Social Statistics Research Centre
University of Southampton

Power of Edge Exclusion Tests in Graphical Gaussian Models

Fatima Salgueiro, Peter Smith and John McDonald

Abstract

Asymptotic multivariate normal approximations to the joint distributions of edge exclusion test statistics for saturated graphical Gaussian models are derived. Non-signed and signed square-root versions of the likelihood ratio, Wald and score test statistics are considered. Non-central chi-squared approximations are also considered for the non-signed versions. These approximations are used to estimate the power of edge exclusion tests and an example is presented.

SSRC Methodology Working Paper M03/02

Power of edge exclusion tests in graphical Gaussian models

By M. FÁTIMA SALGUEIRO

Departamento Métodos Quantitativos, ISCTE, Av. Forças Armadas

1649-026 Lisboa, Portugal

fatima.salgueiro@iscte.pt

PETER W. F. SMITH AND JOHN W. McDONALD

Department of Social Statistics, University of Southampton,

Southampton SO17 1BJ, U.K.

pws@socsci.soton.ac.uk; bigmac@socsci.soton.ac.uk

Summary

Asymptotic multivariate normal approximations to the joint distributions of edge exclusion test statistics for saturated graphical Gaussian models are derived. Non-signed and signed square-root versions of the likelihood ratio, Wald and score test statistics are considered. Non-central chi-squared approximations are also considered for the non-signed versions. These approximations are used to estimate the power of edge exclusion tests and an example is presented.

Some key words: Distributions of test statistics; Likelihood ratio test; Partial correlation; Score test; Signed square-root tests; Wald test.

1. Introduction

Graphical Gaussian models are parametric statistical models for multivariate normal random variables. In these models the independence structure of the variables is displayed using a mathematical graph, the (conditional) independence graph. When searching for a well fitting graphical model it is usual to first test for exclusion of each edge, in turn, from the saturated model, i.e., to perform the first step of a backward elimination model selection procedure. Traditionally the likelihood ratio test has been used, although the Wald or the (efficient) score tests can be used. Closed form expressions for the test statistics for single edge exclusion from the saturated graphical Gaussian model were derived by Smith & Whittaker (1998) and are presented in Table 1, along with their signed square-root versions. The notation used in this paper is explained below. For a introduction to graphical Gaussian models see, for example, Lauritzen (1996) or Whittaker (1990).

[Table 1 about here]

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph, with a finite vertex set $\mathcal{V} = \{1, \ldots, p\}$ and edge set $\mathcal{E}$. Let $\mathcal{W} = \{(r, s) : r, s \in \mathcal{V},\ r \leq s\}$. Roverato & Whittaker (1998) introduced this notation and considered $\mathcal{W}$ as the edge set of the complete graph, although it contains (i, i) for all $i \in \mathcal{V}$. In this paper $\mathcal{T} = \mathcal{W} \setminus \{(i, i) : i \in \mathcal{V}\} = \{(i, j) : i, j \in \mathcal{V},\ i < j\}$ is the edge set of the complete graph. Note that, since $\mathcal{V} = \{1, \ldots, p\}$ has $|\mathcal{V}| = p$ elements, $|\mathcal{W}| = p(p + 1)/2$ and $|\mathcal{T}| = p(p - 1)/2$. The inverse variance matrix of the underlying multivariate normal distribution is denoted by Ω, with rows and columns indexed by the elements of $\mathcal{V}$, i.e., Ω is indexed by $\mathcal{V} \times \mathcal{V}$. Let ω denote the vector of the distinct elements of Ω, with elements $\{\omega\}_{(r,s)} = \omega_{rs}$, indexed by $\mathcal{W}$. The sample inverse variance matrix with divisor n, the sample size, is $\widehat{\Omega}$, the unconstrained maximum likelihood estimator of Ω. The vector of the distinct elements of $\widehat{\Omega}$ is $\widehat{\omega}$, indexed by $\mathcal{W}$, with elements $\{\widehat{\omega}\}_{(r,s)} = \widehat{\omega}_{rs}$. Let $\rho_{rs.R} = -\omega_{rs}(\omega_{rr}\,\omega_{ss})^{-1/2}$, where $R = \mathcal{V} \setminus (r,s)$. Note that, for $i \neq j$, $\rho_{ij.R}$ is the partial correlation coefficient between variables i and j, after conditioning on the remaining variables, indexed by R, and equals minus the off-diagonal elements of the scaled inverse variance matrix. Furthermore, $\rho_{ss.R} = -1$, by definition.

Under the null hypothesis that variables i and j are conditionally independent given the remaining variables in the model, i.e., the edge between i and j is absent from the independence graph, the non-signed versions of the three test statistics for single edge exclusion have, asymptotically, a χ_1^2 distribution and the signed square-root versions have, asymptotically, a $N(0,1)$ distribution. The aim of this paper is to study the distributions of these test statistics under the alternative hypothesis that the saturated model holds, in order to derive asymptotic approximations to power functions. In §2 the delta-method is used to obtain asymptotic normal approximations to the distributions of the test statistics, under this alternative hypothesis, for the general p variables case. Also considered is a non-central chi-squared approximation to the distributions of the non-signed test statistics. In §3 approximations to the power of the single edge exclusion tests are proposed. In §4 the Frets's heads data is used to illustrate power calculations. Section 5 presents some conclusions and possible generalisations of the results.

2. Approximations to the distributions of the test statistics

2.1. *Asymptotic normal approximation*

If $\widehat{\omega}$ is the unconstrained maximum likelihood estimator of ω, then as $n \longrightarrow \infty$,

$$\sqrt{n}(\widehat{\omega} - \omega) \longrightarrow N\{0, \mathrm{Iss}(\Omega)\}$$

in distribution, where $\mathrm{Iss}(\Omega)$, the Isserlis matrix of Ω, is the symmetric matrix indexed by $\mathcal{W} \times \mathcal{W}$ with elements

$$\{\mathrm{Iss}(\Omega)\}_{(r,s),(t,u)} = \omega_{rt}\omega_{su} + \omega_{ru}\omega_{st} \tag{1}$$

(Cox & Wermuth, 1990; Smith & Whittaker, 1998). For further applications of the Isserlis matrix to graphical Gaussian models see Roverato & Whittaker (1998). Because the edge exclusion test statistics are functions of $\widehat{\omega}$, the delta-method can used to obtain normal approximations to the distributions of the test statistics, under the alternative hypothesis that the saturated model holds.

Let the vector $f(\widehat{\Omega})$, indexed by $\mathcal{T}$, have elements $\{f\}_{(i,j)}(\widehat{\Omega}) = f_{ij}(\widehat{\Omega}) = T_{ij}/n$, where T_{ij} is one of the non-signed test statistics for single edge exclusion presented in Table 1. Note that f does not depend on n. If $f(\Omega)$ is differentiable at $\widehat{\Omega}$, then, using the delta-method, the asymptotic distribution of $f(\widehat{\Omega})$ is the normal distribution with mean vector $f(\Omega)$ and variance matrix $n^{-1}V$, i.e., $\sqrt{n}\left[f(\widehat{\Omega}) - f(\Omega)\right] \longrightarrow N(0, V)$ in distribution, where V is indexed by $\mathcal{T} \times \mathcal{T}$, with elements

$$\{V\}_{(i,j),(k,l)} = \sum_{(r,s)} \sum_{(t,u)} \left[\frac{\partial \{f\}_{(i,j)}(\Omega)}{\partial \{\omega\}_{(r,s)}} \{\mathrm{Iss}(\Omega)\}_{(r,s),(t,u)} \frac{\partial \{f\}_{(k,l)}(\Omega)}{\partial \{\omega\}_{(t,u)}} \right]$$

$$= \sum_{(r,s)} \sum_{(t,u)} \left\{ \frac{\partial f_{ij}(\Omega)}{\partial \omega_{rs}} (\omega_{rt}\omega_{su} + \omega_{ru}\omega_{st}) \frac{\partial f_{kl}(\Omega)}{\partial \omega_{tu}} \right\}. \tag{2}$$

Note that (r, s) and $(t, u) \in \mathcal{W}$ and (i, j) and $(k, l) \in \mathcal{T}$. As $T_{ij} = nf_{ij}$, the diagonal elements of nV are the variances of T_{ij} and the off-diagonal elements are the covariances between T_{ij} and T_{kl}.

Let the vector $f^s(\widehat{\Omega})$, indexed by $\mathcal{T}$, have elements $\{f^s\}_{(i,j)}(\widehat{\Omega}) = f^s_{ij}(\widehat{\Omega}) = T^s_{ij}/\sqrt{n}$, so that f^s does not depend on n, where T^s_{ij} is one of the signed square-root versions of the test statistics for single edge exclusion presented in Table 1. Then, the variance matrix of $f^s(\widehat{\Omega})$ is $n^{-1}V$, where V is given by Equation 2, once f is replaced by f^s. Hence, the variances and covariances of the signed square-root versions of the test statistics are, respectively, the diagonal and the off-diagonal elements of V, and do not depend on the sample size n.

For the non-signed likelihood ratio test, T^L_{ij}, the mean of the asymptotic distribution is $AE(T^L_{ij}) = -n \log \left\{ (1 - \omega^2_{ij} (\omega_{ii}\omega_{jj})^{-1} \right\} = -n \log(1 - \rho^2_{ij.R})$. From Equation 2, the variance and covariances simplify, respectively, to $\mathrm{var}(T^L_{ij}) = 4n\omega^2_{ij} (\omega_{ii}\omega_{jj})^{-1} = 4n\rho^2_{ij.R}$ and

$$\mathrm{cov}(T^L_{ij}, T^L_{kl}) = \frac{n\, C_\omega}{\omega_{ii}\omega_{jj}\omega_{kk}\omega_{ll}(\omega_{ii}\omega_{jj} - \omega^2_{ij})(\omega_{kk}\omega_{ll} - \omega^2_{kl})}, \tag{3}$$

where

$$\begin{aligned} C_\omega = {} & 4\omega_{ii}\omega_{jj}\omega_{kk}\omega_{ll}\omega_{ij}\omega_{ik}\omega_{jl}\omega_{kl} + 4\omega_{ii}\omega_{jj}\omega_{kk}\omega_{ll}\omega_{ij}\omega_{il}\omega_{jk}\omega_{kl} - 4\omega_{jj}\omega_{kk}\omega_{ll}\omega^2_{ij}\omega_{ik}\omega_{il}\omega_{kl} \\ & -4\,\omega_{ii}\omega_{kk}\omega_{ll}\omega^2_{ij}\omega_{jk}\omega_{jl}\omega_{kl} - 4\,\omega_{ii}\omega_{jj}\omega_{ll}\omega_{ij}\omega_{ik}\omega_{jk}\omega^2_{kl} - 4\,\omega_{ii}\omega_{jj}\omega_{kk}\omega_{ij}\omega_{il}\omega_{jl}\omega^2_{kl} \\ & +2\,\omega_{ii}\omega_{kk}\omega^2_{ij}\omega^2_{jl}\omega^2_{kl} + 2\,\omega_{ii}\omega_{ll}\omega^2_{ij}\omega^2_{jk}\omega^2_{kl} + 2\,\omega_{jj}\omega_{kk}\omega^2_{ij}\omega^2_{il}\omega^2_{kl} + 2\,\omega_{jj}\omega_{ll}\omega^2_{ij}\omega^2_{ik}\omega^2_{kl}. \end{aligned}$$

Equation 3 holds for all pairs of edges (i, j), $(k, l) \in \mathcal{T}$ and can be simplified when the edges have one vertex in common. It is easy to prove that, starting with Equation 3, collecting terms of the type $-\omega_{pq}(\omega_{pp}\omega_{qq})^{-1/2}$ and replacing them by $\rho_{pq.R}$ gives

$$\mathrm{cov}(T^L_{ij},\, T^L_{kl}) \;=\; \frac{n\,C_\rho}{(1-\rho^2_{ij.R})(1-\rho^2_{kl.R})}, \tag{4}$$

where

$$\begin{aligned} C_\rho \;=\; & 4\,\rho_{ij.R}\rho_{ik.R}\rho_{jl.R}\rho_{kl.R} + 4\,\rho_{ij.R}\rho_{il.R}\rho_{jk.R}\rho_{kl.R} + 4\,\rho^2_{ij.R}\rho_{ik.R}\rho_{il.R}\rho_{kl.R} \\ & +4\,\rho^2_{ij.R}\rho_{jk.R}\rho_{jl.R}\rho_{kl.R} + 4\,\rho_{ij.R}\rho_{ik.R}\rho_{jk.R}\rho^2_{kl.R} + 4\,\rho_{ij.R}\rho_{il.R}\rho_{jl.R}\rho^2_{kl.R} \\ & +2\,\rho^2_{ij.R}\rho^2_{jl.R}\rho^2_{kl.R} + 2\,\rho^2_{ij.R}\rho^2_{jk.R}\rho^2_{kl.R} + 2\,\rho^2_{ij.R}\rho^2_{il.R}\rho^2_{kl.R} + 2\,\rho^2_{ij.R}\rho^2_{kl.R}\rho^2_{ik.R}. \end{aligned}$$

Equation 4 is a general formula for the covariance between T^L_{ij} and T^L_{kl}, in the asymptotic distribution, written as a function of the partial correlation coefficients. For edges with a common vertex, after simplification, values of $\rho_{ss.R}$ are required; recall they are -1, by definition.

Formulae for the means, variances and covariances of all six test statistics, presented in Table 1, are summarised in Table 2. Note that for the signed square-root versions, as $\rho_{ij} \longrightarrow 0$, $AE(T^s_{ij}) \longrightarrow 0$ and $\mathrm{var}(T^s_{ij}) \longrightarrow 1$, the mean and variance of the asymptotic null distribution. For the non-signed versions both means and variances tend to zero, which are not those for a χ^2_1. Hence, for the non-signed versions, the normal approximations will be poor for very small distances from the null.

[Table 2 about here]

2.2. Non-central chi-squared approximation

Local alternatives have been studied in the literature. For a composite hypothesis of the type $H_0 : \psi = \psi_0$ and nuisance parameter λ unspecified, Cox & Hinkley (1974) showed that, under local alternatives $H_a : \psi = \psi_0 + \delta_\psi/\sqrt{n}$, the likelihood ratio test statistic is approximately chi-squared, with degrees of freedom equal to the dimension of ψ and non-centrality parameter $n\delta_\psi^T \, i_{.}(\psi_0 : \lambda) \, \delta_\psi$, where $i_{.}(\psi : \lambda)$ is the inverse of the variance matrix of the asymptotic normal distribution of $\sqrt{n}\widehat{\psi}$. Similar results hold for the Wald and score tests.

For excluding the single edge (i, j) from a saturated graphical Gaussian model the hypotheses are $H_0 : \omega_{ij} = 0$ and $H_a : \omega_{ij} = 0 + \delta_\psi/\sqrt{n}$. From Equation 1, the variance of the asymptotic normal distribution of $\sqrt{n}\widehat{\omega}_{ij}$ is $\{\mathrm{Iss}(\Omega)\}_{(i,j),(i,j)} = \omega_{ii}\omega_{jj} + \omega_{ij}^2$. Hence, the distribution of each T_{ij}, at a local alternative, can be approximated by a non-central χ_1^2, with non-centrality parameter $\gamma_{ij} = n\omega_{ij}^2 (\omega_{ii}\omega_{jj} + \omega_{ij}^2)^{-1} = n\rho_{ij.R}^2 (1 + \rho_{ij.R}^2)^{-1}$, where $\rho_{ij.R}$ is the partial correlation coefficient under the alternative hypothesis.

A simulation study was performed by M. F. Salgueiro in her 2002 University of Southampton Ph.D. thesis to assess the accuracy of the proposed asymptotic normal and non-central chi-squared approximations, as the sample size and the (partial) correlation coefficient vary. The main results are that the normal approximation performs better if the sample size is large and the correlation coefficient is not close to zero. The non-central chi-squared approximation performs better than the normal approximation at small distances from the null, i.e., for small values of $\rho_{ij.R}$, in particular if n is not large.

3. Power of single edge exclusion tests

The power of a hypothesis test is the probability of rejecting the null hypothesis given a particular value of the interest parameter. Define the power of a model selection procedure as the probability of selecting the true model given the specified true model parameters. The traditional definition of power relates to a test of a single null hypothesis. A model selection procedure involves tests of a set of null hypotheses. The fact that more than one hypothesis is tested at a time is an argument against calling the probability of selecting the true model power. However, it has the essence of power in the sense that it is the probability of accepting the 'right hypotheses'.

3.1. Power of non-signed tests

The power of a size α test for excluding edge (i, j) from the saturated model can be estimated, using the asymptotic normal approximation derived in §2.1, as

$$\mathrm{pr}[T_{ij} > \chi^2_{1;1-\alpha} \,|\, \Omega] \doteq \mathrm{pr}\left[Z > \frac{\chi^2_{1;1-\alpha} - AE(T_{ij})}{\sqrt{\mathrm{var}(T_{ij})}}\right],$$

where $Z \sim N(0, 1)$, $\chi^2_{1;1-\alpha}$ is the upper α quantile of a chi-squared distribution on one degree of freedom and the formulae for the the mean and variance are given in Table 2. This power can also be estimated, using the non-central chi-squared approximation derived in §2.2, as $\mathrm{pr}[X > \chi^2_{1;1-\alpha} \,|\, \rho_{ij.R}]$, where $X \sim \chi^2_1(\gamma_{ij})$. Note that, regardless of the number of variables in the model, the power of the test of excluding edge (i, j) depends on $\rho_{ij.R}$ but not on the remaining partial correlation coefficients.

Figure 1 compares the theoretical power of the likelihood ratio test for excluding a single edge from a saturated graphical Gaussian model estimated using the normal approximation (dashed line) and the non-central chi-squared approximation (solid line). The dotted line represents simulated power values (10 000 repetitions). Three sample sizes were used: 50, 200 and 1 000. The horizontal dotted lines correspond to power values of 0 and 0.05. Power is symmetric about zero correlation and increases as $|\rho_{ij.R}|$ increases and as the sample size increases. The asymptotic normal approximation performs well for large sample sizes and values of $\rho_{ij.R}$ not close to zero. For small sample sizes and values of $\rho_{ij.R}$ close to zero, i.e., at small distances from the null, the power of the likelihood ratio test for selecting the saturated model can be accurately approximated by a non-central chi-squared distribution.

[Figure 1 about here]

The probability of excluding neither of the two edges (i, j) and (k, l) from the saturated model, when two separate edge exclusion tests are performed, can be approximated by

$$\text{pr}[\min(T_{ij}, T_{kl}) > \chi^2_{1;1-\alpha} \,|\, \Omega] \doteq \int_{\chi^2_{1;1-\alpha}}^{+\infty} \int_{\chi^2_{1;1-\alpha}}^{+\infty} \phi_2(\mu, \Sigma) dT_{ij} dT_{kl}, \qquad (5)$$

where $\phi_2(\mu, \Sigma)$ is a bivariate normal density with mean vector μ and variance matrix Σ given by the formulae for non-signed tests presented in Table 2. Note that both the mean and the variance of the asymptotic distribution of T_{ij} are a function of n and of $\rho_{ij.R}$, whereas the covariance between T_{ij} and T_{kl} depends not only on $\rho_{ij.R}$ and $\rho_{kl.R}$ but also on $\rho_{ik.R}$, $\rho_{il.R}$, $\rho_{jk.R}$ and $\rho_{jl.R}$. For this reason some non-symmetry and non-monotonicity of the power functions can be observed for certain combinations of values of the partial correlation coefficients.

If there are $|\mathcal{V}| = p$ variables in the saturated model, there are $|\mathcal{T}| = p(p-1)/2$ edges in the graph, and consequently $|\mathcal{T}|$ test statistics for single edge exclusion from the saturated model. The power of selecting the saturated model is then the probability that each of these test statistics is greater than $\chi^2_{1;1-\alpha}$, given the values of all the partial correlation coefficients. A generalisation of Equation 5, with a $|\mathcal{T}|$-dimensional integral, can be used to approximate this power.

3.2. Power of signed square-root tests

For a two-sided test, the null hypothesis that $\rho_{ij.R}$ equals zero is rejected if the absolute value of the signed square-root test statistic is greater than $\Phi(1-\frac{\alpha}{2})$. Hence, the power for the two-sided signed square-root test of excluding edge (i,j) from the saturated model can be estimated as

$$\text{pr}\left[|T^s_{ij}| > \Phi(1-\frac{\alpha}{2}) \,|\, \Omega\right] \simeq \text{pr}\left[Z < \frac{\Phi(\frac{\alpha}{2}) - AE(T^s_{ij})}{\sqrt{\text{var}(T^s_{ij})}}\right] + \text{pr}\left[Z > \frac{\Phi(1-\frac{\alpha}{2}) - AE(T^s_{ij})}{\sqrt{\text{var}(T^s_{ij})}}\right].$$

For a one-sided hypothesis test, the null hypothesis is rejected if the value of the signed square-root test statistic is greater than $\Phi(1-\alpha)$. Hence, the power for the one-sided signed square-root test of excluding edge (i,j) from the saturated model can be estimated as

$$\text{pr}\left[T^s_{ij} > \Phi(1-\alpha) \,|\, \Omega\right] \simeq \text{pr}\left[Z > \frac{\Phi(1-\alpha) - AE(T^s_{ij})}{\sqrt{\text{var}(T^s_{ij})}}\right].$$

Simulation results showed that the normal approximation to the power of the signed square-root test of excluding edge (i,j) from the saturated model is a very good approximation, both when one-sided and two-sided hypothesis tests are used, even for small sample sizes and partial correlations close to zero.

The probability of excluding neither of the two edges (i,j) and (k,l) from the saturated model, when two separate signed square-root edge exclusion tests are performed, can be approximated by

$$\text{pr}\left[\min(T^s_{ij},\, T^s_{kl}) > \Phi(1-\alpha)\,|\,\Omega\right] \doteq \int_{\Phi(1-\alpha)}^{+\infty}\int_{\Phi(1-\alpha)}^{+\infty} \phi_2\left(\mu^s,\, \Sigma^s\right)\, dT^s_{ij}\, dT^s_{kl},$$

for one-sided tests and by

$$\text{pr}\left[|\min(T^s_{ij},\, T^s_{kl})| > \Phi(1-\frac{\alpha}{2})\ \mid\ \Omega\right] \doteq \int_D \phi_2\left(\mu^s,\, \Sigma^s\right)\, dT^s_{ij}\, dT^s_{kl},$$

for two-sided tests. The mean vector μ^s and the variance matrix Σ^s are those for the signed square-root tests presented in Table 2. The domain of integration, D, in the case of two-sided tests, is the region where each of the two test statistics is in $-\infty$ to $\Phi(\frac{\alpha}{2})$ or $\Phi(1-\frac{\alpha}{2})$ to $+\infty$, i.e., $D = \{(-\infty, \Phi(\frac{\alpha}{2})) \cup (\Phi(1-\frac{\alpha}{2}), +\infty)\}^2$. Again, these formulae are easily generalised to estimate the power of selecting the saturated model when there are p variables.

Refer to M.F. Salgueiro 2002 Ph.D. thesis for a comprehensive investigation of the power of a backward elimination model selection procedure for selecting the true (saturated) graphical Gaussian model by simulation, for different number of variables, sample sizes and values of the (partial) correlation coefficient(s).

4. Example

Frets's heads data, presented by Whittaker (1990, p.265), are used to illustrate power calculations. The data consists of measurements of the head length and breadth of the first and second adult sons, in a sample of 25 families. Sample partial correlations are presented in the lower triangle of Table 3. Whittaker (1990) showed how different model selection procedures lead to different chosen models. Based on the six single edge exclusion tests from the saturated model, only edges $(1, 2)$ and $(3, 4)$ are significantly different from zero, at 5%. If all the non-significant edges are removed, the resulting model suggests that the measurements on the two sons are independent, which is not in agreement with partial correlations. The power of these tests is now assessed.

[Table 3 about here]

Under the assumption that the sons are exchangeable, the population within-head partial correlations are equal, i.e., $\rho_{12.R} = \rho_{34.R}$, and so are the between-head, length-to-breadth partial correlations, i.e., $\rho_{23.R} = \rho_{14.R}$. Furthermore, based on the observed data, it seems reasonable to assume that the length-to-length and breadth-to-breadth partial correlations are equal, i.e., $\rho_{13.R} = \rho_{24.R}$, and that these three pairs of partial correlations are in the ratio 1: 0.3: 0.5. Therefore, it is assumed that the population partial correlation matrix is of the form presented in the upper triangle of Table 3, where $0 < \kappa < 5/9$. The upper bound is imposed by the constraint that the variance matrix is positive definite and the lower bound is to rule out negative partial correlations, which do not seem sensible for these variables.

Figure 2 presents the estimated power functions for various sample sizes (from 10 to 1 000) and values of κ (0.3, 0.4 and 0.55) for the one-sided (solid lines) and the two-sided (dashed lines) signed square-root versions of the likelihood ratio test.

[Figure 2 about here]

It is immediately obvious that a sample of 25 would not give enough power to detect the saturated model of the form presented in Table 3, even with the largest possible κ. For this case a sample size of 200 would give a power of 0.62 for a one-sided test and of 0.47 for a two-sided test. One-sided tests have more power and should be used here because all partial correlations are assumed to be positive. As expected, power reduces dramatically as κ decreases. While one may argue that the saturated model is not appropriate for $\kappa = 0.3$, a value of 0.4 has all partial correlations above 0.12 but over 500 observations would be required for a power of 0.7.

5. Discussion

For the non-signed versions of the test statistics considered in this paper, the proposed asymptotic non-central chi-squared approximations perform better than the asymptotic normal approximations for small distances from the null. However, they are worse for large distances from the null and can only be used for single test statistics. The asymptotic normal approximations are more accurate for the signed square-root versions than for the non-signed versions and, therefore, should be preferred.

These approximations are used to estimate the power of edge exclusion tests from a saturated graphical Gaussian model. An example illustrates how the proposed approximations can be used to calculate the sample size required for a certain level of power. Here the required integrations were performed numerically, although for higher dimensional problems Monte Carlo integration may be preferable.

The ideas presented in this paper, in principle, can be generalised to the cases of excluding several edges simultaneously, of excluding one or more edges from a non saturated model and of models involving discrete and mixed variables, provided closed form expressions for (i) the test statistics and (ii) the variance matrix of the asymptotic distribution of the arguments of the test statistics can be derived. However, the calculations are complex and the resulting expressions for the parameters of the asymptotic distribution are usually complicated. Hence, the use of a computer algebra package is recommended.

Consider a single test for excluding several edges from the saturated graphical Gaussian model. In this case the Wald test statistic is a function of $\widehat{\omega}$, the maximum likelihood estimator under the saturated model, whose variance matrix is given by Equation 1. Therefore, for the Wald test statistic, requirements (i) and (ii) are satisfied. If the null model is not decomposable, then it is not possible to express the likelihood ratio or the score test statistics in closed form because the maximum likelihood estimate can only be obtained by iteration. For all decomposable graphical Gaussian models the likelihood ratio test statistic can be expressed as a function of $\widehat{\omega}$ and the score statistic as a function of the maximum likelihood estimator under the null model. Furthermore, an explicit formula for the variance of the asymptotic distribution of the maximum likelihood estimator can be derived (Lauritzen, 1996, p. 141; Roverato & Whittaker, 1998, p. 720). Hence, when the null model is a decomposable graphical Gaussian model requirements (i) and (ii) are also satisfied for the likelihood ratio and score test statistics.

The above arguments can be extended to the situation where there is a set of test statistics, $\{T_i : i = 1, \ldots, m\}$ say, and each T_i is a test statistic for excluding one or more edges from the same graphical model, M_a say. For example, this situation arises at each step of a backward elimination model selection procedure. Here closed form expressions can be obtained for the Wald, the score and the likelihood ratio test statistics when, respectively, M_a, the null model and both models are decomposable. If all the T_i are Wald test statistics and M_a is decomposable, then it is relatively easy to obtain a closed form expression for the variance matrix, since the T_i are functions of the maximum likelihood estimator for a single model, M_a. Note that for the discrete case an explicit expression for the variance matrix of the maximum likelihood estimator for decomposable models is also given by Lauritzen (1996, p. 96). If $m = 1$ and T_1 is a score test statistic, then again only estimates from a single null model are required. However, when there is more than one score test statistic or when there are likelihood ratio test statistics, estimates under more than one model are required and obtaining the variance matrix is more difficult. One possibility is to express the test statistics as a function of the maximum likelihood estimator of M_a. This was what Smith and Whittaker (1998) did to obtain the score test statistic presented in Table 1.

Acknowledgements

The Portuguese grant PRAXIS XXI - BD 19873/99 supported the work of the first author.

References

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

Cox, D.R. & Wermuth, N. (1990). An approximation to maximum likelihood estimates in reduced models. *Biometrika* **77**, 747–61.

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

Roverato, A. & Whittaker, J. (1998). The Isserlis matrix and its applications to non-decomposable graphical Gaussian models. *Biometrika* **85**, 711–25.

Smith, P.W.F. & Whittaker, J. (1998). Edge exclusion tests for graphical Gaussian models. In *Learning in Graphical Models*, Ed. M.I. Jordan. Dordrecht: Kluwer Academic.

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics.* Chichester: Wiley.

Table 1: *Likelihood ratio (L), Wald (W) and score (S) test statistics for the exclusion of edge (i,j) from the saturated graphical Gaussian model*

		non-signed version		signed square-root version
L	T_{ij}^{L}	$= -n\log\left(1-\frac{\widehat{\omega}_{ij}^2}{\widehat{\omega}_{ii}\widehat{\omega}_{jj}}\right)$	T_{ij}^{sL}	$= \sqrt{n}\,\mathrm{sgn}(-\widehat{\omega}_{ij})\left\{-\log\left(1-\frac{\widehat{\omega}_{ij}^2}{\widehat{\omega}_{ii}\widehat{\omega}_{jj}}\right)\right\}^{1/2}$
		$= -n\log(1-\hat{\rho}_{ij.R}^2)$		$= \sqrt{n}\,\mathrm{sgn}(\widehat{\rho}_{ij.R})\left\{-\log(1-\hat{\rho}_{ij.R}^2)\right\}^{1/2}$
W	T_{ij}^{W}	$= n\frac{\widehat{\omega}_{ij}^2}{\widehat{\omega}_{ii}\widehat{\omega}_{jj}+\widehat{\omega}_{ij}^2}$	T_{ij}^{sW}	$= \sqrt{n}\,(-\widehat{\omega}_{ij})\,(\widehat{\omega}_{ii}\widehat{\omega}_{jj}+\widehat{\omega}_{ij}^2)^{-1/2}$
		$= n\frac{\hat{\rho}_{ij.R}^2}{1+\hat{\rho}_{ij.R}^2}$		$= \sqrt{n}\,\widehat{\rho}_{ij.R}\,(1+\hat{\rho}_{ij.R}^2)^{-1/2}$
S	T_{ij}^{S}	$= n\frac{\widehat{\omega}_{ij}^2}{\widehat{\omega}_{ii}\widehat{\omega}_{jj}}$	T_{ij}^{sS}	$= \sqrt{n}\,(-\widehat{\omega}_{ij})\,(\widehat{\omega}_{ii}\widehat{\omega}_{jj})^{-1/2}$
		$= n\hat{\rho}_{ij.R}^2$		$= \sqrt{n}\,\widehat{\rho}_{ij.R}$

Table 2: *Means, variances and covariances of the asymptotic normal distributions for the test statistics (for single edge exclusion, from the saturated graphical Gaussian model)*

	$AE(T_{ij})$	$AE(T^s_{ij})$
L	$-n\log(1-\rho^2_{ij.R})$	$\sqrt{n}\,\mathrm{sgn}(\rho_{ij.R})\left\{-\log(1-\rho^2_{ij.R})\right\}^{1/2}$
W	$n\,\rho^2_{ij.R}\left(1+\rho^2_{ij.R}\right)^{-1}$	$\sqrt{n}\,\rho_{ij.R}\left(1+\rho^2_{ij.R}\right)^{-1/2}$
S	$n\,\rho^2_{ij.R}$	$\sqrt{n}\,\rho_{ij.R}$

	$\mathrm{var}(T_{ij})$	$\mathrm{var}(T^s_{ij})$
L	$4n\rho^2_{ij.R}$	$-\rho^2_{ij.R}\left\{\log(1-\rho^2_{ij.R})\right\}^{-1}$
W	$4n\rho^2_{ij.R}(1-\rho^2_{ij.R})^2(1+\rho^2_{ij.R})^{-4}$	$(1-\rho^2_{ij.R})^2(1+\rho^2_{ij.R})^{-3}$
S	$4n\rho^2_{ij.R}(1-\rho^2_{ij.R})^2$	$(1-\rho^2_{ij.R})^2$

	$\mathrm{cov}(T_{ij},T_{kl})$	$\mathrm{cov}(T^s_{ij},T^s_{kl})$
L	$n\left\{(1-\rho^2_{ij.R})(1-\rho^2_{kl.R})\right\}^{-1}C_\rho$	$\frac{1}{4}\mathrm{sgn}(\rho_{ij.R}\,\rho_{kl.R})\left\{(1-\rho^2_{ij.R})(1-\rho^2_{kl.R})\right\}^{-1} \times\left\{\log(1-\rho^2_{ij.R})\log(1-\rho^2_{kl.R})\right\}^{-1/2}C_\rho$
W	$n\left\{(1+\rho^2_{ij.R})(1+\rho^2_{kl.R})\right\}^{-2}C_\rho$	$\frac{1}{4}(\rho_{ij.R}\rho_{kl.R})^{-1}\left\{(1+\rho^2_{ij.R})(1+\rho^2_{kl.R})\right\}^{-3/2}C_\rho$
S	$n\,C_\rho$	$\frac{1}{4}(\rho_{ij.R}\rho_{kl.R})^{-1}C_\rho$

Table 3: *Observed (lower triangle) and hypothesised (upper triangle) partial correlations for Frets's heads data*

	1	2	3	4
1 - head length, first son		κ	0.5 κ	0.3 κ
2 - head breadth, first son	0.425		0.3 κ	0.5 κ
3 - head length, second son	0.223	0.132		κ
4 - head breadth, second son	0.152	0.225	0.626	

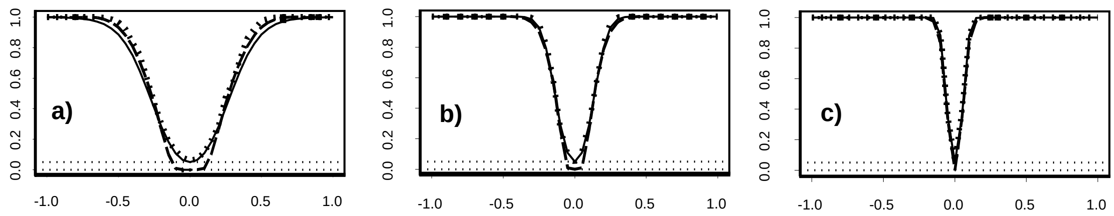

Figure 1: Simulated (dotted) and theoretical power curves for T_{ij}^{L}, with an asymptotic normal approximation (dashed) and a non-central χ_1^2 approximation (solid); n equals: a) 50, b) 200, c) 1 000.

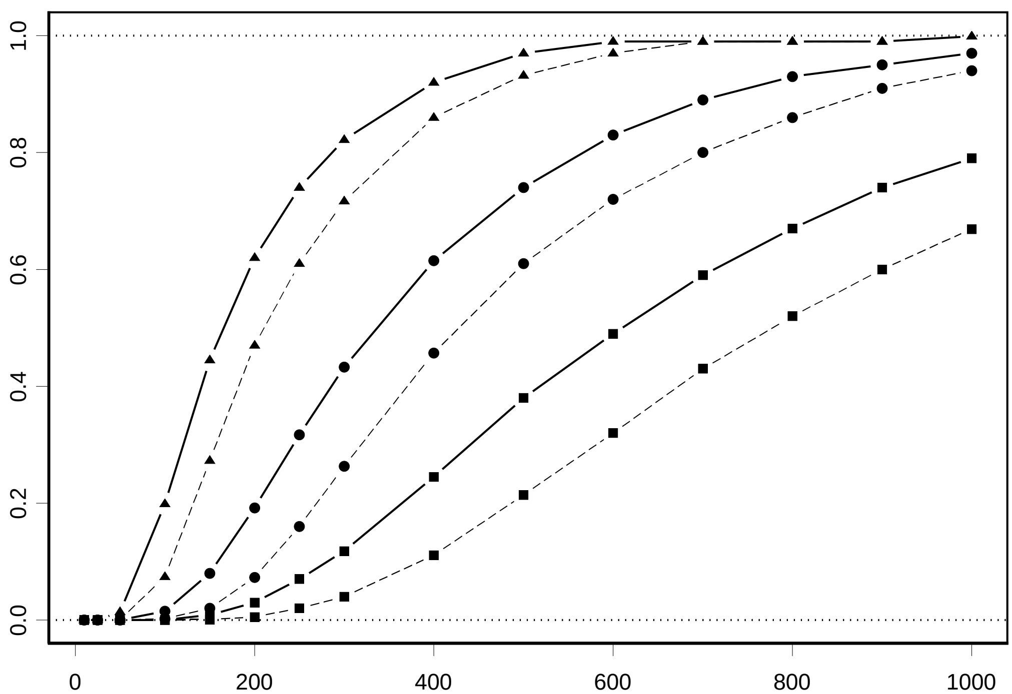

Figure 2: Power functions for hypothesised structure for Frets's heads data, for different sample sizes, using T^{sL}, one-sided (solid line) and two-sided (dashed line). $\kappa = 0.3$ (squares), $\kappa = 0.4$ (circles) and $\kappa = 0.55$ (triangles).