Statistica Sinica **35** (2025), 1-27
doi:https://doi.org/10.5705/ss.202022.0394

DIRECTIONAL TESTS IN GAUSSIAN GRAPHICAL MODELS

Claudia Di Caterina*, Nancy Reid and Nicola Sartori

University of Verona, University of Toronto and University of Padova

Abstract: We develop directional tests to compare incomplete undirected graphs in the general context of covariance selection for Gaussian graphical models. The exactness of the underlying saddlepoint approximation is proved for chordal graphs, and leads to exact control of the size of the tests, given that the only approximation error involved is from the numerical calculation of two scalar integrals. Although exactness is not guaranteed for non-chordal graphs, the ability of the saddlepoint approximation to control the relative error means the proposed method outperforms its competitors even in these cases. The accuracy of our proposal is verified using simulation experiments under challenging scenarios in which inference via standard asymptotic approximations to the likelihood ratio test and some of its higher-order modifications fails. The directional approach is used to illustrate the assessment of Markovian dependencies in a data set from a veterinary trial on cattle. A second example with microarray data shows how to select the graph structure related to genetic anomalies due to acute lymphocytic leukemia.

Key words and phrases: Covariance selection, exponential family, higher-order asymptotics, likelihood ratio test, saddlepoint approximation, undirected graph.

1. Introduction

Undirected graphical models have gained considerable success in a variety of fields, including medicine, social sciences, and physics, owing to their flexibility and easy interpretation. Typically, these probabilistic graphs describe complex multivariate distributions of variables (nodes) using the product of simpler submodels, each referred to a low-dimensional subset of the graph (clique). Book-length expositions on the topic can be found in Lauritzen (1996), Borgelt and Kruse (2002), and Whittaker (2009).

Today, applications of graphical models are challenged by the growth in size and sophistication of modern data. An important question is inferring the structure of large graphs, that is the underlying connections (edges) between the variables under examination. This task is well known in the literature as covariance selection. A popular class of graphical models is that of decomposable models, which describe graphs that contain no chordless cycles of length greater than three. These are called chordal, decomposable, or triangulated graphs

*Corresponding author. E-mail: claudia.dicaterina@univr.it

(Lauritzen, 1996, Sec. 2.1).

For convenience, a graphical model is often expressed using the exponential family form. The Gaussian distribution is particularly suitable for continuous responses, because conditional independence in the graph can be characterized easily in terms of assumptions on model parameters (see Section 3.1).

Likelihood-based inference for covariance selection is discussed in Salgueiro, Smith and McDonald (2005) in the context of testing exclusion of single edges in complete graphs, that is fully saturated models. Córdoba, Bielza and Larrañaga (2020, Sec. 7) review general edge exclusion tests, acknowledging the poor quality of the usual chi-squared approximation to the distribution of the likelihood ratio statistic. They mention that, when testing the removal of r edges, the exact distribution is the product of r Beta random variables (Lauritzen, 1996, Prop. 5.14). However, this result has not received much attention in the literature and seems of limited practical utility. Another strategy is to iteratively perform exclusion tests for single edges based on partial correlation coefficients, with some adjustment needed to account for multiple comparisons.

In this paper, we develop likelihood-based directional tests for covariance selection in Gaussian graphical models, possibly incorporating *a priori* restrictions on the graph structure. Specifically, our method allows one to test hypotheses that involve removing sub-graphs with multiple edges from complete or incomplete graphs. We prove the exactness of the underlying saddlepoint approximation for chordal graphs, and run extensive Monte Carlo simulations to show the null uniform distribution of the directional p-value in challenging scenarios, even when the number of nodes is larger than the sample size. In those settings, the classical approach based on the likelihood ratio statistic or some of its higher-order modifications (Skovgaard, 2001) breaks down. We also show results for a non-chordal graph, where the directional inference is confirmed to be more accurate than that of its competitors. A much simpler problem in covariance selection, limited to testing an incomplete graph versus the saturated model, is studied by Davison et al. (2014, Sec. 5.3), and is shown to be exact in Huang, Di Caterina and Sartori (2022). Our extension involves both theoretical and computational innovations.

Directional inference on a vector-valued parameter of interest was introduced by Fraser and Massam (1985) in nonnormal linear regression models, and then generalized in Skovgaard (1988). Substantial progress from both a methodological and computational perspective was made by Davison et al. (2014), where the computation of the directional p-value by one-dimensional numerical integration proved especially accurate in several settings. The procedure was extended from linear exponential families to nonlinear parameters of interest in general continuous models by Fraser, Reid and Sartori (2016). In addition to its accuracy, the directional approach has been found to coincide with exact results in several classical situations (McCormack et al., 2019).

Section 2 reviews the technique of directional inference for exponential family models. Section 3 presents the new directional testing method for covariance selection. Here, we prove the exactness of the saddlepoint approximation for decomposable Gaussian graphical models in chordal graphs, and develop specific notation also valid in the non-chordal case. A number of computational innovations are presented in Section 4. Simulation results comparing the accuracy of the various methods are shown in Section 5, and Section 6 reports applications to data from a veterinary trial and from a microarray study of altered gene expressions in acute lymphocytic leukemia. Section 7 concludes the paper.

2. Background

2.1. Likelihood ratio tests

Assume that y follows a parametric distribution $f(y;\theta)$, with $\theta \in \mathbb{R}^p$. The log-likelihood function $\ell(\theta) = \ell(\theta; y) = \log f(y;\theta)$ is maximized by the maximum likelihood (ML) estimator $\hat{\theta} = \hat{\theta}(y)$. Possibly after a reparameterization, the model parameter can be typically expressed as $\theta = (\psi, \lambda)$, where $\psi(\theta)$ is the d-dimensional component of interest involved in the hypothesis $H_\psi \colon \psi(\theta) = \psi$. We write $\hat{\theta}_\psi = (\psi, \hat{\lambda}_\psi)$ to denote the constrained ML estimator of θ when the null H_ψ is true.

Under usual regularity conditions (see, e.g., Cox and Hinkley, 1974, Sec. 9.3), the first-order approximation to the distribution of $\hat{\theta}$ is normal with mean θ and estimated covariance matrix $j(\hat{\theta})^{-1}$, with $j(\theta) = -\partial^2\ell(\theta)/\partial\theta\partial\theta^\top$ the observed Fisher information matrix. The hypothesis H_ψ can be tested using the likelihood ratio statistic

$$w(\psi) = 2\{\ell(\hat{\theta}) - \ell(\hat{\theta}_\psi)\}\,, \tag{2.1}$$

which is invariant to reparameterizations, and has an approximate χ^2_d distribution under the null hypothesis H_ψ, where d is the dimension of the parameter of interest ψ.

Skovgaard (2001) introduced two modifications to (2.1),

$$w^*(\psi) = w(\psi)\left\{1 - \frac{\log\gamma(\psi)}{w(\psi)}\right\}^2 \quad \text{and} \quad w^{**}(\psi) = w(\psi) - 2\log\gamma(\psi)\,, \tag{2.2}$$

and showed that the limiting distribution of both test statistics based on the correction factor $\gamma(\psi)$ is also χ^2_d. These modifications were obtained by analogy with the derivation for scalar parameters of interest of modifications to the square root of $w(\psi)$, the so-called r^* approximation of Barndorff-Nielsen (1986), further discussed in Fraser, Reid and Wu (1999). Skovgaard (2001) emphasized not only the simplicity of computation of the adjustment, especially when compared with Bartlett (1937) correction using moments, but also its large-deviation properties.

Tests based on $w(\psi)$, including $w^*(\psi)$, $w^{**}(\psi)$, and the Bartlett-corrected $w(\psi)$, provide omnibus measures of departure of the data from H_ψ: the resulting p-value averages the deviations from the null hypothesis in all potential directions of the parameter space. In the next section, we review the approach of Davison et al. (2014, Sec. 3) for measuring the departure from H_ψ only in the direction indicated by the observed data. For a more complete exposition of the difference between omnibus and directional tests, see Fraser and Reid (2006).

2.2. Directional tests in linear exponential families

Focusing on hypotheses that are linear in the canonical parameter θ of an exponential family model, we summarize here the procedure detailed in Davison et al. (2014, Sec. 3), which involves two steps of dimensionality reduction.

Denoting by $u = u(y)$ the sufficient statistic for the p-dimensional vector parameter θ, we can consider the marginal density of u and the corresponding log-likelihood function $\ell(\theta; u) = \theta^\top u - K(\theta)$, which takes the standard exponential family form. Consistent with the notation established by Davison et al. (2014) and Fraser, Reid and Sartori (2016), we define the observed data $y^0 = (y_1^0, \ldots, y_n^0)$ and the corresponding observed value of the sufficient statistic $u^0 = u(y^0)$. Given the centered statistic $s = u - u^0$ with observed value $s^0 = u^0 - u^0 = 0$, the tilted log-likelihood function is

$$\ell(\theta; s) = \theta^\top s + \ell^0(\theta)\,, \tag{2.3}$$

where $\ell^0(\theta) = \ell(\theta; u = u^0)$.

When the linearity in θ applies to both the interest and the nuisance parameters, meaning $\theta = (\psi, \lambda)$, expression (2.3) can be written as

$$\ell(\theta; s) = \psi^\top s_1 + \lambda^\top s_2 + \ell^0(\psi, \lambda)\,, \tag{2.4}$$

where ψ and s_1 have dimension d. The first dimensionality reduction from p to d follows directly from conditioning on the component of the statistic sufficient for λ. Indeed, the conditional distribution of s_1 given s_2 depends on ψ only, and is still of exponential family form (cf., Lehmann and Romano, 2005, Lem. 2.7.2). Such a conditioning translates into fixing $\hat\theta_\psi = (\psi, \hat\lambda_\psi)$ at the observed value $\hat\theta^0_\psi = (\psi, \hat\lambda^0_\psi)$.

The saddlepoint approximation for this conditional distribution is typically very accurate (Barndorff-Nielsen and Cox, 1979). Following, for instance, Pace and Salvan (1997, Sec. 10.10.2), we can illustrate how the saddlepoint approximation is obtained as the ratio of the saddlepoint approximation for the joint density of $s = (s_1, s_2)$ and the saddlepoint approximation for the marginal density of s_2. Indeed, the former can be expressed as

$$\frac{\exp[\{\theta - \hat\theta(s)\}^\top s + \ell^0(\theta) - \ell^0\{\hat\theta(s)\}]}{(2\pi)^{p/2}|-\ell^0_{\theta\theta}\{\hat\theta(s)\}|^{1/2}} = \frac{\exp[\ell(\theta; s) - \ell\{\hat\theta(s); s\}]}{(2\pi)^{p/2}|j_{\theta\theta}\{\hat\theta(s)\}|^{1/2}}\,, \tag{2.5}$$

where $\hat{\theta}(s)$ solves in θ the score equation from the log-likelihood (2.4), $s = -\ell^0_\theta(\theta) = -\partial\ell^0(\theta)/\partial\theta$, $j_{\theta\theta}(\theta) = -\partial^2\ell(\theta;s)/\partial\theta\partial\theta^\top = -\partial^2\ell^0(\theta)/\partial\theta\partial\theta^\top = -\ell^0_{\theta\theta}(\theta)$, and $|A|$ denotes the determinant of the square matrix A. Similarly, the saddlepoint approximation for the marginal distribution of s_2 is

$$\frac{\exp[\{\lambda - \hat{\lambda}_\psi(s_2)\}^\top s_2 + \ell^0(\theta) - \ell^0\{\hat{\theta}_\psi(s_2)\}]}{(2\pi)^{(p-d)/2}|-\ell^0_{\lambda\lambda}\{\hat{\theta}_\psi(s_2)\}|^{1/2}} = \frac{\exp[\ell(\theta;s) - \ell\{\hat{\theta}_\psi(s_2);s\}]}{(2\pi)^{(p-d)/2}|j_{\lambda\lambda}\{\hat{\theta}_\psi(s_2)\}|^{1/2}}, \quad (2.6)$$

where $\hat{\theta}_\psi(s_2) = (\psi, \hat{\lambda}_\psi(s_2))$ is the solution to the score equation from the log-likelihood (2.4), seen as a function of λ for fixed ψ, $s_2 = -\ell^0_\lambda(\theta) = -\partial\ell^0(\theta)/\partial\lambda$, and $j_{\lambda\lambda}(\theta) = -\partial^2\ell(\theta;s)/\partial\lambda\partial\lambda^\top = -\partial^2\ell^0(\theta)/\partial\lambda\partial\lambda^\top = -\ell^0_{\lambda\lambda}(\theta)$. The ratio of (2.5) and (2.6) when $s_2 = 0$ gives the following saddlepoint approximation for the density of s_1 given $s_2 = 0$, also called double saddlepoint approximation, for the reduced model in $\mathbb{R}^d$:

$$h(s;\psi) = c\exp[\ell(\hat{\theta}^0_\psi;s) - \ell\{\hat{\theta}(s);s\}]\,|j_{\theta\theta}\{\hat{\theta}(s)\}|^{-1/2}, \quad s \in \mathcal{L}^0, \quad (2.7)$$

where the normalizing constant c includes all factors not depending on s_1, and $\mathcal{L}^0$ is the d-dimensional plane described by setting $s_2 = 0$, or equivalently $\hat{\theta}_\psi = \hat{\theta}^0_\psi$. The relative error of the approximation (2.7) is typically of order $O(n^{-1})$, with n denoting the number of independent observations, but it reduces to $O(n^{-3/2})$ after re-normalization. For a comprehensive review of saddlepoint approximations and their statistical applications, see Butler (2007). The following example with a scalar parameter of interest ($d = 1$) illustrates the use of the tilted log-likelihood function (2.4) in the derivation of the saddlepoint approximation (2.7).

Example 1 (Univariate normal distribution). Let $y_1, \ldots, y_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. The log-likelihood function in exponential family form is

$$\ell(\theta) = \ell(\psi,\lambda) = \psi u_1 + \lambda u_2 + \frac{n}{2}\log(-2\psi) + \frac{n\lambda^2}{4\psi},$$

where $\theta = (\psi,\lambda) = (-1/2\sigma^2, \mu/\sigma^2)$ is the canonical parameter and $u = (u_1,u_2) = (\sum_i y_i^2, \sum_i y_i)$ is the minimal sufficient statistic with observed value $u^0 = (u_1^0, u_2^0)$. The tilted log-likelihood (2.4), expressed as a function of the centered sufficient statistic $s = u - u^0$, is

$$\ell(\theta;s) = \ell(\psi,\lambda;s) = \psi(s_1 + u_1^0) + \lambda(s_2 + u_2^0) + \frac{n}{2}\log(-2\psi) + \frac{n\lambda^2}{4\psi}.$$

After some algebra, the unnormalized saddlepoint approximation (2.7) in $\mathcal{L}^0 = \{(s_1,s_2) : s_1 > -u_1^0 + (u_2^0)^2/n, s_2 = 0\}$ can be written as

$$h(s;\psi) \propto \exp\left[\psi\left\{s_1 + u_1^0 - \frac{(u_2^0)^2}{n}\right\}\right]\left\{s_1 + u_1^0 - \frac{(u_2^0)^2}{n}\right\}^{(n-1)/2-1}, \qquad (2.8)$$

where $u_1^0 - (u_2^0)^2/n$ is n times the unadjusted sample variance. In this simple case, the saddlepoint approximation is exact: (2.8) coincides with the kernel of a $\chi^2_{n-1}/(-2\psi)$ distribution, which is the exact conditional distribution of $s_1 = u_1 - u_1^0$ given $s_2 = u_2 - u_2^0 = 0$. This is consistent with the more general result in McCormack et al. (2019).

The second dimensionality reduction from d to one, not needed in the previous example, consists of constructing a one-dimensional conditional distribution for s along the direction indicated by the data. With this aim, denote by s_ψ the expectation of s under model (2.7) if H_ψ holds, that is, the value of s for which $\theta = \hat\theta^0_\psi$ is the constrained ML estimate

$$s_\psi = -\ell^0_\theta(\hat\theta^0_\psi) = \begin{pmatrix} -\ell^0_\psi(\hat\theta^0_\psi) \\ 0 \end{pmatrix}, \qquad (2.9)$$

depending on the observed data point y^0. The line $\mathcal{L}^*$ in $\mathcal{L}^0$, which joins the observed value $s^0 = 0$ and the expected value s_ψ, can be parameterized by a scalar $t \in \mathbb{R}$ as follows:

$$s(t) = s_\psi + t(s^0 - s_\psi) = (1-t)s_\psi\,,$$

and, consequently, the ML estimate $\hat\theta(s)$ in (2.7) can vary with $s(t)$. The approximation (2.7) constrained to $\mathcal{L}^*$ is used to compute the p-value, the probability that $s(t)$ is as far or farther from s_ψ than is the observed value $s^0 = 0$. The directional p-value, which measures the deviation from H_ψ along the line $\mathcal{L}^*$, is thus

$$p(\psi) = \frac{\int_1^{t_{\sup}} t^{d-1} h\{s(t);\psi\}\,\mathrm{d}t}{\int_0^{t_{\sup}} t^{d-1} h\{s(t);\psi\}\,\mathrm{d}t}\,, \qquad (2.10)$$

where $t = 0$ and $t = 1$ correspond respectively to $s = s_\psi$ and to the observed value $s^0 = 0$. The factor t^{d-1} results from the Jacobian of the transformation from the variable $s \in \mathcal{L}^0$ to polar coordinates $(\|s\|, s/\|s\|)$ (Davison et al., 2014, Sec. 3.2). The upper limit of the integrals in (2.10) is the largest value of t for which the ML estimator corresponding to $s(t)$ exists, and in some situations can be determined analytically. The directional p-value in one dimension gives the probability to the right of the observed value, conditional on the observed value being to the right of the expected value under H_ψ, that is, the probability in the right tail of the distribution. In higher dimensions the p-value is the probability of being "further out" on the line connecting the expected value under H_ψ to the observed value, conditional on being on that line (Davison et al., 2014, Sec. 2).

As in Davison et al. (2014, Sec. 3.2), the relative error of formula (2.10) inherits that of the saddlepoint approximation (2.7) after re-normalization, so is typically $O(n^{-3/2})$ in continuous models. When the re-normalized saddlepoint approximation is exact, then the directional test is also exact, because the re-normalization is automatically incorporated in (2.10). McCormack et al. (2019) established this exactness for a number of tests for multivariate normal models, and Huang, Di Caterina and Sartori (2022) were able to prove exactness for the case of testing a saturated Gaussian graphical model in Davison et al. (2014, Sec. 5.3). The exactness in our setting is shown in Section 3.3 for chordal graphs. In addition, numerical results in the last simulation scenario of Section 5 illustrate the extreme accuracy of the directional approach, even when the alternative graph is non-chordal.

Using the notation established in this section, we also give the form of the term $\gamma(\psi)$ appearing in (2.2) under exponential family models. Specifically, equation (13) in Skovgaard (2001) is

$$\gamma(\psi) = \frac{\{(s-s_\psi)^\top j_{\theta\theta}^{-1}(\hat\theta_\psi)(s-s_\psi)\}^{d/2}}{w^{d/2-1}(\hat\theta-\hat\theta_\psi)^\top(s-s_\psi)} \left\{ \frac{|j_{\theta\theta}(\hat\theta_\psi)|}{|j_{\theta\theta}(\hat\theta)|} \right\}^{1/2}, \tag{2.11}$$

evaluated at $s=0$ when computing the corresponding observed p-value.

3. Directional tests for Gaussian graphical models

3.1. Notation and setup

Gaussian graphical models are very useful for describing normal multivariate distributions using the nodes and edges of a related graph. The nodes correspond to variables, and the lack of an edge between two nodes models the conditional independence of the two variables, given the remaining ones. This corresponds to a zero entry in the concentration (inverse covariance) matrix (Lauritzen, 1996), and covariance selection involves identifying these conditional independencies.

Let $y_1, \ldots, y_n$ be a random sample from the q-variate normal distribution $N_q(\mu, \Omega^{-1})$, where the mean is $\mu \in \mathbb{R}^q$ and the $q \times q$ concentration matrix Ω is positive definite. The log-likelihood function for (μ, Ω) is

$$\ell(\mu, \Omega; y) = \frac{n}{2}\log|\Omega| - \frac{1}{2}\mathrm{tr}(\Omega y^\top y) + 1_n^\top y\,\Omega\mu - \frac{n}{2}\mu^\top\Omega\mu\,, \tag{3.1}$$

where y denotes the $n \times q$ matrix with lth row vector $y_l^\top$, and 1_n is a $n \times 1$ vector of ones. The ML estimates of μ and Ω are

$$\hat\mu = \frac{y^\top 1_n}{n}, \quad \hat\Omega = \left(\frac{y^\top y}{n} - \frac{y^\top 1_n 1_n^\top y}{n^2}\right)^{-1}.$$

For covariance selection, the mean parameter is not of direct interest, so we focus instead on the marginal distribution of the ML estimator for the covariance matrix $\hat{\Omega}^{-1} \sim W_q(n-1, \Omega^{-1}/n)$, where W_q denotes the Wishart random variable of order q. The marginal log-likelihood function for Ω

$$\ell(\Omega; y) = \frac{n-1}{2} \log |\Omega| - \frac{n}{2} \mathrm{tr}(\Omega \hat{\Omega}^{-1}) \,,$$

sometimes referred to as restricted log-likelihood or REML, can then be used to carry out inference just on the concentration matrix. The directional p-value for testing constraints on Ω in Section 3.3 is equal to that obtained from the full log-likelihood function (3.1), because of the independence between $\hat{\mu}$ and $\hat{\Omega}$. It is also convenient to exploit the symmetry of the concentration matrix, and express the restricted log-likelihood as

$$\ell(\omega; u) = \frac{n-1}{2} \log |\Omega| - \frac{n-1}{2} \omega^\top J u \,, \tag{3.2}$$

where $\omega = \mathrm{vech}\, \Omega$, $u = n/(n-1)\mathrm{vech}\, \hat{\Omega}^{-1}$, and the matrix $J = G^\top G$ is diagonal with elements equal to either one or two. If A is a $q \times q$ symmetric matrix, $\mathrm{vec}\, A$ is the $q^2 \times 1$ vector that stacks the columns of A, whereas $\mathrm{vech}\, A$ retains only the $q^* = q(q+1)/2$ entries in the lower triangle of A. The two vectors are linked by the relationship $\mathrm{vec}\, A = G\, \mathrm{vech}\, A$, which also gives the $q^2 \times q^*$ duplication matrix G (see, e.g., Abadir and Magnus, 2005, Sec. 11.3).

In the saturated case addressed by Davison et al. (2014, Sec. 5.3), that is, a complete graph in which Ω has no particular a priori structure, the condition $n > q$ is required for the existence of $\hat{\Omega}$ Lauritzen (1996, Thm. 5.1). On the other hand, if the graph is incomplete, with some zero off-diagonal entries in Ω, the ML estimate exists if n is larger than the maximal clique size of the hypothesized graph or its decomposable version (Buhl, 1993; Lauritzen, 1996, Sec. 5.3.2). In what follows, we focus on comparing nested unsaturated models corresponding to nested incomplete graphs. Therefore, we allow the sample size n to be smaller than the number of nodes q, but large enough for the ML estimate of the concentration matrix to exist under the alternative model under study (cf., Sec. 3.2).

3.2. Likelihood quantities for unsaturated models

Suppose some off-diagonal elements Ω_{ij}, for $1 \leq i < j \leq q$, in the concentration matrix are known to be zero, meaning that the underlying graph is known to be incomplete. As in Roverato and Whittaker (1996), we can rearrange the elements of ω, u and the leading diagonal of J to simplify the calculations. Specifically, defining the edge sets

$$k = \{(i,j) : \Omega_{ij} \neq 0, i \leq j\} \quad \text{and} \quad h = \{(i,j) : \Omega_{ij} = 0, i < j\} \,, \tag{3.3}$$

and giving any ordering to k and h such that

$$k = \{k_1, k_2, \ldots, k_p\} \quad \text{and} \quad h = \{h_1, h_2, \ldots, h_w\},$$

it is possible to define

$$\omega = \begin{pmatrix} \omega_k \\ \omega_h \end{pmatrix}, \quad u = \begin{pmatrix} u_k \\ u_h \end{pmatrix}, \quad J = \begin{pmatrix} J_{kk} & 0 \\ 0 & J_{hh} \end{pmatrix}.$$

Because in unsaturated models $\omega_h = 0$, we can write $\Omega = \Omega_k = \Omega(\omega_k)$. Thus, the log-likelihood (3.2) becomes

$$\ell(\omega_k; u_k) = \frac{n-1}{2} \log |\Omega_k| - \frac{n-1}{2} \omega_k^\top J_{kk} u_k \,, \tag{3.4}$$

which is a function of the p-dimensional canonical parameter $\theta = \omega_k$ only, with $p > q$. Differentiation of (3.4) with respect to ω_k leads to the score function

$$\ell_{\omega_k}(\omega_k) = \frac{n-1}{2} J_{kk}(\sigma_k - u_k) \,,$$

where σ_k is the partition of $\sigma = \operatorname{vech} \Omega_k^{-1}$ obtained from (3.3). Solving the score equation leads to $\hat{\sigma}_k = u_k$ and to the corresponding ML estimate $\hat{\omega}_k$, usually derived numerically (see Davison et al., 2014, Sec. 5.3).

Because the observed and expected information matrices are equal in canonical exponential families, from the results in Roverato and Whittaker (1996, Sec. 3) it follows that

$$j_{\omega_k \omega_k}(\omega_k) = \frac{n-1}{4} J_{kk} \operatorname{Iss}(\Omega_k^{-1})_{kk} J_{kk} \,, \tag{3.5}$$

where $\operatorname{Iss}(\Omega_k^{-1})_{kk}$ is a $p \times p$ partition of the Isserlis matrix of the covariance matrix $\Sigma = \Omega_k^{-1}$ (Isserlis, 1918). The entries of $\operatorname{Iss}(\Sigma)_{kk}$ are

$$\operatorname{Cov}(u_{ij}, u_{rs}) = \Sigma_{ir}\Sigma_{js} + \Sigma_{is}\Sigma_{jr} \,,$$

with $(i, j), (r, s) \in k$.

3.3. Comparison of nested unsaturated models

Consider now the partition $\omega_k = (\psi, \lambda)$ of the canonical parameter, where ψ is the component of interest with dimension $d \leq p - q$. The null hypothesis $H_0 : \psi = \psi_0 = 0$ tests whether d additional off-diagonal entries Ω_{ij}, for $i < j$, are zero. Hence, the reduced null model is nested in the alternative unsaturated model of Section 3.2. Starting from (3.4), the log-likelihood ratio statistic for testing H_0 is

$$w(\psi_0) = -(n-1) \log |\hat{\Omega}_k^{-1} \hat{\Omega}_0| \,, \tag{3.6}$$

where $\hat{\Omega}_k = \Omega(\hat{\omega}_k)$ is the ML estimate of Ω obtained from (3.4), and $\hat{\Omega}_0 = \Omega(\hat{\omega}_{k0})$ is its constrained ML estimate under H_0, with $\hat{\omega}_{k0} = (0, \hat{\lambda}_0)$. The null asymptotic distribution of $w(\psi_0)$ is χ^2_d, assuming p and d fixed with n that goes to infinity.

For the directional p-value that discriminates between two nested Gaussian graphical models, as specified in (2.9), we first find the expected value of s under H_0:

$$s_{\psi_0} = -\ell_{\omega_k}(\hat{\omega}_{k0}) = \frac{n-1}{2} J_{kk}(u_k - \hat{\sigma}_{k0}) \,,$$

where $\hat{\sigma}_{k0} = \operatorname{vech} \hat{\Omega}_0^{-1}$. Then, the log-likelihood function (2.3) along the line $s(t) = (1-t)s_{\psi_0}$ follows from (3.4):

$$\ell\{\omega_k; s(t)\} = \frac{n-1}{2} \log|\Omega_k| - \frac{n-1}{2} \omega_k^\top J_{kk}\{\hat{\sigma}_{k0} + t(u_k - \hat{\sigma}_{k0})\} \,. \tag{3.7}$$

The maximization of (3.7) entails that $\hat{\sigma}_k\{s(t)\} = \hat{\sigma}_k(t) = \hat{\sigma}_{k0} + t(u_k - \hat{\sigma}_{k0})$ or, equivalently,

$$\hat{\Omega}_k^{-1}\{s(t)\} = \hat{\Omega}_k^{-1}(t) = t\hat{\Omega}_k^{-1} + (1-t)\hat{\Omega}_0^{-1} \,. \tag{3.8}$$

Given that $\hat{\Omega}_k(t) = \Omega\{\hat{\omega}_k(t)\}$, by taking the inverse of the matrix from the left-hand side of (3.8), the value of $\hat{\omega}_k(t)$ is obtained accordingly. The replacement of ω_k in (3.7) with $\hat{\omega}_k(t)$ and $\hat{\omega}_{k0}$, respectively, delivers the result

$$\begin{aligned}
&\exp[\ell\{\hat{\omega}_{k0}; s(t)\} - \ell\{\hat{\omega}_k(t); s(t)\}] \\
&\propto |\hat{\Omega}_k(t)|^{-(n-1)/2} \exp\left[\frac{n-1}{2}\{\hat{\omega}_k(t) - \hat{\omega}_{k0}\}^\top J_{kk}\hat{\sigma}_k(t)\right] \\
&\propto |\hat{\Omega}_k(t)|^{-(n-1)/2} \,,
\end{aligned}$$

because the function $\{\hat{\omega}_k(t) - \hat{\omega}_{k0}\}^\top J_{kk}\hat{\sigma}_k(t)$ is zero (see proof in Appendix A.2). By (3.5), we obtain $|j_{\omega_k\omega_k}(\omega_k)| \propto |\mathrm{Iss}(\Omega_k^{-1})_{kk}|$ and, consequently,

$$|j_{\omega_k\omega_k}\{\hat{\omega}_k(t)\}|^{-1/2} \propto |\mathrm{Iss}\{\hat{\Omega}_k^{-1}(t)\}_{kk}|^{-1/2} \,.$$

Thus, following expression (2.7), the directional test is based on $p(\psi_0)$ in (2.10), with

$$h\{s(t); \psi_0\} \propto |\hat{\Omega}_k^{-1}(t)|^{(n-1)/2} |\mathrm{Iss}\{\hat{\Omega}_k^{-1}(t)\}_{kk}|^{-1/2} \,, \tag{3.9}$$

and the analytical value of $t_{\sup}$ is calculated as in Section 4.2. If the alternative model were saturated, with q^*-vector $\omega_k = \omega$, then

$$|\mathrm{Iss}\{\hat{\Omega}_k^{-1}(t)\}_{kk}| = |\mathrm{Iss}\{\hat{\Omega}_k^{-1}(t)\}| = 2^q |\hat{\Omega}_k^{-1}(t)|^{q+1} \,,$$

according to the general expression for computing the determinant of the Isserlis matrix (Roverato and Whittaker, 1998, Sec. 2). In this case, (3.9) reduces to

$$h\{s(t); \psi_0\} \propto |\hat{\Omega}_k^{-1}(t)|^{(n-1)/2} |\hat{\Omega}_k^{-1}(t)|^{-(q+1)/2} = |\hat{\Omega}_k^{-1}(t)|^{(n-q-2)/2} \,,$$

which agrees with the simpler result obtained by Davison et al. (2014, Sec. 5.3) for testing the absence of some connections in a complete graph.

Expression (3.9) gives the unnormalized saddlepoint approximation to the distribution of $s(t)$ in $\mathcal{L}^*$. The following theorem, the proof of which is deferred to Appendix A.1, states when (3.9) is also the unnormalized exact null conditional density of $s(t)$ in $\mathcal{L}^*$.

Theorem 1. *Let $Y \sim N_q(\mu, \Omega^{-1})$ denote a Gaussian graphical model with log-likelihood* (3.4)*. If the induced incomplete graph is chordal, then* (3.9) *gives the unnormalized exact conditional density of $s(t)$ in $\mathcal{L}^*$ under $H_0 : \psi = \psi_0 = 0$.*

The normalizing constant simplifies in the ratio of integrals in (2.10), so the approximation error when calculating the directional p-value stems only from the one-dimensional numerical integrations. Thus, in Gaussian graphical models that describe chordal graphs, the saddlepoint approximation to the null conditional density of the sufficient statistic is exact. Consequently, when we test for a reduced graph, the resulting directional p-value is exactly uniformly distributed under the null hypothesis $H_0 : \psi = \psi_0 = 0$.

Monte Carlo experiments in Section 5 support this theoretical result, and empirically show that the directional p-value stays remarkably accurate in the last simulation scenario based on non-chordal graphs. When the exactness does not hold, the relative error of the saddlepoint approximation is still of order $O(n^{-3/2})$, as opposed to the absolute error of order $O(n^{-1})$ of the chi-squared approximation to the distribution of $w(\psi_0)$.

Finally, we give the term $\gamma(\psi)$ in (2.11) from Skovgaard's (2001) modified likelihood ratio statistics (2.2):

$$\gamma(\psi_0) = \frac{2\{(\hat{\sigma}_{k0} - \hat{\sigma}_k)^\top \mathrm{Iss}(\hat{\Omega}_0^{-1})_{kk}^{-1}(\hat{\sigma}_{k0} - \hat{\sigma}_k)\}^{d/2}}{\{-\log|\hat{\Omega}_k^{-1}\hat{\Omega}_0|\}^{d/2-1}(\hat{\omega}_k - \hat{\omega}_{k0})^\top J_{kk}(\hat{\sigma}_{k0} - \hat{\sigma}_k)} \left\{ \frac{|\mathrm{Iss}(\hat{\Omega}_0^{-1})_{kk}|}{|\mathrm{Iss}(\hat{\Omega}_k^{-1})_{kk}|} \right\}^{1/2}. \tag{3.10}$$

4. Computational Aspects

4.1. Calculation of the determinant of the Isserlis matrix

When the dimension p of the canonical parameter ω_k under the alternative model is smaller than q^*, but still relatively large, calculating the determinant of the matrix $\mathrm{Iss}\{\hat{\Omega}_k^{-1}(t)\}_{kk}$ in (3.9) can be computationally quite demanding. It is then advisable to exploit some useful results on the Isserlis matrix in order to speed up the computing time for the directional p-value.

Let A be a $q \times q$ symmetric invertible matrix. Roverato and Whittaker (1998, (15)), for any partition (k', k'') of the edge set k in (3.3) such that $k' \cup k'' = k$ and $k' \cap k'' = \bar{k}$, show that

$$|\mathrm{Iss}(A)_{kk}| = \frac{|\mathrm{Iss}(A)_{k'k'}||\mathrm{Iss}(A)_{k''k''}|}{|\mathrm{Iss}(A)_{\bar{k}\bar{k}}|},$$

which gives a convenient way to reduce the dimensions of the matrices. If, moreover, the graph induced by k is chordal with a vertex set decomposable into cliques $C_1, \ldots, C_K$ and separators $S_2, \ldots, S_K$, according to the definitions in Lauritzen (1996, Sec. 2.1), this can be further simplified to

$$|\mathrm{Iss}(A)_{kk}| = 2^q \frac{\prod_{i=1}^{K} |A_{C_i}|^{n_{C_i}+1}}{\prod_{i=2}^{K} |A_{S_i}|^{n_{S_i}+1}}, \tag{4.1}$$

where n_{C_i} and n_{S_i} denote the number of nodes in the ith clique and ith separator, respectively, and A_{C_i} and A_{S_i} are submatrices of A with rows and columns corresponding to the relative nodes (Roverato and Whittaker, 1998, Eq. 17).

4.2. Numerical integration

The upper bound $t_{\sup}$ in (2.10) is the largest value of t such that the ML estimate $\hat{\Omega}_k(t)$ is positive definite. By the same arguments as in Huang, Di Caterina and Sartori (2022, Lemma 4.1), this upper bound can be obtained explicitly as $t_{\sup} = 1/(1 - \nu_{(1)})$, where $\nu_{(1)}$ is the smallest of the q eigenvalues of $\hat{\Omega}_0\hat{\Omega}_k^{-1}$.

Moreover, writing the integrand in (2.10) as $\exp\{\bar{g}(t;\psi)\}$, where $\bar{g}(t;\psi) = (d-1)\log t + \log h\{s(t);\psi\}$, we can improve the numerical stability of the calculations using the equivalent formula

$$p(\psi) = \frac{\int_1^{t_{\sup}} \exp\{\bar{g}(t;\psi) - \bar{g}(\hat{t};\psi)\}\mathrm{d}t}{\int_0^{t_{\sup}} \exp\{\bar{g}(t;\psi) - \bar{g}(\hat{t};\psi)\}\mathrm{d}t}, \quad \text{where } \hat{t} = \operatorname*{arg\,sup}_{t\in[0,\ t_{\sup}]} \bar{g}(t;\psi).$$

We have also found that the integrand function can be concentrated around its mode, taking nonzero values in a shorter interval $[t_{\min}, t_{\max}] \subseteq [0, t_{\sup}]$. To address this, and deliver more stable numerical results, we use the Gauss–Hermite quadrature (Liu and Pierce, 1994) and integrate over $[t_{\min}, t_{\max}]$ only. As a result, we compute the directional p-value as

$$p(\psi) \;\doteq\; \frac{\int_1^{t_{\max}} \exp\{\bar{g}(t;\psi) - \bar{g}(\hat{t};\psi)\}\mathrm{d}t}{\int_{t_{\min}}^{t_{\max}} \exp\{\bar{g}(t;\psi) - \bar{g}(\hat{t};\psi)\}\mathrm{d}t}. \tag{4.2}$$

The choice $t_{\min} = \max\{0, \hat{t} - c/q(\hat{t};\psi)\}$ and $t_{\max} = \min\{\hat{t} + c/q(\hat{t};\psi), t_{\sup}\}$, where $q(t;\psi) = -\partial^2\bar{g}(t;\psi)/\partial t^2$ and c is a constant to be chosen, is reliable (cf., Huang, Di Caterina and Sartori, 2022, Sec. S1.3). The second derivative of the Isserlis determinant in the last factor of the integrand in (3.9) cannot be derived explicitly, and its numerical approximation may be unstable. In order to choose the width of the integration interval $[t_{\min}, t_{\max}]$, we then set the function $q(t;\psi)$ equal only

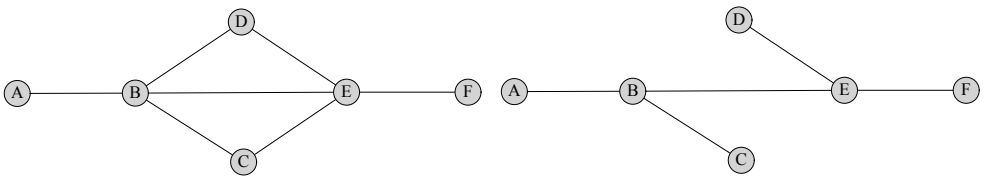

Figure 1. Graphs for the first simulation scenario, where the dimension of the parameter of interest is equal to $d = 3$. The alternative model for the chordal graph on the left is compared against the null model on the right.

to the second derivative of the first factor in (3.9), that is,

$$q(t;\psi) = -\frac{\partial^2 |\hat{\Omega}_k^{-1}(t)|^{(n-1)/2}}{\partial t^2} = \frac{d-1}{t^2} + \frac{n-1}{2}\sum_{i=1}^{q}\frac{(1-\nu_i)^2}{(1-t+t\nu_i)^2}\,.$$

In our numerical experiments, the value of c is chosen for each pair (n, q) by performing preliminary checks to ensure that the integration from $t_{\min}$ to $t_{\max}$ is equal to that over $[0, t_{\sup}]$, and then fixed for further simulations. This simplification was found useful only in settings when $n > q$, and cannot be applied if $\bar{g}(t;\psi)$ is monotonic in $[0, t_{\sup}]$. The directional p-value in that case has to be calculated directly using formula (2.10), but this happened only 21 times in the Monte Carlo experiments discussed below.

5. Simulation Studies

The performance of the directional approach in terms of covariance selection for Gaussian graphical models is examined here using simulation-based experiments. In the first scenario, the focus is on a small chordal graph with $q = 6$ nodes, similar to that in Dawid and Lauritzen (1993, Ex. 7.3). The two models under comparison, differing only by $d = 3$ edges, are presented in Figure 1. Monte Carlo simulations use 100,000 samples of size $n = 8$, generated under the null hypothesis. The empirical p-value distribution of the tests based on $w(\psi_0)$, $w^*(\psi_0)$, $w^{**}(\psi_0)$, and the directional procedure is shown in the left plot of Figure 2 with respect to the reference uniform distribution, focusing on the interval $(0, 0.1)$. The right plot compares the relative errors of the three most accurate methods. Despite the simplicity of the example, the likelihood ratio statistic leads to too many rejections of the null hypothesis, because n is relatively small. The higher-order modifications remedy this, but the directional approach allows an exact control of the size of the test, up to numerical and Monte Carlo errors.

The inferential benefits of our proposal over the omnibus likelihood-based competitors are particularly evident with high magnitudes of q and d. The second scenario is based on the data of Kenward (1987, Tab. 1), from a study on intestinal parasites of 60 calves, where the weight in kilograms of each bovine was

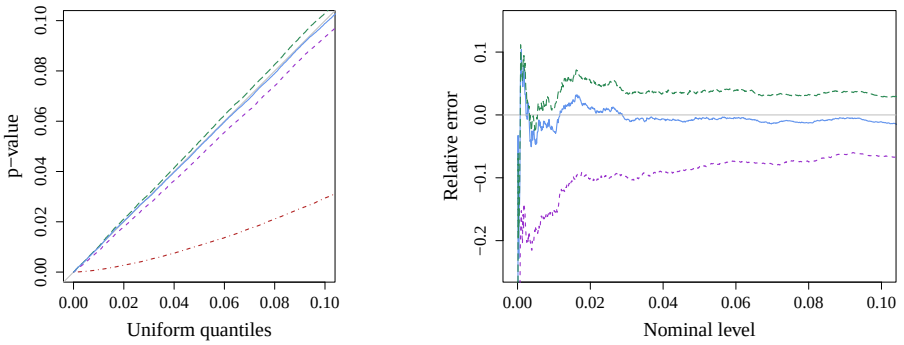

Figure 2. Results based on 100,000 samples simulated under the null model displayed on the right side of Figure 1, with $n = 8$ and $q = 6$. On the left, ordered empirical p-values $\hat{p}_{(i)}$ $(i = 1, \ldots, 100000)$ smaller than 0.1 are compared with the uniform distribution on the diagonal for w (dot-dashed), w^* (dashed), w^{**} (long-dashed), and the directional test (solid). On the right, the corresponding relative errors $\{\hat{p}_{(i)} - (i/n)\}/(i/n)$ are plotted in a similar fashion only for w^*, w^{**}, and the directional method.

recorded on 11 occasions during the grazing season. To enable a comparison with the findings of Davison et al. (2014, Sec. 5.3), who could only test the saturated model, we draw 100,000 samples of size $n = 60$ from a q-variate Gaussian random variable under the hypothesis of first-order Markovian dependence MD(1), with a tridiagonal concentration matrix. For each $q \in \{11, 30, 50\}$, the null hypothesis H_0 : MD(1) is tested against four alternative unsaturated structures, also using $w(\psi_0)$, $w^*(\psi_0)$, and $w^{**}(\psi_0)$. These Markovian dependence models of order m under H_1 : MD(m) with $1 < m < q - 1$ correspond to so-called band concentration matrices, the nonzero entries of which are confined to m diagonals on either side of the main one. The orders m are chosen to check the behavior of the various methods for a wide range of dimensions d of the parameter of interest, and consequently of the nuisance component. Because the Markovian structure induces a chordal graph, the simplification (4.1) is particularly useful for computing the directional p-values with such a high-dimensional parameter of interest.

Table 1 reports our experimental results obtained when $q = 11$, as in the original data set, and Tables 2 and 3 refer to cases with data simulated using a larger covariance matrix, $q = 30$ and $q = 50$, respectively. In line with our theoretical findings, the empirical distribution of the directional p-values is essentially uniform in all settings, and almost unaffected by the size of q and d. The usual likelihood ratio statistic $w(\psi_0)$ is highly sensitive to the dimension of both ψ and λ; its adjustments $w^*(\psi_0)$ and, particularly, $w^{**}(\psi_0)$ seem to suffer from the increasing dimension d of the parameter of interest. Tables 2 and 3 clearly indicate that, as d grows, the test based on $w(\psi_0)$ becomes too liberal, and those based on $w^*(\psi_0)$ and $w^{**}(\psi_0)$ become too conservative. For

Table 1. Empirical p-value distributions (%) based on 100,000 replications. The first-order Markovian model under $H_0 : \mathrm{MD}(1)$ is tested against different Markovian models of orders $m \in \{2, 3, 6, 9\}$ under $H_1 : \mathrm{MD}(m)$, when $n = 60$ observations of a graph with $q = 11$ nodes are available.

Nominal (%)	1.0	2.5	5.0	10.0	25.0	50.0	75.0	90.0	95.0	97.5	99.0
vs MD(2), $d = 9$											
Likelihood ratio, (3.6)	1.4	3.3	6.3	12.0	28.4	53.7	77.5	91.2	95.6	97.8	99.1
Skovgaard's w^*, (3.10)	1.0	2.5	5.1	10.0	25.1	50.2	75.1	89.9	94.9	97.4	99.0
Skovgaard's w^{**}, (3.10)	1.0	2.5	5.1	10.0	25.1	50.2	75.1	89.9	94.9	97.4	99.0
Directional, (4.2)	1.0	2.5	5.1	10.0	25.2	50.3	75.2	90.1	95.0	97.5	99.0
vs MD(3), $d = 17$											
Likelihood ratio, (3.6)	1.8	3.9	7.2	13.5	30.4	56.1	79.3	92.0	96.0	98.1	99.2
Skovgaard's w^*, (3.10)	1.1	2.6	5.0	10.0	24.6	49.6	74.6	89.6	94.7	97.3	98.9
Skovgaard's w^{**}, (3.10)	1.0	2.5	5.0	9.9	24.5	49.5	74.5	89.5	94.7	97.2	98.9
Directional, (4.2)	1.0	2.6	5.1	10.1	25.0	50.3	75.4	90.2	95.0	97.5	99.0
vs MD(6), $d = 35$											
Likelihood ratio, (3.6)	2.5	5.5	9.8	17.4	36.2	62.2	83.3	94.0	97.2	98.6	99.5
Skovgaard's w^*, (3.10)	0.8	2.1	4.3	8.8	22.4	46.4	71.7	87.8	93.6	96.6	98.5
Skovgaard's w^{**}, (3.10)	0.8	2.1	4.2	8.6	22.0	45.9	71.2	87.5	93.4	96.4	98.5
Directional, (4.2)	1.0	2.5	4.9	10.0	25.0	50.3	75.3	90.2	95.1	97.5	99.0
vs MD(9), $d = 44$											
Likelihood ratio, (3.6)	3.3	6.9	12.0	20.6	40.8	66.2	85.9	95.2	97.8	99.0	99.6
Skovgaard's w^*, (3.10)	0.7	1.8	3.7	7.8	20.7	43.7	69.1	86.3	92.6	96.1	98.2
Skovgaard's w^{**}, (3.10)	0.7	1.8	3.6	7.5	20.1	42.8	68.2	85.7	92.2	95.8	98.1
Directional, (4.2)	1.0	2.4	4.9	9.9	25.2	50.0	75.0	90.1	95.1	97.5	99.0
Standard error	0.0	0.0	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.0	0.0

the intermediate case $q = 30$, the leftmost panels of Figure 3 compare the null empirical distribution of the directional p-values with those from $w(\psi_0)$, $w^*(\psi_0)$, and $w^{**}(\psi_0)$. The almost perfect agreement of our proposal with the benchmark uniform distribution given by the diagonal of the panels is apparent.

Before proceeding, we focus on the implementation of formula (4.1) to obtain the determinant of the Isserlis matrix of Ω_k^{-1}, estimated under the alternative hypothesis. When multiplying the determinants of many square matrices of moderate order, some propagation of numerical errors can occur. In our experiments, this is visible, to a certain extent, in the intermediate sections of Tables 2 and 3, when the performance of the directional tests seems slightly worse than in the remaining sections. Indeed, when the null is tested against more extreme Markovian models, the matrices in (4.1) are either many, but small (top section) or large, but few (bottom section). Thus, the final product of their determinants is not overly affected by numerical errors. That being said, note that in all settings the directional approach remains remarkably accurate, significantly improving on the competing testing procedures.

Table 2. Empirical p-value distributions (%) based on 100,000 replications. The first-order Markovian model under $H_0 : \mathrm{MD}(1)$ is tested against different Markovian models of orders $m \in \{2, 9, 18, 28\}$ under $H_1 : \mathrm{MD}(m)$, when $n = 60$ observations of a graph with $q = 30$ nodes are available.

Nominal (%)	1.0	2.5	5.0	10.0	25.0	50.0	75.0	90.0	95.0	97.5	99.0
vs MD(2), $d = 28$											
Likelihood ratio, (3.6)	1.6	3.8	7.2	13.4	30.5	56.4	79.4	92.2	96.2	98.1	99.3
Skovgaard's w^*, (3.10)	1.0	2.5	5.0	10.0	24.9	50.0	75.1	90.1	95.0	97.5	99.0
Skovgaard's w^{**}, (3.10)	1.0	2.5	5.0	10.0	24.9	50.0	75.0	90.0	95.0	97.5	99.0
Directional, (4.2)	1.0	2.4	4.9	10.0	24.9	50.1	75.2	90.2	95.1	97.5	99.0
vs MD(9), $d = 196$											
Likelihood ratio, (3.6)	11.1	19.1	28.4	41.5	64.6	84.6	95.3	98.7	99.5	99.8	99.9
Skovgaard's w^*, (3.10)	0.3	0.9	2.0	4.4	13.3	32.3	57.9	78.5	87.1	92.5	96.4
Skovgaard's w^{**}, (3.10)	0.3	0.8	1.7	3.9	12.1	30.2	55.4	76.5	85.7	91.4	95.7
Directional, (4.2)	0.9	2.3	4.8	9.7	24.7	50.3	75.8	90.5	95.4	97.7	99.1
vs MD(18), $d = 340$											
Likelihood ratio, (3.6)	53.8	66.9	76.9	86.0	95.0	98.8	99.8	100.0	100.0	100.0	100.0
Skovgaard's w^*, (3.10)	0.0	0.1	0.3	0.7	3.0	10.7	27.4	48.8	62.1	72.8	82.9
Skovgaard's w^{**}, (3.10)	0.0	0.0	0.1	0.4	1.7	6.9	19.5	38.2	51.2	62.5	74.3
Directional, (4.2)	0.8	2.2	4.6	9.5	24.7	50.2	76.0	90.8	95.6	97.8	99.2
vs MD(28), $d = 405$											
Likelihood ratio, (3.6)	86.2	92.3	95.6	97.9	99.5	99.9	100.0	100.0	100.0	100.0	100.0
Skovgaard's w^*, (3.10)	0.0	0.0	0.0	0.2	0.9	4.3	13.8	30.0	42.5	53.9	67.0
Skovgaard's w^{**}, (3.10)	0.0	0.0	0.0	0.0	0.2	1.4	5.9	15.5	24.5	33.9	46.4
Directional, (4.2)	1.0	2.4	5.1	10.1	25.2	50.1	75.1	90.1	95.1	97.5	99.0
Standard error	0.0	0.0	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.0	0.0

The third simulation scenario considers a block-diagonal configuration of the concentration matrix under the null hypothesis. Here, 100,000 samples of size $n \in \{40, 60, 90, 120\}$ were drawn from a normal distribution with $q = 50$ components and covariance matrix $\Sigma_0 = \mathrm{diag}\{\Sigma_{01}, \Sigma_{01}\}$, with Σ_{01} a 25×25 sub-matrix with diagonal entries equal to one, and off-diagonal entries equal to 0.5. This condition clearly implies that $\Omega_0 = \Sigma_0^{-1}$ is also block diagonal, so that the first 25 nodes are conditionally (as well as unconditionally) independent of the last 25 nodes in the graph. On the other hand, our alternative model admits the existence of some conditional dependence between the two subsets of nodes. Specifically, in addition to the nonzero elements defined in Ω_0, we also suppose $\Omega_{ij} = \Omega_{ji} \neq 0$, for $i = 16, \ldots, 25$ and $j = 26, \ldots, 50$. It follows that the dimension of the parameter of interest is $d = 250$, and (4.1) can be used to speed up calculations of the Isserlis matrix associated with the chordal alternative incomplete graph.

Simulation results in this framework are presented in Table 4. Given the notable size of d, the relative performance of the approximations under comparison, in terms of the empirical p-value distribution, is analogous to that in the previous experiment, with the only exception that here the version $w^{**}(\psi_0)$ appears to

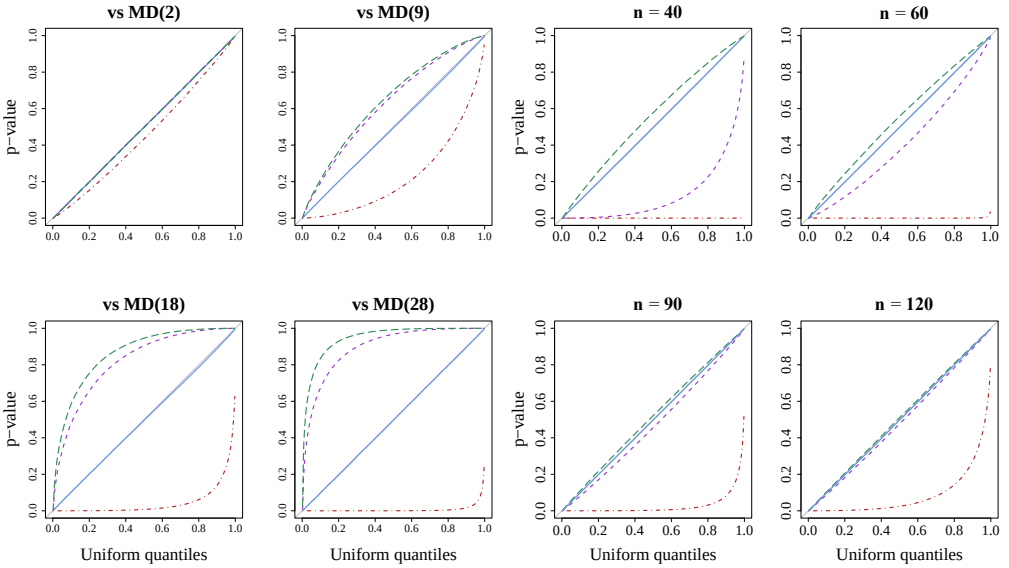

Figure 3. Results based on 100,000 simulated samples. In all eight panels, the empirical p-values obtained via w (dot-dashed), w^* (dashed), w^{**} (long-dashed), and the directional test (solid) are compared with the uniform distribution given by the diagonal. Leftmost panels: the model under H_0 : MD(1) assumes first-order Markovian dependence, with $n = 60$ and $q = 30$. The four panels correspond to different Markovian models under the alternative hypothesis H_1 and related dimensions of ψ: MD(2) and $d = 28$ (top left), MD(9) and $d = 196$ (top right), MD(18) and $d = 340$ (bottom left), and MD(28) and $d = 405$ (bottom right). Rightmost panels: the null model assuming a block-diagonal concentration matrix with $q = 50$ is tested against the same alternative hypothesis implying $d = 250$. The four panels correspond to different sample sizes: $n = 40$ (top left), $n = 60$ (top right), $n = 90$ (bottom left), and $n = 120$ (bottom right).

be, in general, more reliable than $w^*(\psi_0)$. Although the increase in sample size generates some accuracy improvements for all the competitors, as expected, the empirical directional p-value guarantees an almost perfect agreement with its theoretical uniform distribution for all values of n considered. The extreme liberality of the standard likelihood ratio test persists, Skovgaard's $w^*(\psi_0)$ does not correct it enough, and the version $w^{**}(\psi_0)$ overcorrects it. As before, the rightmost panels of Figure 3 show the p-values obtained using the likelihood ratio statistic, its modified versions, and the directional procedure.

As an empirical check of the accuracy of our proposal for non-decomposable models, in the fourth simulation scenario, we consider a small non-chordal graph with $q = 4$ nodes, as in Eriksen (1996, Sec. 4). Figure 4 displays the two models under comparison, which differ only by $d = 2$ edges. Setting the sample size to $n = 7$, 100000 artificial samples are simulated under the null hypothesis. As for the first scenario, the results are presented in the two panels of Figure 5. Because n is small with respect to q and d, the chi-squared approximation to the

Table 3. Empirical p-value distributions (%) based on 100,000 replications. The first-order Markovian model under $H_0 : \mathrm{MD}(1)$ is tested against different Markovian models of orders $m \in \{2, 16, 32, 48\}$ under $H_1 : \mathrm{MD}(m)$, when $n = 60$ observations of a graph with $q = 50$ nodes are available.

Nominal (%)	1.0	2.5	5.0	10.0	25.0	50.0	75.0	90.0	95.0	97.5	99.0
vs MD(2), $d = 48$											
Likelihood ratio, (3.6)	1.8	4.2	7.8	14.5	32.4	58.2	80.9	93.0	96.7	98.4	99.4
Skovgaard's w^*, (3.10)	1.0	2.5	5.0	9.9	25.1	50.1	74.9	90.0	95.1	97.5	99.0
Skovgaard's w^{**}, (3.10)	1.0	2.5	5.0	9.9	25.0	50.0	74.9	89.9	95.1	97.5	99.0
Directional, (4.2)	1.0	2.5	4.9	9.9	25.0	50.1	75.1	90.1	95.2	97.6	99.1
vs MD(16), $d = 615$											
Likelihood ratio, (3.6)	77.9	86.7	92.1	96.0	99.0	99.8	100.0	100.0	100.0	100.0	100.0
Skovgaard's w^*, (3.10)	0.0	0.0	0.1	0.2	1.1	5.1	15.9	33.3	46.2	57.9	70.6
Skovgaard's w^{**}, (3.10)	0.0	0.0	0.0	0.1	0.5	2.6	9.3	22.2	33.2	44.2	57.4
Directional, (4.2)	0.8	2.0	4.3	9.1	24.4	50.4	76.3	91.4	96.1	98.1	99.3
vs MD(32), $d = 1023$											
Likelihood ratio, (3.6)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Skovgaard's w^*, (3.10)	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.7	1.7	3.4	6.7
Skovgaard's w^{**}, (3.10)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2
Directional, (4.2)	0.5	1.4	3.4	8.0	23.5	51.7	78.6	92.8	96.9	98.7	99.5
vs MD(48), $d = 1175$											
Likelihood ratio, (3.6)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Skovgaard's w^*, (3.10)	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.4	1.0	2.0	4.0
Skovgaard's w^{**}, (3.10)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Directional, (4.2)	0.8	2.2	4.7	9.8	25.4	51.1	76.2	90.9	95.5	97.8	99.2
Standard error	0.0	0.0	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.0	0.0

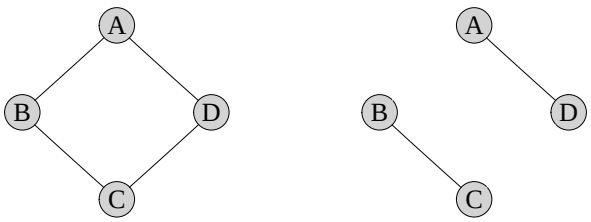

Figure 4. Graphs for the fourth simulation scenario, where the dimension of the parameter of interest is equal to $d = 2$. The alternative model for the non-chordal graph on the left is compared against the null model on the right.

distribution of the likelihood ratio statistic is misleading. The improved versions of w, especially w^* here, are more reliable. However, even in this application to a non-chordal graph, the superiority of the directional approach based on the accurate saddlepoint approximation is evident in terms of the relative error.

6. Applications

First, we examine the data set introduced in the second simulation scenario of Section 5 from the experiment about the control of intestinal parasites in cattle (Kenward, 1987, Tab. 1). However, here we focus separately on the two treatment

Table 4. Empirical p-value distributions (%) based on 100,000 replications. The two-block diagonal structure of the concentration matrix for a graph with $q = 50$ nodes is tested against a more complex structure including $d = 250$ additional edges.

Nominal (%)	1.0	2.5	5.0	10.0	25.0	50.0	75.0	90.0	95.0	97.5	99.0
$n = 40$											
Likelihood ratio, (3.6)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Skovgaard's w^*, (3.10)	27.1	39.1	50.6	63.5	81.7	93.7	98.5	99.7	99.9	100.0	100.0
Skovgaard's w^{**}, (3.10)	0.7	1.7	3.4	7.2	19.3	42.0	68.0	85.9	92.5	96.1	98.3
Directional, (4.2)	1.0	2.5	5.0	10.1	25.2	50.2	75.2	90.0	94.9	97.4	98.9
$n = 60$											
Likelihood ratio, (3.6)	98.4	99.3	99.7	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Skovgaard's w^*, (3.10)	2.4	5.3	9.6	17.3	36.6	62.9	84.2	94.6	97.6	98.9	99.6
Skovgaard's w^{**}, (3.10)	0.6	1.7	3.5	7.5	20.4	43.9	70.3	87.4	93.5	96.7	98.6
Directional, (4.2)	1.0	2.5	5.0	10.0	25.1	50.1	75.2	90.2	95.1	97.6	99.0
$n = 90$											
Likelihood ratio, (3.6)	65.9	77.1	85.0	91.5	97.3	99.4	99.9	100.0	100.0	100.0	100.0
Skovgaard's w^*, (3.10)	1.3	3.2	6.1	12.0	28.5	54.2	78.2	91.7	96.0	98.1	99.2
Skovgaard's w^{**}, (3.10)	0.8	2.1	4.3	8.9	23.0	47.6	73.2	89.0	94.5	97.2	98.8
Directional, (4.2)	0.9	2.5	5.0	10.1	25.0	50.1	75.1	90.1	95.1	97.6	99.0
$n = 120$											
Likelihood ratio, (3.6)	36.6	50.0	61.6	73.6	88.6	96.7	99.3	99.9	100.0	100.0	100.0
Skovgaard's w^*, (3.10)	1.1	2.9	5.6	11.0	26.8	52.2	76.5	90.9	95.5	97.8	99.1
Skovgaard's w^{**}, (3.10)	0.9	2.3	4.6	9.4	24.0	48.7	73.9	89.4	94.6	97.3	98.9
Directional, (4.2)	1.0	2.5	5.0	10.1	25.1	50.1	75.0	90.0	95.0	97.5	99.0
Standard error	0.0	0.0	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.0	0.0

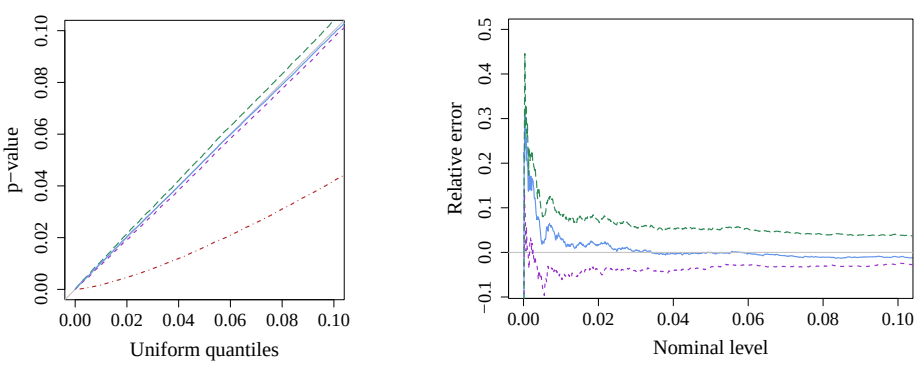

Figure 5. Results based on 100,000 samples simulated under the null model represented by the right graph of Figure 4 with $n = 7$ and $q = 4$. On the left, ordered empirical p-values $\hat{p}_{(i)}$ $(i = 1, \ldots, 100000)$ smaller than 0.1 are compared with the uniform distribution on the diagonal for w (dot-dashed), w^* (dashed), w^{**} (long-dashed), and the directional test (solid). On the right, the corresponding relative errors $\{\hat{p}_{(i)} - (i/n)\}/(i/n)$ are plotted in a similar fashion only for w^*, w^{**}, and the directional method.

groups with equal size $n = 30$ to investigate differences in the underlying temporal dynamics of growth. Recalling that each animal was weighed $q = 11$ consecutive times, we start by assuming a Markovian dependence of order $m = 3$, the simplest model accepted in a test against the saturated one by all the procedures under

analysis and in both groups. This model is then compared against the null hypothesis of first-order dependence, implying $d = 17$. For the calves randomly assigned to the first treatment, the likelihood ratio statistic is $w(\psi_0) = 28.384$ with p-value $= 0.041$, Skovgaard's modifications are $w^*(\psi_0) = 22.977$ with p-value $= 0.150$ and $w^{**}(\psi_0) = 22.691$ with p-value $= 0.160$ and the directional p-value is 0.111. For the second group, we get instead $w(\psi_0) = 31.895$ with p-value $= 0.016$, $w^*(\psi_0) = 30.055$ with p-value $= 0.026$, $w^{**}(\psi_0) = 30.028$ with p-value $= 0.026$ and directional p-value $= 0.029$. The standard likelihood ratio test is the only one to reject the MD(1) model at a 5% significance level for both treatments. Conversely, the other statistics recognize a different time pattern and indicate a more complex dependence of the weights in the second group.

We now consider microarray data from the biostatistical literature (see, e.g., Massa, Chiogna and Romualdi, 2010) that characterize gene expression signatures in acute lymphocytic leukemia cells associated with genotypic abnormalities in adult patients. The normalized version of such data, available in the package topologyGSA (Massa and Sales, 2016) of the R software (R Core Team, 2020), is especially useful for analyzing the B-cell receptor (BCR) signaling pathway, composed of $q = 35$ gene products. The observed samples are classified according to the presence of molecular rearrangements in their genetic profile.

The conversion of biological pathways into graphical models has become standard practice in biostatistics to separate and compare specific portions of the genetic process under examination. Based on the findings of Massa, Chiogna and Romualdi (2010), it is of interest to investigate whether the graph resulting from the well-known BCR signaling pathway in Figure 6 can be simplified further. The restricted graphical model under the null hypothesis in our analysis corresponds to the identified path starting from nodes CD22 and CD72 and ending at AP1, going through RasGRP3, Ras, Raf, MEK1/2, and ERK enzymes. This comparison implies testing the lack of $d = 12$ edges, and can be carried out on the subset of patients not suffering from so-called BCR/ABL rearrangements. With $n = 41$, we obtain $w(\psi_0) = 33.520$ with p-value $= 8.028 \times 10^{-4}$, $w^*(\psi_0) = 32.172$ with p-value $= 13.018 \times 10^{-4}$, $w^{**}(\psi_0) = 32.158$ with p-value $= 13.083 \times 10^{-4}$, and directional p-value $= 13.941 \times 10^{-4}$. Although all four methods indicate that the data are not consistent with the shorter biological path, the p-value from the usual likelihood ratio test $w(\psi_0)$ is relatively much smaller than the other three, and in these types of problems very small p-values are relevant. The agreement of Skovgaard's approximations with the directional p-value is consistent with our simulation results for small values of d with respect to n.

7. Discussion

We have provided theoretical and computational considerations for a likelihood-based approach to covariance selection in unsaturated Gaussian

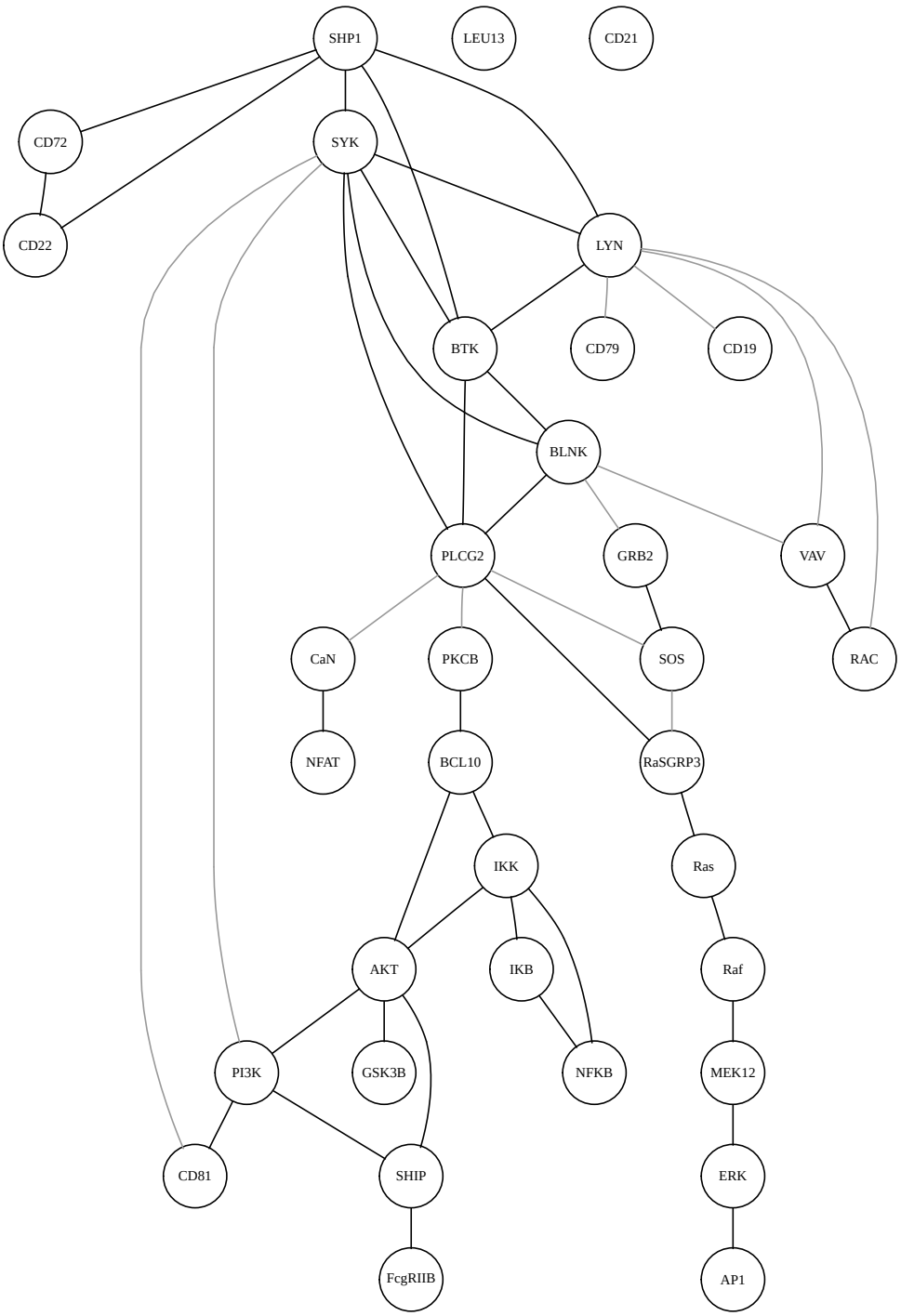

Figure 6. BCR signaling pathway involving $q = 35$ gene products. The interest is in testing whether a simpler path without the $d = 12$ gray edges can be identified.

graphical models. The directional test is based on the saddlepoint approximation to the conditional distribution of sufficient statistics in exponential family models. The saddlepoint approximation to the conditional density was derived explicitly, and proved to be exact within the important class of decomposable models for chordal graphs. Moreover, the computation of the directional p-value using one-dimensional numerical integration is made especially fast, as discussed in Section 4. Simulations in several scenarios, including situations with a high-dimensional parameter of interest and a large number of nuisance parameters, illustrate that the p-values from the directional test are uniformly distributed, up to the approximation error from the one-dimensional numerical integrations. These results confirm the theoretical exactness of the saddlepoint-type method with chordal graphs, even if the number of nodes is greater than the sample size. Our empirical findings suggest also that the saddlepoint approximation, despite not being exact, retains at least the usual accuracy for continuous models when non-chordal graphs are tested.

The likelihood ratio test and its improvements considered here (Skovgaard, 2001) are omnibus tests: the implicit alternative hypothesis is multidimensional. In contrast, the directional test uses information in the data to simplify the testing problem to one dimension. The saddlepoint approximation to this distribution incorporates an adjustment for the estimation of the nuisance parameters that has been found to be very effective in simpler problems (Pierce and Peters, 1992; Tang and Reid, 2020).

A natural question about directional tests is whether they entail a loss of power (Jensen, 2021). This is difficult to assess in simulations, because the alternative hypotheses are very high dimensional. We have concentrated on evaluating the size of the test, which Tables 1–4 show is very well controlled at conventional 0.05 and 0.01 levels, and well into the tails (Figures 2–3). We are not aware of other detailed discussions on the power of the likelihood ratio test for these complex Gaussian graphical models with high-dimensional alternatives. For high-dimensional normal distributions with $q/n \to (0, 1]$, Huang, Di Caterina and Sartori (2022, Sec. 5.3) evaluate the unconditional power of the directional test under a few settings. The performance strongly depends on the specific alternative hypothesis under analysis, so it is impossible to draw generally valid conclusions. Still, in those settings, the directional test proved to be uniformly more powerful than the likelihood ratio test and its modifications considered here. Note too that for simpler testing problems in the multivariate normal model, McCormack et al. (2019) showed that the directional test is equivalent to the uniformly most powerful invariant test based on the F-statistic or Hotelling's T^2-statistic.

The directional approach detailed here can be extended to graphical models for discrete data, such as those in Roverato (2017). However, because discreteness prevents the saddlepoint approximation from being exact, even upon normaliza-

tion, one might reasonably not expect the same accuracy of directional p-values observed in this work, at least in the most challenging testing problems.

The present methodology applies only to situations in which the number of observations is such that the ML estimate exists with probability one under the alternative hypothesis. In particular, the sample size must be greater than the maximal clique size of the hypothesized graph or its decomposable version (Buhl, 1993). The development of reliable likelihood-based testing procedures, omnibus or directional, in circumstances where the number of nodes is much larger than the number of observations is still an open problem, and thus left to future research.

Supplementary Material

Supplementary materials available at `https://github.com/cdicaterina/DirTestGGM.git` provide the data and the R code to reproduce all numerical results in the paper.

Acknowledgments

The authors are grateful to Alberto Roverato for useful discussions and suggestions on the R code to compute the Isserlis matrix. They also thank Davide Risso for his help with the genetic application, and Caizhu Huang for suggesting computational improvements.

Appendices

A.1. Proof of Theorem 1

We want to show that the saddlepoint approximation equals the exact conditional distribution of the sufficient statistic under H_0, up to some constant. The sufficient statistic in our setting is $s = u_k$, i.e. the partition corresponding to the non-zero elements in Ω_k of $u = n/(n-1)\mathrm{vech}\,\hat{\Omega}^{-1}$ where $\hat{\Omega}^{-1} = y^\top y/n - y^\top 1_n 1_n^\top y/n^2$ is the sample covariance matrix.

Substituting in the log-likelihood (3.4) the ML and constrained ML estimates of the canonical parameter ω_k obtained in Section 3, we get

$$\begin{aligned}\exp\{\ell(\hat{\omega}_0; s) - \ell(\hat{\omega}_k; s)\} &= \left(\frac{|\hat{\Omega}_0|}{|\hat{\Omega}_k|}\right)^{(n-1)/2} \exp\left\{\frac{n-1}{2}(\hat{\omega}_k - \hat{\omega}_{k0})^\top J_{kk}\hat{\sigma}_k\right\} \\ &= \left(\frac{|\hat{\Omega}_0|}{|\hat{\Omega}_k|}\right)^{(n-1)/2},\end{aligned}$$

since the exponential equals 1 (see Appendix A.2). Given equation (3.5) in Section 3.2 for $j_{\omega_k\omega_k}(\omega_k)$, we can then write the expression for the saddlepoint

approximation (2.7) in our setting as

$$h(s;\psi_0) \propto \left(\frac{|\hat{\Omega}_0|}{|\hat{\Omega}_k|}\right)^{(n-1)/2} |\mathrm{Iss}(\hat{\Omega}_k^{-1})_{kk}|^{-1/2}\,. \tag{A.1}$$

Consider now the density of $s = u_k$. This is the marginal density of p entries in $\hat{\Omega}^{-1}$, the sample covariance matrix with joint Wishart distribution $W_q(n-1, \Omega^{-1}/n)$. Solving the likelihood equation in Section 3.2 implies that $\hat{\sigma}_k = u_k = s$, hence these entries are the same as those in the corresponding entries of the matrix $\hat{\Omega}_k^{-1}$. We can obtain such a density for chordal graphs with vertex set decomposable into cliques $C_1, \ldots, C_K$ and separators $S_2, \ldots, S_K$ with cardinality n_{C_i} and n_{S_i}, respectively. Combining the results on the factorization of the joint density of $\hat{\Omega}^{-1}$ (Lauritzen, 1996, Eq. 5.45) and on the marginal Wishart distributions for the sub-matrices $\hat{\Omega}_{kC_i}^{-1} = (\hat{\Omega}_k^{-1})_{C_i}$ and $\hat{\Omega}_{kS_i}^{-1} = (\hat{\Omega}_k^{-1})_{S_i}$ (Dawid and Lauritzen, 1993, Sec. 7.3.1), under the null hypothesis $H_0 : \omega_k = (\psi, \lambda) = (0, \lambda)$ the true concentration matrix is Ω_0 and so we have:

$$\begin{aligned} &f(s;\Omega_0^{-1}) = \\ &2^{\{-(n-1)/2\}(\sum_{i=1}^K n_{C_i} - \sum_{i=2}^K n_{S_i})} \frac{\prod_{i=1}^K \Gamma_{n_{C_i}}\{(n-1)/2\}\,|\Omega_{0C_i}^{-1}|^{-(n-1)/2}|\hat{\Omega}_{kC_i}^{-1}|^{(n-2-n_{C_i})/2}}{\prod_{i=2}^K \Gamma_{n_{S_i}}\{(n-1)/2\}\,|\Omega_{0S_i}^{-1}|^{-(n-1)/2}|\hat{\Omega}_{kS_i}^{-1}|^{(n-2-n_{S_i})/2}} \\ &\cdot \exp\left[-\frac{n}{2}\left\{\sum_{i=1}^K \mathrm{tr}\left(\hat{\Omega}_{kC_i}^{-1}\Omega_{0C_i}\right) - \sum_{i=2}^K \mathrm{tr}\left(\hat{\Omega}_{kS_i}^{-1}\Omega_{0S_i}\right)\right\}\right]\,. \end{aligned}$$

Rearranging the factors in the previous formula and neglecting the constants, we can write

$$\begin{aligned} f(s;\Omega_0^{-1}) \propto &\left(\frac{\prod_{i=1}^K |\Omega_{0C_i}^{-1}|}{\prod_{i=2}^K |\Omega_{0S_i}^{-1}|}\right)^{-(n-1)/2} \left(\frac{\prod_{i=1}^K |\hat{\Omega}_{kC_i}^{-1}|}{\prod_{i=2}^K |\hat{\Omega}_{kS_i}^{-1}|}\right)^{(n-1)/2} \frac{\prod_{i=1}^K |\hat{\Omega}_{kC_i}^{-1}|^{-(n_{C_i}+1)/2}}{\prod_{i=2}^K |\hat{\Omega}_{kS_i}^{-1}|^{-(n_{S_i}+1)/2}} \\ &\cdot \exp\left\{-\frac{n}{2}\left[\sum_{i=1}^K \mathrm{tr}\left(\hat{\Omega}_{kC_i}^{-1}\Omega_{0C_i}\right) - \sum_{i=2}^K \mathrm{tr}\left(\hat{\Omega}_{kS_i}^{-1}\Omega_{0S_i}\right)\right]\right\}\,. \end{aligned}$$

We now use the decomposition of the graph (Lauritzen, 1996, p.145) to re-express the first two factors as a ratio of determinants, the result by Roverato and Whittaker (1998) mentioned in Section 4.1 to re-express the third factor as the determinant of the Isserlis matrix, and finally the property of the trace operator to re-express the fourth factor. Hence we have

$$\begin{aligned} f(s;\Omega_0^{-1}) \propto &\left(\frac{|\Omega_0|}{|\hat{\Omega}_k|}\right)^{(n-1)/2} |\mathrm{Iss}(\hat{\Omega}_k^{-1})_{kk}|^{-1/2} \\ &\cdot \exp\left\{-\frac{n}{2}\left[\sum_{i=1}^K \mathrm{tr}\left(\Omega_{0C_i}\hat{\Omega}_{kC_i}^{-1}\right) - \sum_{i=2}^K \mathrm{tr}\left(\Omega_{0S_i}\hat{\Omega}_{kS_i}^{-1}\right)\right]\right\} \end{aligned}$$

$$\propto \left(\frac{|\Omega_0|}{|\hat{\Omega}_k|}\right)^{(n-1)/2} |\mathrm{Iss}(\hat{\Omega}_k^{-1})_{kk}|^{-1/2} \exp\left\{-\frac{n}{2}\left[\mathrm{tr}\left(\Omega_0\hat{\Omega}_k^{-1}\right)\right]\right\},$$

where in the last step we have applied again the decomposition property based on the factorization of the density in chordal graphs (Lauritzen, 1996, Eq. 5.45) to find the final expression in the exponential of the last factor. The null conditional density of the sufficient statistic in $\mathcal{L}_0$ is given by setting $\omega_k = \hat{\omega}_{k0} = (0, \hat{\lambda}_0)$, or equivalently by fixing the concentration matrix under the null hypothesis Ω_0 at its constrained ML estimate $\hat{\Omega}_0$, i.e.

$$\begin{aligned} f(s; \hat{\Omega}_0^{-1}) &\propto \left(\frac{|\hat{\Omega}_0|}{|\hat{\Omega}_k|}\right)^{(n-1)/2} |\mathrm{Iss}(\hat{\Omega}_k^{-1})_{kk}|^{-1/2} \exp\left\{-\frac{n}{2}\left[\mathrm{tr}\left(\hat{\Omega}_0\hat{\Omega}_k^{-1}\right)\right]\right\} \\ &\propto \left(\frac{|\hat{\Omega}_0|}{|\hat{\Omega}_k|}\right)^{(n-1)/2} |\mathrm{Iss}(\hat{\Omega}_k^{-1})_{kk}|^{-1/2}. \end{aligned} \tag{A.2}$$

In the last step we have used $\mathrm{tr}(\hat{\Omega}_0\hat{\Omega}_k^{-1}) = \mathrm{tr}(\hat{\Omega}_0\hat{\Omega}_0^{-1}) = \mathrm{tr}(I_q) = q$ (see Appendix, Section A.2).

Equation (A.2) is equal to equation (A.1), up to a constant. The normalizing constant of $f(s; \hat{\Omega}_0^{-1})$ simplifies in the ratio of integrals in (2.10) for computing the directional p-value. The one-dimensional integration is allowed by further restricting on the line $\mathcal{L}^*$ in $\mathcal{L}^0$, identified by $\hat{\Omega}_k^{-1}(t) = t\hat{\Omega}_k^{-1} + (1-t)\hat{\Omega}_0^{-1}$. As the observed value $\hat{\Omega}_0$ of the concentration matrix under H_0 does not depend on t, we can integrate in the numerator and denominator of (2.10) the function

$$h(s(t); \psi_0) \propto |\hat{\Omega}_k^{-1}(t)|^{(n-1)/2} |\mathrm{Iss}\{\hat{\Omega}_k^{-1}(t)\}_{kk}|^{-1/2},$$

which was given in (3.9).

A.2. Proof of $\mathrm{tr}[\{\hat{\omega}(t) - \hat{\omega}_0\}^\top J\hat{\sigma}(t)] = 0$

We show that the scalar function

$$f(t) = \{\hat{\omega}_k(t) - \hat{\omega}_{k0}\}^\top J_{kk}\hat{\sigma}_k(t)$$

equals zero. Since $f(t) = \mathrm{tr}\{f(t)\}$ and the two models under comparison are nested, it is equivalent to prove that $\mathrm{tr}[\{\hat{\omega}(t) - \hat{\omega}_0\}^\top J\hat{\sigma}(t)]$ is constant in t, where

$$\hat{\omega}(t) = \begin{pmatrix} \hat{\omega}_k(t) \\ 0 \end{pmatrix}, \quad \hat{\omega}_0 = \begin{pmatrix} \hat{\omega}_{k0} \\ 0 \end{pmatrix}, \quad \hat{\sigma}(t) = \begin{pmatrix} \hat{\sigma}_k(t) \\ \hat{\sigma}_h(t) \end{pmatrix},$$

are all vectors of dimension q^*. Letting $\hat{\Omega}_k^{-1}(t) = \Sigma\{\hat{\sigma}(t)\}$, we have:

$$\begin{aligned}\mathrm{tr}[\{\hat{\omega}(t)-\hat{\omega}_0\}^\top J\hat{\sigma}(t)] &= \mathrm{tr}[\mathrm{vech}\,\{\hat{\Omega}_k(t)-\hat{\Omega}_0\}^\top G^\top G\,\mathrm{vech}\,\hat{\Omega}_k^{-1}(t)]\\ &= \mathrm{tr}[\{\hat{\Omega}_k(t)-\hat{\Omega}_0\}^\top\hat{\Omega}_k^{-1}(t)]\\ &= \mathrm{tr}[I_q-\hat{\Omega}_0\{t\hat{\Omega}_k^{-1}+(1-t)\hat{\Omega}_0^{-1}\}]\\ &= \mathrm{tr}(I_q)-t\mathrm{tr}(\hat{\Omega}_0\hat{\Omega}_k^{-1})-(1-t)\mathrm{tr}(I_q)\\ &= q-tq-(1-t)q=0\,.\end{aligned}$$

This uses basic matrix algebra (see, e.g., Abadir and Magnus, 2005) and the equality $\mathrm{tr}(\hat{\Omega}_0\hat{\Omega}_k^{-1}) = \mathrm{tr}(\hat{\Omega}_0\hat{\Omega}_0^{-1}) = \mathrm{tr}(I_q) = q$. The latter is due to the fact that the trace of the product of two symmetric matrices is the sum of the element-wise products and, by the ML equation, $\hat{\Omega}_k^{-1}$ differs from $\hat{\Omega}_0^{-1}$ only when the corresponding entries of $\hat{\Omega}_0$ are zero (cf. Eriksen, 1996, p.278).

In order to derive the same result for the scalar $f(1) = (\hat{\omega}_k - \hat{\omega}_{k0})^\top J_{kk}\hat{\sigma}_k$, the above calculations can be carried out imposing $t = 1$.

References

Abadir, K. M. and Magnus, J. R. (2005). *Matrix Algebra*. Cambridge University press.

Barndorff-Nielsen, O. E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73**, 307–322.

Barndorff-Nielsen, O. E. and Cox, D. R. (1979). Edgeworth and saddle-point approximations with statistical applications (with Discussion). *Journal of the Royal Statistical Society. Series B (Methodological)* **41**, 279–312.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **160**, 268–282.

Borgelt, C. and Kruse, R. (2002). *Graphical Models: Methods for Data Analysis and Mining*. John Wiley & Sons.

Buhl, S. L. (1993). On the existence of maximum likelihood estimators for graphical Gaussian models. *Scandinavian Journal of Statistics* **20**, 263–270.

Butler, R. W. (2007). *Saddlepoint Approximations with Applications*. Cambridge University press.

Córdoba, I., Bielza, C. and Larrañaga, P. (2020). A review of Gaussian Markov models for conditional independence. *Journal of Statistical Planning and Inference* **206**, 127–144.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman & Hall.

Davison, A. C., Fraser, D. A. S., Reid, N. and Sartori, N. (2014). Accurate directional inference for vector parameters in linear exponential families. *Journal of the American Statistical Association* **109**, 302–314.

Dawid, A. P. and Lauritzen, S. L. (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *The Annals of Statistics* **21**, 1272–1317.

Eriksen, P. S. (1996). Tests in covariance selection models. *Scandinavian Journal of Statistics* **23**, 275–284.

Fraser, D. A. S. and Massam, H. (1985). Conical tests: Observed levels of significance and confidence regions. *Statistische Hefte* **26**, 1–17.

Fraser, D. A. S. and Reid, N. (2006). Assessing a vector parameter. *Student* **5**, 247–256.

Fraser, D. A. S., Reid, N. and Sartori, N. (2016). Accurate directional inference for vector parameters. *Biometrika* **103**, 625–639.

Fraser, D. A. S., Reid, N. and Wu, J. (1999). A simple general formula for tail probabilities for frequentist and Bayesian inference. *Biometrika* **86**, 249–264.

Huang, C., Di Caterina, C. and Sartori, N. (2022). Directional testing for high dimensional multivariate normal distributions. *Electronic Journal of Statistics* **16**, 6489–6511.

Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika* **12**, 134–139.

Jensen, J. L. (2021). On the use of saddlepoint approximations in high dimensional inference. *Sankhya A* **83**, 379–392.

Kenward, M. G. (1987). A method for comparing profiles of repeated measurements. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **36**, 296–308.

Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. 3rd Edition. Springer.

Liu, Q. and Pierce, D. A. (1994). A note on Gauss–Hermite quadrature. *Biometrika* **81**, 624–629.

Massa, M. S., Chiogna, M. and Romualdi, C. (2010). Gene set analysis exploiting the topology of a pathway. *BMC Systems Biology* **4**, 121.

Massa, S. and Sales, G. (2016). *topologyGSA: Gene set analysis exploiting pathway topology*. R package version 1.4.6.

McCormack, A., Reid, N., Sartori, N. and Theivendran, S. A. (2019). A directional look at F-tests. *The Canadian Journal of Statistics* **47**, 619–627.

Pace, L. and Salvan, A. (1997). *Principles of Statistical Inference: From a Neo-Fisherian Perspective*. World Scientific.

Pierce, D. A. and Peters, D. (1992). Practical use of higher order asymptotics for multiparameter exponential families. *Journal of the Royal Statistical Society. Series B (Methodological)* **54**, 701–725.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical ComputingVienna, Austria.

Roverato, A. (2017). *Graphical Models for Categorical Data*. Cambridge University press.

Roverato, A. and Whittaker, J. (1996). Standard errors for the parameters of graphical Gaussian models. *Statistics and Computing* **6**, 297–302.

Roverato, A. and Whittaker, J. (1998). The Isserlis matrix and its application to non-decomposable graphical Gaussian models. *Biometrika* **85**, 711–725.

Salgueiro, M. F., Smith, P. W. and McDonald, J. W. (2005). Power of edge exclusion tests in graphical Gaussian models. *Biometrika* **92**, 173–182.

Skovgaard, I. M. (1988). Saddlepoint expansions for directional test probabilities. *Journal of the Royal Statistical Society. Series B (Methodological)* **50**, 269–280.

Skovgaard, I. M. (2001). Likelihood asymptotics. *Scandinavian Journal of Statistics* **28**, 3–32.

Tang, Y. and Reid, N. (2020). Modified likelihood root in high dimensions. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **82**, 1349–1369.

Whittaker, J. (2009). *Graphical Models in Applied Statistics*. John Wiley & Sons.

(Received June 2022; accepted and published online May 2023)