POWER OF EDGE EXCLUSION TESTS FOR GRAPHICAL LOG-LINEAR MODELS

M. FATIMA SALGUEIRO, PETER W.F. SMITH, JOHN W. MCDONALD

ABSTRACT

Asymptotic multivariate normal approximations to the joint distributions of edge exclusion test statistics for saturated graphical log-linear models, with all variables binary, 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 of the test statistics. Simulation results are used to assess the quality of the proposed approximations. These approximations are used to estimate the overall power of edge exclusion tests. Power calculations are illustrated using data on university admissions.

Southampton Statistical Sciences Research Institute
Methodology Working Paper M05/09

Power of edge exclusion tests for graphical log-linear models

M. Fátima Salgueiro *,1

Departamento de Métodos Quantitativos, ISCTE Business School, Av. Forças Armadas 1649-026 Lisboa, Portugal

Peter W.F. Smith and John W. McDonald

Southampton Statistical Sciences Research Institute, University of Southampton, Southampton SO17 1BJ, United Kingdom

Abstract

Asymptotic multivariate normal approximations to the joint distributions of edge exclusion test statistics for saturated graphical log-linear models, with all variables binary, 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 of the test statistics. Simulation results are used to assess the quality of the proposed approximations. These approximations are used to estimate the overall power of edge exclusion tests. Power calculations are illustrated using data on university admissions.

Key words: Edge exclusion test, Graphical log-linear model, Model selection, Odds ratio, Overall power

1 Introduction

We investigate the overall power of the first step of a backward elimination model selection procedure for graphical log-linear models (GLL) with two or

* Corresponding author. Phone: +351-963077654. Fax:+351-217903941
Email addresses: `fatima.salgueiro@iscte.pt` (M. Fátima Salgueiro), `pws@soton.ac.uk` (Peter W.F. Smith), `bigmac@soton.ac.uk` (John W. McDonald).

1 Work undertaken under the Portuguese Grant PRAXIS XXI - BD 19873/99.

three binary variables. We consider non-signed and signed square-root versions of the likelihood ratio, Wald and score test statistics. We derive asymptotic multivariate normal and non-central chi-squared approximations to the joint distributions of the test statistics for single edge exclusion from the saturated GLL model. We illustrate how to estimate power of single edge exclusion tests using the proposed approximations.

In Section 2 we review edge exclusion tests for GLL models. In Section 3 we use the delta method to obtain asymptotic normal approximations to the distributions of the test statistics, under the alternative hypothesis that the saturated model holds. We also consider a non-central chi-squared approximation to the distributions of the non-signed test statistics. In Section 4 the proposed distributions are used to approximate the overall power of the first step of a backward elimination model selection procedure. Conclusions from a simulation study, to assess the quality of such approximations, are given. In Section 5 we illustrate power calculations using data on university admissions. In Section 6 we briefly discuss the difficulties in generalizing the results to higher dimensional contingency tables.

2 Edge Exclusion in Graphical Log-Linear Models

Graphical log-linear models are a subclass of hierarchical log-linear models (see, for example, Agresti [1, pg. 316]) specified by setting a set of two-factor interaction terms (and hence their higher-order relatives) to zero. The parameters of the GLL model are the remaining terms not set to zero. The null hypothesis that the set of two-factor interaction terms, and all higher-order interaction terms including it, are zero is equivalent to the null hypothesis of conditional independence between the two corresponding factors, given the remaining ones. Hence, GLL models can be interpreted solely in terms of conditional independence and the conditional independence structure of the variables can be displayed using an independence graph. For details see Edwards [2], Lauritzen [3] and Whittaker [4].

Consider a p dimensional contingency table, cross-classifying the p dimensional random vector $\mathbf{X}_{\mathcal{V}} = \mathbf{X} = (X_1, X_2, \ldots, X_p)^T$, with $\mathcal{V} = \{1, 2, \ldots, p\}$. Let x_i denote the observed value taken by variable X_i. In this paper all variables are assumed binary and coded 0 and 1. Let $\mathbf{x} = (x_1, x_2, \ldots, x_p)^T$ denote a particular cell in the table, $\mathbf{n}_{\mathcal{V}}(\mathbf{x}_{\mathcal{V}}) = \mathbf{n}(\mathbf{x})$ denote the observed cell counts and $\boldsymbol{\pi}_{\mathcal{V}}(\mathbf{x}_{\mathcal{V}}) = \boldsymbol{\pi}(\mathbf{x})$ denote the probabilities in each cell of the table. Let $\boldsymbol{\pi}_A(\mathbf{x}_A)$ denote the marginal probability of $X_i = x_i : i \in A$. The total sample size equals $n_\emptyset$. The p dimensional random vector $\mathbf{X}$ has a cross-classified multinomial distribution of size one if and only if its density function $f_{\mathcal{V}}$ is given by $f_{\mathcal{V}}(\mathbf{x}) = \boldsymbol{\pi}_{\mathcal{V}}(\mathbf{x})$, assuming that $\boldsymbol{\pi}_{\mathcal{V}}(\mathbf{x}) > 0$ for all $\mathbf{x}$ and that $\sum_{\mathbf{x}} \boldsymbol{\pi}_{\mathcal{V}}(\mathbf{x}) = 1$.

Note that cell probabilities have to be strictly positive to ensure the existence of the log-linear expansions and of the conditional density functions. The family of cross-classified multinomial distributions is closed under marginalization and conditioning.

The log-linear expansion of the cross-classified multinomial distribution density function can be obtained as

$$\log \boldsymbol{\pi}_{\mathcal{V}}(\mathbf{x}) = \sum_{A \subseteq \mathcal{V}} \boldsymbol{\lambda}_A(\mathbf{x}_A),$$

where the summation is over all possible subsets of $\mathcal{V}$, including the empty set $\emptyset$. Each $\boldsymbol{\lambda}_A$ is a function of $\mathbf{x}_A$ and, for reasons of identifiability, corner point constraints are used, setting to zero the λ-term associated with the first category (the reference category coded 0) of each variable in $\mathbf{X}_A$. The log-linear expansion of the saturated graphical log-linear model with two or three binary variables is given by

$$\log \boldsymbol{\pi}_{\mathcal{V}}(\mathbf{x}) = \mathbf{W}^{-1} \begin{pmatrix} \lambda_\emptyset \\ \boldsymbol{\lambda} \end{pmatrix}, \text{ where } \boldsymbol{\lambda}^T = \begin{cases} (\lambda_1, \lambda_2, \lambda_{12}) \text{ if } p = 2 \\ (\lambda_1, \lambda_2, \lambda_{12}, \lambda_3, \lambda_{13}, \lambda_{23}, \lambda_{123}) \text{ if } p = 3 \end{cases}$$

and $\mathbf{W} = \mathbf{W}_1 \otimes \mathbf{W}_2 \otimes \cdots \otimes \mathbf{W}_p$ is the Kronecker product of p $\mathbf{W}_i$ matrices of the form

$$\mathbf{W}_i = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}.$$

Odds ratios are a commonly used measure of association in a contingency table. Let ψ_{ij} denote the marginal odds ratio between X_i and X_j (with $i \neq j$) and $\psi_{ij \cdot k=x}$ denote the conditional odds ratios, given that a third binary variable $X_k = x$. The marginal odds ratio ψ_{ij} is obtained by summing the cell probabilities over both categories of the remaining variables and equals $\psi_{ij} = \{\pi_{ij}(0,0)\,\pi_{ij}(1,1)\}/\{\pi_{ij}(0,1)\,\pi_{ij}(1,0)\}$. The conditional odds ratios, defined for the two categories of $X_k = x$ ($x = 0$ and $x = 1$), are given by $\psi_{ij \cdot k=x} = \{\pi(0,0,x)\,\pi(1,1,x)\}/\{\pi(0,1,x)\,\pi(1,0,x)\}$. If variables X_i and X_j are conditionally independent given the remaining variable X_k, i.e., $X_i \perp\!\!\!\perp X_j \mid X_k$, both conditional odds ratios $\psi_{ij \cdot k=0} = \psi_{ij \cdot k=1} = 1$. For standard sampling schemes, the sample odds ratio $\hat{\psi}_{ij}$ is the maximum likelihood estimator (m.l.e.) of the population odds ratio $\psi_{ij.}$

Backward elimination is a commonly used method for selecting a GLL model. The strategy is to start with the saturated model and test all the pairwise conditional independence statements, using test statistics for single edge exclusion. The likelihood ratio test (LRT) is the most commonly used test; alternatives include the Wald and the efficient score tests. Under the null hypothesis that variables X_i and X_j are conditionally independent given the remaining variables in the model, i.e., the edge between X_i and X_j is absent from the independence graph, the non-signed version of each test statistic is asymptotically chi-squared distributed. In the two variables case signed square-root versions of the test statistics can also be used. Under the null these follow, asymptotically, a standard normal distribution. The test statistics for single edge exclusion from a saturated GLL model are functions of many parameters (representing all higher order interaction terms), the number of parameters depending on the number of variables being considered. Hence, in general, the p variables case is complicated. For binary variables, Salgueiro [5] presented closed form expressions for the test statistics, for $p = 2$ and $p = 3$, as a function of cell probabilities.

In this paper T and T^s denote, respectively, the generic non-signed and signed square-root test statistics. The addition of the superscript L, S or W specifies, respectively, the likelihood ratio, the score or the Wald test statistic. In the two binary variables case $H_0 : X_1 \perp\!\!\!\perp X_2 \Leftrightarrow \lambda_{12} = 0 \Leftrightarrow \psi_{12} = 1$. The three non-signed test statistics for the exclusion of edge (1,2) from the saturated GLL model can be expressed as:

$$T_{12}^L = 2\, n_\emptyset \sum_{x_1, x_2 \in \{0,1\}} \hat{\pi}_{12}(x_1, x_2) \log \left\{ \frac{\hat{\pi}_{12}(x_1, x_2)}{\hat{\pi}_1(x_1)\, \hat{\pi}_2(x_2)} \right\}, \tag{1}$$

$$T_{12}^W = n_\emptyset \left\{ \log \hat{\psi}_{12} \right\}^2 \left\{ \frac{1}{\hat{\pi}(0,0)} + \frac{1}{\hat{\pi}(0,1)} + \frac{1}{\hat{\pi}(1,0)} + \frac{1}{\hat{\pi}(1,1)} \right\}^{-1}, \tag{2}$$

$$T_{12}^S = \frac{n_\emptyset \; \{\hat{\pi}(1,1) - \hat{\pi}_1(1)\, \hat{\pi}_2(1)\}^2}{\hat{\pi}_1(0)\, \hat{\pi}_1(1)\, \hat{\pi}_2(0)\, \hat{\pi}_2(1)}. \tag{3}$$

Signed square-root versions, T_{12}^s, can be obtained by multiplying the sign of the log-odds ratio $\hat{\psi}_{12}$ by the positive square-root of each test statistic T_{12}.

With three binary variables, the non-signed test statistics for excluding edge (i, j) from the saturated GLL model, $(i, j) \in \{(1, 2), (1, 3), (2, 3)\}$, with $H_0 : X_i \perp\!\!\!\perp X_j \mid X_k \Leftrightarrow \lambda_{ij} = \lambda_{ijk} = 0 \Leftrightarrow \psi_{ij \cdot k=0} = \psi_{ij \cdot k=1} = 1$, are:

$$T_{ij}^{L} = 2\, n_{\emptyset} \sum_{x_i,\, x_j,\, x_k \in \{0,1\}} \hat{\pi}_{ijk}(x_i, x_j, x_k)\, \log \left\{ \frac{\hat{\pi}_{ijk}(x_i, x_j, x_k)\, \hat{\pi}_k(x_k)}{\hat{\pi}_{ik}(x_i, x_k)\, \hat{\pi}_{jk}(x_j, x_k)} \right\}, \tag{4}$$

$$T_{ij}^{W} = n_{\emptyset} \left[\frac{\left\{\log(\hat{\psi}_{ij\,.\,k=0})\right\}^2}{\sum_{x_i,\, x_j \in \{0,1\}} \hat{\pi}_{ijk}^{-1}(x_i, x_j, 0)} + \frac{\left\{\log(\hat{\psi}_{ij\,.\,k=1})\right\}^2}{\sum_{x_i,\, x_j \in \{0,1\}} \hat{\pi}_{ijk}^{-1}(x_i, x_j, 1)} \right], \tag{5}$$

$$T_{ij}^{S} = n_{\emptyset} \left[\frac{\hat{\pi}_k(0)\ \{\hat{\pi}_{ijk}(1,1,0)\, \hat{\pi}_k(0) - \hat{\pi}_{ik}(1,0)\, \hat{\pi}_{jk}(1,0)\}^2}{\prod_{x_i,\, x_j \in \{0,1\}} \hat{\pi}_{ik}(x_i, 0)\, \hat{\pi}_{jk}(x_j, 0)} \right.$$
$$\left. + \frac{\hat{\pi}_k(1)\ \{\hat{\pi}_{ijk}(1,1,1)\, \hat{\pi}_k(1) - \hat{\pi}_{ik}(1,1)\, \hat{\pi}_{jk}(1,1)\}^2}{\prod_{x_i,\, x_j \in \{0,1\}} \hat{\pi}_{ik}(x_i, 1)\, \hat{\pi}_{jk}(x_j, 1)} \right]. \tag{6}$$

3 Approximations to the Distributions of the Test Statistics

3.1 Asymptotic normal approximation

The test statistics for single edge exclusion from the saturated GLL model presented in Section 2 can be written as a function of the λ-terms of the log-linear expansion. Also, the asymptotic variance matrix of the m.l.e. of $\boldsymbol{\lambda}$ is known. Smith [6, pg. 73] showed that, for p binary variables cross-classifying the contingency table, the inverse information matrix based on a single observation, $\boldsymbol{K}$, is given by $\boldsymbol{K} = \mathbf{W}^* \operatorname{diag}\{\boldsymbol{\pi}(\mathbf{x})\}^{-1} (\mathbf{W}^*)^T$, where $\mathbf{W}^*$ is obtained from $\mathbf{W}$ (defined in Section 2) by eliminating the first row.

In the two binary variables case, $\boldsymbol{K} = n_{\emptyset}\, \mathrm{var} \left[\hat{\lambda}_1\ \hat{\lambda}_2\ \hat{\lambda}_{12} \right]^T$ equals

$$\begin{bmatrix} \frac{1}{\pi(0,0)} + \frac{1}{\pi(1,0)} & \frac{1}{\pi(0,0)} & -\left\{\frac{1}{\pi(0,0)} + \frac{1}{\pi(1,0)}\right\} \\ \frac{1}{\pi(0,0)} & \frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)} & -\left\{\frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)}\right\} \\ -\left\{\frac{1}{\pi(0,0)} + \frac{1}{\pi(1,0)}\right\} & -\left\{\frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)}\right\} & \frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)} + \frac{1}{\pi(1,0)} + \frac{1}{\pi(1,1)} \end{bmatrix}. \tag{7}$$

Because the edge exclusion test statistics are functions of $\hat{\boldsymbol{\lambda}}$ and the asymptotic variance matrix of $\hat{\boldsymbol{\lambda}}$ is known, Salgueiro [5] used the delta method to derive asymptotic normal approximations to the distributions of the test statistics for single edge exclusion from the saturated GLL model, under the alternative hypothesis that the saturated model holds.

Let $\boldsymbol{\theta} = vec(\boldsymbol{\lambda})$ be the vector of parameters of interest. Its m.l.e., based on $n_{\emptyset}$ observations, is $\hat{\boldsymbol{\theta}} = vec(\hat{\boldsymbol{\lambda}})$ and has an asymptotic normal distribution with mean $\boldsymbol{\theta}$ and variance given by the inverse of the information matrix (see Cox and Hinkley [7, pg. 294]), i.e., $\sqrt{n_\emptyset}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{K})$.

The delta method (see, for example, Bishop et al. [8, pg. 493]) gives, if $f(\boldsymbol{\theta})$ is differentiable at $\boldsymbol{\theta}$,

$$\sqrt{n_\emptyset}\ \left[f(\hat{\boldsymbol{\theta}}) - f(\boldsymbol{\theta})\right] \xrightarrow{\mathcal{D}} N\left(\mathbf{0},\ \left[\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\}^T\ \boldsymbol{K}\ \left\{\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\}\right]\right).$$

In our case let $f_{ij}(\hat{\boldsymbol{\theta}}) = T_{ij}/n_\emptyset$, where T_{ij} is one of the non-signed test statistics given by Equations (1) to (6). For example, in the two variables case, using the LRT, $f_{12}^L(\boldsymbol{\theta}) = 2\sum_{x_1,\, x_2 \in \{0,1\}} \pi_{12}(x_1, x_2) \log\left\{\frac{\pi_{12}(x_1,x_2)}{\pi_1(x_1)\,\pi_2(x_2)}\right\}$. Note that f does not depend on $n_\emptyset$ and is differentiable provided all cell probabilities and all elements of $\boldsymbol{\lambda}$ are different from zero, which is the case for the saturated model.

Hence, the vector of test statistics is asymptotically normal distributed, with means given by $AE(T_{ij}) = n_\emptyset\, f_{ij}(\boldsymbol{\theta})$. For $p = 2$ and 3, respectively, $AE(T_{ij}^L)$ is given by Equations (1) and (4), $AE(T_{ij}^W)$ is given by Equations (2) and (5) and $AE(T_{ij}^S)$ is given by Equations (3) and (6), with estimators replaced by parameters.

The variance matrix of the test statistics, in the asymptotic distribution, is obtained as $n_\emptyset\ \boldsymbol{\Delta}^T\ \boldsymbol{K}\ \boldsymbol{\Delta}$, where $\boldsymbol{K}$ is the inverse of the information matrix based on a single observation and $\boldsymbol{\Delta}$ is the matrix of the derivatives of $f(\boldsymbol{\theta})$ with respect to all elements of $\boldsymbol{\lambda}$. In the two binary variables case $\boldsymbol{K} = n_\emptyset \operatorname{var}\left[\hat{\lambda}_1\ \hat{\lambda}_2\ \hat{\lambda}_{12}\right]^T$ is a 3×3 matrix and $\boldsymbol{\Delta}$ is the vector of the derivatives of $f_{12}(\boldsymbol{\theta})$ with respect to λ_1, λ_2 and λ_{12}. In the three binary variables case $\boldsymbol{K} = n_\emptyset \operatorname{var}\left[\hat{\lambda}_1\ \hat{\lambda}_2\ \hat{\lambda}_{12}\ \hat{\lambda}_3\ \hat{\lambda}_{13}\ \hat{\lambda}_{23}\ \hat{\lambda}_{123}\right]^T$ is a 7×7 matrix and $\boldsymbol{\Delta}$ is a 7×3 matrix, having in each column the derivatives of each of the three $f_{ij}(\boldsymbol{\theta})$ with respect to the seven λ-terms.

Once $\boldsymbol{\Delta}$ is written as a function of the cell probabilities, multiplying $n_\emptyset\ \boldsymbol{\Delta}^T\ \boldsymbol{K}\ \boldsymbol{\Delta}$ and simplifying the resulting expression gives, for the two variables case:

$$\mathrm{var}(T_{12}^L) = 4\, n_\emptyset \sum_{x_1,\, x_2 \in \{0,1\}} \pi_{12}(x_1,x_2) \log^2 \left\{ \frac{\pi_{12}(x_1,x_2)}{\pi_1(x_1)\, \pi_2(x_2)} \right\} - \frac{1}{n_\emptyset} \{AE(T_{12}^L)\}^2,$$

$$\mathrm{var}(T_{12}^W) = 4AE(T_{12}^W) \left[1 + \frac{\log \psi_{12} \left[\frac{1}{\{\pi(0,0)\}^2} - \frac{1}{\{\pi(0,1)\}^2} - \frac{1}{\{\pi(1,0)\}^2} + \frac{1}{\{\pi(1,1)\}^2} \right]}{\left\{ \frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)} + \frac{1}{\pi(1,0)} + \frac{1}{\pi(1,1)} \right\}^2} \right]$$

$$+ \frac{1}{n_\emptyset} \{AE(T_{12}^W)\}^2 \left[\frac{\frac{1}{\{\pi(1,0)\}^3} + \frac{1}{\{\pi(1,1)\}^3}}{\left\{ \frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)} + \frac{1}{\pi(1,0)} + \frac{1}{\pi(1,1)} \right\}^2} - 1 \right].$$

It has not yet been possible to obtain a simplified formula for $\mathrm{var}(T_{12}^S)$.

In the three binary variables case variances and covariances of the likelihood ratio test in the asymptotic distribution simplify to:

$$\mathrm{var}(T_{ij}^L) = 4\, n_\emptyset \sum_{x_i,\, x_j,\, x_k \in \{0,1\}} \pi_{ijk}(x_i,x_j,x_k)\, \log^2 \left\{ \frac{\pi_{ijk}(x_i,x_j,x_k)\, \pi_k(x_k)}{\pi_{ik}(x_i,x_k)\, \pi_{jk}(x_j,x_k)} \right\}$$

$$- \frac{1}{n_\emptyset} \{AE(T_{ij}^L)\}^2, \tag{8}$$

$$\mathrm{cov}(T_{ij}^L, T_{ik}^L) = -\tfrac{1}{n_\emptyset} \{AE(T_{ij}^L)\} \{AE(T_{ik}^L)\} + 4\, n_\emptyset \tag{9}$$

$$\sum_{x_i,\, x_j,\, x_k} \left[\pi_{ijk}(x_i,x_j,x_k) \log \left\{ \frac{\pi_{ijk}(x_i,x_j,x_k)\pi_j(x_j)}{\pi_{ij}(x_i,x_j)\pi_{kj}(x_k,x_j)} \right\} \log \left\{ \frac{\pi_{ijk}(x_i,x_j,x_k)\pi_k(x_k)}{\pi_{ik}(x_i,x_k)\pi_{jk}(x_j,x_k)} \right\} \right].$$

Again it has not yet been possible to obtain simplified formulas for $\mathrm{var}(T_{ij}^W)$, $\mathrm{var}(T_{ij}^S)$, $\mathrm{cov}(T_{ij}^W, T_{ik}^W)$ and $\mathrm{cov}(T_{ij}^S, T_{ik}^S)$.

In order to derive asymptotic normal approximations to the distributions of the signed square-root test statistics for edge $(1,2)$ exclusion from the saturated GLL model with two binary variables, let $f_{12}^s(\hat{\boldsymbol{\theta}}) = T_{12}^s/\sqrt{n_\emptyset}$, so that f^s does not depend on $n_\emptyset$. Note that $T_{12}^s = sign(\log \hat{\psi}_{12})\, \sqrt{T_{12}}$, where T_{12} is one of the non-signed test statistics given by Equations (1) to (3).

By using the delta method, each signed square-root test statistic for edge $(1,2)$ exclusion from the saturated GLL model is asymptotically normal distributed, with mean

$$AE(T_{12}^s) = \sqrt{n_\emptyset}\; f_{12}^s(\boldsymbol{\theta}) = \sqrt{n_\emptyset}\; sign(\log \psi_{12})\, \sqrt{f_{12}(\boldsymbol{\theta})} = sign(\log \psi_{12})\, \sqrt{AE(T_{12})}$$

and variance

$$\mathrm{var}(T_{12}^s) = (\sqrt{n_\emptyset})^2 \,\mathrm{var}\{f_{12}^s(\hat{\boldsymbol{\theta}})\} = n_\emptyset \left\{ \frac{sign(\log\psi_{12})}{2\sqrt{f_{12}(\boldsymbol{\theta})}} \right\}^2 \mathrm{var}\{f_{12}(\hat{\boldsymbol{\theta}})\} = \frac{\mathrm{var}(T_{12})}{4AE(T_{12})}.$$

Note that, under the alternative hypothesis that the saturated model holds, the asymptotic distribution of T_{12} (T_{ij}, in the three binary variables case) tends to a normal distribution as $n_\emptyset$ tends to infinity. At $\lambda_{12} = 0 \Leftrightarrow \psi_{12} = 1$ ($\psi_{12\cdot k=0} = \psi_{12\cdot k=1} = 1$, in the three binary variables case) the asymptotic distribution is degenerate with mean zero and variance zero. Hence, for the non-signed versions, the normal approximations will be poor for very small distances from the null.

3.2 Non-central chi-squared approximation

Local alternatives have been studied in the literature. For a composite hypothesis of the type $H_0 : \boldsymbol{\varphi} = \boldsymbol{\varphi}_0$ and nuisance parameter $\boldsymbol{\nu}$ unspecified, Cox and Hinkley [7] showed that, under local alternatives $H_a : \boldsymbol{\varphi} = \boldsymbol{\varphi}_0 + \boldsymbol{\delta}_{\boldsymbol{\varphi}}/\sqrt{n_\emptyset}$, the non-signed likelihood ratio test statistic is approximately chi-squared, with degrees of freedom equal to the dimension of $\boldsymbol{\varphi}$ and non-centrality parameter $n_\emptyset\, \boldsymbol{\delta}_{\boldsymbol{\varphi}}{}^T\, i_.(\boldsymbol{\varphi}_0 : \boldsymbol{\nu})\, \boldsymbol{\delta}_{\boldsymbol{\varphi}}$. Here $i_.(\boldsymbol{\varphi} : \boldsymbol{\nu})$ is the inverse of the variance matrix of the asymptotic normal distribution of $\sqrt{n_\emptyset}\,\hat{\boldsymbol{\varphi}}$, where $\hat{\boldsymbol{\varphi}}$ denotes the m.l.e. of $\boldsymbol{\varphi}$. Similar results hold for the non-signed Wald and score tests.

For excluding edge $(1,2)$ from a saturated graphical log-linear model with two binary variables, the hypotheses are $H_0 : \lambda_{12} = 0 \Leftrightarrow \log\psi_{12} = 0$ and $H_a : \log\psi_{12} = 0 + \delta_{\psi_{12}}/\sqrt{n_\emptyset}$. From Equation (7), the variance of the asymptotic normal distribution of $\sqrt{n_\emptyset}\hat{\lambda}_{12}$ is $\boldsymbol{K}[3,3] = \frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)} + \frac{1}{\pi(1,0)} + \frac{1}{\pi(1,1)}$. Hence, the distribution of each T_{12}, at a local alternative, can be approximated by $\chi_1^2(\gamma_{12})$, a non-central χ_1^2 with non-centrality parameter

$$\begin{aligned}\gamma_{12} &= (\sqrt{n_\emptyset}\log\psi_{12})^2\ (\boldsymbol{K}[3,3])^{-1} \\ &= \frac{n_\emptyset \log^2\psi_{12}}{\{\pi(0,0)\}^{-1} + \{\pi(0,1)\}^{-1} + \{\pi(1,0)\}^{-1} + \{\pi(1,1)\}^{-1}},\end{aligned}$$

where $\log\psi_{12}$ is the log-odds ratio under the alternative hypothesis. Note that the non-centrality parameter equals the expected value of the Wald test statistic in the asymptotic normal distribution, i.e., $\gamma_{12} = AE(T_{12}^W)$.

A simulation study was performed by Salgueiro [5] to assess the accuracy of the proposed asymptotic normal and non-central chi-squared approximations, as the sample size, the odds ratio and the marginal cell probabilities vary.

As expected, the main results are that the normal approximation performs better for large values of $n_\emptyset$, odds ratio values not close to independence and marginal probabilities not close to zero or one. The non-central chi-squared approximation performs better than the normal approximation at small distances from the null, i.e., for values of ψ_{12} close to one, particularly if $n_\emptyset$ is not large.

4 Power of Single Edge Exclusion Tests

The asymptotic approximations to the distributions of the test statistics for single edge exclusion presented in Section 3 can be used to estimate the power of the first step of a backward elimination model selection procedure for selecting the saturated GLL model. Recall that the power of a hypothesis test is the probability of rejecting the null hypothesis given a particular value of the interest parameter(s). Also recall that when testing for the presence of an edge in an independence graph, the null hypothesis corresponds to (conditional) independence. For a valid interpretation of missing edges in the independence graph using the Markov properties, it is crucial to be reasonably sure that a missing edge in the graph indeed corresponds to conditional independence. Therefore, power calculations are particularly important in the context of graphical models.

In the cross-tabulation of three binary variables there are eight cell probabilities that total one. Hence, the parameter space is seven dimensional. In the two binary variables case the parameter space has dimension three. Let $\boldsymbol{\xi}$ denote the vector of the chosen parameters, either cell probabilities or combinations of conditional odds ratios and marginal probabilities that uniquely define the contingency table under analysis, depending on the information available.

4.1 Power of non-signed tests

The power of a size α test for excluding edge (i, j) from the saturated GLL model with two or three binary variables can be estimated, using the asymptotic normal approximations derived in Section 3.1, as

$$P\left[T_{ij} > \chi^2_{d;1-\alpha} \mid \boldsymbol{\xi}\right] \simeq P\left[Z > \frac{\chi^2_{d;1-\alpha} - AE(T_{ij})}{\sqrt{\text{var}(T_{ij})}}\right], \tag{10}$$

where T_{ij} is the test statistic of interest, with mean and variance, in the asymptotic distribution, given by $AE(T_{ij})$ and $\text{var}(T_{ij})$, $Z \sim N(0, 1)$ and $\chi^2_{d;1-\alpha}$ is the upper α quantile of a χ^2_d distribution. The degrees of freedom d

are 1 and 2, respectively in the two and in the three binary variables cases. Also recall that formulas for $AE(T_{ij})$ and $\text{var}(T_{ij})$ are different in the two and in the three binary variables cases. In the two binary variables case this power can also be estimated, using the non-central chi-squared approximation derived in Section 3.2, as $P[X > \chi^2_{1;1-\alpha} \,|\, \boldsymbol{\xi}]$, where $X \sim \chi^2_1(\gamma_{12})$.

Figure 1 compares the power of the non-signed tests for excluding edge $(1, 2)$ from the saturated GLL model with two binary variables, using the normal approximation (dashed line) and the non-central chi-squared approximation (solid line), for different combinations of marginal probabilities and odds ratio values. The dotted line is the estimated exact power, used as the standard for comparison, based on 1 000 simulations. Note that there are only three curves in each panel, rather than nine, because the power functions are essentially the same for each test statistic. A sample size of 1 000 was used. The horizontal dotted lines correspond to power values of 0 and 0.05. In each plot the odds ratio, on the horizontal axis, varies from 1 to 4. The marginal probability $\pi_1(0)$ takes the values 0.1, 0.5 and 0.9 in the plots in rows 1 to 3, respectively. The marginal probability $\pi_2(0)$ takes the values 0.1, 0.2 and 0.3 in the plots in columns a) to c), respectively.

From Figure 1 it is possible to conclude that, even for a sample size of 1 000, the normal approximation is a poor approximation for values of the odds ratio close to one. Indeed, at an alternative close to the null the non-central chi-squared approximation performs much better. When the odds ratio value is far from one and $\pi_1(0)$ is around 0.9, the normal approximation performs better than the non-central chi-squared approximation (see the plots in row 3). Note that, in such cases, the minimum expected cell counts can be very small: in plot a.3) values of 3.8 and 2.9 are reached for ψ_{12} equal to 3 and 4, respectively. The chi-square approximation is very poor then. The non-central chi-squared approximation performs very well when $\pi_1(0)$ is around 0.5 (see the plots in row 2).

For a sample size of 10 000 (considered a very large sample size for a GLL model with two binary variables) the normal approximation is still a poor approximation for odds ratio values close to one. Also the non-central chi-squared approximation performs better than the normal approximation. For further details see Salgueiro [5].

The traditional definition of power relates to a test of a single null hypothesis. If there are three or more binary variables in the GLL model, the first step of a backward elimination model selection procedure may involve testing a set of null hypotheses. As the true model is the saturated model, all the hypotheses in this set are false. The probability of rejecting all these null hypotheses,

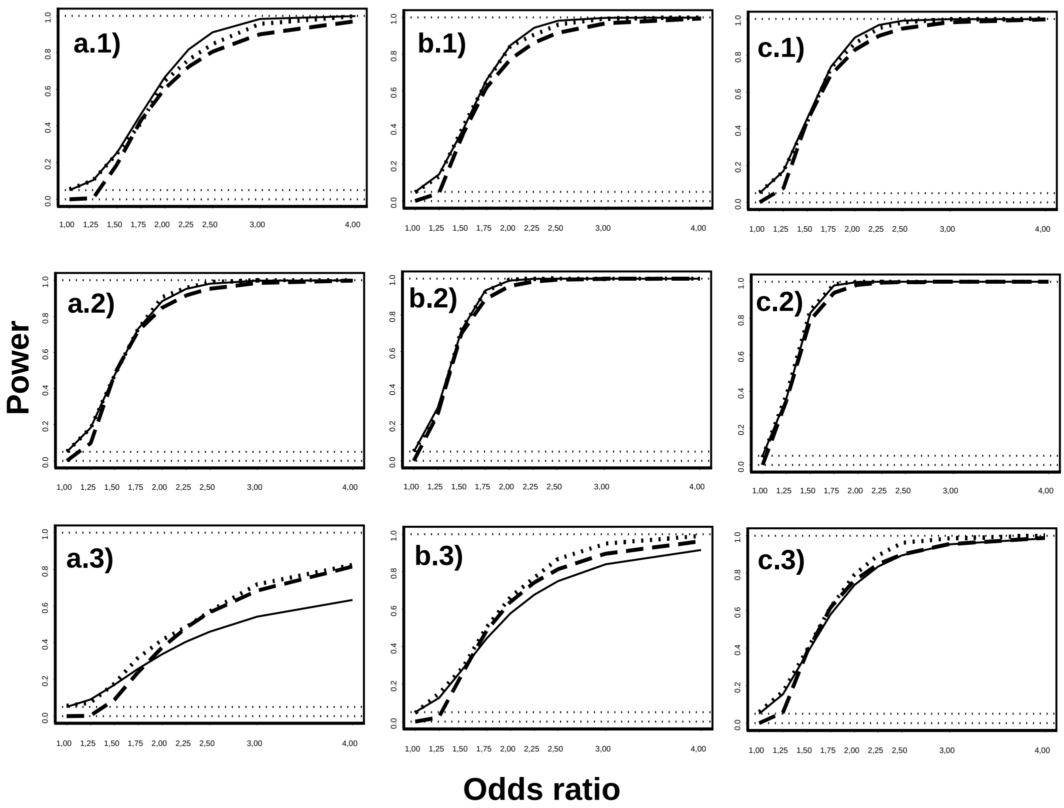

Fig. 1. Simulated (dotted line) and theoretical power values for T_{12}, with an asymptotic normal approximation (dashed line) and a non-central χ^2_1 approximation (solid line); $n_\emptyset = 1\,000$. Odds ratio ψ_{12} from 1 to 4 in each plot. $\pi_1(0)$ equals: 1) 0.1, 2) 0.5 and 3) 0.9. $\pi_2(0)$ equals: a) 0.1, b) 0.2 and c) 0.3. The horizontal dotted lines correspond to power values of 0, 0.05 and 1.

i.e., of selecting the true (saturated) model has been termed overall power in the multiple comparisons literature (see, for example, Hochberg and Tamhane [9]).

In the three binary variables case the probability of excluding neither of the two edges (i, j) and (i, k) from the saturated GLL model, when two separate edge exclusion tests are performed, can be approximated by

$$P\left[min(T_{ij}, T_{ik}) > \chi^2_{2;1-\alpha} \,|\, \boldsymbol{\xi}\right] \simeq \int_{\chi^2_{2;1-\alpha}}^{\infty} \int_{\chi^2_{2;1-\alpha}}^{\infty} \phi_2\left(\boldsymbol{\mu},\, \boldsymbol{\Sigma}\right)\; \mathrm{d}T_{ij}\; \mathrm{d}T_{ik}, \qquad (11)$$

where $\phi_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a bivariate normal density with mean vector $\boldsymbol{\mu}$ and variance matrix $\boldsymbol{\Sigma}$ given by the formulas for non-signed tests presented in Section 3.1. If the LRT is used, for example, means, variances and covariances are given, respectively, by Equations (4) (with estimators replaced by parameters), (8) and (9).

The power of selecting the saturated GLL model with three binary variables is the probability that each of the test statistics T_{12}, T_{13} and T_{23} is greater than $\chi^2_{2;1-\alpha}$, given the values of the chosen parameters in $\boldsymbol{\xi}$. A generalization of Equation (11), with a three-dimensional integral, can be used to approximate this power.

4.2 Power of signed square-root tests

Recall that, in the two binary variables case, signed square-root test statistics T^s_{12} equal $sign(\log\hat{\psi}_{12})\sqrt{T_{12}}$, where T_{12} is one of the non-signed test statistics given by Equations (1) to (3).

For a two-sided test of size α, the null hypothesis that $\lambda_{12} = 0 \Leftrightarrow \psi_{12} = 1$ is rejected if the absolute value of the signed square-root test statistic is greater than $z_{1-\alpha/2}$, where $z_{1-\alpha}$ is the upper α quantile of the standard normal distribution. Hence, the power for the two-sided size α signed square-root test of excluding edge (1,2) from the saturated GLL model with two binary variables can be approximated by

$$P\left[|T^s_{12}| > z_{1-\alpha/2}\,|\,\boldsymbol{\xi}\right] \simeq P\left[Z < \frac{z_{\alpha/2} - AE(T^s_{12})}{\sqrt{\text{var}(T^s_{12})}}\right] + P\left[Z > \frac{z_{1-\alpha/2} - AE(T^s_{12})}{\sqrt{\text{var}(T^s_{12})}}\right], \tag{12}$$

where T^s is the signed square-root version of the test statistic of interest, with mean and variance, in the asymptotic distribution, given by $AE(T^s_{12})$ and $\text{var}(T^s_{12})$. The power for a one-sided test of size $\alpha/2$ is approximated by either the first or the second term on the right-hand side of Equation (12), depending on the direction of the alternative hypothesis.

Simulation results (Salgueiro [5]) showed that the normal approximation to the power of the signed square-root tests of excluding edge $(1,2)$ from the saturated model is a very good approximation, even for moderate sample sizes, marginal probabilities close to zero or one and odds ratio values close to one. Her simulation results also showed that the asymptotic normal approximations are more accurate for the signed square-root versions than for the non-signed versions, suggesting that, when there is a choice, signed square-root test statistics should be preferred.

5 An Example: University Admissions

Data on graduate admissions to the University of California at Berkeley in 1973, presented by Agresti [1, pg. 63], are used to illustrate power calculations. In particular, the associations between admission (A: y or n), gender (G: m

or f) and department (D: 3 or 4) are investigated. For these data, $\hat{\psi}_{GA} = 1.02$ and, conditioning on D, $\hat{\psi}_{GA\cdot D=3} = 1.13$ and $\hat{\psi}_{GA\cdot D=4} = 0.92$. For $n_{\emptyset} = 1710$, the LRT statistic for $H_0 : G \perp\!\!\!\perp A|D$ is 1.05, with a p-value of 0.59, and a backward elimination model selection procedure chooses model GD, A ($\alpha = 0.05$). Hence, there is no evidence of gender discrimination in the admission process for departments 3 and 4.

To investigate the power associated with this LRT and this model selection procedure, values of $\hat{\psi}_{GA\cdot D=3}$ and $\hat{\psi}_{GA\cdot D=4}$ more extreme than the observed are considered. The five remaining parameters in $\boldsymbol{\xi}$ are selected to be the marginal probability of $D = 3$, $\pi_D(3)$, the probabilities of $G = m$ given $D = d$, $\pi_{G\cdot D}(m, d)$, and the probabilities of $A = y$ given $D = d$, $\pi_{A\cdot D}(y, d)$. These five parameters are set close to their observed values: $\pi_D(3) = 0.54$, $\pi_{G\cdot D}(m, 3) = 0.35$, $\pi_{G\cdot D}(m, 4) = 0.53$ and $\pi_{A\cdot D}(y, 3) = \pi_{A\cdot D}(y, 4) = 0.35$.

For the LRT of $H_0 : G \perp\!\!\!\perp A|D$, the power is greater than 0.62 (0.88) if one (both) of $\hat{\psi}_{GA\cdot D=3}$ and $\hat{\psi}_{GA\cdot D=4}$ is (are) outside $(0.67, 1.50)$. Hence, a sample of 1710 has enough power to detect a substantively interesting (conditional) association between G and A. For the power of selecting the saturated model the picture is less clear, as can be seen from Table 1. If one of the conditional odds ratios is less than 0.67 and the other is greater than 1.50 then the power is greater than 0.87. However, if they are both less than 0.67 or both greater than 1.50 then the power can be much lower. This is because for such values of $\hat{\psi}_{GA\cdot D}$ and the remaining values of $\boldsymbol{\xi}$ set close to their observed values, the induced conditional association between A and D is small and hence the corresponding edge is not required in the model. The results in Table 1 highlight the need for care when specifying the values in $\boldsymbol{\xi}$ to ensure that power calculations relevant to the hypotheses of interest are being performed.

6 Discussion

Presented in this paper are methods for estimating the power of single edge exclusion tests and a backward elimination model selection procedure for a GLL model with two or three binary variables. The methodology presented in this paper, in principle, can be used for GLL models with four or more binary variables. However, there is currently no straightforward way of generalizing the formulas presented, due to the complexity and dimensionality of the parameter space. In contrast, in the graphical Gaussian framework generalizations are straightforward, as shown by Salgueiro et al. [10].

Table 1
Power of selecting the saturated model for various values of $\hat{\psi}_{GA\cdot D=3}$ (in rows) and $\hat{\psi}_{GA\cdot D=4}$ (in columns); $n_{\emptyset} = 1710$.

	0.25	0.33	0.50	0.67	0.90	1.10	1.50	2.00	3.00	4.00
0.25	0.45	0.49	0.82	0.96	0.99	0.99	**0.99**	**0.99**	**1.00**	**1.00**
0.33	0.50	0.26	0.49	0.81	0.96	0.99	**0.99**	**0.99**	**0.99**	**1.00**
0.50	0.83	0.50	0.03	0.16	0.64	0.85	**0.98**	**0.99**	**0.99**	**0.99**
0.67	0.96	0.82	0.16	0.00	0.10	0.45	**0.87**	**0.98**	**0.99**	**0.99**
0.90	0.99	0.96	0.65	0.10	0.00	0.00	0.47	0.86	0.99	0.99
1.10	0.99	0.99	0.86	0.46	0.00	0.00	0.11	0.66	0.97	0.99
1.50	**0.99**	**0.99**	**0.98**	**0.88**	0.48	0.12	0.00	0.17	0.82	0.96
2.00	**0.99**	**0.99**	**0.99**	**0.98**	0.87	0.68	0.18	0.03	0.50	0.85
3.00	**1.00**	**0.99**	**0.99**	**0.99**	0.99	0.97	0.83	0.52	0.29	0.54
4.00	**1.00**	**1.00**	**0.99**	**0.99**	0.99	0.99	0.97	0.86	0.56	0.51

References

[1] A. Agresti, *Categorical Data Analysis*, John Wiley & Sons, New Jersey, 2nd edition, 2002.

[2] D. Edwards, *Introduction to Graphical Modelling*, Springer-Verlag, New York, 2nd edition, 2000.

[3] S.L. Lauritzen, *Graphical Models*, Clarendon Press, Oxford, 1996.

[4] J. Whittaker, *Graphical Models in Applied Multivariate Statistics*, John Wiley & Sons, Chichester, 1990.

[5] M.F. Salgueiro, *Distributions of Test Statistics for Edge Exclusion for Graphical Models*, Ph.D. Thesis, University of Southampton, 2003.

[6] P.W.F. Smith, *Edge Exclusion and Model Selection in Graphical Models*, Ph.D. Thesis, Lancaster University, 1990.

[7] D. Cox, D. Hinkley, *Theoretical Statistics*, Chapman & Hall, 1974.

[8] Y. Bishop, S. Fienberg, P. Holland, *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, Massachusetts, 1975.

[9] Y. Hochberg, A. Tamhane, *Multiple Comparison Procedures*, John Wiley & Sons, Chichester, 1987.

[10] M.F. Salgueiro, P.W.F. Smith, J.W. McDonald, Power of edge exclusion tests in graphical Gaussian models, *Biometrika*, 92, 2005, to appear.