multivariate distribution with univariate marginal distributions $K_1, \ldots, K_p$.

Models for multivariate analysis can be produced at will by specifying independently a copula, which contains all the information concerning the dependence among variables, and the marginal distributions. The copula associated with an elliptical distribution is termed an elliptical copula.

As a first example, the t copula with ν degrees of freedom and linear correlation matrix R is

$$C_{p,\nu,R}(u_1, \ldots, u_p) = H_{p,\nu,R}\left\{F_\nu^{-1}(u_1), \ldots, F_\nu^{-1}(u_p)\right\},$$

where $H_{p,\nu,R}$ is the distribution function of the $t_{p,\nu}(0, R)$ distribution, whose density is

$$h(z) = c_{p,\nu}(1 + z^{\mathrm{T}} R^{-1} z/\nu)^{-(\nu+p)/2}$$

for some constant $c_{p,\nu}$, and F_ν is the univariate distribution function of the t distribution with ν degrees of freedom.

A second example is the Gaussian copula

$$C_{p,R}(u_1, \ldots, u_p) = \Phi_{p,R}\left\{\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_p)\right\}, \tag{10}$$

where $\Phi_{p,R}$ is the distribution function of the $N_p(0, R)$ distribution.

Investigations on meta-elliptical distributions were initiated by Fang, Fang, & Kotz (2002) and their dependence properties studied further by Abdous, Genest, & Rémillard (2005).

Definition 1. *The random vector $X = (X^{(1)}, \ldots, X^{(p)})$ with continuous marginals K_i $(i = 1, \ldots, p)$ is meta-elliptically distributed with density generator g, and positive definite linear correlation matrix R, if the joint distribution of the variables $Z^{(i)} = F^{-1}\left\{K_i(X^{(i)})\right\}$ $(i = 1, \ldots, p)$, where F is given by Equation (3), is elliptical with density h given by Equation (4).*

When h is the $N_p(0, R)$ density, the resulting distribution is the meta-Gaussian distribution due to Kelly & Krzysztofowicz (1997). Liu, Lafferty, & Wasserman (2009) defined nonparanormal distributions. The copula of a nonparanormal distribution with monotone functions is the Gaussian copula. This means that meta-Gaussian and nonparanormal distributions constitute only one family.

Results on Kendall's tau and Spearman's rho correlation coefficients in dimension two are now presented. Without any condition on moments, they are

$$\tau = 4E\left\{H(Z^{(1)}, Z^{(2)})\right\} - 1$$

$$\rho = 12E\left\{F(Z^{(1)})F(Z^{(2)})\right\} - 3,$$

where F is the common marginal in H; see Equations (3) and (4).

Let r be the linear correlation coefficient of a meta-elliptical distribution of dimension 2. In Lindskog, McNeil, & Schmock (2003) and Fang, Fang, & Kotz (2002), the expression $\tau = (2/\pi)\arcsin(r)$ for Kendall's tau is independent of the density generator g. Therefore, it holds in the general family of meta-elliptical distributions. Moreover, Kendall's tau corresponds to Blomqvist's medial correlation (Abdous, Genest, & Rémillard, 2005). Spearman's rho is more cumbersome since it generally depends on both g and r. A closed-form expression for elliptical distributions is not available, apart from some exceptions such as the meta-Gaussian distributions for which $\rho = (6/\pi)\arcsin(r/2)$, and another distribution in Hult & Lindskog (2002).

4.1. Conditional Distribution of Meta-Elliptical Distributions

The next result establishes that conditional distributions in meta-elliptical distributions are meta-elliptical.

Proposition 1. *Assume $(X^{(1)}, \ldots, X^{(p)})$ follows a meta-elliptical distribution with continuous marginals K_i ($i = 1, \ldots, p$), density generator g, and positive definite linear correlation matrix R. Define $Z^{(i)} = F^{-1}\left\{K_i(X^{(i)})\right\}$ ($i = 1, \ldots, p$), where F is the distribution function in Equation (3). The following two statements hold.*

1. *Conditionally on $(Z^{(3)}, \ldots, Z^{(p)})$, the variables $(Z^{(i)} - \delta_i)/\gamma_{ii}^{1/2}$ ($i = 1, 2$), where δ_i and γ_{ii} are defined in Equations (5) and (6), have the same distribution function, say $\tilde{F}$.*
2. *Conditionally on $(X^{(3)}, \ldots, X^{(p)})$, the distribution of $(X^{(1)}, X^{(2)})$ is meta-elliptical with the linear correlation r^{12} in Equation (7), and marginal distribution functions*

$$\tilde{K}_i(x^{(i)}) = \tilde{F}\left[\frac{F^{-1}\left\{K_i(x^{(i)})\right\} - \delta_i}{\gamma_{ii}^{1/2}}\right] \quad (i = 1, 2). \tag{11}$$

The partial Kendall/Spearman correlation between $X^{(i)}$ and $X^{(j)}$ is the Kendall/Spearman correlation in the conditional distribution of $(X^{(i)}, X^{(j)})$, given all remaining variables. It follows immediately from Proposition 1 that $\tau^{ij} = (2/\pi)\arcsin(r^{ij})$ is the partial Kendall's tau correlation. Moreover, if uncorrelatedness is interpreted in the sense of Kendall or Spearman, then $X^{(i)}$ and $X^{(j)}$ are conditionally uncorrelated, given all remaining variables, if and only if $\tau^{ij} = r^{ij} = \theta_{ij} = 0$. This statement is emphasized in the following corollary.

Corollary 1. *Assume $X = (X^{(1)}, \ldots, X^{(p)})$ follows a meta-elliptical distribution. Then, $X^{(i)}$ and $X^{(j)}$ are conditionally uncorrelated, given all remaining variables, in the sense of Kendall or Spearman if and only if θ_{ij}, the element in position (i, j) of R^{-1}, vanishes.*

For meta-Gaussian distributions since the Gaussian copula factorizes when the linear correlation vanishes, the conditionally uncorrelated statement can be replaced by the stronger conditionally independent expression. Proposition 1 has implications for the interpretation of non-edges in undirected graphs. For example, Finegold & Drton (2011) and Vogel & Fried (2011) assumed a t distribution with finite second moments in order to interpret conditional uncorrelatedness in terms of partial Pearson correlations. Proposition 1 states that conditional uncorrelatedness can be interpreted without assuming any moment. Therefore, one can even assume, for example, a Cauchy distribution.

Meta-elliptical distributions provide a big leap in generality over meta-Gaussian distributions. However, independence among marginals in meta-elliptical distributions is only possible in the sub-family of meta-Gaussian distributions. Hence, one is forced to replace the conditional independence between two variables by the weaker notion of conditional uncorrelatedness, except when the meta-elliptical distribution is meta-Gaussian. As graphical models, meta-elliptical distributions are also much more general than elliptical distributions in Finegold & Drton (2011). All marginals of an elliptical distribution have the same distribution apart from location and scatter parameters. This restriction does not hold for meta-elliptical distributions since the marginals are arbitrary.

5. MOTIVATION FOR THE META-ELLIPTICAL LASSO

The meta-elliptical lasso introduced in Section 2 is now motivated. Assume $X = (X^{(1)}, \ldots, X^{(p)})$ follows a meta-elliptical distribution with known density generator g, unknown positive definite linear correlation matrix R, and continuous marginals K_i $(i = 1, \ldots, p)$. The joint distribution of $Z^{(i)} = F^{-1}\left\{K_i(X^{(i)})\right\}$, where F is given by Equation (3), is elliptical with density h in Equation (4). Therefore, when marginals K_i are known and $n \geq p$, efficient estimation of R for elliptical distributions is made possible by Kent & Tyler (1991) .

5.1. Efficient Estimator with Known Marginals and $n \geq p$

In the classical asymptotic theory, an efficient estimation of Θ and R in terms of Fisher's information, provided by the solution to the scale-only problem of Kent & Tyler (1991), is obtained from the fixed point algorithm

$$\Theta_{m+1}^{-1} = \frac{1}{n}\sum_{l=1}^{n} u(Z_l^{\mathrm{T}}\Theta_m Z_l)Z_l Z_l^{\mathrm{T}} \quad (m = 1, 2, \ldots), \tag{12}$$

where the function $u(s) = -2g'(s)/g(s)$ acts as a weight function in this iterative re-weighted estimate. Kent & Tyler (1991) also showed that if $n \geq p$, $u(s) \geq 0$ and $u(s)$ is continuous and non-increasing, and $su(s)$ is strictly increasing and bounded, then there exists a unique solution to

$$\Theta^{-1} = \frac{1}{n}\sum_{l=1}^{n} u(Z_l^{\mathrm{T}}\Theta Z_l)Z_l Z_l^{\mathrm{T}}.$$

Moreover, the fixed point algorithm in Equation (12) converges to the solution regardless of the initial positive definite matrix Θ_1 selected.

For example, the weight function of the multivariate t distribution in dimension p with ν degrees of freedom is $u(s) = (\nu + p)/(\nu + s)$.

5.2. The Case of Unknown Marginals

When marginals K_i are unknown, one must resort to pseudo-observations

$$\hat{Z}_l^{(i)} = F^{-1}\left\{R_l^{(i)}/(n+1)\right\} \quad (l = 1, \ldots, n;\ i = 1, \ldots, p).$$

The fixed point algorithm is then the Equation (12) with the unobservable variables $Z_l = (Z_l^{(1)}, \ldots, Z_l^{(p)})$ replaced by the pseudo-observations $\hat{Z}_l = (\hat{Z}_l^{(1)}, \ldots, \hat{Z}_l^{(p)})$. However, if $n < p$, the weighted covariance matrix of Equation (12) becomes singular after the first iteration. For high dimensional problems and sparse matrix Θ, it is proposed to solve the estimating equation

$$\Theta^{-1} - \frac{1}{n}\sum_{l=1}^{n} u(\hat{Z}_l^{\mathrm{T}}\Theta\hat{Z}_l)\hat{Z}_l\hat{Z}_l^{\mathrm{T}} - \lambda\Gamma(\Theta) = 0 \tag{13}$$

by a fixed point algorithm. Let Θ_1 be the initial value. Until convergence, for $m = 1, 2, \ldots$ compute

$$\hat{R}_{m+1} = \frac{1}{n} \sum_{l=1}^{n} u(\hat{Z}_l^{\mathrm{T}} \Theta_m \hat{Z}_l) \hat{Z}_l \hat{Z}_l^{\mathrm{T}} \tag{14}$$

and find the solution Θ_{m+1} to

$$\Theta_{m+1}^{-1} - \hat{R}_{m+1} - \lambda \Gamma(\Theta_{m+1}) = 0. \tag{15}$$

The solution Θ_{m+1} to Equation (15) is recognized as the minimizer of the Gaussian lasso problem with $\hat{R}_{m+1}$ used as input. Iterations stop when weights have stabilized between successive iterations.

It is not known whether the fixed point algorithm defined by Equations (14) and (15) always converges. However, if it converges, it provides a solution to Equation (13). In simulations, the solution to Equation (13), over a sequence of decreasing values of regularization parameters λ, could require 30 iterations for the first value, and only 2 or 3 iterations for later values using warm starts. This heavy computational burden, compared to the Kendall lasso in Liu et al. (2012), may be reduced by using as an initial estimate the solution Θ_1 of the Kendall lasso and, afterwards, performing only one re-weighting iteration. Because Equation (14) is not a correlation matrix, it should be noted that the matrix Θ_{m+1}^{-1} which solves Equation (15) is not guaranteed to be a correlation matrix. The meta-elliptical lasso (resp. re-weighted Kendall lasso) proposed in Section 2 uses the re-weighted Pearson (resp. back transformed Kendall) correlation matrix in Step 4 of the algorithm which guarantees a correlation matrix at the output.

It should be stated that the five-step algorithm in Section 2 really consists of two big steps:

(a) Obtain as an initial estimate Θ_1 the Kendall lasso of Liu et al. (2012).
(b) Obtain a refined estimate Θ_2 by re-weighting.

This two-step algorithm (a) and (b) does not exactly implement the fixed point iteration (14) and (15); it is nonetheless motivated by the fixed point iteration.

5.3. Relation to the Meta-Gaussian Lasso

Consider the Step 1 of the meta-elliptical lasso of Section 2 with $F = \Phi$. The pseudo-observations reduce to Gaussian scores which could be Winsorized as suggested by Liu, Lafferty, & Wasserman (2009)

$$\hat{Z}_l^{(i)} = \Phi^{-1} \left\{ T_{\delta_n} \left[R_l^{(i)}/(n+1) \right] \right\},$$

where

$$T_{\delta_n}(x) = \delta_n \, I(x < \delta_n) + x \, I(\delta_n \le x \le 1 - \delta_n) + (1 - \delta_n) \, I(x > 1 - \delta_n)$$

and $\delta_n = 1/(4n^{1/4}\sqrt{\pi \log n})$ is a truncation parameter to achieve the desired rate of convergence in the high dimensional setting. The choice $\delta_n = 1/(n+1)$ corresponds to no Winsorization. If Step 2 is replaced by the computation of the Pearson correlation matrix, then Step 3 computes the meta-Gaussian lasso of Liu, Lafferty, & Wasserman (2009). Next, for meta-Gaussian distributions, the density generator $g(s) = \exp(-s/2)$ yields the weight function $u(s) = 1$. In this case, re-weighting has no effect.

5.4. Relation to the EM Algorithm for the t Lasso

The fixed point algorithm defined by Equations (14) and (15) is related to the EM algorithm of Finegold & Drton (2011) who assumed that $Z = (Z^{(1)}, \ldots, Z^{(p)})$ follows a t distribution with ν degrees of freedom, location vector μ and positive definite inverse scatter matrix Θ. For our purpose, the location vector is assumed known, $\mu = 0$. Their t lasso is the solution to the (non-convex) optimization problem

$$\min_{\Theta \succeq 0} - \log\det\Theta - \frac{2}{n}\sum_{l=1}^{n} \log g(Z_l^{\mathrm{T}}\Theta Z_l) + \lambda||\Theta||_1,$$

where $g(s) \propto (1 + s/\nu)^{-(\nu+p)/2}$.

The distribution for Z is a Gaussian mixture obtained by assuming Z given τ is distributed as $N_p(0, \Theta^{-1}/\tau)$ and τ is distributed as Gamma$(\nu/2, \nu/2)$. Since the optimization cannot be done by solving a sub-gradient, they proposed a modified EM algorithm in which $\tau_1, \ldots, \tau_n$ are treated as missing variables. The E-step is

$$\tau_{m+1,l} = u(Z_l^{\mathrm{T}}\Theta_m Z_l) = (\nu + p)/(\nu + Z_l^{\mathrm{T}}\Theta_m Z_l)\ (l = 1, \ldots, n)$$

and the M-step seeks the solution Θ_{m+1} to

$$\min_{\Theta \succeq 0} - \log\det\Theta + \mathrm{tr}\left\{\frac{1}{n}\sum_{l=1}^{n} \tau_{m+1,l} Z_l Z_l^{\mathrm{T}}\Theta\right\} + \lambda||\Theta||_1. \tag{16}$$

Equation (15) is simply the sub-gradient of the convex Gaussian lasso in Equation (16) in which observations Z_l are replaced by pseudo-observations $\hat{Z}_l$.

6. SIMULATION STUDY

All simulations were run with `dpglasso` package version 1.0. The figures are best visualized in colour. Simulations compare receiver operating characteristic (ROC) curves which is a graphical plot of true positive rates versus false positive rates as the regularization parameter λ varies. Each ROC curve is an average over 100 trials. The sparse matrix Θ was generated using the procedure described in Finegold & Drton (2011):

1. θ_{ij}, $i > j$, are independently distributed variables taking values -1, 0, and 1 with probability 0.01, 0.98, and 0.01;
2. $\theta_{ij} = \theta_{ji}$, $i < j$;
3. $\theta_{ii} = 1 + h$, where h is the number of non-zero elements in the ith row of Θ.

Step 3 ensures a positive definite matrix since it is strictly diagonally dominant. The diagonal elements are then reduced by a common factor to strengthen partial correlations. This factor is as large as possible while maintaining positive definiteness.

6.1. Simulation 1

The first simulation shows the effect of three transformations of marginals. The simulated data follow one of six meta-elliptical distributions obtained by a combination of one among three marginal transformations applied to one among two distributions (multivariate Gaussian, $N_p(0, R)$, and multivariate Cauchy, $t_{p,1}(0, R)$). The three transformations are the identity which means no transform is done, in which case the simulated distribution of the data is truly Gaussian or Cauchy, the cumulative distribution function (cdf) which yields uniform marginals, and the power function

$\text{sign}(x)|x|^3$ taken from Liu, Lafferty, & Wasserman (2009). In each case, the same transformation is applied to all p marginals. The first two lassos are the Kendall lasso and the Spearman lasso. When the distribution is meta-Gaussian, they are compared to the meta-Gaussian lasso without Winsorization ($\delta_n = 1/(n+1)$) and the Gaussian lasso. On the other hand, when the distribution is meta-Cauchy, they are compared to the meta-Cauchy lasso and the Cauchy lasso. The meta-Cauchy lasso is a special case of the meta-elliptical lasso of Section 2 with $u(s) = (1+p)/(1+s)$, and the Cauchy lasso is a special case of the t lasso for the same choice of the weight function u. For the Cauchy lasso and the meta-Cauchy lasso, only one re-weighting iteration of the algorithm is performed after starting with the initial estimate Θ_1 produced by the Kendall lasso.

Kendall, Spearman, meta-Cauchy, and meta-Gaussian lassos are computed from ranks which are invariant to monotone transformations of marginals. Hence, their ROC curves at a given distribution remain the same regardless of the applied transformation. Only ROC curves of the Gaussian lasso and the Cauchy lasso are affected. For meta-Gaussian distributions, Figure 2 confirms the findings of Liu et al. (2012). These findings are that Kendall, Spearman and meta-Gaussian lassos are nearly as efficient as the Gaussian lasso when the distribution is truly Gaussian. The performance of the Gaussian lasso can be poor for distributions which are meta-Gaussian, but not Gaussian. A similar conclusion is obtained for meta-Cauchy distributions; the performance of the Cauchy lasso can be poor when the distribution is meta-Cauchy, but not Cauchy. The meta-Cauchy lasso is nearly as efficient as the Cauchy lasso when the distribution is truly Cauchy. None of the simulations reported earlier in the literature considered distributions other than meta-Gaussian or t, apart from some contaminated versions thereof. Figure 2 reveals a new and interesting fact: the meta-Cauchy lasso outperforms the Kendall/Spearman lasso when the distribution is meta-Cauchy.

6.2. Simulation 2

Given the poor performance of the Gaussian lasso and the Cauchy lasso for some transformations of marginals, the second simulation reported in Figure 3 considered only lassos based on ranks which are invariant to such transformations. Instead of transforming marginals, the effect of some contaminated data is now investigated. As in the first simulation, two distributions (multivariate Gaussian, $N_p(0, R)$, and multivariate Cauchy, $t_{p,1}(0, R)$) are simulated. The first distribution has univariate $N(0, 1)$ marginals, whereas the second has univariate Cauchy marginals with much longer tails. The monotone transformation $\Phi^{-1}[F_1(x)]$ (without any effect on invariant lassos), where F_1 is the cdf of a univariate Cauchy distribution, is applied to the marginals of the multivariate Cauchy resulting in a meta-Cauchy distribution with $N(0, 1)$ marginals. Outliers will thus be of the same magnitude for both distributions. Deterministic contaminated data as in Liu et al. (2012) consist of replacing $\lfloor nr \rfloor$ observations by a vector $(5, -5, 5, -5, \ldots)$ of length p, where $r = 0$, 0.01 or 0.05 is the contamination level. Four invariant lassos are compared: the Kendall lasso and the Spearman lasso as in the first simulation, the Winsorized meta-Gaussian lasso and the new re-weighted Kendall lasso described in Section 2 with weight function $u(s) = (1+p)/(1+s)$. For contaminated meta-Gaussian distributions, Liu et al. (2012) found that the Winsorized meta-Gaussian lasso performs better than the non-Winsorized version. The simulation for meta-Gaussian distributions confirms the findings of Liu et al. (2012): the Kendall/Spearman lasso is nearly as good as the (Winsorized) meta-Gaussian lasso when there is no contamination. However, the Kendall lasso and, to a lesser extent, the Spearman lasso outperform the meta-Gaussian lasso in presence of contamination. Interestingly, the re-weighted Kendall lasso and the Kendall lasso have almost identical curves, which means that re-weighting did not have ill effects for meta-Gaussian distributions. For meta-Cauchy distributions the findings are very different. The re-weighted Kendall lasso outperforms all three other lassos when there is no ($r = 0$) or small ($r = 0.01$) contamination. At the higher level ($r = 0.05$) of contamination, the

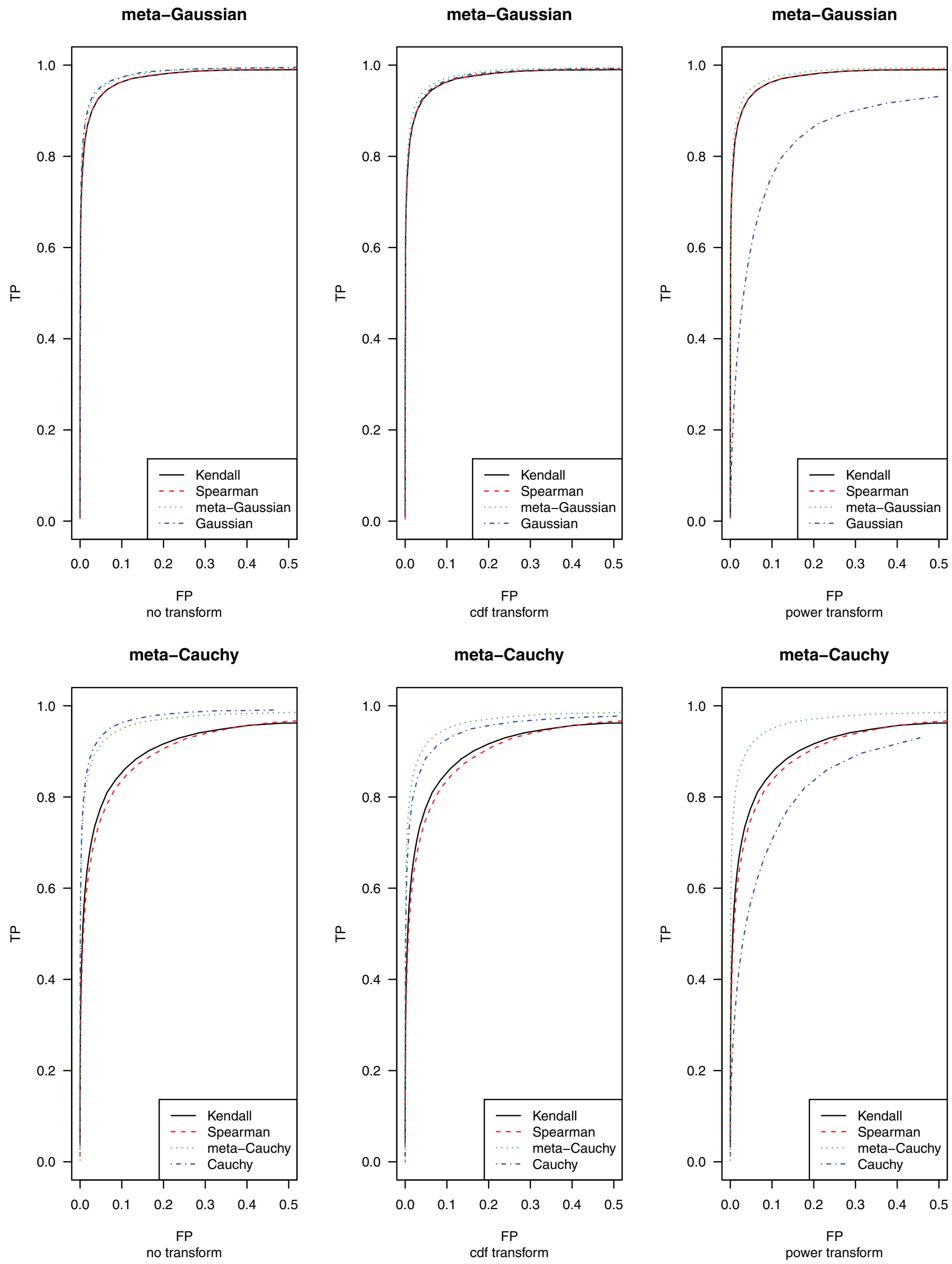

FIGURE 2: ROC curves for the identity (left), cdf (middle), and power transformations (right) for meta-Gaussian (top) and meta-Cauchy (bottom) distributions. Here $n = 100$ and $p = 100$. The lassos are: Kendall and Spearman from Liu et al. (2012); meta-Gaussian with $\delta_n = 1/(n+1)$ from Liu, Lafferty, & Wasserman (2009); meta-Cauchy of Section 2; Gaussian; Cauchy of Finegold & Drton (2011). [Colour figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

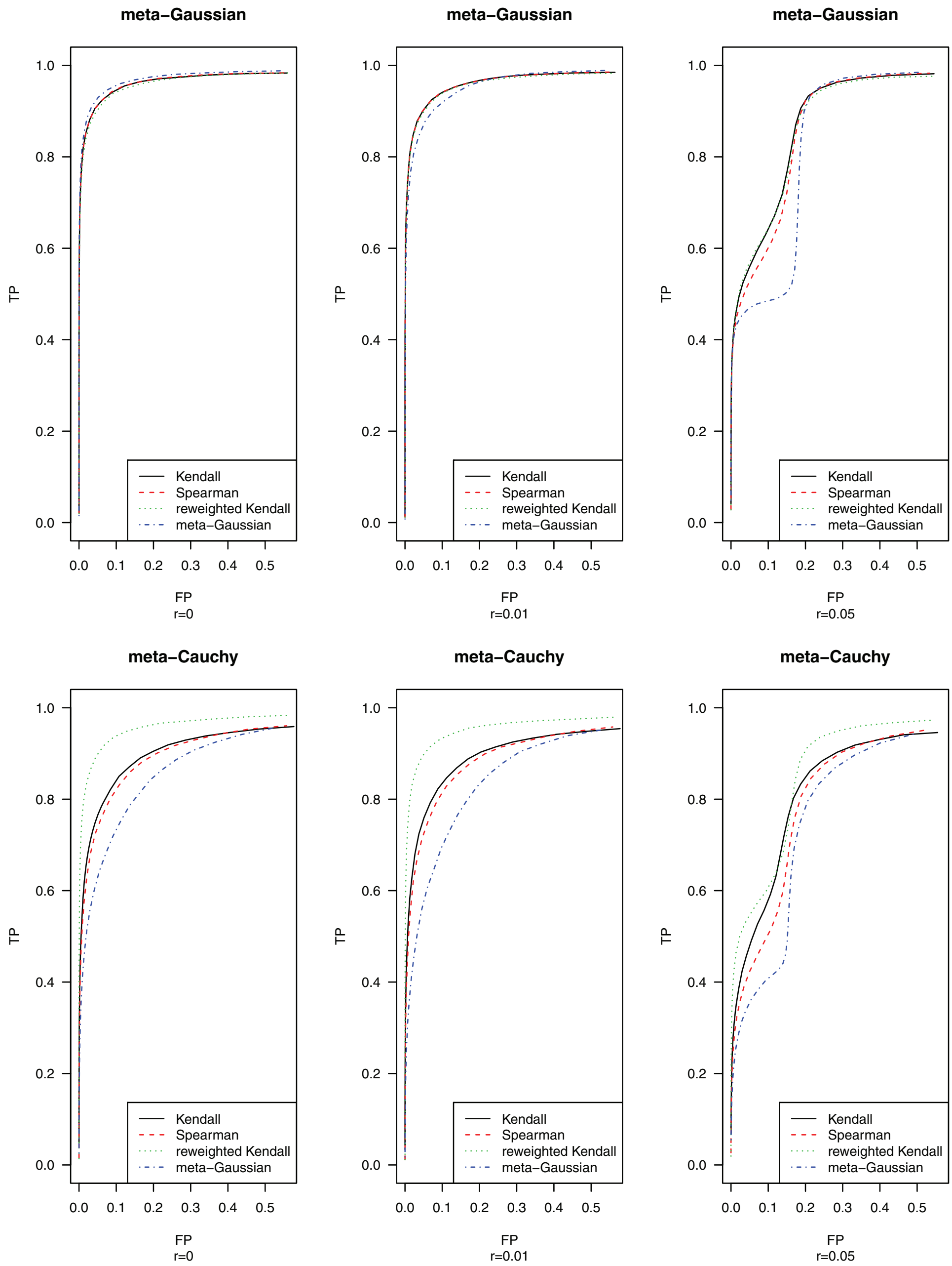

FIGURE 3: ROC curves for deterministic contamination of $r = 0$ (left), $r = 0.01$ (middle) and $r = 0.05$ (right) for meta-Gaussian (top) and meta-Cauchy (bottom) distributions. Here $n = 100$ and $p = 100$. The lassos are: Kendall and Spearman from Liu et al. (2012); Winsorized meta-Gaussian from Liu, Lafferty, & Wasserman (2009); re-weighted Kendall of Section 2 with $u(s) = (1 + p)/(1 + s)$. [Colour figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

Kendall lasso performs closer but still not as well as the re-weighted Kendall lasso. It should be remarked at this point that it would be difficult in practice to distinguish between meta-Gaussian and meta-Cauchy distributions when both have the same marginals. Goodness-of-fit tests for copulas are available in Genest & Rémillard (2008) but are only feasible in large n and small p

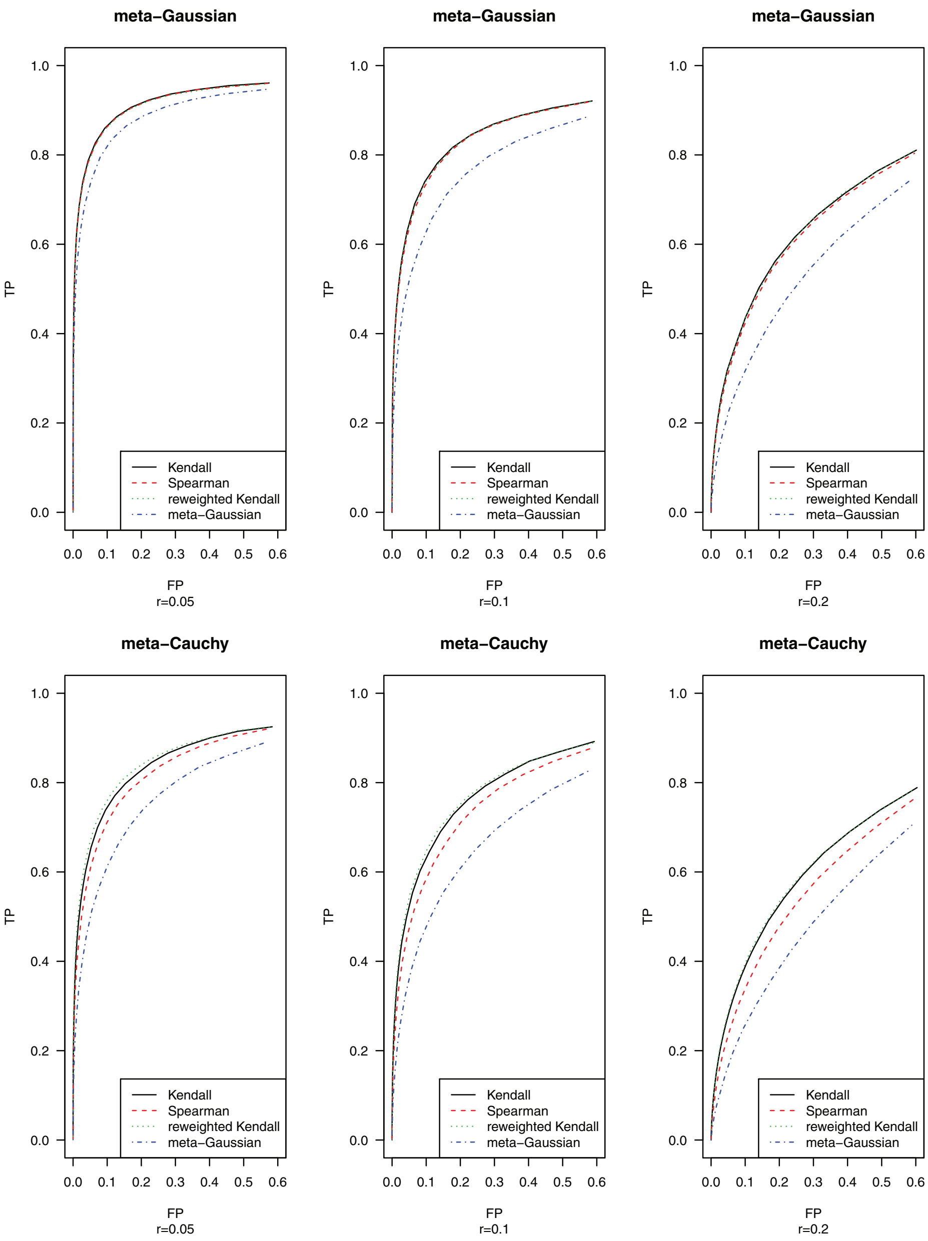

FIGURE 4: ROC curves for random contamination of $r = 0.05$ (left), $r = 0.1$ (middle), and $r = 0.2$ (right) for meta-Gaussian (top) and meta-Cauchy (bottom) distributions. Here $n = 100$ and $p = 100$. The lassos are: Kendall and Spearman from Liu et al. (2012); Winsorized meta-Gaussian from Liu, Lafferty, & Wasserman (2009); re-weighted Kendall of Section 2 with $u(s) = (1 + p)/(1 + s)$. [Colour figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

situations. Instead of guessing which copula best fits the data in order to select the best lasso, a more pragmatic approach is to choose the lasso which performs best under the largest family of copulas. The meta-Cauchy ($\nu = 1$) and meta-Gaussian ($\nu \to \infty$) distributions considered in Figures 3 and 4 are the limiting cases of the family of meta-t distributions. The re-weighted Kendall lasso is thus expected to perform well under all intermediate meta-t distributions.

6.3. Simulation 3

The last simulation reported in Figure 4 is identical to the second except that the contamination is now random. Random contamination is more realistic and not as severe as deterministic contamination which can really hurt a non-robust lasso. Three levels of random contamination $r = 0$, $r = 0.1$ or $r = 0.2$ are now considered. A number $\lfloor nr \rfloor$ of entries of each variable are selected at random (according to a uniform distribution) and replaced by either 5 or -5 with equal probability. The re-weighted Kendall lasso of Section 2 with weight function $u(s) = (1 + p)/(1 + s)$ performs as well as Kendall/Spearman lasso and outperforms the (Winsorized) meta-Gaussian lasso under contaminated meta-Gaussian distributions. For contaminated meta-Cauchy distributions, the re-weighted Kendall lasso performs slightly better than the Kendall lasso and outperforms the Spearman lasso and the (Winsorized) meta-Gaussian lasso.

7. CONCLUSION

The Kendall lasso of Liu et al. (2012) assumed a meta-Gaussian distribution for the observations. Since goodness-of-fit for a meta-Gaussian distribution is never tested in high dimensional settings, the following question arises: How does the Kendall lasso performs if the distribution of the data is not meta-Gaussian, or some contaminated versions thereof? This question motivated the study of graphical lassos adapted to observations following a meta-elliptical distribution. It was established in Corollary 1 that non-edges in meta-elliptical graphs can be interpreted in terms of conditional uncorrelatedness between two variables, given all other variables. In terms of receiver operating characteristic curves, the re-weighted Kendall lasso introduced in Section 2 provides substantial gains in efficiency when the distribution is meta-elliptical, but not meta-Gaussian. Moreover, it suffers only marginal losses in efficiency when the distribution is meta-Gaussian. The re-weighted Kendall lasso can thus be used as a safe replacement to the Kendall lasso, and it only requires roughly twice the amount of computation.

In the original submission, the fixed point algorithm was initialized with the identity matrix and was iterated until convergence. This made the algorithm slow because computation time is roughly proportional to the number of iterates. Although convergence was not a problem, a comment of a referee on the infeasibility of the algorithm in high dimensional settings prompted a modified approach consisting of two steps: (a) start with a good preliminary estimate, the Kendall lasso, and (b) perform only one re-weighting iteration. This approach has been very successful.

APPENDIX

Proof of Proposition 1. The conditional distribution function

$$pr(X^{(1)} \le x^{(1)}, X^{(2)} \le x^{(2)} \mid X^{(k)} = x^{(k)}, \quad k \ne 1, 2)$$

is equal to

$$pr(Z^{(1)} \le z^{(1)}, Z^{(2)} \le z^{(2)} \mid Z^{(k)} = z^{(k)}, \quad k \ne 1, 2),$$

where $z^{(i)} = F^{-1}\left\{K_i(x^{(i)})\right\}$. The latter is the distribution function of an elliptical distribution with location δ in Equation (5) and scatter matrix $R_{11.2}$. Since the linear correlation in $R_{11.2}$ is r^{12}, the copula is an elliptical copula with linear correlation r^{12}. Upon using Equations (5) and (6), the conditional distribution function of $Z^{(i)}$ is

$$\tilde{F}\left[\frac{z^{(i)} - \delta_i}{\gamma_{ii}^{1/2}}\right] \ (i = 1, 2).$$

■

BIBLIOGRAPHY

Abdous, B., Genest, C., & Rémillard, B (2005). Dependence properties of meta-elliptical distributions. In *Statistical Modeling and Analysis for Complex Data Problems*, GERAD 25th Anniv. Ser., 1, Springer, New York, pp. 1–15.

Becker, C. (2005). Iterative proportional scaling based on a robust start estimator. In *Classification—The Ubiquitous Challenge*, Springer, Heidelberg, pp. 248–255.

Boyd, S. & Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge.

Chen, Y., Wiesel, A., Eldar, Y. C., & Hero, A. O. (2010). Shrinkage algorithms for MMSE covariance estimation. *IEEE Transactions on Signal Processing*, 58, 5016–5029.

Cox, D. R. & Wermuth, N. (1996). *Multivariate Dependencies: Models, Analysis and Interpretation*. Monographs on statistics and applied probability. Chapman & Hall, London.

Fang, H.-B., Fang, K.-T., & Kotz, S. (2002). The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis*, 82, 1–16.

Finegold, M. & Drton, M. (2011). Robust graphical modeling of gene networks using classical and alternative t-distributions. *Annals of Applied Statistics*, 5, 1057–1080.

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics Oxford England*, 9, 432–441.

Genest C. & Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'Institut Henri Poincaré: Probabilités et Statistiques*, 44, 1096–1127.

Gottard, A. & Pacillo, S. (2010). Robust concentration graph model selection. *Computational Statistics & Data Analysis*, 54, 3070–3079.

Hult H. & Lindskog, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied Probability*, 34, 587–608.

Kelly K. S. & Krzysztofowicz, R. (1997). A bivariate meta-gaussian density for use in hydrology. *Stochastic Hydrology and Hydraulics*, 11, 17–31.

Kent, J. T. & Tyler, D. E. (1991). Redescending M-estimates of multivariate location and scatter. *Annals of Statistics*, 19, 2102–2119.

Knight, W. R. (1966). A computer method for calculating Kendall's tau with ungrouped data. *Journal of the American Statistical Association*, 61, 436–439.

Lauritzen, S. L. (1996). *Graphical Models*. Oxford Statistical Science Series. Oxford University Press, New York, USA.

Ledoit, O. & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88, 365–411.

Lindskog, F., McNeil, A., & Schmock, U. (2003). Kendall's tau for elliptical distributions. In *Credit Risk: Measurement, Evaluation, and Management*, Springer, Heidelberg, pp. 149–156.

Liu, H., Han, F., Yuan, M., Lafferty, J., & Wasserman, L. (2012). High dimensional semiparametric gaussian copula graphical models. *Annals of Statistics*, 40, 2293–2326.

Liu, H., Lafferty, J., & Wasserman, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research*, 10, 2295–2328.

Mazumder, R. & Hastie, T. (2012a). The graphical lasso: New insights and alternatives. *Electronic Journal of Statistics*, 6, 2125–2149.

Mazumder, R. & Hastie, T. (2012b). Exact covariance thresholding into connected components for large-scale graphical lasso. *Journal of Machine Learning Research*, 13, 781–794.

Meinshausen, N. & Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *Annals of Statistics*, 34, 1436–1462.

Miyamura, M. & Kano, Y. (2006). Robust Gaussian graphical modeling. *Journal of Multivariate Analysis*, 97, 1525–1550.

Schäfer, J. & Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 4, 1–30.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de statistique de l'Université de Paris*, 8, 229–231.

Tenenhaus, A., Guillemot,V., Gidrol, X., & Frouin, V. (2010). Gene association networks from microarray data using a regularized estimation of partial correlation based on pls regression. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 7, 251–262.

Tyler, D. E. (1983). Robustness and efficiency properties of scatter matrices. *Biometrika*, 70, 411– 420.

Vogel, D. & Fried, R. (2011). Elliptical graphical modelling. *Biometrika*, 98, 935–951.

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Wiley Series in Probability & Statistics, John Wiley & Sons, Chichester.

Witten, D. M., Friedman, J. H., & Simon, N. (2011). New insights and faster computations for the graphical lasso. *Journal of Computational and Graphical Statistics*, 20, 892–900.

Yuan, M. & Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94, 19–35.

Received 28 May 2012
Accepted 19 February 2014