UNIVERSITY OF SOUTHAMPTON

Distributions of Test Statistics for Edge Exclusion for Graphical Models

by Maria de Fátima Ramalho Fernandes Salgueiro

Doctor of Philosophy

Department of Social Statistics

Faculty of Social Sciences

November 2002

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL SCIENCES

SOCIAL STATISTICS

Doctor of Philosophy

DISTRIBUTION OF TEST STATISTICS FOR EDGE EXCLUSION FOR GRAPHICAL MODELS

by Maria de Fátima Ramalho Fernandes Salgueiro

Three test statistics for single edge exclusion from the saturated model are considered: the likelihood ratio test, the Wald test and the efficient score test. Non-signed and signed square-root versions are used. Their distributions are investigated, in particular under the alternative hypothesis that the saturated model holds. The delta-method is used to derive approximating asymptotic normal distributions. A non-central chi-square approximation is also proposed.

The power of the three test statistics for single edge exclusion is studied in detail, both for graphical Gaussian models with p variables and for graphical log-linear models with two and three binary variables. Theoretical asymptotic power functions are derived for the non-signed and the signed square-root versions of each test statistic. The normal and the non-central chi-square approximations, previously derived, are used. The quality of the approximations is assessed by simulation.

The single-factor model and the latent class model (with all variables binary) are parameterised, within the framework of graphical models, as a graphical Gaussian model and as a graphical log-linear model, respectively. The implications of such parameterisations are discussed, in particular concerning the parameter space and its admissible regions. It is proved that marginalising over the latent variable, both in a single-factor graphical Gaussian model and in a single-factor graphical log-linear model, induces no conditional independencies between manifest variables and, therefore, an independence graph that is complete. Consequently, starting with the saturated model and performing backwards elimination model selection is suggested as the most appropriate way for the data analyst to detect the presence of a latent variable. However, since model selection is subject to type II error, it is recommended that the power of the test statistics for single edge exclusion from the saturated model is taken into account.

A parallel is made between results obtained for graphical Gaussian models and for graphical log-linear models and some guidelines/recommendations are given to the data analyst.

Contents

List of Tables

List of Figures

Acknowledgements

A big thank you is first due to my supervisors, Dr. Peter W. F. Smith and Dr. John W. McDonald, for their encouragement and support throughout the project. In particular, the commitment of Dr. Peter Smith to the project has been outstanding.

The friendly, research oriented, atmosphere in the Department has also been very beneficial. A special thank you to all those who attended my departmental presentations and gave me feedback on my work, to Dr. Sandy Mackinnon for his readiness in sorting computer troubles and, obviously, to my colleagues for the interesting discussions and the sharing of different experiences throughout the process.

Acknowledgement is also due to Professor James Vickers for his help in deriving the general formula in Equation 3.4.

The Department has also given me the opportunity of presenting my work outside the University, namely in three Research Students Conferences in Probability and Statistics. Attending the 2000 RSS Conference in Reading and, therefore, meeting Professor Nanny Wermuth was made possible by a University of Southampton grant, whereas a RSS grant allowed me to attend the 2002 RSS Conference in Plymouth, and present my work there.

This research project, undertaken while on sabbatical leave from my university in Lisbon - *ISCTE*, would not have been possible without the financial support of a Portuguese grant from *Fundação para a Ciência e Tecnologia - PRAXIS XXI* - BD 19873/99. I am also indebted to both organisations for the opportunity to present my research at the 2001 International Workshop on Statistical Modelling, in Denmark, where I was able to discuss my work with Dr. David Edwards, and at the 2002 Conference of the Society for Multivariate Analysis in Behavioural Sciences, in the Netherlands.

Finally, a special *'muito obrigada'* to all those 'anonymous faces' that, in one way or another, stimulated and supported me during my three years stay in Southampton. *'SGD'*!

Glossary

$X = (X_1, \ldots, X_p)$	vector of full set of random variables. X_i is the i-th element of X
X_A	sub-vector of X, containing X_i, for $i \in A$
(X_A, X_B)	is a partition of $X_{A \cup B} = X_{AB}$
$f_{XY}(x, y)$	joint density function of random vectors X and Y
$f_X(x)$	marginal density of X
$f_{X\|Y}(x\|y)$	conditional density of X given $Y = y$
AE	asymptotic expectation
GG	graphical Gaussian model
GLL	graphical log-linear model
LRT_{ij}	likelihood ratio test statistic for edge ij exclusion from the saturated GLL model
n, $n_\emptyset$	sample size
$n(x)$	cell counts, in a contingency table
O	sample scaled inverse variance matrix (with the negatives of the sample partial correlations)
S	sample variance matrix with divisor n-1 and elements s_{ij}
T	population scaled inverse variance matrix (with the negatives of the population partial correlation coefficients)
$Score_{ij}$	efficient score test statistic for edge ij exclusion from the saturated GLL model
$T^\bullet_{ij}$	test statistic for edge ij exclusion from the saturated GG model: T^L_{ij} likelihood ratio; T^W_{ij} Wald; T^S_{ij} efficient score
U	inverse of the matrix T
V	sample variance matrix, with divisor n and elements v_{ij}
V_{AA}	sample variance matrix of X_A
V_{AB}	sample covariance matrix between X_A and X_B

$Wald_{ij}$	Wald test statistic for edge ij exclusion from the saturated GLL model
Ω	population inverse variance matrix, with elements ω_{ij}
$\pi(x)$	cell probabilities, in a contingency table
ψ_{ij}	odds ratio between X_i and X_j (binary variables)
$\psi_{ij.k}$	conditional odds ratio between X_i and X_j, for the levels of X_k
ρ_{ij}	population correlation coefficient
$\rho_{ij.rest}$	population partial correlation coefficient
Σ	population variance matrix, with elements σ_{ij}
$\hat{}$	the unconstrained m.l.e.

Chapter 1

Introduction

Graphical modelling is a form of multivariate analysis that uses mathematical graphs to represent models. It has emerged as a statistical technique explicitly based on the concept of conditional independence: the emphasis is on the assessment of the relationships that might exist between pairs of variables, conditioning on (controlling for) the remaining variables under analysis. Graphs are used to display the (conditional) independence structure of the variables. Each variable is represented by a vertex (node), associations between variables being represented by edges: either lines or directed arrows. The interpretation of the association structure among the variables can be directly read from the graph, using the Markov properties. In brief: two vertices are connected if there is a direct association between them; two vertices are not connected if the corresponding variables are conditionally independent, given (the) other variables. Thus, interpretation and model fitting are based purely on conditional independence relationships.

The conditional independence between two variables, given all the other variables in the model, can be tested using test statistics for single edge exclusion. Rejecting the null hypothesis of conditional independence between variables X_i and X_j (given the remaining variables) corresponds to having edge ij present in the (conditional independence) graph. The usual first step in fitting graphical models consists in considering the saturated model and calculating the test statistics for single edge exclusion, in order to test for the exclusion of each of the edges present in the graph.

Different test statistics for single edge exclusion have been used in the past. This thesis considers three: *the likelihood ratio test, the Wald test and the efficient score*

test statistics. It is known that under the null hypothesis of conditional independence between X_i and X_j, given the remaining variables, subject to certain regularity conditions, these test statistics are asymptotically chi-square distributed (see, for example, Whittaker, 1990, page 187). The first two aims of this thesis are to study the distribution of these three test statistics under the alternative hypothesis that the saturated model holds and, consequently, to derive asymptotic power functions for these test statistics.

The variables in the graphical model can be either *manifest* (observed) variables or *latent* (unobserved) variables. A third aim of the thesis is to investigate the parameterisation of graphical models with a single latent variable, both in the case all variables are assumed multivariate normal distributed and in the case all variables are binary, the manifest variables cross-classifying a contingency table. The former case has a parallel to factor analysis models with a single-factor, and the latter has a parallel to latent class analysis models.

This introductory chapter has four main sections. Section 1.1 provides a historical background of graphical models and graphical modelling, since their roots in areas such as statistical physics (where graphs were used to represent relationships) and in path analysis, up to recent developments, with different areas of application. Section 1.2 briefly summarises the use of latent variables in factor analysis models and in latent class analysis models. Because the focus of this research is on graphical models, a wider coverage is given to the literature on this framework, whereas references to the use of latent variables are restricted to what was considered essential for the understanding of this thesis. Section 1.3 reflects the motivation for the work undertaken and now summarised in this thesis: it lists citations from the literature that explicitly mention the 'need' to include latent variables within the framework defined by graphical models. Finally, Section 1.4 gives an overview of the structure of the thesis.

1.1 Historical Background of Graphical Models

1.1.1 The roots: graphs and path analysis

The roots for graphical models can be found in the beginning of the 20th century, in different scientific areas such as physics and genetics. In 1902 Gibbs, a physician, used

an *undirected graph* to represent the relationships and interactions in a large system of particles (possibly the atoms of a gas or a solid), as explained by Lauritzen (1996). Gibbs was interested in the interaction of groups of particles and he used undirected graphs to identify sites of particles that interact with one another.

In genetics, graphical models go back to Wright (1934) (and previous papers in 1921 and 1923), who introduced *path analysis* - which represents one of the early attempts to meet the need for methods to analyse structures. The idea is that heritable properties of natural species can be studied using a graph with directed relations, with arrows moving from parent to child. The vertices of the graph represent continuous random variables and edges represent correlations and causations. The 1934 paper classified the *'method of path coefficients as a flexible means of relating the correlation coefficients between variables in a multiple system to the functional relations among them'*. Path analysis became very popular and started being used in several areas such as economics and other social sciences - see, for example, Wold (1954), Blalock (1971) and Duncan (1975).

1.1.2 Conditional independence: an important tool for modelling data

During the 60s and 70s, there was an increased recognition of *conditional independence* as an important tool for modelling multivariate data. Dawid (1979) gave a general discussion of conditional independence in statistical theory, and introduced the notation, now widely used.

Dawid claimed that *'rather than just being another useful tool in the statistician's kitbag, conditional independence offers a new language for the expression of statistical concepts and a framework for their study'*. The framework is illustrated by an examination of the role of conditional independence in several diverse areas of statistics. Dawid wanted to show that independence and particularly conditional independence are fundamental concepts in the theory of statistical inference. Many important concepts of statistics (sufficiency, ancillarity, etc.) can be regarded as expressions of conditional independence and many results and theorems concerning these concepts are just applications of some simple general properties of conditional independence. According to Dawid, by taking conditional independence as basic, and expressing other properties

in terms of it, an unification of many areas of statistics (which appeared unrelated) is achieved. For proofs see Dawid's 1980 paper, the purpose of which was to construct a rigorous general theory of conditional independence. Yet, in the discussion of the 1979 paper, some authors were skeptical that the notion of conditional independence could be regarded as such a fundamental tool, that it could contribute to statistical inference.

1.1.3 Developments in contingency tables analysis

One of the reasons for the increased recognition of conditional independence were the advances in the field of analysis of multivariate categorical data cross-classified in *contingency tables*, commonly designated as *log-linear modelling*. Indeed, the standard statistical practice of analysing categorical data altered radically during the 70s and the development of log-linear models made it possible to formulate complex models for the dependencies between the variables cross-classified in a contingency table.

Starting points in the new approach of analysis of contingency tables include work by Birch, Bishop, Goodman and Haberman. A brief summary of their main contributions follows. In an attempt to define log-linear models for contingency tables, Birch (1963) defined interactions in three-way and many-way contingency tables as certain linear combinations of the logarithms of the expected frequencies. He then discussed maximum-likelihood estimation for many-way tables and gave solutions for the three-way contingency tables case. Birch (1964) derived a criterion for testing the null hypothesis of conditional independence of two dichotomous random variables. In Birch (1965), the detection of partial association was generalised and three different criteria for testing the null hypothesis of conditional independence of two random variables were presented. These criteria include situations when both variables are qualitative, both variables are quantitative and one variable is qualitative and the other quantitative. Bishop (1969) discussed three methods for fitting log-linear models to multivariate contingency-table data with one dichotomous variable: logit analysis, the split-table method and the full contingency-table method. The last one can be regarded as a generalised approach: basically, it is the same method as described by Birch (1963), with the advantage that an iterative fitting algorithm may be used.

Goodman (1969) presented a method of partitioning a χ^2 statistic (the statistic

based upon the likelihood-ratio criterion for testing the null hypothesis that the three variables, cross-classifying the three-way table, are mutually independent), for the $I \times J \times K$ contingency table, into additive components. These can be used to test: i) the null hypothesis of zero three factor interaction, ii) the null hypothesis that the partial association between two of the variables in the table is zero, iii) some null hypotheses concerning the two-way marginal distributions. Goodman's test statistic will have the appropriate asymptotic χ^2 distribution when a sample of n observations ($n \longrightarrow \infty$) is drawn from the population $I \times J \times K$ table. Goodman (1970) discussed, for the m-way contingency table, both the direct estimation of the multiplicative interactions among the m variables and the indirect testing of hypotheses pertaining to these interactions. He considered hierarchical hypothesis, including those that can be expressed in terms of concepts such as independence and equiprobability and conditional properties (such as conditional independence). Methods of partitioning those hypotheses were introduced, providing insight into the relationships between tests applied to the m-way table and tests applied to marginal tables formed from it. Goodman (1971) can be viewed as an addendum to the two papers mentioned above, on the analysis of marginal tables and on the partitioning of the chi-square.

Haberman (1974) provided a general, unified treatment of log-linear models for frequency data by means of a coordinate free method of linear algebra and differential calculus. This method of definition permited the development of a general theory. Previous studies had, indeed, a much more limited scope: each new type of model that had been proposed required a new examination of the properties of the maximum-likelihood estimates and a new computational procedure to find them. Statisticians had to devise new proofs of the uniqueness and existence of their estimates, and had to show that their algorithms converged. The general log-linear model has the advantage of eliminating these problems for a large class of models for frequency data.

The authors mentioned above can be considered the base for a much wider scope of recent literature and advances in this statistical area of log-linear modelling. According to Wermuth and Lauritzen (1983) the most appealing features of a hierarchical log-linear model are that: i) it has a set of minimal sufficient statistics, which is a set of proper marginal contingency tables, and ii) that each of the jointly sufficient tables matches exactly the corresponding table derived from the maximum likelihood

estimate of the joint table. Variable sets corresponding to tables listed in the set of minimal sufficient statistics are the important ones in each given model, because their observed marginal tables contain the relevant information for the joint distribution of all variables.

1.1.4 Covariance selection models and zero partial associations

Dempster (1972) pioneered the development of graphical models theory for continuous variables, assuming normality. The covariance fitting technique presented involves the exponential family of normal distributions with unknown covariance structure. He called those models *covariance selection models*. Instead of modelling the variance matrix, the structure of the inverse variance matrix is modelled. Dempster showed that the covariance structure of a multivariate normal population can be simplified by setting to zero elements of the inverse variance matrix (or elements of the inverse correlation matrix).

Wermuth (1976*a*) made use of analogies between log-linear models, which exhibit conditional independence, and the covariance selection models of Dempster and introduced models for contingency tables which are analogous to covariance selection models. She concentrated on a class of patterns of association between variables characterised by *zero partial association*, that is, pairs of variables that are conditionally independent, given the other variables. Wermuth discussed different possible patterns of association for the variables, characterised by different zero partial associations, and presented rules for the interpretation of such patterns and for computing test statistics. She stated that testing whether an element of the inverse variance matrix is zero is a test for zero partial association, that is, for conditional independence.

Wermuth (1976*b*) proposed a non-iterative model search technique to find simple patterns of association for several variables. The analysis was restricted to *multiplicative models*, therefore all patterns under consideration were interpretable in terms of zero partial association of variable pairs. In the case of a multinomial distribution, multiplicative models are a subclass of log-linear models (Birch, 1963), and in the case of a multivariate normal distribution they are a subclass of covariance selection models (Dempster, 1972). A model is defined as multiplicative if the joint distribution of sev-

eral variables can be factored into the marginal distributions of subgroups of variables (and vice versa). Wermuth showed how multiplicative models, both for contingency tables and as a subclass of covariance selection models, may be used, in a similar manner, to study simple patterns of association. However, neither Dempster nor Wermuth mention the idea of using a graph to summarise the results of an analysis.

1.1.5 The beginning of modern development of graphical models

The first key paper for the modern development of graphical models is the paper by Darroch, Lauritzen and Speed (1980), in which a way of constructing a graph that has a well defined probabilistic interpretation was proposed. That graph was called the *conditional independence graph* (or independence graph, for short). The authors used the close connection between the theory of Markov fields and that of log-linear interaction models for contingency tables to define a new class of models for such tables: *graphical models* - a subclass of the so-called *hierarchical models* that can be represented by a simple, undirected graph on as many vertices as the dimension of the corresponding table. All these models can be given an interpretation in terms of conditional independence and the interpretation can be read directly off the graph in the form of a Markov property. The class of graphical models contains that of *decomposable models* and the authors gave a simple criterion for decomposability of a given graphical model: first check whether the model is graphical and then, if it is, check whether the graph is decomposable, that is, whether there are no cyclic subgraphs of length greater than three.

Lauritzen and Wermuth (1989) is the second key paper for the modern development of graphical models, together with Wermuth and Lauritzen (1990). Early work in the field involved just one type of variable (discrete or continuous) at a time. The 1989 paper introduced *mixed graphical association models* (based on Conditional-Gaussian (CG) distributions, allowing to simultaneously consider quantitative and qualitative variables), and *graphical chain models.* Graphical chain representations were suggested to display complex association structures among variables (qualitative or quantitative). The word 'association' was used broadly, to include both symmetric associations for variables treated on an equal footing (undirected edges) and directed relations con-

cerned with the dependence of a response on explanatory variables (directed edges). Symmetric associations occur not only when there are no response variables at all, but also when some variables are joint responses or joint influences, or when they are joint intermediate variables (responses to one set of variables and influences to another set). The *chain graph* consists of a block of variables separated, for example, by time. The chain structure is supplied from subject-matter knowledge about responses and potential influences. Several examples were given in the 1990 paper. The first version of Lauritzen and Wermuth (1989) was a Research Report published in 1984 (University of Aalborg). In the meantime, Wermuth (1985) illustrated, by using several examples, how conditional independencies are reflected in data of CG distributions and showed representations of graphical chain models.

Since the paper of Darroch, Lauritzen and Speed (1980), there has been an increasing interest in graphical models, both by data analysts and by statisticians. Some examples of remarkable developments observed in this area are now given. Edwards and Kreiner (1983) considered strategies for model selection. The advised strategy was, at least at preliminary stages, to restrict model search to models which can be understood purely in terms of conditional independence relationships, i.e., to graphical models. The authors suggested that a natural first step of a graphical model selection procedure is to start with the saturated model and then test each edge (exclusion) for significance.

The possibility of reinterpreting models represented by undirected full-line graphs as dependence models in fully directed graphs was obtained by Wermuth (1980) for Gaussian systems, by Wermuth and Lauritzen (1983) for contingency tables and by Frydenberg (1990) for conditional Gaussian systems. Wermuth (1980) showed the equivalence of covariance selection models to systems of linear recursive equations with independent errors. Wermuth and Lauritzen (1983) introduced recursive models and directed graphs for contingency tables. The aim was to specify decomposable models as the intersecting class of hierarchical log-linear models with recursive models. Frydenberg and Lauritzen (1989) focused on the decomposition of the likelihood in mixed graphical association (interaction) models and Frydenberg (1990) dealt with collapsibility in such models.

In recent years new types of graphs have been proposed in the literature, and new

areas of application have been found. Since these do not directly relate to the work underlying this thesis, only a brief reference is made to them. *Graphs with two types of edge* and *covariance graphs* were introduced by Cox and Wermuth (1993) (see also Wermuth and Cox, 1992). The two types of edges (full or dashed) indicate different types of conditional analysis. A covariance matrix with some zero off-diagonal entries can be represented by an undirected graph with the corresponding edges missing: a covariance graph. Koster (1996) introduced *reciprocal graphs*, which are a generalisation of chain graphs, allowing for double-headed arrows to be present, representing feedback loops (to establish a relation with nonrecursive equation systems). Didelez (1999) defined *local independence graphs*, based on the concept of *local independence.* Such a graph represents the dynamic dependence structure of several continuous time processes which, jointly, form a so-called composable Markov process.

For a review of the application of directed acyclic graphs to probabilistic expert systems, see Spiegelhalter, Dawid, Lauritzen and Cowell (1993). The idea of handling the uncertainty in expert systems in a coherent probabilistic manner, using directed acyclic graphs to represent causal networks, had been introduced by Lauritzen and Spiegelhalter (1988). More recently (1999), a book on *Probabilistic Networks and Expert Systems* (by Cowell, Dawid, Lauritzen and Spiegelhalter) was published, and Lauritzen (2001) reviewed a number of modern applications of graphical models from different areas such as decision support systems, telecommunications and machine learning.

Software currently available for fitting graphical models includes DIGRAM, by Svend Kreiner for the analysis of contingency tables, GraphFitI ('Graph'ical models 'Fit'ting 'I'nteractions), by Angelika Bauth, for model selection, particularly in graphical chain models and MIM ('M'ixed 'I'nteraction 'M'odelling), developed by David Edwards, who claims MIM to be *'currently the world's leading software for graphical modeling'* (http://www.hypergraph.dk/). MIM is a commercial software and the latest version (version 3.1) supports undirected, directed and chain graphs, for continuous, discrete and mixed models. MIM was the software used in this research project. Throughout the thesis several references are made to MIM notation and results.

Finally, it must be noted that there has been much development in graphical models in recent years, both in the statistical and the artificial intelligence literature. Most of it is beyond the scope of this thesis and, therefore, is not reported here.

1.2 Latent Variables in Factor Analysis and in Latent Class Analysis

The concept of *latent variable* used in this thesis refers to an unobservable variable that cannot be measured directly, but only by means of indicators - the *manifest variables.* Latent variable models are statistical models specifying the joint distribution of a set of random variables, some of which are unobservable (latent), and provide an important tool for the analysis of multivariate data. The classical basic idea is that the manifest variables (indicators) are conditionally independent given the latent variable. In other words, the latent variable totally accounts for the observed relationships among the manifest variables (assumption of local independence). Several reasons why latent variables should be introduced in a model can be invoked. One possible reason, common to many techniques of multivariate analysis, is to *reduce dimensionality.* This is the idea behind *factor analysis.* The purpose is to convey the information contained in the interrelationships of many variables, with a good approximation, in a much smaller set of variables, i.e., to reduce the dimensionality of the data, preserving as much as possible of the initial structure. Another possible reason for using latent variables is related to the fact that in many fields to which statistical methods are applied, social sciences for example, central entities are handled as if they were measurable quantities, but no measurement instruments exist for them (motivation, satisfaction, quality of life, general intelligence, ...). *Factors*, latent or hidden variables represent, therefore, something underlying what is observed.

As far as level of measurement is concerned, variables can be classified as *metrical* and *categorical.* Metrical variables have realised values in the set of real numbers, and may be discrete or continuous. Categorical variables assign individuals to one of a set of categories, which may be ordered or unordered. Such classification applies to manifest as well as to latent variables. When all variables (manifest and latent) are metrical, *factor analysis methods* can be used. When all variables are categorical, *latent class analysis methods* should be used.

A brief historical background of these latent variables methods follows. Francis Galton (1822-1911) can be considered the direct forerunner of factor analysis and Charles Spearman (1863-1945) the father. After coming up with the concepts of *correlation*

and *regression*, in the late eighties Galton wrote two articles containing what could be regarded as the embryo of formal factor analysis. Spearman (1904), a psychologist who acquired statistical skills, performed several investigations that included mental variables, and considered these mental variables as being indicators of the same phenomenon. This lead him to invent factor analysis, and to introduce a *'one-factor model' for 'general intelligence'*. Yet, neither did he speak of a model, nor was the model-concept established in statistics at that time. The 'tetrad difference theorem' is also due to Spearman.

Louis Leon Thurstone (1887-1955) performed extensive factor analyses, formally what one today could call *exploratory factor analysis*. An important mark in the theoretical development of the factor analysis model was the 1956 paper by Anderson and Rubin, considered by the authors as *'an unified exposition from the viewpoint of the mathematical statistician'*. A major contribution for the recent development of factor analysis has been the work by Karl Jöreskog (1935 -). His major contributions include the introduction of *confirmatory factor analysis* (also known as *structural equation modelling*) and the development of LISREL ('LI'near 'S'tructural 'REL'ations), a software package widely used for fitting such models.

At an early stage, it became clear that factor analysis models were inappropriate to deal with non-quantitative data, and the growth of social research had greatly increased the amount of such data. Louis Guttman (1916-1987), a psychometrician, emphasised the need to apply new mathematical methods to qualitative data. The foundation was launched for the development of latent class models. Latent class analysis was mainly developed within the social and political sciences. The latent class model was introduced by Lazarsfeld and Henry (1968) as a way of formulating latent attitudinal variables from dichotomous survey items. The methodology was formalised and extended to nominal variables by Goodman (1974), who also developed the maximum likelihood algorithm that serves as the basis of several current latent class software programmes. The increasing popularity of latent class models was also due to the important work of Haberman (1979), who represented such models as log-linear models.

Variants have been proposed for ordinal variables (see, for example, Heinen, 1996) and for combinations of variables of different scale types (Moustaki, 1996). Important recent developments include also the work by Vermunt and Magidson, which has lead to the recent development of the *Latent GOLD* software.

1.3 Latent Variables in the Graphical Models Literature

This section includes references from the literature that point out the use and need of including latent variables in the framework defined by graphical models. Such references are presented in chronological order.

In the discussion of Dawid (1979), a key paper for the understanding of conditional independence in statistical theory, Professor Novick referred to latent trait theory, latent variables and conditional independence. He argued that Dawid's paper did not contain any reference to Psychometrika, or to psychometric literature in general, in which the concept of conditional independence is used extensively. Novick claimed psychometric latent trait theory is about the attainment of conditional independence through the construction of latent variables.

When reviewing structural analysis of multivariate data, Kiiveri and Speed (1982), mentioned factor analysis and latent class analysis as techniques in which considerable research had been done in recent years. The aim of their paper was to review that research, expressing it in what the authors considered a *'unified and to some extent novel way'*: in terms of (probabilistic) independence. They showed that many confirmatory factor analysis models can be expressed in terms of independence. Several examples were given, the third one being related to the single-factor model and to the latent class model. Their approach was to view the observed data as incomplete and work in the framework laid down in Dempster, Laird and Rubin (1977) for the analysis of incomplete data from exponential families (the EM algorithm).

Hodapp and Wermuth (1983) described decomposable models and claimed such models are well-suited to study multivariate relationships and to obtain a unified overview of complex variable relationships. They contrasted the graphical models framework with what they considered *'probably the best known models of dependencies'*: linear structural equation models. The authors argued that if the latter are restricted to having only observed variables, linear structural equations are obtained, of which decomposable models are a sub-class. However, if latent variables are introduced, that is, for LISREL models in general, no sufficient conditions and no necessary and sufficient conditions are known for parameter identification, which the authors regarded as a serious problem. A clear argument was made for the conditional inde-

pendence framework. Graphs for decomposable models were presented and examples from psychology were given, but no latent variables were used.

Lauritzen (1989) is a key paper surveying the mathematical and statistical theory of mixed graphical (association) models. In the discussion of the paper Jöreskog compared the LISREL methodology with mixed graphical models, claiming LISREL is primarily designed for models with latent variables, and Lauritzen was not dealing with such type of variables. In his reply to the discussion, Lauritzen pointed out that *'the graphical models described in the paper are in their infancy and far from being fully developed'*, and said that, although *'the latent problem is a question of treating the statistical problems with missing data, I am sure that this will be treated in the near future'*.

Wermuth and Lauritzen (1990) also referred to the use of latent variables in the framework defined by graphical models. The authors mentioned that graphical chain models can be viewed as extending path analysis in different directions, one of them being the possibility that some of the variables may be latent (meaning that information on them has to be obtained indirectly, with the help of other directly observable variables). Section 6 of their paper stated that the factor analysis model and the latent class model can be considered special cases of graphical models. In the discussion of the paper Kiiveri argued that estimation procedures for graphical chain models with latent variables needed to be developed, the difficult problems of identifiability that will arise having to be taken into account. The authors replied saying that *'though graphical chain models permit the inclusion of latent variables, the estimation theory for such models is basically undeveloped. We agree that this extension is essential for applications in the social sciences, that the EM algorithm may be of help, but we also believe that a new approach to solving the problem of overparameterisation is needed, which occurs for models with latent variables, but not for other conditional Gaussian chain models'*.

In the discussion section of Wermuth and Cox (1992), when referring to open problems for future research, the authors said *'the role of latent variables needs more study, both in connection with errors in measurement and with the incorporation of hidden variables'*. Wermuth (1993) is a paper with several examples on graphical chain representations, without latent variables. Nevertheless, Wermuth agreed that *'the statistical theory for models with latent variables which is needed in many applications is not yet well developed'*, which she considered *'one of the main disadvantages'* of graphical

chain models (when compared to linear structural equation models).

Whittaker (1993) contrasted the relationship between the diagrams used in latent variable modelling and the conditional independence graphs of graphical modelling and stated that *'the exact relationship between the conditional independence graph and the latent variable diagram requires some puzzling out'*. An example was given, with two observed explanatory variables and two latent dependent variables (one measured by two indicators and the other measured by a single indicator). The associated chain independence graph has three blocks corresponding to the explanatory variables, the latent variables and the observed response variables (the three indicators). However, there is no reference in the paper either to identification or to estimation of such a model. In Whittaker's words, *'it can be seen that the latent variable diagram has a very intimate relationship with the associated directed conditional independence graph, with the corollary that latent variable modelling is just a form of graphical modelling that incorporates latent variables'*. Yet, Whittaker does not seem very supportive of the use of latent variables when he says *'if it is not possible to make direct measurements on the variables of interest, latent variable modelling is the only recourse. However, there is a price to pay in that latent variable diagrams are no longer necessarily verifiable against data, and the independence structure of the latent variable model is a product of assumption rather than of empirical observation'*.

In brief: these references have one point in common - further work is required within the graphical models framework in order to incorporate latent variables, so widely used in many areas of application, particularly in the social sciences. This thesis aims to contribute to such work by investigating the parameterisation of the single-factor model as a graphical Gaussian model and the parameterisation of the latent class model as a graphical log-linear model. Special attention is also devoted to the parameter spaces and their admissible regions. Some advice is given to the data analyst, regarding strategies to be used when trying to detect the presence of a latent variable, based on the conditional independence structure of the manifest variables (indicators). Indeed, starting with the saturated model of the manifest variables and testing for all possible single edge exclusions seems an obvious way to detect the presence of a single latent variable, both in a graphical Gaussian and in a graphical log-linear model. However, the test statistics for single edge exclusion are subject to type II errors. This

suggested investigating the distribution of the test statistics for single edge exclusion, in particular under the alternative hypothesis that the saturated model holds, in order that theoretical asymptotic power functions could be derived.

The trail of thought underlying the research project that gave rise to the current thesis has just been presented. The next section gives an overview of the thesis, describing the main structure of each chapter.

1.4 Overview of the thesis

Besides this introductory chapter, the thesis has five chapters: Chapter 2 is mainly a review chapter, Chapters 3 to 5 present the contributions made by the current research project and Chapter 6 summarises the concluding remarks. A more detailed description of the contents of each chapter follows.

Chapter 2 clarifies concepts and definitions and reviews the theory underlying graphical models, required for the work undertaken and for the understanding of the thesis. Although its contents are taken from existing literature, it was necessary to summarise them using an unifying notation. Particular attention is devoted to the representation of graphical models by means of graphs, and to the Markov properties that allow conditional independence statements to be read from the graph. Graphical models based on the multivariate normal distribution (graphical Gaussian models) and graphical models for categorical variables cross-classified in a contingency table (graphical log-linear models) are explained separately. A reference is made to model selection, with an emphasis on backwards elimination. The chapter concludes by presenting the three test statistics for single edge exclusion from the saturated model that are used throughout the thesis: the likelihood ratio, the Wald and the efficient score test statistics.

Chapter 3 proposes normal approximations to the distributions of the three test statistics for single edge exclusion, under the alternative hypothesis that the saturated model holds. The delta-method is used to derive asymptotic normal approximations. Non-signed and signed square-root versions of the three test statistics are considered.

Results are given for the general p variables case for graphical Gaussian models, and for the two and the three binary variables cases for graphical log-linear models. Alternatively, in the two variables case, a non-central chi-square approximation is proposed to the distribution of the likelihood ratio test, under the alternative hypothesis of non-independence. The quality of the two approximations is assessed by simulation and some guidelines are given regarding the comparative performance of each of the two approximations.

Chapter 4 derives asymptotic normal approximations to the power of selecting the saturated model, using the non-signed and the signed square-root versions of the three tests for single edge exclusion, both in graphical Gaussian models (with p variables) and in graphical log-linear models (with two or three binary variables). In the two variables case, a non-central chi-square approximation to the power of the LRT for selecting the saturated model is also investigated. A simulation study is conducted to assess the quality of the proposed approximating power functions, for various values of the sample size and for different values of the measure of association between variables: partial correlation coefficients, in graphical Gaussian models, and odds ratio and marginal probabilities, in graphical log-linear models.

Chapter 5 parameterises the single-factor model as a graphical Gaussian model and the latent class model (all variables binary) as a graphical log-linear model and suggests using conditional independence graphs to represent such models. Parameter spaces are investigated. The conditional independence structure between manifest variables, arising from marginalising over the latent variable, either in the single-factor graphical Gaussian model or in the latent class graphical log-linear model, is studied in detail. Recommendations are given to the data analyst, regarding strategies for detecting models compatible with a single-factor graphical Gaussian model and with a latent class graphical log-linear model, taking into account power results.

Chapter 6 summarises the main conclusions of the thesis, points out its main contributions and suggests areas of further research.

Chapter 2

Graphical Gaussian and Graphical Log-linear Models

Whittaker (1990) defined a *graphical model* as *'a family of probability density functions that incorporates a specific set of conditional independence constraints listed in an independence graph'* and *graphical modelling* as *'the statistical activity of fitting graphical models to data'*. Hence, graphical modelling is another statistical technique in the framework of parametric statistical modelling.

The present chapter reviews some of the theory behind graphical models and graphical modelling considered essential for the understanding of the work undertaken in this thesis. All its contents are well established in the literature. However, it was necessary to summarise the extensive existing literature, being consistent in the notation used. The key concepts of independence and conditional independence are presented (Section 2.1), as well as the main notions and definitions of graph theory (Section 2.2) required for the understanding of the thesis. Section 2.3 explains the use of graphs to represent graphical models and Section 2.4 presents the Markov properties which allow conditional independence statements to be read from graphs. Sections 2.5 and 2.6 deal with two different families of graphical models, namely graphical Gaussian models (when all variables are multivariate normal distributed) and graphical log-linear models (when all variables are categorical, from a cross-classified multinomial distribution). Section 2.8 is devoted to the process of model selection. Section 2.9 presents the test statistics for single edge exclusion from the saturated model that are used in this thesis, both in the graphical Gaussian and in the graphical log-linear frameworks:

the likelihood ratio, the Wald and the efficient score test statistics.

2.1 Independence and Conditional Independence: Key Concepts

This section defines independence and conditional independence, first of events and then of random variables and vectors. Important lemmas are presented (for further details see, for example, Whittaker, 1990, Chapter 2). Independence and conditional independence are then defined in terms of (partial) correlation coefficients for normal distributed variables (Section 2.1.1), and in terms of (conditional) odds ratios and dependence ratios for binary variables cross-classified in contingency tables (Section 2.1.2).

Two events A and B are defined as *independent* if and only if their joint probability factorises into the product of the marginal probabilities, that is, if $P(A \cap B) = P(A)\,P(B)$. This independence is denoted by $A \perp\!\!\!\perp B$, following Dawid (1979) notation. The conditional probability of A given B, only defined when $P(B) > 0$, is given by $P(A|B) = \frac{P(A \cap B)}{P(B)}$. Three events A, B and C are defined as *mutually independent* if and only if each of the three pairs of events are independent and $P(A \cap B \cap C) = P(A)\,P(B)\,P(C)$. Indeed, the fact that the three pairs are marginally independent is not enough to guarantee mutual independence. However, mutual independence implies pairwise independence for the three pairs of events. The events A and B are *conditionally independent* given the event C, assuming that $P(C) > 0$, if and only if $P(A \cap B|C) = P(A|C)\,P(B|C)$. This conditional independence is denoted by $A \perp\!\!\!\perp B|C$. It is a symmetric relationship in the sense that $A \perp\!\!\!\perp B|C$ implies $B \perp\!\!\!\perp A|C$.

These concepts can be directly extended to random variables and vectors. Let X and Y be random vectors. The joint probability density function of (X, Y) is denoted by $f_{XY}(x, y)$, the marginal density of X by $f_X(x)$ and the conditional density of X given $Y = y$ by $f_{X|Y}(x|y)$. X and Y are *independent*, denoted by $X \perp\!\!\!\perp Y$, if and only if their joint probability density function factorises into the product of the two marginal densities, that is, if $f_{XY}(x, y) = f_X(x)\,f_Y(y)$, for all values of x and y. The independence relationship is symmetric in X and Y, that is, if $X \perp\!\!\!\perp Y$, then $Y \perp\!\!\!\perp X$. Equivalent formulations of the property $X \perp\!\!\!\perp Y$ are:

- $f_{X|Y}(x|y) = f_X(x)$, for all x (which means the conditional and the marginal density functions are identical);
- more generally, $f_{X|Y}(x|y) = g(x)$, for all x;
- $f_{XY}(x,y) = g(x)\,h(y)$, for all x and y. This *factorisation criterion* states that the random vectors X and Y are independent if and only if such functions g and h exist, without necessarily being the marginal densities.

The *reduction lemma* states that joint independence implies marginal independence, that is if (X, Y, Z) is a partitioned random vector, then $X \perp\!\!\!\perp (Y, Z)$ implies $X \perp\!\!\!\perp Y$ and $X \perp\!\!\!\perp Z$, but the converse is not necessarily true. In other words, it is not true that if X is marginally independent of Y and of Z, then X is jointly independent of Y and Z.

X and Y are *conditionally independent* given Z, denoted by $X \perp\!\!\!\perp Y|Z$, if and only if $f_{XY|Z}(x,y|z) = f_{X|Z}(x|z)\,f_{Y|Z}(y|z)$, for all values of x and y and for all z for which $f_Z(z) > 0$. Two equivalent formulations of this definition are the following:

- $f_{X|YZ}(x|y,z) = f_{X|Z}(x|z)$;
- $f_{XYZ}(x,y,z) = \frac{f_{XZ}(x,z)\,f_{YZ}(y,z)}{f_Z(z)}$ (which is a way of writing the conditional independence statement in terms of marginal densities).

The *factorisation criterion* for conditional independence states that X and Y are conditionally independent given Z if and only if there exist functions g and h such that, for all x and y and for all z for which $f_Z(z) > 0$,

$$X \perp\!\!\!\perp Y|Z = g(x,z)\,h(y,z).$$

The *reduction lemma* says that if (X, Y, Z_1, Z_2) is a partitioned random vector, then $Y \perp\!\!\!\perp (Z_1, Z_2)|X$ implies $Y \perp\!\!\!\perp Z_1|X$. If f is positive, the *block independence lemma* states that

$$Y \perp\!\!\!\perp (Z_1, Z_2)|X \;\Leftrightarrow\; [Y \perp\!\!\!\perp Z_1|(X, Z_2) \;\text{ and }\; Y \perp\!\!\!\perp Z_2|(X, Z_1)]\,.$$

Alternatively, the following equivalence can also be proved:

$$Y \perp\!\!\!\perp (Z_1, Z_2)|X \;\Leftrightarrow\; [Y \perp\!\!\!\perp Z_2|(X, Z_1) \;\text{ and }\; Y \perp\!\!\!\perp Z_1|X]\,.$$

There is a symmetry property in conditional independence, in the sense that if $X \perp\!\!\!\perp Y|Z$, then $Y \perp\!\!\!\perp X|Z$. Additional implications of the pairs of properties i) $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp Z|Y$ or ii) $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp Y$ are given in Dawid (1980).

2.1.1 Independence and conditional independence for normal distributed variables

If (X_1, X_2) has a bivariate normal distribution, then X_1 and X_2 are independent if and only if the *correlation coefficient* $\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\,\sigma_{22}}}$ is zero (σ_{11} and σ_{22} are, respectively, the variance of X_1 and of X_2 and σ_{12} is the covariance between X_1 and X_2). If (X_1, X_2, X_3) has a trivariate normal distribution and X_1 and X_2 are conditionally independent given X_3, then the off-diagonal element ω_{12} in the inverse variance matrix Ω is equal to zero. Also, the partial correlation coefficient $\rho_{12.3}$ given by $\frac{\rho_{12}-\rho_{13}\,\rho_{23}}{\sqrt{(1-\rho_{13}^2)\,(1-\rho_{23}^2)}}$ is equal to zero. Therefore, the pairwise marginal correlations satisfy $\rho_{12} - \rho_{13}\,\rho_{23} = 0$. When there are four random variables the partial correlation coefficient between X_1 and X_2 given X_3 and X_4 can be obtained as

$$\rho_{12.34} = \frac{\rho_{12.3} - \rho_{14.3}\,\rho_{24.3}}{\sqrt{(1-\rho_{14.3}^2)\,(1-\rho_{24.3}^2)}}.$$

The partial correlation coefficients are minus the off-diagonal elements of the scaled (to have a unit diagonal) inverse variance matrix, that is, $\rho_{ij.rest} = -\frac{\omega_{ij}}{\sqrt{\omega_{ii}\,\omega_{jj}}}$ (as Whittaker (1990, page 143) proves). Additional information on graphical models for variables with a normal distribution is given in Section 2.5.

2.1.2 Independence and conditional independence for binary variables

Suppose X_1 and X_2 are two binary variables. In the 2×2 contingency table the cell probabilities, $\pi(x)$, and the cell counts, $n(x)$, are

		X_2		
		0	1	
X_1	0	$\pi(0,0)$	$\pi(0,1)$	$\pi_1(0)$
	1	$\pi(1,0)$	$\pi(1,1)$	$\pi_1(1)$
		$\pi_2(0)$	$\pi_2(1)$	1

		X_2		
		0	1	
X_1	0	$n(0,0)$	$n(0,1)$	$n_1(0)$
	1	$n(1,0)$	$n(1,1)$	$n_1(1)$
		$n_2(0)$	$n_2(1)$	$n_\emptyset$

Note that $\pi(0,0) = \pi_{12}(0,0)$: the subscript is omitted, for simplicity of notation, whenever it is clear that it refers to all variables in the model. It is required throughout this thesis that probabilities in each cell are strictly positive: no structurally empty cells are allowed.

The *odds ratio* (also known as cross-product ratio) is a commonly used measure of association in a contingency table. Let

$$\psi_{12} = \frac{\pi(0,0)\,\pi(1,1)}{\pi(0,1)\,\pi(1,0)}$$

denote the odds ratio in a 2×2 contingency table. It can assume any positive value, and equals one when the variables X_1 and X_2 are independent. For standard sampling schemes, the sample odds ratio, $\hat{\psi}_{12}$, is the m.l.e. estimator of the true odds ratio ψ_{12} (see Agresti, 1996, page 24), where

$$\hat{\psi}_{12} = \frac{n(0,0)\,n(1,1)}{n(0,1)\,n(1,0)} = \frac{\hat{\pi}(0,0)\,\hat{\pi}(1,1)}{\hat{\pi}(0,1)\,\hat{\pi}(1,0)}.$$

The log transform of the sample odds ratio $\hat{\psi}_{12}$ has a sampling distribution close to normality: the asymptotic approximating normal distribution has mean $\log \psi_{12}$. As $n_\emptyset$ increases, the asymptotic mean square error for $\sqrt{n_\emptyset}\left(\log \hat{\psi}_{12} - \log \psi_{12}\right)$ is given by $\frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)} + \frac{1}{\pi(1,0)} + \frac{1}{\pi(1,1)}$, which can be estimated by $\frac{n_\emptyset}{n(0,0)} + \frac{n_\emptyset}{n(0,1)} + \frac{n_\emptyset}{n(1,0)} + \frac{n_\emptyset}{n(1,1)}$. Thus (see Sen and Singer, 1993, page 263),

$$\sqrt{n_\emptyset}\left[\log \hat{\psi}_{12} - \log \psi_{12}\right] \Big/ \sqrt{\frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)} + \frac{1}{\pi(1,0)} + \frac{1}{\pi(1,1)}} \xrightarrow{\mathcal{D}} N(0,\ 1).$$

Another possible parameterisation, proposed by Ekholm, Smith and McDonald (1995), is based on the *dependence ratio*. If X_1 and X_2 are binary variables and in both cases the 'success' is the category 1, the dependence ratio between X_1 and X_2, denoted as τ_{12}, is given by

$$\tau_{12} = \frac{P[X_1 = \text{success and } X_2 = \text{success}]}{P[X_1 = \text{success}]\,P[X_2 = \text{success}]} = \frac{P[X_1 = 1 \text{ and } X_2 = 1]}{P[X_1 = 1]\,P[X_2 = 1]} = \frac{\pi(1,1)}{\pi_1(1)\,\pi_2(1)}.$$

If X_1 and X_2 are independent, $\tau_{12} = 1$. The dependence ratio is asymmetric, i.e., it is not invariant to the coding of the variables. For convenience of notation, in the remainder of the thesis the dependence ratio for (X_1, X_2) will be denoted as $\tau_{12}(1,1)$. Here the 1 indicates the 'success' category. Consequently, $\tau_{12}(0,0)$ will denote the dependence ratio for $(1 - X_1, 1 - X_2)$, $\tau_{12}(1,0)$ the dependence ratio for $(X_1, 1 - X_2)$ and $\tau_{12}(0,1)$ the dependence ratio for $(1 - X_1, X_2)$.

If X_1, X_2 and X_3 are three binary variables, cross-classified in a $2 \times 2 \times 2$ contingency table, there are eight cell probabilities, denoted as $\pi(i,j,k)$, where $i,j,k \in \{0,1\}$. Note

that, as in the two variables case, $\pi(0,0,0) = \pi_{123}(0,0,0)$, the subscript being omitted for simplicity of notation.

A possible display of the contingency table follows

$X_3 = 0$

		X_2 = 0	X_2 = 1	
X_1	0	$\pi(0,0,0)$	$\pi(0,1,0)$	$\pi_{13}(0,0)$
	1	$\pi(1,0,0)$	$\pi(1,1,0)$	$\pi_{13}(1,0)$
		$\pi_{23}(0,0)$	$\pi_{23}(1,0)$	$\pi_3(0)$

$X_3 = 1$

		X_2 = 0	X_2 = 1	
X_1	0	$\pi(0,0,1)$	$\pi(0,1,1)$	$\pi_{13}(0,1)$
	1	$\pi(1,0,1)$	$\pi(1,1,1)$	$\pi_{13}(1,1)$
		$\pi_{23}(0,1)$	$\pi_{23}(1,1)$	$\pi_3(1)$

The marginal probabilities are obtained as

$$\pi_i(j) = P[X_i = j] = \sum_{x:\, x_i = j} \pi(x).$$

In the three binary variables case *marginal* and *conditional odds ratios* can be defined. The marginal odds ratio between X_i and X_j (with i and j from 1 to 3 and $i \neq j$), ψ_{ij} is obtained by summing the cell probabilities over both categories of the remaining variable. For example, the marginal odds ratio between X_1 and X_3 equals

$$\psi_{13} = \frac{\pi_{13}(0,0)\,\pi_{13}(1,1)}{\pi_{13}(0,1)\,\pi_{13}(1,0)}.$$

The conditional odds ratios between X_i and X_j are defined for the two categories of X_k ($k = 0$ and $k = 1$), i.e.,

$$\psi_{ij\,.\,k} = \frac{\pi(0,0,k)\,\pi(1,1,k)}{\pi(0,1,k)\,\pi(1,0,k)}.$$

If $X_i \perp\!\!\!\perp X_j \mid X_k$ both conditional odds ratios $\psi_{ij\,.\,k}$ are equal to one.

There is a direct relationship between $\log \psi_{ij}$ and the λ term(s) in the classical log-linear model related to the interaction between X_i and X_j. Assuming corner point constraints are used, in the 2×2 case,

$$\log \psi_{12} = \lambda_{12}.$$

Therefore, if X_1 and X_2 are independent, both λ_{12} and $\log \psi_{12}$ are zero. In the $2 \times 2 \times 2$ case,

$$\log(\psi_{ij\,.\,k=0}) = \lambda_{ij} \text{ and } \log(\psi_{ij\,.\,k=1}) = \lambda_{ij} + \lambda_{ijk}.$$

X_i and X_j are conditionally independent, given X_k if and only if the two-way interaction term involving X_i and X_j (λ_{ij}) as well as the three-way interaction term (λ_{ijk}) are zero, i.e.,

$$X_i \perp\!\!\!\perp X_j \mid X_k \;\Leftrightarrow\; \log(\psi_{ij\,.\,k=0}) = \log(\psi_{ij\,.\,k=1}) = \lambda_{ij} = \lambda_{ijk} = 0.$$

Additionally, $\log(\psi_{ik\,.\,j=0}) = \log(\psi_{ik\,.\,j=1}) = \lambda_{ik}$.

Using the dependence ratio parameterisation, with 1 as the 'success' category, $\tau_{123}(1,1,1)$ denotes the dependence ratio for (X_1, X_2, X_3), and is given by

$$\tau_{123}(1,1,1) = \frac{\pi(1,1,1)}{\pi_1(1)\,\pi_2(1)\,\pi_3(1)}.$$

Analogously, $\tau_{123}(0,0,1)$ denotes the dependence ratio for $(1-X_1, 1-X_2, X_3)$, and is given by

$$\tau_{123}(0,0,1) = \frac{\pi(0,0,1)}{\pi_1(0)\,\pi_2(0)\,\pi_3(1)}.$$

More details on graphical models for binary variables are given in Section 2.6.

2.2 Some Notes on Graph Theory

As previously stated, the key tool in graphical modelling is the *graph* of the model. All books on graphical models have a revision of graph theory. The current section summarises some of those revisions, by presenting the main concepts and objects of graph theory that are required for the understanding of the thesis. For further details see Whittaker (1990, Section 3.1) Cox and Wermuth (1996, Section 2.2) and Edwards (2000, Sections 1.2 and 7.1).

A *graph*, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, is a mathematical object consisting of a finite set $\mathcal{V}$ of *vertices* and a finite set $\mathcal{E}$ of *edges* (or arcs) between these vertices. Vertices correspond to the variables in the model. Each pair of variables can have no or one edge between them, which can be undirected or directed. There is a *directed edge* or *arrow* between vertices v_i and v_j in $\mathcal{V}$ if the set $\mathcal{E}$ contains the ordered pair (v_i, v_j), vertex v_i being a *parent* of vertex v_j and vertex v_j being a *child* of vertex v_i, but not the ordered pair (v_j, v_i). There is an *undirected edge* or *line* between these vertices if $\mathcal{E}$ contains both pairs (v_i, v_j) and (v_j, v_i). The graph is *undirected* if all edges are undirected and is *directed* if it only contains directed edges.

Two vertices $v_i, v_j \in \mathcal{V}$ are *adjacent*, written as $v_i \sim v_j$, if there is an undirected edge between them. Vertices v_i and v_j are also adjacent if there is an arrow (directed

edge) connecting them. Any subset $\mathcal{U} \subseteq \mathcal{V}$ induces a *subgraph* of $\mathcal{G}$. This is the graph $\mathcal{G}_{\mathcal{U}} = (\mathcal{U}, \mathcal{F})$ whose edge set $\mathcal{F}$ consists of those edges in $\mathcal{E}$ where both endpoints are in $\mathcal{U}$. The subgraph $\mathcal{G}_{\mathcal{U}}$ can be obtained by deleting all the vertices not in $\mathcal{U}$ from the graph on $\mathcal{V}$, together with all edges that do not join two elements of $\mathcal{U}$. A graph or subgraph is *complete* if all vertices are joined with either directed or undirected edges.

A *clique* of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a subset of vertices which induce a complete subgraph, but for which the addition of a further vertex from $\mathcal{V}$ makes the induced subgraph incomplete; in other words, a clique is maximally complete. A *path* is a sequence of vertices $\{v_1, v_2, \ldots, v_k\}$ such that v_i is adjacent to v_{i+1}, for each $i = 1, \ldots, (k-1)$. If there is an arrow from v_i to v_{i+1} (for each $i = 1, \ldots, k-1$), then there is a *directed path* from v_1 to v_k. The path is a *cycle* if the end points are the same, that is $v_1 = v_k$. In directed graphs the directed path is called a *directed cycle*. The cycle is *chordless* if no other than successive pairs of vertices in the cycle are adjacent. Two vertices, v_i and v_j are *connected* if there is a path from v_i to v_j and a path from v_j to v_i, and a graph is connected if all pairs of vertices are connected.

A graph is *triangulated* if it has no chordless cycles of length greater than or equal to four. If A, B, C are three disjoint subsets of $\mathcal{V}$, then C *separates* A from B in $\mathcal{G}$ if every path from any vertex in A to any in B contains at least one vertex from C.

Let $\mathcal{U} \subset \mathcal{V}$ denote a subset of vertices of the graph. The *neighbours* of $\mathcal{U}$ are all those vertices in $\mathcal{V} \setminus \mathcal{U}$ that are adjacent to a vertex in $\mathcal{U}$. The set of *parents* of $\mathcal{U}$ is $\mathrm{pa}(\mathcal{U})$, the set of all those vertices in $\mathcal{V} \setminus \mathcal{U}$ that have a child in $\mathcal{U}$. The *boundary* of $\mathcal{U}$ is $\mathrm{bd}(\mathcal{U})$, the union of the neighbours and the parents of $\mathcal{U}$. In an undirected graph the boundary and the set of neighbours are the same.

2.3 The Use of Graphs to Represent Graphical Models

This section deals with using different types of graphs to represent graphical models, in particular *conditional independence graphs* (with undirected edges), *directed acyclic graphs* (with directed edges) and *chain graphs* (with directed and undirected edges).

2.3.1 The conditional independence graph

Let X be a vector of random variables, of dimension p. The corresponding set of vertices is given by $\mathcal{V} = \{1, 2, \ldots, p\}$. An undirected graph is the *conditional independence graph* for X if there is no edge between X_i and X_j when X_i and X_j are conditionally independent given the remaining $p - 2$ variables ('the rest'), that is,

$$X_i \perp\!\!\!\perp X_j \,|\, X_{\mathcal{V}\setminus(i,j)} \Leftrightarrow (i,j) \notin \mathcal{E}.$$

The resulting graph depicts the pattern of associations between the variables in X since it is constructed from selected independencies between pairs of variables, conditioned on all the remaining variables in X.

The diagram of a graph is a picture in which vertices are drawn as dots (representing discrete variables) or circles (continuous variables) and edges are drawn as lines (undirected edges). Figure 2.1 gives an example of a conditional independence graph with five continuous variables.

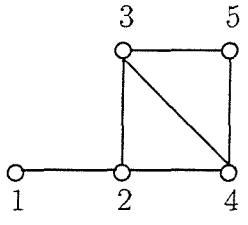

Figure 2.1: An example of a conditional independence graph.

Using the definition of a conditional independence graph, the graph in Figure 2.1 can be interpreted as follows

$$1 \perp\!\!\!\perp 3 \,|\, \{2,4,5\};\ 1 \perp\!\!\!\perp 4 \,|\, \{2,3,5\};\ 1 \perp\!\!\!\perp 5 \,|\, \{2,3,4\};\ 2 \perp\!\!\!\perp 5 \,|\, \{1,3,4\}.$$

For example, 1 is not adjacent to 3, 4 or 5. The graph is triangulated, but that would not be the case if edge $(3,4)$ was not present. The triangulated property of a graph is closely related to the existence of closed-form maximum likelihood estimates. The cliques in the graph are given by $\{1,2\}$, $\{2,3,4\}$ and $\{3,4,5\}$. $\{1,2\}$ induces a complete subgraph, and $\{2,3,4,5\}$ induces an incomplete subgraph (since edge $(2,5)$ is not present).

The *complementary graph* $\bar{\mathcal{G}}$ of a graph $\mathcal{G}$ has the same vertex set $\mathcal{V}$, the edge

set being formed by those edges not present in $\mathcal{E}$. For example, Figure 2.2 shows an independence graph and its complementary graph.

Figure 2.2: An example of an independence graph and corresponding complementary graph.

2.3.2 The directed acyclic independence graph - DAG

Directed independence graphs allow for the representation of the lack of symmetry in the roles played by the variables, due to sequence in time, to a natural ordering of the variables or to some other notion of causality. Edges are, therefore, directed (represented by single-headed arrows), but directed cycles (for example Figure 2.3) are not allowed (that is why they are named directed acyclic graphs) because there is no suitable joint probability distribution to model such a situation (see Whittaker, 1990, page 72).

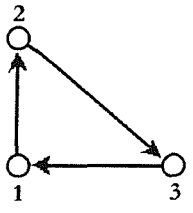

Figure 2.3: An example of a non-allowed directed graph with a directed cycle.

Specifying a complete ordering of the vertices in the graph guarantees no directed cycles. This order has to be assumed a priori, by presupposition of the data analyst, and means that any edge in the graph can only have one possible direction. As a consequence of the ordering, each variable (vertex) has a past, a present and a future.

Let $\mathcal{V}(j) = \{1, 2, \ldots, j\}$ denote the past and present of variable X_j. X is a vector of random variables, of dimension p, and $\mathcal{V} = \{1, 2, \ldots, p\}$. The symbol $\prec$ denotes the complete ordering. The *directed independence graph of* X is the directed graph $\mathcal{G}^{\prec} = (\mathcal{V}, \mathcal{E}^{\prec})$, where the edge (i, j), with $i \prec j$, is *not* in the edge set $\mathcal{E}^{\prec}$ if and only if $j \perp\!\!\!\perp i \mid \mathcal{V}(j) \setminus \{i, j\}$. In other words, the directed edge (i, j) is missing in the graph

if X_j and X_i are conditionally independent given the past (the prior variables). One should note that, whereas in undirected independence graphs the conditioning was on the rest, in directed acyclic graphs the conditioning set is restricted to the past.

Consider the directed acyclic graph $\mathcal{G}^{\prec} = (\mathcal{V}, \mathcal{E}^{\prec})$. The associated undirected graph is defined as $\mathcal{G}^{u} = (\mathcal{V}, \mathcal{E}^{u})$. $\mathcal{G}^{u}$ has the same vertex set as $\mathcal{G}^{\prec}$, each directed edge in $\mathcal{E}^{\prec}$ being replaced by an undirected edge to obtain $\mathcal{E}^{u}$. A directed graph is said to satisfy the *Wermuth condition* if no subgraph has the *unmarried parents* configuration presented in Figure 2.4.

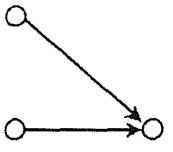

Figure 2.4: The forbidden configuration defined by Wermuth condition.

The definition of 'moral graph' is required for establishing the Markov properties of directed acyclic graphs (defined in Section 2.4.2), and follows. The *moral graph* associated with the directed graph $\mathcal{G}^{\prec} = (\mathcal{V}, \mathcal{E}^{\prec})$ is the undirected graph $\mathcal{G}^{m} = (\mathcal{V}, \mathcal{E}^{m})$ on the same vertex set $\mathcal{V}$, and with an edge set $\mathcal{E}^{m}$ obtained by including all edges in $\mathcal{E}^{\prec}$ together will all edges required to eliminate forbidden Wermuth configurations from $\mathcal{G}^{\prec}$ (that is, 'marrying parents'). Figure 2.5 shows an example of a directed acyclic graph, and associated moral graph, obtained replacing directed edges by undirected edges and 'marrying' vertices 2 and 3, since they are parents of vertex 4, by creating an undirected association between them.

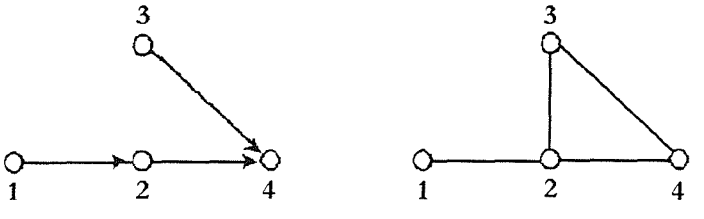

Figure 2.5: An example of a DAG and the associated moral graph.

2.3.3 The chain independence graph

Chain independence graphs have a mixture of directed and undirected edges, and conditional independence graphs and DAGs can be viewed as special cases of chain graphs, when just one type of edge (undirected or directed, respectively) is present.

Instead of the concept of complete ordering imposed in DAGs, the concept of partial ordering, denoted by the symbol $\preceq$, is used. Let us suppose the vertex set $\mathcal{V}$ is partitioned into subsets, called *blocks*, which are completely ordered, forming a chain. The induced partial order, on the vertices of $\mathcal{V}$, is that:

- $i \prec j$, whenever $i \in b_r$ and $j \in b_s$, with $r < s$, i.e., two elements from different blocks are joined by a directed edge;
- $i \preceq j$ whenever $i, j \in b_r$, i.e., two elements from the same block are joined by an undirected edge.

The parents of vertex i (when i is in block b_r) are drawn from the past, that is from blocks $b_1 \cup b_2 \cup \ldots \cup b_{r-1}$, and are joined to i by single-headed arrows pointing to i. The direction of the arrows connecting vertices in different blocks is determined by the ordering of the blocks. This block formulation excludes not only graphs with directed cycles (as was the case with DAGs), but also graphs with cycles containing at least one directed edge.

The definition of chain independence graph follows. Let $\mathcal{V}(j) = b_1 \cup b_2 \cup \ldots \cup b_r(j)$ denote the set of all past and present variables with respect to X_j (where $b_r(j)$ is the index of the block containing X_j). X is a vector of random variables, of dimension p, and $\mathcal{V} = \{1, 2, \ldots, p\}$. The symbol $\preceq$ denotes the partial ordering. The *chain independence graph of* X is the graph $\mathcal{G}^{\preceq} = (\mathcal{V}, \mathcal{E}^{\preceq})$, where the edge (i, j), with $i \preceq j$, is *not* in the edge set $\mathcal{E}^{\preceq}$ if and only if $j \perp\!\!\!\perp i \mid \mathcal{V}(j) \setminus \{i, j\}$. If this condition fails and $i \prec j$, then the edge between i and j, present in the graph, is directed and only $(i, j) \in \mathcal{E}^{\preceq}$. Otherwise, the edge between i and j, present in the graph, is undirected, and both $(i, j) \in \mathcal{E}^{\preceq}$ and $(j, i) \in \mathcal{E}^{\preceq}$. Lauritzen and Wermuth (1989) also called these *block recursive graphs.* The convention adopted to ensure an unambiguous interpretation of each pairwise relation, is that the conditioning variables of each pair are the remaining *concurrent variables.*

As in the case of DAGs, the moral graph plays a crucial role for establishing the Markov properties of chain graphs (defined in Section 2.4.3). The moral graph is also

obtained by replacing directed edges by undirected edges and by 'marrying' parents. However, the set of possible parents has to be enlarged, so that it also includes all parents of a connected subset of 'children'. Figure 2.6 gives an example of a chain graph and the associated moral graph.

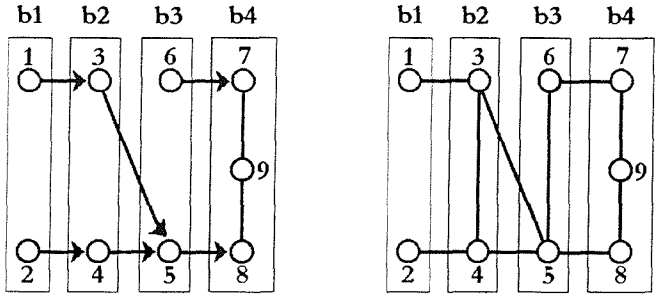

Figure 2.6: An example of a chain independence graph and the associated moral graph.

Some comments on Figure 2.6. There are four blocks: $b_1 = \{1, 2\}$, $b_2 = \{3, 4\}$, $b_3 = \{5, 6\}$ and $b_4 = \{7, 8, 9\}$. The moral graph on the right joins the parents: vertices 3 and 4 are connected because they are parents of vertex 5, vertices 5 and 6 are connected because they are parents of a connected subset of 'children' (vertices 7, 8 and 9).

Chain graphs provide the conditional independence framework for discussing multivariate regression and simultaneous equation models. Other types of graph have been suggested in the literature, including:

- *local independence graphs*, which apply the ideas of graphical models to multivariate stochastic processes (representing continuous time systems). The key concept is local independence: a component of the process is independent of the past of another component, given its own past and possibly the past of the remaining components. Didelez (1999) studied the Markov properties of local independence graphs, and her PhD thesis (Didelez, 2000) is on graphical models for event history data based on local independence;

- *covariance graphs*, introduced by Cox and Wermuth (1993), have edges drawn as *dashed lines*, displaying the marginal independence structure of a set of variables. The associated models are dual to graphical Gaussian models, in the sense that they constrain a set of elements of the variance matrix to be zero (whereas GG models constrain elements of the inverse variance matrix to be zero);

- *reciprocal graphs* generalise chain graphs by allowing double-headed arrows representing feedback loops, such as those arising in nonrecursive equation systems. Koster (1996) introduced reciprocal graphs, derived pairwise, local and global Markov properties for such class of graphs, and clarified when it is legitimate to interpret LISREL path diagrams as conditional independence graphs.

These other types of graphs will not be used in this thesis.

2.4 Markov Properties

Markov properties are important because they allow one to relate a random vector X to a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, and to interpret the latter. These properties have been presented in the literature in different ways. The decision was made to follow Whittaker's definition and not Lauritzen's, the definition of the latter having been considered very mathematical.

2.4.1 Markov properties for undirected graphs

A vector X is, with respect to a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$,

- *pairwise Markov*: if an edge missing in the graph corresponds to a conditional independence statement, i.e., $(i,j) \notin \mathcal{E} \Rightarrow X_i \perp\!\!\!\perp X_j \,|\, X_{\mathcal{V}\setminus(i,j)}$;

- *local Markov*: if any variable is independent of all the remaining variables conditional only on its boundary, i.e., for all $X_i \in \mathcal{V}$, $X_i \perp\!\!\!\perp X_{\mathcal{V}\setminus\{i,\mathrm{bd}(i)\}} \,|\, \mathrm{bd}(i)$;

- *global Markov*: if any two subsets of variables X_A and X_B separated by a third subset X_C is independent conditionally only on the variables in X_C (with X_A, X_B and X_C disjoint subsets of $\mathcal{V}$). In other words, X_C separates X_A from X_B $\Rightarrow$ $X_A \perp\!\!\!\perp X_B \,|\, X_C$.

The global Markov property implies the local Markov property, which in turn implies the pairwise Markov property.

The *separation theorem* states that, under the condition that the density function is positive, if X_A, X_B and X_C are vectors containing disjoint subsets of variables from X, and if in the conditional independence graph of X each vertex in A is separated

from each vertex in B by the subset C, then $X_A \perp\!\!\!\perp X_B \mid X_C$. Therefore, the separation theorem asserts that the pairwise property (recall that the definition of conditional independence graph (Section 2.3.1) was given based on the pairwise Markov property) implies the global Markov property ($X_A \perp\!\!\!\perp X_B \mid X_C$). As a result, the three properties are equivalent if the density is positive (which is the case with the multivariate normal distribution).

One consequence of the separation theorem is that some of the variables in the conditioning set may become *redundant*. A conditional independence between a pair of variables is *minimal* if it is not possible to apply the separation theorem to eliminate any variable from the conditioning set. Going back to the example presented in Figure 2.1 it is possible to conclude that:

- using the pairwise Markov property, there are four conditional independence statements, which are equivalent to those associated with the definition of conditional independence graph; as written in Section 2.3.1,

$$1 \perp\!\!\!\perp 3 \mid \{2,4,5\};\ 1 \perp\!\!\!\perp 4 \mid \{2,3,5\};\ 1 \perp\!\!\!\perp 5 \mid \{2,3,4\};\ 2 \perp\!\!\!\perp 5 \mid \{1,3,4\};$$

- using the local Markov property, since there are five vertices, there are five conditional independence statements, namely

$$1 \perp\!\!\!\perp \{3,4,5\} \mid 2;\ 2 \perp\!\!\!\perp 5 \mid \{1,3,4\};\ 3 \perp\!\!\!\perp 1 \mid \{2,4,5\};\ 4 \perp\!\!\!\perp 1 \mid \{2,3,5\};\ 5 \perp\!\!\!\perp \{1,2\} \mid \{3,4\};$$

- using the global Markov property, and the separation theorem, the conditional independence statements can be summarised as

$$1 \perp\!\!\!\perp 3 \mid \{2,4\};\ 1 \perp\!\!\!\perp 4 \mid \{2,3\};\ 1 \perp\!\!\!\perp 5 \mid \{2,3,4\};\ 2 \perp\!\!\!\perp 5 \mid \{3,4\}.$$

  Comparing these conditional independence statements with those given by the pairwise Markov properties it is possible to conclude that, for example, X_5 is redundant in the first two conditioning sets and variable X_1 is redundant in the last set. There are nine minimal independence statements, namely

$$1 \perp\!\!\!\perp 3 \mid 2;\ 1 \perp\!\!\!\perp 3 \mid 4;\ 1 \perp\!\!\!\perp 4 \mid 2;\ 1 \perp\!\!\!\perp 4 \mid 3;\ 1 \perp\!\!\!\perp 5 \mid 2;\ 1 \perp\!\!\!\perp 5 \mid 3;\ 1 \perp\!\!\!\perp 5 \mid 4;\ 2 \perp\!\!\!\perp 5 \mid 3;\ 2 \perp\!\!\!\perp 5 \mid 4.$$

2.4.2 Markov properties for directed graphs

When variables are naturally ordered (according to time or to some causal relationship) and directed independence graphs are used, the conditioning set has to be limited to

the 'past' and the Markov properties have to be redefined. The concept of moral graph (introduced in Section 2.3.2) is crucial for this redefinition.

A corollary of a Markov theorem for directed independence graphs states that *when the moral graph $\mathcal{G}^m$ is identical to the undirected graph $\mathcal{G}^u$ (so that no 'marrying' is required), the Markov properties of the directed graph $\mathcal{G}^{\prec}$ are exactly identical to those of the moral graph $\mathcal{G}^m$* (Whittaker, 1990, Corollary 3.5.3).

Let us consider the example of the single-factor model with three manifest variables. The classical formulation of this model is given in Section 5.1.1, whereas Section 5.2.1 presents the single-factor model in the framework of graphical models. The single-factor model implies that, given the latent variable (denoted by L), the manifest variables (denoted by 1, 2 and 3) are conditionally independent. A possible representation of such a model is given in Figure 2.7 a), by using a directed acyclic graph, with directed edges from the latent variable to each of the manifest variables, the former being the parent of the latter. The moral graph associated with this DAG is shown in Figure 2.7 b).

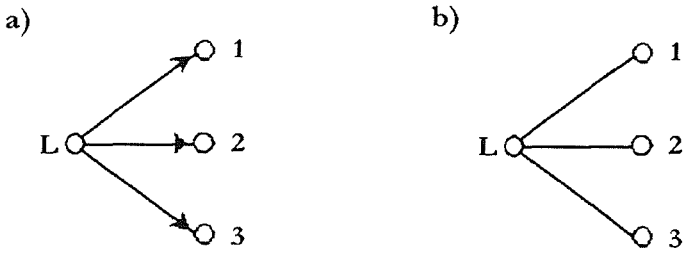

Figure 2.7: A DAG representing a classical single-factor model (in panel a)) and the corresponding moral graph (in panel b)).

Applying the corollary presented above to the directed graph in Figure 2.7 a) it is possible to conclude that, based on the Markov properties of the moral graph in Figure 2.7 b), the manifest variables are conditionally independent, given the latent variable. Consequently, the conditional independence statements, read from the Markov properties, of either an undirected or a directed representation of a classical single-factor model are exactly the same. For this reason, for simplicity and to use the notation MIM uses, undirected graphs are used in this thesis to represent single-factor models.

2.4.3 Markov properties for chain independence graphs

When variables can be partitioned into disjoint sets (blocks) which are completely ordered, forming a chain, the conditioning set for each pairwise independent statement is given by all the variables in the 'past' blocks and all the remaining variables in the 'present' block.

A Markov theorem for chained block independence graphs states that *the directed independence graph $\mathcal{G}^{\preceq}$ possesses the Markov properties of its associated moral graph, $\mathcal{G}^m$* (Whittaker, 1990, Theorem 3.6.1).

Let us now consider the example of a single-factor model with correlated residuals between two manifest variables. Although these models are not included in this thesis, they are very common in the LISREL literature. Basically, they generalise the single-factor model by relaxing the axiom of local independence, i.e., the latent variable does not account for all the associations between the manifest variables. Recent work has been undertaken by Stanghellini (1997) and Vicard (2000), regarding the issue of identification of a single-factor model with correlated residuals, using graphical rules. In the graphical models framework a single-factor model with correlated residuals can be represented in two ways:

a) as a chain independence graph with two blocks: block one includes the latent variable, block two the manifest variables. Within block two, manifest variables with correlated residuals will have an undirected edge connecting them, whereas directed arrows will connect the latent variable to the manifest variables, the arrows pointing to the latter;

b) as an undirected graph, with undirected edges connecting the latent variable with each of the manifest variables, and undirected edges connecting manifest variables which have correlated residual terms.

An example of a model with four manifest variables, the residual terms of 3 and 4 being correlated, follows. Figure 2.8 shows the two possible representations of this single-factor model with correlated residuals.

The Markov properties for chain graphs guarantee that the conditional independence statements that can be read from the two types of graph are the same. The moral graph associated with the chain graph of the single-factor model with correlated

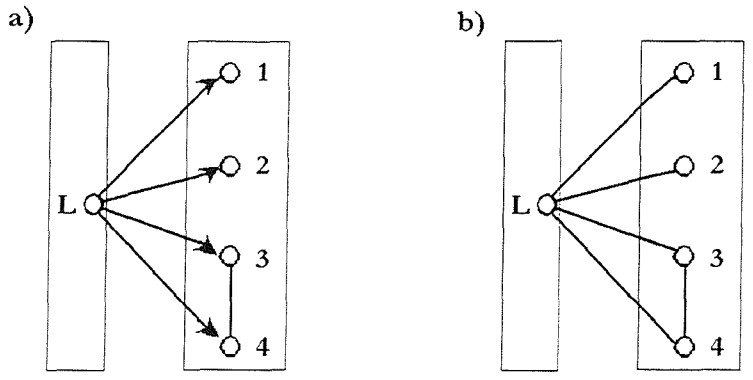

Figure 2.8: A chain graph representing a single-factor model with correlated residuals (in panel a)) and the corresponding moral graph (in panel b)).

residuals (Figure 2.8 a)) is, indeed, the conditional independence graph of that factor model (Figure 2.8 b)).

2.5 Some Notes on Graphical Gaussian Models

Graphical models based on the multivariate normal distribution are called *graphical Gaussian models* (called GG models in this thesis) or, for historical reasons, *covariance selection models.* The assumption is that the continuous random variables follow a joint multivariate normal distribution. Recall that the multivariate normal distribution is closed under marginalisation and conditioning, that is, the marginal and the conditional distributions of a multivariate normal are also multivariate normal. Indeed, suppose (X_A, X_B) is a partitioned vector with a normal distribution, with parameters mean vector (μ_A, μ_B) and variance matrix

$$\Sigma = \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{bmatrix}.$$

Then the marginal distribution of X_A is normal with mean μ_A and variance Σ_{AA}, and the conditional distribution of X_A given $X_B = x_B$ is also normal with mean

$$\mu_{A\,.\,B} = \mu_A + \Sigma_{AB}\Sigma_{BB}^{-1}(x_B - \mu_B)$$

and with variance

$$\Sigma_{AA\,.\,B} = \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}.$$

A zero element in the inverse variance matrix $\Omega = \Sigma^{-1}$, say $\omega_{ij} = 0$, determines a conditional independence statement between the two variables i and j, given all the remaining variables, implying $\rho_{ij.rest} = 0$ and, therefore, the absence of the edge between i and j in the independence graph. Hence, a GG model is a family of normal distributions for X satisfying the pairwise conditional independence restrictions underlying the independence graph. Such constraints are equivalent to specifying zeros in the inverse variance parameters that correspond to edges absent in the independence graph.

A maximum likelihood procedure is then applied in order to fit the specified model. Suppose $X = (X_1, \ldots, X_p)^T$ is a p dimensional random variable, with multivariate normal distribution, with mean μ and variance matrix Σ. The density of X, parameterised using $\Omega = \Sigma^{-1}$, can be written as

$$f_X(x; \mu, \Omega) = \frac{1}{(2\pi)^{p/2}} |\Omega|^{1/2} \exp\left\{-\frac{1}{2}(x-\mu)^T \Omega (x-\mu)\right\}.$$

Therefore, the log density equals

$$\log f_X(x; \mu, \Omega) = -\frac{p}{2}\log(2\pi) + \frac{1}{2}\log|\Omega| - \frac{1}{2}(x-\mu)^T \Omega (x-\mu).$$

The log-likelihood function for the mean μ and the inverse variance Ω, of the multivariate normal distribution based on a random sample of size n is given by

$$\begin{aligned} l(\mu, \Omega; x) &= -\tfrac{np}{2}\log(2\pi) + \tfrac{n}{2}\log|\Omega| - \tfrac{1}{2}\textstyle\sum_{i=1}^{n}(x_i-\mu)^T \Omega (x_i-\mu) \\ &= -\tfrac{np}{2}\log(2\pi) + \tfrac{n}{2}\log|\Omega| - \tfrac{n}{2}\operatorname{tr}(\Omega V) - \tfrac{n}{2}(\bar{x}-\mu)^T \Omega (\bar{x}-\mu), \end{aligned}$$

where V is the sample variance matrix with divisor n. The sampling distribution of V is Wishart. It can be proved that when Σ is unconstrained the unique maximum likelihood estimators of the mean μ and of the variance Σ, from a sample of independent normal observations, are the sample mean $\bar{x}$ and the sample variance V. Consequently, $\hat{\mu}$ is given by $\bar{x}$ and $\hat{\Omega}$, the unconstrained m.l.e. of Ω, is given by V^{-1}. For a GG model with graph $(\mathcal{V}, \mathcal{E})$, the maximum likelihood estimators of the parameters are given by the following equations

$$\begin{aligned} \hat{\sigma}_{ii} &= v_{ii} \quad i = 1, \ldots, p \\ \hat{\sigma}_{ij} &= v_{ij} \quad i \neq j \text{ and } (i,j) \in \mathcal{E} \\ \hat{\omega}_{ij} &= 0 \quad i \neq j \text{ and } (i,j) \notin \mathcal{E}. \end{aligned}$$

In brief: the estimated and the sample variances and covariances are identical for the subsets of the variables corresponding to the cliques in the graph, and the estimated

inverse variances corresponding to absent edges are set to zero. An iterative algorithm for computing maximum likelihood estimates for graphical Gaussian models was described by Speed and Kiiveri (1986).

A generalised log-likelihood ratio test statistic, the deviance, can be used to test the goodness of fit of a model and to compare different nested GG models. The *deviance* of a model M is twice the difference between the unconstrained maximum of the log likelihood (saturated model) and the maximum taken over M (the model under consideration). Because $\hat{\mu} = \bar{x}$, the last term in the expression of the log-likelihood function vanishes, i.e. $(\bar{x} - \hat{\mu})^T \Omega (\bar{x} - \hat{\mu}) = 0$. Therefore,

$$\begin{aligned}
\mathrm{dev}(M) &= 2\left\{(\max \hat{l}_{\mathrm{saturated}}) - (\max \hat{l}_{\mathrm{M}})\right\} \\
&= 2\left\{\left(\tfrac{n}{2}\log|V^{-1}| - \tfrac{n}{2}\,\mathrm{tr}(V^{-1}V)\right) - \left(\tfrac{n}{2}\log|\hat{\Omega}| - \tfrac{n}{2}\,\mathrm{tr}(\hat{\Omega}V)\right)\right\} \\
&= (n\log|V^{-1}| - np) - (n\log|\hat{\Omega}| - np) \\
&= n\log\tfrac{|V^{-1}|}{|\hat{\Omega}|} \\
&= -n\log|V\hat{\Omega}|.
\end{aligned}$$

Under the null hypothesis that model M holds, the deviance has an asymptotic chi-square distribution, with degrees of freedom equal to the number of parameters set to zero (i.e., the number of edges missing in the independence graph). Two nested models, $M_1 \subseteq M_2$, can be compared by the deviance difference (d) between them, which is then a generalised log-likelihood ratio and has an asymptotic chi-square distribution under the null hypothesis that M_1 holds. It can be calculated as $d = n\log\left(\frac{|\hat{\Sigma}_1|}{|\hat{\Sigma}_2|}\right)$, where $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ are the m.l.e. of Σ under M_1 and M_2, respectively. Alternatively, it can be calculated as $d = -n\log\left(\frac{|\hat{\Omega}_1|}{|\hat{\Omega}_2|}\right)$, where $\hat{\Omega}_1$ and $\hat{\Omega}_2$ are the m.l.e. of Ω under M_1 and M_2, respectively. The degrees of freedom equal the difference in free parameters (edges) between M_2 and M_1. The deviance difference associated with removing edge ij from the full (saturated) model can be simplified to $-n\log(1-\hat{\rho}^2_{ij.rest})$ (see Whittaker, 1990, page 189). Also, the deviance of a model entirely specified by $X_B \perp\!\!\!\perp X_C \mid X_A$ (where X_A, X_B and X_C are individual variables, or distinct groups of variables, forming a partition of X) is given by $-n\log\left(\frac{|V|\,|V_{AA}|}{|V_{A\cup B,\,A\cup B}||V_{A\cup C,\,A\cup C}|}\right)$ (see Whittaker, 1990, page 179). V_{AA} denotes the sample variance matrix, with divisor n, of the variables in X_A and $V_{A\cup B,\,A\cup B}$ denotes the sample variance matrix of the variables in the partitions X_A and X_B.

More details on test statistics for single edge exclusion, in GG models, are given in Section 2.9; model selection strategies are dealt with in Section 2.8.

2.6 Some Notes on Graphical Log-linear Models

The aim of this section is to briefly explain how to fit graphical models to multi-way contingency tables based on sampling from the cross-classified multinomial distribution. As will become clear, graphical log-linear models are a subclass of the hierarchical log-linear models, specified by parameterising the density in terms of the coefficients of its log-linear expansion. First, the notation to be used is established, and the cross-classified multinomial distribution is defined. The concepts of hierarchical log-linear model and graphical log-linear model are presented. Finally, formulae are given for obtaining the log-likelihood function and the deviance of a model.

Using *coordinate projection notation*, let us consider a p dimensional contingency table, cross-classifying the p dimensional random vector $X_{\mathcal{V}} = X = (X_1, X_2, \ldots, X_p)^T$, with $\mathcal{V} = \{1, 2, \ldots, p\}$. Let x_i denote the observed value taken by variable X_i, which has r_i categories. Let $x = (x_1, x_2, \ldots, x_p)^T$ denote a particular cell in the table, $n_{\mathcal{V}}(x_{\mathcal{V}}) = n(x)$ denote the observed cell counts and $\pi_{\mathcal{V}}(x_{\mathcal{V}}) = \pi(x)$ denote the probabilities in each cell of the table. Consider the partitioned observation $X = (X_A, X_B)$, with $B = \mathcal{V} \setminus A$. The values of X_A, denoted by x_A, are cells in a marginal table, with marginal cell counts $n_A(x_A)$ and marginal probabilities $\pi_A(x_A)$, given by $\pi_A(x_A) = \sum_{x_B} \pi_{\mathcal{V}}(x_A, x_B)$.

The definition of a cross-classified multinomial distribution follows. The p dimensional random vector X has a *cross-classified multinomial distribution of size one* if and only its density function $f_{\mathcal{V}}$ is given by $f_{\mathcal{V}}(x) = \pi_{\mathcal{V}}(x)$, assuming that $\pi_{\mathcal{V}}(x) > 0$ for all x and that $\sum_x \pi_{\mathcal{V}}(x) = 1$. 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 marginalisation and conditioning. As already mentioned, the marginal distribution of X_A, denoted by $f_A(x_A)$, is obtained by summing over the values of X_B. Consequently π_A is also positive for all x_A (since $\pi_{\mathcal{V}}(x)$ is) and sums to one (since $\pi_{\mathcal{V}}(x)$ does). In conclusion, $f_A(x_A)$ is also multinomial of size one. The conditional density function of X_B given

$X_A = x_A$, denoted by $f_{B|A}(x_B|x_A)$, is given by $\frac{\pi_{AB}(x_A,x_B)}{\pi_A(x_A)}$, for all x_B. Since the ratio $\frac{\pi_{AB}(x_A,x_B)}{\pi_A(x_A)}$ is positive and adds to one for each fixed value of x_A, the conditional distribution $f_{B|A}(x_B|x_A)$ is multinomial of size one.

Independence can be defined as follows: consider the partitioned multinomial random vector $X = (X_A, X_B)$, with $B = \mathcal{V} \setminus A$. X_A and X_B are independent if and only if the joint probability factorises as $\pi_{AB}(x) = \pi_A(x_A)\,\pi_B(x_B)$. Similarly, consider the partitioned random vector $X = (X_A, X_B, X_C)$, with $\mathcal{V} = \{A \cup B \cup C\}$. X_B and X_C are *conditionally independent* given X_A if and only if $\pi_{ABC}(x) = \frac{\pi_{AB}(x_A,x_B)\,\pi_{AC}(x_A,x_C)}{\pi_A(x_A)}$.

The aim now is to write down the density function $f_{\mathcal{V}}$ as a log-linear expansion. The *log-linear expansion of the cross-classified multinomial distribution density function* can be obtained as

$$\log f_{\mathcal{V}}(x) = \sum_{A \subseteq \mathcal{V}} \lambda_A(x_A),$$

where the summation is over all possible subsets of $\mathcal{V}$, including the empty set $\emptyset$. Each λ_A is a function of x_A and, for reasons of identifiability, corner point constraints are used, setting to zero the λ associated with the first category of each variable x_A (the reference category).

Conditional independence can easily be defined in terms of the λ. Indeed, if (X_A, X_B, X_C) is a partitioned multinomial random vector, then $X_B \perp\!\!\!\perp X_C \mid X_A$ if and only if all λ with one or more coordinate in B and one or more coordinate in C are zero. Hence, the definition of graphical log-linear model can be established. Following Whittaker (1990, page 207), given an independence graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the cross-classified multinomial distribution for the random vector X is a *multinomial graphical model for* X if the distribution of X is arbitrary, apart from the constraints that, for all pairs of coordinates not in the edge set of the graph, the λ containing the selected coordinates are zero. In other words, the density of a multinomial graphical model is given by $\log f_{\mathcal{V}}(x) = \sum_{A \subseteq \mathcal{V}} \lambda_A(x_A)$, subject to the constraints that $\lambda_A = 0$ if $(i,j) \subseteq A$ and $(i,j) \notin \mathcal{E}$. The parameters of the graphical model are the remaining λ that are not set to zero. Since in the thesis corner point constraints are used and all variables are binary $(0,1)$, $\lambda_A(x) = 0$ when $x \neq (1,1,\ldots 1)$. Therefore, and for notational simplicity, $\lambda_A(1,1,\ldots 1) = \lambda_A$.

These concepts are now illustrated by an example. Consider $X = (X_1, \ldots, X_5)^T$ a vector of five binary variables, and the log-linear expansion of the corresponding multinomial graphical model for X given by

$$\log f_{12345} = \lambda_\emptyset + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 + \lambda_{12} + \lambda_{23} + \lambda_{24} + \lambda_{34} + \lambda_{35} + \lambda_{45} + \lambda_{234} + \lambda_{345}.$$

The graphical representation of the model is given in Figure 2.9. Since $\lambda_{13} = 0$ (and, consequently, all higher order interactions involving coordinates 1 and 3 are zero) vertices 1 and 3 are not connected in the graph, corresponding to the conditional independence between 1 and 3, given the remaining variables. Since λ_{23}, λ_{24} and λ_{34} are not constrained to be zero, the three-way interaction λ_{234} is also not constrained to be zero. Note that λ_{235} was set to zero since λ_{25} is zero. Therefore, there is no edge in the graph connecting vertices 2 and 5 (2 and 5 are conditionally independent, given the rest).

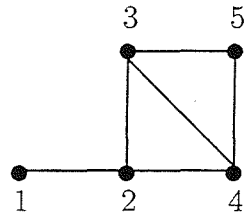

Figure 2.9: The graph of a multinomial graphical model.

Hierarchical log-linear models are a larger class of models. Indeed, a log-linear model is *hierarchical* if whenever a particular λ is constrained to be zero, all higher order λ terms are also set to zero. This means that all models that are graphical are hierarchical, but not all hierarchical log-linear models are graphical. Consider, for example the following two log-linear expansions:

i) $\log f_{123} = \lambda_\emptyset + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_{12} + \lambda_{13} + \lambda_{23}$;

ii) $\log f_{123} = \lambda_\emptyset + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_{12} + \lambda_{13} + \lambda_{23} + \lambda_{123}$.

Log-linear model i) (model 12, 13, 23) is hierarchical but it is not graphical, because the constraint $\lambda_{123} = 0$ does not correspond to a pairwise conditional independence, whereas model ii) (model 123) is hierarchical and graphical. One should note that, in both cases (model i) and model ii)), the independence graph has the representation of

Figure 2.10. Therefore, an independence graph does not correspond to a unique log-linear model; many different log-linear models may have the same independence graph, as long as they contain the same two factor interactions. However, every independence graph corresponds to a unique graphical log-linear model, which can be read from the graph by identifying the generating class as the set of all cliques in the graph. The graph in Figure 2.10 has the clique {1,2,3}, therefore it corresponds to the graphical log-linear model 123 (in MIM notation).

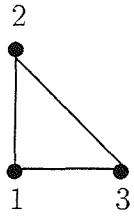

Figure 2.10: The graphical log-linear model 123.

In brief: a hierarchical log-linear model is *graphical* if and only if its maximal λ terms (also known as *generators*) correspond to cliques in the independence graph.

The *log-likelihood function*, based on a random sample of $n_\emptyset$ multinomial random observations, can be written as a function of the λ as

$$l(\lambda, n) = \sum_A \sum_{x_A} n_A(x_A)\, \lambda_A(x_A).$$

The *deviance of a model M* can be obtained as

$$dev(M) = 2\sum_x n(x) \log\left(\frac{n(x)}{n_\emptyset \hat{\pi}(x)}\right) = 2\sum_x n(x) \log\left(\frac{\pi(x)}{\hat{\pi}(x)}\right),$$

where $\hat{\pi}$ is the m.l.e. of π. Under the null hypothesis that model M holds, the deviance has an asymptotic chi-squared distribution, with degrees of freedom given by the number of parameters set to zero. Two nested models, $M_1 \subseteq M_2$, can be compared using the deviance difference, which is a generalised log-likelihood ratio statistic and has an asymptotic chi-square distribution under the null hypothesis that M_1 holds. The deviance difference equals $2\sum_x n(x) \log\left(\frac{\hat{\pi}_2(x)}{\hat{\pi}_1(x)}\right)$, where $\hat{\pi}_1(x)$ and $\hat{\pi}_2(x)$ are the m.l.e. of $\pi(x)$ under M_1 and M_2, respectively. The deviance difference associated with removing edge ij from the saturated model is given by (see Whittaker, 1990, page 224)

$$dev(X_i \perp\!\!\!\perp X_j \mid X_{V\setminus\{i,j\}}) = 2 \sum_{\text{all cells}} n_V(x_V) \log\left(\frac{n_V(x_V)\ n_{V\setminus\{i,j\}}(x_{V\setminus\{i,j\}})}{n_{V\setminus i}(x_{V\setminus\{i\}})\ n_{V\setminus j}(x_{V\setminus\{j\}})}\right).$$

Under the null hypothesis that $X_i \perp\!\!\!\perp X_j \mid X_{\mathcal{V}\setminus\{i,j\}}$, the deviance difference has an asymptotic chi-square distribution with degrees of freedom given by

$$df = r_{\mathcal{V}\setminus\{i,j\}}\,(r_i - 1)\,(r_j - 1).$$

More details on test statistics for single edge exclusion, in GLL models, are given in Section 2.9 and Section 2.8 deals with models selection strategies.

2.7 The Inverse of the Information Matrix

2.7.1 The inverse of the information matrix in GG models

Smith (1990, page 21) showed that, for GG models, the inverse information matrix (or asymptotic variance matrix of the m.l.e. of ω) can be written as

$$n\,\mathrm{cov}(\hat{\omega}_{ij}, \hat{\omega}_{rs}) = \omega_{ir}\omega_{js} + \omega_{is}\omega_{jr}, \tag{2.1}$$

where $\hat{\omega}$ is the unconstrained m.l.e. of ω. Similar result was obtained by Cox and Wermuth (1990), by arguing that if Y (of dimension p) is multivariate normal distributed with mean μ and variance matrix Σ (with elements σ) then, asymptotically, the m.l.e. of the σ are normal distributed with mean σ and variance $\frac{1}{n}\,\mathrm{Iss}(\Sigma)$, i.e., $\sqrt{n}(\hat{\sigma} - \sigma) \xrightarrow{\mathcal{D}} N(0, \mathrm{Iss}(\Sigma))$, where $\mathrm{Iss}(\Sigma)$, the Isserlis matrix of Σ (Isserlis, 1918) is the symmetric matrix with elements $n\,\mathrm{cov}(\hat{\sigma}_{ij}, \hat{\sigma}_{rs}) = \sigma_{ir}\sigma_{js} + \sigma_{is}\sigma_{jr}$. Consequently, it can be shown that the asymptotic variance matrix of the m.l.e. of the ω can be obtained as $\frac{1}{n}\,\mathrm{Iss}(\Omega)$, where $\Omega = \Sigma^{-1}$ and $\sqrt{n}(\hat{\omega} - \omega) \xrightarrow{\mathcal{D}} N(0, \mathrm{Iss}(\Omega))$.

Roverato and Whittaker (1998) described some properties of $\mathrm{Iss}(\Sigma)$ and derived the zero structure of its inverse. A novel edge set notation was used, allowing for the symmetry between Σ and $\mathrm{Iss}(\Sigma)$ to be highlighted. From Equation 20 of Roverato and Whittaker (1998) the asymptotic variance of the m.l.e. of the canonical parameters of the saturated GG model becomes equal to $\frac{1}{n}\,\mathrm{Iss}(\Sigma^{-1})$, which is exactly the result previously derived, in alternative ways, by Smith (1990) and Cox and Wermuth (1990).

Let K denote the inverse information matrix based on a single observation and, therefore, taking values that do not depend on the sample size n. In the two variables case, and using Equation 2.1,

$$K = n \operatorname{var} \begin{bmatrix} \hat{\omega}_{12} \\ \hat{\omega}_{11} \\ \hat{\omega}_{22} \end{bmatrix} = \begin{bmatrix} \omega_{11}\omega_{22} + \omega_{12}^2 & 2\omega_{11}\omega_{12} & 2\omega_{22}\omega_{12} \\ 2\omega_{11}\omega_{12} & 2\omega_{11}^2 & 2\omega_{12}^2 \\ 2\omega_{22}\omega_{12} & 2\omega_{12}^2 & 2\omega_{22}^2 \end{bmatrix}. \tag{2.2}$$

Similar reasoning can be used when three or more variables are present. For example, in the three variables case,

$$K = n \operatorname{var} \begin{bmatrix} \hat{\omega}_{12} \\ \hat{\omega}_{13} \\ \hat{\omega}_{23} \\ \hat{\omega}_{11} \\ \hat{\omega}_{22} \\ \hat{\omega}_{33} \end{bmatrix} =$$

$$= \begin{bmatrix} \omega_{11}\omega_{22} + \omega_{12}^2 & \omega_{11}\omega_{23} + \omega_{13}\omega_{12} & \omega_{12}\omega_{23} + \omega_{13}\omega_{22} & 2\omega_{11}\omega_{12} & 2\omega_{12}\omega_{22} & 2\omega_{13}\omega_{23} \\ \omega_{11}\omega_{23} + \omega_{13}\omega_{12} & \omega_{11}\omega_{33} + \omega_{13}^2 & \omega_{12}\omega_{33} + \omega_{13}\omega_{23} & 2\omega_{11}\omega_{13} & 2\omega_{12}\omega_{23} & 2\omega_{13}\omega_{33} \\ \omega_{12}\omega_{23} + \omega_{13}\omega_{22} & \omega_{12}\omega_{33} + \omega_{13}\omega_{23} & \omega_{22}\omega_{33} + \omega_{23}^2 & 2\omega_{13}\omega_{12} & 2\omega_{22}\omega_{23} & 2\omega_{23}\omega_{33} \\ 2\omega_{11}\omega_{12} & 2\omega_{11}\omega_{13} & 2\omega_{13}\omega_{12} & 2\omega_{11}^2 & 2\omega_{12}^2 & 2\omega_{13}^2 \\ 2\omega_{12}\omega_{22} & 2\omega_{12}\omega_{23} & 2\omega_{22}\omega_{23} & 2\omega_{12}^2 & 2\omega_{22}^2 & 2\omega_{23}^2 \\ 2\omega_{13}\omega_{23} & 2\omega_{13}\omega_{33} & 2\omega_{23}\omega_{33} & 2\omega_{13}^2 & 2\omega_{23}^2 & 2\omega_{33}^2 \end{bmatrix}.$$

2.7.2 The inverse of the information matrix in GLL models

Smith (1990, page 73) showed that the inverse information matrix for a sample of size $n_\emptyset$ (or asymptotic variance matrix of the m.l.e. of λ) can be written as $\frac{1}{n_\emptyset} W^* \operatorname{diag}(\pi(x))^{-1} (W^*)^T$, where W^* is obtained from W by eliminating the first row. If p is the number of binary variables cross-classifying the contingency table, $W = W_1 \otimes W_2 \otimes \ldots \otimes W_p$ is the Kronecker product of p W_i matrices of the form

$$W_i = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}.$$

Keeping the notation used in Section 2.7.1, the inverse information matrix based on a single observation, K, is given by

$$K = W^* \operatorname{diag}(\pi(x))^{-1} (W^*)^T. \tag{2.3}$$

In the two binary variables case,

$$W = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 1 & -1 & -1 & 1 \end{bmatrix}. \text{ Hence, } W^* = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

Consequently,

$$\begin{aligned} K &= n_\emptyset \operatorname{var} \begin{bmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \\ \hat{\lambda}_{12} \end{bmatrix} = W^* \begin{bmatrix} \pi(0,0) & 0 & 0 & 0 \\ 0 & \pi(1,0) & 0 & 0 \\ 0 & 0 & \pi(0,1) & 0 \\ 0 & 0 & 0 & \pi(1,1) \end{bmatrix}^{-1} (W^*)^T = \\ &= \begin{bmatrix} \frac{1}{\pi(0,0)} + \frac{1}{\pi(1,0)} & \frac{1}{\pi(0,0)} & -\left(\frac{1}{\pi(0,0)} + \frac{1}{\pi(1,0)}\right) \\ \frac{1}{\pi(0,0)} & \frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)} & -\left(\frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)}\right) \\ -\left(\frac{1}{\pi(0,0)} + \frac{1}{\pi(1,0)}\right) & -\left(\frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)}\right) & \frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)} + \frac{1}{\pi(1,0)} + \frac{1}{\pi(1,1)} \end{bmatrix}. \end{aligned} \tag{2.4}$$

Analogously, in the three binary variables case,

$$W = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$$

and

$$W^* = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 \end{bmatrix}.$$

Therefore,

$$K = n_\emptyset \text{ var} \begin{bmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \\ \hat{\lambda}_{12} \\ \hat{\lambda}_3 \\ \hat{\lambda}_{13} \\ \hat{\lambda}_{23} \\ \hat{\lambda}_{123} \end{bmatrix} =$$

$$= W^* \begin{bmatrix} \pi(0,0,0) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \pi(1,0,0) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \pi(0,1,0) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \pi(1,1,0) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \pi(0,0,1) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \pi(1,0,1) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \pi(0,1,1) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \pi(1,1,1) \end{bmatrix}^{-1} (W^*)^T. \tag{2.5}$$

After the calculations K, a symmetric 7×7 matrix, is obtained, with the following diagonal elements (corresponding to the variances of the $\hat{\lambda}$, respectively $\hat{\lambda}_1$, $\hat{\lambda}_2$, $\hat{\lambda}_{12}$, $\hat{\lambda}_3$, $\hat{\lambda}_{13}$, $\hat{\lambda}_{23}$ and $\hat{\lambda}_{123}$):

$$K[1,1] = \left(\frac{1}{\pi(0,0,0)} + \frac{1}{\pi(1,0,0)}\right); \qquad K[2,2] = \left(\frac{1}{\pi(0,0,0)} + \frac{1}{\pi(0,1,0)}\right);$$

$$K[3,3] = \left(\frac{1}{\pi(0,0,0)} + \frac{1}{\pi(0,1,0)} + \frac{1}{\pi(1,0,0)} + \frac{1}{\pi(1,1,0)}\right); \qquad K[4,4] = \left(\frac{1}{\pi(0,0,0)} + \frac{1}{\pi(0,0,1)}\right);$$

$$K[5,5] = \left(\frac{1}{\pi(0,0,0)} + \frac{1}{\pi(1,0,0)} + \frac{1}{\pi(0,0,1)} + \frac{1}{\pi(1,0,1)}\right);$$

$$K[6,6] = \left(\frac{1}{\pi(0,0,0)} + \frac{1}{\pi(0,1,0)} + \frac{1}{\pi(0,0,1)} + \frac{1}{\pi(0,1,1)}\right);$$

$$K[7,7] = \left(\frac{1}{\pi(0,0,0)} + \frac{1}{\pi(0,1,0)} + \frac{1}{\pi(1,0,0)} + \frac{1}{\pi(1,1,0)} + \frac{1}{\pi(0,0,1)} + \frac{1}{\pi(0,1,1)} + \frac{1}{\pi(1,0,1)} + \frac{1}{\pi(1,1,1)}\right).$$

The off-diagonal elements are the covariances between the $\hat{\lambda}$, as follows:

$$K[2,1] = K[4,1] = -K[6,1] = K[4,2] = -K[5,2] = -K[4,3] = \frac{1}{\pi(0,0,0)};$$

$$K[3,1] = K[5,1] = -K[7,1] = -K[5,3] = -\left(\frac{1}{\pi(0,0,0)} + \frac{1}{\pi(1,0,0)}\right);$$

$$K[3,2] = K[6,2] = -K[7,2] = -K[6,3] = -\left(\frac{1}{\pi(0,0,0)} + \frac{1}{\pi(0,1,0)}\right);$$

$$K[7,3] = -\left(\frac{1}{\pi(0,0,0)} + \frac{1}{\pi(0,1,0)} + \frac{1}{\pi(1,0,0)} + \frac{1}{\pi(1,1,0)}\right);$$

$$K[5,4] = K[6,4] = -K[7,4] = -K[6,5] = -\left(\frac{1}{\pi(0,0,0)} + \frac{1}{\pi(0,0,1)}\right);$$

$$K[7,5] = -\left(\frac{1}{\pi(0,0,0)} + \frac{1}{\pi(1,0,0)} + \frac{1}{\pi(0,0,1)} + \frac{1}{\pi(1,0,1)}\right);$$

$$K[7,6] = -\left(\frac{1}{\pi(0,0,0)} + \frac{1}{\pi(0,1,0)} + \frac{1}{\pi(0,0,1)} + \frac{1}{\pi(0,1,1)}\right).$$

2.8 Some Notes on Model Selection

Several methods can be used to perform model selection both in GG and in GLL models. For a review on the topic see, for example, Cox and Wermuth (1996, Chapter 8) and Edwards (2000, Chapter 6).

Taking into account the *lattice* of all possible models, strategies to select a graphical model may include incremental or global search procedures, based on significance tests, or procedures that optimise an information criteria. The information criteria to be minimised can be the Akaike's Information criteria (AIC) or the Bayesian Information criteria (BIC), given by

$$AIC = -2\log L - 2f \quad \text{and} \quad BIC = -2\log L - \sqrt{nf},$$

where L is the maximised likelihood under the model, n is the number of observations and f is the number of free parameters in the model. Global search procedures aim to detect minimally adequate models. A model is adequate if its deviance is sufficiently small. From all the adequate models, the one with the fewest parameters should be considered. A model is minimally adequate if there is no other model nested in it that is also adequate. Edwards and Havánek (1985, 1987) considered models that are minimally adequate and proposed a fast procedure for model selection based on the *principle of coherence*. This principle, due to Gabriel (1969), states that, for any two nested models, $M_0 \subseteq M_1$, if M_1 is rejected, then M_0 must also be rejected. Conversely, if M_0 is accepted, then M_1 must also be accepted. Models are selected or rejected on the basis of the overall goodness of fit, and not on the basis of deviance differences.

Incremental search procedures include backwards elimination, forward selection and stepwise procedures that alternate between a backwards step and a forward step. Backwards elimination starts with the saturated model and tests for all pairwise conditional independence statements using test statistics for single edge exclusion (the deviance is the most commonly used statistic). The least significant edge is removed and the

procedure continues until all edges are considered significant given the specified significance level. When performing model selection using backwards elimination, the principle of coherence just means that if the removal of an edge is rejected at one step, then the edge is not subsequently eligible for removal. Forward selection starts with the independence model and includes in the model edges that significantly improve the fit.

Stepwise model selection can also be performed taking into account only *decomposable models*. At any step, the edges whose exclusion (backwards procedure) or inclusion (forward procedure) would result in a non-decomposable model are considered non-eligible for removal (inclusion). The classical definition of a decomposable model is given by Haberman (1974): a model is decomposable if either (i) it is complete or (ii) it is reducible to two decomposable components. This recursive definition proves equivalent to stating that a model is decomposable if and only if it has complete irreducible components. For Whittaker (1990, page 381) a random vector X is *reducible*, i.e., there exists a decomposition of X, if and only if there exists a partition of X into (X_A, X_B, X_C) such that $X_B \perp\!\!\!\perp X_C \mid X_A$ (and neither B nor C are empty) and the subgraph on X_A is complete. If such a decomposition exists, the *components* of X are $X_{AB} = (X_A, X_B)$ and $X_{AC} = (X_A, X_C)$. If such a decomposition does not exist the vector X is *irreducible*. Decomposable random vectors have independence graphs consisting entirely of complete subgraphs. Hence, the maximal irreducible components of a decomposable model are the cliques of the graph. Decomposable models are *multiplicative*: every density function in the model fully factorises into the product of marginal density functions. Besides, decomposable models have triangulated independence graphs, i.e., graphs with chordless cycles with no more than three vertices (which gives the possibility of an immediate visual check on the decomposability of a given graphical model) and closed-form maximum likelihood estimates. In brief: decomposable models are graphical models with triangulated graphs.

The purpose of the current section is not to describe in detail all possible methods that can be used to perform model selection but to justify backwards elimination as the obvious method to use, particularly when trying to detect the presence of a latent variable.

It is current practice in graphical modelling to start with the saturated model and

test for all possible single edge exclusions. Indeed Whittaker (1990, page 252) stated that *'backwards elimination methods starting from the maximal model directly test the conditional independencies'*, since the graphical model is a model for the joint distribution of the set of variables under study, simplified by conditional independence constraints. According to Edwards (2000) the usual argument for backwards elimination (versus forward selection) is that backwards methods start with a complex model (likely to be consistent with the data) and forward methods start with a very simple model (unlikely to be consistent with the data).

Calculating all possible edge exclusion tests (from the saturated model) and comparing the obtained test statistics with a chi-square distribution is a procedure asymptotically correct, requiring large samples. According to Porteous (1985) it may be quite poor for small sample sizes. In this case, if available, exact tests should be used, as suggested by Davison, Smith and Whittaker (1991).

In this thesis only backwards elimination is considered. The test statistics for single edge exclusion used are presented in detail in the next section.

2.9 Test Statistics for Single Edge Exclusion in GG and in GLL Models

Smith (1990) studied in detail edge exclusion tests, for conditional independence, in GG and in GLL models. The Wald and the efficient score tests for single edge exclusion where constructed and compared to the traditionally used likelihood ratio test statistic. Particular attention was devoted to the adequacy of the chi-square approximation to the distribution of each of this three test statistics, under the null hypothesis of conditional independence. This thesis builds upon Smith's (1990) work: the three test statistics for single edge exclusion are considered, the focus being on deriving an approximating distribution, under the alternative hypothesis that the saturated model holds (which is done in Chapter 3). Section 2.9.1 reviews the rationale for the derivation of the three tests, Section 2.9.2 provides an overview of these test statistics in the GG models framework, whereas Section 2.9.3 presents the three test statistics for GLL models.

2.9.1 The likelihood ratio, the Wald and the score tests

The likelihood ratio test is very often used in statistics, particularly when the null hypothesis is composite, i.e., when the parameter space constrained under the null hypothesis is more than a single point. Two other tests, also used for composite hypothesis are the Wald test and the Lagrange multiplier test, also known as the efficient score test. Buse (1982) gave a discussion of the geometry of the three tests and demonstrated that if the log-likelihood function is quadratic (which happens for normally distributed data) the three test statistics have chi-square distributions, for all sample sizes, under the null hypothesis. If the log-likelihood departures from the quadratic shape, the distributions of the three test statistics are asymptotically chi-square.

Suppose $\phi = \{\theta, \psi\}$ is the vector of unknown parameters, where θ is the vector of the r parameters of interest and ψ is the vector of nuisance parameters. Consider the null hypothesis $H_0 : \theta = \theta_0$, ψ unspecified and the alternative hypothesis $H_A : \theta \neq \theta_0$, ψ unspecified. The likelihood ratio test compares twice the difference between the maximum of the log-likelihood under H_A and H_0 to the critical value of a chi-square distribution on r degrees of freedom, i.e.,

$$LR = 2\left[l(\hat{\phi}) - l(\tilde{\phi})\right] = 2\left[l(\hat{\phi}) - l(\theta_0, \tilde{\psi})\right],$$

where $\hat{\phi}$ denotes the unrestricted m.l.e of ϕ and $\tilde{\psi}$ denotes the m.l.e. of ψ restricted by H_0. Indeed, by applying a Taylor series expansion to LR about the parameter point (θ_0, ψ), Cox and Hinkley (1974, page 323) showed that the limiting distribution of the likelihood ratio is chi-squared, with degrees of freedom equal to the dimension of θ, the distribution being central under H_0 (similar proof was given by Sen and Singer, 1993, pages 114–115). It was also shown that the Taylor series expansion of LR leads to two asymptotically equivalent test statistics: W, the Wald test statistic and S, the score test statistic. The Wald test statistic is given by

$$W = n\left(\hat{\theta} - \theta_0\right)^T \left[\hat{K}_{\theta\theta}\right]^{-1} \left(\hat{\theta} - \theta_0\right),$$

where $\hat{K}_{\theta\theta}$ is the variance matrix of the limiting normal distribution of $\hat{\theta}$, i.e, the submatrix of the inverse information matrix K corresponding to θ, evaluated at the m.l.e. Note that the asymptotic variance of $\hat{\phi}$ equals $\frac{1}{n}K$.

The efficient score test statistic is given by

$$S = n\,[U_\theta(\theta_0, \tilde{\psi})]^T\, \tilde{K}_{\theta\theta}\,[U_\theta(\theta_0, \tilde{\psi})],$$

where $\tilde{K}_{\theta\theta}$ is the inverse information matrix evaluated using $\tilde{\phi}$ (the restricted, under H_0, m.l.e. of ϕ) and $U_\theta(\theta_0, \tilde{\psi})$ is the *efficient score* (Cox and Hinkley, 1974, page 107) and equals the derivative of the log-likelihood function with respect to the elements of θ, here constrained under H_0. Recall that, in regular problems, the efficient score $U(\theta)$ has expectation zero and variance given by the Fisher information matrix. Also, in regular problems, there is a close connection between the maximum likelihood estimate and the efficient score $U(\theta)$, since $\hat{\theta}$, the m.l.e. of θ, satisfies $U(\hat{\theta}) = 0$ (Cox and Hinkley, 1974, page 280).

Because of the asymptotic equivalence of LR, W and S, and since LR is approximately central chi-square distributed under H_0, W and S also have approximate central chi-square distributions under the null, with degrees of freedom equal to the dimension of θ. When the parameter of interest θ is a scalar, the signed square-root of the likelihood ratio is given by

$$LR^{sign} = \text{sign}(\hat{\theta} - \theta_0)\,\sqrt{LR}.$$

LR^{sign} is asymptotically distributed according to a standard normal distribution (see for example Severini, 2000, pages 117 and 121). Similar reasoning applies to the Wald and the score tests.

Local alternatives can also be considered. Suppose the case of a simple null hypothesis $H_0 : \theta = \theta_0$ and alternatives that are local, i.e., sequences $\{\theta_n\}$ such that θ_n converges to $\theta_0 (\Leftrightarrow \theta_n = \theta_0 + \frac{\delta}{\sqrt{n}})$. Cox and Hinkley (1974, pages 317–318) showed that if $\sqrt{n}(\theta_n - \theta_0)$ converges to δ, then the likelihood ratio is approximately chi-squared, with degrees of freedom equal to the dimension of θ and non-centrality parameter $\varphi = \delta^T\,[K_{\theta_0\theta_0}]^{-1}\,\delta$. This result extends to the case of composite null hypothesis $H_0 : \theta = \theta_0, \psi$ unspecified. For local alternatives $\theta_n = \theta_0 + \frac{\delta_\theta}{\sqrt{n}}$ the non-centrality parameter is $\varphi = \delta_\theta^T\,[K_{\theta_0\theta_0}]^{-1}\,\delta_\theta$, where $[K_{\theta_0\theta_0}]^{-1}$ is the partition of the information matrix corresponding to the restriction in H_0. The number of degrees of freedom equals the dimension of θ. The same result applies to the Wald and score tests. For a detailed proof see also Severini (2000, pages 117–119).

2.9.2 Likelihood ratio, Wald and score test statistics for single edge exclusion in GG models

The likelihood ratio test statistic for single edge ij exclusion from the saturated GG model (see Whittaker, 1990, page 189) denoted in this thesis as T_{ij}^{L} is given by $T_{ij}^{L} = -n\log(1-\hat{\rho}_{ij.rest}^{2})$, where $\hat{\rho}_{ij.rest}$ equals the sample partial correlation coefficient between X_i and X_j given the remaining variables in the model. From Smith (1990, Sections 3.2.3 and 3.2.4), the Wald and the score test statistics for the exclusion of edge ij from the saturated model, here denoted as T_{ij}^{W} and T_{ij}^{S}, are given by $T_{ij}^{W} = n\frac{\hat{\rho}_{ij.rest}^{2}}{1+\hat{\rho}_{ij.rest}^{2}}$ and $T_{ij}^{S} = n\hat{\rho}_{ij.rest}^{2}$.

Smith and Whittaker (1998, Appendix D) suggested signed square-root versions of these three test statistics can be obtained by multiplying the sign of $\hat{\rho}_{ij.rest}$ by the square-root value of the test statistic. Table 2.1 summarises the formulae for the non-signed and for the signed square-root versions of the three test statistics, for single edge ij exclusion from the saturated model, that are used in this thesis: the likelihood ratio, the Wald and the score test statistics. They are presented both as a function of the sample partial correlation coefficients and of the elements of the inverse variance matrix.

Under the null hypothesis of conditional independence between X_i and X_j (i.e, $\rho_{ij.rest} = 0$) the non-signed versions of the three test statistics are chi-square distributed on one degree of freedom (see Smith, 1990) and the signed square-root versions of the three test statistics are normal distributed (Smith and Whittaker, 1998).

	non-signed version	signed square-root version
likelihood ratio	$T_{ij}^{L} =$	$T_{ij}^{signL} =$
test	$= -n\log(1-\frac{\hat{\omega}_{ij}^2}{\hat{\omega}_{ii}\hat{\omega}_{jj}})$	$= sign(\frac{-\hat{\omega}_{ij}}{\sqrt{\hat{\omega}_{ii}\hat{\omega}_{jj}}})\sqrt{-n\log(1-\frac{\hat{\omega}_{ij}^2}{\hat{\omega}_{ii}\hat{\omega}_{jj}})}$
	$= -n\log(1-\hat{\rho}_{ij.rest}^2))$	$= sign(\hat{\rho}_{ij.rest})\sqrt{-n\log(1-\hat{\rho}_{ij.rest}^2)}$
Wald test	$T_{ij}^{W} =$	$T_{ij}^{signW} =$
	$= n\frac{\frac{\hat{\omega}_{ij}^2}{\hat{\omega}_{ii}\hat{\omega}_{jj}}}{1+\frac{\hat{\omega}_{ij}^2}{\hat{\omega}_{ii}\hat{\omega}_{jj}}}$	$= \frac{-\hat{\omega}_{ij}}{\sqrt{\hat{\omega}_{ii}\hat{\omega}_{jj}}}\sqrt{\frac{n}{1+\frac{\hat{\omega}_{ij}^2}{\hat{\omega}_{ii}\hat{\omega}_{jj}}}}$
	$= n\frac{\hat{\rho}_{ij.rest}^2}{1+\hat{\rho}_{ij.rest}^2}$	$= \hat{\rho}_{ij.rest}\sqrt{\frac{n}{1+\hat{\rho}_{ij.rest}^2}}$
score test	$T_{ij}^{S} =$	$T_{ij}^{signS} =$
	$= n\frac{\hat{\omega}_{ij}^2}{\hat{\omega}_{ii}\hat{\omega}_{jj}}$	$= \frac{-\hat{\omega}_{ij}}{\sqrt{\hat{\omega}_{ii}\hat{\omega}_{jj}}}\sqrt{n}$
	$= n\hat{\rho}_{ij.rest}^2$	$= \hat{\rho}_{ij.rest}\sqrt{n}$

Table 2.1: Test statistics for single edge ij exclusion from the saturated GG model.

2.9.3 Likelihood ratio, Wald and score test statistics for single edge exclusion in GLL models

This section presents the three test statistics for single edge exclusion, within the GLL models framework, in a general way. In Section 3.4 they are specified for the two and for the three binary variables cases. The reason for doing so follows. In the case of a GG model, each test statistic for single edge exclusion, from the saturated model, depends on one parameter, the partial correlation coefficient between two variables given all the remaining variables, irrespective of the total number of variables in the model. However, in a GLL model, the test statistics for single edge exclusion, from the saturated model, are a function of different parameters (representing all higher order interaction terms), the number of parameters depending on the number of variables being considered. Hence, generalisations for the p variables case are very easy in GG

models, but can be very complicated in GLL models.

Smith (1990, Section 5.3) constructed the likelihood ratio, the Wald and the efficient score tests for the general null hypothesis that $H_0 : \lambda_A = 0$, and the alternative hypothesis that λ_A is unconstrained (where λ_A is a vector, of dimension r, containing the λ parameters of interest). Under the null hypothesis that $\lambda_A = 0$ all three test statistics are chi-square distributed on r degrees of freedom (the number of λ terms set to zero). Smith (1990, Section 5.5) presented the three test statistics for independence in a 2×2 contingency table, as a function of the observed cell counts. Here cell probabilities are used instead. The likelihood ratio test statistic for H_0 (independence), denoted in this thesis as LRT, equals

$$LRT = 2\, n_\emptyset \sum_x \hat{\pi}(x) \log\left(\frac{\hat{\pi}(x)}{\tilde{\pi}(x)}\right),$$

where $\hat{\pi}(x)$ are the unconstrained m.l.e. of the cell probabilities $\pi(x)$ and $\tilde{\pi}(x)$ are the m.l.e. of $\pi(x)$ constrained under H_0.

The Wald test statistic for H_0 equals

$$Wald = n_\emptyset\, (\hat{\lambda}_A)^T \left[\hat{K}_{AA}\right]^{-1} \hat{\lambda}_A,$$

where $\hat{K}_{AA}$ is the m.l.e. of the asymptotic variance matrix of $\hat{\lambda}_A$ based on a single observation. In the two variables case, when $r = 1$, the Wald test statistic simplifies to

$$Wald_{12} = \frac{n_\emptyset\, \hat{\lambda}_{12}^2}{\hat{K}[3,3]},$$

where $\hat{\lambda}_{12} = \log \hat{\psi}_{12}$ and $\hat{K}[3,3] = \hat{\text{var}}(\hat{\lambda}_{12}) = \frac{1}{\hat{\pi}(0,0)} + \frac{1}{\hat{\pi}(0,1)} + \frac{1}{\hat{\pi}(1,0)} + \frac{1}{\hat{\pi}(1,1)}$ (recall the inverse information matrix K is given by Equation 2.4).

The score test statistic equals

$$Score = n_\emptyset\, (\tilde{l}'_A)^T \left[\tilde{K}_{AA}\right] \tilde{l}'_A,$$

where $\tilde{l}'_A$ is the derivative of the log-likelihood function (of the log-linear model), with respect to λ_A, constrained under H_0, evaluated at the m.l.e.. In the two variables case, when $r = 1$, the score test statistic simplifies to

$$Score_{12} = n_\emptyset\, (\tilde{l}'_{12})^2\, \tilde{K}[3,3],$$

where $\tilde{l}'_{12} = \hat{\pi}(1,1) - \hat{\pi}_1(1)\, \hat{\pi}_2(1)$ and $\tilde{K}[3,3] = \frac{1}{\tilde{\pi}(0,0)} + \frac{1}{\tilde{\pi}(0,1)} + \frac{1}{\tilde{\pi}(1,0)} + \frac{1}{\tilde{\pi}(1,1)}$.

In conclusion: GG and GLL models have been presented. Considering the saturated model and calculating all possible single edge exclusion tests is claimed to be the usual procedure for starting model selection, particularly if the data analyst is interested in detecting the presence of a latent variable. The formulae for the three test statistics for single edge exclusion used in this dissertation have been summarised.

In Chapter 3 the distributions of these test statistics are investigated, in particular under the alternative hypothesis that the saturated model holds.

Chapter 3

Distributions of the Test Statistics for Single Edge Exclusion

The three test statistics for single edge exclusion used in this thesis (likelihood ratio, Wald and score test statistics) are summarised in Section 2.9. Under the null hypothesis that X_i is conditionally independent from X_j given the remaining variables in the model, i.e., the edge between X_i and X_j is absent from the independence graph of the variables, the three test statistics for single edge exclusion are asymptotic chi-square distributed. The number of degrees of freedom is given by the number of parameters in the model set to zero. In a graphical Gaussian model there is just one parameter associated with each edge exclusion: $\omega_{ij} = 0 \Leftrightarrow \rho_{ij.rest} = 0$. In a graphical log-linear model, for each edge exclusion from the saturated model, the number of λ terms set to zero depends on the number of variables in the model: when all variables are binary, if $p = 2$ there is just one two-way interaction term to be set to zero (λ_{12}), whereas if $p = 3$ two λ terms must be set to zero - the two-way and the three-way interaction terms λ_{ij} and λ_{ijk}. Thus, in GLL models the complexity increases considerably with the number of variables, which does not happen in GG models.

The aim of this chapter is to study the distributions of the three test statistics for single edge exclusion under the alternative hypothesis that the saturated model holds. Indeed, an approximation to the distributions of the test statistics for single edge exclusion, under the alternative hypothesis, is required for obtaining the asymptotic power functions derived in Chapter 4. Of particular use are the vectors of the means and the matrices of the variances and covariances of the test statistics in the asymptotic

normal distributions.

The current chapter has two main parts: the first part deals with deriving approximations to the distributions of the test statistics for single edge exclusion, from the saturated model, in the framework of GG models, whereas the second part is developed within the framework of GLL models.

The *delta-method* is used (Section 3.1.1) to obtain asymptotic normal approximations to the distributions of the test statistics: in the GG models framework results are derived for the general case of p variables, and presented as function of ω, the elements of the inverse variance matrix (Section 3.1.2), and as function of ρ, the elements of the scaled inverse variance matrix (Sections 3.1.3 and 3.1.4). Approximations to the distributions of the signed square-root versions of the test statistics are derived in Section 3.1.5. In the GLL models framework the two and the three variables cases are considered (Sections 3.4.1 to 3.4.3). Although the methodology used can be applied to contingency tables with higher dimensions, the number of parameters involved makes the calculations very messy, making it almost impossible to derive general 'simplified' formulae. Results are obtained for the non-signed (Sections 3.5.1 and 3.5.2) and for the signed square-root versions (Section 3.5.3) of the test statistics.

In the two variables case, the possibility of using a non-central chi-square approximation to the distributions of the test statistics is also analysed: in Section 3.2 for GG models and in Section 3.6 for GLL models. Both main parts of the chapter end with some guidelines as to when each of the approximating distributions performs better: simulation results are used to assess the quality of the proposed approximations, as the sample size varies (Sections 3.3 and 3.7).

3.1 Normal Approximations to the Distributions of the Test Statistics in GG Models

The test statistics for the exclusion of edge ij from the saturated model, and corresponding signed square-root versions, in the GG models framework, are summarised in Section 2.9.2, Table 2.1. The aim now is to obtain approximations to the distributions of these test statistics, under the alternative hypothesis that the saturated model holds. In Section 2.7.1, following Smith (1990, page 21) the asymptotic variance matrix

of the m.l.e. of ω was presented. Because the test statistics are functions of the $\hat{\omega}$, the *delta-method* can be applied to Smith's result, in order to derive asymptotic normal approximations to the distributions of the test statistics. A brief explanation of the method is given in Section 3.1.1. For further details see, for example, Bishop, Fienberg and Holland (1975, page 493) and Sen and Singer (1993, pages 131–137).

3.1.1 Using the delta-method to obtain asymptotic normal approximations

Let $\hat{\theta}$, a random column vector, be the m.l.e of θ based on n observations. It is a well known result that, under certain regularity conditions, $\hat{\theta}$ has an asymptotic normal distribution with mean θ and variance given by the inverse of the information matrix (see Cox and Hinkley, 1974, page 294). In other words,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} N(0, K),$$

where $K = \mathcal{I}^{-1}$ is the inverse information matrix based on a single observation. One should note that $\text{var}(\hat{\theta}) = \frac{1}{n}K$. If $f(\theta)$ is differentiable at $\hat{\theta}$, then, using the delta-method, the approximating distribution to $f(\hat{\theta})$ is the normal distribution, with mean $f(\theta)$ and variance matrix given by $(\frac{\partial f}{\partial \theta})^T \text{var}(\theta) (\frac{\partial f}{\partial \theta})$, i.e.,

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

In the case under study, let V^{-1} denote the unconstrained m.l.e. of Ω, i.e. the sample inverse variance matrix with divisor n, a symmetric matrix with elements denoted by $\hat{\omega}_{ij}$. The vector of the distinct elements of V^{-1} is denoted by $\text{vec}(V^{-1})$. It includes first the off-diagonal elements $\hat{\omega}_{ij}$ and finally the diagonal elements $\hat{\omega}_{ii}$. For example, for the three variables case, $\hat{\theta} = \text{vec}(V^{-1}) = (\hat{\omega}_{12}\ \hat{\omega}_{13}\ \hat{\omega}_{23}\ \hat{\omega}_{11}\ \hat{\omega}_{22}\ \hat{\omega}_{33})^T$ and $\theta = \text{vec}(\Omega) = (\omega_{12}\ \omega_{13}\ \omega_{23}\ \omega_{11}\ \omega_{22}\ \omega_{33})^T$. The function f depends on the test statistic being used. Suppose the test statistic is the likelihood ratio test: $f_{ij}^L(\hat{\theta}) = \frac{1}{n} T_{ij}^L$. Consequently, $f_{ij}^L(\theta) = -\log(1 - \frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}})$, since $T_{ij}^L = -n\log(1 - \frac{\hat{\omega}_{ij}^2}{\hat{\omega}_{ii}\hat{\omega}_{jj}})$. Provided all elements of the $\hat{\theta}$ vector are different from zero, f_{ij}^L is differentiable at $\hat{\theta}$. Then,

$$\sqrt{n}\left[f\left(\text{vec}(V^{-1})\right) - f\left(\text{vec}(\Omega)\right)\right] \xrightarrow{\mathcal{D}} N\left(0, \left[\left(\frac{\partial f(\text{vec}(\Omega))}{\partial \text{vec}(\Omega)}\right)^T K \left(\frac{\partial f(\text{vec}(\Omega))}{\partial \text{vec}(\Omega)}\right)\right]\right),$$

i.e. $f\left(\text{vec}(V^{-1})\right)$ is asymptotically normal distributed.

Section 3.1.2 shows how this formula is applied, both in the case of two and in the case of four variables. In the remainder of the chapter $\text{var}(T_{ij}^{\bullet})$ and $\text{cov}(T_{ij}^{\bullet}, T_{kl}^{\bullet})$ denote, respectively, the variance and the covariance of the test statistics in the asymptotic distribution, where, for convenience of notation, $T_{ij}^{\bullet} = n\, f_{ij}^{\bullet}(\hat{\theta})$ and $T_{ij}^{sign\,\bullet} = \sqrt{n}\, f_{ij}^{sign\,\bullet}(\hat{\theta})$.

3.1.2 Asymptotic distribution of the LRT as a function of ω

In this section formulae for the variance and covariance of the likelihood ratio test statistic (for single edge exclusion, from the saturated model) in the asymptotic distribution are derived. First, the two variables case is presented in detail; next, the four variables case is considered. Results are then generalised to the situation of p variables.

The two variables case

Let us consider the two variables case and derive the mean $AE[T_{12}^L]$ and the variance $\text{var}(T_{12}^L)$ of the test statistic T_{12}^L, in the asymptotic distribution.

Using the delta-method, and considering $\hat{\theta} = \text{vec}(V^{-1}) = (\hat{\omega}_{12}\ \hat{\omega}_{11}\ \hat{\omega}_{22})^T$, $\theta = \text{vec}(\Omega) = (\omega_{12}\ \omega_{11}\ \omega_{22})^T$, $f_{12}^L(\theta) = -\log(1 - \frac{\omega_{12}^2}{\omega_{11}\omega_{22}})$ and $f_{12}^L(\hat{\theta}) = -\log(1 - \frac{\hat{\omega}_{12}^2}{\hat{\omega}_{11}\hat{\omega}_{22}})$, the mean of T_{12}^L in the asymptotic distribution is given by

$$AE[T_{12}^L] = n\, f_{12}^L(\theta) = -n \log(1 - \frac{\omega_{12}^2}{\omega_{11}\omega_{22}}).$$

The variance in the asymptotic distribution equals

$$\text{var}(T_{12}^L) = n\, \Delta^T\, K\, \Delta,$$

where

$$\Delta = \begin{bmatrix} \partial f_{12}^L/\partial\omega_{12} \\ \partial f_{12}^L/\partial\omega_{11} \\ \partial f_{12}^L/\partial\omega_{22} \end{bmatrix} = \begin{bmatrix} \frac{2\omega_{12}}{\omega_{11}\omega_{22}-\omega_{12}^2} \\ \frac{-\omega_{12}^2}{\omega_{11}(\omega_{11}\omega_{22}-\omega_{12}^2)} \\ \frac{-\omega_{12}^2}{\omega_{22}(\omega_{11}\omega_{22}-\omega_{12}^2)} \end{bmatrix},$$

and K is a 3×3 matrix, given by Equation 2.2. Substituting in the above equation gives the simplified result

$$\text{var}(T_{12}^L) = \frac{4n\omega_{12}^2}{\omega_{11}\omega_{22}},$$

which can be written as a function of the (partial) correlation coefficient as $\text{var}(T_{12}^L) = 4n\rho_{12}^2$. The n in the numerator implies that, as the sample size increases, the variance of the test statistic increases. One should note that the n also appears in the expected value of the test statistic.

Analogous reasoning can be followed for the three and four variables cases. Calculations were done separately for each of this cases. However, only the results for four variables are presented since the three variables case (as well as the two variables case) is a particular case of the four variables situation.

The four variables case

There are now six T_{ij}^L (with $i < j$, from 1 to 4) test statistics, for single edge exclusion from the saturated model, to be considered.

The vector of means of the six test statistics is given by

$$\begin{bmatrix} AE[T_{12}^L] \\ AE[T_{13}^L] \\ AE[T_{14}^L] \\ AE[T_{23}^L] \\ AE[T_{24}^L] \\ AE[T_{34}^L] \end{bmatrix} = \begin{bmatrix} -n\log(1-\frac{\omega_{12}^2}{\omega_{11}\omega_{22}}) \\ -n\log(1-\frac{\omega_{13}^2}{\omega_{11}\omega_{33}}) \\ -n\log(1-\frac{\omega_{14}^2}{\omega_{11}\omega_{44}}) \\ -n\log(1-\frac{\omega_{23}^2}{\omega_{22}\omega_{33}}) \\ -n\log(1-\frac{\omega_{24}^2}{\omega_{22}\omega_{44}}) \\ -n\log(1-\frac{\omega_{34}^2}{\omega_{33}\omega_{44}}) \end{bmatrix}.$$

The variance matrix equals $n\,\Delta^T\,K\,\Delta$. The 10×10 matrix $K = n\,\mathrm{var}[\mathrm{vec}(V^{-1})]$ is calculated, as in the two variables case, using Equation 2.1. The 10×6 matrix Δ has the derivatives of the six $f^L(\theta)$ (f_{12}^L, f_{13}^L, f_{14}^L, f_{23}^L, f_{24}^L, f_{34}^L, in columns) with respect to the ten distinct ω (ω_{12}, ω_{13}, ω_{14}, ω_{23}, ω_{24}, ω_{34}, ω_{11}, ω_{22}, ω_{33}, ω_{44}, in rows). For example, column three equals $(0\ 0\ \partial f_{14}^L/\partial\omega_{14}\ 0\ 0\ 0\ \partial f_{14}^L/\partial\omega_{11}\ 0\ 0\ \partial f_{14}^L/\partial\omega_{44})^T$. Performing the substitutions induces a 6×6 variance matrix with the variances of the test statistics, in the asymptotic distribution, on the main diagonal, the off-diagonal elements being the covariances between the test statistics, in the asymptotic distribution, $(T_{ij}^L,\ T_{kl}^L)$, with $i < j$ and $k < l$, as follows

$$\mathrm{var}(T_{ij}^L,\ T_{kl}^L) = \begin{bmatrix} \mathrm{var}(T_{12}^L) & & & & & \\ \mathrm{cov}(T_{12}^L, T_{13}^L) & \mathrm{var}(T_{13}^L) & & & & \\ \mathrm{cov}(T_{12}^L, T_{14}^L) & \mathrm{cov}(T_{13}^L, T_{14}^L) & \mathrm{var}(T_{14}^L) & & & \\ \mathrm{cov}(T_{12}^L, T_{23}^L) & \mathrm{cov}(T_{13}^L, T_{23}^L) & \mathrm{cov}(T_{14}^L, T_{23}^L) & \mathrm{var}(T_{23}^L) & & \\ \mathrm{cov}(T_{12}^L, T_{24}^L) & \mathrm{cov}(T_{13}^L, T_{24}^L) & \mathrm{cov}(T_{14}^L, T_{24}^L) & \mathrm{cov}(T_{23}^L, T_{24}^L) & \mathrm{var}(T_{24}^L) & \\ \mathrm{cov}(T_{12}^L, T_{34}^L) & \mathrm{cov}(T_{13}^L, T_{34}^L) & \mathrm{cov}(T_{14}^L, T_{34}^L) & \mathrm{cov}(T_{23}^L, T_{34}^L) & \mathrm{cov}(T_{24}^L, T_{34}^L) & \mathrm{var}(T_{34}^L) \end{bmatrix},$$

where

$$\mathrm{var}(T_{12}^L) = \frac{4n\omega_{12}^2}{\omega_{11}\omega_{22}},\ \mathrm{var}(T_{13}^L) = \frac{4n\omega_{13}^2}{\omega_{11}\omega_{33}},\ \mathrm{var}(T_{14}^L) = \frac{4n\omega_{14}^2}{\omega_{11}\omega_{44}},$$

$$\mathrm{var}(T_{23}^L) = \frac{4n\omega_{23}^2}{\omega_{22}\omega_{33}},\ \mathrm{var}(T_{24}^L) = \frac{4n\omega_{24}^2}{\omega_{22}\omega_{44}},\ \mathrm{var}(T_{34}^L) = \frac{4n\omega_{34}^2}{\omega_{33}\omega_{44}},$$

and, for example,

- $\mathrm{cov}(T_{12}^L,\, T_{13}^L) = \frac{n}{\omega_{11}^2\omega_{22}\omega_{33}(\omega_{11}\omega_{22}-\omega_{12}^2)(\omega_{11}\omega_{33}-\omega_{13}^2)}\,[4\,\omega_{11}^3\omega_{22}\omega_{33}\omega_{12}\omega_{13}\omega_{23} - 4\,\omega_{11}^2\omega_{33}\omega_{12}^3\omega_{13}\omega_{23} - 2\,\omega_{11}^2\omega_{22}\omega_{33}\omega_{12}^2\omega_{13}^2 - 4\,\omega_{11}^2\omega_{22}\omega_{12}\omega_{13}^3\omega_{23} + 2\,\omega_{11}^2\omega_{12}^2\omega_{23}^2\omega_{13}^2 + 2\,\omega_{11}\omega_{33}\omega_{12}^4\omega_{13}^2 + 2\,\omega_{11}\omega_{22}\omega_{12}^2\omega_{13}^4];$

- $\mathrm{cov}(T_{12}^L,\, T_{24}^L) = \frac{n}{\omega_{11}\omega_{22}^2\omega_{44}(\omega_{11}\omega_{22}-\omega_{12}^2)(\omega_{22}\omega_{44}-\omega_{24}^2)}\,[4\,\omega_{11}\omega_{22}^3\omega_{44}\omega_{12}\omega_{14}\omega_{24} - 4\,\omega_{22}^2\omega_{44}\omega_{12}^3\omega_{14}\omega_{24} - 2\,\omega_{11}\omega_{22}^2\omega_{44}\omega_{12}^2\omega_{24}^2 - 4\,\omega_{11}\omega_{22}^2\omega_{12}\omega_{14}\omega_{24}^3 + 2\,\omega_{11}\omega_{22}\omega_{12}^2\omega_{24}^4 + 2\,\omega_{22}^2\omega_{12}^2\omega_{14}^2\omega_{24}^2 + 2\,\omega_{22}\omega_{44}\omega_{12}^4\omega_{24}^2];$

- $\mathrm{cov}(T_{13}^L,\, T_{23}^L) = \frac{n}{\omega_{11}\omega_{22}\omega_{33}^2(\omega_{11}\omega_{33}-\omega_{13}^2)(\omega_{22}\omega_{33}-\omega_{23}^2)}\,[4\,\omega_{11}\omega_{33}^3\omega_{22}\omega_{13}\omega_{12}\omega_{23} - 4\,\omega_{33}^2\omega_{22}\omega_{13}^3\omega_{12}\omega_{23} - 4\,\omega_{11}\omega_{33}^2\omega_{13}\omega_{12}\omega_{23}^3 - 2\,\omega_{11}\omega_{33}^2\omega_{22}\omega_{13}^2\omega_{23}^2 + 2\,\omega_{11}\omega_{33}\omega_{13}^2\omega_{23}^4 + 2\,\omega_{33}\omega_{22}\omega_{13}^4\omega_{23}^2 + 2\,\omega_{33}^2\omega_{13}^2\omega_{12}^2\omega_{23}^2];$

- $\mathrm{cov}(T_{12}^L,\, T_{34}^L) = \frac{n}{\omega_{11}\omega_{22}\omega_{33}\omega_{44}(\omega_{11}\omega_{22}-\omega_{12}^2)(\omega_{33}\omega_{44}-\omega_{34}^2)}\,[4\,\omega_{11}\omega_{22}\omega_{33}\omega_{44}\omega_{12}\omega_{13}\omega_{24}\omega_{34} + 4\,\omega_{11}\omega_{22}\omega_{33}\omega_{44}\omega_{12}\omega_{14}\omega_{23}\omega_{34} - 4\,\omega_{22}\omega_{33}\omega_{44}\omega_{12}^2\omega_{13}\omega_{14}\omega_{34} - 4\,\omega_{11}\omega_{33}\omega_{44}\omega_{12}^2\omega_{23}\omega_{24}\omega_{34} - 4\,\omega_{11}\omega_{22}\omega_{44}\omega_{12}\omega_{13}\omega_{23}\omega_{34}^2 - 4\,\omega_{11}\omega_{22}\omega_{33}\omega_{12}\omega_{14}\omega_{24}\omega_{34}^2 + 2\,\omega_{11}\omega_{33}\omega_{12}^2\omega_{24}^2\omega_{34}^2 + 2\,\omega_{11}\omega_{44}\omega_{12}^2\omega_{23}^2\omega_{34}^2 + 2\,\omega_{22}\omega_{33}\omega_{12}^2\omega_{14}^2\omega_{34}^2 + 2\,\omega_{22}\omega_{44}\omega_{12}^2\omega_{13}^2\omega_{34}^2];$

- $\mathrm{cov}(T_{13}^L,\, T_{24}^L) = \frac{n}{\omega_{11}\omega_{33}\omega_{22}\omega_{44}(\omega_{11}\omega_{33}-\omega_{13}^2)(\omega_{22}\omega_{44}-\omega_{24}^2)}\,[4\,\omega_{11}\omega_{33}\omega_{22}\omega_{44}\omega_{13}\omega_{12}\omega_{34}\omega_{24} + 4\,\omega_{11}\omega_{33}\omega_{22}\omega_{44}\omega_{13}\omega_{14}\omega_{23}\omega_{24} - 4\,\omega_{33}\omega_{22}\omega_{44}\omega_{13}^2\omega_{12}\omega_{14}\omega_{24} - 4\,\omega_{11}\omega_{22}\omega_{44}\omega_{13}^2\omega_{23}\omega_{34}\omega_{24} - 4\,\omega_{11}\omega_{33}\omega_{44}\omega_{13}\omega_{12}\omega_{23}\omega_{24}^2 - 4\,\omega_{11}\omega_{33}\omega_{22}\omega_{13}\omega_{14}\omega_{34}\omega_{24}^2 + 2\,\omega_{11}\omega_{22}\omega_{13}^2\omega_{34}^2\omega_{24}^2 + 2\,\omega_{11}\omega_{44}\omega_{13}^2\omega_{23}^2\omega_{24}^2 + 2\,\omega_{33}\omega_{22}\omega_{13}^2\omega_{14}^2\omega_{24}^2 + 2\,\omega_{33}\omega_{44}\omega_{13}^2\omega_{12}^2\omega_{24}^2].$

In brief: a general term for the mean of the test statistic T_{ij}^L $(i < j)$ in the asymptotic distribution can be written as

$$AE[T_{ij}^L] = -n\log(1 - \frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}}), \tag{3.1}$$

a general term for the variance of T_{ij}^L in the asymptotic distribution can be written as

$$\mathrm{var}(T_{ij}^L) = \frac{4n\omega_{ij}^2}{\omega_{ii}\omega_{jj}} \tag{3.2}$$

and a general term for the covariance of the test statistics T_{ij}^L, T_{kl}^L $(i < j,\, k < l)$, in the asymptotic distribution, can be written as

$$
\begin{aligned}
\operatorname{cov}(T_{ij}^L,\, T_{kl}^L) = \; & \frac{n}{\omega_{ii}\omega_{jj}\omega_{kk}\omega_{ll}(\omega_{ii}\omega_{jj}-\omega_{ij}^2)(\omega_{kk}\omega_{ll}-\omega_{kl}^2)} \\
& [4\,\omega_{ii}\omega_{jj}\omega_{kk}\omega_{ll}\omega_{ij}\omega_{ik}\omega_{jl}\omega_{kl} + 4\,\omega_{ii}\omega_{jj}\omega_{kk}\omega_{ll}\omega_{ij}\omega_{il}\omega_{jk}\omega_{kl} \\
& -4\,\omega_{jj}\omega_{kk}\omega_{ll}\omega_{ij}^2\omega_{ik}\omega_{il}\omega_{kl} - 4\,\omega_{ii}\omega_{kk}\omega_{ll}\omega_{ij}^2\omega_{jk}\omega_{jl}\omega_{kl} \\
& -4\,\omega_{ii}\omega_{jj}\omega_{ll}\omega_{ij}\omega_{ik}\omega_{jk}\omega_{kl}^2 - 4\,\omega_{ii}\omega_{jj}\omega_{kk}\omega_{ij}\omega_{il}\omega_{jl}\omega_{kl}^2 \\
& +2\,\omega_{ii}\omega_{kk}\omega_{ij}^2\omega_{jl}^2\omega_{kl}^2 + 2\,\omega_{ii}\omega_{ll}\omega_{ij}^2\omega_{jk}^2\omega_{kl}^2 \\
& +2\,\omega_{jj}\omega_{kk}\omega_{ij}^2\omega_{il}^2\omega_{kl}^2 + 2\,\omega_{jj}\omega_{ll}\omega_{ij}^2\omega_{ik}^2\omega_{kl}^2].
\end{aligned}
\tag{3.3}
$$

For simplicity of notation, in the remainder of this chapter this Equation 3.3 is also going to be written as

$$
\operatorname{cov}(T_{ij}^L,\, T_{kl}^L) = \frac{n}{\omega_{ii}\omega_{jj}\omega_{kk}\omega_{ll}(\omega_{ii}\omega_{jj}-\omega_{ij}^2)(\omega_{kk}\omega_{ll}-\omega_{kl}^2)}\, C_\omega.
$$

Equation 3.3 holds for the cases of i, j, k, l from 1 to 4, (both neighbour and non-neighbour vertices), but also, as the next section will prove, for any case and any number of variables ($i < j$ and $k < l$, from 1 to p). In practice, the mathematical package MAPLE was used to perform the calculations for the five variables case, and results show Equation 3.3 holds. Neighbour vertices are a particular case of the general formula, that is, situations when $i = k$, or $i = l$, or $j = k$, or $j = l$. For example, supposing $i = k$. In order to obtain $\operatorname{cov}(T_{12}^L,\, T_{13}^L)$, it is only required to start with Equation 3.3, replace i by 1, j by 2, k by 1 and l by 3, and simplify the results.

The p variables case

The proof that Equation 3.3 is a general formula that holds for any number of variables and both for neighbour and non-neighbour vertices follows. Expressing the delta method in coordinate form gives, in general,

$$
\begin{aligned}
\operatorname{cov}(T_{ij}^L,\, T_{kl}^L) \; & = \sum_{p,q,r,s} \left\{ n\, \frac{\partial f_{ij}^L}{\partial \omega_{pq}}\, \frac{\partial f_{kl}^L}{\partial \omega_{rs}}\, K_{pqrs} \right\} \\
& = \sum_{p,q,r,s} \left\{ n\, \frac{\partial f_{ij}^L}{\partial \omega_{pq}}\, \frac{\partial f_{kl}^L}{\partial \omega_{rs}}\, (\omega_{pr}\omega_{qs} + \omega_{ps}\omega_{qr}) \right\},
\end{aligned}
\tag{3.4}
$$

where K_{pqrs} is the inverse of the information matrix based on a single observation. The proof that starting with this general formula and performing all the calculations induces Equation 3.3 follows. Recall that $f_{ij}^L = -\log(1 - \frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}})$, and that $i < j$ and

$k < l$. The derivative of f_{ij}^L with respect to a generic element ω_{pq} takes the form

$$\begin{aligned}\frac{\partial f_{ij}^L}{\partial \omega_{pq}} &= \frac{-1}{1-\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}}} \left[\delta_{ip}\delta_{jq}\left(\frac{-2\omega_{ij}}{\omega_{ii}\omega_{jj}}\right) + \delta_{ip}\delta_{iq}\left(\frac{\omega_{ij}^2}{\omega_{ii}^2\omega_{jj}}\right) + \delta_{jp}\delta_{jq}\left(\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}^2}\right)\right] \\ &= \frac{-1}{(\omega_{ii}\omega_{jj})\,(\omega_{ii}\omega_{jj}-\omega_{ij}^2)} \left[\delta_{ip}\delta_{iq}\,\omega_{ij}^2\omega_{jj} + \delta_{jp}\delta_{jq}\,\omega_{ij}^2\omega_{ii} - 2\delta_{ip}\delta_{jq}\,\omega_{ii}\omega_{jj}\omega_{ij}\right].\end{aligned} \tag{3.5}$$

Analogously,

$$\frac{\partial f_{kl}^L}{\partial \omega_{rs}} = \frac{-1}{(\omega_{kk}\omega_{ll})\,(\omega_{kk}\omega_{ll}-\omega_{kl}^2)} \left[\delta_{kr}\delta_{ks}\,\omega_{kl}^2\omega_{ll} + \delta_{lr}\delta_{ls}\,\omega_{kl}^2\omega_{kk} - 2\delta_{kr}\delta_{ls}\,\omega_{kk}\omega_{ll}\omega_{kl}\right]. \tag{3.6}$$

Substituting Equations 3.5 and 3.6 in Equation 3.4 gives

$$\begin{aligned}\mathrm{cov}(T_{ij}^L,\, T_{kl}^L) &= \sum_{p,q,r,s}\left\{n\,\frac{\partial f_{ij}^L}{\partial \omega_{pq}}\,\frac{\partial f_{kl}^L}{\partial \omega_{rs}}\,K_{pqrs}\right\} \\ &= n\,\frac{-1}{\omega_{ii}\omega_{jj}(\omega_{ii}\omega_{jj}-\omega_{ij}^2)}\,\frac{-1}{\omega_{kk}\omega_{ll}(\omega_{kk}\omega_{ll}-\omega_{kl}^2)} \\ &\quad \times \sum_{p,q,r,s}\{[\delta_{ip}\delta_{iq}\,\omega_{ij}^2\omega_{jj} + \delta_{jp}\delta_{jq}\,\omega_{ij}^2\omega_{ii} - 2\delta_{ip}\delta_{jq}\,\omega_{ii}\omega_{jj}\omega_{ij}] \\ &\quad \times [\delta_{kr}\delta_{ks}\,\omega_{kl}^2\omega_{ll} + \delta_{lr}\delta_{ls}\,\omega_{kl}^2\omega_{kk} - 2\delta_{kr}\delta_{ls}\,\omega_{kk}\omega_{ll}\omega_{kl}] \times K_{pqrs}\} \\ &= \frac{n}{\omega_{ii}\omega_{jj}\omega_{kk}\omega_{ll}(\omega_{ii}\omega_{jj}-\omega_{ij}^2)(\omega_{kk}\omega_{ll}-\omega_{kl}^2)} \times \sum_{p,q,r,s}\{\ldots\},\end{aligned}$$

where, for the result to hold, in the last expression the summation between curly brackets has to equal the expression previously denoted by C_ω. The proof continues. Using Equation 2.1,

$$\begin{aligned}\sum_{p,q,r,s}\{\ldots\} = \sum_{p,q,r,s}\{\;&\delta_{ip}\delta_{iq}\delta_{kr}\delta_{ks}\omega_{ij}^2\omega_{jj}\omega_{kl}^2\omega_{ll}\,[\omega_{pr}\omega_{qs}+\omega_{ps}\omega_{qr}] \\ &+\delta_{ip}\delta_{iq}\delta_{lr}\delta_{ls}\omega_{ij}^2\omega_{jj}\omega_{kl}^2\omega_{kk}\,[\omega_{pr}\omega_{qs}+\omega_{ps}\omega_{qr}] \\ &-2\delta_{ip}\delta_{iq}\delta_{kr}\delta_{ls}\omega_{ij}^2\omega_{jj}\omega_{kk}\omega_{ll}\omega_{kl}\,[\omega_{pr}\omega_{qs}+\omega_{ps}\omega_{qr}] \\ &+\delta_{jp}\delta_{jq}\delta_{kr}\delta_{ks}\omega_{ij}^2\omega_{ii}\omega_{kl}^2\omega_{ll}\,[\omega_{pr}\omega_{qs}+\omega_{ps}\omega_{qr}] \\ &+\delta_{jp}\delta_{jq}\delta_{lr}\delta_{ls}\omega_{ij}^2\omega_{ii}\omega_{kl}^2\omega_{kk}\,[\omega_{pr}\omega_{qs}+\omega_{ps}\omega_{qr}] \\ &-2\delta_{jp}\delta_{jq}\delta_{kr}\delta_{ls}\omega_{ij}^2\omega_{ii}\omega_{kk}\omega_{ll}\omega_{kl}\,[\omega_{pr}\omega_{qs}+\omega_{ps}\omega_{qr}] \\ &-2\delta_{ip}\delta_{jq}\delta_{kr}\delta_{ks}\omega_{ii}\omega_{jj}\omega_{ij}\omega_{kl}^2\omega_{ll}\,[\omega_{pr}\omega_{qs}+\omega_{ps}\omega_{qr}] \\ &-2\delta_{ip}\delta_{jq}\delta_{lr}\delta_{ls}\omega_{ii}\omega_{jj}\omega_{ij}\omega_{kl}^2\omega_{kk}\,[\omega_{pr}\omega_{qs}+\omega_{ps}\omega_{qr}] \\ &+4\delta_{ip}\delta_{jq}\delta_{kr}\delta_{ls}\omega_{ii}\omega_{jj}\omega_{ij}\omega_{kk}\omega_{ll}\omega_{kl}\,[\omega_{pr}\omega_{qs}+\omega_{ps}\omega_{qr}]\;\}\end{aligned}$$

$$\begin{aligned}= \omega_{ij}^2\omega_{jj}\omega_{kl}^2\omega_{ll}\,[\omega_{ik}\omega_{ik}+\omega_{ik}\omega_{ik}] + \omega_{ij}^2\omega_{jj}\omega_{kl}^2\omega_{kk}\,[\omega_{il}\omega_{il}+\omega_{il}\omega_{il}] \\ -2\omega_{ij}^2\omega_{jj}\omega_{kk}\omega_{ll}\omega_{kl}\,[\omega_{ik}\omega_{il}+\omega_{il}\omega_{ik}] + \omega_{ij}^2\omega_{ii}\omega_{kl}^2\omega_{ll}\,[\omega_{jk}\omega_{jk}+\omega_{jk}\omega_{jk}] \\ +\omega_{ij}^2\omega_{ii}\omega_{kl}^2\omega_{kk}\,[\omega_{jl}\omega_{jl}+\omega_{jl}\omega_{jl}] - 2\omega_{ij}^2\omega_{ii}\omega_{kk}\omega_{ll}\omega_{kl}\,[\omega_{jk}\omega_{jl}+\omega_{jl}\omega_{jk}] \\ -2\omega_{ii}\omega_{jj}\omega_{ij}\omega_{kl}^2\omega_{ll}\,[\omega_{ik}\omega_{jk}+\omega_{ik}\omega_{jk}] - 2\omega_{ii}\omega_{jj}\omega_{ij}\omega_{kl}^2\omega_{kk}\,[\omega_{il}\omega_{jl}+\omega_{il}\omega_{jl}] \\ +4\omega_{ii}\omega_{jj}\omega_{ij}\omega_{kk}\omega_{ll}\omega_{kl}\,[\omega_{ik}\omega_{jl}+\omega_{il}\omega_{jk}]\end{aligned}$$

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

which equals the expression C_ω in Equation 3.3, hence the proof is complete.

In conclusion: Equation 3.4 presents a general formula for the covariance of the test statistics T_{ij}^L, T_{kl}^L $(i < j,\ k < l)$ in the asymptotic distribution, when the likelihood ratio test statistic for single edge exclusion, from the saturated model, is used.

3.1.3 Asymptotic distribution of the LRT as a function of ρ

In Section 3.1.2 a normal approximation to the distribution of the likelihood ratio test statistic is derived, with the means, variances and covariances given as a function of ω, the elements of the inverse variance matrix, by Equations 3.1, 3.2 and 3.3. The purpose of this section is to express those means, variances and covariances as a function of ρ, the partial correlation coefficients.

The mean of the test statistic T_{ij}^L, in the asymptotic distribution, is given by Equation 3.1 as $AE[T_{ij}^L] = -n\log(1 - \frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}})$. Because $\rho_{ij.rest} = \frac{-\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}$, this mean can be written as a function of the partial correlation coefficient, as

$$AE[T_{ij}^L] = -n\log(1 - \rho_{ij.rest}^2). \tag{3.7}$$

Analogously, the variance in the asymptotic distribution can be written as a function of the partial correlation coefficient as

$$\mathrm{var}(T_{ij}^L) = 4n\rho_{ij.rest}^2. \tag{3.8}$$

A general formula for the covariances of the test statistics T_{ij}^L, T_{kl}^L in the asymptotic distribution is given by Equation 3.3. An easy way to obtain this formula as a function of the partial correlation coefficients is, for example, to set $\omega_{ii} = \omega_{jj} = \omega_{kk} = \omega_{ll} = 1$ and all six elements of the type ω_{pq} equal to $-\rho_{pq.rest}$. One should note that there are different combinations of values of the ω that will lead to the same set of values of ρ, and one of them was chosen. Alternatively, it can easily be proved that, starting with Equation 3.3 and replacing all elements of the type $\frac{-\omega_{pq}}{\sqrt{\omega_{pp}\omega_{qq}}}$ by $\rho_{pq.rest}$ gives Equation 3.9,

which is a general formula for the covariances of the test statistics, in the asymptotic distribution, written as a function of the partial correlation coefficients:

$$\begin{aligned}\mathrm{cov}(T_{ij}^L,\, T_{kl}^L) = \;& \frac{n}{(1-\rho_{ij.rest}^2)(1-\rho_{kl.rest}^2)} \\ & [4\,\rho_{ij.rest}\rho_{ik.rest}\rho_{jl.rest}\rho_{kl.rest} + 4\,\rho_{ij.rest}\rho_{il.rest}\rho_{jk.rest}\rho_{kl.rest} \\ & +4\,\rho_{ij.rest}^2\rho_{ik.rest}\rho_{il.rest}\rho_{kl.rest} + 4\,\rho_{ij.rest}^2\rho_{jk.rest}\rho_{jl.rest}\rho_{kl.rest} \\ & +4\,\rho_{ij.rest}\rho_{ik.rest}\rho_{jk.rest}\rho_{kl.rest}^2 + 4\,\rho_{ij.rest}\rho_{il.rest}\rho_{jl.rest}\rho_{kl.rest}^2 \\ & +2\,\rho_{ij.rest}^2\rho_{jl.rest}^2\rho_{kl.rest}^2 + 2\,\rho_{ij.rest}^2\rho_{jk.rest}^2\rho_{kl.rest}^2 \\ & +2\,\rho_{ij.rest}^2\rho_{il.rest}^2\rho_{kl.rest}^2 + 2\,\rho_{ij.rest}^2\rho_{kl.rest}^2\rho_{ik.rest}^2].\end{aligned} \tag{3.9}$$

For simplicity of notation, in the remainder of this chapter this Equation 3.9 is also going to be written as

$$\mathrm{cov}(T_{ij}^L,\, T_{kl}^L) = \frac{n}{(1-\rho_{ij.rest}^2)(1-\rho_{kl.rest}^2)}\, C_\rho.$$

The proof that Equation 3.3 implies Equation 3.9 follows. Equation 3.3 can be written as

$$\mathrm{cov}(T_{ij}^L,\, T_{kl}^L) = \frac{n}{\omega_{ii}\omega_{jj}\omega_{kk}\omega_{ll}\left[\omega_{ii}\omega_{jj}(1-\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}})\omega_{kk}\omega_{ll}(1-\frac{\omega_{kl}^2}{\omega_{kk}\omega_{ll}})\right]}\, C_\omega =$$

$$= \frac{n}{(1-\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}})(1-\frac{\omega_{kl}^2}{\omega_{kk}\omega_{ll}})\omega_{ii}^2\omega_{jj}^2\omega_{kk}^2\omega_{ll}^2}\, C_\omega =$$

$$= \frac{n}{(1-\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}})(1-\frac{\omega_{kl}^2}{\omega_{kk}\omega_{ll}})} [4\,\frac{\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}\frac{\omega_{ik}}{\sqrt{\omega_{ii}\omega_{kk}}}\frac{\omega_{jl}}{\sqrt{\omega_{jj}\omega_{ll}}}\frac{\omega_{kl}}{\sqrt{\omega_{kk}\omega_{ll}}} + 4\,\frac{\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}\frac{\omega_{il}}{\sqrt{\omega_{ii}\omega_{ll}}}\frac{\omega_{jk}}{\sqrt{\omega_{jj}\omega_{kk}}}\frac{\omega_{kl}}{\sqrt{\omega_{kk}\omega_{ll}}}$$

$$-4\,\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}}\frac{\omega_{ik}}{\sqrt{\omega_{ii}\omega_{kk}}}\frac{\omega_{il}}{\sqrt{\omega_{ii}\omega_{ll}}}\frac{\omega_{kl}}{\sqrt{\omega_{kk}\omega_{ll}}} - 4\,\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}}\frac{\omega_{jk}}{\sqrt{\omega_{jj}\omega_{kk}}}\frac{\omega_{jl}}{\sqrt{\omega_{jj}\omega_{ll}}}\frac{\omega_{kl}}{\sqrt{\omega_{kk}\omega_{ll}}}$$

$$-4\,\frac{\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}\frac{\omega_{ik}}{\sqrt{\omega_{ii}\omega_{kk}}}\frac{\omega_{jk}}{\sqrt{\omega_{jj}\omega_{kk}}}\frac{\omega_{kl}^2}{\omega_{kk}\omega_{ll}} - 4\,\frac{\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}\frac{\omega_{il}}{\sqrt{\omega_{ii}\omega_{ll}}}\frac{\omega_{jl}}{\sqrt{\omega_{jj}\omega_{ll}}}\frac{\omega_{kl}^2}{\omega_{kk}\omega_{ll}}$$

$$+2\,\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}}\frac{\omega_{jl}^2}{\omega_{jj}\omega_{ll}}\frac{\omega_{kl}^2}{\omega_{kk}\omega_{ll}} + 2\,\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}}\frac{\omega_{jk}^2}{\omega_{jj}\omega_{kk}}\frac{\omega_{kl}^2}{\omega_{kk}\omega_{ll}} + 2\,\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}}\frac{\omega_{il}^2}{\omega_{ii}\omega_{ll}}\frac{\omega_{kl}^2}{\omega_{kk}\omega_{ll}} + 2\,\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}}\frac{\omega_{ik}^2}{\omega_{ii}\omega_{kk}}\frac{\omega_{kl}^2}{\omega_{kk}\omega_{ll}}].$$

Finally, replacing all elements $\frac{-\omega_{pq}}{\sqrt{\omega_{pp}\omega_{qq}}}$ by $\rho_{pq.rest}$ gives Equation 3.9. Using this general formula, the calculation of the covariance of the test statistics for non-neighbour vertices (i.e., i, j, k and l all different) is straightforward: it just requires replacing i, j, k and l by the number of the vertices one is interested in. The situation of neighbour vertices is similar. However, after the substitution is performed, values of $\rho_{ii.rest}$, $\rho_{jj.rest}$, $\rho_{kk.rest}$ or

$\rho_{ll.rest}$ are obtained, which must be set to -1. Indeed, the ρ can be thought of as minus the elements of the scaled inverse variance matrix. Therefore, the off-diagonal elements correspond to the partial correlations, whereas the diagonal elements correspond to -1. Some examples of the final versions of the formulae follow:

- $\text{cov}(T^L_{12}, T^L_{34}) =$
$\frac{n}{(1-\rho^2_{12.rest})(1-\rho^2_{34.rest})}$ $[4\,\rho_{12.rest}\rho_{13.rest}\rho_{24.rest}\rho_{34.rest} + 4\,\rho_{12.rest}\rho_{14.rest}\rho_{23.rest}\rho_{34.rest} +$
$4\,\rho^2_{12.rest}\rho_{13.rest}\rho_{14.rest}\rho_{34.rest} + 4\,\rho^2_{12.rest}\rho_{23.rest}\rho_{24.rest}\rho_{34.rest} +$
$4\,\rho_{12.rest}\rho_{13.rest}\rho_{23.rest}\rho^2_{34.rest} + 4\,\rho_{12.rest}\rho_{14.rest}\rho_{24.rest}\rho^2_{34.rest} + 2\,\rho^2_{12.rest}\rho^2_{24.rest}\rho^2_{34.rest} +$
$2\,\rho^2_{12.rest}\rho^2_{23.rest}\rho^2_{34.rest} + 2\,\rho^2_{12.rest}\rho^2_{14.rest}\rho^2_{34.rest} + 2\,\rho^2_{12.rest}\rho^2_{34.rest}\rho^2_{13.rest}]$;

- $\text{cov}(T^L_{12}, T^L_{13}) = \frac{n}{(1-\rho^2_{12.rest})(1-\rho^2_{13.rest})}$ $[-4\,\rho_{12.rest}\rho_{13.rest}\rho_{23.rest} - 2\,\rho^2_{12.rest}\rho^2_{13.rest} +$
$4\,\rho^3_{12.rest}\rho_{13.rest}\rho_{23.rest} + 4\,\rho_{12.rest}\rho^3_{13.rest}\rho_{23.rest} + 2\,\rho^4_{12.rest}\rho^2_{13.rest} + 2\,\rho^2_{12.rest}\rho^4_{13.rest} +$
$2\,\rho^2_{12.rest}\rho^2_{13.rest}\rho^2_{23.rest}]$;

- $\text{cov}(T^L_{12}, T^L_{24}) = \frac{n}{(1-\rho^2_{12.rest})(1-\rho^2_{24.rest})}$ $[-4\,\rho_{12.rest}\rho_{24.rest}\rho_{14.rest} - 2\,\rho^2_{12.rest}\rho^2_{24.rest} +$
$4\,\rho^3_{12.rest}\rho_{24.rest}\rho_{14.rest} + 4\,\rho_{12.rest}\rho^3_{24.rest}\rho_{14.rest} + 2\,\rho^2_{12.rest}\rho^4_{24.rest} + 2\,\rho^4_{12.rest}\rho^2_{24.rest} +$
$2\,\rho^2_{12.rest}\rho^2_{24.rest}\rho^2_{14.rest}]$;

- $\text{cov}(T^L_{13}, T^L_{23}) = \frac{n}{(1-\rho^2_{13.rest})(1-\rho^2_{23.rest})}$ $[-4\,\rho_{13.rest}\rho_{23.rest}\rho_{12.rest} - 2\,\rho^2_{13.rest}\rho^2_{23.rest} +$
$4\,\rho^3_{13.rest}\rho_{23.rest}\rho_{12.rest} + 4\,\rho_{13.rest}\rho^3_{23.rest}\rho_{12.rest} + 2\,\rho^2_{13.rest}\rho^4_{23.rest} + 2\,\rho^4_{13.rest}\rho^2_{23.rest} +$
$2\,\rho^2_{13.rest}\rho^2_{23.rest}\rho^2_{12.rest}]$.

3.1.4 Asymptotic distributions of the Wald and of the score test statistics

Using the delta-method, and following the reasoning of Section 3.1.2, this section proposes normal approximations to the distributions of the Wald test statistic and of the score test statistic for single edge exclusion from the saturated model. Formulae for the means, variances and covariances, in the asymptotic distributions, are given as a function of the elements of the inverse variance matrix and as a function of the partial correlation coefficients. The section ends with a table summarising the formulae for the variances and covariances, in the asymptotic distribution, of the three test statistics for single edge exclusion from the saturated model used in this thesis.

Let $\text{var}(T_{ij}^W)$, $\text{cov}(T_{ij}^W, T_{kl}^W)$ denote the variances and the covariances of the Wald test statistic in the asymptotic distribution, and $\text{var}(T_{ij}^S)$ and $\text{cov}(T_{ij}^S, T_{kl}^S)$ denote the variances and covariances of the score test statistic, for single edge exclusion from the saturated model, in the asymptotic distribution. $AE[T_{ij}^W]$ and $AE[T_{ij}^S]$ denote the means, in the asymptotic distribution, respectively of the Wald and of the score test statistics.

These means can be written as a function of the partial correlation coefficient, as

$$AE[T_{ij}^W] = n \frac{\rho_{ij.rest}^2}{1 + \rho_{ij.rest}^2}, \tag{3.10}$$

in the case of the Wald test statistic, and as

$$AE[T_{ij}^S] = n \rho_{ij.rest}^2, \tag{3.11}$$

in the case of the score test statistic. The relationship between the variance (and the covariance) of the test statistics using the likelihood ratio test and the Wald test can easily be justified. In fact, a second use of the delta method would require f_{ij}^W to be written as a function of f_{ij}^L and, therefore, the new Δ matrix would also include $\frac{\partial f_{ij}^W}{\partial f_{ij}^L}$, which is equivalent to $\frac{\partial f_{ij}^W / \partial \omega_{ij}}{\partial f_{ij}^L / \partial \omega_{ij}}$. As a result,

$$\begin{aligned}
\text{var}(T_{ij}^W) &= \text{var}(T_{ij}^L) \frac{\partial f_{ij}^W / \partial \omega_{ij}}{\partial f_{ij}^L / \partial \omega_{ij}} \frac{\partial f_{ij}^W / \partial \omega_{ij}}{\partial f_{ij}^L / \partial \omega_{ij}} \\
&= \text{var}(T_{ij}^L) \frac{(1 - \frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}})^2}{(1 + \frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}})^4} \\
&= \text{var}(T_{ij}^L) \frac{(1 - \rho_{ij.rest}^2)^2}{(1 + \rho_{ij.rest}^2)^4},
\end{aligned}$$

$$\begin{aligned}
\text{and } \text{cov}(T_{ij}^W, T_{kl}^W) &= \text{cov}(T_{ij}^L, T_{kl}^L) \frac{\partial f_{ij}^W / \partial \omega_{ij}}{\partial f_{ij}^L / \partial \omega_{ij}} \frac{\partial f_{kl}^W / \partial \omega_{kl}}{\partial f_{kl}^L / \partial \omega_{kl}} \\
&= \text{cov}(T_{ij}^L, T_{kl}^L) \frac{(1 - \frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}})(1 - \frac{\omega_{kl}^2}{\omega_{kk}\omega_{ll}})}{(1 + \frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}})^2 (1 + \frac{\omega_{kl}^2}{\omega_{kk}\omega_{ll}})^2} \\
&= \text{cov}(T_{ij}^L, T_{kl}^L) \frac{(1 - \rho_{ij.rest}^2)(1 - \rho_{kl.rest}^2)}{(1 + \rho_{ij.rest}^2)^2 (1 + \rho_{kl.rest}^2)^2}.
\end{aligned}$$

Analogously, for the score test

$$\begin{aligned}
\text{var}(T_{ij}^S) &= \text{var}(T_{ij}^L) \frac{\partial f_{ij}^S/\partial \omega_{ij}}{\partial f_{ij}^L/\partial \omega_{ij}} \frac{\partial f_{ij}^S/\partial \omega_{ij}}{\partial f_{ij}^L/\partial \omega_{ij}} \\
&= \text{var}(T_{ij}^L) \ (1 - \frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}})^2 \\
&= \text{var}(T_{ij}^L) \ (1 - \rho_{ij.rest}^2)^2, \\
\text{and } \text{cov}(T_{ij}^S, T_{kl}^S) &= \text{cov}(T_{ij}^L, T_{kl}^L) \frac{\partial f_{ij}^S/\partial \omega_{ij}}{\partial f_{ij}^L/\partial \omega_{ij}} \frac{\partial f_{kl}^S/\partial \omega_{kl}}{\partial f_{kl}^L/\partial \omega_{kl}} \\
&= \text{cov}(T_{ij}^L, T_{kl}^L) \ (1 - \frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}})(1 - \frac{\omega_{kl}^2}{\omega_{kk}\omega_{ll}}) \\
&= \text{cov}(T_{ij}^L, T_{kl}^L) \ (1 - \rho_{ij.rest}^2)(1 - \rho_{kl.rest}^2).
\end{aligned}$$

Alternatively, general formulae for the covariance of the test statistics, in the asymptotic distribution, can be obtained from Equation 3.4, replacing f_{ij}^L by f_{ij}^W (in the case of the Wald test statistic) or by f_{ij}^S (in the case of the score test statistic). Consequently,

$$\text{cov}(T_{ij}^W, T_{kl}^W) = \sum_{p,q,r,s} \left\{ n \frac{\partial f_{ij}^W}{\partial \omega_{pq}} \frac{\partial f_{kl}^W}{\partial \omega_{rs}} (\omega_{pr}\omega_{qs} + \omega_{ps}\omega_{qr}) \right\}, \tag{3.12}$$

and

$$\text{cov}(T_{ij}^S, T_{kl}^S) = \sum_{p,q,r,s} \left\{ n \frac{\partial f_{ij}^S}{\partial \omega_{pq}} \frac{\partial f_{kl}^S}{\partial \omega_{rs}} (\omega_{pr}\omega_{qs} + \omega_{ps}\omega_{qr}) \right\}. \tag{3.13}$$

Table 3.1 summarises the derived formulae for the variances and covariances of the three test statistics, for single edge exclusion from the saturated GG model, in the asymptotic distribution.

	variance	covariance
likelihood	$\text{var}(T_{ij}^L) =$	$\text{cov}(T_{ij}^L, T_{kl}^L) =$
ratio test	$= \frac{4n\omega_{ij}^2}{\omega_{ii}\omega_{jj}}$	$= \frac{n}{\omega_{ii}\omega_{jj}\omega_{kk}\omega_{ll}(\omega_{ii}\omega_{jj}-\omega_{ij}^2)(\omega_{kk}\omega_{ll}-\omega_{kl}^2)}\ C_\omega$
	$= 4n\rho_{ij.rest}^2$	$= \frac{n}{(1-\rho_{ij.rest}^2)(1-\rho_{kl.rest}^2)}\ C_\rho$
Wald test	$\text{var}(T_{ij}^W) =$	$\text{cov}(T_{ij}^W, T_{kl}^W) =$
	$= \frac{4n\omega_{ij}^2(1-\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}})^2}{\omega_{ii}\omega_{jj}(1+\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}})^4}$	$= \frac{n}{(\omega_{ii}\omega_{jj}+\omega_{ij}^2)^2(\omega_{kk}\omega_{ll}+\omega_{kl}^2)^2}\ C_\omega$
	$= \frac{4n\rho_{ij.rest}^2(1-\rho_{ij.rest}^2)^2}{(1+\rho_{ij.rest}^2)^4}$	$= \frac{n}{(1+\rho_{ij.rest}^2)^2(1+\rho_{kl.rest}^2)^2}\ C_\rho$
score test	$\text{var}(T_{ij}^S) =$	$\text{cov}(T_{ij}^S, T_{kl}^S) =$
	$= 4n\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}}(1-\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}})^2$	$= \frac{n}{\omega_{ii}^2\omega_{jj}^2\omega_{kk}^2\omega_{ll}^2}\ C_\omega$
	$= 4n\rho_{ij.rest}^2(1-\rho_{ij.rest}^2)^2$	$= n\ C_\rho$

Table 3.1: Variances and covariances of the three test statistics (for single edge exclusion, from the saturated GG model) in the asymptotic distribution.

3.1.5 Asymptotic distributions of the signed square-root versions of the test statistics

The signed square-root versions of the test statistics are presented in Section 2.9.2, Table 2.1. The aim now is to obtain asymptotic normal approximations to the distributions of the signed square-root versions of the test statistics, once more using the delta-method. A summary table with the derived formulae for the variances and covariances, in the asymptotic distribution, is presented at the end of the section.

The means can be written as a function of the partial correlation coefficient, as

$$AE[T_{ij}^{signL}] = sign(\rho_{ij.rest})\sqrt{-n\log(1-\rho_{ij.rest}^2)}, \tag{3.14}$$

in the case of the LRT statistic, as

$$AE[T_{ij}^{signW}] = \rho_{ij.rest}\sqrt{\frac{n}{1+\rho_{ij.rest}^2}}, \tag{3.15}$$

in the case of the Wald test statistic, and as

$$AE[T_{ij}^{signS}] = \rho_{ij.rest}\sqrt{n}, \tag{3.16}$$

in the case of the score test statistic.

The signed square-root versions can be written as a function of the corresponding non-signed ones, and so the delta-method can be applied to obtain the variance matrix of the signed versions, based on the variance matrix of the corresponding non-signed versions. The new Δ matrix has the derivative of the signed square-root version with respect to the non-signed version. For example, when the likelihood ratio test is used,

$$\begin{aligned}
\frac{\partial f_{ij}^{signL}}{\partial f_{ij}^{L}} &= \frac{\partial\left(sign(\frac{-\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}})\sqrt{f_{ij}^{L}}\right)}{\partial f_{ij}^{L}} \\
&= sign(\tfrac{-\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}) \times \tfrac{1}{2\sqrt{f_{ij}^{L}}} \\
&= sign(\tfrac{-\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}) \times \frac{1}{2\sqrt{-\log(1-\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}})}} \\
&= sign(\rho_{ij.rest}) \times \frac{1}{2\sqrt{-\log(1-\rho_{ij.rest}^2)}}.
\end{aligned}$$

Hence, since $T_{ij}^{signL} = \sqrt{n}\, f_{ij}^{signL}$, the variance of the signed square-root version of the likelihood ratio test statistic, in the asymptotic distribution, can be written as

$$\begin{aligned}
\mathrm{var}(T_{ij}^{signL}) &= (\sqrt{n})^2 \times sign^2\left(\tfrac{-\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}\right)\left(\frac{1}{2\sqrt{-\log(1-\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}})}}\right)^2 \times \tfrac{1}{n^2} \times \mathrm{var}(T_{ij}^{L}) \\
&= \mathrm{var}(T_{ij}^{L}) \times \frac{-1}{4n\log\left(1-\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}}\right)} \\
&= \mathrm{var}(T_{ij}^{L}) \times \frac{-1}{4n\log(1-\rho_{ij.rest}^2)},
\end{aligned}$$

and the covariance can be written as

$$\begin{aligned}\operatorname{cov}(T_{ij}^{signL},\, T_{kl}^{signL}) &= \operatorname{cov}(T_{ij}^{L},\, T_{kl}^{L}) \times \frac{sign\left(\frac{-\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}\right) sign\left(\frac{-\omega_{kl}}{\sqrt{\omega_{kk}\omega_{ll}}}\right)}{4n\sqrt{-\log\left(1-\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}}\right)}\sqrt{-\log\left(1-\frac{\omega_{kl}^2}{\omega_{kk}\omega_{ll}}\right)}} \\ &= \operatorname{cov}(T_{ij}^{L},\, T_{kl}^{L}) \times \frac{sign(\rho_{ij.rest})\, sign(\rho_{kl.rest})}{4n\sqrt{-\log(1-\rho_{ij.rest}^2)}\sqrt{-\log(1-\rho_{kl.rest}^2)}}.\end{aligned}$$

Similar reasoning can be followed for the signed square-root versions of the Wald and of the score test statistics. Table 3.2 summarises the proposed formulae for the variances and covariances of the signed square-root versions of the test statistics, for single edge exclusion from the saturated model, in the asymptotic distribution, for the likelihood ratio, the Wald and the score test statistics.

It is worth noting that both the variances and the covariances of the signed square-root versions of the test statistics do not depend on the sample size n. Additionally, the variances simplify to

$$\operatorname{var}(T_{ij}^{signL}) = \frac{-\rho_{ij.rest}^2}{\log(1-\rho_{ij.rest}^2)}$$

and

$$\operatorname{var}(T_{ij}^{signW}) = \frac{(1-\rho_{ij.rest}^2)^2}{(1+\rho_{ij.rest}^2)^3},$$

for the signed square-root likelihood ratio and Wald test statistics. A very 'neat' expression is obtained for the variance of the signed square-root score test:

$$\operatorname{var}(T_{ij}^{signS}) = (1-\rho_{ij.rest}^2)^2.$$

	variance	covariance
likelihood	$\text{var}(T_{ij}^{signL}) =$	$\text{cov}(T_{ij}^{signL}, T_{kl}^{signL}) =$
ratio test	$= \text{var}(T_{ij}^{L}) \times \frac{-1}{4n \log\left(1-\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}}\right)}$	$= \text{cov}(T_{ij}^{L}, T_{kl}^{L}) \times \frac{sign\left(\frac{-\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}\right) sign\left(\frac{-\omega_{kl}}{\sqrt{\omega_{kk}\omega_{ll}}}\right)}{4n\sqrt{-\log\left(1-\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}}\right)}\sqrt{-\log\left(1-\frac{\omega_{kl}^2}{\omega_{kk}\omega_{ll}}\right)}}$
	$= \text{var}(T_{ij}^{L}) \times \frac{-1}{4n \log(1-\rho_{ij.rest}^2)}$	$= \text{cov}(T_{ij}^{L}, T_{kl}^{L}) \times \frac{sign(\rho_{ij.rest})\, sign(\rho_{kl.rest})}{4n\sqrt{-\log(1-\rho_{ij.rest}^2)}\sqrt{-\log(1-\rho_{kl.rest}^2)}}$
Wald test	$\text{var}(T_{ij}^{signW}) =$	$\text{cov}(T_{ij}^{signW}, T_{kl}^{signW}) =$
	$= \text{var}(T_{ij}^{W}) \times \frac{1+\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}}}{4n\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}}}$	$= \text{cov}(T_{ij}^{W}, T_{kl}^{W}) \times \frac{\sqrt{1+\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}}}\sqrt{1+\frac{\omega_{kl}^2}{\omega_{kk}\omega_{ll}}}}{4n\frac{\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}\frac{\omega_{kl}}{\sqrt{\omega_{kk}\omega_{ll}}}}$
	$= \text{var}(T_{ij}^{W}) \times \frac{1+\rho_{ij.rest}^2}{4n\,\rho_{ij.rest}^2}$	$= \text{cov}(T_{ij}^{W}, T_{kl}^{W}) \times \frac{\sqrt{1+\rho_{ij.rest}^2}\sqrt{1+\rho_{kl.rest}^2}}{4n\,\rho_{ij.rest}\rho_{kl.rest}}$
score test	$\text{var}(T_{ij}^{signS}) =$	$\text{cov}(T_{ij}^{signS}, T_{kl}^{signS}) =$
	$= \text{var}(T_{ij}^{S}) \times \frac{1}{4n\frac{\omega_{ij}^2}{\omega_{ii}\omega_{jj}}}$	$= \text{cov}(T_{ij}^{S}, T_{kl}^{S}) \times \frac{1}{4n\frac{\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}\frac{\omega_{kl}}{\sqrt{\omega_{kk}\omega_{ll}}}}$
	$= \text{var}(T_{ij}^{S}) \times \frac{1}{4n\,\rho_{ij.rest}^2}$	$= \text{cov}(T_{ij}^{S}, T_{kl}^{S}) \times \frac{1}{4n\,\rho_{ij.rest}\rho_{kl.rest}}$

Table 3.2: Variances and covariances of the signed square-root versions of the test statistics (for single edge exclusion, from the saturated GG model) in the asymptotic distribution.

3.2 Non-central χ^2 Approximation to the Distribution of the LRT in a GG Model with Two Variables

The delta-method was used in Section 3.1 to derive normal approximations to the distributions of the test statistics, since each test statistic is a 'well behaved' function of the $\hat{\omega}$, the unconstrained m.l.e. of the elements of the inverse variance matrix, differentiable provided all $\hat{\omega}$ are different from zero. Because each statistic tests the null hypothesis that $\omega_{ij} = 0 (\Leftrightarrow \rho_{ij} = 0)$, the normal approximations should be poor at very small distances from the null.

However, power at *local alternatives* has been studied in the literature. In Section 2.9.1, when reviewing the rationale for the derivation of the likelihood ratio, Wald and score tests, it was mentioned that the three test statistics are chi-square distributed under the null hypothesis, and non-central chi-square distributed for local alternatives of the type $\theta_n = \theta_0 + \frac{\delta_\theta}{\sqrt{n}}$ (with nuisance parameters ψ unspecified), i.e., when $\sqrt{n}(\theta_n - \theta_0)$ converges to δ_θ. References to the topic of local alternatives include Cox and Hinkley (1974, page 324), Sen and Singer (1993, page 238), Ferguson (1996, pages 148–149) and Severini (2000, pages 117–119).

In Chapter 4 of this thesis power is calculated using the non central chi-square approximation. The aim of the current section is to derive the non central chi-square approximation to the distribution of the LRT statistic for single edge exclusion from a saturated GG model with two variables, at a local alternative. Section 3.3 compares the quality of the non central chi-square approximation to that of the normal approximation previously derived, in particular as the sample size varies.

Denoting the true value of the parameter by θ_n and the parameter point under H_0 closest to θ_n by θ_0, the *distance from the null* can be defined as $\delta_\theta = \sqrt{n}\,(\theta_n - \theta_0)$. In the simplest case of just one restriction under H_0, the noncentrality parameter φ can be obtained as

$$\varphi = \delta_\theta^T \, [K_{\theta_0\theta_0}]^{-1} \, \delta_\theta,$$

where $[K_{\theta_0\theta_0}]^{-1}$ is the partition of the information matrix associated with the restriction in H_0. This result is going to be used to obtain an asymptotic non-central chi-square approximation to the distribution of the non-signed version of the LRT for single edge exclusion from a saturated GG model with two variables, at an alternative close to the null. As mentioned in Section 2.7.1, the corresponding inverse information matrix is given by $K_{\theta_0\theta_0} = \omega_{11}\omega_{22} + \omega_{12}^2$, which can be expressed as a function of ρ_{12} as $1 + \rho_{12}^2$. Consequently, $[K_{\theta_0\theta_0}]^{-1} = 1/(1 + \rho_{12}^2)$.

It is, therefore, proposed that the distribution of the LRT for single edge exclusion from a saturated GG model with two variables, at a local alternative, is approximated by a non-central chi-square distribution, with noncentrality parameter given by

$$\varphi = [\sqrt{n}\,(\rho_{12} - 0)] \times \left[\frac{1}{1 + \rho_{12}^2}\right] \times [\sqrt{n}\,(\rho_{12} - 0)] = \frac{n\,\rho_{12}^2}{1 + \rho_{12}^2}.$$

In the next section the quality of such an approximation is assessed, as n and ρ_{12} vary.

3.3 Assessing the Quality of the Approximations, in GG Models

In Section 3.1 asymptotic normal approximations to the distributions of the test statistics for single edge exclusion from the saturated GG model were derived, and in Section 3.2 a non central chi-square approximation was proposed. The purpose of the current section is to analyse the quality of the two approximations, for various values of n, by using a simulation study. The two variables case is considered and the LRT statistic is used.

Recall that, under the alternative hypothesis that the saturated model holds, T_{12}^L is asymptotically normal distributed, with mean $AE[T_{12}^L] = -n\log(1-\rho_{12}^2)$ and variance $\text{var}(T_{12}^L) = 4n\rho_{12}^2$. At $\rho_{12} = 0$ the asymptotic distribution of T_{12}^L is degenerate, with mean zero and variance zero. Hence, the normal approximation holds for n at infinity, but it is poor for n finite. Indeed, for $\rho_{12} = 0$ and n finite, a chi-square distribution on one degree of freedom is a better approximation.

The results of the simulation study are now presented. First, it is shown how the mean and the variance of T_{12}^L vary as a function of both the correlation coefficient ρ_{12} and the sample size n. Then, the empirical distribution of the LRT statistic is plotted. Histograms are produced for different values of ρ_{12}, and three different sample sizes are used. The curve of the theoretical normal approximation derived in Section 3.1.3 overlaps each histogram, so that the quality of the approximation can be assessed. Simulated results for the mean, variance and covariance of the test statistic(s) were obtained as follows. Different values were assigned to the (partial) correlation coefficient, and samples of different sizes were generated for a normal distribution. Likelihood ratio test statistics (for single edge exclusion, from the saturated model) were calculated and stored at each step. After 1 000 repetitions, the mean and the variance values of the test statistics, were calculated.

Figure 3.1 displays simulation results. The mean values of the test statistic T_{12}^L are displayed in plots a), whereas plots b) represent the variances. The correlation coefficient varies between -0.9 and 0.9, on the horizontal axis. Plots 1) are based on a sample size of 50 and in plots 2), 3) and 4) n equals 200, 500 and 1 000, respectively. In each plot there are two lines: the blue line corresponds to the values obtained using the derived theoretical formulae, and the red line corresponds to the simulated values.

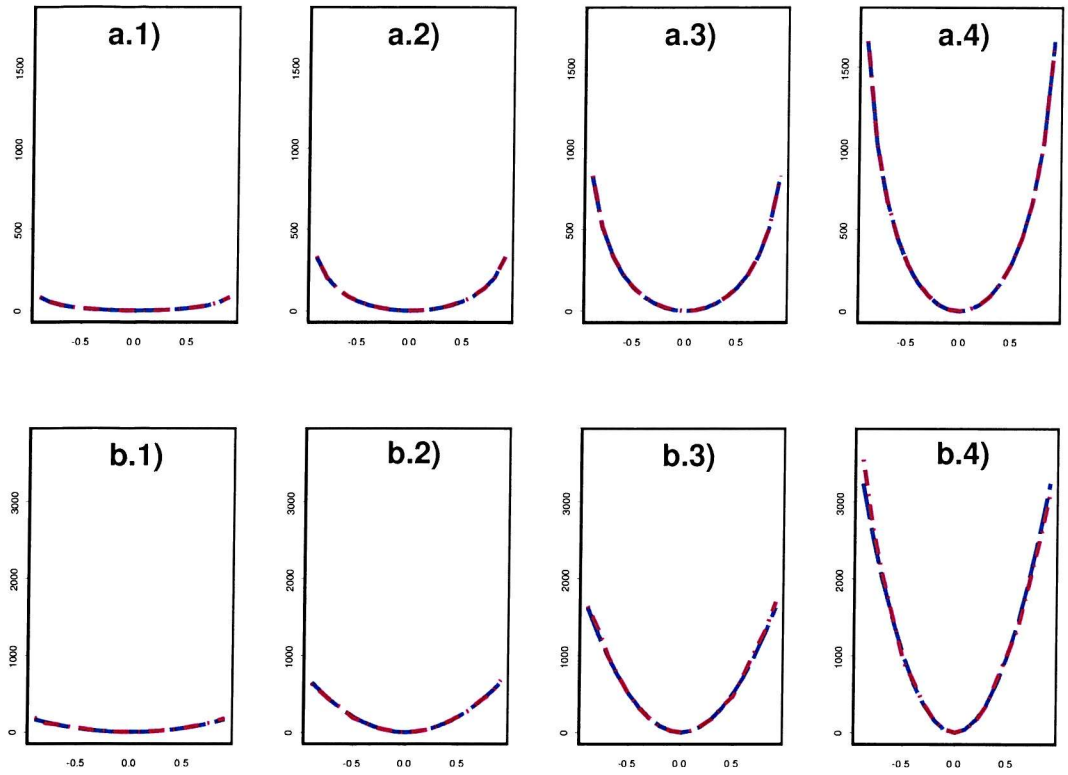

Figure 3.1: Mean (plots a) and variance (plots b) values of the likelihood ratio test statistic T_{12}^L, in the asymptotic normal distribution, as a function of ρ_{12}. Values were calculated using proposed formulae (blue lines) and by simulation (red lines), for different sample sizes: 1) $n = 50$, 2) $n = 200$, 3) $n = 500$, 4) $n = 1\,000$.

The fact that the two lines almost overlap shows the agreement between simulated results and formulae derived in Section 3.1. From these plots it is clear that both the mean and the variance of T_{12}^L, in the asymptotic distribution, increase as the sample size increases and as the absolute value of the correlation coefficient increases. Both are symmetric about zero correlation.

Figure 3.2 shows the histograms of the empirical distribution of the likelihood ratio test statistic T_{12}^L. Results are given for three different sample sizes: $n = 50$ (plots in blue), $n = 200$ (plots in red) and $n = 1\,000$ (plots in green). For each sample size there are ten different histograms, corresponding to the values of ρ_{12} ranging from 0 to 0.9 (with an interval of 0.1). Recall that, for $\rho_{12} = 0$, the asymptotic distribution of the test statistic for single edge exclusion is chi-squared on one degree of freedom (the green line overlapping the corresponding histograms represents the density of a χ_1^2 distribution). For $\rho_{12} \neq 0$, the asymptotic distribution of T_{12}^L tends to the normal distribution as n

tends to infinity (the black line overlapping most histograms represents the density of a normal distribution with mean $-n\log(1-\rho_{12}^2)$ and variance $4n\rho_{12}^2$). The histograms show that, as expected, the distribution looks chi-squared on one degree of freedom for $\rho_{12}=0$ (particularly noteworthy for large sample sizes). For $\rho_{12}\neq 0$, the distribution of T_{12}^L tends to the normal distribution, faster for larger sample sizes: for $n=1000$ the histograms start showing a normal shape at $\rho_{12}=0.2$, whereas for $n=50$ this happens at $\rho_{12}=0.6$.

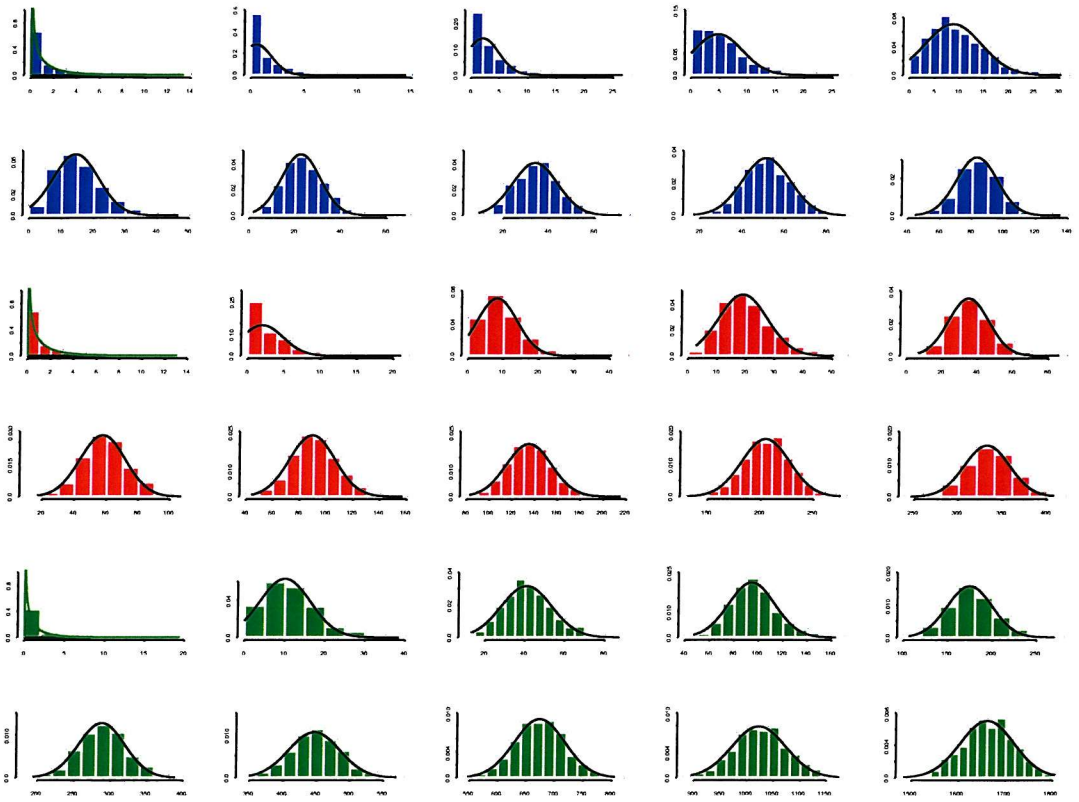

Figure 3.2: Histograms of the empirical distribution of the likelihood ratio test statistic T_{12}^L, for different sample sizes: $n=50$ (plots in blue), $n=200$ (plots in red) and $n=1\,000$ (plots in green). Normal density overlapping. For each sample size ρ_{12} ranges from 0 to 0.9 (with an interval of 0.1)

In brief: the normal approximation to the distribution of the LRT statistic for single edge exclusion from the saturated GG model with two variables is a good approximation if n is large and ρ_{12} is not close to zero. The approximation is poor for small sample sizes and values of the correlation coefficient close to zero.

In Section 3.2 a non central chi-square approximation to the distribution of the

LRT statistic for single edge exclusion from the saturated GG model was proposed. The aim now is to assess the quality of such an approximation and contrast the normal approximation with the non central chi-square approximation. For that, p-p plots were produced, comparing the observed cumulative probabilities (obtained in the simulation study) with those cumulative probabilities that would be expected if the simulated values of the LRT statistic were asymptotically normal or non-central χ^2 distributed. The red line represents an exact agreement between observed and expected, the blue curve represents the asymptotic normal approximation (with mean $-n\log(1-\rho_{12}^2)$ and variance $4n\rho_{12}^2$) and the green curve represents the non central chi-square approximation (with noncentrality parameter $\frac{n\,\rho_{12}^2}{1+\rho_{12}^2}$).

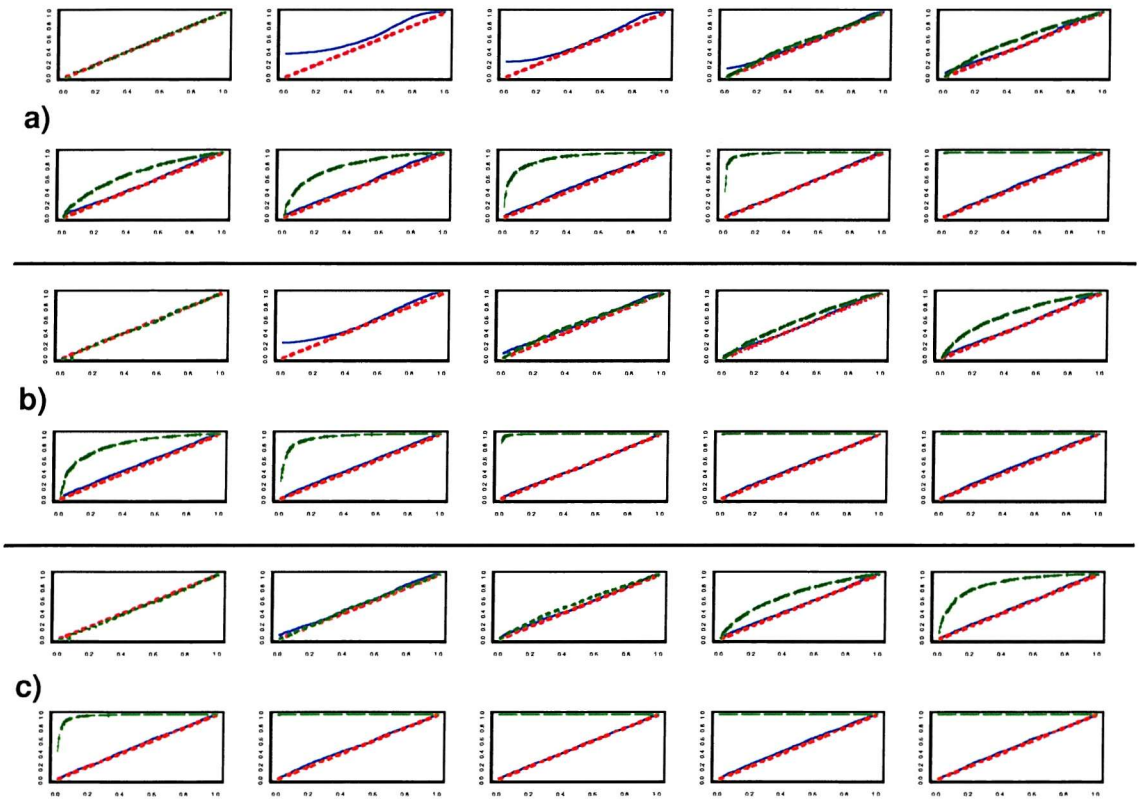

Figure 3.3: P-P plots of the distribution of the likelihood ratio test statistic T_{12}^L, for different sample sizes: $n = 50$ (plots a)), $n = 200$ (plots b)) and $n = 1\,000$ (plots c)). For each sample size ρ_{12} ranges from 0 to 0.9 (with an interval of 0.1)

Figure 3.3 shows some of the p-p plots that were obtained. Some conclusions can be drawn:

- the normal approximation works well for large values of n (1 000), even for small correlations (above 0.2). As stated before, for n small the normal approximation

is poor. The approximations seems to hold for $\rho_{12} > 0.3$ when $n = 200$, whereas with $n = 50$ it just holds for $\rho_{12} > 0.5$;

- the non-central chi-square approximation performs better for small values of ρ_{12}, in particular if the sample size is not large. The approximation does not hold for correlations distant from zero, becoming even worse as n increases.

These two conclusions relate to the issue of the distance from the null - recall that the use of the non-central chi-square approximation is suggested *at an alternative close to the null.* This distance being measured by $\delta = \sqrt{n}(\rho_{12} - 0)$ means that, as n increases, the approximation performs better when ρ_{12} becomes closer to zero.

In brief: in the two variables case the normal approximation is a better approximation if the sample size is large and the correlation coefficient is not close to zero. The non central chi-square approximation performs better than the normal at small distance from the null, i.e., if $n = 1\,000$ and $\rho_{12} \leq 0.1$, $n = 200$ and $\rho_{12} \leq 0.2$ or $n = 50$ and $\rho_{12} \leq 0.3$.

The first three sections of the current chapter concerned GG models. In the remainder of the chapter GLL models are considered.

3.4 Test Statistics for Single Edge Exclusion in GLL Models: Two and Three Variables Cases

This second part of Chapter 3 is devoted to the study of the distributions of the test statistics for single edge exclusion, from the saturated model, in graphical log-linear models. Attention is restricted to the two and the three variables cases, due to the fact that the number of parameters to be considered increases substantially as the number of variables increases. So does the complexity of the notation to be used. This did not happen in GG models, where results could easily be generalised to the p variables case. The structure of this second part of the chapter is the following: first, the likelihood ratio test statistic is expressed in cell probabilities, using a log-linear formulation (Section 3.4.1), and so are the Wald and the score test statistics (Section 3.4.2) and corresponding signed square-root versions (Section 3.4.3). Then, asymptotic normal approximations to the distributions of the test statistics are derived:

for the LRT statistic in Section 3.5.1, for the Wald and the score test statistics in Section 3.5.2, and for the corresponding signed square-root versions in Section 3.5.3. Section 3.6 proposes a non central chi-square approximation to the distribution of the LRT statistic, for single edge exclusion from the saturated GLL model and Section 3.7 assesses the quality of the two derived approximations.

3.4.1 LRT statistic for single edge exclusion from the saturated GLL model

The two variables case

Suppose X_1 and X_2 are two binary variables. Using a log-linear expansion, the saturated model can be written as $\log f_{12} = \lambda_\emptyset + \lambda_1 + \lambda_2 + \lambda_{12}$.

The m.l.e. of the cell probabilities under the saturated model, $\hat{\pi}(x)$, and under the constrained model of independence, $\tilde{\pi}(x)$, are

$\hat{\pi}(0,0)$	$= \exp\{\hat{\lambda}_\emptyset\}$	$\tilde{\pi}(0,0)$	$= \hat{\pi}_1(0) \times \hat{\pi}_2(0)$
$\hat{\pi}(0,1)$	$= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_2\}$	$\tilde{\pi}(0,1)$	$= \hat{\pi}_1(0) \times \hat{\pi}_2(1)$
$\hat{\pi}(1,0)$	$= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_1\}$	$\tilde{\pi}(1,0)$	$= \hat{\pi}_1(1) \times \hat{\pi}_2(0)$
$\hat{\pi}(1,1)$	$= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_1 + \hat{\lambda}_2 + \hat{\lambda}_{12}\}$	$\tilde{\pi}(1,1)$	$= \hat{\pi}_1(1) \times \hat{\pi}_2(1)$
$\hat{\pi}_1(0)$	$= \exp\{\hat{\lambda}_\emptyset\} \times (1 + \exp\{\hat{\lambda}_2\})$	$\tilde{\pi}_1(0)$	$= \hat{\pi}_1(0)$
$\hat{\pi}_1(1)$	$= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_1\} \times (1 + \exp\{\hat{\lambda}_2 + \hat{\lambda}_{12}\})$	$\tilde{\pi}_1(1)$	$= \hat{\pi}_1(1)$
$\hat{\pi}_2(0)$	$= \exp\{\hat{\lambda}_\emptyset\} \times (1 + \exp\{\hat{\lambda}_1\})$	$\tilde{\pi}_2(0)$	$= \hat{\pi}_2(0)$
$\hat{\pi}_2(1)$	$= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_2\} \times (1 + \exp\{\hat{\lambda}_1 + \hat{\lambda}_{12}\})$	$\tilde{\pi}_2(1)$	$= \hat{\pi}_2(1)$

Note that under independence the marginal probabilities do not change and the cell probabilities are the product of the corresponding marginal probabilities. From Section 2.9.3, the likelihood ratio test statistic for single edge exclusion from the saturated model, LRT_{12}, equals

$$\begin{aligned} LRT_{12} &= 2\, n_\emptyset \sum_x \hat{\pi}(x) \log\left(\frac{\hat{\pi}(x)}{\tilde{\pi}(x)}\right) = 2\, n_\emptyset \sum_{x_1,x_2\in\{0,1\}} \hat{\pi}_{12}(x_1,x_2) \log\left(\frac{\hat{\pi}_{12}(x_1,x_2)}{\hat{\pi}_1(x_1)\,\hat{\pi}_2(x_2)}\right) \\ &= 2\, n_\emptyset \left[\hat{\pi}(0,0) \log\left(\frac{\hat{\pi}(0,0)}{\hat{\pi}_1(0)\hat{\pi}_2(0)}\right) + \hat{\pi}(0,1) \log\left(\frac{\hat{\pi}(0,1)}{\hat{\pi}_1(0)\hat{\pi}_2(1)}\right)\right. \\ &\qquad \left. + \hat{\pi}(1,0) \log\left(\frac{\hat{\pi}(1,0)}{\hat{\pi}_1(1)\hat{\pi}_2(0)}\right) + \hat{\pi}(1,1) \log\left(\frac{\hat{\pi}(1,1)}{\hat{\pi}_1(1)\hat{\pi}_2(1)}\right)\right]. \end{aligned} \tag{3.17}$$

It is suggested that the likelihood ratio test statistic is expressed as a function of the m.l.e. of the λ terms in the log-linear expansion of f_{12}, as

$$\begin{aligned} LRT_{12} &= 2\,n_\emptyset \Big[-\exp\{\hat{\lambda}_\emptyset\} \log\Big(\exp\{\hat{\lambda}_\emptyset\} \times (1+\exp\{\hat{\lambda}_1\}) \times (1+\exp\{\hat{\lambda}_2\})\Big) \\ &\quad -\exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_2\} \log\Big(\exp\{\hat{\lambda}_\emptyset\} \times (1+\exp\{\hat{\lambda}_2\}) \times (1+\exp\{\hat{\lambda}_1 + \hat{\lambda}_{12}\})\Big) \\ &\quad -\exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_1\} \log\Big(\exp\{\hat{\lambda}_\emptyset\} \times (1+\exp\{\hat{\lambda}_1\}) \times (1+\exp\{\hat{\lambda}_2 + \hat{\lambda}_{12}\})\Big) \\ &\quad +\exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_1 + \hat{\lambda}_2 + \hat{\lambda}_{12}\} \Big(\log(\exp\{\hat{\lambda}_{12}\}) \\ &\quad -\log(\exp\{\hat{\lambda}_\emptyset\} \times (1+\exp\{\hat{\lambda}_2 + \hat{\lambda}_{12}\}) \times (1+\exp\{\hat{\lambda}_1 + \hat{\lambda}_{12}\})\Big) \Big]. \end{aligned}$$

Since the sum of the four cell probabilities equals one, $\lambda_\emptyset$ is a function of λ_1, λ_2 and λ_{12}, i.e., $\hat{\lambda}_\emptyset = -\log(1+\exp\{\hat{\lambda}_1\} + \exp\{\hat{\lambda}_2\} + \exp\{\hat{\lambda}_1 + \hat{\lambda}_2 + \hat{\lambda}_{12}\})$.

The likelihood ratio test statistic for single edge exclusion from the saturated model can also be written as a function of the dependence ratios, as

$$\begin{aligned} LRT_{12} &= 2\,n_\emptyset \textstyle\sum_{x_1,x_2\in\{0,1\}} \hat{\tau}_{12}(x_1,x_2)\,\hat{\pi}_1(x_1)\,\hat{\pi}_2(x_2) \log\left(\hat{\tau}_{12}(x_1,x_2)\right) \\ &= 2\,n_\emptyset \left[\hat{\tau}_{12}(0,0)\,\hat{\pi}_1(0)\,\hat{\pi}_2(0) \log\left(\hat{\tau}_{12}(0,0)\right) + \hat{\tau}_{12}(0,1)\,\hat{\pi}_1(0)\,\hat{\pi}_2(1) \log\left(\hat{\tau}_{12}(0,1)\right)\right. \\ &\quad \left. +\hat{\tau}_{12}(1,0)\,\hat{\pi}_1(1)\,\hat{\pi}_2(0) \log\left(\hat{\tau}_{12}(1,0)\right) + \hat{\tau}_{12}(1,1)\,\hat{\pi}_1(1)\,\hat{\pi}_2(1) \log\left(\hat{\tau}_{12}(1,1)\right)\right]. \end{aligned}$$

The three variables case

Suppose X_1, X_2 and X_3 are three binary variables, cross-classified in a $2 \times 2 \times 2$ contingency table. Using a log-linear expansion, the saturated model can be written as $\log f_{123} = \lambda_\emptyset + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_{12} + \lambda_{13} + \lambda_{23} + \lambda_{123}$. Since there are three variables, there are three possible conditional independencies to test for, namely that:

- $X_1 \perp\!\!\!\perp X_2 | X_3 \Leftrightarrow H_0 : \lambda_{12} = \lambda_{123} = 0$; $H_a : \lambda_{12}, \lambda_{123}$ unconstrained;
- $X_1 \perp\!\!\!\perp X_3 | X_2 \Leftrightarrow H_0 : \lambda_{13} = \lambda_{123} = 0$; $H_a : \lambda_{13}, \lambda_{123}$ unconstrained;
- $X_2 \perp\!\!\!\perp X_3 | X_1 \Leftrightarrow H_0 : \lambda_{23} = \lambda_{123} = 0$; $H_a : \lambda_{23}, \lambda_{123}$ unconstrained.

Consequently, there are three different likelihood ratio test statistics for single edge exclusion from the saturated model, denoted as LRT_{12}, LRT_{13} and LRT_{23}. In the

three variables case, a general expression for the likelihood ratio test statistic for the exclusion of edge ij from the saturated model is given by

$$LRT_{ij} = 2\,n_{\emptyset} \sum_{x_i,x_j,x_k \in \{0,1\}} \hat{\pi}_{ijk}(x_i,x_j,x_k)\, \log\left(\frac{\hat{\pi}_{ijk}(x_i,x_j,x_k)\,\hat{\pi}_k(x_k)}{\hat{\pi}_{ik}(x_i,x_k)\,\hat{\pi}_{jk}(x_j,x_k)}\right). \tag{3.18}$$

Hence, the values of the three test statistics can be obtained as

$$\begin{aligned} LRT_{12} = 2\,n_{\emptyset} \Big[& \hat{\pi}(0,0,0) \log\left(\tfrac{\hat{\pi}(0,0,0)\hat{\pi}_3(0)}{\hat{\pi}_{13}(0,0)\hat{\pi}_{23}(0,0)}\right) + \hat{\pi}(0,1,0) \log\left(\tfrac{\hat{\pi}(0,1,0)\hat{\pi}_3(0)}{\hat{\pi}_{13}(0,0)\hat{\pi}_{23}(1,0)}\right) \\ & + \hat{\pi}(1,0,0) \log\left(\tfrac{\hat{\pi}(1,0,0)\hat{\pi}_3(0)}{\hat{\pi}_{13}(1,0)\hat{\pi}_{23}(0,0)}\right) + \hat{\pi}(1,1,0) \log\left(\tfrac{\hat{\pi}(1,1,0)\hat{\pi}_3(0)}{\hat{\pi}_{13}(1,0)\hat{\pi}_{23}(1,0)}\right) \\ & + \hat{\pi}(0,0,1) \log\left(\tfrac{\hat{\pi}(0,0,1)\hat{\pi}_3(1)}{\hat{\pi}_{13}(0,1)\hat{\pi}_{23}(0,1)}\right) + \hat{\pi}(0,1,1) \log\left(\tfrac{\hat{\pi}(0,1,1)\hat{\pi}_3(1)}{\hat{\pi}_{13}(0,1)\hat{\pi}_{23}(1,1)}\right) \\ & + \hat{\pi}(1,0,1) \log\left(\tfrac{\hat{\pi}(1,0,1)\hat{\pi}_3(1)}{\hat{\pi}_{13}(1,1)\hat{\pi}_{23}(0,1)}\right) + \hat{\pi}(1,1,1) \log\left(\tfrac{\hat{\pi}(1,1,1)\hat{\pi}_3(1)}{\hat{\pi}_{13}(1,1)\hat{\pi}_{23}(1,1)}\right) \Big]; \end{aligned}$$

$$\begin{aligned} LRT_{13} = 2\,n_{\emptyset} \Big[& \hat{\pi}(0,0,0) \log\left(\tfrac{\hat{\pi}(0,0,0)\hat{\pi}_2(0)}{\hat{\pi}_{12}(0,0)\hat{\pi}_{23}(0,0)}\right) + \hat{\pi}(0,1,0) \log\left(\tfrac{\hat{\pi}(0,1,0)\hat{\pi}_2(1)}{\hat{\pi}_{12}(0,1)\hat{\pi}_{23}(1,0)}\right) \\ & + \hat{\pi}(1,0,0) \log\left(\tfrac{\hat{\pi}(1,0,0)\hat{\pi}_2(0)}{\hat{\pi}_{12}(1,0)\hat{\pi}_{23}(0,0)}\right) + \hat{\pi}(1,1,0) \log\left(\tfrac{\hat{\pi}(1,1,0)\hat{\pi}_2(1)}{\hat{\pi}_{12}(1,1)\hat{\pi}_{23}(1,0)}\right) \\ & + \hat{\pi}(0,0,1) \log\left(\tfrac{\hat{\pi}(0,0,1)\hat{\pi}_2(0)}{\hat{\pi}_{12}(0,0)\hat{\pi}_{23}(0,1)}\right) + \hat{\pi}(0,1,1) \log\left(\tfrac{\hat{\pi}(0,1,1)\hat{\pi}_2(1)}{\hat{\pi}_{12}(0,1)\hat{\pi}_{23}(1,1)}\right) \\ & + \hat{\pi}(1,0,1) \log\left(\tfrac{\hat{\pi}(1,0,1)\hat{\pi}_2(0)}{\hat{\pi}_{12}(1,0)\hat{\pi}_{23}(0,1)}\right) + \hat{\pi}(1,1,1) \log\left(\tfrac{\hat{\pi}(1,1,1)\hat{\pi}_2(1)}{\hat{\pi}_{12}(1,1)\hat{\pi}_{23}(1,1)}\right) \Big]; \end{aligned}$$

$$\begin{aligned} LRT_{23} = 2\,n_{\emptyset} \Big[& \hat{\pi}(0,0,0) \log\left(\tfrac{\hat{\pi}(0,0,0)\hat{\pi}_1(0)}{\hat{\pi}_{12}(0,0)\hat{\pi}_{13}(0,0)}\right) + \hat{\pi}(0,1,0) \log\left(\tfrac{\hat{\pi}(0,1,0)\hat{\pi}_1(0)}{\hat{\pi}_{12}(0,1)\hat{\pi}_{13}(0,0)}\right) \\ & + \hat{\pi}(1,0,0) \log\left(\tfrac{\hat{\pi}(1,0,0)\hat{\pi}_1(1)}{\hat{\pi}_{12}(1,0)\hat{\pi}_{13}(1,0)}\right) + \hat{\pi}(1,1,0) \log\left(\tfrac{\hat{\pi}(1,1,0)\hat{\pi}_1(1)}{\hat{\pi}_{12}(1,1)\hat{\pi}_{13}(1,0)}\right) \\ & + \hat{\pi}(0,0,1) \log\left(\tfrac{\hat{\pi}(0,0,1)\hat{\pi}_1(0)}{\hat{\pi}_{12}(0,0)\hat{\pi}_{13}(0,1)}\right) + \hat{\pi}(0,1,1) \log\left(\tfrac{\hat{\pi}(0,1,1)\hat{\pi}_1(0)}{\hat{\pi}_{12}(0,1)\hat{\pi}_{13}(0,1)}\right) \\ & + \hat{\pi}(1,0,1) \log\left(\tfrac{\hat{\pi}(1,0,1)\hat{\pi}_1(1)}{\hat{\pi}_{12}(1,0)\hat{\pi}_{13}(1,1)}\right) + \hat{\pi}(1,1,1) \log\left(\tfrac{\hat{\pi}(1,1,1)\hat{\pi}_1(1)}{\hat{\pi}_{12}(1,1)\hat{\pi}_{13}(1,1)}\right) \Big]. \end{aligned}$$

Note that, for example, $\pi_{231}(0,1,1) = \pi_{123}(1,0,1)$ and $\pi_{31}(0,1) = \pi_{13}(1,0)$. In each of these three formulae the m.l.e. of the cell probabilities under the saturated model, $\hat{\pi}(x)$, are

$$
\begin{aligned}
\hat{\pi}(0,0,0) &= \exp\{\hat{\lambda}_\emptyset\} \\
\hat{\pi}(0,0,1) &= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_3\} \\
\hat{\pi}(0,1,0) &= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_2\} \\
\hat{\pi}(0,1,1) &= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_2 + \hat{\lambda}_3 + \hat{\lambda}_{23}\} \\
\hat{\pi}(1,0,0) &= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_1\} \\
\hat{\pi}(1,0,1) &= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_1 + \hat{\lambda}_3 + \hat{\lambda}_{13}\} \\
\hat{\pi}(1,1,0) &= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_1 + \hat{\lambda}_2 + \hat{\lambda}_{12}\} \\
\hat{\pi}(1,1,1) &= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_1 + \hat{\lambda}_2 + \hat{\lambda}_3 + \hat{\lambda}_{12} + \hat{\lambda}_{13} + \hat{\lambda}_{23} + \hat{\lambda}_{123}\}
\end{aligned}
$$

$$
\begin{aligned}
\hat{\pi}_1(0) &= \exp\{\hat{\lambda}_\emptyset\} \times \left(1 + \exp\{\hat{\lambda}_2\} + \exp\{\hat{\lambda}_3\} + \exp\{\hat{\lambda}_2 + \hat{\lambda}_3 + \hat{\lambda}_{23}\}\right) \\
\hat{\pi}_1(1) &= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_1\} \times \left(1 + \exp\{\hat{\lambda}_2 + \hat{\lambda}_{12}\} + \exp\{\hat{\lambda}_3 + \hat{\lambda}_{13}\} + \exp\{\hat{\lambda}_2 + \hat{\lambda}_3 + \hat{\lambda}_{12} + \hat{\lambda}_{13} + \hat{\lambda}_{23} + \hat{\lambda}_{123}\}\right) \\
\hat{\pi}_2(0) &= \exp\{\hat{\lambda}_\emptyset\} \times \left(1 + \exp\{\hat{\lambda}_1\} + \exp\{\hat{\lambda}_3\} + \exp\{\hat{\lambda}_1 + \hat{\lambda}_3 + \hat{\lambda}_{13}\}\right) \\
\hat{\pi}_2(1) &= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_2\} \times \left(1 + \exp\{\hat{\lambda}_1 + \hat{\lambda}_{12}\} + \exp\{\hat{\lambda}_3 + \hat{\lambda}_{23}\} + \exp\{\hat{\lambda}_1 + \hat{\lambda}_3 + \hat{\lambda}_{12} + \hat{\lambda}_{13} + \hat{\lambda}_{23} + \hat{\lambda}_{123}\}\right) \\
\hat{\pi}_3(0) &= \exp\{\hat{\lambda}_\emptyset\} \times \left(1 + \exp\{\hat{\lambda}_1\} + \exp\{\hat{\lambda}_2\} + \exp\{\hat{\lambda}_1 + \hat{\lambda}_2 + \hat{\lambda}_{12}\}\right) \\
\hat{\pi}_3(1) &= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_3\} \times \left(1 + \exp\{\hat{\lambda}_1 + \hat{\lambda}_{13}\} + \exp\{\hat{\lambda}_2 + \hat{\lambda}_{23}\} + \exp\{\hat{\lambda}_1 + \hat{\lambda}_2 + \hat{\lambda}_{12} + \hat{\lambda}_{13} + \hat{\lambda}_{23} + \hat{\lambda}_{123}\}\right)
\end{aligned}
$$

$$
\begin{aligned}
\hat{\pi}_{12}(0,0) &= \exp\{\hat{\lambda}_\emptyset\} \times (1 + \exp\{\hat{\lambda}_3\}) \\
\hat{\pi}_{12}(0,1) &= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_2\} \times (1 + \exp\{\hat{\lambda}_3 + \hat{\lambda}_{23}\}) \\
\hat{\pi}_{12}(1,0) &= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_1\} \times (1 + \exp\{\hat{\lambda}_3 + \hat{\lambda}_{13}\}) \\
\hat{\pi}_{12}(1,1) &= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_1 + \hat{\lambda}_2 + \hat{\lambda}_{12}\} \times (1 + \exp\{\hat{\lambda}_3 + \hat{\lambda}_{13} + \hat{\lambda}_{23} + \hat{\lambda}_{123}\})
\end{aligned}
$$

$$
\begin{aligned}
\hat{\pi}_{13}(0,0) &= \exp\{\hat{\lambda}_\emptyset\} \times (1 + \exp\{\hat{\lambda}_2\}) \\
\hat{\pi}_{13}(0,1) &= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_3\} \times (1 + \exp\{\hat{\lambda}_2 + \hat{\lambda}_{23}\}) \\
\hat{\pi}_{13}(1,0) &= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_1\} \times (1 + \exp\{\hat{\lambda}_2 + \hat{\lambda}_{12}\}) \\
\hat{\pi}_{13}(1,1) &= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_1 + \hat{\lambda}_3 + \hat{\lambda}_{13}\} \times (1 + \exp\{\hat{\lambda}_2 + \hat{\lambda}_{12} + \hat{\lambda}_{23} + \hat{\lambda}_{123}\})
\end{aligned}
$$

$$
\begin{aligned}
\hat{\pi}_{23}(0,0) &= \exp\{\hat{\lambda}_\emptyset\} \times (1 + \exp\{\hat{\lambda}_1\}) \\
\hat{\pi}_{23}(0,1) &= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_3\} \times (1 + \exp\{\hat{\lambda}_1 + \hat{\lambda}_{13}\}) \\
\hat{\pi}_{23}(1,0) &= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_2\} \times (1 + \exp\{\hat{\lambda}_1 + \hat{\lambda}_{12}\}) \\
\hat{\pi}_{23}(1,1) &= \exp\{\hat{\lambda}_\emptyset + \hat{\lambda}_2 + \hat{\lambda}_3 + \hat{\lambda}_{23}\} \times (1 + \exp\{\hat{\lambda}_1 + \hat{\lambda}_{12} + \hat{\lambda}_{13} + \hat{\lambda}_{123}\})
\end{aligned}
$$

Since the sum of the eight cell probabilities equals one, $\lambda_\emptyset$ is a function of the remaining λ ($\lambda_1, \lambda_2, \lambda_3, \lambda_{12}, \lambda_{13}, \lambda_{23}$ and λ_{123}), i.e.,

$$
\begin{aligned}
\hat{\lambda}_\emptyset = -\log\Big(1 &+ \exp\{\hat{\lambda}_1\} + \exp\{\hat{\lambda}_2\} + \exp\{\hat{\lambda}_3\} + \exp\{\hat{\lambda}_1 + \hat{\lambda}_2 + \hat{\lambda}_{12}\} + \exp\{\hat{\lambda}_1 + \hat{\lambda}_3 + \hat{\lambda}_{13}\} \\
&+ \exp\{\hat{\lambda}_2 + \hat{\lambda}_3 + \hat{\lambda}_{23}\} + \exp\{\hat{\lambda}_1 + \hat{\lambda}_2 + \hat{\lambda}_3 + \hat{\lambda}_{12} + \hat{\lambda}_{13} + \hat{\lambda}_{23} + \hat{\lambda}_{123}\}\Big).
\end{aligned}
$$

The likelihood ratio test statistic for single edge exclusion from the saturated model can also be written as a function of the dependence ratios, as

$$LRT_{ij} = 2\, n_{\emptyset} \sum_{x_i,x_j,x_k \in \{0,1\}} \hat{\pi}_{ijk}(x_i,x_j,x_k) \log\left(\frac{\hat{\tau}_{ijk}(x_i,x_j,x_k)}{\hat{\tau}_{ik}(x_i,x_k)\,\hat{\tau}_{jk}(x_j,x_k)}\right). \tag{3.19}$$

3.4.2 Wald and score test statistics for single edge exclusion from the saturated GLL model

A general expression for the Wald and the score test statistics for single edge exclusion, in GLL models, is given in Section 2.9.3, following Smith (1990, Section 5.5). Recall that, under the null hypothesis of conditional independence, the test statistics are χ^2 distributed, the number of degrees of freedom being given by the number of parameters set to zero. The two test statistics are now written as a function of the cell probabilities, for the two and the three binary variables cases.

The Wald test statistic

In the two variables case, the Wald test statistic for single edge exclusion from the saturated model is

$$Wald_{12} = \frac{n_{\emptyset}\left[\log \hat{\psi}_{12}\right]^2}{\frac{1}{\hat{\pi}(0,0)}+\frac{1}{\hat{\pi}(0,1)}+\frac{1}{\hat{\pi}(1,0)}+\frac{1}{\hat{\pi}(1,1)}} = \frac{n_{\emptyset}\left[\log\left(\frac{\hat{\pi}(0,0)\,\hat{\pi}(1,1)}{\hat{\pi}(0,1)\,\hat{\pi}(1,0)}\right)\right]^2}{\frac{1}{\hat{\pi}(0,0)}+\frac{1}{\hat{\pi}(0,1)}+\frac{1}{\hat{\pi}(1,0)}+\frac{1}{\hat{\pi}(1,1)}}. \tag{3.20}$$

The null hypothesis is that variables 1 and 2 are independent, i.e., $\psi_{12} = 1 \Leftrightarrow \log \psi_{12} = 0 \Leftrightarrow \lambda_{12} = 0$ and the alternative hypothesis is that λ_{12} is unconstrained. An extension to the three variables case is now proposed. There are three possible Wald test statistics for single edge exclusion from the saturated model, denoted as $Wald_{12}$, $Wald_{13}$ and $Wald_{23}$. The Wald test statistic for the exclusion of edge 12 from the saturated model, with null hypothesis that both λ_{12} and λ_{123} are zero (and alternative that both are unconstrained), equals

$$Wald_{12} = n_{\emptyset} \begin{bmatrix} \hat{\lambda}_{12} \\ \hat{\lambda}_{123} \end{bmatrix}^T \begin{bmatrix} \hat{K}[3,3] & \hat{K}[7,3] \\ \hat{K}[7,3] & \hat{K}[7,7] \end{bmatrix}^{-1} \begin{bmatrix} \hat{\lambda}_{12} \\ \hat{\lambda}_{123} \end{bmatrix},$$

where $\hat{K}[3,3] = \frac{1}{\hat{\pi}(0,0,0)} + \frac{1}{\hat{\pi}(0,1,0)} + \frac{1}{\hat{\pi}(1,0,0)} + \frac{1}{\hat{\pi}(1,1,0)}$; $\hat{K}[7,3] = -\hat{K}[3,3]$ and $\hat{K}[7,7] = \hat{K}[3,3] + \frac{1}{\hat{\pi}(0,0,1)} + \frac{1}{\hat{\pi}(0,1,1)} + \frac{1}{\hat{\pi}(1,0,1)} + \frac{1}{\hat{\pi}(1,1,1)}$ (recall the inverse information matrix K is given by Equation 2.5).

Substituting in the above equation gives

$$Wald_{12} = n_\emptyset \left\{ \frac{\left(\hat{\lambda}_{12}\right)^2}{\frac{1}{\hat{\pi}(0,0,0)} + \frac{1}{\hat{\pi}(0,1,0)} + \frac{1}{\hat{\pi}(1,0,0)} + \frac{1}{\hat{\pi}(1,1,0)}} + \frac{\left(\hat{\lambda}_{12} + \hat{\lambda}_{123}\right)^2}{\frac{1}{\hat{\pi}(0,0,1)} + \frac{1}{\hat{\pi}(0,1,1)} + \frac{1}{\hat{\pi}(1,0,1)} + \frac{1}{\hat{\pi}(1,1,1)}} \right\}.$$

This simplified formula for the Wald test statistic for testing the conditional independence between variables 1 and 2 given variable 3 has two terms. Each term corresponds to a category of the binary variable 3, the variable being conditioned on. The numerators are the square of the conditional log odds ratio between variables 1 and 2 for each level of variable 3. The denominators are asymptotic variances (of the m.l.e. of the λ) for each category of variable 3.

More generally, it is suggested that the Wald test statistic for the exclusion of edge ij from the saturated GLL model (with three binary variables) is given by

$$\begin{aligned} Wald_{ij} &= n_\emptyset \left\{ \frac{\left[\log(\hat{\psi}_{ij\,.\,k=0})\right]^2}{\sum_{x_i,\, x_j \in \{0,1\}} \frac{1}{\hat{\pi}_{ijk}(x_i,x_j,x_k=0)}} + \frac{\left[\log(\hat{\psi}_{ij\,.\,k=1})\right]^2}{\sum_{x_i,\, x_j \in \{0,1\}} \frac{1}{\hat{\pi}_{ijk}(x_i,x_j,x_k=1)}} \right\} \\ &= n_\emptyset \left\{ \frac{\left(\hat{\lambda}_{ij}\right)^2}{\sum_{x_i,\, x_j \in \{0,1\}} \frac{1}{\hat{\pi}_{ijk}(x_i,x_j,x_k=0)}} + \frac{\left(\hat{\lambda}_{ij} + \hat{\lambda}_{ijk}\right)^2}{\sum_{x_i,\, x_j \in \{0,1\}} \frac{1}{\hat{\pi}_{ijk}(x_i,x_j,x_k=1)}} \right\} \end{aligned} \tag{3.21}$$

The score test statistic

From Section 2.9.3, in the two variables case, the score test statistic for single edge exclusion from the saturated model is

$$Score_{12} = n_\emptyset \, (\tilde{l}'_{12})^2 \, \tilde{K}[3,3],$$

where $\tilde{l}'_{12} = \hat{\pi}(1,1) - \hat{\pi}_1(1)\,\hat{\pi}_2(1)$ and $\tilde{K}[3,3] = \frac{1}{\hat{\pi}(0,0)} + \frac{1}{\hat{\pi}(0,1)} + \frac{1}{\hat{\pi}(1,0)} + \frac{1}{\hat{\pi}(1,1)}$. Constrained under the null hypothesis of independence between variables 1 and 2, the last term can be further simplified as

$$\begin{aligned} \tilde{K}[3,3] &= \frac{1}{\hat{\pi}_1(0)\,\hat{\pi}_2(0)} + \frac{1}{\hat{\pi}_1(0)\,\hat{\pi}_2(1)} + \frac{1}{\hat{\pi}_1(1)\,\hat{\pi}_2(0)} + \frac{1}{\hat{\pi}_1(1)\,\hat{\pi}_2(1)} \\ &= \frac{\hat{\pi}_1(1)\,\hat{\pi}_2(1) + \hat{\pi}_1(1)\,\hat{\pi}_2(0) + \hat{\pi}_1(0)\,\hat{\pi}_2(1) + \hat{\pi}_1(0)\,\hat{\pi}_2(0)}{\hat{\pi}_1(0)\,\hat{\pi}_1(1)\,\hat{\pi}_2(0)\,\hat{\pi}_2(1)} = \ldots = \frac{1}{\hat{\pi}_1(0)\,\hat{\pi}_1(1)\,\hat{\pi}_2(0)\,\hat{\pi}_2(1)}. \end{aligned}$$

Therefore,

$$Score_{12} = \frac{n_\emptyset \, [\hat{\pi}(1,1) - \hat{\pi}_1(1)\,\hat{\pi}_2(1)]^2}{\hat{\pi}_1(0)\,\hat{\pi}_1(1)\,\hat{\pi}_2(0)\,\hat{\pi}_2(1)}. \tag{3.22}$$

One should note that

$$\begin{aligned}\hat{\pi}(1,1) - \hat{\pi}_1(1)\,\hat{\pi}_2(1) &= \hat{\pi}(1,1) - [\hat{\pi}(1,1) + \hat{\pi}(1,0)]\,[\hat{\pi}(0,1) + \hat{\pi}(1,1)] \\ &= \ldots = \hat{\pi}(1,1)\,[1 - \hat{\pi}(0,1) - \hat{\pi}(1,1) - \hat{\pi}(1,0)] - \hat{\pi}(1,0)\,\hat{\pi}(0,1) \\ &= \hat{\pi}(0,0)\,\hat{\pi}(1,1) - \hat{\pi}(0,1)\,\hat{\pi}(1,0).\end{aligned}$$

Consequently, the score test statistic can also be written as

$$Score_{12} = \frac{n_\emptyset\ [\hat{\pi}(0,0)\,\hat{\pi}(1,1) - \hat{\pi}(0,1)\,\hat{\pi}(1,0)]^2}{\hat{\pi}_1(0)\,\hat{\pi}_1(1)\,\hat{\pi}_2(0)\,\hat{\pi}_2(1)},$$

which is Pearson's chi-square statistic for independence, usually written in cell counts (see Agresti, 1996, page 52), here expressed in cell probabilities.

An extension to the three binary variables case is now derived. The score test statistic for the exclusion of edge 12 from the saturated model, with null hypothesis that both λ_{12} and λ_{123} are zero (and alternative hypothesis that both are unconstrained) equals

$$Score_{12} = n_\emptyset \begin{bmatrix} \tilde{l}'_{12} \\ \tilde{l}'_{123} \end{bmatrix}^T \begin{bmatrix} \tilde{K}[3,3] & \tilde{K}[7,3] \\ \tilde{K}[7,3] & \tilde{K}[7,7] \end{bmatrix} \begin{bmatrix} \tilde{l}'_{12} \\ \tilde{l}'_{123} \end{bmatrix},$$

where

$$\begin{aligned}\tilde{K}[3,3] &= \frac{1}{\hat{\pi}(0,0,0)} + \frac{1}{\hat{\pi}(0,1,0)} + \frac{1}{\hat{\pi}(1,0,0)} + \frac{1}{\hat{\pi}(1,1,0)} \\ &= \frac{\hat{\pi}_3(0)}{\hat{\pi}_{13}(0,0)\,\hat{\pi}_{23}(0,0)} + \frac{\hat{\pi}_3(0)}{\hat{\pi}_{13}(0,0)\,\hat{\pi}_{23}(1,0)} + \frac{\hat{\pi}_3(0)}{\hat{\pi}_{13}(1,0)\,\hat{\pi}_{23}(0,0)} + \frac{\hat{\pi}_3(0)}{\hat{\pi}_{13}(1,0)\,\hat{\pi}_{23}(1,0)} \\ &= \ldots = \frac{[\hat{\pi}_3(0)]^3}{\hat{\pi}_{13}(0,0)\,\hat{\pi}_{13}(1,0)\,\hat{\pi}_{23}(0,0)\,\hat{\pi}_{23}(1,0)}\end{aligned}$$

$$\tilde{K}[7,3] = -\tilde{K}[3,3]$$

$$\begin{aligned}\tilde{K}[7,7] &= \tilde{K}[3,3] + \frac{\hat{\pi}_3(1)}{\hat{\pi}_{13}(0,1)\,\hat{\pi}_{23}(0,1)} + \frac{\hat{\pi}_3(1)}{\hat{\pi}_{13}(0,1)\,\hat{\pi}_{23}(1,1)} + \frac{\hat{\pi}_3(1)}{\hat{\pi}_{13}(1,1)\,\hat{\pi}_{23}(0,1)} + \frac{\hat{\pi}_3(1)}{\hat{\pi}_{13}(1,1)\,\hat{\pi}_{23}(1,1)} \\ &= \ldots = \frac{[\hat{\pi}_3(0)]^3}{\hat{\pi}_{13}(0,0)\,\hat{\pi}_{13}(1,0)\,\hat{\pi}_{23}(0,0)\,\hat{\pi}_{23}(1,0)} + \frac{[\hat{\pi}_3(1)]^3}{\hat{\pi}_{13}(0,1)\,\hat{\pi}_{13}(1,1)\,\hat{\pi}_{23}(0,1)\,\hat{\pi}_{23}(1,1)}\end{aligned}$$

and

$$\begin{aligned}\tilde{l}'_{12} &= \hat{\pi}_{12}(1,1) - \left[\frac{\hat{\pi}_{13}(1,0)\,\hat{\pi}_{23}(1,0)}{\hat{\pi}_3(0)} + \frac{\hat{\pi}_{13}(1,1)\,\hat{\pi}_{23}(1,1)}{\hat{\pi}_3(1)}\right] \\ \tilde{l}'_{123} &= \hat{\pi}(1,1,1) - \frac{\hat{\pi}_{13}(1,1)\,\hat{\pi}_{23}(1,1)}{\hat{\pi}_3(1)},\end{aligned}$$

once evaluated at the m.l.e. Performing the substitutions and simplifying the results gives

$$Score_{12} = n_{\emptyset} \left\{ \frac{[\hat{\pi}(1,1,0)\,\hat{\pi}_3(0) - \hat{\pi}_{13}(1,0)\,\hat{\pi}_{23}(1,0)]^2\,\hat{\pi}_3(0)}{\hat{\pi}_{13}(0,0)\,\hat{\pi}_{13}(1,0)\,\hat{\pi}_{23}(0,0)\,\hat{\pi}_{23}(1,0)} + \frac{[\hat{\pi}(1,1,1)\,\hat{\pi}_3(1) - \hat{\pi}_{13}(1,1)\,\hat{\pi}_{23}(1,1)]^2\,\hat{\pi}_3(1)}{\hat{\pi}_{13}(0,1)\,\hat{\pi}_{13}(1,1)\,\hat{\pi}_{23}(0,1)\,\hat{\pi}_{23}(1,1)} \right\}.$$

The proposed formula for the score test statistic, for testing the null hypothesis of conditional independence between variables 1 and 2 given variable 3, has also two terms, corresponding to the two categories of the variable being conditioned on. The denominators are the product of all marginal probabilities for each category of variable 3.

More generally, it is suggested that the score test statistic for the exclusion of edge ij from the saturated GLL model (with three binary variables) is given by

$$Score_{ij} = n_{\emptyset} \left\{ \frac{\left[\hat{\pi}_{ijk}(1,1,0)\,\hat{\pi}_k(0) - \hat{\pi}_{ik}(1,0)\,\hat{\pi}_{jk}(1,0)\right]^2 \hat{\pi}_k(0)}{\prod_{x_i,x_j \in \{0,1\}} \hat{\pi}_{ik}(x_i,x_{k=0})\,\hat{\pi}_{jk}(x_j,x_{k=0})} + \frac{\left[\hat{\pi}_{ijk}(1,1,1)\,\hat{\pi}_k(1) - \hat{\pi}_{ik}(1,1)\,\hat{\pi}_{jk}(1,1)\right]^2 \hat{\pi}_k(1)}{\prod_{x_i,x_j \in \{0,1\}} \hat{\pi}_{ik}(x_i,x_{k=1})\,\hat{\pi}_{jk}(x_j,x_{k=1})} \right\}. \tag{3.23}$$

3.4.3 Signed square-root versions of the tests statistics for single edge exclusion in GLL models

Only the two variables case is considered when studying the signed square-root versions of the test statistics in GLL models, since in the presence of three (or more) variables the test statistic for single edge exclusion from the saturated model depends on two (or more) parameters. Therefore, it becomes unclear for which parameter to consider the sign. One-sided tests should be used when the analyst is interested in testing for a positive association between the two variables, which corresponds to an odds ratio greater than one (and a log odds ratio greater than zero).

In GG models the parameter of interest is the partial correlation coefficient and its sign is considered. In GLL models the logarithm of the odds ratio ψ_{12} is the parameter of interest. A positive association corresponds to $\psi_{12} > 1 \Leftrightarrow \log \psi_{12} > 0$. The sign of the test statistic is negative if $\psi_{12} < 1 \Leftrightarrow \log \psi_{12} < 0$.

It is proposed that the signed square-root version of the LRT statistic for single edge exclusion from the saturated model, denoted as LRT_{12}^{sign}, is obtained as

$$LRT_{12}^{sign} = sign[\log \hat{\psi}_{12}] \sqrt{2\, n_{\emptyset} \sum_{x_1,x_2 \in \{0,1\}} \hat{\pi}_{12}(x_1,x_2) \log\left(\frac{\hat{\pi}_{12}(x_1,x_2)}{\hat{\pi}_1(x_1)\,\hat{\pi}_2(x_2)}\right)}. \tag{3.24}$$

Similarly, the signed square-root version of the Wald test statistic for single edge exclusion from the saturated model, denoted as $Wald_{12}^{sign}$, is given by

$$Wald_{12}^{sign} = \log \hat{\psi}_{12} \sqrt{\frac{n_\emptyset}{\frac{1}{\hat{\pi}(0,0)} + \frac{1}{\hat{\pi}(0,1)} + \frac{1}{\hat{\pi}(1,0)} + \frac{1}{\hat{\pi}(1,1)}}}, \tag{3.25}$$

and the signed square-root version of the score test statistic equals

$$Score_{12}^{sign} = [\hat{\pi}(1,1) - \hat{\pi}_1(1)\,\hat{\pi}_2(1)] \sqrt{\frac{n_\emptyset}{\hat{\pi}_1(0)\,\hat{\pi}_1(1)\,\hat{\pi}_2(0)\,\hat{\pi}_2(1)}}. \tag{3.26}$$

As in the Gaussian case, under the null hypothesis of independence the signed square-root versions of the three test statistics are normally distributed.

3.5 Normal Approximations to the Distributions of the Test Statistics in GLL Models with Two and Three Variables

The aim of this section is to derive asymptotic normal approximations to the distributions of the test statistics for single edge exclusion from the saturated model (LRT, Wald and score test statistics) in graphical log-linear models with two and three binary variables, under the alternative hypothesis that the saturated model holds. The delta-method is used.

3.5.1 Using the LRT statistic

A reasoning similar to that followed in Section 3.1.1 is now going to be used to obtain the asymptotic distribution of the LRT statistic for single edge exclusion from the saturated model, under the alternative hypothesis that the saturated GLL model holds, using the delta-method.

The vector of parameters of interest is now $\theta = vec(\lambda)$, and its m.l.e., based on $n_\emptyset$ observations is $\hat{\theta} = vec(\hat{\lambda})$. Recall that, since $\hat{\theta}$ is the m.l.e. of θ, it has an asymptotic normal distribution with mean θ and variance given by the inverse of the information matrix. If K is the inverse information matrix based on a single observation (as defined in Section 2.7.2), $\text{var}(\hat{\theta}) = \frac{1}{n_\emptyset} K$. Using the delta-method, if $f(\theta)$ is differentiable at $\hat{\theta}$,

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

In this chapter $f_{\bullet}^{LRT}(\hat{\theta}) = \frac{1}{n_\emptyset} LRT_{\bullet}$. For example, in the two variables case,

$$f_{12}^{LRT}(\theta) = 2 \sum_{x_1,x_2 \in \{0,1\}} \pi_{12}(x_1,x_2) \log\left(\frac{\pi_{12}(x_1,x_2)}{\pi_1(x_1)\,\pi_2(x_2)}\right),$$

since LRT_{12} is given by Equation 3.17.

The function $f_{\bullet}^{LRT}$ is differentiable at $\hat{\theta}$ provided all cell probabilities and all λ are different from zero. In the remainder of the chapter $\text{var}(LRT_{\bullet})$ and $\text{cov}(LRT_{\bullet}, LRT_{\bullet})$ denote, respectively, the variance and the covariance of the test statistics in the asymptotic distribution where, for notational convenience, $LRT_{\bullet} = n_\emptyset\, f_{\bullet}^{LRT}(\hat{\theta})$ and $LRT_{\bullet}^{sign} = \sqrt{n_\emptyset}\, f_{\bullet}^{signLRT}(\hat{\theta})$.

In the two variables case, $\hat{\theta} = vec(\hat{\lambda}) = [\hat{\lambda}_1\ \hat{\lambda}_2\ \hat{\lambda}_{12}]^T$ and $\theta = vec(\lambda) = [\lambda_1\ \lambda_2\ \lambda_{12}]^T$. Therefore, applying the delta-method, the expected value of the LRT statistic, in the asymptotic distribution, equals

$$AE[LRT_{12}] \ = 2\, n_\emptyset \textstyle\sum_{x_1,x_2 \in \{0,1\}} \pi_{12}(x_1,x_2) \log\left(\frac{\pi_{12}(x_1,x_2)}{\pi_1(x_1)\,\pi_2(x_2)}\right), \qquad (3.27)$$

and the variance in the asymptotic distribution is obtained as

$$\text{var}(LRT_{12}) = n_\emptyset\, \Delta^T\, K\, \Delta,$$

where K is given by Equation 2.4 and Δ is the vector of the derivatives of f_{12}^{LRT} with respect to λ_1, λ_2 and λ_{12}. In order to obtain such derivatives it is necessary to calculate the derivatives of each of the four cell probabilities, as well as of each of the four logarithmic terms, with respect to the three λ. Note that, $\lambda_\emptyset$ is also a function of λ_1, λ_2 and λ_{12}. All these derivatives can be expressed in cell probabilities, as shown in appendix Tables A.1 and A.2. When all the derivatives of f_{12}^{LRT} with respect to the λ are written as a function of the cell probabilities, the vector Δ is of the following form

$$\Delta = \begin{bmatrix} \partial f_{12}^{LRT}/\partial\lambda_1 \\ \partial f_{12}^{LRT}/\partial\lambda_2 \\ \partial f_{12}^{LRT}/\partial\lambda_{12} \end{bmatrix} =$$

$$= \begin{bmatrix} \frac{-\pi_1(1)}{n_\emptyset} AE[LRT_{12}] + 2\left[\pi(1,0)\log\left(\frac{\pi(1,0)}{\pi_1(1)\pi_2(0)}\right) + \pi(1,1)\log\left(\frac{\pi(1,1)}{\pi_1(1)\pi_2(1)}\right)\right] \\ \frac{-\pi_2(1)}{n_\emptyset} AE[LRT_{12}] + 2\left[\pi(0,1)\log\left(\frac{\pi(0,1)}{\pi_1(0)\pi_2(1)}\right) + \pi(1,1)\log\left(\frac{\pi(1,1)}{\pi_1(1)\pi_2(1)}\right)\right] \\ \frac{-\pi(1,1)}{n_\emptyset} AE[LRT_{12}] + 2\left[\pi(1,1)\log\left(\frac{\pi(1,1)}{\pi_1(1)\pi_2(1)}\right)\right] \end{bmatrix}. \qquad (3.28)$$

Multiplying $n_\emptyset\ \Delta^T\ K\ \Delta$ and simplifying the resulting expression by writing it as a function of cell probabilities, π, gives the following result

$$\begin{aligned}\mathrm{var}(LRT_{12}) = 4\,n_\emptyset \Big[&\pi(0,0)\log^2\left(\tfrac{\pi(0,0)}{\pi_1(0)\pi_2(0)}\right) + \pi(0,1)\log^2\left(\tfrac{\pi(0,1)}{\pi_1(0)\pi_2(1)}\right) \\ &+\pi(1,0)\log^2\left(\tfrac{\pi(1,0)}{\pi_1(1)\pi_2(0)}\right) + \pi(1,1)\log^2\left(\tfrac{\pi(1,1)}{\pi_1(1)\pi_2(1)}\right)\Big] \\ &-\tfrac{1}{n_\emptyset}\,(AE[LRT_{12}])^2 \\ = 4\,n_\emptyset \textstyle\sum_{x_1,x_2\in\{0,1\}} &\pi_{12}(x_1,x_2)\log^2\left(\tfrac{\pi_{12}(x_1,x_2)}{\pi_1(x_1)\,\pi_2(x_2)}\right) - \tfrac{1}{n_\emptyset}\,(AE[LRT_{12}])^2.\end{aligned} \tag{3.29}$$

In the three variables case, $\hat{\theta} = vec(\hat{\lambda}) = [\hat{\lambda}_1\ \hat{\lambda}_2\ \hat{\lambda}_{12}\ \hat{\lambda}_3\ \hat{\lambda}_{13}\ \hat{\lambda}_{23}\ \hat{\lambda}_{123}]^T$ and $\theta = vec(\lambda) = [\lambda_1\ \lambda_2\ \lambda_{12}\ \lambda_3\ \lambda_{13}\ \lambda_{23}\ \lambda_{123}]^T$. Applying the delta-method, the expected values for each of the three LRT statistics for single edge exclusion from the saturated model, in the asymptotic distribution, are given by

$$AE[LRT_{ij}] = 2\,n_\emptyset \sum_{x_i,x_j,x_k\in\{0,1\}} \pi_{ijk}(x_i,x_j,x_k)\,\log\left(\frac{\pi_{ijk}(x_i,x_j,x_k)\,\pi_k(x_k)}{\pi_{ik}(x_i,x_k)\,\pi_{jk}(x_j,x_k)}\right). \tag{3.30}$$

The variance matrix of the three likelihood ratio test statistics, for single edge exclusion from the saturated GLL model, is a 3×3 symmetric matrix and equals $n_\emptyset\ \Delta^T\ K\ \Delta$, where K is given by Equation 2.5, and Δ is a 7×3 matrix, having in each column the derivatives of f_{ij}^{LRT} with respect to the seven λ. The derivatives of $\lambda_\emptyset$, as well as those of each of the eight cell probabilities, and those of the logarithmic terms of each of the f_{ij}^{LRT}, with respect to the vector of seven λ, once expressed in cell probabilities, are presented in appendix Tables A.3, A.4 and A.5. So are the various elements of the matrix Δ, written as a function of the cell probabilities (Tables A.6 and A.7).

In conclusion: it is proposed that the variance (in the asymptotic distribution) of the LRT statistic for the exclusion of edge ij (with i and j distinct, from 1 to 3) from the saturated GLL model is expressed in cell probabilities, as

$$\begin{aligned}\mathrm{var}(LRT_{ij}) = 4\,n_\emptyset \textstyle\sum_{x_i,x_j,x_k\in\{0,1\}} &\pi_{ijk}(x_i,x_j,x_k)\,\log^2\left(\tfrac{\pi_{ijk}(x_i,x_j,x_k)\,\pi_k(x_k)}{\pi_{ik}(x_i,x_k)\,\pi_{jk}(x_j,x_k)}\right) \\ &-\tfrac{1}{n_\emptyset}(AE[LRT_{ij}])^2.\end{aligned} \tag{3.31}$$

Analogously, the covariance (in the asymptotic distribution) between the LRT statistic for the exclusion of edge ij and the LRT statistic for the exclusion of edge ik (with i, j and k distinct, from 1 to 3) from the saturated GLL model is expressed in cell probabilities, as

$$\begin{aligned}
&\mathrm{cov}(LRT_{ij}, LRT_{ik}) = \\
&= 4\, n_\emptyset \sum_{x_i,x_j,x_k \in \{0,1\}} \left[\pi_{ijk}(x_i,x_j,x_k) \log\left(\frac{\pi_{ijk}(x_i,x_j,x_k)\,\pi_k(x_k)}{\pi_{ik}(x_i,x_k)\,\pi_{jk}(x_j,x_k)}\right) \log\left(\frac{\pi_{ijk}(x_i,x_j,x_k)\,\pi_j(x_j)}{\pi_{ij}(x_i,x_j)\,\pi_{kj}(x_k,x_j)}\right)\right] \\
&\quad -\tfrac{1}{n_\emptyset}\,(AE[LRT_{ij}])\,(AE[LRT_{ik}]).
\end{aligned} \tag{3.32}$$

A small simulation study was performed, validating the formulae presented for the variance structure (in the asymptotic distribution) of the three test statistics for single edge exclusion from the saturated model. One thousand contingency tables with a certain cell probability structure were generated. Two different sample sizes, $n_\emptyset$, were used: 1 000 and 10 000. For each of the tables, the values of the three likelihood ratio test statistics for single edge exclusion from the saturated model were calculated and stored. Then, for each sample size, the variance matrices of the one thousand values of the three test statistics were obtained.

The vector of cell probabilities used, in standard cell order, is

$$\pi(x) = [0.05\ 0.05\ 0.1\ 0.1\ 0.2\ 0.1\ 0.15\ 0.25]^T.$$

The theoretical values obtained using the six formulae proposed above can be compared to the simulated values:

	theoretical values	simulated values	**theoretical** values	simulated values
	$n_\emptyset = 1\,000$	$n_\emptyset = 1\,000$	$n_\emptyset = 10\,000$	$n_\emptyset = 10\,000$
$\mathrm{var}(LRT_{12})$	**98.400**	102.792	**984.004**	991.077
$\mathrm{var}(LRT_{13})$	**69.317**	74.935	**693.170**	682.474
$\mathrm{var}(LRT_{23})$	**227.909**	224.088	**2 279.089**	2 308.754
$\mathrm{cov}(LRT_{12}, LRT_{13})$	**68.042**	70.386	**680.421**	677.314
$\mathrm{cov}(LRT_{12}, LRT_{23})$	**68.263**	71.101	**682.630**	692.646
$\mathrm{cov}(LRT_{13}, LRT_{23})$	**66.516**	69.199	**665.164**	651.256

Additionally, for both sample sizes, ten batches of 1 000 repetitions each were performed, and for each of them the variances and covariances of the three test statistics were calculated and stored. Finally, 95 and 99% confidence intervals for the mean variances and mean covariances were obtained, as follows:

CI for μ	95% $n_\emptyset = 1\,000$	99% $n_\emptyset = 1\,000$	95% $n_\emptyset = 10\,000$	99% $n_\emptyset = 10\,000$
var (LRT_{12})	(99.79; 108.40)	**(97.91; 110.28)**	**(946.03; 1 019.89)**	**(929.91; 1 036.01)**
var (LRT_{13})	(72.03; 78.0)	(70.73; 79.31)	**(675.36; 716.28)**	**(666.43; 725.21)**
var (LRT_{23})	**(224.53; 246.25)**	**(219.79; 250.99)**	**(2 242.27; 2 334.95)**	**(2 222.03; 2 355.19)**
cov $(LRT_{12},\ LRT_{13})$	(68.52; 74.17)	**(67.29; 75.40)**	**(653.98; 703.54)**	**(643.16; 714.36)**
cov $(LRT_{12},\ LRT_{23})$	**(65.73; 73.30)**	**(64.08; 74.96)**	**(624.48; 723.71)**	**(602.81; 745.37)**
cov $(LRT_{13},\ LRT_{23})$	**(65.27; 73.22)**	**(63.53; 74.95)**	**(629.44; 698.59)**	**(614.35; 713.69)**

The proposed theoretical values lie inside the boundaries of the confidence intervals, not only at a 99% confidence level, but also at 95%, when the sample size is big $n_\emptyset = 10\,000$. For a sample size of 1 000 theoretical values are inside the limits of the confidence intervals at a 99% confidence level, but some values are outside when a 95% confidence level is considered. These results confirm that the proposed formulae for the variances and the covariances of the likelihood ratio test statistics, for single edge exclusion from the saturated model, hold asymptotically.

In conclusion: the LRT statistic for single edge exclusion from a saturated GLL model with two or three binary variables, under the alternative hypothesis that the saturated model holds, has an asymptotic normal distribution with means given by Equations 3.27 and 3.30, and variances and covariances given by Equations 3.29, 3.31 and 3.32.

3.5.2 Using the Wald and the score test statistics

In the case the Wald test statistic is used, with two variables,

$$f_{12}^{Wald} = \frac{\left[\log\left(\frac{\pi(0,0)\,\pi(1,1)}{\pi(0,1)\,\pi(1,0)}\right)\right]^2}{\frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)} + \frac{1}{\pi(1,0)} + \frac{1}{\pi(1,1)}} =$$

$$= \lambda_{12}^2 \;/\; [\exp\{-\lambda_\emptyset\}\ (1 + \exp\{-\lambda_2\} + \exp\{-\lambda_1\} + \exp\{-\lambda_1 - \lambda_2 - \lambda_{12}\})].$$

Applying the delta-method, the expected value of the Wald test statistic, in the asymptotic distribution, equals

$$AE[Wald_{12}] = \frac{n_\emptyset \left[\log\left(\frac{\pi(0,0)\,\pi(1,1)}{\pi(0,1)\,\pi(1,0)}\right)\right]^2}{\frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)} + \frac{1}{\pi(1,0)} + \frac{1}{\pi(1,1)}}, \tag{3.33}$$

and the variance in the asymptotic distribution is obtained as

$$\text{var}(Wald_{12}) = n_\emptyset \, \Delta^T \, K \, \Delta,$$

where K is given by Equation 2.4 and Δ is the vector of the derivatives of f_{12}^{Wald} with respect to λ_1, λ_2 and λ_{12}, expressed in cell probabilities as follows

$$\Delta = \begin{bmatrix} \partial f_{12}^{Wald}/\partial\lambda_1 \\ \partial f_{12}^{Wald}/\partial\lambda_2 \\ \partial f_{12}^{Wald}/\partial\lambda_{12} \end{bmatrix} =$$

$$= \begin{bmatrix} \frac{1}{n_\emptyset} AE[Wald_{12}] \left[-\pi_1(1) + \frac{\frac{1}{\pi(1,0)} + \frac{1}{\pi(1,1)}}{\frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)} + \frac{1}{\pi(1,0)} + \frac{1}{\pi(1,1)}} \right] \\ \frac{1}{n_\emptyset} AE[Wald_{12}] \left[-\pi_2(1) + \frac{\frac{1}{\pi(0,1)} + \frac{1}{\pi(1,1)}}{\frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)} + \frac{1}{\pi(1,0)} + \frac{1}{\pi(1,1)}} \right] \\ \frac{1}{n_\emptyset} AE[Wald_{12}] \left[-\pi(1,1) + \frac{\frac{1}{\pi(1,1)}}{\frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)} + \frac{1}{\pi(1,0)} + \frac{1}{\pi(1,1)}} \right] + \frac{2 \log \psi_{12}}{\frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)} + \frac{1}{\pi(1,0)} + \frac{1}{\pi(1,1)}} \end{bmatrix}. \tag{3.34}$$

It is suggested that, in the two variables case, the variance of the Wald test statistic, in the asymptotic distribution, simplifies to

$$\begin{aligned} \text{var}(Wald_{12}) = \; & 4\, AE[Wald_{12}] \left[1 + \frac{\log \psi_{12} \left[\frac{1}{(\pi(0,0))^2} - \frac{1}{(\pi(0,1))^2} - \frac{1}{(\pi(1,0))^2} + \frac{1}{(\pi(1,1))^2} \right]}{\left(\frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)} + \frac{1}{\pi(1,0)} + \frac{1}{\pi(1,1)} \right)^2} \right] \\ & + \frac{1}{n_\emptyset} \left(AE[Wald_{12}] \right)^2 \left[\frac{\frac{1}{(\pi(0,0))^3} + \frac{1}{(\pi(0,1))^3} + \frac{1}{(\pi(1,0))^3} + \frac{1}{(\pi(1,1))^3}}{\left(\frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)} + \frac{1}{\pi(1,0)} + \frac{1}{\pi(1,1)} \right)^2} - 1 \right]. \end{aligned} \tag{3.35}$$

In the three variables case, when applying the delta-method, the expected values, in the asymptotic distribution, for each of the three Wald test statistics for single edge exclusion from the saturated model are given by

$$AE[Wald_{ij}] = n_\emptyset \left\{ \frac{\left[\log(\hat{\psi}_{ij.k=0}) \right]^2}{\sum_{x_i, x_j \in \{0,1\}} \frac{1}{\pi_{ijk}(x_i, x_j, x_k=0)}} + \frac{\left[\log(\hat{\psi}_{ij.k=1}) \right]^2}{\sum_{x_i, x_j \in \{0,1\}} \frac{1}{\pi_{ijk}(x_i, x_j, x_k=1)}} \right\} = f_{ij}^{Wald}. \tag{3.36}$$

The variance matrix, in the asymptotic distribution of the three Wald test statistics for single edge exclusion from the saturated model, is a 3×3 symmetric matrix and equals $n_\emptyset \, \Delta^T \, K \, \Delta$, where K is given by Equation 2.5, and Δ is a 7×3 matrix, having in

each column the derivatives of f_{ij}^{Wald} with respect to the seven λ. Each of the elements of the matrix Δ can be written as a function of the cell probabilities, as presented in appendix Tables A.8 to A.10.

It was not possible to obtain a nice simplified formula for the variances and covariances of the three Wald test statistics. For further calculations the several pages long formulae produced by MAPLE were used.

In the case of the score test statistic, with two variables,

$$f_{12}^{Score} = \frac{[\pi(1,1)-\pi_1(1)\,\pi_2(1)]^2}{\pi_1(0)\,\pi_1(1)\,\pi_2(0)\,\pi_2(1)} = \frac{[\pi(1,1)-\pi_1(1)\,\pi_2(1)]^2}{[1-\pi_1(1)]\,\pi_1(1)\,[1-\pi_2(1)]\,\pi_2(1)}.$$

Expressing directly f_{12}^{Score} as a function of the three λ and differentiating it with respect to λ_1, λ_2 and λ_{12} was not easy because of the complicated expressions that were obtained. The decision was made to obtain such derivatives indirectly, by differentiating f_{12}^{Score} with respect to the π and then the π with respect to the λ, i.e.,

$$\frac{\partial f_{12}^{Score}}{\partial \lambda} = \frac{\partial \pi}{\partial \lambda}\,\frac{\partial f_{12}^{Score}}{\partial \pi},$$

where π and λ are 3×1 vectors with $\pi_1(1), \pi_2(1), \pi(1,1)$ and $\lambda_1, \lambda_2, \lambda_{12}$, respectively. For example,

$$\frac{\partial f_{12}^{Score}}{\partial \lambda_1} = \frac{\partial \pi_1(1)}{\partial \lambda_1}\,\frac{\partial f_{12}^{Score}}{\partial \pi_1(1)} + \frac{\partial \pi_2(1)}{\partial \lambda_1}\,\frac{\partial f_{12}^{Score}}{\partial \pi_2(1)} + \frac{\partial \pi(1,1)}{\partial \lambda_1}\,\frac{\partial f_{12}^{Score}}{\partial \pi(1,1)}.$$

Analogously for the derivatives with respect to λ_2 and λ_{12}. After some calculations these derivatives can be expressed in cell probabilities, as follows

$$\frac{\partial f_{12}^{Score}}{\partial \pi_1(1)} = \left[\frac{-2\{\pi(1,1)-\pi_1(1)\,\pi_2(1)\}}{\pi_1(1)\,[1-\pi_1(1)]\,[1-\pi_2(1)]}\right] - f_{12}^{Score}\,\frac{1-2\pi_1(1)}{\pi_1(1)\,[1-\pi_1(1)]}$$

$$\frac{\partial f_{12}^{Score}}{\partial \pi_2(1)} = \left[\frac{-2\{\pi(1,1)-\pi_1(1)\,\pi_2(1)\}}{[1-\pi_1(1)]\,\pi_2(1)\,[1-\pi_2(1)]}\right] - f_{12}^{Score}\,\frac{1-2\pi_2(1)}{\pi_2(1)\,[1-\pi_2(1)]}$$

$$\frac{\partial f_{12}^{Score}}{\partial \pi(1,1)} = 2\,\frac{\pi_{12}-\pi_1(1)\,\pi_2(1)}{\pi_1(1)\,[1-\pi_1(1)]\,\pi_2(1)\,[1-\pi_2(1)]}$$

Consequently,

$$\Delta = \begin{bmatrix} \partial f_{12}^{Score}/\partial\lambda_1 \\ \partial f_{12}^{Score}/\partial\lambda_2 \\ \partial f_{12}^{Score}/\partial\lambda_{12} \end{bmatrix} =$$

$$= \begin{bmatrix} \frac{1}{n_\emptyset} AE[Score_{12}] \left[1 - 2\pi_1(1) - \frac{\{\pi(1,1) - \pi_1(1)\,\pi_2(1)\}\,[1-2\pi_2(1)]}{\pi_2(1)\,[1-\pi_2(1)]}\right] \\ \frac{1}{n_\emptyset} AE[Score_{12}] \left[1 - 2\pi_2(1) - \frac{\{\pi(1,1) - \pi_1(1)\,\pi_2(1)\}\,[1-2\pi_1(1)]}{\pi_1(1)\,[1-\pi_1(1)]}\right] \\ \frac{1}{n_\emptyset} AE[Score_{12}]\,\pi(1,1) \left[4 - \frac{1}{\pi_1(1)} - \frac{1}{\pi_2(1)} + \frac{2\,\{1-\pi_1(1)-\pi_2(1)-\pi(1,1)+2\,\pi_1(1)\,\pi_2(1)\}}{\pi(1,1)-\pi_1(1)\,\pi_2(1)}\right] \end{bmatrix}. \tag{3.37}$$

Applying the delta-method, the expected value of the score test statistic, in the asymptotic distribution, simplifies to

$$AE[Score_{12}] = \frac{n_\emptyset\,[\pi(1,1) - \pi_1(1)\,\pi_2(1)]^2}{\pi_1(0)\,\pi_1(1)\,\pi_2(0)\,\pi_2(1)}, \tag{3.38}$$

and the variance, in the asymptotic distribution, is obtained as

$$\text{var}(Score_{12}) = n_\emptyset\,\Delta^T\,K\,\Delta,$$

where K is given by Equation 2.4 and Δ is given by Equation 3.37. Calculations where made by hand and using MAPLE. However, it was not possible to obtain a nice simplified formula for the variance of the score test statistic (as had been in the case of the LRT and the Wald test statistics). For that reason a final expression for the variance is not presented here. Additionally, the decision was made not to consider the three variables case, due to the complexity already existent in the two variables case. Hence, the specific calculations for obtaining the variance of the score test statistic, in the three variables case, were not performed. Yet, the expected values, in the asymptotic distribution, for each of the three score test statistics for single edge exclusion from the saturated model are given by

$$AE[Score_{ij}]$$

$$= n_\emptyset \left\{ \frac{\left[\pi_{ijk}(1,1,0)\,\pi_k(0) - \pi_{ik}(1,0)\,\pi_{jk}(1,0)\right]^2 \pi_k(0)}{\prod_{x_i,x_j\in\{0,1\}} \pi_{ik}(x_i,x_{k=0})\,\pi_{jk}(x_j,x_{k=0})} + \frac{\left[\pi_{ijk}(1,1,1)\,\pi_k(1) - \pi_{ik}(1,1)\,\pi_{jk}(1,1)\right]^2 \pi_k(1)}{\prod_{x_i,x_j\in\{0,1\}} \pi_{ik}(x_i,x_{k=1})\,\pi_{jk}(x_j,x_{k=1})} \right\}. \tag{3.39}$$

3.5.3 Using signed square-root versions of the test statistics, in the two variables case

The aim of this section is to derive the asymptotic distributions of the signed square-root versions of the test statistics for single edge exclusion from the saturated GLL model, in the two variables case, under the alternative hypothesis of non-independence. The delta-method is used, as in previous sections.

In the case of the signed square-root version of the LRT statistic,

$$f_{12}^{signLRT} = sign[\log \psi_{12}] \sqrt{2 \sum_{x_1,x_2 \in \{0,1\}} \pi_{12}(x_1,x_2) \log \left(\frac{\pi_{12}(x_1,x_2)}{\pi_1(x_1)\,\hat{\pi}_2(x_2)} \right)}.$$

In the case of the Wald test statistic,

$$f_{12}^{signWald} = \log \psi_{12} \sqrt{\frac{1}{\frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)} + \frac{1}{\pi(1,0)} + \frac{1}{\pi(1,1)}}},$$

and in the case of the score test statistic

$$f_{12}^{signScore} = [\pi(1,1) - \pi_1(1)\,\pi_2(1)] \sqrt{\frac{1}{\pi_1(0)\,\pi_1(1)\,\pi_2(0)\,\pi_2(1)}}.$$

Using the delta-method, and under the alternative hypothesis that the saturated model holds, the signed square-root versions of the three test statistics for single edge exclusion, in GLL models, are asymptotically normal distributed:

$$LRT_{12}^{sign} \xrightarrow{\mathcal{D}} N\left(AE[LRT_{12}^{sign}], var(LRT_{12}^{sign})\right),$$

$$Wald_{12}^{sign} \xrightarrow{\mathcal{D}} N\left(AE[Wald_{12}^{sign}], var(Wald_{12}^{sign})\right),$$

$$Score_{12}^{sign} \xrightarrow{\mathcal{D}} N\left(AE[Score_{12}^{sign}], var(Score_{12}^{sign})\right),$$

where

$$AE[LRT_{12}^{sign}] = \sqrt{n_\emptyset}\, f_{12}^{signLRT}, \quad AE[Wald_{12}^{sign}] = \sqrt{n_\emptyset}\, f_{12}^{signWald},$$
$$AE[Score_{12}^{sign}] = \sqrt{n_\emptyset}\, f_{12}^{signScore}.$$

As in Section 3.1.5, the variance of the signed square-root versions of the test statistics can be obtained by applying the delta-method a second time, the new Δ matrix having the derivative of the signed square-root version of the test statistic being considered with respect to the non signed one. For example, in the case of the LRT,

$$\frac{\partial f_{12}^{signLRT}}{\partial f_{12}^{LRT}} = \frac{\partial \left(sign[\log \psi_{12}] \sqrt{f_{12}^{LRT}} \right)}{\partial f_{12}^{LRT}} = sign[\log \psi_{12}] \frac{1}{2\sqrt{f_{12}^{LRT}}}.$$

Hence,

$$var(LRT_{12}^{sign}) = var(LRT_{12}) \frac{1}{4\,AE[LRT_{12}]}, \tag{3.40}$$

where $var(LRT_{12})$ is given by Equation 3.29 and $AE[LRT_{12}]$ is given by Equation 3.27. Analogously,

$$var(Wald_{12}^{sign}) = var(Wald_{12}) \frac{1}{4\,AE[Wald_{12}]}, \tag{3.41}$$

where $var(Wald_{12})$ is given by Equation 3.35 and $AE[Wald_{12}]$ is given by Equation 3.33 and

$$var(Score_{12}^{sign}) = var(Score_{12}) \frac{1}{4\,AE[Score_{12}]}, \tag{3.42}$$

where $AE[Score_{12}]$ is given by Equation 3.38.

In brief: asymptotic normal approximations have been derived to the distributions of the non-signed and the signed square-root versions of the three test statistics for single edge exclusion from the saturated GLL model. The possibility of an alternative approximation is studied in the next section.

3.6 Non-central χ^2 Approximation to the Distribution of the LRT in a GLL Model with Two Variables

Similarly to what was undertaken in Section 3.2, the current section proposes a non-central chi-square approximation to the distribution of the likelihood ratio test statistic for single edge exclusion from the saturated GLL model with two binary variables, at a local alternative. Section 3.7 compares the quality of the non-central chi-square approximation to that of the normal approximation derived in Section 3.5.1.

A brief explanation on how to obtain the non-centrality parameter φ follows. The null hypothesis is that $\lambda_{12} = 0 \Leftrightarrow \psi_{12} = 1 \Leftrightarrow \log\psi_{12} = 0$. Consequently, the distance from the null is measured as $\delta = \sqrt{n_\emptyset}\,(\log\psi_{12} - 0)$. The non-centrality parameter equals

$$\varphi = \delta^T \ (K[3,3])^{-1} \ \delta,$$

where $K[3,3]$ is the element of the inverse information matrix associated with the restriction in H_0, i.e., associated with λ_{12}. From Section 2.7.2, $K[3,3] = \frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)} +$

$\frac{1}{\pi(1,0)} + \frac{1}{\pi(1,1)}$. Therefore,

$$\varphi = \frac{n_\emptyset \log^2 \psi_{12}}{\frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)} + \frac{1}{\pi(1,0)} + \frac{1}{\pi(1,1)}} = \left[\frac{AE[\log \hat{\psi}_{12}]}{ASE(\log \hat{\psi}_{12})}\right]^2. \tag{3.43}$$

Note that φ can also be obtained as the square of the ratio of the mean of the log transform of the sample odds ratio, in the asymptotic distribution, over its asymptotic standard error. Also note that the non-centrality parameter equals the mean of the Wald test statistic in the asymptotic normal distribution, i.e., $\varphi = AE[Wald_{12}]$. Although the three variables case is not considered, it is possible to conclude that the non-centrality parameter of a non-central chi-square approximation to the LRT statistic, under the alternative hypothesis that edge ij is present in the model, equals $AE[Wald_{ij}]$, given by Equation 3.36.

3.7 Quality of the Two Approximations, in GLL Models

In the two binary variables case, under the null hypothesis of independence ($\lambda_{12} = 0 \Leftrightarrow \psi_{12} = 1$), the asymptotic distribution of the likelihood ratio test statistic for single edge exclusion from the saturated GLL model is χ^2_1. Under the alternative hypothesis that the saturated model holds, (using the delta-method) the asymptotic distribution of the likelihood ratio test tends to the normal distribution, as derived in Section 3.5.1, as $n_\emptyset$ tends to infinity. At $\lambda_{12} = 0$ (or $\psi_{12} = 1$) the asymptotic distribution is degenerate, with mean zero and variance zero. The normal approximation holds for $n_\emptyset$ at infinity, but it is poor for $n_\emptyset$ finite.

As in Section 3.3 (for GG models), a simulation study was used to assess the quality of the two proposed approximations (normal and non-central chi-square) to the distribution of the LRT statistic for single edge exclusion from the saturated GLL model with two binary variables, for various values of $n_\emptyset$. First, a brief explanation on how the simulation study was conducted is given. Then, the histograms of the empirical distribution of the LRT statistic are plotted, with a normal curve overlapping, corresponding to the proposed theoretical normal approximation. Two different sample sizes and several combinations of odds ratios and marginal probabilities are used. The quality of the normal approximation is assessed. Finally, p-p plots are produced,

allowing the comparison between the quality of the normal and of the non-central chi-square approximations. Some guidelines are given as to when each of the two approximations performs better.

The parameters for the simulation study were chosen as follows. In a 2×2 contingency table there are four cell probabilities that add up to one, therefore three unknowns. Alternatively, one can consider two fixed marginal probability values, say $\pi_1(0)$ and $\pi_2(0)$, and a given odds ratio value (ψ_{12}) and obtain the corresponding cell probabilities. Indeed, for a given combination of $\pi_1(0)$, $\pi_2(0)$ and ψ_{12} the four cell probabilities can be written as: $\pi(0,1) = \pi_1(0) - \pi(0,0)$; $\pi(1,0) = \pi_2(0) - \pi(0,0)$; $\pi(1,1) = 1 - \pi_1(0) - \pi_2(0) + \pi(0,0)$. If $\psi_{12} = 1$, then $\pi(0,0) = \pi_1(0)\pi_2(0)$. If $\psi_{12} \neq 1$, since

$$\psi_{12} = \frac{\pi(0,0)\pi(1,1)}{\pi(0,1)\pi(1,0)} = \frac{\pi(0,0)\,(1 - \pi_1(0) - \pi_2(0) + \pi(0,0))}{(\pi_1(0) - \pi(0,0))\,(\pi_2(0) - \pi(0,0))},$$

$$\pi(0,0) = \frac{\pi_1(0) + \pi_2(0)}{2} - \frac{1}{2(1-\psi_{12})} + \frac{\sqrt{[(1-\psi_{12})\,(\pi_1(0) + \pi_2(0)) - 1]^2 + 4\,(1-\psi_{12})\,[\psi_{12}\,\pi_1(0)\pi_2(0)]}}{2(1-\psi_{12})}.$$

The decision was made to fix the values of $\pi_1(0)$, $\pi_2(0)$ and ψ_{12}.

Because of the way the two marginal probabilities were chosen, there are several types of symmetries that can be taken into account when using $\pi_1(0)$, $\pi_2(0)$ and ψ_{12} to calculate either the likelihood ratio test statistic (from Equation 3.17) or the probability of selecting the saturated model (Section 4.2.1). In brief:

- symmetry due to the swapping of $\pi_1(0)$ and $\pi_2(0)$: for a given ψ_{12}, the value of LRT_{12} is the same for $\pi_1(0) = x$; $\pi_2(0) = y$ and for $\pi_1(0) = y$; $\pi_2(0) = x$;
- symmetry related to complementary marginal probability values: for a given ψ_{12}, the value of LRT_{12} is the same for $\pi_1(0) = x$; $\pi_2(0) = y$ and for $\pi_1(0) = 1 - y$; $\pi_2(0) = 1 - x$;
- symmetry when one of the marginal probabilities equals 0.5: for a given ψ_{12}, the value of LRT_{12} is the same for $\pi_1(0) = 0.5$; $\pi_2(0) = y$ and $\pi_1(0) = 0.5$; $\pi_2(0) = 1 - y$; or for $\pi_1(0) = x$; $\pi_2(0) = 0.5$ and $\pi_1(0) = 1 - x$; $\pi_2(0) = 0.5$.

Besides, there is a 'reciprocal (inverse)' symmetry due to the value of the odds ratio. Indeed, the value of LRT_{12} is the same for $\psi_{12} = z$, $\pi_1(0) = x$, $\pi_2(0) = y$ and for $\psi_{12} = 1/z$, $\pi_1(0) = 1 - x$, $\pi_2(0) = y$.

Due to these symmetries, only some combinations of marginal probabilities and odds ratio are considered. ψ_{12} takes values from 1 to 4, $\pi_1(0)$ from 0.1 to 0.9 and $\pi_2(0)$ from 0.1 to 0.2. Figure 3.4 shows histograms of the empirical distribution of the

likelihood ratio test statistic, for the chosen combinations of marginal cell probabilities and odds ratio values. The sample size equals 1 000 observations. Each of the eight rows corresponds to a different combination of cell marginal probabilities $\pi_1(0)$ and $\pi_2(0)$: from top to bottom of the plot, 0.1 and 0.1; 0.3 and 0.1; 0.5 and 0.1; 0.7 and 0.1; 0.9 and 0.1; 0.3 and 0.2; 0.5 and 0.2; and 0.7 and 0.2. Two different colours are used to highlight a change in marginal probabilities. Each of the six columns corresponds to a different odds ratio value: from left to right, $\psi_{12} = 1, 1.5, 2, 2.5, 3$ and 4. In each histogram the horizontal axis corresponds to the range of values obtained for LRT_{12} for the combination of cell probabilities used. In each plot there is a blue curve overlapping: it represents the density of a normal distribution with mean and variance given by Equations 3.27 and 3.29. The aim is to detect how close the simulated values are to the theoretical normal approximation proposed. The green line in the first column plots represents the density of a chi-square distribution on one degree of freedom. The histograms show that, as expected, the distribution of LRT_{12} looks chi-squared on one degree of freedom for $\psi_{12} = 1$. For $\psi_{12} \neq 1$ the distribution of LRT_{12} tends to the normal as the odds ratio increases, faster for less unbalanced combinations of marginal probabilities (as happens for $\pi_1(0) = 0.3$ or 0.5 and $\pi_2(0) = 0.2$). Indeed, for less unbalanced combinations of marginal probabilities, normality seems to be achieved for an odds ratio value around 2. An odds ratio of at least 2.5 seems to be required for more unbalanced combinations of $\pi_1(0)$ and $\pi_2(0)$.

It is worth taking into account the minimum expected counts in each cell, for the different combinations of marginal probabilities and odds ratio values, with a sample size of 1 000. A minimum expected cell count of 10 is obtained with an odds ratio of 1 and $\pi_1(0) = \pi_2(0) = 0.1$. When $\pi_2(0) = 0.1$ and $\pi_1(0) = 0.9$ a minimum expected cell count of 7.1 is obtained with $\psi_{12} = 1.5$, whereas values of ψ_{12} of 3 and 4 lead to minimum expected cell counts of 3.8 and 2.9, respectively. For this reason odds ratio values greater than 4 were not considered in the simulation study. For more balanced combinations of marginal probabilities, for example 0.3 and 0.2, or 0.5 and 0.2, the minimum expected cell counts equal 60 (for $\psi_{12} = 1$) and 48.86 (for $\psi_{12} = 4$), respectively. In brief: the most unbalanced combinations of marginal probabilities are the ones that should be looked at more carefully when drawing conclusions, since they are more likely to correspond to situations of very small expected cell frequencies. Expected cell counts associated with the different combinations of $\pi_1(0)$, $\pi_2(0)$ and ψ_{12}

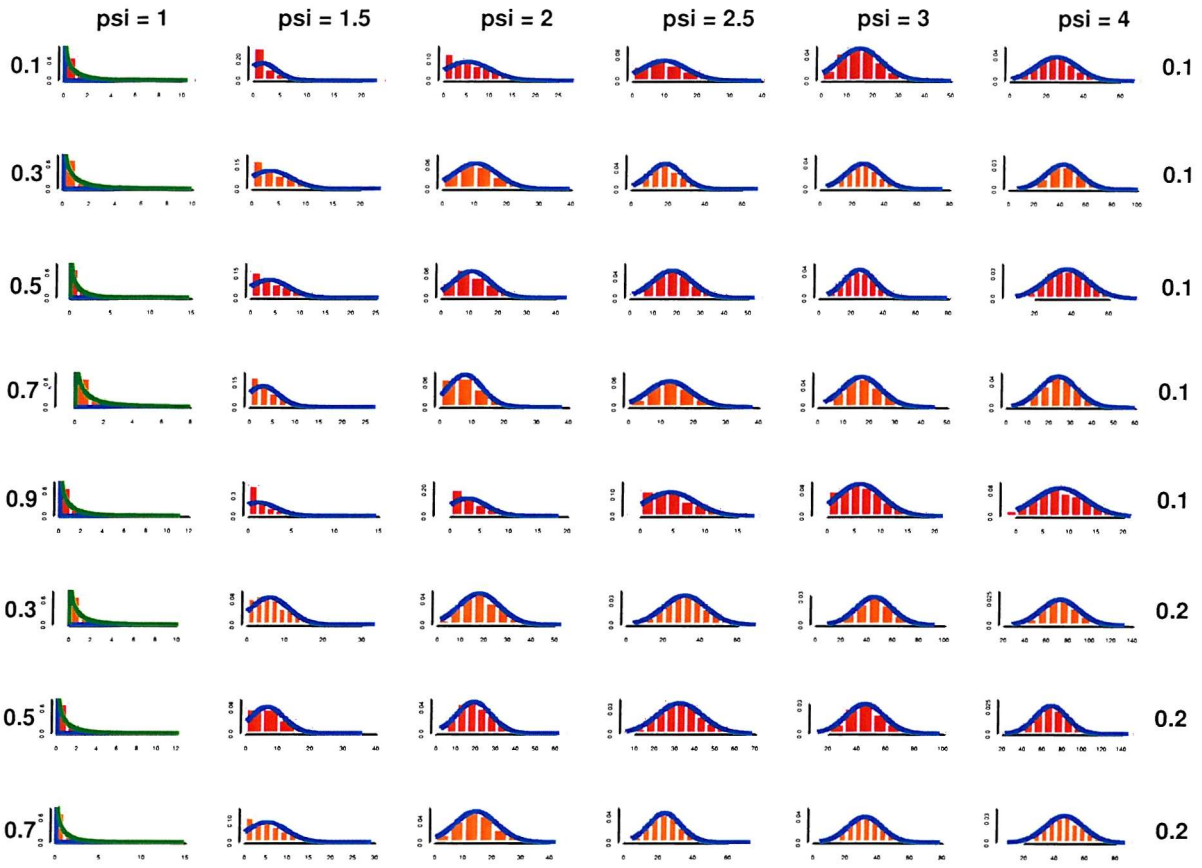

Figure 3.4: Histograms of the empirical distribution of LRT_{12}, for $n_\emptyset = 1\,000$ for different combinations of cell marginal probabilities (in rows: values for $\pi_1(0)$ on the left, values for $\pi_2(0)$ on the right of the plot) and odds ratios (in columns: values for ψ_{12} on top).

used in the current section (for $n_\emptyset = 1\,000$) are given in appendix Table B.1.

Figure 3.5 highlights the better performance of the normal approximation for large values of the sample size: $n_\emptyset = 10\,000$. The same combinations of marginal probabilities were chosen. To facilitate the interpretation of the plots only values of the odds ratio from 1 to 2.5 are displayed, the reason being normality is now achieved for smaller values of the odds ratio, even for more unbalanced combinations of marginal probabilities. This confirms the normal approximation holds asymptotically.

In brief: the normal approximation to the distribution of the LRT statistic for single edge exclusion from the saturated model is a good approximation for $n_\emptyset$ large, an odds ratio not close to one and balanced combinations of marginal probabilities. One should note that in a GG model $n = 1\,000$ can be considered a large sample size. However, that is not the case in a GLL model: for a 2×2 contingency table $n_\emptyset = 10\,000$ can be considered a large sample size.

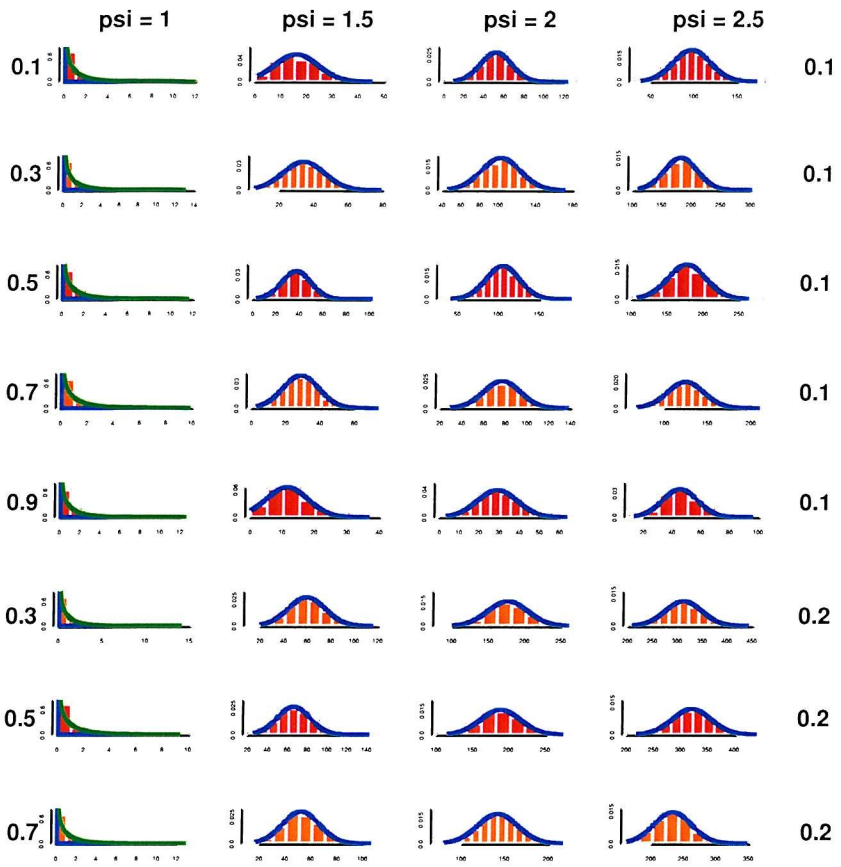

Figure 3.5: Histograms of the empirical distribution of LRT_{12}, for $n_\emptyset = 10\,000$ for different combinations of cell marginal probabilities (in rows: values for $\pi_1(0)$ on the left, values for $\pi_2(0)$ on the right of the plot) and odds ratios (in columns: values for ψ_{12} on top).

The quality of the non-central chi-square approximation proposed in Section 3.6 is now asessed, and contrasted to that of the normal approximation, by using p-p plots. Figure 3.6 shows p-p plots of the distribution of the likelihood ratio test statistic, obtained for the combinations of marginal probabilities $\pi_1(0)$, $\pi_2(0)$ and odds ratio ψ_{12} used in Figure 3.4. The sample size is $n_\emptyset = 1\,000$. As in Section 3.3 (for GG models), the red line represents an exact agreement between observed and expected cumulative probabilities, the blue curve represents the asymptotic normal approximation (with mean and variance given by Equations 3.27 and 3.29) and the green curve the non central chi-square approximation (with non-centrality parameter given by Equation 3.43). From Figure 3.6, with a sample size of 1 000, it is possible to conclude that, as expected, for an odds ratio of one the chi-square distribution is a good approximation (the non-centrality parameter, is zero, so the non central chi-square becomes a central chi-square on one degree of freedom). For values of the odds ratio close to one (1.5,

and even 2) the non central chi-square is a better approximation than the normal. For higher values of the odds ratio (namely 3 or 4) the non central chi-square becomes a poor approximation, except in the cases of balanced tables ($\pi_1(0) = 0.3$; $\pi_2(0) = 0.1$ and $\pi_1(0) = 0.3$; $\pi_2(0) = 0.2$).

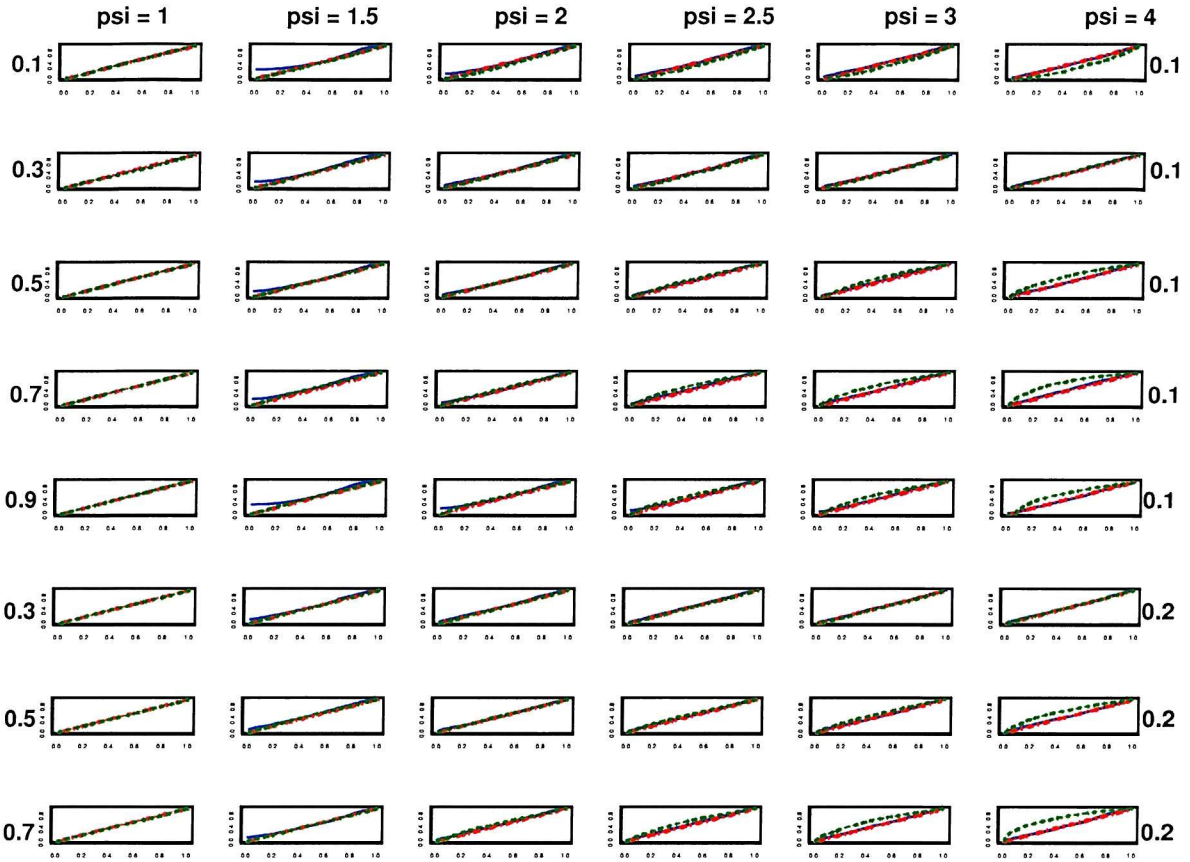

Figure 3.6: P-P plots of the distribution of LRT_{12}, for different combinations of cell marginal probabilities (in rows: values for $\pi_1(0)$ on the left, values for $\pi_2(0)$ on the right of the plot) and odds ratios (in columns: values for ψ_{12} on top). $n_\emptyset = 1\,000$

At this point it was necessary to clarify the concept of balance in a 2×2 contingency table. Since no measure of balance was found in the literature, the *Balance Index* (BI) is proposed. In the 2×2 case,

$$BI = \frac{1}{\pi(0,0)} + \frac{1}{\pi(0,1)} + \frac{1}{\pi(1,0)} + \frac{1}{\pi(1,1)}.$$

The balance index has the following two desirable properties:

i) it relates to information, i.e., it comes from the information matrix and equals the asymptotic variance of $\hat{\lambda}_{12}$ (from Section 2.7.2);

ii) it has a minimum value of 16 when all four cell probabilities are equal. Big BI values can be reached if one of the cell probabilities is very small, by comparison with the others. Higher values of BI are, therefore, associated with contingency tables with small minimum (expected) cell counts. Since all cell probabilities are assumed positive, BI always has finite values.

Table 3.3 shows the values of the balance index for the different combinations of odds ratios and marginal probabilities used in Figures 3.4 and 3.6. For $\psi_{12} = 3$ or 4, the smallest values of BI are obtained for $\pi_1(0) = 0.3$, $\pi_2(0) = 0.1$ and $\pi_1(0) = 0.3$, $\pi_2(0) = 0.2$, the combinations of marginal probabilities for which the non-central chi-square approximation performs best.

$\pi_1(0)$ / $\pi_2(0)$	$\psi_{12} = 1$	$\psi_{12} = 1.5$	$\psi_{12} = 2$	$\psi_{12} = 2.5$	$\psi_{12} = 3$	$\psi_{12} = 4$
0.1 / 0.1	123.5	97.3	84.5	76.9	72.1	66.4
0.3 / 0.1	52.9	47.8	45.9	**45.5**	**45.7**	**47.1**
0.5 / 0.1	44.4	45.8	48.5	51.7	55.2	62.6
0.7 / 0.1	52.9	61.9	71.4	81.1	90.9	110.7
0.9 / 0.1	123.5	163.1	202.9	242.9	282.8	362.7
0.3 / 0.2	29.8	27.6	26.8	**26.6**	**26.7**	**27.2**
0.5 / 0.2	25	25.4	26.6	27.9	29.3	32.2
0.7 / 0.2	29.7	33.5	37.5	41.6	45.7	54.1

Table 3.3: Balance index values for different combinations of marginal probabilities and odds ratios

The proposed measure of balance needs to be further investigated, so that the relationship between the value of the balance index, the value of the odds ratio and the quality of the derived approximations to the distribution of the LRT statistic can be better understood.

The quality of the two approximations (normal and non-central chi-square) for a sample size of 10 000 can be assessed by inspecting Figure 3.7. For $\psi_{12} \neq 1$, the blue line and the red line almost overlap, confirming the conclusions drawn from the histograms in Figure 3.5: for $n_\emptyset = 10\,000$ the normal approximation is a good approximation, even for small values of the odds ratio and for less balanced combinations of marginal cell

probabilities. This confirms the normal approximation holds asymptotically. However, the situation is quite different in the case of the non-central chi-square approximation. The green and the red lines almost overlap for values of ψ_{12} up to 1.5, particularly for more balanced combinations of $\pi_1(0)$ and $\pi_2(0)$. Yet, as the distance from the null increases (i.e., as $\sqrt{n}\log(\psi_{12})$ increases) the non-central chi-square approximation becomes very poor, except in the cases of balanced tables ($\pi_1(0) = 0.3; \pi_2(0) = 0.1$ and $\pi_1(0) = 0.3; \pi_2(0) = 0.2$, as before). This is shown by the green line, in the p-p plots corresponding to ψ_{12} equal to 3 or 4.

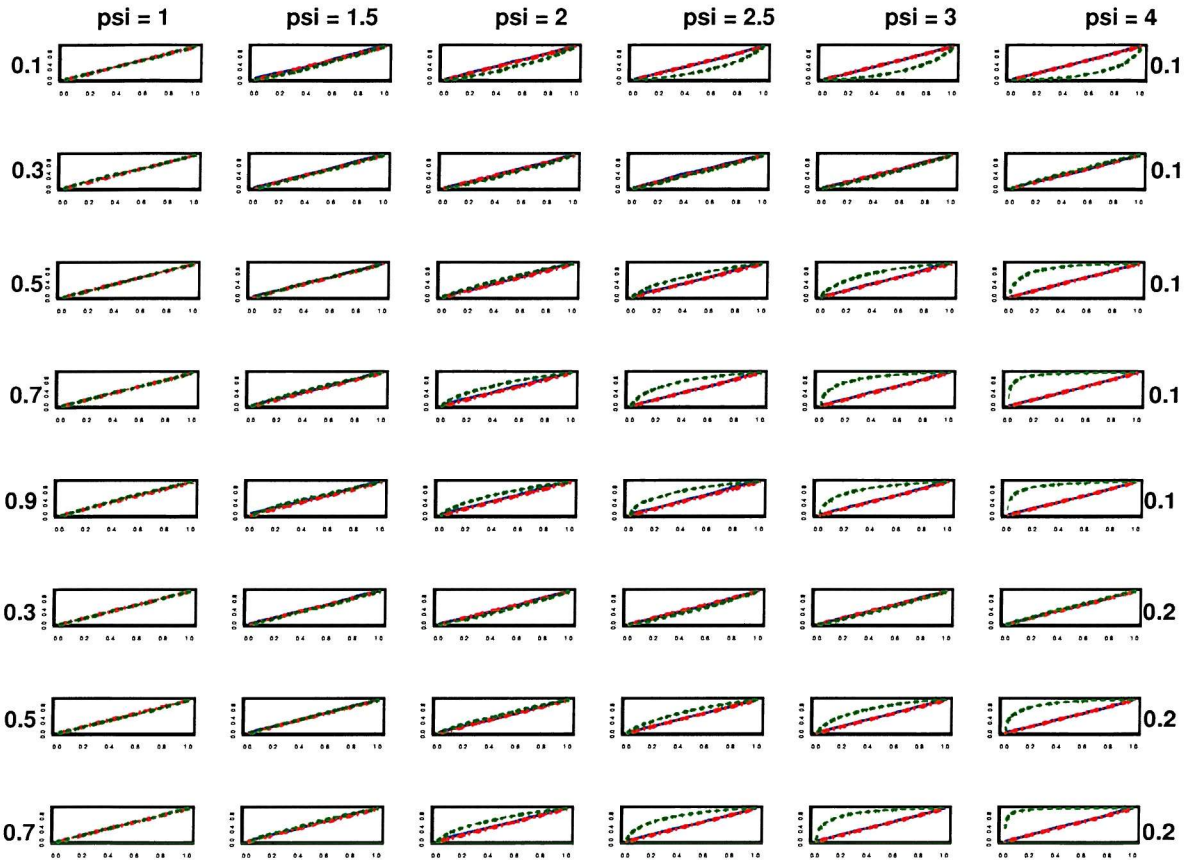

Figure 3.7: P-P plots of the distribution of LRT_{12}, for different combinations of cell marginal probabilities (in rows: values for $\pi_1(0)$ on the left, values for $\pi_2(0)$ on the right of the plot) and odds ratios (in columns: values for ψ_{12} on top). $n_\emptyset = 10\,000$

Although plots are not presented, a smaller sample size of $n_\emptyset = 500$ was also investigated. As expected, the normal approximation performs much worse than in the case $n_\emptyset = 1\,000$. The approximation only seems acceptable for values of ψ_{12} of 3 or 4 and balanced tables. The non-central chi-square is a better approximation for small values of ψ_{12}.

In conclusion: two different approximations to the distributions of the non-signed and signed square-root versions of the likelihood ratio test statistic for single edge exclusion, under the alternative hypothesis that the saturated model holds, have been derived, both in GG and in GLL models. The non-central chi-square approximation can be used in the two variables case and is a good approximation when the distance from the null is small (as n increases, either ρ_{12} or $\log \psi_{12}$ have to be close to zero). For small sample sizes, it performs better than the normal approximation, the distance from the null being small. The normal approximation holds asymptotically, i.e., performs better for large sample sizes.

In the framework of GG models $n = 1\,000$ can already be considered a large sample size and the normal approximation is a good one, even for small values of the (partial) correlation coefficient. In GLL models, 1 000 observations is not a large sample size and the asymptotic normal approximation has some limitations, particularly for small values of the odds ratio and less balanced combinations of marginal probabilities. In the 2×2 case, 10 000 observations can already be considered a large sample size and the normal approximation seems a good approximation to the distribution of the likelihood ratio test statistic for single edge exclusion under the alternative hypothesis that the saturated model holds.

Asymptotic normal approximations to the distributions of the Wald and of the score test statistics for single edge exclusion from the saturated model have also been derived. In GG models the conclusions apply to the general p variables case, whereas results for GLL models are restricted to the two and the three binary variables cases. Conclusions are very similar to those associated with the LRT statistic.

In Chapter 4 all these theoretical approximations are used to derive the power of selecting the saturated model for the three test statistics for single edge exclusion.

Chapter 4

The Power of the Test Statistics for Single Edge Exclusion

In Chapter 3 theoretical approximations have been derived to the distributions of the test statistics for single edge exclusion, under the alternative hypothesis that the saturated model holds. The aim of the current chapter is to investigate the power of the three test statistics for single edge exclusion (likelihood ratio, Wald and score), both in GG and in GLL models. Non-signed and signed square-root versions of the test statistics are considered.

The *power* of an hypothesis test is the probability of rejecting the null hypothesis when it is false (the alternative hypothesis being true) and equals one minus the type II error. The definition of *power of a model selection procedure* used in this thesis follows Smith (1992): the power of the model selection procedure refers to the probability of selecting the true model given the specified true model parameters. The traditional definition of power is associated with a single hypothesis test with a single null hypothesis. However, a model selection procedure involves a sequence of tests with a sequence of null hypotheses. The fact that more than one hypothesis is tested at a time may raise an argument for not calling the probability of selecting the true model *power*. Yet, because it has the essence of power, in the sense that it is the probability of accepting the 'right hypotheses', the decision was made to keep the term power.

Simulation studies are conducted to estimate the power of selecting the saturated model when using the test statistics for single edge exclusion. Theoretical power functions are also derived, and the quality of the approximations is assessed. Throughout

the chapter each individual hypothesis test is carried out considering a size (type I error) of 5%. This does not mean, however, that each model selection procedure has an overall size of 5% because of multiple testing.

A more detailed description of the structure of the current chapter is now given. The chapter has two main parts: the first part (Section 4.1) aims at studying the power of the test statistics for single edge exclusion in GG models; the second part (Section 4.2) is concerned with such a study in GLL models.

The power of selecting the saturated GG model with two variables, using the likelihood ratio test statistic, is studied in Section 4.1.1. Asymptotic normal and non-central chi-square approximations are derived. The quality of the two approximations is assessed by simulation, particularly as the sample size varies. Section 4.1.2 derives asymptotic normal approximations to the power of the Wald and of the score test statistics for single edge exclusion from a saturated GG model with two variables. Section 4.1.3 investigates the positive definiteness constraint on the 3×3 scaled inverse variance matrix. In Section 4.1.4 the power of selecting the saturated GG model with three variables, using the LRT statistic, is estimated by simulation. The corresponding theoretical normal approximation is derived in Section 4.1.5. Section 4.1.6 shows how the proposed theoretical approximations to the power of the test statistics for single edge exclusion, in GG models, can be generalised to the p variables case.

The power of selecting the saturated GLL model, with two binary variables, using the likelihood ratio test statistic, is investigated in Section 4.2.1. Asymptotic normal and non-central chi-square approximations are derived. The quality of the two approximations is assessed by simulation, as the sample size varies. Asymptotic normal approximations to the power of the Wald and of the score test statistics (in the two binary variables case) are derived in Section 4.2.2. The power of the signed square-root versions of the three test statistics is studied in Section 4.2.3. Finally, Section 4.2.4 aims to generalise previous results, regarding the power of the test statistics for single edge exclusion from the saturated GLL model, to the three binary variables case.

4.1 Power of Single Edge Exclusion in GG Models

In this first part of the chapter GG models are considered and the power of single edge exclusion tests is analysed.

4.1.1 Power of the LRT statistic in the two variables case: normal versus non-central χ^2 approximations

The power for the LRT of excluding edge 12 from the saturated GG model with two variables, also called the power of selecting the saturated GG model with two variables, is first estimated by simulation. Theoretical power functions are then derived and compared to the simulation results.

Simulated power functions

The power of a backwards elimination model selection procedure for selecting the true (saturated) model is now investigated by a simulation study. In GG modelling, the association between manifest variables is measured by the (partial) correlation coefficients. Power is calculated for different values of the correlation coefficient (41 different values of ρ_{12} ranging from -1 to 1, with an interval of 0.05, are used). In other words, the probability of selecting the saturated model given each of the specified 'true' values for the correlation coefficient, is estimated. To study how power varies as a function of sample size n, simulations are repeated for $n = 50, 100, 200, 500, 1\,000$. All the simulation results presented were obtained using the likelihood ratio test for single edge exclusion. Simulations were also carried out using the Wald test statistic and the score test statistic and equivalent results were obtained - for that reason the corresponding plots are omitted.

A more detailed explanation of the simulation procedure is now given. For each sample size n and for each chosen combination of values of the (partial) correlation coefficient(s) in the population, 1 000 samples were generated from a normal distribution with that pre-defined covariance structure. For each sample, the test statistic for single edge exclusion from the saturated model was calculated and stored. The probability of selecting the saturated model was then estimated, as the number of times, out of 1 000, that the saturated model was chosen, that is $P[T^L_{ij} > \chi^2_{1;\,0.95}]$, for each edge ij.

Figure 4.1 summarises the results of the simulations for the two variables case, obtained after 1 000 iterations. It is worth noting that:

- the power function is symmetric about zero correlation;
- the power increases as $|\rho_{12}|$ increases;
- the power also increases as the sample size increases.

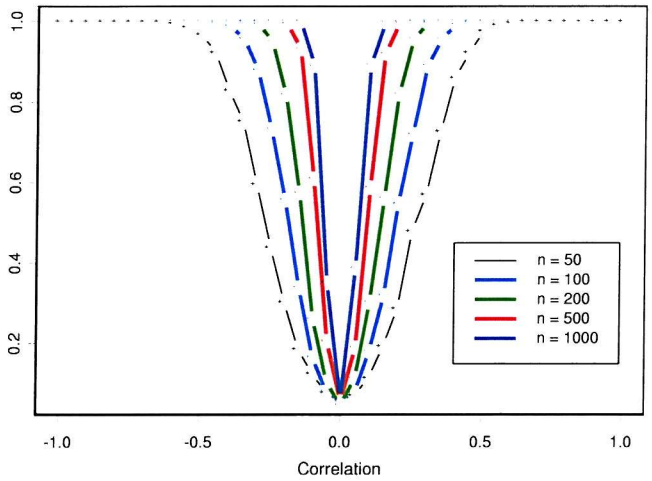

Figure 4.1: Power functions for the saturated GG model - two variables, five different sample sizes.

Theoretical power functions using the LRT statistic

Recall that testing that edge 12 is not present in the model is testing the null hypothesis, H_0, that $\rho_{12} = 0$. Under H_0, the test statistic T_{12}^L is chi-square distributed, on one degree of freedom. The probability of selecting a model with edge 12 present (corresponding to rejecting H_0) is calculated as $P[T_{12}^L > \chi^2_{1;\,0.95} \mid \rho_{12}]$, for different values of ρ_{12}.

Theoretical power functions are derived both for the non-signed and for the signed square-root versions of the LRT statistic. The case of the non-signed version of the LRT is now considered. As derived in Section 3.1.3, under the alternative hypothesis that the saturated model holds (i.e., $\rho_{12} \neq 0$), the likelihood ratio test statistic for single edge exclusion, is asymptotically normal distributed, with mean $AE[T_{12}^L] = -n\log(1-\rho_{12}^2)$ and variance $\text{var}(T_{12}^L) = 4n\rho_{12}^2$. Therefore, an asymptotic normal approximation to the power for the LRT of excluding edge 12 from the saturated GG model with two

variables can be obtained as

$$P[T_{12}^L > 3.8414 \mid \rho_{12}] \stackrel{a}{=} P\left[z > \frac{3.8414 + n\log(1-\rho_{12}^2)}{2\sqrt{n}\,|\rho_{12}|}\right], \tag{4.1}$$

where $z \sim N(0,\ 1)$.

Alternatively, a non-central χ^2 approximation can be used. In Section 3.2 it was proposed that the distribution of the likelihood ratio test statistic for single edge exclusion from the saturated model can be approximated by a non-central χ^2 distribution, with non-centrality parameter $\varphi = \frac{n\,\rho_{12}^2}{1+\rho_{12}^2}$. The null hypothesis that $\rho_{12} = 0$ is rejected if $P[T_{12}^L > \chi^2_{1;\,0.95} \mid \rho_{12}]$. Hence, the theoretical power functions can be obtained by calculating one minus the cumulative probability for a non-central chi-square distribution with one degree of freedom, for a quantile value of 3.8414 (when ρ_{12} varies between -1 and 1).

Figure 4.2 compares the simulated power values (in red, corresponding to the five different lines in Figure 4.1) with the theoretical values calculated using the asymptotic normal approximation (in blue) and the non-central chi-square approximation (in green), for the non-signed version of the likelihood ratio test statistic for single edge exclusion from the saturated model. The five different plots correspond to the five sample sizes used: 50, 100, 200, 500 and 1 000 observations. The horizontal dotted lines correspond to power values of 0 and 0.05.

Figure 4.2 highlights a difference in shape between the red and the blue curves, particularly for small values of ρ_{12}, indicating the theoretical normal approximation performs poorly. Note that if $\rho_{12} = 0$, $AE[T_{12}^L] = 0$ and $var(T_{12}^L) = 0$. The z value is infinity. Using a (partial) correlation value close to zero induces a very small value for the variance of T_{12}^L and, consequently, z becomes very big and the corresponding theoretical probability (blue curve) is zero. For large sample sizes ($n = 1\,000$) the asymptotic normal approximation performs well for values of ρ_{12} not close to zero. For sample sizes of, at least, 200 observations, the green and the red lines overlap. This indicates power calculations for the non-signed version of the LRT statistic for single edge exclusion from the saturated model can be accurately approximated by a non-central chi-square distribution.

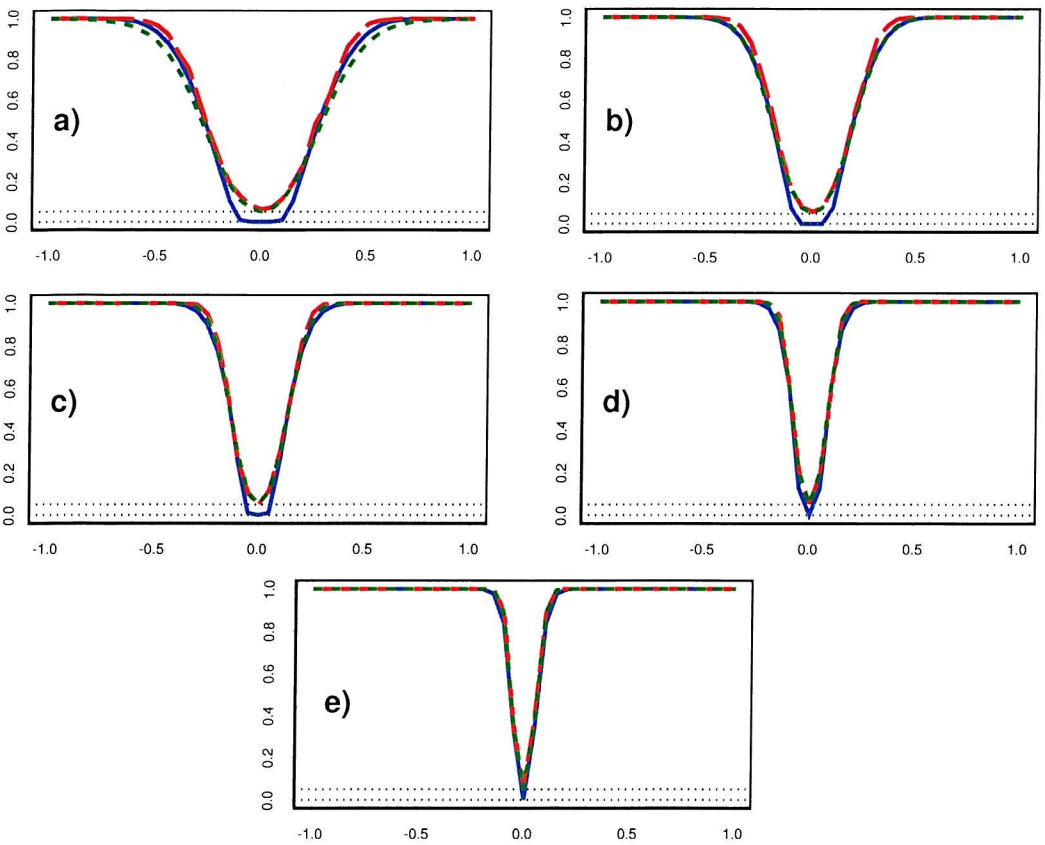

Figure 4.2: Simulated (in red) and theoretical power curves using the likelihood ratio test statistic for single edge exclusion, T_{12}^{L}, with an asymptotic normal approximation (in blue) and a non-central chi-square approximation (in green), for five different sample sizes: a) $n = 50$, b) $n = 100$, c) $n = 200$, d) $n = 500$, and e) $n = 1\,000$.

The case of the theoretical power calculations when using the signed square-root version of the LRT is now considered. Recall that, from Section 2.9.2,

$$T_{12}^{signL} = sign(\hat{\rho}_{12})\sqrt{-n\log(1-\hat{\rho}_{12}^2)},$$

and that, from Section 3.1.5, in the asymptotic normal distribution,

$$\text{var}(T_{12}^{signL}) = 4n\rho_{12}^2 \times \frac{-1}{4n\log(1-\rho_{12}^2)} = \frac{-\rho_{12}^2}{\log(1-\rho_{12}^2)}$$

and

$$AE[T_{12}^{signL}] = sign(\rho_{12})\sqrt{-n\log(1-\rho_{12}^2)}.$$

If $\rho_{12} = 0$, T_{12}^{signL} is asymptotically normal distributed, with mean zero and variance one. Indeed, if $\rho_{12} = 0$, then $sign(\rho_{12})\sqrt{-n\log(1-\rho_{12}^2)} = 0$ and, applying Hôpital's rule, $\frac{-\rho_{12}^2}{\log(1-\rho_{12}^2)}$ tends to one as ρ_{12} tends to zero. If $\rho_{12} \neq 0$, the

signed square-root version of the likelihood ratio test statistic for single edge exclusion, from the saturated model, is asymptotically normal distributed with mean $AE[T_{12}^{signL}] = sign(\rho_{12})\sqrt{-n\log(1-\rho_{12}^2)}$ and variance $\text{var}(T_{12}^{signL}) = \frac{-\rho_{12}^2}{\log(1-\rho_{12}^2)}$.

In the case of a two-sided hypothesis test, the null hypothesis that ρ_{12} is zero is rejected if the absolute value of the signed version of the LRT statistic is greater than 1.96, for the different values of ρ_{12}. That is, asymptotically, the power for the two-sided signed square-root LRT of excluding edge 12 from the saturated GG model with two variables can be obtained as

$$P[|T_{12}^{signL}| > 1.96 \mid \rho_{12}] \stackrel{a}{=} P\left[z < \frac{-1.96 - sign(\rho_{12})\sqrt{-n\log(1-\rho_{12}^2)}}{|\rho_{12}|/\sqrt{-\log(1-\rho_{12}^2)}}\right] + P\left[z > \frac{1.96 - sign(\rho_{12})\sqrt{-n\log(1-\rho_{12}^2)}}{|\rho_{12}|/\sqrt{-\log(1-\rho_{12}^2)}}\right], \tag{4.2}$$

where $z \sim N(0,\ 1)$.

ρ_{12} varies between -1 and 1. Figure 4.3 compares the simulated power values (in red) with the theoretical values of the normal approximation (in blue) obtained for the two-sided signed square-root version of the likelihood ratio test statistic for single edge exclusion from the saturated model. As previously, the five plots correspond to the five different sample sizes used. The horizontal dotted line corresponds to a power value of 0.05.

In all five plots the blue and the red lines overlap, validating the theoretical normal approximation of the power functions presented above, even for small values of the sample size, and small (partial) correlation coefficients.

In the case of a one-sided hypothesis test, the null hypothesis that ρ_{12} is zero is rejected if the value of the signed version of the LRT statistic is greater than 1.645, for the different values of ρ_{12}. That is, asymptotically, the power for the one-sided signed square-root LRT for excluding edge 12 from the saturated GG model with two variables can be obtained as

$$P[T_{12}^{signL} > 1.645 \mid \rho_{12}] \stackrel{a}{=} P\left[z > \frac{1.645 - sign(\rho_{12})\sqrt{-n\log(1-\rho_{12}^2)}}{|\rho_{12}|/\sqrt{-\log(1-\rho_{12}^2)}}\right], \tag{4.3}$$

where $z \sim N(0,\ 1)$.

ρ_{12} varies between zero and one. Figure 4.4 compares the simulated power values (in red) with the theoretical values of the normal approximation (in blue) obtained for

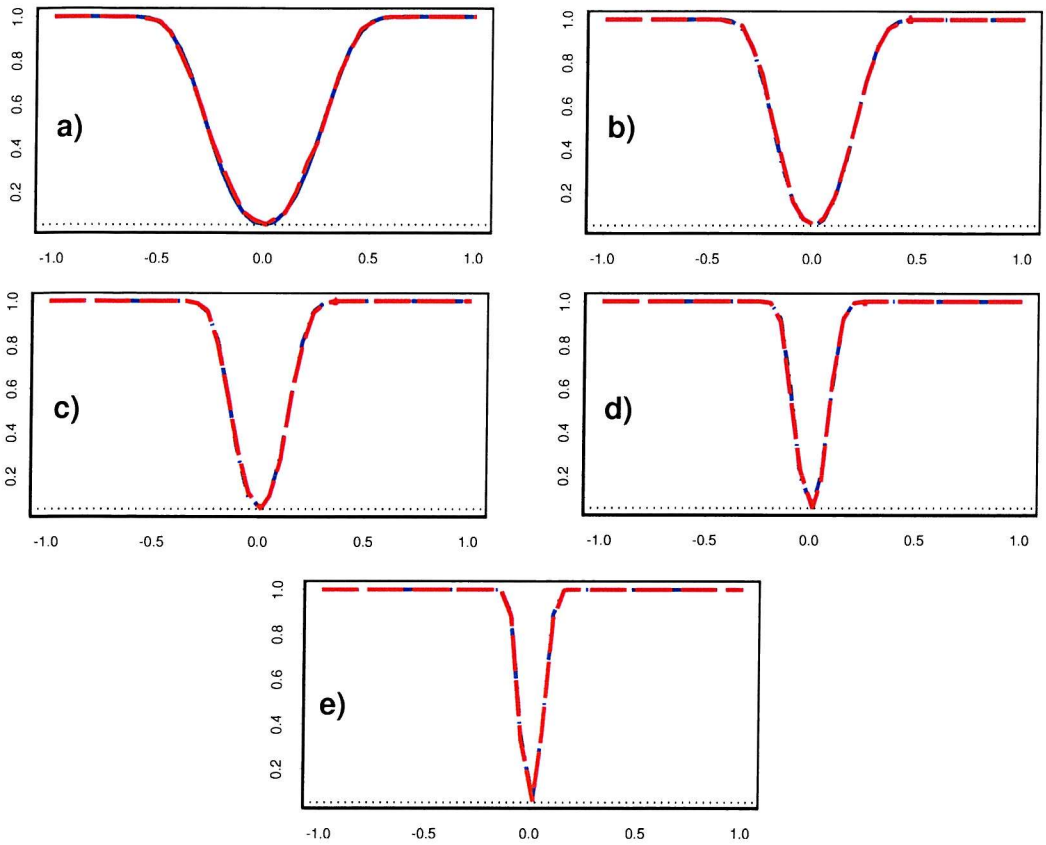

Figure 4.3: Simulated (in red) and theoretical asymptotic normal (in blue) power curves using the two-sided signed square-root version of the likelihood ratio test statistic for single edge exclusion, T_{12}^{signL}, for five different sample sizes: a) $n = 50$, b) $n = 100$, c) $n = 200$, d) $n = 500$, and e) $n = 1\,000$.

the one-sided signed square-root version of the likelihood ratio test statistic for single edge exclusion from the saturated model.

Results of the theoretical derivations of the power functions presented are validated by the overlapping of the blue and the red lines in all five plots. One should note, that for negative values of ρ_{12}, power either equals zero or, for small sample sizes, a value very close to zero. For example, for $n = 50$ power equals 0.0091 for $\rho_{12} = -0.1$ and equals 0.00005 for $\rho_{12} = -0.3$.

In brief: the normal approximation to the power for the signed square-root LRT of excluding edge 12 from the saturated GG model with two variables is a good approximation, both when one-sided and two-sided hypothesis tests are used, even for small values of n and (partial) correlation coefficients close to zero.

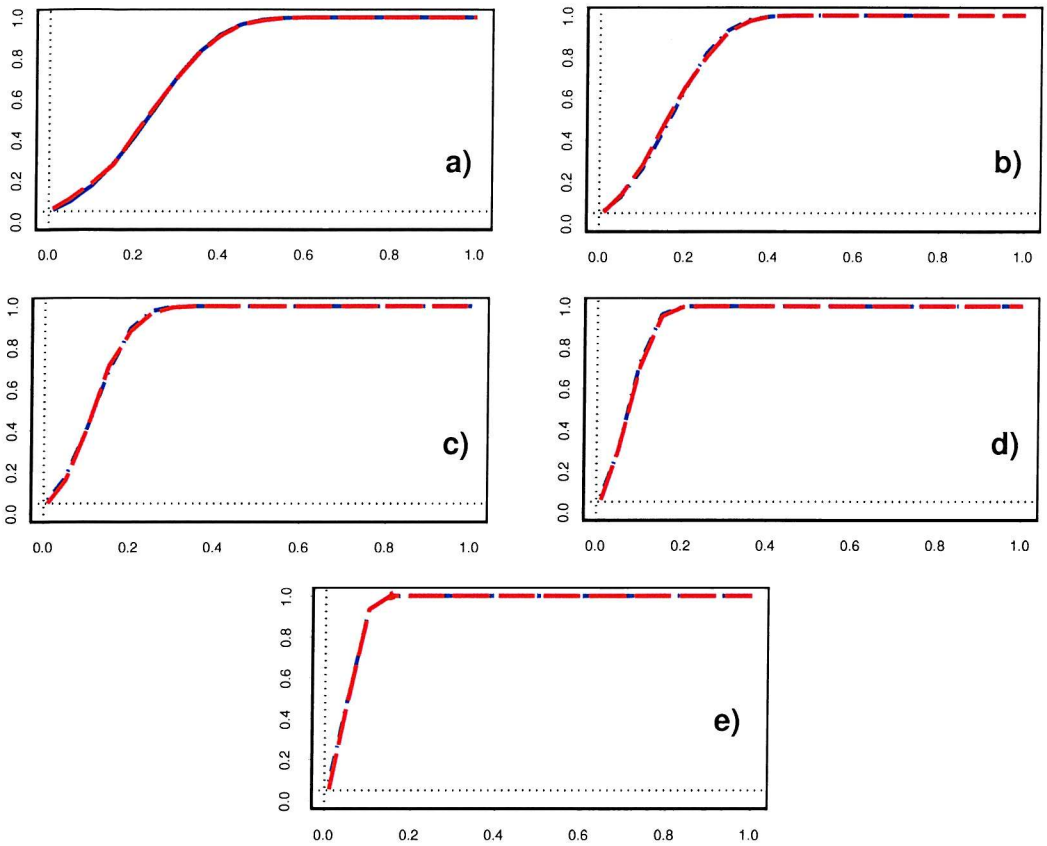

Figure 4.4: Simulated (in red) and theoretical asymptotic normal (in blue) power curves using the one-sided signed square-root version of the likelihood ratio test statistic for single edge exclusion, T_{12}^{signL}, for five different sample sizes: a) $n = 50$, b) $n = 100$, c) $n = 200$, d) $n = 500$, and e) $n = 1\,000$.

In conclusion, the power of selecting the saturated GG model with two variables, using the likelihood ratio test statistic for single edge exclusion:

- increases as $|\rho_{12}|$ increases. The value of $-n\log(1-\rho_{12}^2)$ is always non-negative. For a given sample size, if ρ_{12} increases in absolute value, the numerator of Equation 4.1 becomes more negative and the corresponding z values become more negative. Consequently, power increases. Similar reasoning applies to Equations 4.2 and 4.3;

- increases as n increases. n appears in the numerator of Equation 4.1 and $\sqrt{n}$ in its denominator. Therefore, for a given ρ_{12}, as n increases the numerator becomes more negative, faster than the denominator becomes more positive. Hence, the z values become more negative and power increases. In the signed square-root versions n only appears in the numerator of Equations 4.2 and 4.3, and a similar

reasoning applies;

- for two-sided hypothesis tests power is symmetric about zero correlation: both in Equations 4.1 and 4.2 either the absolute value of ρ_{12} or its square-root value are considered, leading to the same z values for positive and negative ρ_{12}, and the standardised normal distribution is symmetric about zero.

4.1.2 Power of the Wald and score test statistics in the two variables case: normal approximations

From Section 3.1.4 recall that, if $\rho_{12} \neq 0$ (i.e., the saturated model, with two variables, holds) the Wald and the score test statistics for single edge exclusion in GG models are asymptotically normal distributed, with mean $AE[T_{12}^W] = n\,\frac{\rho_{12}^2}{1+\rho_{12}^2}$; $AE[T_{12}^S] = n\,\rho_{12}^2$ and variance $\text{var}(T_{12}^W) = \frac{4\,n\,\rho_{12}^2\,(1-\rho_{12}^2)^2}{(1+\rho_{12}^2)^4}$; $\text{var}(T_{12}^S) = 4\,n\,\rho_{12}^2\,(1-\rho_{12}^2)^2$. Therefore, as in Section 4.1.1, theoretical power functions for the non-signed versions of the Wald and the score test statistics can be obtained, using a normal approximation, as

$$P[T_{12}^W > 3.8414 \mid \rho_{12}] \stackrel{a}{=} P\left[z > \frac{3.8414 - n\,\rho_{12}^2\,(1+\rho_{12}^2)^{-1}}{2\,\sqrt{n}\,|\rho_{12}|\,(1-\rho_{12}^2)\,(1+\rho_{12}^2)^{-2}}\right]$$

and

$$P[T_{12}^S > 3.8414 \mid \rho_{12}] \stackrel{a}{=} P\left[z > \frac{3.8414 - n\,\rho_{12}^2}{2\,\sqrt{n}\,|\rho_{12}|\,(1-\rho_{12}^2)}\right],$$

where $z \sim N\,(0,\ 1)$.

Signed square-root versions of the Wald and of the score test statistics can also be considered. From Section 3.1.5 recall that, if $\rho_{12} \neq 0$, the signed square-root versions of the Wald and of the score test statistics for single edge exclusion from the saturated model are asymptotically normal distributed, with mean $AE[T_{12}^{signW}] = \rho_{12}\sqrt{\frac{n}{1+\rho_{12}^2}}$; $AE[T_{12}^{signS}] = \rho_{12}\sqrt{n}$ and variance $\text{var}(T_{12}^{signW}) = \frac{(1-\rho_{12}^2)^2}{(1+\rho_{12}^2)^3}$; $\text{var}(T_{12}^{signS}) = (1-\rho_{12}^2)^2$.

Asymptotically, the power for the two-sided signed square-root Wald test of excluding edge 12 from the saturated GG model with two variables can be obtained as

$$P[|T_{12}^{signW}| > 1.96 \mid \rho_{12}] \stackrel{a}{=} P\left[z < \frac{-1.96-\rho_{12}\,\sqrt{n\,(1+\rho_{12}^2)^{-1}}}{(1-\rho_{12}^2)\,(1+\rho_{12}^2)^{-3/2}}\right] + P\left[z > \frac{1.96-\rho_{12}\,\sqrt{n\,(1+\rho_{12}^2)^{-1}}}{(1-\rho_{12}^2)\,(1+\rho_{12}^2)^{-3/2}}\right]$$

and for the two-sided signed square-root score test statistic as

$$P[|T_{12}^{signS}| > 1.96 \mid \rho_{12}] \stackrel{a}{=} P\left[z < \frac{-1.96-\rho_{12}\,\sqrt{n}}{(1-\rho_{12}^2)}\right] + P\left[z > \frac{1.96-\rho_{12}\,\sqrt{n}}{(1-\rho_{12}^2)}\right],$$

where $z \sim N(0,\ 1)$.

In the case of a one-sided hypothesis test, the power of selecting the saturated GG model with two variables, using the signed square-root versions of the Wald and the score test statistics for single edge exclusion is given by

$$P[T_{12}^{signW} > 1.645 \mid \rho_{12}] \stackrel{a}{=} P\left[z > \frac{1.645 - \rho_{12}\sqrt{n\,(1+\rho_{12}^2)^{-1}}}{(1-\rho_{12}^2)\,(1+\rho_{12}^2)^{-3/2}}\right]$$

and

$$P[T_{12}^{signS} > 1.645 \mid \rho_{12}] \stackrel{a}{=} P\left[z > \frac{1.645 - \rho_{12}\sqrt{n}}{(1-\rho_{12}^2)}\right],$$

where $z \sim N(0,\ 1)$.

Plots referring to this section are not presented for the simple reason that the power functions of the Wald and of the score test statistics are very similar to those of the LRT statistic: differences can hardly be detected when observing the corresponding plots. The conclusions drawn in Section 4.1.1, for the LRT statistic, also apply for the Wald and the score test statistics. In short: in the two variables case power increases as n increases, as $|\rho_{12}|$ increases and is symmetric about zero correlation for two-sided hypothesis tests. Additionally, the asymptotic normal approximation is a good approximation to the power of the signed square-root versions of the Wald and of the score test statistics (even if n is not large), whereas for the non-signed versions the approximation is only acceptable for large samples and values of the (partial) correlation coefficient not close to zero.

4.1.3 The shape of the scaled inverse variance matrix

The study of the power functions in the three (or more) variables case requires some thought when specifying the domain, since the scaled inverse variance matrix is constrained to be positive definite. The aim of this section is to express the positive definiteness constraint in terms of partial correlations and to emphasise some properties of the set of all possible combinations of the three partial correlation coefficients.

Rousseeuw and Molenberghs (1994) studied the shape of correlation matrices. They stated that the correlation matrix of three variables is positive definite if its determinant is strictly positive, i.e., if

$$1 + 2\,\rho_{12}\,\rho_{13}\,\rho_{23} - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 > 0. \tag{4.4}$$

They constructed a three-dimensional graph of the boundary set of all the combinations of the three correlation coefficients that satisfy this condition as an equality and showed that the different combinations of the three ρ_{ij} form a convex set (the interior of the boundary set) which is symmetric with respect to rotations corresponding to permutations of the components of $(\rho_{12},\ \rho_{13},\ \rho_{23})$. Additionally, they showed that any horizontal cross section of this surface is an ellipse.

These results are now extended to partial correlation coefficients. Recall that the scaled inverse variance matrix has ones on the main diagonal, the off-diagonal elements being the negative of the partial correlation coefficients. In the three variables case

$$scaled\,(\Omega) = \begin{bmatrix} 1 & -\rho_{12.3} & -\rho_{13.2} \\ -\rho_{12.3} & 1 & -\rho_{23.1} \\ -\rho_{13.2} & -\rho_{23.1} & 1 \end{bmatrix}.$$

For the determinant of this matrix to be strictly positive, it is required that

$$1 - 2\,\rho_{12.3}\,\rho_{13.2}\,\rho_{23.1} - \rho^2_{12.3} - \rho^2_{13.2} - \rho^2_{23.1} > 0. \tag{4.5}$$

As happened with correlation coefficients, the different combinations of the three $\rho_{ij.k}$ form a convex set which is symmetric with respect to rotations corresponding to permutations of the components of $(\rho_{12.3},\ \rho_{13.2},\ \rho_{23.1})$. Figure 4.5 supports the statement that any horizontal cross section of this surface is an ellipse. In all panels $\rho_{13.2}$ is on the horizontal axis, $\rho_{23.1}$ is on the vertical axis and $\rho_{12.3}$ takes arbitrary positive values of 0.1 in a), 0.5 in b), 0.7 in c), 0.9 in d) and negative values of - 0.1 in e), -0.5 in f), -0.7 in g) and - 0.9 in h).

Because of the existing symmetries, only certain combinations of values of the three partial correlations need to be considered in the next section, when studying the power of the LRT statistic for single edge exclusion from a saturated GG model with three variables.

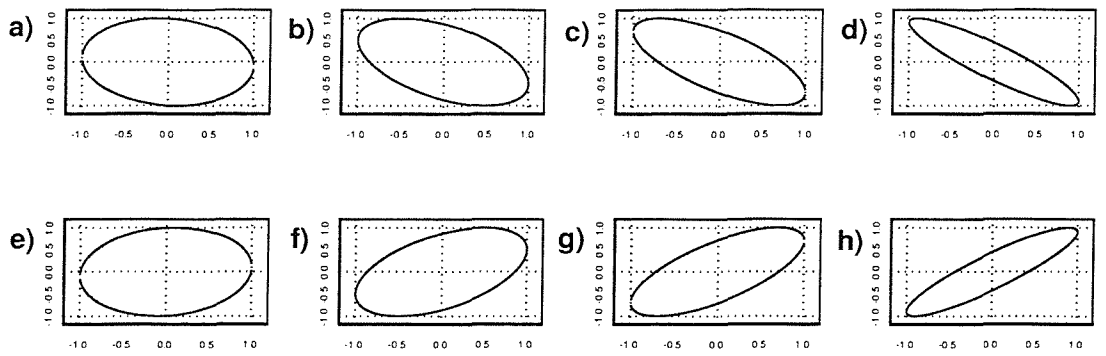

Figure 4.5: The positive definiteness constraint for partial correlation coefficients: $\rho_{13.2}$ on the horizontal axis, $\rho_{23.1}$ on the vertical axis and $\rho_{12.3}$ taking values of: a) 0.1, b) 0.5, c) 0.7, d) 0.9, e) -0.1, f) -0.5, g) -0.7, h) -0.9.

4.1.4 Simulated power of the LRT statistic in the three variables case

The aim of this section is to study the power of the test statistics for single edge exclusion from the saturated GG model with three variables. A simulation study is used. Theoretical normal approximations to the power functions are derived in Section 4.1.5. The quality of the approximations is then assessed. The LRT statistic is used.

The power for the three tests of excluding edge ij, edge ik and edge jk, here called *power of selecting the saturated model*, is estimated as the probability that each of the three test statistics for single edge exclusion is greater than a critical value from a chi-square distribution on one degree of freedom, given certain values for the three partial correlation coefficients, i.e.,

$$P[T^L_{12} > \chi^2_{1;\,0.95} \text{ and } T^L_{13} > \chi^2_{1;\,0.95} \text{ and } T^L_{23} > \chi^2_{1;\,0.95} \mid \rho_{12.3},\ \rho_{13.2},\ \rho_{23.1}].$$

This corresponds to the three null hypotheses that each of the partial correlation coefficients is zero, and to the three alternative hypotheses that each of them is different from zero (saturated model). Power is, therefore, the overall probability of rejecting the three null hypotheses that the partial correlation coefficients are zero, when these hypotheses are false and the saturated model holds. For each of the three tests a size of 5% is considered. When the true model is the independence model, the overall probability of accepting the saturated model equals 0.05^3. Consequently, the overall size of the selection procedure can range between 0.05^3 and 0.05.

In this simulation study the three partial correlation coefficients vary between -1 and 1, subject to the positive definiteness constraint defined by Equation 4.5. In practice $\rho_{13.2}$ and $\rho_{23.1}$ vary from 0.1 to 0.9 (with an interval of 0.1) and $\rho_{12.3}$ varies between -0.9 and 0.9, within the region of positive definiteness. Figure 4.6 shows the power functions for a sample size of 200, using the likelihood ratio test for edge exclusion. Plot a) was obtained after 20 000 repetitions (in order to reduce Monte Carlo error) and the remaining three plots after 1 000 repetitions.

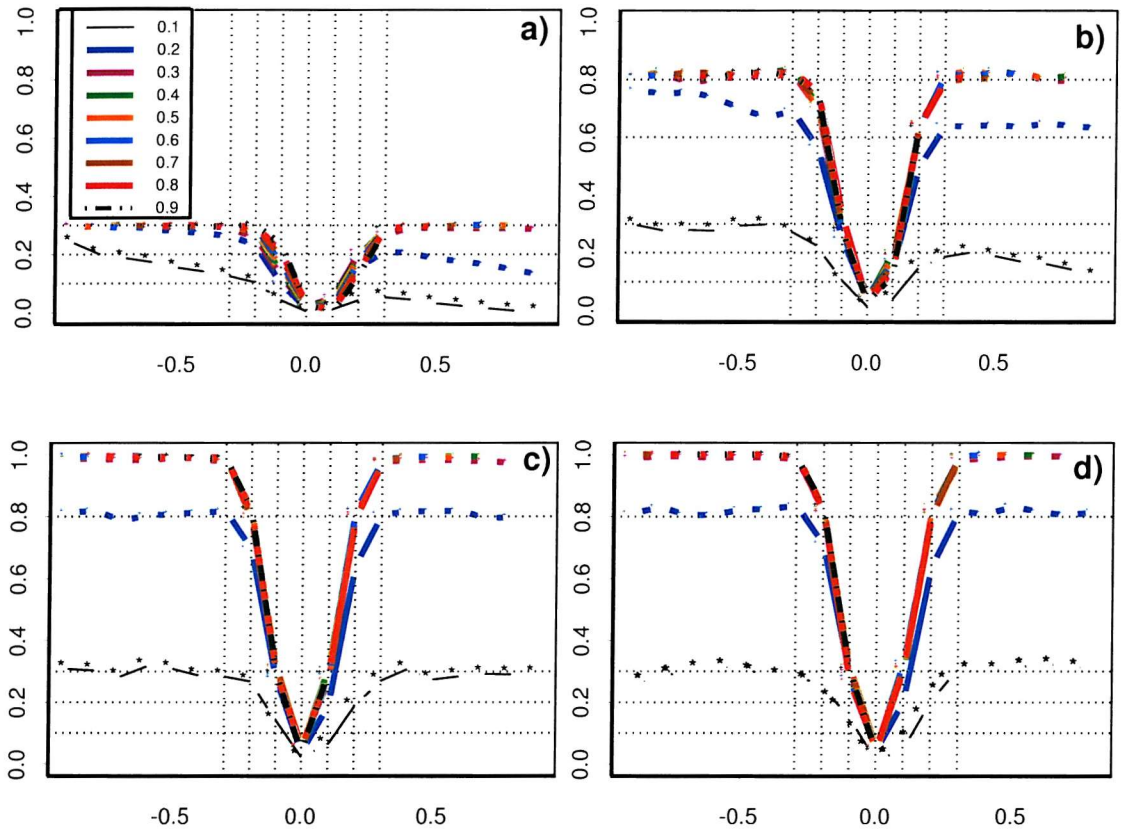

Figure 4.6: Power functions for the GG saturated model, with 3 variables, using the LRT statistic, for a sample size of 200. $\rho_{12.3}$ on horizontal axis. In each plot $\rho_{23.1}$ from 0.1 to 0.9. $\rho_{13.2}$ equals 0.1 in a), 0.2 in b), 0.3 in c) and 0.4 in d).

Some conclusions can be drawn from Figure 4.6:

- generally, power increases as partial correlations increase. However, the black and blue lines (corresponding to values of 0.1 and 0.2 of $\rho_{23.1}$), in panel b), and especially in panel a), show some non-monotonicity with increasing positive partial correlations;

- for $n = 200$, the probability of selecting the saturated model has a maximum value

of $\simeq 0.3$ when one of the $\rho_{ij.k} \simeq 0.1$, even if the other two partial correlations are large. This probability goes up to $\simeq 0.8$, or almost 1, when the minimum $\rho_{ij.k} \simeq 0.2$ or 0.3, respectively;

- there is a lack of symmetry in the power function about zero partial correlation, particularly noteworthy for small values (note the black and blue lines corresponding to 0.1 and 0.2 in panels a) and b)).

Simulations were also carried out using the Wald test statistic and the score test statistic and very similar results were obtained. For that reason the corresponding plots are omitted.

In order to try to understand and explain the non-monotonicity and non-symmetry of the power functions two different types of analysis were carried out. First, simulations were done to estimate both the power for the test of excluding edge ij and the power for the two tests of excluding edge ij and edge ik. The results of the simulation follow. Secondly, a theoretical asymptotic normal approximation to the power functions, in the presence of three variables, was derived. Results are given in Section 4.1.5.

Power for the test of excluding edge ij

The *power for the test of excluding edge* ij is estimated as the probability that the test statistic T_{ij}^L is greater than a critical value from a chi-square distribution on one degree of freedom, given certain values for the partial correlation coefficients, i.e.,

$$P[T_{ij}^L > \chi^2_{1;\,0.95} \mid \rho_{12.3},\ \rho_{13.2},\ \rho_{23.1}].$$

The null hypothesis is that $\rho_{ij.k} = 0$, and the alternative is that it is not zero (assuming the other two edges are present, since edge exclusion is from the saturated model). The two variables case is an example of the power for the test of excluding edge ij (and Figure 4.1 shows a symmetric power function).

Results of the simulation study are presented in Figure 4.7, and it is possible to conclude that power functions are symmetric, as expected. One thousand repetitions were made and $n = 200$. In all plots the horizontal axis corresponds to $\rho_{12.3}$, and the different lines in each plot correspond to values of $\rho_{23.1}$. Plots in panel a) are associated with testing for the exclusion of edge 12, plots in panel b) with the exclusion of edge 13

and plots in panel c) with the exclusion of edge 23. Plots 1) correspond to a value of 0.1 of $\rho_{13.2}$, whereas plots 2) and 3) correspond to values of 0.2 and 0.3, respectively. Plots for higher values of this partial correlation coefficient are not presented for simplicity, since they have a similar pattern.

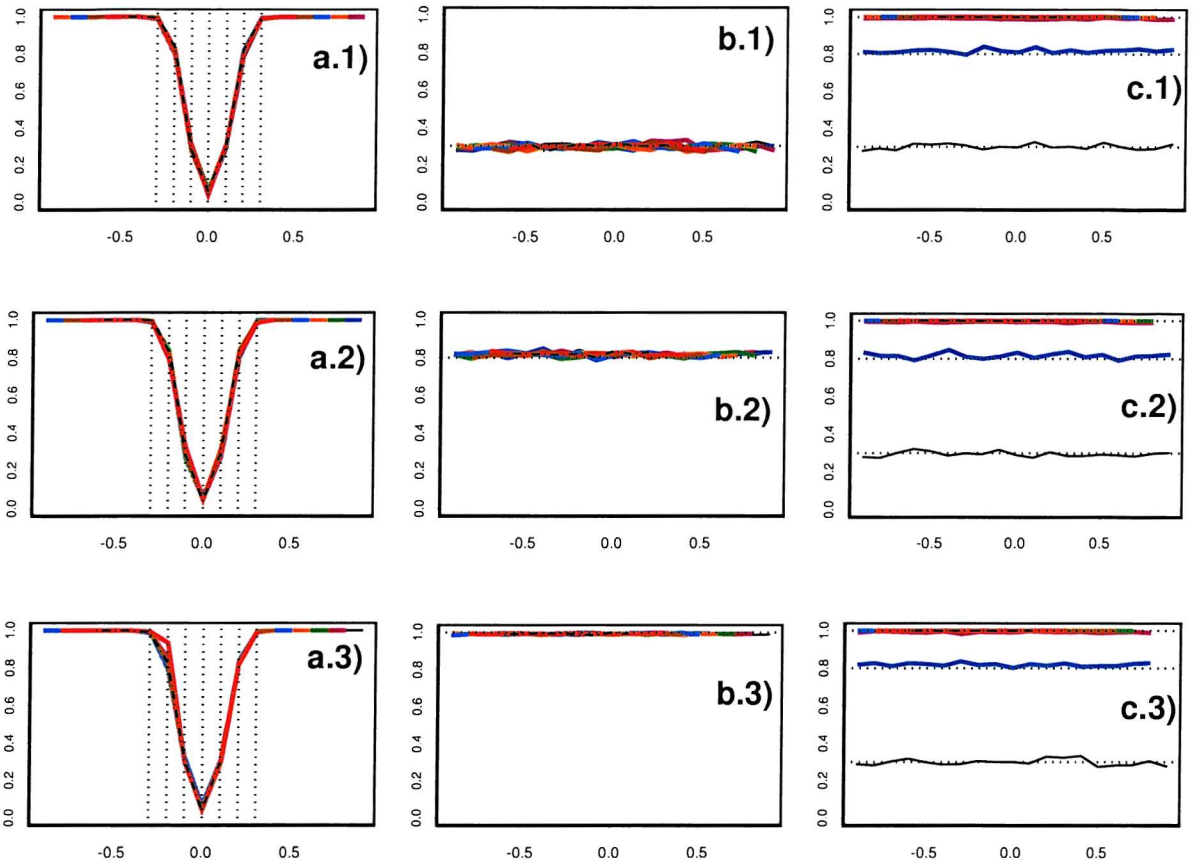

Figure 4.7: Power functions for the test of excluding edge ij, LRT statistic, n=200. Plots a) edge 12, plots b) edge 13, plots c) edge 23. In all plots $\rho_{12.3}$ on the horizontal axis, the different lines corresponding to values of $\rho_{23.1}$. Plots 1) $\rho_{13.2} = 0.1$, plots 2) $\rho_{13.2} = 0.2$, plots 3) $\rho_{13.2} = 0.3$.

In brief:

- plots in panel a) correspond to power calculations for the test of excluding edge 12. The pattern is the same in all three plots, clearly determined by the values of $\rho_{12.3}$ and neither depending on $\rho_{13.2}$ (plots are alike when this coefficient takes values 0.1, 0.2 or 0.3) nor on $\rho_{23.1}$ (lines almost coincident, in each plot);

- plots in panel b) show the power for the test of excluding edge 13. The pattern is always the same, neither depending on $\rho_{12.3}$ (the lines are horizontal), nor on

$\rho_{23.1}$ (lines almost coincident, in each plot). When $\rho_{13.2}$ increases power increases, and therefore lines move horizontally;

- in panel c) plots, power for the test of excluding edge 23, power functions vary because $\rho_{23.1}$ varies. Indeed they do not depend either on $\rho_{12.3}$ (all the lines are horizontal) or on $\rho_{13.2}$ (the three plots are very similar);

- it seems possible to conclude that the power for the test of excluding edge ij (given the remaining edges are present) just depends on $\rho_{ij.k}$: power increases as $|\rho_{ij.k}|$ increases, and is symmetric, about zero, in $\rho_{ij.k}$.

A theoretical explanation for the behaviour of the power for the test of excluding edge ij is given in Section 4.1.5, where it is shown that, asymptotically, both the mean and the variance of the test statistic for single edge ij exclusion only depend on the partial correlation coefficient of interest: $\rho_{ij.rest}$.

Power for the tests of excluding edge ij and edge ik

The *power for the two tests of excluding edge ij and edge ik* is estimated as the probability that each of the two test statistics for single edge exclusion T^L_{ij} and T^L_{ik} is greater than a critical value from a chi-squared distribution on one degree of freedom, given certain values for the three partial correlation coefficients, i.e.,

$$P[T^L_{ij} > \chi^2_{1;\,0.95} \text{ and } T^L_{ik} > \chi^2_{1;\,0.95} \mid \rho_{12.3},\ \rho_{13.2},\ \rho_{23.1}].$$

The two null hypotheses are that $\rho_{ij.k}$ and $\rho_{ik.j}$ are zero, and the two alternative hypotheses are that they are different from zero. In both cases it is assumed the remaining two edges are present, since edge exclusion is from the saturated model. Power denotes the overall probability of rejecting the two null hypotheses that $\rho_{ij.k}$ and $\rho_{ik.j}$ are zero when these are false and the saturated model holds. The overall size of the selection procedure is expected to range between 0.05^2 and 0.05, since for each of the two tests a size of 5% is considered.

Results are presented in Figure 4.8. One thousand repetitions were made and $n = 200$. In all plots the horizontal axis corresponds to $\rho_{12.3}$, and the different lines in each plot correspond to values of $\rho_{23.1}$. It is possible to conclude that there exists some non-symmetry and some non-monotonicity in the power functions. In brief:

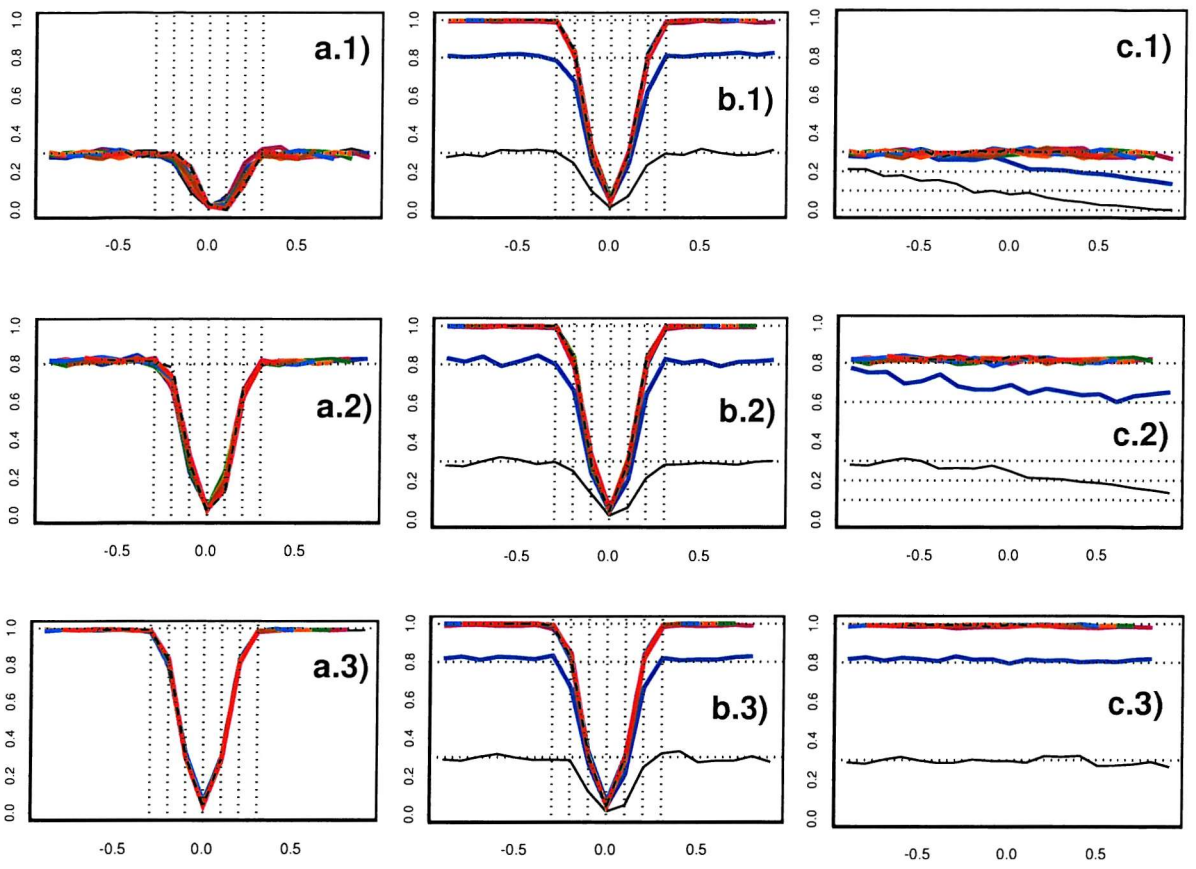

Figure 4.8: Power functions for the tests of excluding edge ij and edge ik, LRT statistic, n=200. Plots a) edges 12, 13, plots b) edges 12, 23, plots c) edges 13, 23. In all plots $\rho_{12.3}$ on the horizontal axis, the different lines corresponding to values of $\rho_{23.1}$. Plots 1) $\rho_{13.2} = 0.1$, plots 2) $\rho_{13.2} = 0.2$, plots 3) $\rho_{13.2} = 0.3$.

- plots in panel a) - power for the tests of excluding edge 12 and edge 13 - power does not seem to vary with $\rho_{23.1}$ (lines almost coincident, in the three plots). Power increases as $|\rho_{12.3}|$ and or $|\rho_{13.2}|$ increase; does not exceed 0.3 if one of the ρ is around 0.1 (even if the other is large). Power almost reaches one if $|\rho_{12.3}| \geq 0.3$ and or $|\rho_{13.2}| \geq 0.3$. Some non-symmetry is present for small values of $\rho_{12.3}$ and or $\rho_{13.2}$, but the power function becomes symmetric about zero when $\rho_{13.2}$ increases (compare plots a.1), a.2) and a.3)), even if $\rho_{12.3}$ is small;

- plots in panel b) - power for the tests of excluding edge 12 and edge 23 - power does not seem to vary with $\rho_{13.2}$ (the three plots are similar). Power increases as $|\rho_{12.3}|$ and or $|\rho_{23.1}|$ increase. Although globally the three plots look symmetric, there seems to be some non-symmetry, and also some non-monotonicity, in the black line ($\rho_{23.1} = 0.1$), which become more evident as $\rho_{13.2}$ increases (compare

plots b.1), b.2) and b.3));

- plots in panel c) - power for the tests of excluding edge 13 and edge 23 - power functions clearly non-symmetric and non-monotonic for small values of $\rho_{23.1}$, particularly noteworthy when $\rho_{13.2}$ is also small (black and blue lines in plots c.1) and c.2));

- the power for the two tests of excluding edge ij and edge ik varies mainly as a function of $\rho_{ij.k}$ and $\rho_{ik.j}$: it increases as $|\rho_{ij.k}|$ and or $|\rho_{ik.j}|$ increase. A combination with certain values of $\rho_{jk.i}$ seems to lead to some non-symmetry and some non-monotonicity.

Hence, it is possible to conclude that the problems of non-symmetry and non-monotonicity of the power functions are already present in the case of two tests for single edge exclusion. In the next section power functions are studied theoretically and an asymptotic normal approximation is derived. It will then become clear that the mean and correlation structures of the test statistics justify the non-symmetry and non-monotonicity of the power functions for certain combinations of partial correlation coefficients and sample sizes.

4.1.5 Normal approximation to the power of the test statistics for single edge exclusion in the three variables case

In Section 3.1.3 an asymptotic normal approximation to the distribution of the LRT statistic for single edge exclusion from the saturated GG model was derived. In the three variables case, under the alternative hypothesis that the saturated model holds (i.e, the three partial correlation coefficients are different from zero) the vector of test statistics T_{ij}^{L} is asymptotically normal distributed, with means given by $AE[T_{ij}^{L}] = -n\log(1-\rho_{ij.k}^{2})$, variances given by $\text{var}(T_{ij}^{L}) = 4n\rho_{ij.k}^{2}$, and covariances given by Equation 3.9, when the non-signed version of the LRT statistic is used. In the case the signed square-root version of the LRT statistic is used, formulae for the means $AE[T_{ij}^{signL}]$, variances $\text{var}(T_{ij}^{signL})$ and covariances $\text{cov}(T_{ij}^{signL}, T_{ik}^{signL})$ are summarised in Table 3.2. Hence, as in the two variables case, the asymptotic power for the LRT of selecting the saturated GG model with three variables can be obtained, using a trivariate normal approximation, as

$$P[T_{12}^L > 3.8414 \text{ and } T_{13}^L > 3.8414 \text{ and } T_{23}^L > 3.8414 \mid \rho_{12.3},\ \rho_{13.2},\ \rho_{23.1}] \stackrel{a}{=}$$

$$\int_{3.8414}^{+\infty} \int_{3.8414}^{+\infty} \int_{3.8414}^{+\infty} \Phi(\mu,\ \Sigma)\ dT_{12}^L\ dT_{13}^L\ dT_{23}^L, \tag{4.6}$$

where $\Phi(\mu,\ \Sigma)$ is the joint cumulative distribution of the joint trivariate normal density of the three LRT statistics with vector of means μ and variance matrix Σ given by $AE[T_{ij}^L]$, $\text{var}(T_{ij}^L)$ and $\text{cov}(T_{ij}^L, T_{ik}^L)$. As mentioned in Section 4.1.4, the variance matrix of the test statistics has to be positive definite.

The values of the integral of the joint cumulative distribution function can not be obtained directly with S-Plus, since its *pmvnorm* function only works if the variance matrix is the identity matrix, which is not the case. The *pmvnorm* function of the *mvtnorm* package of the programme R' was used to compute the distribution function of the multivariate normal distribution for the required limits, mean vectors and variance matrices. The function is based on algorithms by Genz and Bretz, which incorporate a sequence of transformations of the original integral before applying numerical integration.

Figure 4.9 compares simulated power values (in red) with theoretical power values obtained with the normal approximation just derived. Three different sample sizes are used: $n = 200$ (plots in panel a)), $n = 500$ (plots in panel b)) and $n = 1\,000$ (plots in panel c)). In each plot $\rho_{12.3}$, on the horizontal axis, varies between -0.9 and 0.9, within the region of positive definiteness. Four different combinations of values for $\rho_{13.2}$ and $\rho_{23.1}$ were chosen: 0.1 and 0.1 (plots 1)), 0.1 and 0.2 (plots 2)), 0.2 and 0.2 (plots 3)) and 0.2 and 0.3 (plots 4)). For other combinations of partial correlation coefficients, obtained power patterns are similar to those presented here. Some conclusions can be drawn from Figure 4.9: as expected, power increases as the sample size increases, and as the absolute values of the partial correlation coefficients increase. Theoretical values confirm simulation results from Section 4.1.4 that there is some non-symmetry and some non-monotonicity in the power functions, particularly for small sample sizes and values of one of the partial correlation coefficients close to zero (plots a.1) to a.3) and b.1)).

The non-symmetry and non-monotonicity of the power functions is accounted for by the structure of the means and of the correlations, in the asymptotic distribution, of

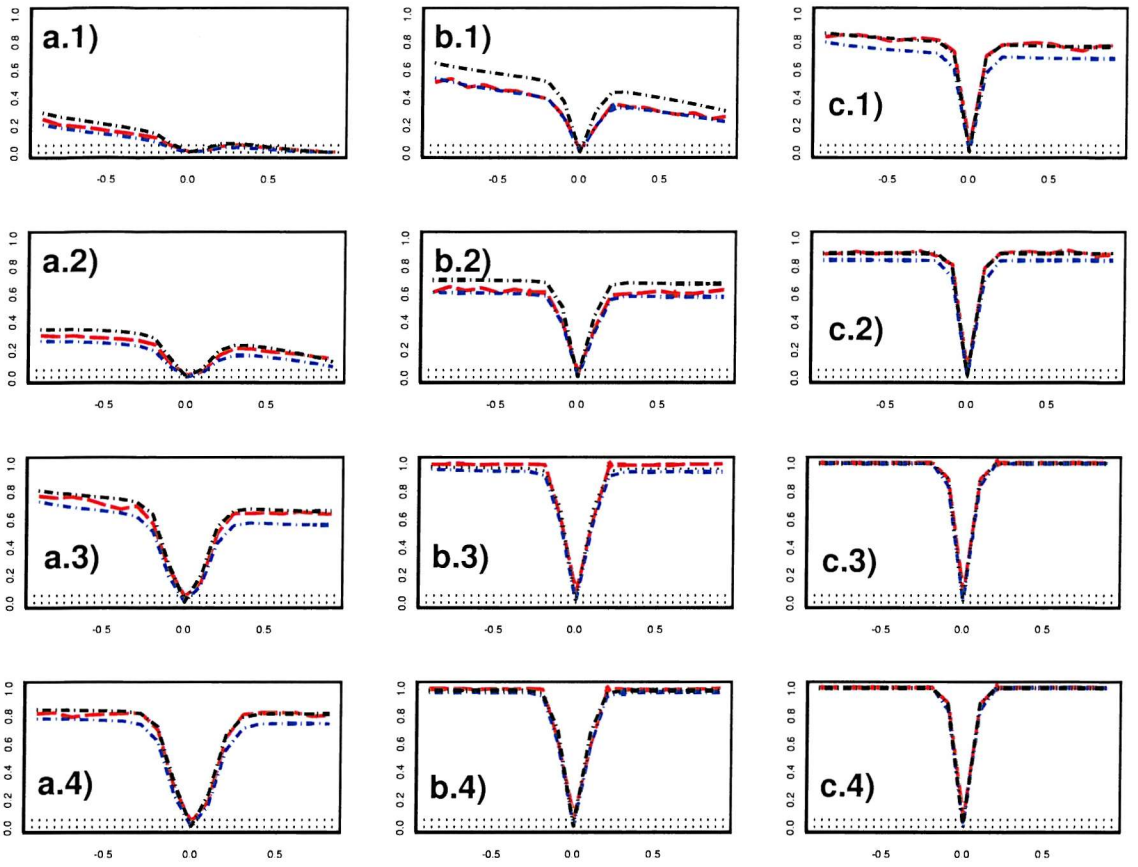

Figure 4.9: Simulated (in red) and theoretical normal (in blue) and truncated normal (in black) power curves, for saturated model, using the LRT statistic. In all plots $\rho_{12.3}$ on the horizontal axis. Plots a) $n = 200$, plots b) $n = 500$, plots c) $n = 1000$. Different combinations of values for $\rho_{13.2}$ and $\rho_{23.1}$: plots 1) 0.1 and 0.1, plots 2) 0.1 and 0.2, plots 3) 0.2 and 0.2, plots 4) 0.2 and 0.3.

the three test statistics. Consider the case of plot b.1) in Figure 4.9. The sample size equals 500, $\rho_{13.2} = \rho_{23.1} = 0.1$ and $\rho_{12.3}$ ranges from -0.9 to 0.9. The expected values of the three likelihood ratio test statistics T_{ij}^L, in the asymptotic distribution, for certain combinations of partial correlation values are:

$\rho_{12.3}$; $\rho_{13.2}$; $\rho_{23.1}$	-0.6; 0.1; 0.1	-0.1; 0.1; 0.1	0.2; 0.1; 0.1	0.9; 0.1; 0.1
$AE[T_{12}^L]$	223.140	5.025	20.410	830.400
$AE[T_{13}^L]$	5.025	5.025	5.025	5.025
$AE[T_{23}^L]$	5.025	5.025	5.025	5.025

The corresponding correlation structure for the three test statistics, in the asymptotic distribution, follows:

$\rho_{12.3}$; $\rho_{13.2}$; $\rho_{23.1}$	-0.6; 0.1; 0.1	-0.1; 0.1; 0.1	0.2; 0.1; 0.1	0.9; 0.1; 0.1
$cor[T_{12}^L, T_{13}^L]$	0.070	0.095	- 0.111	- 0.136
$cor[T_{12}^L, T_{23}^L]$	0.070	0.095	- 0.111	- 0.136
$cor[T_{13}^L, T_{23}^L]$	0.597	0.095	- 0.205	-0.900

The correlations between the three test statistics are always positive if $\rho_{12.3}$ is negative and both $\rho_{13.2}$ and $\rho_{23.1}$ are positive. The correlations are always negative if the three partial correlation coefficients are positive. However, the expected values are always positive. A correlation coefficient of -0.9, associated with expected values of $\simeq 0.5$, makes it more likely for the values of T_{13}^L and T_{23}^L to be below the critical value of 3.8414 than does a correlation coefficient of 0.597 associated with the same expected values. All this accounts for the non-symmetry and non-monotonicity of the power functions in plot b.1). When the value of one of the partial correlation coefficients increases, say $\rho_{23.1} = 0.2$, as in plot b.2), the expected value of T_{23}^L, in the asymptotic distribution, increases to 20.41. Although the correlation structure changes slightly, it becomes less likely for the values of T_{23}^L to be below the critical value of 3.8414 for different combinations of $\rho_{12.3}$. For that reason there is no non-symmetry and no non-monotonicity in plot b.2) Figure 4.9.

When assessing the quality of the normal approximation to the power of selecting the saturated model, using the LRT statistic, it is possible to conclude the approximation holds asymptotically, i.e., it performs better for larger sample sizes (see plots in panel c), associated with $n = 1\,000$) and for values of the partial correlation coefficients not close to zero (plots c.3) and c.4)). Recall that a similar conclusion was drawn in the two variables case, Section 4.1.1.

The reason why, in Figure 4.9, plots 1 (corresponding to values of $\rho_{13.2}$ and $\rho_{23.1}$ close to zero), the normal approximation seems to perform better for small rather than for large sample sizes (plots a.1) versus c.1)) is not yet clear. It is also not clear why, for $n = 200$ (plots in panel a) the curves of the simulated and theoretical values seem to get further apart, indicating a poorer approximation, as $\rho_{13.2}$ increases (plots a.2) and a.3)). In order to try to account for this non-justified pattern of the power functions, a significance level of 10% was also considered and corresponding simulated and theoretical power values were obtained. The obtained plots are not presented because they are very similar to those in Figure 4.9, where a size of 5% is considered.

Therefore, varying the size level was not conclusive. A larger sample size was also used, $n = 2\,000$ and in this case the simulated and the theoretical power curves are very close, even for $\rho_{13.2}$ and $\rho_{23.1}$ equal to 0.1 (the case of plots 1). This confirms the proposed normal approximation holds asymptotically.

Additionally, the *total non-admissible region* was quantified. Indeed, a trivariate normal approximation has been proposed to the power of selecting the saturated model, with three variables, using the LRT. However, the values of the test statistic are always non-negative. Thus, the total non-admissible region (denoted as NAR) corresponds to all situations when at least one test statistic is negative and is obtained as one minus the probability that all three test statistics are greater, or equal, to zero. Values of this total NAR were calculated for the different sample sizes and the different combinations of partial correlation coefficients used in Figure 4.9. Results indicate that, as expected, NAR values are higher for smaller sample sizes, particularly for smaller partial correlation coefficients. Appendix Table C.1 lists the NAR values. Such values were used to obtain a 'truncated' normal approximation to the power functions. Indeed, since the total admissible region should be one, and is not, theoretical power values given by the normal approximation (represented by the blue curves in Figure 4.9) where divided by the actual admissible region (equal to one minus NAR). The corrected power values are represented by the black curves in Figure 4.9. This correction has lead theoretical power values to increase, improving the quality of the asymptotic normal approximation in the case the sample size is large, even for small values of the partial correlation coefficients - note the overlapping of the red and of the black curves in panels c.1) and c.2). The proposed truncated normal approximation is of no use for small sample sizes and small partial correlation coefficients - plots 1) and 2) in panels a) and b).

The asymptotic power for the signed square-root LRT of selecting the saturated GG model with three variables can be obtained, using a trivariate normal approximation, as follows:

- in the case of a one-sided hypothesis test

$$P[T_{12}^{signL} > 1.645 \text{ and } T_{13}^{signL} > 1.645 \text{ and } T_{23}^{signL} > 1.645 \mid \rho_{12.3},\ \rho_{13.2},\ \rho_{23.1}] \stackrel{a}{=} \int_{1.645}^{+\infty}\int_{1.645}^{+\infty}\int_{1.645}^{+\infty} \Phi\left(\mu^{sign},\ \Sigma^{sign}\right)\, dT_{12}^{signL}\, dT_{13}^{signL}\, dT_{23}^{signL}; \qquad (4.7)$$

- in the case of a two-sided hypothesis test

$$P[|T_{12}^{signL}| > 1.96 \text{ and } |T_{13}^{signL}| > 1.96 \text{ and } |T_{23}^{signL}| > 1.96 \mid \rho_{12.3},\ \rho_{13.2},\ \rho_{23.1}] \stackrel{a}{=}$$

$$\int\int\int_D \Phi\left(\mu^{sign},\ \Sigma^{sign}\right)\, dT_{12}^{signL}\, dT_{13}^{signL}\, dT_{23}^{signL}; \tag{4.8}$$

where $\Phi\left(\mu^{sign},\ \Sigma^{sign}\right)$ is the joint cumulative distribution of the joint trivariate normal density of the signed square-root versions of the three LRT statistics with vector of means μ^{sign} and variance matrix Σ^{sign} given by $AE[T_{ij}^{signL}]$, $\text{var}(T_{ij}^{signL})$ and $\text{cov}(T_{ij}^{signL}, T_{ik}^{signL})$. The variance matrix is constrained to be positive definite. D, the domain of integration in the case of a two-sided hypothesis test, is the region where each of the three test statistics is in $-\infty$ to -1.96 or 1.96 to $+\infty$, i.e., $D = \{(-\infty, -1.96) \cup (1.96, +\infty)\}^3$.

Figure 4.10 compares simulated power values (in red) with theoretical power values (in blue), using a one-sided hypothesis test with the signed square-root LRT statistic. Three different sample sizes are used: $n = 200$ (plots in panel a)), $n = 500$ (plots in panel b)) and $n = 1\,000$ (plots in panel c)). In each plot $\rho_{12.3}$, on the horizontal axis, varies between 0 and 0.9, within the region of positive definiteness. As before, four different combinations of values for $\rho_{13.2}$ and $\rho_{23.1}$ were chosen: 0.1 and 0.1 (plots 1)), 0.1 and 0.2 (plots 2)), 0.2 and 0.2 (plots 3)) and 0.2 and 0.3 (plots 4)). Some comments on Figure 4.10: the normal approximation is a good approximation to the power of the one-sided signed square-root LRT statistic for selecting the saturated model, even for a sample size of 200 (plots in panel a) and partial correlation values close to zero (plots 1 and 2). Power increases as the sample size increases and as the partial correlation coefficients increases. However, some non-monotonicity seems to exist for small sample sizes ($n = 200$) and small values of ρ (see plots a.1) and a.2)). The reason is the following: the correlations between the three test statistics, in the asymptotic distribution, are always negative and, for a sample size of 200, the expected values of the test statistics, in the asymptotic distribution, equal 1.42 and 2.86, respectively for partial correlations coefficients of 0.1 and 0.2. It is, therefore, very likely that values of the test statistics are below the critical value of 1.96, and even decrease as $\rho_{12.3}$ increases (because of the negative correlations). Hence power decreases, as in plot a.1).

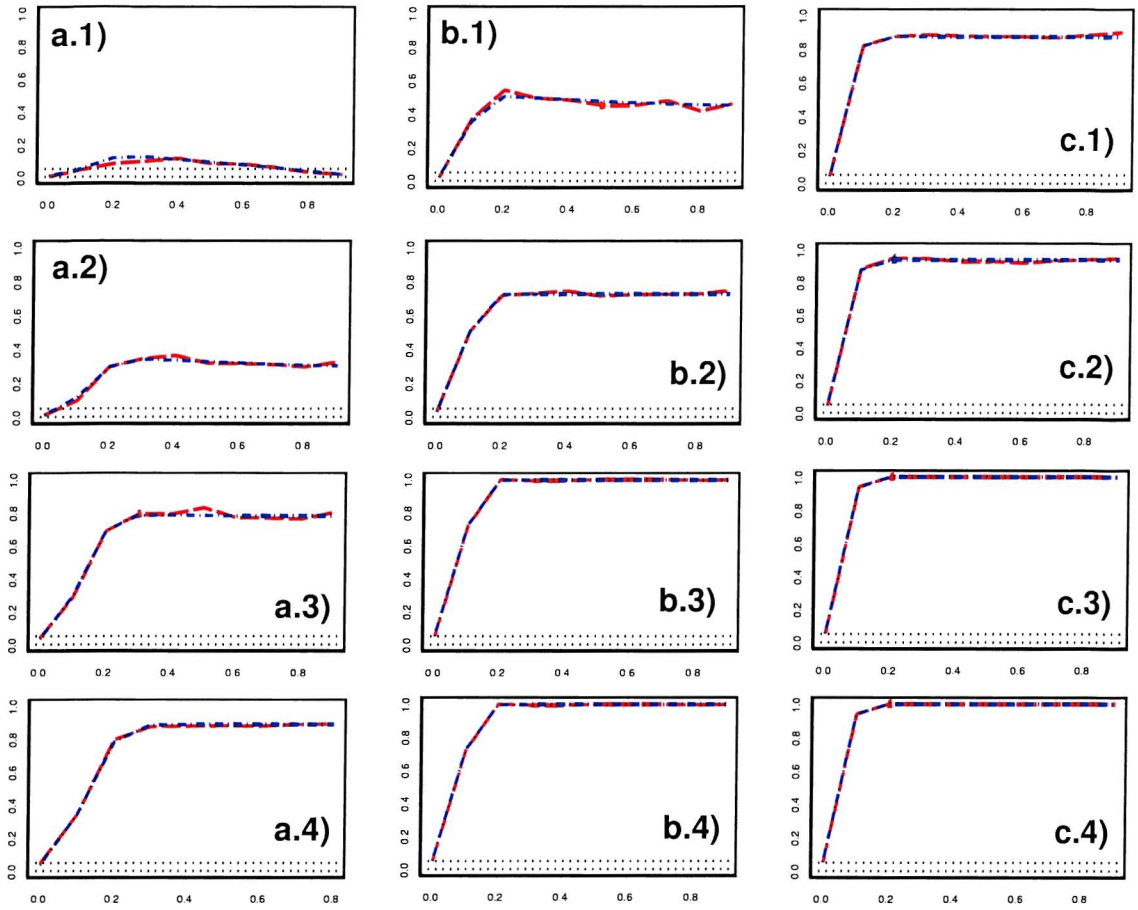

Figure 4.10: Simulated (in red) and theoretical normal (in blue) power curves, for saturated model, using the signed square-root LRT statistic, one-sided hypothesis test. In all plots $\rho_{12.3}$ on the horizontal axis. Plots a) $n = 200$, plots b) $n = 500$, plots c) $n = 1000$. Different combinations of values for $\rho_{13.2}$ and $\rho_{23.1}$: plots 1) 0.1 and 0.1, plots 2) 0.1 and 0.2, plots 3) 0.2 and 0.2, plots 4) 0.2 and 0.3.

Analogously, Figure 4.11 compares simulated power values (in red) with theoretical power values (in blue), using a two-sided hypothesis test with the signed square-root LRT statistic. In all plots $\rho_{12.3}$, on horizontal axis, varies between -0.9 and 0.9, within the region of positive definiteness. The normal approximation is a good approximation to the power of selecting the saturated model using the two-sided signed square-root LRT, even for small sample sizes and small partial correlation coefficients. Note similar conclusions were drawn in the two variables case (Section 4.1.1).

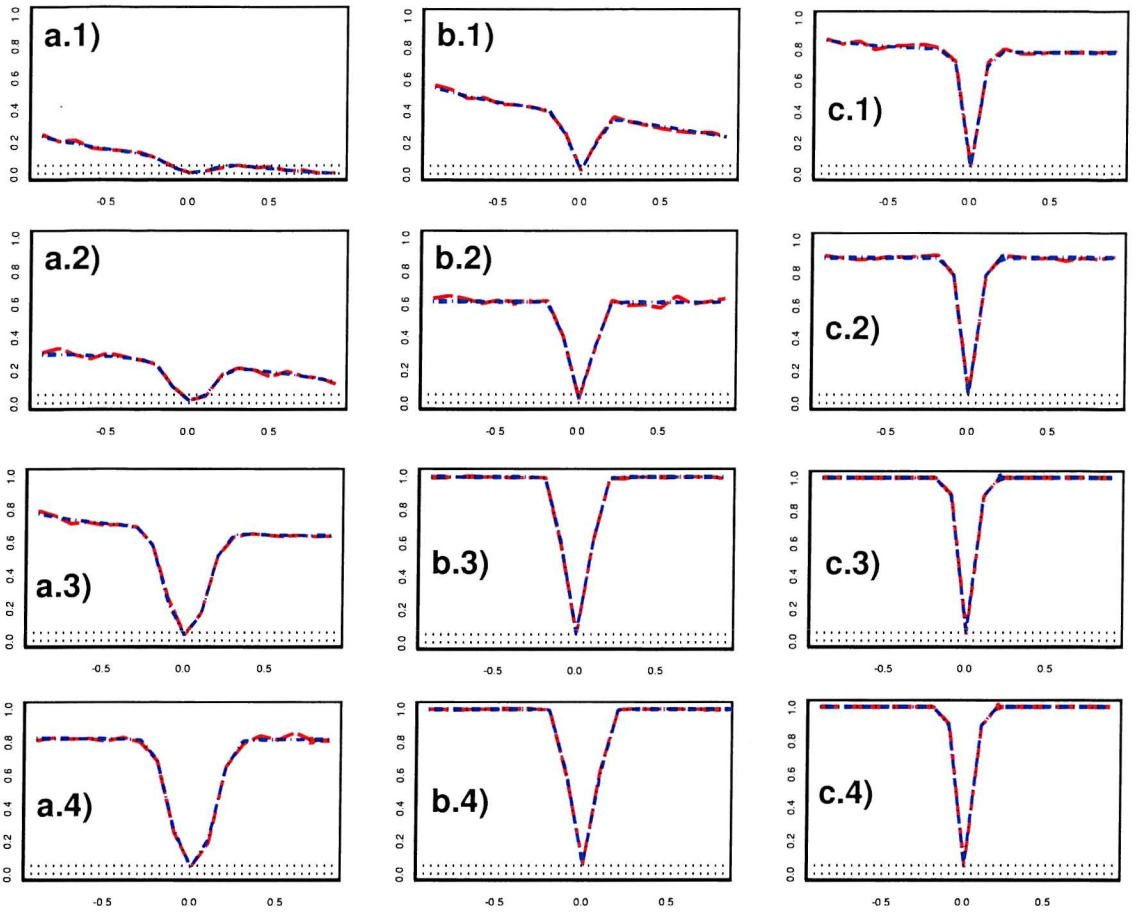

Figure 4.11: Simulated (in red) and theoretical normal (in blue) power curves, for saturated model, using the signed square-root LRT statistic, two-sided hypothesis test. In all plots $\rho_{12.3}$ on the horizontal axis. Plots a) $n = 200$, plots b) $n = 500$, plots c) $n = 1000$. Different combinations of values for $\rho_{13.2}$ and $\rho_{23.1}$: plots 1) 0.1 and 0.1, plots 2) 0.1 and 0.2, plots 3) 0.2 and 0.2, plots 4) 0.2 and 0.3.

Until now the current section has dealt with deriving power functions for selecting the saturated model, using the LRT statistic. In Section 4.1.4 a simulation study was used to estimate the power for the test of excluding edge ij and the power for the two tests of excluding edge ij and edge ik, using the non signed version of the LRT statistic. The results of the simulation, for a sample size of 200, are summarised in Figures 4.7 and 4.8. A theoretical asymptotic normal approximation to the power for the test(s) of excluding edge ij (and edge ik) is now derived.

Even if there are three variables in the model, the power for the test of excluding edge ij is obtained, as in the two variables case (with the correlation coefficient ρ_{12}

being replaced by the partial correlation coefficient $\rho_{ij.k}$), as

$$P[T_{ij}^L > 3.8414 \mid \rho_{12.3},\ \rho_{13.2},\ \rho_{23.1}] \stackrel{a}{=} P\left[z > \frac{3.8414 + n\log(1-\rho_{ij.k}^2)}{2\sqrt{n}\,|\rho_{ij.k}|}\right],$$

where $z \sim N(0,\ 1)$ and is not dependent on the values of any of the remaining partial correlation coefficients. Such behaviour was pointed out when commenting on simulated values presented in Figure 4.7. In brief: the power for the test of excluding edge ij increases as n increases, as $|\rho_{ij.k}|$ increases, and is symmetric, about zero, in $\rho_{ij.k}$.

An asymptotic bivariate normal approximation to the power for the tests of excluding edge ij and edge ik , when the LRT statistic is used, is now proposed as

$$P[T_{ij}^L > 3.8414 \text{ and } T_{ik}^L > 3.8414 \mid \rho_{12.3},\ \rho_{13.2},\ \rho_{23.1}] \stackrel{a}{=}$$

$$\int_{3.8414}^{+\infty}\int_{3.8414}^{+\infty} \Phi(\mu,\ \Sigma)\ dT_{ij}^L\ dT_{ik}^L,$$

where $\Phi(\mu,\ \Sigma)$ is the joint cumulative distribution of the joint bivariate normal density of the two test statistics with vector of means μ and variance matrix Σ given by $AE[T_{ij}^L]$, $\text{var}(T_{ij}^L)$ and $\text{cov}(T_{ij}^L, T_{ik}^L)$. The variance matrix has to be positive definite. Note that the mean and the variance of T_{ij}^L are a function of n and of $\rho_{ij.k}$, whereas the covariance between T_{ij}^L and T_{ik}^L depends not only on $\rho_{ij.k}$ and $\rho_{ik.j}$ but also on $\rho_{jk.i}$. For that reason some non-symmetry and non-monotonicity of the power functions can be observed for certain combinations of values of the three partial correlation coefficients, as detected in Figure 4.9.

Figure 4.12 compares simulated power values (in red) with theoretical values (in blue) obtained with the normal approximation just derived, for the tests of excluding edge 13 and edge 23, using the LRT statistic. As before, three different sample sizes are used: $n = 200$ (plots in panel a)), $n = 500$ (plots in panel b)) and $n = 1\,000$ (plots in panel c)). In each plot $\rho_{12.3}$, on the horizontal axis, varies between -0.9 and 0.9, within the region of positive definiteness. The four different combinations of values for $\rho_{13.2}$ and $\rho_{23.1}$ are: 0.1 and 0.1 (plots 1)), 0.1 and 0.2 (plots 2)), 0.2 and 0.2 (plots 3)) and 0.2 and 0.3 (plots 4)). Some comments on Figure 4.12: the power for the two tests of excluding edge 13 and edge 23 depends not only on $\rho_{13.2}$ and $\rho_{23.1}$ (it clearly increases as these increase) but it also depends on $\rho_{12.3}$. Indeed, not all the lines are horizontal, even for large sample sizes: see for example the case of $\rho_{13.2} = 0.1$ and $\rho_{23.1} = 0.1$, when $n = 1\,000$ (plot c.1)) and also the case of $\rho_{13.2} = 0.1$ and $\rho_{23.1} = 0.2$,

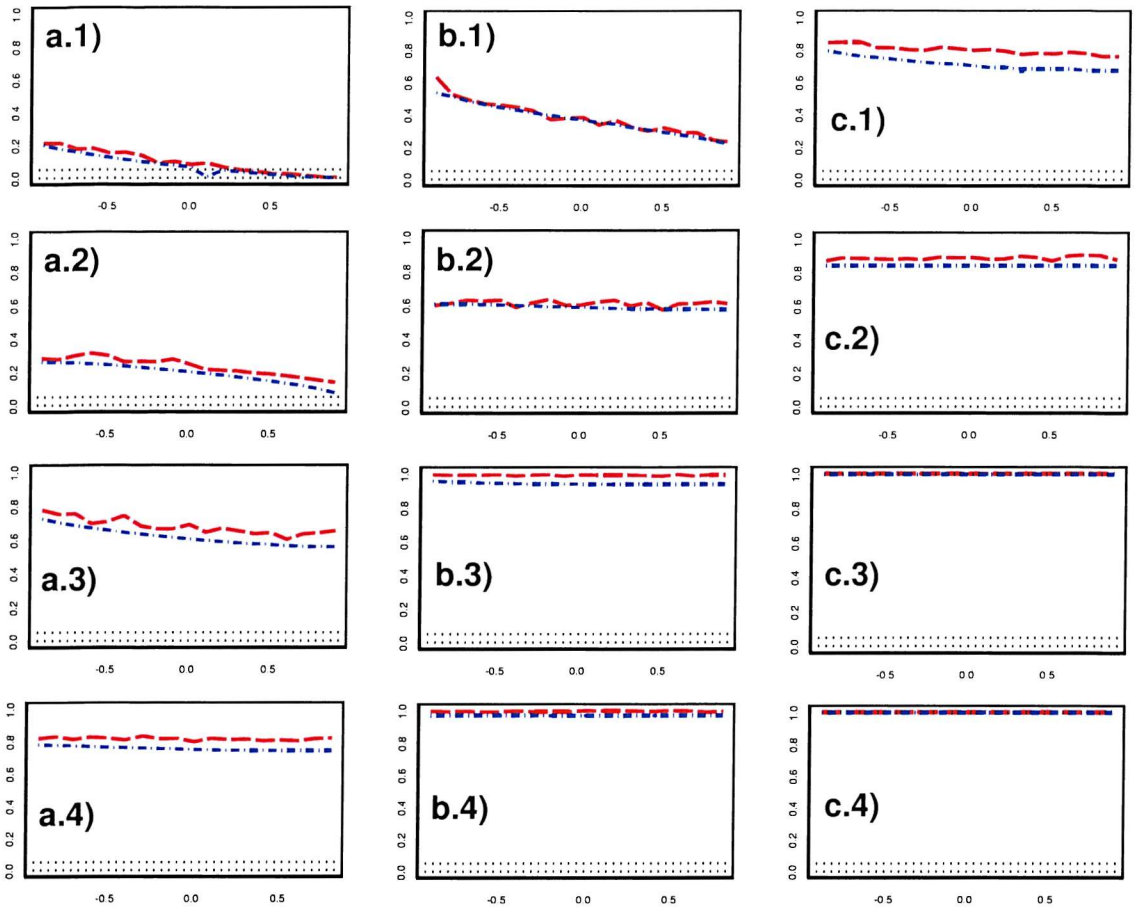

Figure 4.12: Simulated (in red) and theoretical normal (in blue) power curves, for the tests of excluding edge 13 and edge 23, using the LRT statistic. In all plots $\rho_{12.3}$ on the horizontal axis. Plots a) $n = 200$, plots b) $n = 500$, plots c) $n = 1000$. Different combinations of values for $\rho_{13.2}$ and $\rho_{23.1}$: plots 1) 0.1 and 0.1, plots 2) 0.1 and 0.2, plots 3) 0.2 and 0.2, plots 4) 0.2 and 0.3.

when $n = 500$ (plot b.2)). The normal approximation performs better asymptotically, i.e., for large sample sizes, particularly for partial correlation coefficients not close to zero (plots c.3) and c.4)).

The justification for the non-symmetry and non-monotonicity of the power functions in Figure 4.12, plots a.1), a.2) a.3) and b.1), is basically the same as the one given when commenting on Figure 4.9. Since only two test statistics are now used a graphical display is also presented. Consider the case of plot b.1) in Figure 4.12. The sample size equals 500 and $\rho_{13.2} = \rho_{23.1} = 0.1$. Consider three possible values for $\rho_{12.3}$: -0.6, 0.2 and 0.9. The expected values of the two test statistics T_{13}^L and T_{23}^L, in the asymptotic distribution, equal 5.025, not depending on the value of $\rho_{12.3}$. However,

the correlation coefficient between T_{13}^L and T_{23}^L is a function of $\rho_{12.3}$ and equals 0.597, -0.205 and -0.9, respectively for $\rho_{12.3}$ equal to -0.6, 0.2 and 0.9. When $\rho_{23.1} = 0.2$, plot b.2) in Figure 4.12 (where the non-symmetry and non-monotonicity are no longer present), the expected value of T_{23}^L, in the asymptotic distribution, equals 20.41 and the correlation coefficients between T_{13}^L and T_{23}^L change to 0.594, -0.21 and -0.9. Recall that the red lines in plots b.1) and b.2) correspond to simulated power values, i.e., to the overall probability that both T_{13}^L and T_{23}^L are greater than the critical value of 3.8414, for the different combinations of the three partial correlation coefficients. Consequently, plotted values are the number of times, out of 1 000 (repetitions), that both test statistics are greater than 3.8414, for the 1 000 samples generated from a bivariate normal distribution with the specified correlation structure.

Figure 4.13 shows the 1 000 pairs of simulated values of T_{13}^L and T_{23}^L, for the six combinations of $\rho_{12.3}; \rho_{13.2}; \rho_{23.1}$ chosen above: plot a) -0.6; 0.1; 0.1, plot b) 0.2; 0.1; 0.1, plot c) 0.9; 0.1; 0.1, plot d) -0.6; 0.1; 0.2, plot e) 0.2; 0.1; 0.2 and plot f) 0.9; 0.1; 0.2. The values of T_{13}^L are displayed on the horizontal axis, the vertical axis corresponding to values of T_{23}^L. The red lines represent the critical value of a chi-square distribution on one *df*, for a 5% size: 3.8414. In each plot the number of dots in the region where both test statistics are greater then 3.8414, out of 1 000, gives the overall power for the two tests of excluding edges 13 and 23. Plots a), b) and c) correspond to three power values in the red line of plot b.1), Figure 4.12: respectively 0.466, 0.362 and 0.229. Power in plot c), where $\text{cor}(T_{13}^L, T_{23}^L) = -0.9$, is much lower than in plot a), where $\text{cor}(T_{13}^L, T_{23}^L) = 0.597$. The sign of the correlation coefficient between the two test statistics, together with small ($\simeq 5$) expected values for the test statistics, accounts for the non-symmetry and non-monotonicity of the power function. Plots d), e) and f) correspond to power values of 0.62, 0.62 and 0.61. Although $\text{cor}(T_{13}^L, T_{23}^L)$ is the same in plots c) and f), the fact that in plot f) $\rho_{23.1} = 0.2$ has lead to an increase in the expected value of T_{23}^L (in the asymptotic distribution), making it less likely that values of T_{23}^L are under 3.8414. As a result, power values do not decrease in plot f) as they do in plot c).

In brief: the mean and correlation structures of the test statistics accounts for the non-symmetry and non-monotonicity of the power functions, noteworthy for small values of the partial correlation coefficients.

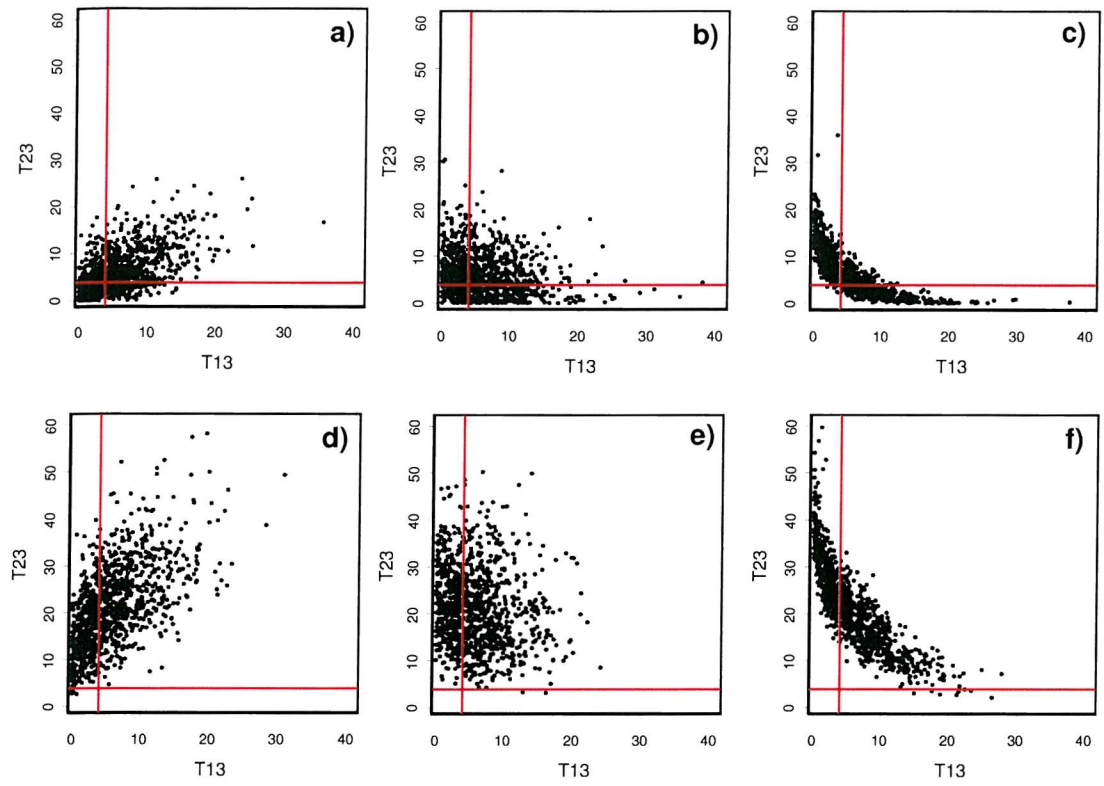

Figure 4.13: T_{13}^L and T_{23}^L simulated values, for six combinations of $\rho_{12.3}; \rho_{13.2}; \rho_{23.1}$: a) -0.6; 0.1; 0.1, b) 0.2; 0.1; 0.1, c) 0.9; 0.1; 0.1, d) -0.6; 0.1; 0.2, e) 0.2; 0.1; 0.2, f) 0.9; 0.1; 0.2.

Until now normal approximations to the power of the likelihood ratio test statistic have been derived, both concerning selecting the saturated model (with three variables) and test(s) of exluding edge ij (and edge ik). Non-signed and signed square-root versions of the LRT have been considered. Similar reasoning can be applied to derive asymptotic power functions for the Wald and the score test statistics. Consequently, asymptotic power functions (saturated model) for the Wald or the score test statistics can be obtained by replacing in Equation 4.6, respectively

- T_{ij}^L by T_{ij}^W or T_{ij}^S;
- $AE[T_{ij}^L]$ by $AE[T_{ij}^W]$ or $AE[T_{ij}^S]$;
- $\text{var}(T_{ij}^L)$ by $\text{var}(T_{ij}^W)$ or $\text{var}(T_{ij}^S)$;
- $\text{cov}(T_{ij}^L, T_{ik}^L)$ by $\text{cov}(T_{ij}^W, T_{ik}^W)$ or $\text{cov}(T_{ij}^S, T_{ik}^S)$.

Analogously, signed square-root versions of the Wald and score test statistics can be considered, once the appropriate replacements are made in Equation 4.7 for one-sided

hypothesis tests and in Equation 4.8 for two-sided hypothesis tests. The conclusions are expected to be similar to those drawn so far, using the LRT, as happened in the two variables case.

4.1.6 Power of single edge exclusion tests, in GG models: the p variables case

The normal approximation to the power of the test statistics for single edge exclusion from the saturated GG model, with three variables, derived in Section 4.1.5, can be generalised to the p variables case, using a multivariate normal approximation, by generalising Equations 4.6, 4.7 and 4.8. One should note, however, that a model with four variables implies six partial correlation coefficients (and six test statistics for single edge exclusion from the saturated model) and a model with five variables implies ten (i.e., $p(p-1)/2$) test statistics for single edge exclusion. Therefore, calculating the power of the test statistics for single edge exclusion in a model with $p = 5$ variables requires using a multivariate normal distribution of dimension ten, i.e., a ten dimensional integral.

When p is large and numerical integration becomes impractical it is always possible to perform Monte Carlo integration. This would first require generating a large number of observations (say, one million) from a $p(p-1)/2$ - dimensional multivariate normal distribution with vector of means $AE[T^{\bullet}_{ij}]$ (the expected values of the test statistics, in the asymptotic distribution) and variance matrix given by $\mathrm{cov}(T^{\bullet}_{ij}, T^{\bullet}_{kl})$ (the variances and covariances of the test statistics, in the asymptotic distribution). Generating a multivariate normal distribution of dimension $p(p-1)/2$ is straightforward: it requires generating $p(p-1)/2$ univariate standard normals, combining them in a vector, premultiplying that vector by the positive square-root of the variance matrix $\mathrm{cov}(T^{\bullet}_{ij}, T^{\bullet}_{kl})$ and adding the result to the vector of means $AE[T^{\bullet}_{ij}]$ (see Mardia, Kent and Bibby, 1979, Section 2.5.1). The power of the model selection procedure, i.e., the Monte Carlo probability of selecting the saturated model is obtained as the number of values, out of a million, in the region between zero and the density of the multivariate normal distribution, from the upper limits of integration to infinity.

The complexity of the problem increases considerably with the number of variables in the model. Because of this complexity, the decision was made not to investigate

the quality of the normal approximation when four or more variables are present. Furthermore, performing simulations becomes much more complex when the number of variables increases (in the sense that there are many possible combinations of values of the ρ, with the additional constraint of a positive definite scaled inverse variance matrix). Also, while it is possible to visualise the associations between three variables in a three dimensional space, this becomes no longer possible when six or ten associations are present. However, it was decided to consider the four and the five variables cases, restricting the attention to models with equal partial correlation coefficients between all variables, and to obtain power functions by simulation.

First, it is required to express the region of positive definiteness as a function of the partial correlation coefficients. An *equicorrelation matrix* (see Mardia, Kent and Bibby, 1979, page 461) is a $p \times p$ matrix of the type

$$P = (1 - \rho)I + \rho J, \tag{4.9}$$

that is, a matrix with ones on the main diagonal, all off-diagonal elements being equal to the correlation coefficient ρ. J denotes a $p \times p$ matrix of ones. The determinant of P is given by the product of its eigenvalues, where $\lambda_1 = 1+(p-1)\rho$ and $\lambda_2 = \ldots = \lambda_p = 1-\rho$. P is positive definite when all eigenvalues are positive, that is, when $1 - \rho > 0$ and $1 + (p-1)\rho > 0$. Consequently, P is positive definite when $\rho \in \left(\frac{-1}{p-1}\,,\ 1\right)$.

Because the focus is on partial correlation coefficients, existing results for an equicorrelation matrix had to be adapted. Indeed, the scaled inverse variance matrix, with ones on the main diagonal and off-diagonal elements being minus the partial correlation coefficients, can be written as

$$T = sc(P^{-1}) = (1 + \rho_{ij.rest})I - \rho_{ij.rest}J \tag{4.10}$$

The determinant of T is given by the product of its eigenvalues, where $\lambda_1 = 1 - (p-1)\rho_{ij.rest}$ and $\lambda_2 = \ldots = \lambda_p = 1 + \rho_{ij.rest}$. T is positive definite when $1 + \rho_{ij.rest} > 0$ and $1 - \rho_{ij.rest}(p-1) > 0$. Because $\rho_{ij.rest} \in (-1\,,\ 1)$, $1 + \rho_{ij.rest}$ is always positive and $[1 - \rho_{ij.rest}(p-1)]$ is strictly positive if $\rho_{ij.rest} < \frac{1}{p-1}$. In other words, it is suggested that the positive definiteness constraint in the scaled inverse variance matrix with p variables, when all partial correlation coefficients $\rho_{ij.rest}$ are equal, is that $\rho_{ij.rest} \in \left(-1\,,\ \frac{1}{p-1}\right)$.

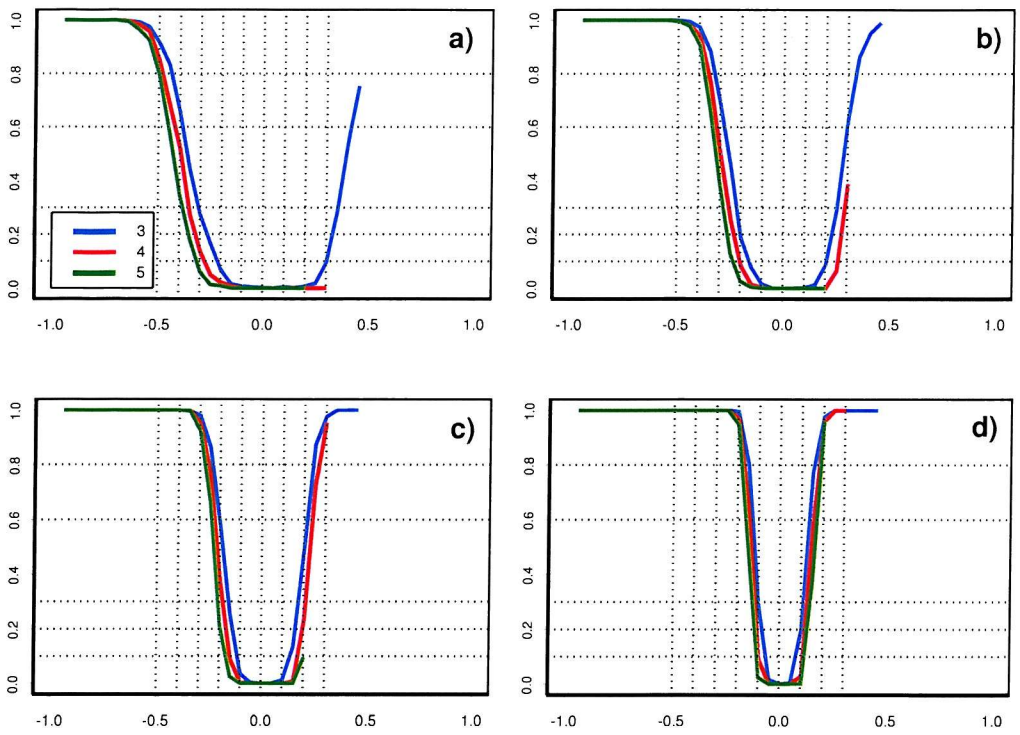

Figure 4.14: Power of selecting the saturated GG model (with three, four or five variables) when all partial correlation coefficients are equal. Four different sample sizes: a) $n = 50$, b) $n = 100$, c) $n = 200$, d) $n = 500$.

Simulations were performed for the cases of three, four and five manifest variables, in order to estimate the power for selecting the saturated model, i.e., the probability of selecting the saturated model given the specified values for $\rho_{ij.rest}$. All partial correlation coefficients are assumed equal, and varying from -0.95 to $\frac{1}{p-1}$ with an interval of 0.05. In each case 1 000 repetitions were done. Four different sample sizes are considered: 50, 100, 200 and 500. The likelihood ratio test for single edge exclusion is used. The results obtained are presented in Figure 4.14.

Some conclusions can be drawn:

- for a certain value of the partial correlation coefficient, the smaller the number of variables, the larger power is. This effect is particular noteworthy for small absolute values of the partial correlation, whereas for large absolute values the differences vanish;

- with large sample sizes, power tends to increase faster as the absolute value of the partial correlation coefficient increases, and, therefore, differences in power

between models with three, four and five variables can only be detected for small absolute values of $\rho_{ij.rest}$: less than |0.2| with a sample size of 500 (see panel d)).

The LRT statistic was used. Similar results should be obtained using either the Wald or the score test statistics.

4.2 Power of Single Edge Exclusion in GLL Models

The power of the test statistics for single edge exclusion from the saturated GLL model is investigated. Non-signed and signed square-root versions of the test statistics are considered. The two binary variables case is analysed and some insight is given regarding three binary variables. As in the continuous case, the power of a backwards elimination model selection procedure for selecting the saturated GLL model is defined as the probability of selecting the true (saturated) GLL model given the specified true model parameters. In the continuous case the association between variables is measured by a single parameter, the (partial) correlation coefficient, whereas in the case of two or more binary variables cross-classified in a contingency table, more parameters are required, the total number of parameters depending on the number of binary variables in the GLL model.

4.2.1 Power of the LRT in a GLL model with two binary variables

The odds ratio, ψ_{12}, is a commonly used measure of association between two binary variables, but additional information is required. As in Section 3.7, besides ψ_{12}, $\pi_1(0)$ and $\pi_2(0)$ are considered fixed. The LRT statistic is used. Simulated power values are presented. The theoretical power of the LRT statistic is then derived, using asymptotic normal and non-central chi-square approximations. The quality of the two approximations is assessed.

Simulated power values

For the two binary variables case, power is calculated for different combinations of the three chosen parameters ψ_{12}, $\pi_1(0)$ and $\pi_2(0)$. Because of the symmetries explained in Section 3.7, $\pi_1(0)$, takes values between 0.1 and 0.9 (an interval of 0.1 is used), $\pi_2(0)$

takes values between 0.1 and 0.5 and the odds ratio ψ_{12} takes values greater or equal to one (1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 3 and 4). A total sample size of 1 000 observations is considered. The likelihood ratio test statistic for single edge exclusion from the saturated model, given by Equation 3.17, is used. Recall that the null hypothesis is that the independence model holds, the alternative being that the saturated model holds. Power is estimated as the number of times, out of 1 000, that the saturated model is chosen, that is, that $P\left[LRT_{12} > \chi^2_{1;\,0.95} \mid \pi_1(0),\, \pi_2(0),\, \psi_{12}\right]$.

Figure 4.15 shows simulated power values, for $n_\emptyset = 1\,000$, obtained for the different combinations of odds ratio (on the horizontal axis of each plot) and marginal probabilities. In each plot each of the nine lines corresponds to a value of $\pi_1(0)$ (from 0.1 to 0.9). $\pi_2(0)$ has a value of 0.1 in panel a), 0.2 in b), 0.3 in c), 0.4 in d) and 0.5 in e).

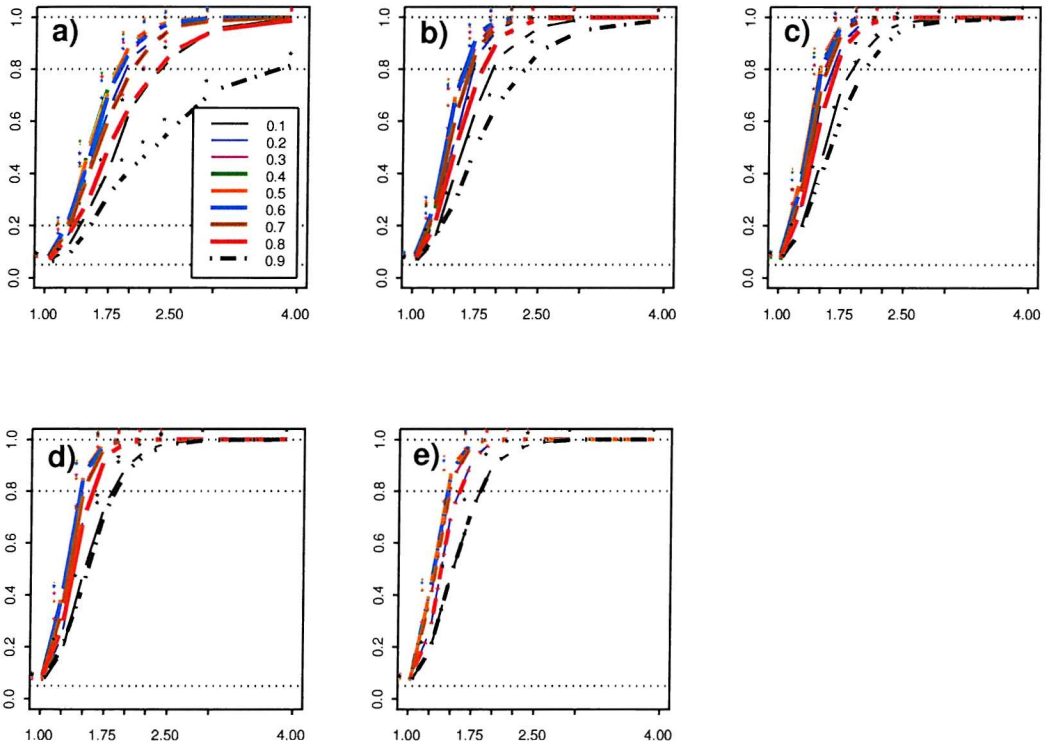

Figure 4.15: Simulated power of the saturated model, with two binary variables, for a sample size $n_\emptyset = 1\,000$ and different combinations of ψ_{12} (from 1 to 4, on the x axis), $\pi_1(0)$ (from 0.1 to 0.9 in each plot) and $\pi_2(0)$ (from 0.1 to 0.5 in panels a) to e), respectively).

Some conclusions can be drawn from Figure 4.15:

- for a given combination of marginal probabilities, power increases as the odds

ratio deviates from one;

- power increases faster when one of the marginal probabilities is around 0.4 or 0.5, particularly if the other marginal probability also takes that value;
- if $\pi_2(0) = 0.1$ (plot in panel a) a power value of $\simeq 0.8$ is achieved with $\psi_{12} = 2.25$, for values of $\pi_1(0)$ of 0.8, whereas if $\pi_2(0) = 0.3$ (plot in panel c) similar values of power are achieved for $\psi_{12} = 1.5$. Indeed, more balanced combinations of marginal probabilities $\pi_1(0)$, $\pi_2(0)$ lead to higher power values, even for smaller odds ratios ψ_{12}.

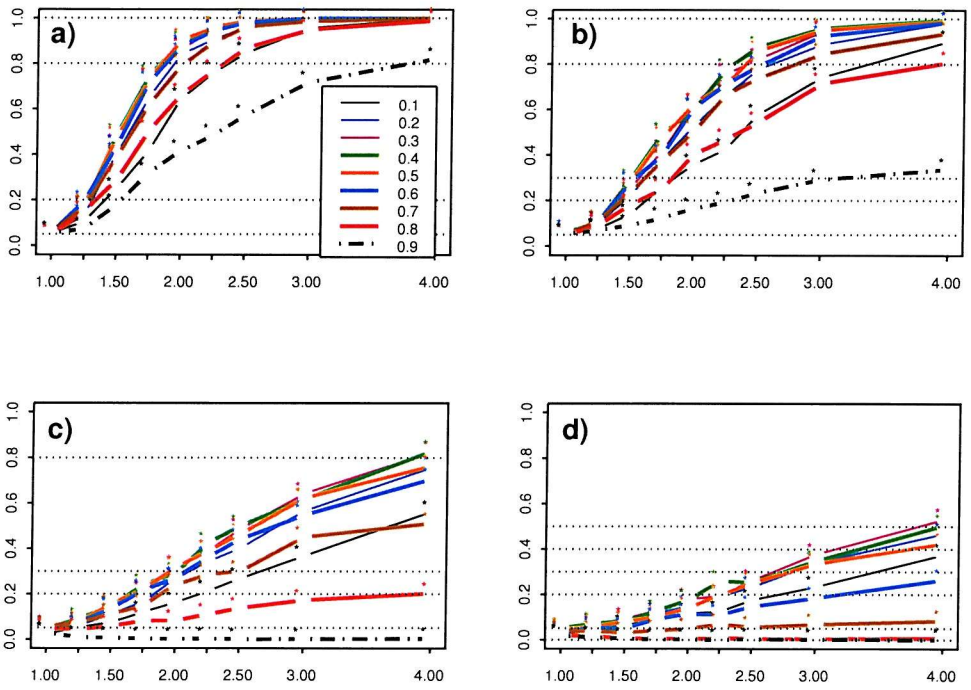

Figure 4.16: Simulated power of the saturated model for different sample sizes: a) $n_\emptyset = 1000$, b) $n_\emptyset = 500$, c) $n_\emptyset = 200$, d) $n_\emptyset = 100$. In all plots $\pi_2(0) = 0.1$, $\pi_1(0)$ from 0.1 to 0.9 and ψ_{12} from 1 to 4.

The effect of reducing the total sample size on the probability of selecting the saturated model seems more drastic than in the continuous case. Figure 4.16 shows the simulated power of the saturated model for different sample sizes: 1 000, 500, 200 and 100, for an unbalanced combination of marginal probabilities: $\pi_1(0)$ from 0.1 to 0.9 in each plot and $\pi_2(0) = 0.1$ in all four plots. It is possible to conclude that power decreases with decreasing sample size. With $n_\emptyset = 1\,000$ (plot in panel a) and

$\pi_2(0) = 0.1$ a power of 0.8 can be obtained with $\psi_{12} \simeq 2.25$ for all $\pi_1(0)$ between 0.1 and 0.8, whereas with $n_\emptyset = 500$ (plot in panel b) that power value is only achieved if $\pi_1(0) \simeq 0.4$. If $n_\emptyset = 200$ (plot in panel c) an odds ratio of 4 is required, and if $n_\emptyset = 100$ (plot in panel d), even with $\psi_{12} = 4$, the maximum value for power is $\simeq 0.5$.

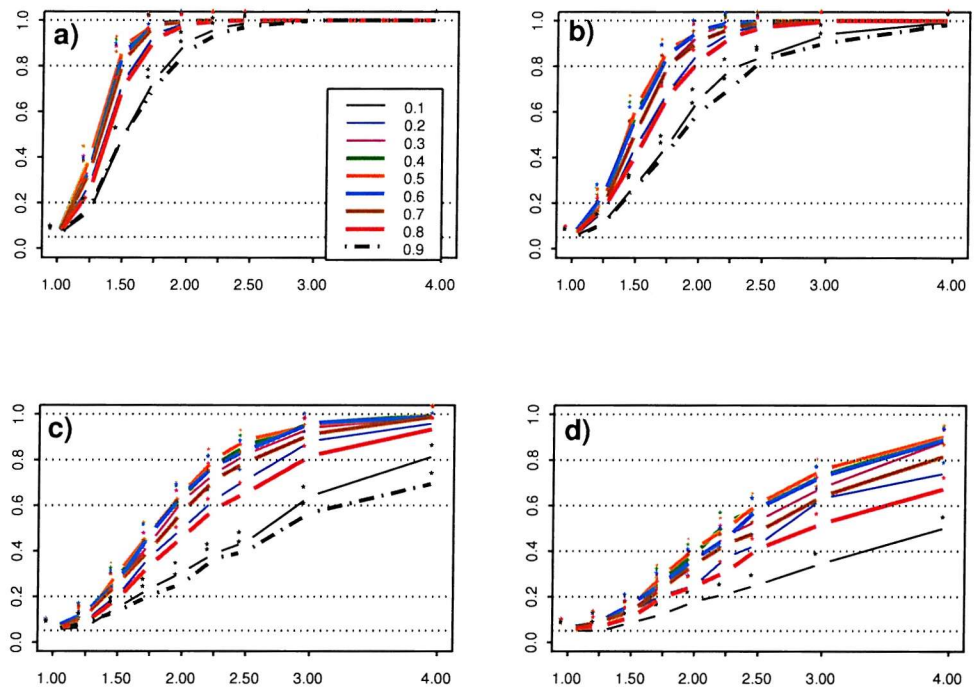

Figure 4.17: Simulated power of the saturated model for different sample sizes: a) $n_\emptyset = 1000$, b) $n_\emptyset = 500$, c) $n_\emptyset = 200$, d) $n_\emptyset = 100$. In all plots $\pi_2(0) = 0.4$, $\pi_1(0)$ from 0.1 to 0.9 and ψ_{12} from 1 to 4.

Figure 4.17 shows the effect of reducing the sample size if $\pi_2(0) = 0.4$. By comparison with Figure 4.16 it is possible to conclude that, when the sample size decreases, the effect on the power of selecting the saturated model is not so severe if marginal probabilities are more balanced. Thus, when varying the sample size, power is influenced not only by the value of the odds ratio, but also by how balanced the marginal probabilities are.

In brief: in the two binary variables case, power increases as the sample size increases and as the value of the odds ratio gets further from one, faster for more balanced marginal probabilities combinations.

Theoretical power values

Theoretical asymptotic power functions are now derived. As mentioned in Section 3.5.1, if $\psi_{12} \neq 1$, i.e., if the saturated model with two binary variables holds, the distribution of the likelihood ratio test statistic for single edge exclusion from the saturated model can be asymptotically approximated by a normal distribution, with mean $AE[LRT_{12}]$ (Equation 3.27) and variance $\text{var}(LRT_{12})$ (Equation 3.29). Therefore, asymptotically, the power of selecting the saturated GLL model with two binary variables, using the LRT, can be calculated using a normal approximation as

$$P\left[LRT_{12} > 3.8414 \mid \pi_1(0),\, \pi_2(0),\, \psi_{12}\right] \stackrel{a}{=} P\left[z > \frac{3.8414 - AE[LRT_{12}]}{\sqrt{\text{var}(LRT_{12})}}\right], \qquad (4.11)$$

where $z \sim N(0,\, 1)$.

A non-central chi-square approximation to the power of the LRT statistic can also be used, with non-centrality parameter φ given by Equation 3.43. The null hypothesis of independence is rejected if $P\left[LRT_{12} > \chi^2_{1;\,0.95} \mid \pi_1(0),\, \pi_2(0),\, \psi_{12}\right]$. Hence, the theoretical power functions can be obtained by calculating one minus the cumulative probability for a non-central chi-square distribution with one degree of freedom, and non-centrality parameter φ, for a quantile value of 3.8414 (and given values of $\pi_1(0)$, $\pi_2(0)$ and ψ_{12}).

The quality of the two approximations to the power of the LRT statistic can now be assessed. Figure 4.18 compares the simulated power values (in red) with the theoretical values calculated using the asymptotic normal and the non-central chi-square approximations presented above, for different combinations of marginal probabilities and odds ratio values. The asymptotic normal approximation corresponds to the blue curve and the non-central chi-square approximation to the green curve. A sample size of 1 000 is used. In each plot the odds ratio is represented on the horizontal axis and varies from 1 to 4. The marginal probability $\pi_1(0)$ takes the values 0.1, 0.3, 0.5, 0.7 and 0.9, respectively in plots 1 to 5. The marginal probability $\pi_2(0)$ takes the values 0.1, 0.2 and 0.3, respectively in plots in panels a), b) and c).

Some comments on Figure 4.18. Even for a sample size of 1 000, the normal approximation seems a poor approximation for values of the odds ratio close to one. In such cases, at an alternative close to the null, i.e., ψ_{12} close to one, the non-central

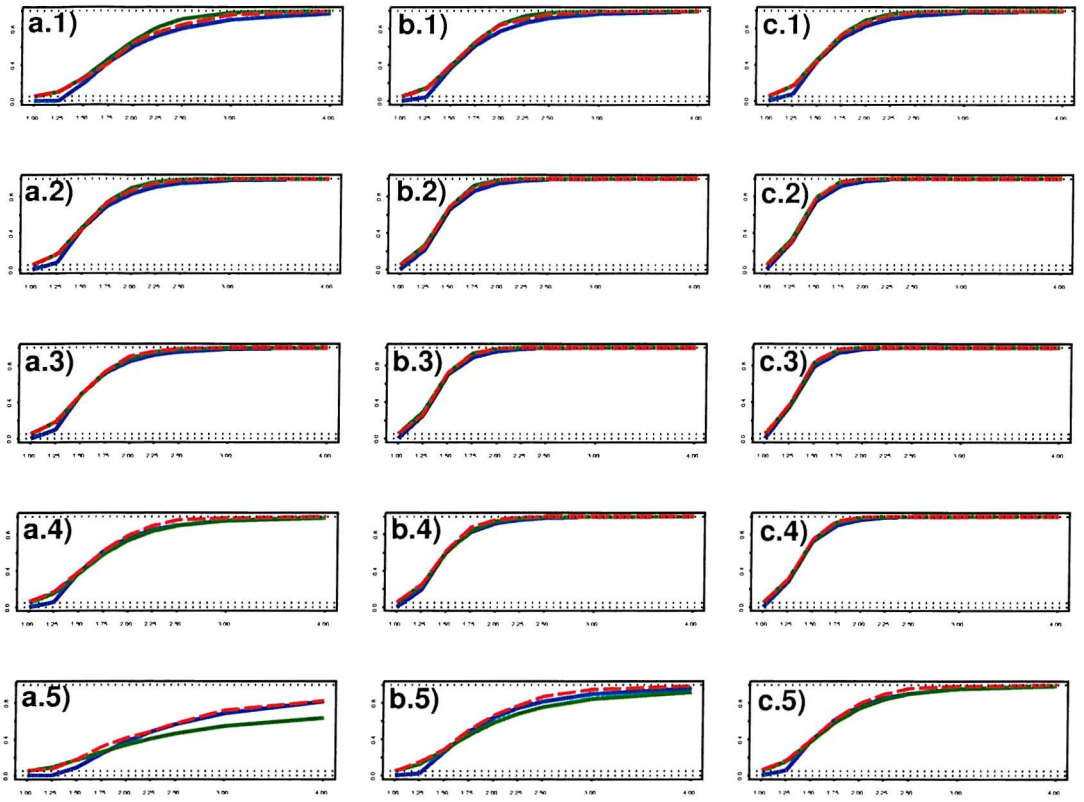

Figure 4.18: Simulated (in red) and theoretical power values using an asymptotic normal (in blue) and a non-central chi-square (in green) approximation. $n_\emptyset = 1\,000$. ψ_{12} from 1 to 4 in each plot. $\pi_1(0)$ equals: 1) 0.1, 2) 0.3, 3) 0.5, 4) 0.7 and 5) 0.9. $\pi_2(0)$ equals: a) 0.1, b) 0.2 and c) 0.3.

chi-square approximation performs much better. Note that if $\psi_{12} = 1$, $AE[LRT_{12}] = 0$ and $\text{var}(LRT_{12}) = 0$. The z value is infinity. Using an odds ratio value close to one ($\psi_{12} = 1.0001$) induces a very small value for the variance of LRT and, therefore, z becomes very big, the corresponding theoretical probability is zero and the normal approximation is poor (a value of 0.05 was expected). However, when the odds ratio value is far from one and the contingency table is unbalanced, i.e., the values of the balance index are high (Table 3.3), the normal approximation seems to perform better. That is the case of plots b.5) and especially a.5). Note that, in these cases, the minimum expected cell counts (appendix Table B.1) can be very small: in plot a.5) values of 3.8 and 2.9 are reached, respectively for ψ_{12} equal to 3 and 4. The chi-square approximation is very poor then.

For a sample size of 10 000 (considered a large sample size for a GLL model with two binary variables) the normal approximation is still a poor approximation to the

power of the LRT statistic for odds ratio values close to one. Indeed, for a large sample size, the non-central chi-square approximation to the power of the non-signed version of the LRT statistic (saturated GLL model) performs better than the normal approximation. These conclusions were drawn from Figure 4.19. Note these conclusions are in agreement with those drawn in Section 4.1.1, for GG models with two variables.

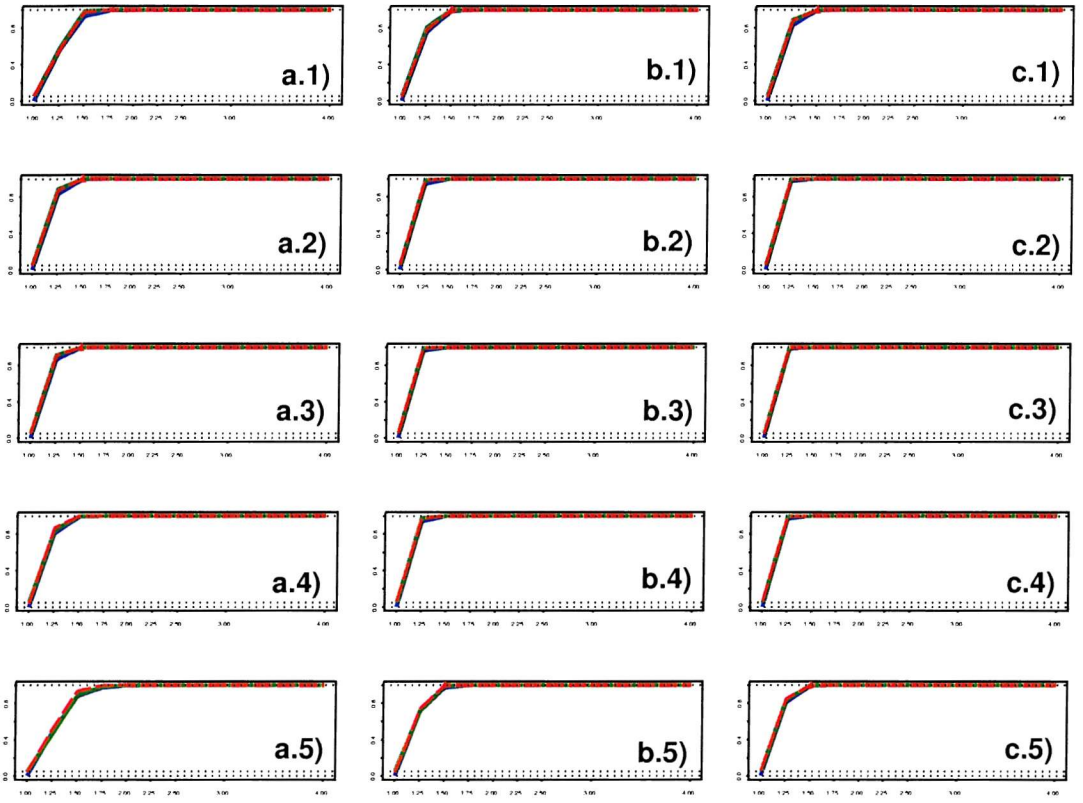

Figure 4.19: Simulated (in red) and theoretical power values using an asymptotic normal (in blue) and a non-central chi-square (in green) approximation. $n_{\emptyset} = 10\,000$. ψ_{12} from 1 to 4 in each plot. $\pi_1(0)$ equals: 1) 0.1, 2) 0.3, 3) 0.5, 4) 0.7 and 5) 0.9. $\pi_2(0)$ equals: a) 0.1, b) 0.2 and c) 0.3.

4.2.2 Normal approximations to the power of the Wald and the score test statistics in the two binary variables case

The power of selecting the saturated GLL model with two binary variables, using the Wald and the score test statistics was estimated by simulation, as in Section 4.2.1 when the LRT statistic was used. Very similar results were obtained and for that reason the

corresponding plots are omitted.

As for the LRT, theoretical power values can be obtained using a normal approximation. The power of selecting the saturated GLL model with two binary variables, using the Wald test is given by

$$P\left[Wald_{12} > 3.8414 \mid \pi_1(0),\, \pi_2(0),\, \psi_{12}\right] \stackrel{a}{=} P\left[z > \frac{3.8414 - AE[Wald_{12}]}{\sqrt{\text{var}(Wald_{12})}}\right], \quad (4.12)$$

where $z \sim N(0,\, 1)$, and $AE[Wald_{12}]$ and $\text{var}(Wald_{12})$ are given by Equations 3.33 and 3.35, respectively.

The power of selecting the saturated GLL model using the score test, is given by

$$P\left[Score_{12} > 3.8414 \mid \pi_1(0),\, \pi_2(0),\, \psi_{12}\right] \stackrel{a}{=} P\left[z > \frac{3.8414 - AE[Score_{12}]}{\sqrt{\text{var}(Score_{12})}}\right], \quad (4.13)$$

where $z \sim N(0,\, 1)$, $AE[Score_{12}]$ is given by Equation 3.38 and the methodology to obtain $\text{var}(Score_{12})$ is explained in Section 3.5.2.

The conclusions derived in Section 4.2.1, regarding the normal approximation to the power of the LRT statistic, also apply to the power of the Wald and score test statistics. In brief: the asymptotic normal approximation to the power of the test statistics, in GLL models with two binary variables, is a good approximation for large sample sizes and odds ratio values not close to one.

4.2.3 Power of the signed square-root versions of the test statistics, in GLL models with two binary variables

The signed square-root versions of the three test statistics for single edge exclusion from a saturated GLL model, with two binary variables, are presented in Section 3.4.3. In Section 3.5.3 normal approximations to the distributions of the signed square-root versions of the three test statistics are derived, under the alternative hypothesis that the saturated GLL model holds. The formulae for the means and variances of LRT_{12}^{sign}, $Wald_{12}^{sign}$ and $Score_{12}^{sign}$, in the asymptotic distribution, are also given in Section 3.5.3. All these can be used to derive theoretical power functions.

In the case of a two-sided hypothesis test, the null hypothesis that $\log\psi_{12} = 0$ is rejected if the absolute value of the signed square-root version of the test statistic being used is greater than 1.96, for the different values of $\log\psi_{12}$, $\pi_1(0)$ and $\pi_2(0)$. In

other words, asymptotically, the power of selecting the saturated GLL model with two binary variables, using the two-sided signed square-root version of each of the three test statistics for single edge exclusion, can be obtained as

$$P[|Test_{12}^{signL}| > 1.96 \mid \pi_1(0), \pi_2(0), \log\psi_{12}] \stackrel{a}{=} P\left[z < \frac{-1.96 - AE[Test_{12}^{sign}]}{\sqrt{\text{var}(Test_{12}^{sign})}}\right] + P\left[z > \frac{1.96 - AE[Test_{12}^{sign}]}{\sqrt{\text{var}(Test_{12}^{sign})}}\right],$$

where $z \sim N(0, 1)$ and $Test_{12}^{sign}$ can equal LRT_{12}^{sign}, $Wald_{12}^{sign}$ or $Score_{12}^{sign}$, with corresponding formulae for means and variances, in the asymptotic distribution, given in Section 3.5.3.

In the case of a one-sided hypothesis test, the null hypothesis that $\log\psi_{12} = 0$ is rejected if the value of the signed square-root version of the test statistic being used is greater than 1.645, for the different values of $\log\psi_{12}$, $\pi_1(0)$ and $\pi_2(0)$. Consequently, the power of selecting the saturated GLL model with two binary variables, using the one-sided signed square-root version of each of the three test statistics for single edge exclusion, can be obtained as

$$P[Test_{12}^{signL} > 1.645 \mid \pi_1(0), \pi_2(0), \log\psi_{12}] \stackrel{a}{=} P\left[z < \frac{1.645 - AE[Test_{12}^{sign}]}{\sqrt{\text{var}(Test_{12}^{sign})}}\right].$$

The quality of the asymptotic normal approximation is now assessed. Since only odds ratio values greater or equal to one have been used so far, only one-sided hypothesis tests are considered. The three test statistics for single edge exclusion were used. Because the pattern of their power functions is very similar only the results associated with the power of the one-sided signed square-root likelihood ratio test statistic are presented. Figure 4.20 compares simulated power values with theoretical power values obtained with the proposed approximation, for a sample size of 1 000. Four different unbalanced combinations of marginal probabilities were chosen: $\pi_2(0) = 0.1$ in all plots and $\pi_1(0)$ equals 0.1 in panel a), 0.2 in b), 0.8 in c) and 0.9 in d). The choice is based on the fact that previous results (Section 4.2.1) indicate approximations tend to perform worse for unbalanced combinations of marginal probabilities rather than for more balanced combinations.

In conclusion: the asymptotic normal approximation is a good approximation to the power of the one-sided signed square-root version of the test statistics for single edge

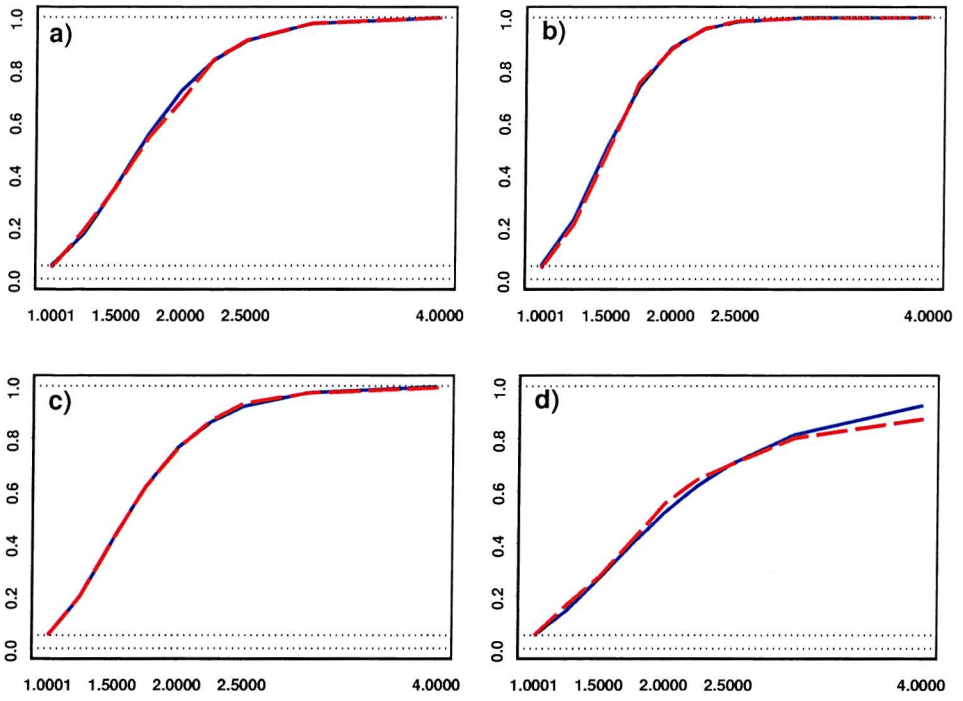

Figure 4.20: Simulated (in red) and theoretical asymptotic normal (in blue) power curves using the one-sided signed square-root likelihood ratio test statistic. $n_{\emptyset} = 1\,000$. ψ_{12} from 1 to 4 in each plot. $\pi_2(0)$ equals 0.1. $\pi_1(0)$ equals: a) 0.1, b) 0.2, c) 0.8, and d) 0.9.

exclusion, from a saturated GLL model with two binary variables, even for moderate sample sizes ($n_{\emptyset} = 1\,000$), less balanced combinations of marginal probabilities and odds ratio values close to one. These results are in agreement with those obtained in Section 4.1.1, for GG models with two variables.

4.2.4 Power of the test statistics for single edge exclusion in GLL models with three binary variables

In a $2 \times 2 \times 2$ contingency table there are eight cell probabilities that add up to one. Consequently, the parameter space is seven dimensional. For this reason a comprehensive investigation of the power functions for the different combinations of values of the seven parameters is not possible. The aim of the current section is, therefore, to provide some insight on how to obtain approximating theoretical power values for a specific contingency table.

Let θ be the 7×1 vector of the chosen parameters, either seven cell probabilities or

combinations of conditional odds ratios and marginal probabilities that uniquely define the contingency table under analysis, depending on the information available to the data analyst. Following a reasoning similar to that of Section 4.1.5, asymptotic normal approximations to the power of selecting the saturated GLL model with three binary variables, using the test statistics (LRT, Wald or score) for single edge exclusion from the saturated model, are now derived.

From Section 3.5.1, in the three binary variables case, under the alternative hypothesis that the saturated model holds, the vector of the three likelihood ratio test statistics LRT_{ij} is asymptotically normal distributed, with means given by Equation 3.30, variances given by Equation 3.31 and covariances given by Equation 3.32. If the Wald or the score test statistics are used instead, the vectors of test statistics $Wald_{ij}$ and $Score_{ij}$ are asymptotically normal distributed, with means given by Equations 3.36 and 3.39, respectively. Procedures for obtaining the variances and covariances are given in Section 3.5.2. Hence, the asymptotic power for the LRT of selecting the saturated GLL model with three binary variables can be obtained, using a trivariate normal approximation, as

$$P[LRT_{12} > 3.8414 \text{ and } LRT_{13} > 3.8414 \text{ and } LRT_{23} > 3.8414 \mid \theta] \stackrel{a}{=} \int_{3.8414}^{+\infty} \int_{3.8414}^{+\infty} \int_{3.8414}^{+\infty} \Phi(\mu, \Sigma)\, dLRT_{12}\, dLRT_{13}\, dLRT_{23}, \tag{4.14}$$

where $\Phi(\mu, \Sigma)$ is the joint cumulative distribution of the joint trivariate normal density of the three LRT statistics with vector of means μ and positive definite variance matrix Σ, given by Equations 3.30, 3.31 and 3.32. Similar reasoning applies to the Wald and to the score test statistics.

In practical terms, in order to obtain the asymptotic normal approximation to the power functions, the data analyst has to choose the test statistic for single edge exclusion to be used. Then, for the set of cell probabilities of interest, the means, variances and covariances of the test statistics have to be calculated (using formulae from Section 3.5.2). Finally, the *pmvnorm* function of R can be used to compute the corresponding triple integral defined by Equation 4.14. From the results obtained in the two variables case, it is expected that the normal approximation performs well

for large sample sizes, more balanced contingency tables with no small values for the minimum expected cell counts and conditional odds ratios not close to one.

In conclusion: it is proposed that the theoretical power of selecting the saturated GG or GLL model, using the test statistics for single edge exclusion from the saturated model, is obtained with asymptotic normal and non-central chi-square approximations to the distributions of the test statistics. The non-central chi-square approximation can be used in the two variables case and is a good approximation, both in GG and in GLL models, particularly at an alternative close to the null (ρ_{12} close to zero or ψ_{12} close to one). In the single edge case, if the signed square-root versions of the test statistics are used, the normal approximation is a good approximation to the power of the test statistics, even if the sample size is not very large and the correlation coefficient or the log odds ratio, are close to zero. However, if the non-signed version of the test statistics is used, the asymptotic normal approximation requires larger sample sizes, and remains a poor approximation for correlation coefficients close to zero (in GG models) and odds ratio values close to one (in GLL models with two binary variables). As highlighted in the conclusion of Chapter 3, one thousand observations can be considered a large sample size in a GG model, whereas in a GLL model, for a sample to be considered large, many more observations are required.

Chapter 5

Single-Factor GG Model and Latent Class GLL Model

As mentioned in Chapter 1, several references have been made in the literature to the use and importance of incorporating latent variables in the *graphical models* framework. Recent contributions in this area include Stanghellini (1997), Vicard (2000) and Giudici and Stanghellini (2001), all being related to the identification of factor analysis models with correlated residuals. Stanghellini (1997) presented the independence graph of the residuals and derived a sufficient condition for global identification of such a model. Vicard (2000) used the complementary graph of the independence graph of the residuals to derive a necessary and sufficient condition for identification. The single-factor model with correlated residuals is represented by a chain graph, following Cox and Wermuth (1996, Section 8.3). Giudici and Stanghellini (2001) defined a *graphical factor analysis model* as a factor model with correlated residuals and gave a sufficient condition for the identification of a factor model with an arbitrary number of factors, somehow generalising Stanghellini (1997). A Bayesian approach was adopted to tackle the issue of model comparison. The graphical factor analysis model is represented as a chain graph, but dashed arrows are used between boxes, thus representing the marginal associations between pairs of variables.

Edwards (1995, Section 7.1.2 and 2000, Section 4.6.2) used conditional independence graphs to represented both a single-factor analysis model and a latent class model and claimed both models can be fitted using MIM. Yet, it is recommended in this thesis that the data analyst is very cautious when using the software since not only

does MIM have no built in check for identification, but it also manages to give parameter estimates for models with 'negative degrees of freedom'. Besides, the relationship between the estimates produced by MIM and those produced by other packages that fit classical single-factor models, such as SPSS or LISREL, is often not clear. A deeper understanding of these models is, therefore, required.

The focus of the current chapter is neither on identification nor on estimation. Apparently taking a step back, the emphasis is on *model parameterisation*, both of the single-factor model as a graphical Gaussian model and of the latent class model as a graphical log-linear model. Models are represented by conditional independence graphs, with associations between each of the manifest variables and the latent variable measured either by partial correlation coefficients or by conditional log odds ratios.

Although traditionally factor analysis models and latent class models have been developed and treated quite separately, even as far as users and software packages for model fitting are concerned, some recent efforts have been made to create a theoretical unified framework and to work within it. These include Bartholomew and Knott (1999), Bartholomew, Steele, Moustaki and Galbraith (2002) and work by Vermunt leading to Latent GOLD software. The fact that, in this thesis, both models are treated in the same chapter corresponds to the belief that graphical models provide a unified framework for including both Gaussian and categorical latent variables. Furthermore, although *mixed models* are out of the scope of this thesis, once conditional-Gaussian distributions are considered, models parallel to latent trait and latent profile analysis can also be included in the unified framework. It is certainly a topic for further research.

The structure of this chapter is as follows. The classical factor analysis model and the classical latent class model are reviewed in Sections 5.1 and 5.4. Section 5.2 parameterises the single-factor GG model using partial correlations, relates the classical to the proposed parameterisation and investigates the admissible regions of the parameter space. Section 5.3 focus on detecting a model consistent with a single-factor GG model and gives some recommendations to the data analyst. Finally, Section 5.5 parameterises the latent class GLL model and investigates the conditional independence structure of the manifest variables arising from a latent class GLL model.

5.1 The Classical Factor Analysis Model

This section reviews the classical parameterisation of the factor analysis model, particular attention being devoted to the single-factor model. The question of model identification is addressed and a brief reference is made to model estimation.

5.1.1 The classical parameterisation of the single-factor model

The factor analysis model can be written as

$$X = \Lambda\xi + \delta, \tag{5.1}$$

where X is a vector of p manifest variables, ξ is a vector of m underlying factors (latent variables) such that $m < p$, Λ is a $p \times m$ matrix of factor loadings relating the manifest variables to the underlying factors, and δ is a vector of p variables representing random measurement error and indicator specificity. Variables are considered to be measured as deviations from their means, that is $E[X] = E[\xi] = 0$. The model assumes that $E[\xi\delta^T] = 0$, $E[\delta] = 0$, var $[\delta] = \Theta_\delta$ (diagonal) and that X, ξ and δ are multivariate normal distributed.

The variance matrix for X, with elements denoted by σ_{ij}, is

$$\Sigma_{XX} = \Lambda\Phi\Lambda^T + \Theta_\delta, \tag{5.2}$$

where Φ is the $m \times m$ variance matrix of ξ and Θ_δ is the $p \times p$ diagonal variance matrix of δ. In the classical factor analysis model the matrix Θ_δ is diagonal, i.e., all associations among the manifest variables are explained by the unobserved factors, whereas in the confirmatory factor analysis framework residual terms are allowed to be correlated. If Φ is the identity matrix the m factors are orthogonal and the solution will be unique up to rotation, i.e., up to post-multiplication by an orthogonal matrix.

One of the crucial issues associated with factor models is that of *identification.* An unknown parameter is identified if it can be written as a function of one or more elements of the variance matrix Σ_{XX} (leading to a unique solution). A model is identified if all parameters to be estimated from the data are identified. Restrictions have, therefore, to be imposed on the model defined by Equations 5.1 and 5.2 to ensure that the model is identified, i.e., that the elements of Θ_δ and of Λ can be uniquely (up to

rotation) expressed as a function of the elements in Σ_{XX}. A usual first step to avoid a basic problem of identification is to fix the scale of the latent variables, i.e., to set the variances of the latent variables to one. Anderson and Rubin (1956, Section 5) dealt with the problem of identification and gave a necessary and sufficient condition for the identification of a single-factor model in Theorem 5.5: at least three factor loadings have to be non-zero. However, no necessary and sufficient condition exists to ensure the identification of models with two or more latent variables, when specific sets of factor loadings are set to zero and others have to be estimated from the data (confirmatory factor analysis framework). Some rules have been proposed in the literature to try to assess the identification of a confirmatory factor analysis model - for a summary see, for example, Bollen (1989, pages 238-251). Since only single-factor models are considered in this thesis, no problems of identification exist provided there are three or more manifest variables in each model and the latent variable is scaled to have unit variance.

The single-factor model

A particular case of the model defined by Equation 5.1 is the classical *single-factor model*, with p manifest variables and one factor (latent variable) ξ. In this case, Equations 5.1 and 5.2 simplify to

$$X = \lambda\xi + \delta \tag{5.3}$$

$$\Sigma_{XX} = \lambda\lambda^T + \Theta_\delta. \tag{5.4}$$

The latent variable ξ has been scaled to have unit variance and λ is a $p \times 1$ vector of factor loadings. Each λ_i represents the magnitude of the expected change in the manifest variable X_i for a unit change in the latent variable ξ. As mentioned earlier, the model is identified provided p, the number of manifest variables, is at least three. The aim is to express λ and the elements of the diagonal matrix Θ_δ in terms of the observable Σ_{XX}, as a solution to 5.4. In order to guarantee the existence of a solution that is real (non-complex) and yields non-negative variances, Σ_{XX} has to meet some requirements.

Anderson and Rubin (1956, Theorem 4.2) stated that '*a necessary and sufficient condition that Σ_{XX} be a variance matrix of a factor analysis model with one factor is*

that $p(p-1)/2-p$ *independent tetrad conditions are satisfied and*

$$0 \leq \frac{\sigma_{ki}\sigma_{ij}}{\sigma_{kj}} \leq \sigma_{ii}$$

for one pair $(j \neq k)$ *for each* i.'

There are $p(p+1)/2$ distinct elements in the observable Σ_{XX} and p elements in the diagonal of Θ_δ and p elements in the vector λ (forming a total of $2p$ unknown parameters). Therefore, the system has $p(p+1)/2$ equations and $2p$ unknown parameters. Let $C = p(p+1)/2 - 2p = \ldots = p(p-1)/2 - p$ denote the number of equations minus the number of unknowns to be determined. If $C > 0$, Σ_{XX} must satisfy some $C = p(p-1)/2-p$ conditions for a solution to exist. The $p(p-1)/2-p$ *tetrad conditions* are given by

$$\frac{\sigma_{12}\sigma_{13}}{\sigma_{23}} = \frac{\sigma_{12}\sigma_{14}}{\sigma_{24}} = \ldots = \frac{\sigma_{12}\sigma_{1p}}{\sigma_{2p}} = \frac{\sigma_{13}\sigma_{14}}{\sigma_{34}} = \ldots = \frac{\sigma_{13}\sigma_{1p}}{\sigma_{3p}} = \ldots = \frac{\sigma_{1(p-1)}\sigma_{1p}}{\sigma_{(p-1)p}}.$$

In the case $p = 3$, $p(p-1)/2 - p = 0$ and so the solution to the system of six equations (with six unknowns) is possible and determined, without any additional constraints having to be satisfied. If $p = 4$ the system has ten equations and eight unknowns, $p(p-1)/2-p = 2$ and two tetrad conditions have to be satisfied. They are given by

$$\frac{\sigma_{12}\sigma_{13}}{\sigma_{23}} = \frac{\sigma_{12}\sigma_{14}}{\sigma_{24}} = \frac{\sigma_{13}\sigma_{14}}{\sigma_{34}} \Leftrightarrow \sigma_{12}\sigma_{13}\sigma_{24}\sigma_{34} = \sigma_{12}\sigma_{14}\sigma_{23}\sigma_{34} = \sigma_{13}\sigma_{14}\sigma_{23}\sigma_{24} \Leftrightarrow$$

$$\left\{ \begin{array}{l} \sigma_{13}\sigma_{24} = \sigma_{14}\sigma_{23} \\ \sigma_{12}\sigma_{34} = \sigma_{14}\sigma_{23} \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \sigma_{12}\sigma_{34} - \sigma_{14}\sigma_{23} = 0 \\ \sigma_{13}\sigma_{24} - \sigma_{14}\sigma_{23} = 0 \end{array} \right\}.$$

Similar reasoning can be followed for $p \geq 5$.

Anderson and Rubin (1956, Theorem 4.1) gave a necessary and sufficient condition for Σ_{XX} to be a variance matrix of a single-factor model, i.e., '*there has to exist a diagonal matrix* Θ_δ^* *(with non-negative elements) such that* $\Sigma_{XX} - \Theta_\delta^*$ *is positive semidefinite of rank one.*' In order to prove Anderson and Rubin's Theorem 4.2, based on Theorem 4.1, it is required to prove that $\Sigma_{XX} - \Theta_\delta^*$ being positive semidefinite of rank one is equivalent to satisfying $p(p-1)/2 - p$ independent tetrad conditions and having $0 \leq \frac{\sigma_{ki}\sigma_{ij}}{\sigma_{kj}} \leq \sigma_{ii}$, for one pair $(j \neq k)$ for each i. One has to be able to subtract non-negative elements from the diagonal of Σ_{XX} to get a positive semidefinite matrix of rank one.

It is known that $\Sigma_{XX} - \Theta_\delta$ is of rank one if and only if Θ_δ can be chosen so that all *second-order minors* of $\Sigma_{XX} - \Theta_\delta$ are zero. Consequently:

- if $\Sigma_{XX} - \Theta_\delta$ is of rank one, then all its second-order minors are zero. A second-order minor which does not include a diagonal element, known as a *tetrad*, is of the form

$$\begin{vmatrix} \sigma_{hi} & \sigma_{hj} \\ \sigma_{ki} & \sigma_{kj} \end{vmatrix} = \sigma_{hi}\sigma_{kj} - \sigma_{hj}\sigma_{ki} \ (h, i, k, j \text{ different}).$$

  Setting all to zero implies that $\sigma_{hi}\sigma_{kj} - \sigma_{hj}\sigma_{ki} = 0$, for all possible combinations of different h, i, j, k. A second-order minor, which includes one diagonal element, is given by

$$\begin{vmatrix} \sigma_{ii} - \theta_{\delta_{ii}} & \sigma_{ij} \\ \sigma_{ki} & \sigma_{kj} \end{vmatrix} = (\sigma_{ii} - \theta_{\delta_{ii}})\sigma_{kj} - \sigma_{ij}\sigma_{ki} \ (i, k, j \text{ different}).$$

  Setting this expression to zero implies that $\theta_{\delta_{ii}}$ has to be chosen as $\theta_{\delta_{ii}} = \sigma_{ii} - \frac{\sigma_{ki}\sigma_{ij}}{\sigma_{kj}}$. The conditions that the solution be consistent (i.e., independent of the pair j, k) are known as the *tetrad conditions*.

  In brief: all second-order minors being zero imply that the tetrad conditions are satisfied.

- If all second-order minors of $\Sigma_{XX} - \Theta_\delta$ are zero, then the matrix $\Sigma_{XX} - \Theta_\delta$ is of rank one. If the tetrad conditions are satisfied, all second-order minors of $\Sigma_{XX} - \Theta_\delta$ are zero, and consequently this matrix is of rank one.

For the matrix $\Sigma_{XX} - \Theta_\delta$ to be positive semidefinite its diagonal elements have to be non-negative, i.e., $\sigma_{ii} - \theta_{\delta_{ii}} \geq 0$. Since $\theta_{\delta_{ii}}$ has to equal $\sigma_{ii} - \frac{\sigma_{ki}\sigma_{ij}}{\sigma_{kj}}$,

$$\sigma_{ii} - \theta_{\delta_{ii}} \geq 0 \Leftrightarrow \sigma_{ii} - (\sigma_{ii} - \frac{\sigma_{ki}\sigma_{ij}}{\sigma_{kj}}) \geq 0 \Leftrightarrow \frac{\sigma_{ki}\sigma_{ij}}{\sigma_{kj}} \geq 0.$$

For the matrix Θ_δ to be positive semidefinite, all the diagonal elements have to be non-negative, and so

$$\theta_{\delta_{ii}} \geq 0 \Rightarrow \sigma_{ii} \geq \frac{\sigma_{ki}\sigma_{ij}}{\sigma_{kj}}.$$

As a result,

$$0 \leq \frac{\sigma_{ki}\sigma_{ij}}{\sigma_{kj}} \leq \sigma_{ii}.$$

Hence, Anderson and Rubin's Theorem 4.2 has just been proven.

5.1.2 The parameter space for a single-factor model with three or four manifest variables

In this section the parameter space for the single-factor model, with three or four manifest variables, is investigated. Permissible regions for the correlation coefficients are derived.

The single-factor model with three manifest variables

In the three manifest variables case Equation 5.4 can be written as

$$\Sigma_{XX} = \lambda\lambda^T + \Theta_\delta \Leftrightarrow \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ & \sigma_{22} & \sigma_{23} \\ & & \sigma_{33} \end{bmatrix} = \begin{bmatrix} \lambda_1^2 + \theta_{\delta_{11}} & \lambda_1\lambda_2 & \lambda_1\lambda_3 \\ & \lambda_2^2 + \theta_{\delta_{22}} & \lambda_2\lambda_3 \\ & & \lambda_3^2 + \theta_{\delta_{33}} \end{bmatrix}.$$

Solving for λ and Θ_δ is equivalent to solving a system with six equations, and 6 unknown parameters $\lambda_1, \lambda_2, \lambda_3, \theta_{\delta_{11}}, \theta_{\delta_{22}}, \theta_{\delta_{33}}$, (indeed the model is identified and saturated) as follows

$$\left\{ \begin{array}{c} \lambda_1^2 + \theta_{\delta_{11}} = \sigma_{11};\quad \lambda_2^2 + \theta_{\delta_{22}} = \sigma_{22};\quad \lambda_3^2 + \theta_{\delta_{33}} = \sigma_{33}; \\ \lambda_1\lambda_2 = \sigma_{12};\quad \lambda_1\lambda_3 = \sigma_{13};\quad \lambda_2\lambda_3 = \sigma_{23} \end{array} \right\}.$$

Solving these equations with respect to the unknown parameters it is possible to obtain

$$\left\{ \begin{array}{c} \lambda_1^2 = \frac{\sigma_{12}\sigma_{13}}{\sigma_{23}};\quad \lambda_2^2 = \frac{\sigma_{12}\sigma_{23}}{\sigma_{13}};\quad \lambda_3^2 = \frac{\sigma_{13}\sigma_{23}}{\sigma_{12}}; \\ \theta_{\delta_{11}} = \sigma_{11} - \lambda_1^2 = \sigma_{11} - \frac{\sigma_{12}\sigma_{13}}{\sigma_{23}};\quad \theta_{\delta_{22}} = \ldots = \sigma_{22} - \frac{\sigma_{12}\sigma_{23}}{\sigma_{13}};\quad \theta_{\delta_{33}} = \ldots = \sigma_{33} - \frac{\sigma_{13}\sigma_{23}}{\sigma_{12}} \end{array} \right\}.$$

Notice that because $\theta_{\delta_{ii}}$ has to be non-negative, $\frac{\sigma_{ij}\sigma_{ik}}{\sigma_{jk}} \leq \sigma_{ii}$ and because λ_i^2 is non-negative, $\frac{\sigma_{ij}\sigma_{ik}}{\sigma_{jk}} \geq 0$. Indeed, this is precisely the condition required by Anderson and Rubin (1956, Theorem 4.2) previously shown. Because there are three manifest variables $p(p-1)/2 - p = 0$, and no tetrad conditions need to be satisfied.

When, instead of the variance matrix, the population correlation matrix, denoted by P with elements ρ_{ij}, is used, the corresponding results hold:

- $\rho_{ij} = \lambda_i\lambda_j$ and $\lambda_i^2 = \frac{\rho_{ij}\rho_{ik}}{\rho_{jk}}$;
- $\theta_{\delta_{ii}} = 1 - \lambda_i^2 = 1 - \frac{\rho_{ij}\rho_{ik}}{\rho_{jk}}$;
- $\theta_{\delta_{ii}} \geq 0 \Rightarrow \lambda_i^2 \leq 1$ and $\frac{\rho_{ij}\rho_{ik}}{\rho_{jk}} \leq 1$;

- $\lambda_i^2 \geq 0 \Rightarrow \frac{\rho_{ij}\rho_{ik}}{\rho_{jk}} \geq 0.$

Consequently $0 \leq \frac{\rho_{ij}\rho_{ik}}{\rho_{jk}} \leq 1$, which is the condition imposed by Anderson and Rubin in Theorem 4.2, expressed in terms of correlations. However, for this expression to hold for all manifest variables, all three ρ have to be different from zero, so that the three λ are finite. Therefore, the product of the three correlation coefficients has to be positive and the correlation matrix P is taken as positive definite. This is what Dijkstra (1992) stated as Proposition 1. Since one of the aims of this thesis is to relate graphical Gaussian models to factor models, and GG models literature assumes *positive definite matrices*, it was decided to consider only positive definite (and not positive semidefinite) variance and correlation matrices.

In conclusion, for a 3×3 positive definite correlation matrix to be suitable for a single-factor analysis model it is required that

$$0 < \frac{\rho_{12}\rho_{13}}{\rho_{23}} \leq 1; \quad 0 < \frac{\rho_{12}\rho_{23}}{\rho_{13}} \leq 1; \quad 0 < \frac{\rho_{13}\rho_{23}}{\rho_{12}} \leq 1.$$

Figure 5.1 represents the allowable values of the three correlations. The ellipse represents the positive definiteness constraint defined by Equation 4.4. An arbitrary value of 0.5 was chosen for one of the correlation coefficients, whereas the other two vary between -1 and 1, within the region of positive definiteness. The fact that the product of the three correlation coefficients has to be positive restricts the allowable values to quadrants one and three. Finally, the three conditions $\frac{\rho_{ij}\rho_{ik}}{\rho_{jk}} \leq 1$ determine the allowable shaded area.

The single-factor model with four manifest variables

In the four manifest variables case Equation 5.4 can be written as a system of ten equations, with ten known parameters and eight unknown parameters $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \theta_{\delta_{11}}, \theta_{\delta_{22}}, \theta_{\delta_{33}}, \theta_{\delta_{44}})$. That is, $\Sigma_{XX} = \lambda\lambda^T + \Theta_\delta$ is equivalent to

$$\left\{ \begin{array}{c} \lambda_1^2 + \theta_{\delta_{11}} = \sigma_{11}; \quad \lambda_2^2 + \theta_{\delta_{22}} = \sigma_{22}; \quad \lambda_3^2 + \theta_{\delta_{33}} = \sigma_{33}; \quad \lambda_4^2 + \theta_{\delta_{44}} = \sigma_{44}; \\ \lambda_1\lambda_2 = \sigma_{12}; \quad \lambda_1\lambda_3 = \sigma_{13}; \quad \lambda_1\lambda_4 = \sigma_{14}; \quad \lambda_2\lambda_3 = \sigma_{23}; \quad \lambda_2\lambda_4 = \sigma_{24}; \quad \lambda_3\lambda_4 = \sigma_{34} \end{array} \right\}.$$

Solving these ten equations with respect to, for example, λ_1, it is possible to obtain

$$\lambda_1^2 = \frac{\sigma_{12}\sigma_{13}}{\sigma_{23}} \text{ and } \lambda_1^2 = \frac{\sigma_{12}\sigma_{14}}{\sigma_{24}} \text{ and } \lambda_1^2 = \frac{\sigma_{13}\sigma_{14}}{\sigma_{34}}.$$

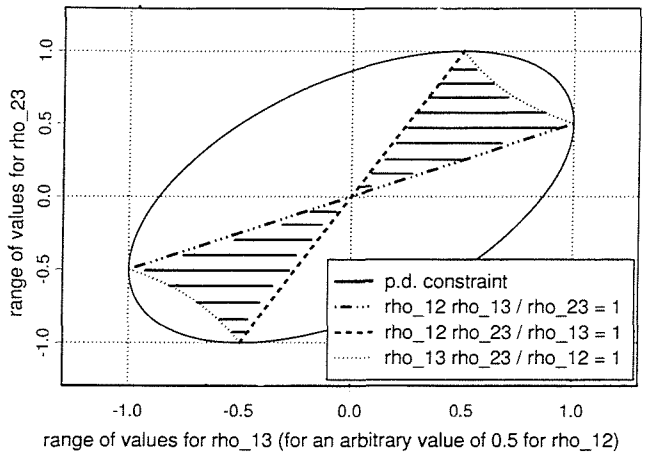

Figure 5.1: Possible values of three correlation coefficients to define a positive definite matrix suitable for a single-factor analysis model

In order to have consistent solutions for the parameter λ_1 (and similar reasoning can be used for all the other parameters) some equality constraints have to be imposed, namely that

$$\frac{\sigma_{12}\sigma_{13}}{\sigma_{23}} = \frac{\sigma_{12}\sigma_{14}}{\sigma_{24}} = \frac{\sigma_{13}\sigma_{14}}{\sigma_{34}}.$$

These simplify to

$$\sigma_{12}\sigma_{34} - \sigma_{14}\sigma_{23} = 0; \quad \sigma_{13}\sigma_{24} - \sigma_{14}\sigma_{23} = 0, \tag{5.5}$$

which are exactly the $p(p-1)/2 - p = 2$ tetrad conditions that have to be satisfied. For a more detailed example see Harman (1967, page 73).

It was shown that in the $p = 4$ variables case the two tetrad conditions imply $\sigma_{ij} = \lambda_i\lambda_j$ (or $\rho_{ij} = \lambda_i\lambda_j$ if correlations are used instead). Moreover, having $\sigma_{ij} = \lambda_i\lambda_j$, $(i \neq j)$ implies the tetrad conditions are satisfied, as mentioned by Cox and Wermuth (1996, page 192), since then

$$\sigma_{12}\sigma_{34} - \sigma_{14}\sigma_{23} = 0 \Leftrightarrow \lambda_1\lambda_2\lambda_3\lambda_4 - \lambda_1\lambda_4\lambda_2\lambda_3 = 0 \text{ and}$$
$$\sigma_{13}\sigma_{24} - \sigma_{14}\sigma_{23} = 0 \Leftrightarrow \lambda_1\lambda_3\lambda_2\lambda_4 - \lambda_1\lambda_4\lambda_2\lambda_3 = 0.$$

In brief, for a 4×4 positive definite correlation matrix to be suitable for a single-factor analysis model it is required that:

- $\rho_{ij} = \lambda_i\lambda_j$ (i and j distinct, from 1 to 4), so that $\frac{\rho_{ij}\rho_{ik}}{\rho_{jk}} = \frac{\lambda_i\lambda_j\lambda_i\lambda_k}{\lambda_j\lambda_k} = \lambda_i^2$;
- $0 < \frac{\rho_{ij}\rho_{ik}}{\rho_{jk}} \leq 1$ (for all possible combinations of distinct i, j and k, from 1 to 4).

These results will be extended in Section 5.2.1, once the single-factor model is parameterised in terms of partial correlation coefficients.

5.1.3 Some notes on the estimation of the factor model

Maximum likelihood is the method commonly used for estimating the factor analysis model. The log-likelihood for a sample of size n equals $-\frac{n}{2}\log(2\pi) - \frac{n}{2}\left[\log|\Sigma_{XX}| + tr(S\Sigma_{XX}^{-1})\right]$. Within the context of confirmatory factor analysis, instead of maximising the logarithm of the likelihood function it is convenient to minimise the fitting function (denoted as F_{ML}), also known as discrepancy function, (see Jöreskog (1967, page 457) for further details), defined as

$$F = \log|\Sigma_{XX}| + tr(S\Sigma_{XX}^{-1}) - \log|S| - p, \tag{5.6}$$

where S is the unbiased sample variance matrix (with divisor $n-1$). The EM algorithm is often the iterative technique used for the optimisation.

The first goodness-of-fit measure widely used was the probability associated with the chi-square likelihood ratio test, which is given by $n-1$ times the minimum value of the fitting function F_{ML} obtained for the specified model. Under the assumption of multivariate normal X, $(n-1)\,F_{ML}$ is asymptotically chi-square distributed. If the model is correct, and the sample size is sufficiently large, the χ^2 measure is the likelihood ratio test for testing the model against the alternative that Σ_{XX} is unconstrained. The associated degrees of freedom (df) for χ^2 are given by $df = \{\frac{1}{2}p(p+1)\} - t$, where t is the number of independent parameters to be estimated (equal to $2p$ in the classical single-factor model). The probability level for the resulting chi-square value is the probability of obtaining a larger chi-square, given that the model is correct. Hence, small chi-square values, with corresponding large probability levels, indicate good fit. This measure has been criticised in the literature, partially because it is sensitive to the sample size, and several other measures of goodness-of-fit have been suggested; see, for example, Anderson and Gerbing (1984) and Bollen (1989, Chapter 7).

In the three manifest variables case $df = 0$ and the solution is unique. In the presence of four or more manifest variables an iterative procedure is required, either to maximise the log-likelihood function or to minimise the discrepancy function. Some problems can, therefore, occur, namely *non-convergence* and *improper solutions*.

Solutions are considered *non-convergent* when a certain estimation algorithm, within a previously set number of iterations, is unable to satisfy a prescribed convergence criteria. Solutions are *improper* when the estimates have values that are impossible (or implausible) in the population (Bollen, 1987): for example one or more of the unique variances (elements of the diagonal matrix Θ_δ) are negative or correlations are greater than one.

Several studies have been carried out to understand the circumstances under which these problems occur, including van Driel (1978), Anderson and Gerbing (1984), Boomsma (1985) and Boomsma and Hoogland (2001). Key reasons for non-convergence and improper solutions are population parameters near the boundaries of proper solutions (small values of the error variances $\Theta_{\delta ii}$, or correlations close to one in the population), misspecification of the model, the existence of outliers and influential observations (Bollen, 1987), inconsistent variates and sampling fluctuations. Non-convergence tends to decrease with larger factor loadings, more indicators per factor and larger samples. For further details see Bartholomew and Knott (1999).

5.2 The Single-Factor Graphical Gaussian Model

Section 5.1 reviewed the classical parameterisation of the single-factor model. In the current section the single-factor model is parameterised as a GG model, and the relationship between the two parameterisations is studied in detail. Indeed, parameterising the single-factor model as a GG model allows a normal distributed latent variable to be included in the framework of graphical Gaussian models.

In the classical single-factor model $X = \lambda\xi + \delta$ the manifest variables are conditionally independent given the latent variable ξ (normally distributed with zero mean and unit variance). This conditional model can be interpreted in the regression framework, considering the manifest variables the response and the latent variable the explanatory variable. Consequently, given ξ, X is normal distributed with mean $\lambda\xi$ and variance matrix Θ_δ. In fact,

$$\mu_{X.\xi} = \mu_X + \Sigma_{X\xi}(\Sigma_{\xi\xi})^{-1}(\xi - \mu_\xi) = \mu_X + \lambda(\xi - \mu_\xi) = 0 + \lambda(\xi - 0) = \lambda\xi \qquad (5.7)$$

and $\Sigma_{XX.\xi} = var(\delta) = \Theta_\delta$. The vector of the p regression coefficients, λ, can be obtained

as

$$\lambda = \Sigma_{X\xi}(\Sigma_{\xi\xi})^{-1} = \Sigma_{X\xi} \tag{5.8}$$

and the diagonal matrix Θ_δ as

$$\Theta_\delta = \Sigma_{XX.\xi} = diag\{\Sigma_{XX} - \Sigma_{X\xi}\Sigma_{\xi\xi}^{-1}\Sigma_{\xi X}\} = diag\{\Sigma_{XX} - \Sigma_{X\xi}\Sigma_{\xi X}\}. \tag{5.9}$$

For simplicity of notation in the remainder of this chapter the latent variable will be denoted by L and the vector of manifest variables by M.

5.2.1 Parameterising the single-factor GG model using partial correlations

This section gives a detailed explanation on how to parameterise a single-factor model as a GG model, i.e., using partial correlations. The conditional independence structure between manifest variables arising from a single-factor GG model (by marginalising over the latent variable L) is investigated. The admissible region for the partial correlation coefficients between manifest variables, compatible with a single-factor model, is then derived.

In this thesis the single-factor GG model is represented by an independence graph, as justified in Section 2.4.2. Figure 5.2 displays the independence graph associated with a single-factor model with three manifest random variables (1, 2 and 3) and one latent variable (L), all assumed normal distributed.

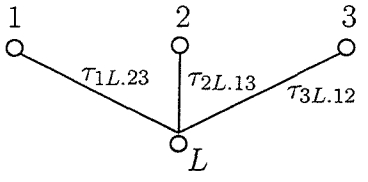

Figure 5.2: The independence graph of a single-factor model: the manifest variables 1, 2 and 3 are conditionally independent given L. $\tau_{iL.rest}$ represents the partial correlation between manifest variable i and latent variable L, given the remaining manifest variables.

The joint distribution of the manifest and latent variables is multivariate normal with positive definite variance matrix denoted by Σ, with inverse denoted by Ω. In the

case of three manifest variables and a single latent variable,

$$\Omega = \left[\begin{array}{ccc|c} \omega_{11} & 0 & 0 & \omega_{1L} \\ 0 & \omega_{22} & 0 & \omega_{2L} \\ 0 & 0 & \omega_{33} & \omega_{3L} \\ \hline \omega_{1L} & \omega_{2L} & \omega_{3L} & \omega_{LL} \end{array}\right],$$

where the diagonal elements are positive and $\omega_{iL} \neq 0$, $i \in M = \{1, 2, 3\}$. When Ω is scaled to have unit values on the main diagonal, which is denoted by $sc(\Omega)$, the matrix T is obtained. Whittaker (1990, page 144) showed that the off-diagonal elements of the scaled inverse variance matrix are the negative partial correlations of the corresponding elements, given the rest. Therefore, $\tau_{iL.jk}$, the negative of the non-zero off-diagonal element, represents the partial correlation between the manifest variable i and the latent variable L, given the remaining two manifest variables, and is obtained as $-\tau_{iL.jk} = \frac{\omega_{iL}}{\sqrt{\omega_{ii}\omega_{LL}}}$. Consequently,

$$T = sc(\Omega) = \left[\begin{array}{ccc|c} 1 & 0 & 0 & -\tau_{1L.23} \\ 0 & 1 & 0 & -\tau_{2L.13} \\ 0 & 0 & 1 & -\tau_{3L.12} \\ \hline -\tau_{1L.23} & -\tau_{2L.13} & -\tau_{3L.12} & 1 \end{array}\right] = \left[\begin{array}{cc} I & T_{ML} \\ T_{LM} & 1 \end{array}\right].$$

Let T^M denote the scaled inverse variance matrix of the manifest variables and U, with elements ν_{ij}, denote the inverse of T. Indeed U is a 'modified' Σ in the sense that there exists a symmetric square matrix G such that $GUG^T = \Sigma$. The matrix U can be partitioned as

$$U = \left[\begin{array}{ccc|c} \nu_{11} & \nu_{12} & \nu_{13} & \nu_{1L} \\ \nu_{12} & \nu_{22} & \nu_{23} & \nu_{2L} \\ \nu_{13} & \nu_{23} & \nu_{33} & \nu_{3L} \\ \hline \nu_{1L} & \nu_{2L} & \nu_{3L} & \nu_{LL} \end{array}\right] = \left[\begin{array}{cc} U_{MM} & U_{ML} \\ U_{LM} & \nu_{LL} \end{array}\right].$$

The aim now is to investigate the conditional independence structure between the three manifest variables arising from marginalising the joint distribution of all variables over the latent variable. It is proved that marginalising over the latent variable L in Figure 5.2 yields the saturated model for the three manifest variables, and therefore induces an independence graph that is complete. This is equivalent to stating that $1 \perp\!\!\!\perp 2|L$, $1 \perp\!\!\!\perp 3|L$ and $2 \perp\!\!\!\perp 3|L$ implies no zero entries in T^M, the scaled inverse variance matrix of the manifest variables.

A proof follows, in four steps. First T, the scaled inverse variance matrix of all the variances is inverted and U, a 'modified' variance matrix of all the variables is obtained. Then, by marginalising over the latent variable L, the matrix U_{MM}, a 'modified' variance matrix of the manifest variables, is obtained. Inverting and scaling this matrix induces T^M, the scaled inverse variance matrix of the manifest variables. Recall that the off-diagonal elements of this matrix are the negatives of the partial correlations between manifest variables, and all these are shown to be different from zero. Let

$$U = T^{-1} = \begin{bmatrix} I & T_{ML} \\ T_{LM} & 1 \end{bmatrix}^{-1} = \begin{bmatrix} (I - T_{ML}T_{LM})^{-1} & -(I - T_{ML}T_{LM})^{-1}T_{ML} \\ -T_{LM}(I - T_{ML}T_{LM})^{-1} & (1 - T_{LM}I^{-1}T_{ML})^{-1} \end{bmatrix}. \tag{5.10}$$

Because the multivariate normal distribution is closed under marginalisation, the marginal distribution of the manifest variables is multivariate normal, with 'modified' variance matrix given by $U_{MM} = (I - T_{ML}T_{LM})^{-1}$. Inverting this matrix gives ${U_{MM}}^{-1} = I - T_{ML}T_{LM}$ which, once scaled, induces

$$T^M = \begin{bmatrix} 1 & -\frac{\tau_{1L.23}\tau_{2L.13}}{\sqrt{1-\tau^2_{1L.23}}\sqrt{1-\tau^2_{2L.13}}} & -\frac{\tau_{1L.23}\tau_{3L.12}}{\sqrt{1-\tau^2_{1L.23}}\sqrt{1-\tau^2_{3L.12}}} \\ -\frac{\tau_{1L.23}\tau_{2L.13}}{\sqrt{1-\tau^2_{1L.23}}\sqrt{1-\tau^2_{2L.13}}} & 1 & -\frac{\tau_{2L.13}\tau_{3L.12}}{\sqrt{1-\tau^2_{2L.13}}\sqrt{1-\tau^2_{3L.12}}} \\ -\frac{\tau_{1L.23}\tau_{3L.12}}{\sqrt{1-\tau^2_{1L.23}}\sqrt{1-\tau^2_{3L.12}}} & -\frac{\tau_{2L.13}\tau_{3L.12}}{\sqrt{1-\tau^2_{2L.13}}\sqrt{1-\tau^2_{3L.12}}} & 1 \end{bmatrix}.$$

Because $\tau_{iL.jk} \neq 0$, $-\frac{\tau_{iL.jk}\tau_{jL.ik}}{\sqrt{1-\tau^2_{iL.jk}}\sqrt{1-\tau^2_{jL.ik}}} \neq 0$. Consequently T^M has no zero elements; there are no zero entries in the scaled inverse variance matrix of the manifest variables. For T^M to be positive definite the variances $1 - \tau^2_{iL.jk}$ have to be positive, which implies $0 < \tau^2_{iL.jk} < 1$. Note that whereas in the classical factor model the factor loadings $\lambda_i^2 \in [0, 1]$ (since positive semidefinite matrices are allowed), in the current thesis it is suggested $\tau^2_{iL.test} \in (0, 1)$ (since only positive definite matrices are considered).

The off-diagonal elements of the scaled inverse variance matrix of the manifest variables are the negatives of the partial correlation coefficients between the manifest variables. Therefore,

$$\rho_{ij.k} = \frac{\tau_{iL.jk}\tau_{jL.ik}}{\sqrt{1 - \tau^2_{iL.jk}}\sqrt{1 - \tau^2_{jL.ik}}} \quad \text{with distinct } i, j, k \in M. \tag{5.11}$$

When solving the system of three equations defined by Equation 5.11 with respect

to the three $\tau_{iL.jk}$ the following expression is obtained

$$\tau^2_{iL.jk} = \frac{1}{1 + \frac{\rho_{jk.i}}{\rho_{ij.k}\rho_{ik.j}}}. \tag{5.12}$$

Since $0 < \tau^2_{il.jk} < 1$, then $\frac{\rho_{jk.i}}{\rho_{ij.k}\rho_{ik.j}} > 0$.

Hence, the following result concerning the conditional independence structure between three manifest variables compatible with a single-factor GG model is proposed: three manifest variables can define a single-factor model if and only if their scaled inverse variance matrix is positive definite and the product of the three partial correlation coefficients is positive.

The implications of this result in terms of admissible regions for the three partial correlation coefficients between manifest variables, compatible with a single-factor GG model, are now derived. Figure 5.3 shows a graphical display of the allowable region for the three partial correlation coefficients. The ellipse represents the positive definiteness constraint on T^M defined by Equation 4.5. Additionally, there is now the constraint of a positive product of the three partial correlation coefficients. In practice, $\rho_{13.2}$ and $\rho_{23.1}$ vary between -1 and 1 (excluding zero values), whereas $\rho_{12.3}$ was given positive arbitrary values of 0.1, 0.5, 0.7 and 0.9, respectively in panels a), b), c) and d).

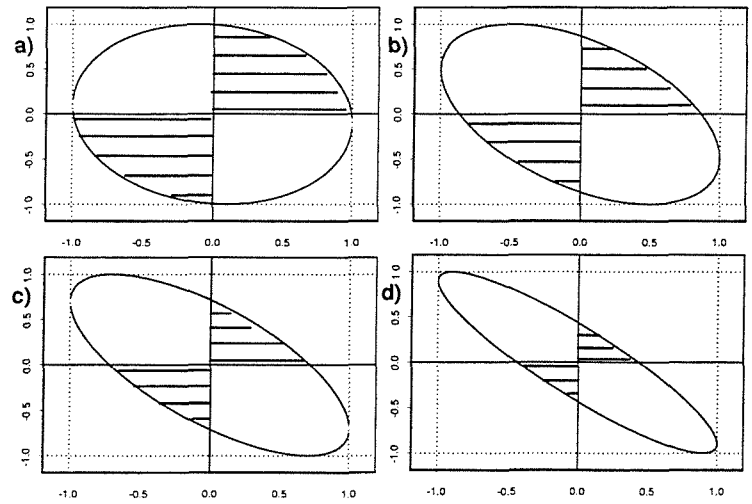

Figure 5.3: The admissible region for the three partial correlation coefficients compatible with a single-factor GG model: $\rho_{13.2}$ on the horizontal axis, $\rho_{23.1}$ on the vertical axis and $\rho_{12.3}$ taking positive arbitrary values of: a) 0.1, b) 0.5, c) 0.7, d) 0.9.

When there are more than three manifest variables in the model, additional constraints have to be imposed in order to have consistent solutions (as in Section 5.1.1). In the four manifest variables case, the system of equations defined by Equation 5.11 will have four τ and six ρ. For the solutions for each parameter to be consistent, the constraint that $\rho_{12.34}\rho_{34.12} = \rho_{13.24}\rho_{24.13} = \rho_{14.23}\rho_{23.14}$ is required. Therefore, an equivalent version of the two tetrad conditions presented in Equation 5.5 can be expressed in terms of partial correlations as

$$\rho_{12.34}\rho_{34.12} - \rho_{14.23}\rho_{23.14} = 0; \quad \rho_{13.24}\rho_{24.13} - \rho_{14.23}\rho_{23.14} = 0. \tag{5.13}$$

Setting $\rho_{ij.rest}$ equal to $\frac{\tau_{iL.rest}\tau_{jL.rest}}{\sqrt{1-\tau^2_{iL.rest}}\sqrt{1-\tau^2_{jL.rest}}}$ (a generalisation of Equation 5.11) guarantees the two conditions defined by Equation 5.13 are satisfied, since then

$$\rho_{12.34}\rho_{34.12} - \rho_{14.23}\rho_{23.14} = 0 \Leftrightarrow$$

$$\frac{\tau_{1L.234}\tau_{2L.134}}{\sqrt{1-\tau^2_{1L.234}}\sqrt{1-\tau^2_{2L.134}}}\frac{\tau_{3L.124}\tau_{4L.123}}{\sqrt{1-\tau^2_{3L.124}}\sqrt{1-\tau^2_{4L.123}}} - \frac{\tau_{1L.234}\tau_{4L.123}}{\sqrt{1-\tau^2_{1L.234}}\sqrt{1-\tau^2_{4L.123}}}\frac{\tau_{2L.134}\tau_{3L.124}}{\sqrt{1-\tau^2_{2L.134}}\sqrt{1-\tau^2_{3L.124}}} = 0.$$

Similar reasoning can be followed for the second tetrad condition. One should note that, when there are four or more manifest variables, the proof previously presented still holds, in the sense that marginalising over the latent variable induces a scaled inverse variance matrix of the manifest variables with no zero elements. However, because of the restrictions imposed by the tetrad conditions, although the model for the manifest variables has a complete independence graph, it is not necessarily the saturated model.

Hence, the following result is proposed: for a 4×4 positive definite scaled inverse variance matrix to be suitable for a single-factor GG model it is required that:

- $\rho_{ij.rest} = \frac{\tau_{iL.rest}\tau_{jL.rest}}{\sqrt{1-\tau^2_{iL.rest}}\sqrt{1-\tau^2_{jL.rest}}}$, so that $\frac{1}{1+\frac{\rho_{jk.rest}}{\rho_{ij.rest}\rho_{ik.rest}}} =$

  $$= \frac{1}{1+\frac{(\tau_{jL.rest}\tau_{kL.rest})/(\sqrt{1-\tau^2_{jL.rest}}\sqrt{1-\tau^2_{kL.rest}})}{(\tau_{iL.rest}\tau_{jL.rest}\tau_{iL.rest}\tau_{kL.rest})/(\sqrt{1-\tau^2_{iL.rest}}\sqrt{1-\tau^2_{jL.rest}}\sqrt{1-\tau^2_{iL.rest}}\sqrt{1-\tau^2_{kL.rest}})}} = \ldots$$

  $$= \tau^2_{iL.rest};$$

- $\frac{\rho_{jk.rest}}{\rho_{ij.rest}\rho_{ik.rest}} > 0$ (for all possible combinations of distinct i, j and k, from 1 to 4).

These conditions imply certain patterns of signs for the ρ, which are analysed in detail in Section 5.2.3.

5.2.2 The relationship between the classical and the proposed parameterisation of the single-factor model

The purpose of this section is to establish a relationship between the classical parameterisation of the single-factor model reviewed in Section 5.1.1 and the parameterisation of the single-factor GG model proposed in Section 5.2.1.

By Equation 5.8, $\lambda = \Sigma_{ML}$. Equation 5.10 shows a 'modified' Σ_{ML} (denoted by U_{ML}) can be obtained as $-(I-T_{ML}T_{LM})^{-1}T_{ML}$. Therefore, a possible non-standardised solution for the vector of factor loadings can be obtained as

$$\lambda = -(I - T_{ML}T_{LM})^{-1}T_{ML}. \tag{5.14}$$

This equation gives non-standardised λ_i as function of the p different τ. In the three manifest variables case this can be written as

$$\lambda_i = \frac{\tau_{iL.jk}}{1 - \tau^2_{iL.jk} - \tau^2_{jL.ik} - \tau^2_{kL.ij}}. \tag{5.15}$$

More generally, in the p variables case

$$\lambda_i = \frac{\tau_{iL.rest}}{1 - \sum_{q=1}^{p} \tau^2_{qL.rest}}. \tag{5.16}$$

The proof follows. From Equation 5.14, $\lambda = -(I-T_{ML}T_{LM})^{-1}T_{ML}$. Mardia, Kent and Bibby (1979, Property A.2.4f, page 459) stated that, if all necessary inverses exist,

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

where $A_{(p\times p)}$, $B_{(p\times n)}$, $C_{(n\times n)}$ and $D_{(n\times p)}$. Considering $A_{(p\times p)} = I$, $B_{(p\times 1)} = -T_{ML}$, $C_{(1\times 1)} = I$ and $D_{(1\times p)} = T_{LM}$, then

$$\begin{aligned}(I - T_{ML}T_{LM})^{-1} &= I^{-1} + I^{-1}T_{ML}\{1 - T_{LM}I^{-1}T_{ML}\}^{-1}T_{LM}I^{-1} \\ &= I + T_{ML}\{1 - T_{LM}T_{ML}\}^{-1}T_{LM}.\end{aligned}$$

Since $T_{LM}T_{ML}$ is a scalar and $\{1 - T_{LM}T_{ML}\}^{-1} = \frac{1}{1-\sum_{q=1}^{p}\tau^2_{qL.rest}}$,

$$\begin{aligned}\lambda &= -\left[T_{ML} + \{1 - T_{LM}T_{ML}\}^{-1}T_{ML}T_{LM}T_{ML}\right] \\ &= -\left[T_{ML} + \frac{\sum_{q=1}^{p}\tau^2_{qL.rest}}{1-\sum_{q=1}^{p}\tau^2_{qL.rest}}T_{ML}\right] \\ &= -\frac{1}{1-\sum_{q=1}^{p}\tau^2_{qL.rest}}T_{ML} = \left\{\frac{\tau_{iL.rest}}{1-\sum_{q=1}^{p}\tau^2_{qL.rest}}\right\}.\end{aligned}$$

Because the denominator is always positive, λ_i has the same sign as the corresponding $\tau_{iL.rest}$.

The 'classical standardised solution' for λ_i is obtained by dividing the factor loading by the standard deviations of the corresponding manifest variable and of the latent variable. The standardised solution is, therefore, given by

$$\lambda_i^{sc} = \frac{\lambda_i}{\sqrt{\nu_{ii}\,\nu_{LL}}}. \tag{5.17}$$

More generally,

$$\lambda^{sc} = \{diag\,\{U_{MM}\}\}^{-1/2}\ \lambda\ \nu_{LL}^{-1/2} \tag{5.18}$$

which is equivalent to scaling U and taking the corresponding partition scU_{ML}. By Equation 5.9, $\Theta_\delta = diag\{\Sigma_{XX} - \Sigma_{X\xi}\Sigma_{\xi X}\}$. Since $\lambda = \Sigma_{ML}$, it is possible to obtain

$$\Theta_\delta = diag\{\Sigma_{MM} - \lambda\lambda^T\}. \tag{5.19}$$

The corresponding classical standardised solution is given by

$$\Theta_\delta^{sc} = diag\{I - \lambda^{sc}(\lambda^{sc})^T\}, \tag{5.20}$$

which is equivalent to scaling U and calculating $\Theta_\delta^{sc} = diag\{I - (scU_{ML})(scU_{LM})^T\}$.

All the derived relationships between the classical parameterisation of the single-factor model and the parameterisation of the single-factor GG model hold theoretically, for the general p manifest variables case once the tetrad conditions are fulfilled.

5.2.3 Patterns of signs for the ρ compatible with a single-factor GG model

In Section 5.2.1 the admissible regions for the partial correlation coefficients between manifest variables arising from a single-factor GG model were investigated. The results derived then suggest that only certain patterns of signs for the ρ are compatible with a single-factor GG model. In brief: for all possible combinations of distinct i, j and k, $\frac{\rho_{jk.rest}}{\rho_{ij.rest}\,\rho_{ik.rest}} > 0$ (and the tetrad conditions have to be satisfied).

The aim of the current section is to derive a general rule for defining the patterns of signs of the partial correlations between manifest variables that are compatible with a single-factor GG model. As a matter of fact, the specification (construction) of a

single-factor model based on substantive knowledge has to take into account models that are compatible with the inverse variance structure of the manifest variables in the population. In other words, when the researcher specifies the manifest variables she or he wants to incorporate in the model, believing they have a certain structure of partial correlations in the population, certain patterns for the factor loadings (partial correlations between the manifest variables and the latent variable) can be expected. Also, the researcher has to be aware that there are patterns of signs in the inverse variance matrix that are not compatible with a model arising from a single-factor model.

From the generalisation of Equations 5.11 and 5.12, it is possible to conclude that the same value of $\rho_{ij.rest}$ can be associated with different combinations of signs for $\tau_{iL.rest}$ and $\tau_{jL.rest}$. Indeed, since $\rho_{ij.rest} = \frac{\tau_{iL.rest}\tau_{jL.rest}}{\sqrt{1-\tau^2_{iL.rest}}\sqrt{1-\tau^2_{jL.rest}}}$, the sign of $\rho_{ij.rest}$ equals the product of the signs of $\tau_{iL.rest}$ and $\tau_{jL.rest}$.

In the three manifest variables case, there are $2^3 = 8$ different combinations of signs for the three τ, as shown in Table 5.1. The first three columns of the table include the eight different combinations of signs for the τ (the partial correlation coefficients between each of the manifest variables and the latent variable). The last three columns show the corresponding signs for the ρ (the partial correlation coefficients between the three manifest variables). Recall that the sign of $\rho_{ij.k}$ equals the product of the signs of $\tau_{iL.jk}$ and $\tau_{jL.ik}$ (from Equation 5.11).

$\tau_{1L.23}$	$\tau_{2L.13}$	$\tau_{3L.12}$	$\rho_{12.3}$	$\rho_{13.2}$	$\rho_{23.1}$
+	+	+	+	+	+
+	+	−	+	−	−
+	−	+	−	+	−
−	+	+	−	−	+
−	−	−	+	+	+
−	−	+	+	−	−
−	+	−	−	+	−
+	−	−	−	−	+

Table 5.1: Different combinations for the signs of the τ and of the ρ in the three manifest variables case.

Various conclusions can be drawn from Table 5.1:

- either all three ρ are positive, or one is positive and two are negative: it is not possible to have a model with three manifest variables arising from a single factor model in which just one or all three ρ are negative;
- when $\rho_{ij.k}$ is positive and the other two ρ are negative, two different combinations of positive and negative values of τ can occur: either $\tau_{iL.rest}$ and $\tau_{jL.rest}$ are positive and $\tau_{kL.rest}$ is negative or $\tau_{iL.rest}$ and $\tau_{jL.rest}$ are negative and $\tau_{kL.rest}$ is positive;
- if $\tau_{iL.rest}$ is negative and the other two τ are positive (or the opposite), then all ρ in column (row) i of the scaled inverse variance matrix will be negative, and all remaining ρ will be positive. This condition defines an allowable pattern for the ρ compatible with a single-factor model. Table 5.2 helps visualising it. The table should be interpreted as follows. The first column displays the vector of the signs of the τ, whereas the second column shows the signs of the corresponding ρ. Then if, for example, $\tau_{1L.23}$ is negative and the other two τ are positive, all ρ in column one and all ρ in row one will be negative, and the remaining ρ will be positive.

signs $\tau =$		signs $\rho =$	
$\begin{bmatrix} \text{sign}\tau_{1L.23} \\ \text{sign}\tau_{2L.13} \\ \text{sign}\tau_{3L.12} \end{bmatrix}$		$\begin{bmatrix} & \text{sign}\rho_{12.3} & \text{sign}\rho_{13.2} \\ \text{sign}\rho_{12.3} & & \text{sign}\rho_{23.1} \\ \text{sign}\rho_{13.2} & \text{sign}\rho_{23.1} & \end{bmatrix}$	
$\begin{bmatrix} - \\ + \\ + \end{bmatrix}$	or $\begin{bmatrix} + \\ - \\ - \end{bmatrix}$	$\begin{bmatrix} & - & - \\ - & & + \\ - & + & \end{bmatrix}$	
$\begin{bmatrix} + \\ - \\ + \end{bmatrix}$	or $\begin{bmatrix} - \\ + \\ - \end{bmatrix}$	$\begin{bmatrix} & - & + \\ - & & - \\ + & - & \end{bmatrix}$	
$\begin{bmatrix} + \\ + \\ - \end{bmatrix}$	or $\begin{bmatrix} - \\ - \\ + \end{bmatrix}$	$\begin{bmatrix} & + & - \\ + & & - \\ - & - & \end{bmatrix}$	

Table 5.2: The allowable pattern of two negatives and one positive ρ in the three manifest variables case.

In the four manifest variables case, there are $2^4 = 16$ different combinations of signs for the four τ. The major conclusions are:

- for four manifest variables to arise from a single-factor model either all six partial correlations are positive or, if some are negative, the number of negative ρ can be three or four, and certain patterns have to be met. It is not possible that only one, only two, five or six of the ρ are negative;
- the six ρ will be positive, if either the four τ are all positive or all negative;
- if there are two positive and two negative τ, there will be 2 positive and 4 negative ρ;
- solutions with three positive and three negative ρ correspond to models with either one positive and three negative τ or models with one negative and three positive τ;
- if $\tau_{iL.rest}$ is negative and the remaining three τ are positive (or the opposite), then all ρ in column (row) i of the scaled inverse variance matrix will be negative, and all remaining ρ will be positive, as happened in the three variable case (recall the results in Table 5.2);
- if $\tau_{iL.rest}$ and $\tau_{jL.rest}$ are negative and the remaining two τ are positive (or the opposite), then all ρ in column (row) i except those in column (row) j (and all those in column (row) j except those in column (row) i) will be negative, and the remaining two ρ will be positive. Table 5.3 illustrates this result.

When five manifest variables are present there are $2^5 = 32$ different combinations of signs for the five τ. Similarly to the four variable case, different patterns of signs for the ρ are allowed. In short:

- if $\tau_{iL.rest}$ is negative and the remaining four τ are positive (or the opposite), then all ρ in column (row) i of the scaled inverse variance matrix will be negative, and all remaining ρ will be positive;
- if $\tau_{iL.rest}$ and $\tau_{jL.rest}$ are negative and the remaining three τ are positive (or the opposite), then all ρ in column (row) i except those in column (row) j (and all those in column (row) j except those in column (row) i) will be negative, and the remaining ρ will be positive.

signs $\tau =$		signs $\rho =$		
$\text{sign}\tau_{1L.234}$			$\text{sign}\rho_{12.34}$	$\text{sign}\rho_{13.24}$	$\text{sign}\rho_{14.23}$
$\text{sign}\tau_{2L.134}$		$\text{sign}\rho_{12.34}$		$\text{sign}\rho_{23.14}$	$\text{sign}\rho_{24.13}$
$\text{sign}\tau_{3L.124}$		$\text{sign}\rho_{13.24}$	$\text{sign}\rho_{23.14}$		$\text{sign}\rho_{34.12}$
$\text{sign}\tau_{4L.123}$		$\text{sign}\rho_{14.23}$	$\text{sign}\rho_{24.13}$	$\text{sign}\rho_{34.12}$	
$\begin{bmatrix} - \\ - \\ + \\ + \end{bmatrix}$	or $\begin{bmatrix} + \\ + \\ - \\ - \end{bmatrix}$	$\begin{bmatrix} & + & - & - \\ + & & - & - \\ - & - & & + \\ - & - & + & \end{bmatrix}$		
$\begin{bmatrix} - \\ + \\ - \\ + \end{bmatrix}$	or $\begin{bmatrix} + \\ - \\ + \\ - \end{bmatrix}$	$\begin{bmatrix} & - & + & - \\ - & & - & + \\ + & - & & - \\ - & + & - & \end{bmatrix}$		
$\begin{bmatrix} - \\ + \\ + \\ - \end{bmatrix}$	or $\begin{bmatrix} + \\ - \\ - \\ + \end{bmatrix}$	$\begin{bmatrix} & - & - & + \\ - & & + & - \\ - & + & & - \\ + & - & - & \end{bmatrix}$		

Table 5.3: The allowable pattern of two positive and four negative ρ in the four manifest variables case.

The rules presented above for the three, four and five manifest variables cases define the theoretical patterns of signs of the partial correlation coefficients between manifest variables that are compatible with a single-factor model. All these rules can be further simplified, as follows. Let $signs\,\tau$ be a $p \times 1$ column vector with the signs of the partial correlation coefficients between each of the p manifest variables X and the latent variable L. Let $signs\,\rho$ be a $p \times p$ symmetric matrix with the signs of the partial correlation coefficients between the p manifest variables as off-diagonal elements. The elements in the main diagonal are not of interest.

The following general rule is proposed: the off-diagonal elements of $signs\,\rho$ represent a pattern of signs for the ρ that is compatible with a single-factor GG model, with patterns of signs for the τ given by $signs\,\tau$, if and only if

$$(signs\,\tau)\,(signs\,\tau)^T = (signs\,\rho)$$

(for the off-diagonal elements). This rule refers to population parameters and assumes the tetrad conditions are satisfied.

5.3 Detecting a Model Consistent with a Single-Factor GG Model

The goodness-of-fit likelihood ratio test presented in Section 5.1.3 is often used to test the null hypothesis that m factors (latent variables) are sufficient to describe the data, against the alternative hypothesis that Σ_{XX} is unconstrained. In practice, this allows the data analyst to decide how many factors to fit to the data, if any. The usual strategy is to start with m equal to zero or one and perform a sequence of hypotheses tests, by increasing the number of factors by one, until the fit of the model is judged to be adequate. For further details see Mardia, Kent and Bibby (1979, page 268). However, as mentioned in Section 5.1.3, the procedure has been criticised in the literature: the fit always improves when m increases, making it unclear where to stop; because of the sequential character of the tests it is argued the p-value of a test should be regarded as a measure of the adequacy of the model; and although corrections have been suggested to improve the chi-square approximation of the test, it remains sensitive to the sample size (see also Bartholomew and Knott, 1999, Section 3.8). A different approach is suggested in this thesis, in order to detect a model consistent with a single-factor model, taking into account the power of the model selection procedure.

The focus of Section 5.2 is on the parameterisation of the single-factor GG model using partial correlations. Admissible regions for the ρ were investigated. In particular, patterns of signs of the population partial correlation coefficients were studied in order to help specify a single-factor GG model based on subject-matter knowledge. Additionally, in Section 5.2.1 it was shown that marginalising over the latent variable in the single-factor model induces no conditional independencies between manifest variables and, therefore, an independence graph that is complete.

All this suggests that when trying to detect (identify) a model consistent with a single-factor model, i.e., to detect the presence of a normally distributed latent variable, the data analyst should be looking at a GG model for the manifest variables with a complete independence graph. In fact, in Section 2.8 backwards elimination was recommended as the model selection procedure when trying to detect the presence of a latent variable in the context of GG models. Because the main interest is to identify strong associations between the manifest variables, the saturated model is first considered and afterwards a sequence of single edge exclusion tests is performed

(simultaneous multiple edge exclusion testing is not considered in this thesis). In practice, when performing model selection, type II errors (i.e. false acceptances of the null hypothesis) can occur and, therefore, the observed association structure will often not correspond to the true (saturated) model. Recall that, for each edge exclusion test, the null hypothesis is that the partial correlation is zero, the alternative hypothesis being that it is different from zero. Consequently, one or more edges will be missing in the selected independence graph for the manifest variables. Therefore, if the null hypothesis just presented is not rejected, the single-factor model does not hold, unless a type II error was made. In the case of three manifest variables there are three possible tests for single edge exclusion from the saturated model and, consequently, three possibilities of making a type II error. Hence, the power of the test statistics for single edge exclusion has to be taken into account when trying to select a model consistent with a single-factor GG model. Section 5.3.1 addresses this issue.

Also, it may happen that the sample available does not have the pattern of partial correlation coefficients signs the population has. That being the case, the general rule proposed in Section 5.2.3, to obtain patterns of signs for the ρ compatible with a single-factor GG model, may no longer apply, particularly when there are more than three manifest variables. The aim of Section 5.3.2 is, therefore, to suggest some recommendations to the data analyst when trying to identify a model consistent with a single-factor model.

5.3.1 The power of selecting a model consistent with a single-factor GG model

The power of selecting the saturated model when using the LRT statistic, in the three manifest variables case, was studied in detail in Section 4.1.4, using simulation. An asymptotic normal approximation to the power functions was derived in Section 4.1.5 and it was then explained how to obtain asymptotic power functions for the Wald and the score test statistics. Section 4.1.6 considered generalising such approximations to the situation of four, or more, manifest variables. The only constraint imposed then is that the scaled inverse variance matrix has to be positive definite.

Once a model arising from a single-factor GG model is considered, the additional constraint that, for all possible combinations of distinct i, j and k, $\frac{\rho_{jk.rest}}{\rho_{ij.rest}\,\rho_{ik.rest}} > 0$ has

to be imposed (as shown in Sections 5.2.1 and 5.2.3). In practical terms, in the three manifest variables case, this additional constraint means that the product of the three partial correlation coefficients has to be positive, i.e., either all three ρ are positive, or one is positive and the other two are negative.

Theoretical power values can be obtained using the asymptotic normal approximation derived in Section 4.1.5, namely Equation 4.6 for the three variables case. The three partial correlation coefficients are allowed to vary within the region of positive definiteness defined by Equation 4.5, with the additional constraint that their product has to be positive.

Simulated power values calculated in Section 4.1.4 and displayed in Figure 4.6 still hold for the power of selecting a model arising from a single-factor GG model, provided $\rho_{12.3}$, on the horizontal axis, just takes positive values. The main conclusion derived then still holds for the power of selecting a model compatible with a single-factor GG model with three manifest variables: power increases as the partial correlation coefficients increase. For $n = 200$, the probability of selecting the saturated model, using the LRT statistic, has a maximum value of $\simeq 0.3$ when one of the $\rho_{ij.k} \simeq 0.1$, even if the other two partial correlations are large. This probability goes up to $\simeq 0.8$, or almost 1, when the minimum $\rho_{ij.k} \simeq 0.2$ or 0.3, respectively.

The one-sided signed square-root version of the likelihood ratio test statistic can be used, assuming one is only interested in the positive values of the three partial correlation coefficients (believing that the three factor loadings of the underlying factor model have the same sign). Although, for simplicity, empirical power plots are not presented, their pattern is very similar to that obtained using a non-signed version of the likelihood ratio test statistic. Power has a maximum value of 0.4 when one of the ρ is small, even if the other two are large, which is higher than the 0.3 value obtained when a non-signed version of the test statistic was used. Globally power values are higher when the one-sided signed square-root version of the test statistic is used, by comparison with either the non-signed version or the two-sided signed square-root version, because one-sided tests tend to be more powerful than two-sided tests.

Additional simulations were carried out to investigate the effect of changing the sample size, in the three variables case. The LRT statistic was used. As in Section 4.1.4,

power was calculated as

$$P[T_{12}^L > \chi^2_{1;\,0.95} \text{ and } T_{13}^L > \chi^2_{1;\,0.95} \text{ and } T_{23}^L > \chi^2_{1;\,0.95} \mid \rho_{12.3},\ \rho_{13.2},\ \rho_{23.1}].$$

When $\rho_{12.3}$ takes positive values, power curves are exactly the same whether $\rho_{13.2}$ and $\rho_{23.1}$ are both positive or both negative. One thousand repetitions and four different sample sizes were considered. Figure 5.4 summarises the main results. Plots in panel a) correspond to $n = 50$, in panel b) to $n = 100$, in panel c) to $n = 500$ and in panel d) to $n = 1\,000$. In all panels $\rho_{12.3}$ is represented on the horizontal axis. The nine different lines in each plot correspond to the values of $\rho_{23.1}$ between 0.1 and 0.9. $\rho_{13.2}$ has a value of 0.1 in plots 1), of 0.2 in plots 2) and of 0.3 in plots 3).

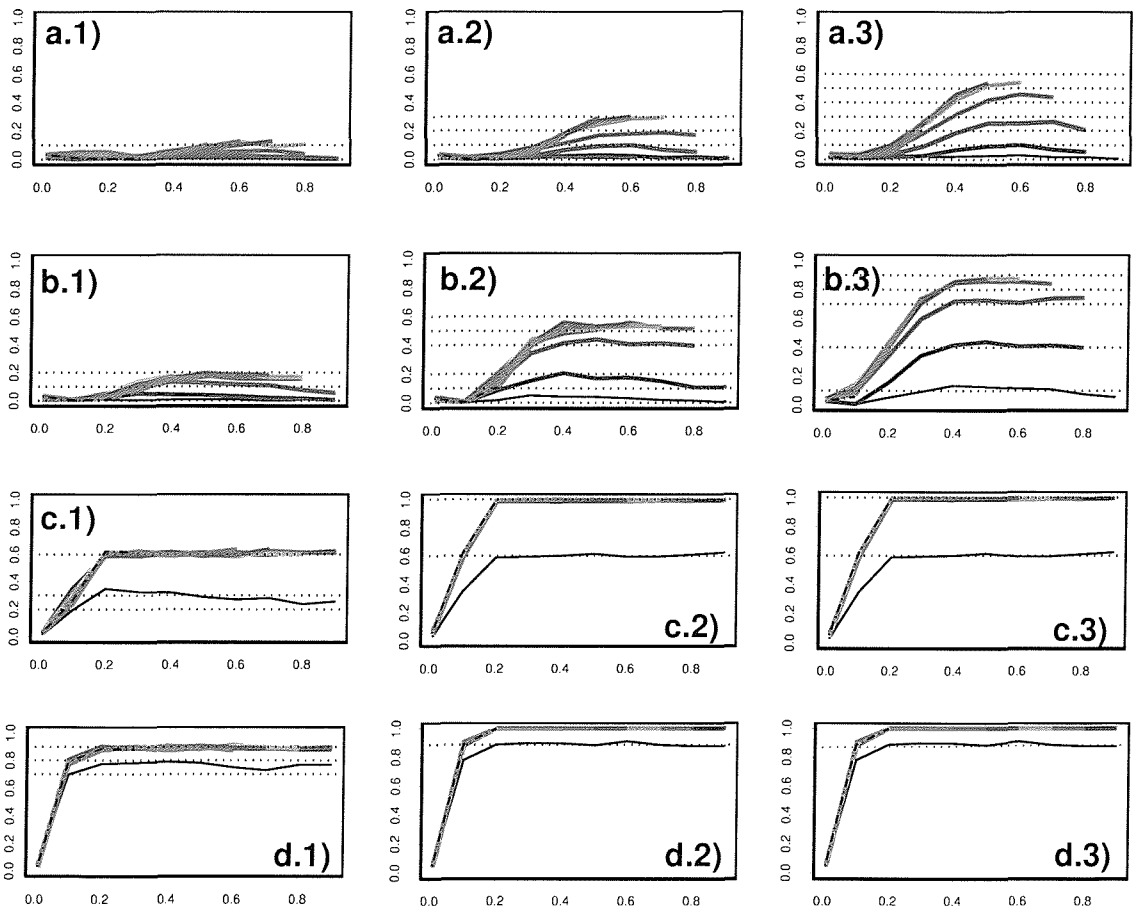

Figure 5.4: Power functions for the saturated model arising from a single-factor model with three manifest variables, using different sample sizes: a) $n = 50$, b) $n = 100$, c) $n = 500$, d) $n = 1\,000$. $\rho_{12.3}$ on the horizontal axis. $\rho_{23.1}$ from 0.1 to 0.9 in each plot. $\rho_{13.2}$ from 0.1 to 0.3 in plots 1), 2) and 3).

By comparison with the results for the case $n = 200$, previously considered, it is possible to conclude that, as n increases, power tends to rapidly increase, and consequently the non-monotonicity tends to disappear (see panels c) and d)). Note that when one of the partial correlation coefficients is around 0.1 (even if the other two are large) a maximum power value of 0.3 can be reached with a sample size of 200, whereas for $n = 500$ or $n = 1\,000$ power values of 0.6 and 0.9, respectively, can be obtained. As n decreases, see panels a) and b), the non-monotonicity effect becomes more evident than that in Figure 4.6, even for larger partial correlation coefficients. Power values tend to become very low, especially for small values of the partial correlation coefficients.

In Section 5.2.2 some formulae were derived relating the classical and the proposed parameterisations of the single-factor model. In the three manifest variables case, when $df = 0$ and no tetrad conditions have to be satisfied, the values of $\tau_{iL.rest}$, the partial correlations between each of the manifest variables and the latent variable, can be directly obtained from the observed partial correlation coefficients between manifest variables using Equation 5.12. The classical factor loadings λ_i can then be obtained from the $\tau_{iL.rest}$ using Equation 5.15. Using this mapping between possible values for the three partial correlation coefficients and corresponding expected factor loadings, it is possible to understand the association between the magnitude of the λ and the level of power (the probability of selecting the saturated model for the manifest variables) as a function of the sample size. Indeed, the awareness existing in the factor analysis literature as to what constitutes a large value of λ was considered important here.

Table 5.4 displays several possible combinations of values for the three partial correlation coefficients, the corresponding standardised factor loadings and power values for different sample sizes. The first three columns have the values for the ρ (small, intermediate and large values were used, although values very close to the boundary of positive definiteness were not considered to avoid improper solutions). The fourth, fifth and six column have the corresponding values of the λ (the standardised factor loadings that would be obtained if the single-factor model was fitted). The last five columns display the values of power for the five different sample sizes that were used in the simulations.

$\rho_{13.2}$	$\rho_{23.1}$	$\rho_{12.3}$	λ_1	λ_2	λ_3	$n = 50$	$n = 100$	$n = 200$	$n = 500$	n = 1000
0.1	0.1	0.1	0.33	0.33	0.33	0.001	0.000	0.012	0.192	0.697
0.1	0.1	0.7	0.85	0.85	0.28	0.015	0.002	0.017	0.282	0.729
0.1	0.2	0.5	0.66	0.81	0.36	0.012	0.040	0.189	0.615	0.883
0.1	0.3	0.8	0.90	0.97	0.66	0.014	0.076	0.290	0.613	0.886
0.1	0.4	0.6	0.75	0.94	0.62	0.055	0.155	0.327	0.617	0.898
0.1	0.6	0.6	0.85	0.98	0.85	0.100	0.188	0.312	0.596	0.893
0.2	0.2	0.2	0.50	0.50	0.50	0.005	0.085	0.499	0.986	1
0.2	0.3	0.4	0.65	0.75	0.56	0.066	0.415	0.816	0.995	1
0.2	0.4	0.5	0.74	0.87	0.68	0.166	0.505	0.832	0.999	1
0.2	0.5	0.6	0.87	0.95	0.83	0.285	0.531	0.822	0.994	1
0.3	0.3	0.3	0.65	0.65	0.65	0.089	0.597	0.977	1	1
0.3	0.4	0.5	0.82	0.87	0.77	0.417	0.857	0.993	1	1
0.3	0.5	0.5	0.86	0.92	0.86	0.522	0.881	0.992	1	1
0.4	0.4	0.4	0.82	0.82	0.82	0.549	0.965	1	1	1
0.4	0.5	0.5	0.93	0.95	0.93	0.776	0.985	1	1	1

Table 5.4: Power values as a function of the ρ, the λ and the sample size.

From Table 5.4 (and from plots in Figure 5.4) it is possible to conclude that, for a given combination of ρ and, consequently, of λ, power values are highly determined by the sample size. In other words, the probability of selecting the saturated model, arising from a single-factor model, varies according to the degree of association between manifest variables (the values of the ρ), but conclusions can be quite misleading if the sample size is small. For example, let us consider a combination of partial correlation coefficients of 0.3, 0.4, 0.5. The corresponding standardised factor loadings equal 0.82, 0.87, 0.77 (which, in the literature of factor analysis models, are considered large values of standardised factor loadings). If the sample size is small, say 50 observations, the power of selecting the saturated model for the manifest variables is only around 0.42, whereas it goes up to around 0.86 with a sample size of 100 and almost reaches 1 with a sample size of 200. The data analyst must, therefore, be very careful when dealing with small sample sizes. Indeed, LISREL literature on the robustness of the software recommends a minimum sample size of 100 or even 200, depending on the number of variables and parameters to be estimated.

5.3.2 Some recommendations to the data analyst

When trying to select a model consistent with a single-factor GG model the data analyst has to take into account the power of the test statistic for edge exclusion. As justified in Section 5.3.1, although the magnitude of the partial correlation coefficients is important, the question of the sample size is crucial and small samples can be quite misleading. Power values can, consequently, be low and the data analyst must be prepared to fit the single-factor GG model, even if there is one or more conditional independencies between the manifest variables. Furthermore, although only the three manifest variables case was considered in detail in Section 5.3.1, these conclusions should also apply when four or more manifest variables are present. Indeed, in Section 4.1.6, when power functions were compared for the three, the four and the five variables cases, considering all partial correlation coefficients equal, the conclusion was drawn that, for a given value of ρ and for a given sample size, the larger the number of variables, the lower power values are. In other words, the more manifest variables there are in the model, the more likely it is that the data analyst has to consider independence graphs with more edges missing, when trying to fit a single-factor GG model.

Another important aspect to be taken into account is that, due to sampling fluctuations, the pattern of signs of the sample partial correlation coefficients may differ from the pattern of signs of the ρ in the population from which the sample was drawn. One should note that, once it is believed a single-factor GG model holds, it is assumed that the manifest variables are drawn from the 'true' single-factor model and, therefore, a certain pattern of signs is expected for the sample partial correlation coefficients between manifest variables. The question of the patterns of signs for the ρ compatible with a single-factor GG model was investigated in Section 5.2.3, and a general rule was proposed for detecting compatible patterns, once the tetrad conditions are satisfied. In the three manifest variables case, $df = 0$, there is a unique solution for the single-factor model (up to rotation), no iteration is required and no tetrad conditions have to be satisfied. Consequently, trying to fit a single-factor model to a sample with an incompatible pattern of signs will lead to either a non-convergent or to an improper solution.

When the number of manifest variables is greater than three, the number of un-

knowns (τ and elements in Θ_δ) in the single-factor GG model (as presented in Section 5.2.1) is smaller than the number of known partial correlation coefficients between the manifest variables. And although the tetrad conditions are satisfied in the population, i.e., for the scaled inverse variance matrix of the manifest variables of the 'true' single-factor GG model, there is no guarantee that they will be fulfilled in the sample. Let us suppose an example of a 4×4 sample scaled inverse variance matrix, with a pattern of signs for the ρ that, according to the rules presented in Section 5.2.3, is not compatible with a single-factor model. The data analyst may decide not to fit such a model. Nevertheless, it may happen that, if the sample scaled inverse variance matrix is not close to satisfying the tetrad conditions, once the analyst tries to fit a single-factor model, she or he actually manages to obtain sensible estimates. This is due to the fact that current software for estimating factor models, aiming at either maximise the log likelihood function or minimise the discrepancy function, do not incorporate any additional constraint, rather than the fact that the sample variance matrix has to be positive definite. If the sample is not close to satisfying the tetrad conditions, it can happen that the software manages 'a way out' and is able to fit a single-factor model using a sample that has a 'non-allowable' structure, i.e., patterns of signs of the partial correlation coefficients that, in principle, are not compatible with a single-factor model. The data analyst may consequently be 'surprised' by the fact that convergence was achieved or that parameter estimates have a structure of signs different from the one that was expected. It is, therefore, important for the analyst to be aware of what is happening.

The question of incompatible sign patterns of sample covariances of manifest variables (not inverse variance matrices, as studied above) was mentioned by Boomsma (1985) and by Boomsma and Hoogland (2001) in their study on the robustness of maximum likelihood estimation with LISREL, trying to account for non-convergence and improper solutions. A simulation study was used to predict non-convergence in situations where the sign pattern of the sample covariances between manifest variables is incompatible with the signs of products of possible factor loadings in the population model. The following results were obtained: *'in the case of three manifest variables, for a sample size of 25 (with 400 repetitions) the correct-prediction rate of (non)convergence was 98%, whereas for a sample size of 50 it was 99%. However, us-*

ing similar prediction criteria for a model with 4 manifest variables was unsuccessful'. Boomsma and Hoogland (2001) stated that *'for some models, inadmissible sign patterns of the observed sample covariances linked to the same factor have good predicted value for non-convergence.'*

Yet, besides not mentioning why in the four manifest variables case the prediction criteria was unsuccessful, in his study Boomsma does not seem to take into account the constraint imposed by the tetrad conditions. In this thesis it it proposed that compatible sign patterns of the population partial correlation coefficients should be used to construct (specify) the theoretical single-factor model. Next, the pattern of signs of the sample partial correlations should be considered. In the three variables case, this pattern is a good criteria as to whether it is possible to fit a single-factor model (particularly if the sample size is not very small). When four or more manifest variables are present, the analyst should be aware that having an incompatible pattern of signs for the sample partial correlation coefficients may lead to fitting a single-factor model which does not have the expected combination of signs for the factor loadings or partial correlations between manifest variables and the latent variable, if the sample being analysed is not close to fulfilling the tetrad conditions.

Some brief recommendations to the data analyst, when trying to detect the presence of a latent variable are:

- first, sample partial correlation coefficients should be calculated. If an incompatible pattern of signs is found, there is strong evidence that a single-factor model cannot be fitted with such a sample. If the data analyst believes that in the factor model all factor loadings should be positive, and finds evidence of that from the sample (by obtaining three positive ρ) then, in a second step, a signed square-root version of a test statistic should be used to perform one-sided tests for single-edge exclusion from the saturated model. If the analyst thinks some of the factor loadings could be negative, and finds evidence for that from the sample (an allowed pattern of signs for the ρ is obtained), then a signed square-root version of a test statistic for single-edge exclusion should be used, in a second step, to perform a two-sided hypothesis test;

- in a second step the most appropriate test statistic for single edge exclusion from

the saturated model should be used. If the independence graph of the manifest variables is complete, the variables under analysis can be considered indicators of a single-factor GG model. If the model with a complete independence graph does not hold, and particularly in the cases of small values of one or more ρ and small sample sizes, associated with small power of selecting the saturated model, the analyst should consider obtaining a larger sample size. That not being possible, she or he should still try to fit a single-factor model, and be cautious when interpreting the results.

In summary, Section 5.3 has provided some evidence that when performing model selection to detect a single-factor GG model, under certain patterns of the observed partial correlation coefficients, the data analyst must still consider an observed association structure that is not necessarily the one induced by the saturated model of the manifest variables, i.e. the true model, particularly if the sample size is small.

5.4 The Classical Parameterisations of the Latent Class Model

The aim of latent class analysis is to define a latent variable as a set of classes within which the manifest categorical variables are locally independent. There are two *central assumptions* in latent class models. One is that the population consists of a set of mutually exclusive and exhaustive homogeneous subpopulations, which make up a latent classification that is discrete by definition. The other is local independence, i.e., within a given latent subpopulation, all manifest indicators will be statistically independent. In other words, manifest variables are conditionally independent given the categories of the latent variable. The latent class analysis can be either exploratory or confirmatory. In the former case there are no *a priori* restrictions on the parameters of the model, whereas in the latter case restrictions can be imposed.

There are two main *classical parameterisations* of the latent class model: one is based on conditional probabilities (Goodman, 1974) and the other uses a log-linear model formulation (Haberman, 1979). The notation adopted in the thesis is used to present the latent class model under both parameterisations, for the general case of four

manifest variables and one latent variable, all with two or more categories (classes). Let X_1, X_2, X_3 and X_4 be four manifest variables, with A, B, C and D classes, respectively, and L be a latent variable with T classes. The observed information about X_1, X_2, X_3 and X_4 can be summarised in a $A \times B \times C \times D$ contingency table. Let $\pi_{1234}(a, b, c, d)$ denote the probability that an individual will be at level (a, b, c, d) with respect to the joint variable (X_1, X_2, X_3, X_4). All π are assumed positive.

Using Goodman's *conditional probabilities parameterisation*, the observed variables X_1, X_2, X_3 and X_4 are conditionally independent given the level l of the latent variable L if

$$\pi_{1234L}(a, b, c, d, l) = \pi_L(l)\, \pi_{1|L}(a, l)\, \pi_{2|L}(b, l)\, \pi_{3|L}(c, l)\, \pi_{4|L}(d, l) \tag{5.21}$$

and

$$\pi_{1234}(a, b, c, d) = \sum_{l=1}^{T} \pi_{1234L}(a, b, c, d, l). \tag{5.22}$$

Also,

$$\sum_{l=1}^{T} \pi_L(l) = \sum_{a=1}^{A} \pi_{1|L}(a, l) = \sum_{b=1}^{B} \pi_{2|L}(b, l) = \sum_{c=1}^{C} \pi_{3|L}(c, l) = \sum_{d=1}^{D} \pi_{4|L}(d, l) = 1$$

The *latent class probabilities*, $\pi_L(l)$, and the *conditional probabilities*, $\pi_{1|L}(a, l)$, $\pi_{2|L}(b, l)$, $\pi_{3|L}(c, l)$, $\pi_{4|L}(d, l)$, are the two fundamental quantities of latent class analysis under Goodman's parameterisation. In other words, Equation 5.21 states that the probability that a randomly selected case will be located in cell (a, b, c, d, l) equals the product of the probability of a randomly selected case being at level l of the latent variable L times the conditional probabilities that a case in class l of the latent variable will be located at a certain category of each of the manifest variables. The conditional probabilities represent a measure of the degree of association between each of the manifest variables and each of the latent classes and can be compared to the factor loadings in factor analysis.

Haberman (1979) presented the unrestricted *latent class model* as a *log-linear model*

$$\begin{aligned} \log m_{1234L}(a, b, c, d, l) = \; & \lambda_\emptyset + \lambda_1(a) + \lambda_2(b) + \lambda_3(c) + \lambda_4(d) + \lambda_L(l) \\ & + \lambda_{1L}(a, l) + \lambda_{2L}(b, l) + \lambda_{3L}(c, l) + \lambda_{4L}(d, l), \end{aligned} \tag{5.23}$$

where $m_{1234L}(a, b, c, d, l)$ are the expected counts of $n_{1234L}(a, b, c, d, l)$, not known because L is unobserved. The $\lambda_\emptyset$ term is a normalising constant, to ensure that the sum of the expected counts over all possible combinations of cells equals the sample size $n_\emptyset$.

As far as the process of *identification* of a latent class model is concerned, the first step consists of checking for a non-negative number of degrees of freedom. Indeed, a necessary condition for identifiability is that the number of df is non-negative. The total number of degrees of freedom is given by

$$\text{df} = (A \times B \times C \times D - 1) - [(T-1) + T(A-1) + T(B-1) + T(C-1) + T(D-1)],$$

where $A \times B \times C \times D - 1$ is the available number of degrees of freedom in the cross-tabulation of the manifest variables and $(T-1)+T(A-1)+T(B-1)+T(C-1)+T(D-1)$ is the total number of parameters to be estimated in the unrestricted model, defined either by Equation 5.21 or by Equation 5.23.

Yet, Goodman (1974) showed that, even when df is positive, the model may not be identified. Unidentifiable models can be made identifiable by imposing restrictions on one or more of the parameters of the model. Goodman (1974) provided a necessary and sufficient condition for determining the *local identifiability* of a latent class model: '*the matrix of partial derivatives of the nonredundant observed probabilities with respect to the nonredundant model parameters must be of full column rank* (in the four manifest variables case equal to [(T - 1) + T (A - 1) + T (B - 1) + T (C -1) + T (D -1)]), *i.e., there must be no linearly dependent columns*'. This condition has to be satisfied for the specific set of data being analysed: there is no a priori way of guaranteeing the identification of a model. In practice, existing software for estimating latent class models, such as lEM and Latent GOLD, have a built in check for identification (based on the information matrix), giving a warning message to the data analyst whenever the specified model is not identified.

The two classical parameterisations can be easily related. The marginal cell probabilities $\pi_{1234}(a,b,c,d)$ can be derived from Equation 5.23 as

$$\pi_{1234}(a,b,c,d) = \frac{1}{n_\emptyset} \sum_l m_{1234L}(a,b,c,d,l). \tag{5.24}$$

Hence, the parameters of Equation 5.21 can be derived from Equations 5.23 and 5.24 as

$$\pi_L(l) = \tfrac{1}{n_\emptyset}\, m_{1234L}(l)$$

$$\pi_{1|L}(a,l) = \tfrac{m_{1234L}(a,l)}{n_\emptyset\ \pi_L(l)} \qquad \pi_{2|L}(b,l) = \tfrac{m_{1234L}(b,l)}{n_\emptyset\ \pi_L(l)}$$

$$\pi_{3|L}(c,l) = \tfrac{m_{1234L}(c,l)}{n_\emptyset\ \pi_L(l)} \qquad \pi_{4|L}(d,l) = \tfrac{m_{1234L}(d,l)}{n_\emptyset\ \pi_L(l)}$$

If both the manifest and the latent variable are binary, the odds ratio between the manifest variable X_M and the latent variable L can be defined as

$$\psi_{ML} = \frac{\pi_{M|L}(0,0)\,\pi_{M|L}(1,1)}{\pi_{M|L}(0,1)\,\pi_{M|L}(1,0)} = \frac{\pi_{M|L}(0,0)\,\left(1-\pi_{M|L}(0,1)\right)}{\pi_{M|L}(0,1)\,\left(1-\pi_{M|L}(0,0)\right)}.$$

Consequently, $\log \psi_{ML} = \lambda_{ML}$.

The probability that an individual belonging to latent class l of L will respond to item X_M in category m equals

$$\pi_{M|L}(m,l) = \frac{\pi_{ML}(m,l)}{\pi_L(l)} = \frac{\exp\{\lambda_M(m) + \lambda_{ML}(m,l)\}}{\sum_m \exp\{\lambda_M(m) + \lambda_{ML}(m,l)\}}.$$

Therefore, if both X_M and L are binary, the logits of category 1 of X_M versus category 0 of X_M (i.e., the tendency to answer item M in category 1 rather than in category 0) can be expressed as

$$\log \frac{\pi_{M|L}(1,0)}{\pi_{M|L}(0,0)} = \lambda_M \quad \text{and} \quad \log \frac{\pi_{M|L}(1,1)}{\pi_{M|L}(0,1)} = \lambda_M + \lambda_{ML}.$$

As far as *confirmatory latent class analysis* is concerned, there are mainly two types of constraints that can be imposed on the parameters of the model, namely:

- conditional probabilities constraints, that can either be equality constraints or specific value restrictions. Setting $\pi_{M|L}(m,l) = x$ is restricting the value of the conditional probability. For example, setting $\pi_{M|L}(0,0) = \pi_{M|L}(1,0)$ means that respondents in category 0 of L are equally likely to answer in categories 0 or 1 of X_M. If the latent class has three or more categories it is possible to constrain two of them to have the same probability of response in one of the categories of a manifest variable, i.e., $\pi_{M|L}(m,1) = \pi_{M|L}(m,2)$. One should note, however, that it does not make sense to impose such a restriction if the latent variable is binary, since it would imply independence between the manifest variable and the latent, and in latent class analysis manifest variables are assumed to depend on the latent variable;

- latent class probabilities constraints, that can either be testing whether the probability equals a specified value, or whether $T-1$ of the T classes of the latent variable are equiprobable or have some relationship between them. One should note that, when the latent variable is binary, it is not sensible to test either if $\pi_L(l) = 0$ or if $\pi_L(l) = 1$, since it implies all the population is in one class and the latent variable does not make sense.

This thesis considers unconstrained latent class models.

As far as procedures for estimating the parameters are concerned, Haberman (1979) suggested the use of iterative proportional fitting, but similar results can be obtained using the EM algorithm. Indeed, existing software for fitting latent class models such as lEM and Latent GOLD use the EM algorithm to fit latent class models.

5.5 The Latent Class Graphical Log-Linear Model

Section 5.4 reviewed the two classical parameterisations of the latent class model. In this section the latent class model is parameterised as a GLL model, allowing a categorical latent variable to be included in the graphical log-linear model framework. All variables, both manifest (M) and latent (L), are assumed binary, categories being coded as 0 and 1. Corner point constraints are used in the log-linear formulation, i.e., $\lambda_L(0) = 0$, $\lambda_M(0) = 0$, $\lambda_{ML}(0,1) = 0$ and $\lambda_{ML}(1,0) = 0$, where M stands for any manifest variable in the model. For simplicity of notation $\lambda_L(1) = \lambda_L$, $\lambda_M(1) = \lambda_M$ and $\lambda_{ML}(1,1) = \lambda_{ML}$.

Throughout the section the notion of *sensible model* is used several times. It is assumed that, for the latent class model to hold, it has to be sensible. A latent class GLL model is defined as sensible if:

- none of the two categories of the latent variable is empty, i.e., both $\pi_L(0)$ and $\pi_L(1)$ are different from zero and, consequently, from one. This implies that in the log-linear formulation $\lambda_L \neq 0$;

- there are no structural zeros, either in the cross-tabulation of the manifest variables or in the cross-tabulation between the latent variable and each of the manifest variables. All cell probabilities are assumed positive. In the log-linear expansion, all this implies that, for all manifest variables M, $\lambda_M \neq 0$ and $\lambda_{ML} \neq 0$;

- there is an association between the latent variable and each of the manifest variables, i.e., the odds ratio between each manifest variable and the latent variable has to be different from one, implying all $\log \psi_{ML} \neq 0$ and all $\lambda_{ML} \neq 0$. Otherwise it would make no sense to have that manifest variable in the model - recall that in the latent class model the manifest variable is assumed to depend on the

unobserved latent variable.

The section starts with the proposed parameterisation of the latent class GLL model. The conditional independence structure between the manifest variables arising from a latent class GLL model, by marginalising over the latent variable L, is then investigated.

5.5.1 Parameterising the latent class GLL model

A natural way of including a categorical (binary) latent variable in a graphical model, the remaining manifest variables also being categorical (binary), is to parameterise the latent class model as a graphical log-linear model, using a log-linear expansion. It is proposed that the latent class GLL model is represented by a conditional independence graph, as justified in Section 2.4.2. Figure 5.5 displays the independence graphs associated with two latent class models, one with two manifest random variables, 1 and 2 (in panel a) and the other with three manifest random variables, 1, 2 and 3 (in panel b), each of them with a single latent variable (L). All variables are binary.

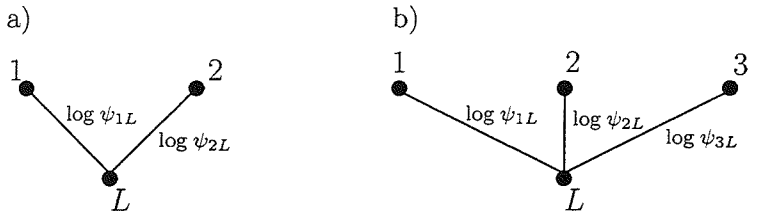

Figure 5.5: Examples of latent class models: a) the binary manifest variables 1 and 2 are conditionally independent, given the binary latent variable L; b) 1, 2 and 3 are conditionally independent given L. The edges are associated with the conditional log odds ratio between the latent and each of the manifest variables, given the remaining variables are at level 1.

The independence graph in panel a) corresponds to the following latent class model

$$\log m_{12L}(a, b, l) = \lambda_\emptyset + \lambda_1 + \lambda_2 + \lambda_L + \lambda_{1L} + \lambda_{2L}, \tag{5.25}$$

where $\log \psi_{1L} = \lambda_{1L}$ and $\log \psi_{2L} = \lambda_{2L}$. The independent graph in panel b) corresponds to the model

$$\log m_{123L}(a, b, c, l) = \lambda_\emptyset + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_L + \lambda_{1L} + \lambda_{2L} + \lambda_{3L}, \tag{5.26}$$

where $\log \psi_{1L} = \lambda_{1L}$, $\log \psi_{2L} = \lambda_{2L}$ and $\log \psi_{3L} = \lambda_{3L}$. Thus, generalising the log-linear expansion to four or more manifest variables is straightforward.

Regarding the identification of these models, it was stated earlier that, although there exists no a priori condition to guarantee models are identified, a first requirement is a non-negative number of degrees of freedom. In the two binary manifest variables case, with a single binary latent variable, the latent class GLL model is not identified. In fact, there are three knowns (the four cell probabilities that add up to one) and five parameters to be estimated: λ_1, λ_2, λ_L, λ_{1L} and λ_{2L} ($\lambda_\emptyset$ is a normalising constant). Consequently, the number of degrees of freedom is not non-negative, and the model is not identified unless two additional constraints are imposed on the parameters. In the three binary manifest variables case, with a binary latent variable, the necessary condition for identifiability is satisfied: $df = 0$ (and so is non-negative). There are now seven knowns (the eight cell probabilities that add up to one) and seven parameters to be estimated: λ_1, λ_2, λ_3, λ_L, λ_{1L}, λ_{2L} and λ_{3L}. The number of degrees of freedom is zero; there is a one to one correspondence between the cell probabilities and the seven λ terms: each of them can be uniquely obtained from the observed values. In the case of a latent class model with four or more binary manifest variables and a single binary latent variable the number of degrees of freedom is always positive, the necessary condition for identifiability being satisfied.

5.5.2 The conditional independence structure of the manifest variables arising from a latent class GLL model

The aim of this section is to investigate the conditional independence structure between the binary manifest variables arising from marginalising the latent class GLL model over the latent variable. It is proved that marginalising over the latent variable L, either in panel a) or in panel b) of Figure 5.5, yields an independence structure between the manifest variables with no conditional independencies and complete conditional independence graphs. Besides the two and the three binary manifest variables cases, the four variables case is also considered and it is shown how results apply to higher dimensional contingency tables.

The two manifest variables latent class model

As mentioned before, it is assumed that the sensible latent class GLL model holds, which implies that all cell probabilities are assumed positive, the two manifest variables depend on the latent variable, and so all five lambda parameters have to be different from zero, i.e., $\lambda_1 \neq 0$, $\lambda_2 \neq 0$, $\lambda_L \neq 0$, $\lambda_{1L} \neq 0$ and $\lambda_{2L} \neq 0$. Also, for identifiability reasons, the values of two of these parameters have to be constrained.

The aim now is to prove that, if the two binary manifest variables latent class GLL model holds and is sensible, marginalising over the latent variable L implies that the manifest variables, 1 and 2, are not independent: the odds ratio ψ_{12} is different from one and the independence graph of the two manifest variables is complete. The proof requires expressing the odds ratio between 1 and 2 as a function of the λ terms of the log-linear expansion of the latent class model and showing that, if the latent class model is sensible and holds, such odds ratio is different from one. The proof follows. The expected cell counts in each of the four cells of the 2×2 contingency table of the two manifest variables can be obtained, using Equation 5.25, as

$$
\begin{aligned}
m_{12}(0,0) &= m_{12L}(0,0,0) + m_{12L}(0,0,1) = \exp\{\lambda_\emptyset\}\,(1+\exp\{\lambda_L\}) \\
m_{12}(0,1) &= m_{12L}(0,1,0) + m_{12L}(0,1,1) = \exp\{\lambda_\emptyset + \lambda_2\}\,(1+\exp\{\lambda_L + \lambda_{2L}\}) \\
m_{12}(1,0) &= m_{12L}(1,0,0) + m_{12L}(1,0,1) = \exp\{\lambda_\emptyset + \lambda_1\}\,(1+\exp\{\lambda_L + \lambda_{1L}\}) \\
m_{12}(1,1) &= m_{12L}(1,1,0) + m_{12L}(1,1,1) = \exp\{\lambda_\emptyset + \lambda_1 + \lambda_2\}\,(1+\exp\{\lambda_L + \lambda_{1L} + \lambda_{2L}\})\,.
\end{aligned}
$$

Consequently, the odds ratio ψ_{12} can be expressed as

$$
\psi_{12} = \frac{m_{12}(0,0)\,m_{12}(1,1)}{m_{12}(0,1)\,m_{12}(1,0)} = \ldots = \frac{(1+\exp\{\lambda_L\})\,(1+\exp\{\lambda_L + \lambda_{1L} + \lambda_{2L}\})}{(1+\exp\{\lambda_L + \lambda_{1L}\})\,(1+\exp\{\lambda_L + \lambda_{2L}\})}.
$$

Since all cell probabilities are assumed positive, the model is sensible and $\lambda_1 \neq 0$, $\lambda_2 \neq 0$, $\lambda_L \neq 0$, $\lambda_{1L} \neq 0$ and $\lambda_{2L} \neq 0$, the value of the odds ratio, ψ_{12}, has to be different from one. Recall that $\log \psi_{12} = \lambda_{12}$. If $\psi_{12} \neq 1$, then $\log \psi_{12} \neq 0$ and $\lambda_{12} \neq 0$ and so 1 and 2 are not independent. One should note that ψ_{12} can still tend to one, in the limit, if λ_L is very large in absolute value by comparison with λ_{1L} and λ_{2L}. That being the case, the values in the reference category of the latent variable will tend to zero, corresponding to empty cells and to a non sensible latent class model.

In conclusion: if a two binary manifest variables latent class model is sensible and holds, marginalising over the latent variable implies a complete independence graph for the manifest variables, which are not independent.

The three manifest variables latent class model

If a sensible latent class GLL model, with three binary manifest variables holds, because all cell probabilities are assumed positive and the three manifest variables depend on the latent variable, all seven lambda parameters are different from zero, i.e., $\lambda_1 \neq 0$, $\lambda_2 \neq 0$, $\lambda_3 \neq 0$, $\lambda_L \neq 0$, $\lambda_{1L} \neq 0$, $\lambda_{2L} \neq 0$ and $\lambda_{3L} \neq 0$. There are zero degrees of freedom; it is a saturated model.

It is proposed that, if the three manifest variables latent class model is sensible and holds, marginalising over the latent variable L induces the saturated model for the three manifest variables, implying no conditional independencies between 1, 2 and 3. In other words, all six conditional odds ratios between the three binary manifest variables have to be different from one. The proof requires expressing the six conditional odds ratios as a function of the λ terms of the log-linear expansion of the latent class model and showing that all six conditional odds ratio are different from one.

The expected cell counts in each of the eight cells of the $2 \times 2 \times 2$ contingency table of the three manifest variables can be obtained, using Equation 5.26, as

$$
\begin{aligned}
m_{123}(0,0,0) &= m_{123L}(0,0,0,0) + m_{123L}(0,0,0,1) = \exp\{\lambda_\emptyset\}\,(1+\exp\{\lambda_L\}) \\
m_{123}(0,0,1) &= \ldots = \exp\{\lambda_\emptyset + \lambda_3\}\,(1+\exp\{\lambda_L + \lambda_{3L}\}) \\
m_{123}(0,1,0) &= \ldots = \exp\{\lambda_\emptyset + \lambda_2\}\,(1+\exp\{\lambda_L + \lambda_{2L}\}) \\
m_{123}(0,1,1) &= \ldots = \exp\{\lambda_\emptyset + \lambda_2 + \lambda_3\}\,(1+\exp\{\lambda_L + \lambda_{2L} + \lambda_{3L}\}) \\
m_{123}(1,0,0) &= \ldots = \exp\{\lambda_\emptyset + \lambda_1\}\,(1+\exp\{\lambda_L + \lambda_{1L}\}) \\
m_{123}(1,0,1) &= \ldots = \exp\{\lambda_\emptyset + \lambda_1 + \lambda_3\}\,(1+\exp\{\lambda_L + \lambda_{1L} + \lambda_{3L}\}) \\
m_{123}(1,1,0) &= \ldots = \exp\{\lambda_\emptyset + \lambda_1 + \lambda_2\}\,(1+\exp\{\lambda_L + \lambda_{1L} + \lambda_{2L}\}) \\
m_{123}(1,1,1) &= \ldots = \exp\{\lambda_\emptyset + \lambda_1 + \lambda_2 + \lambda_3\}\,(1+\exp\{\lambda_L + \lambda_{1L} + \lambda_{2L} + \lambda_{3L}\})\,.
\end{aligned}
$$

Consequently, the conditional odds ratios $\psi_{12.3=0}$ and $\psi_{12.3=1}$ can be expressed as

$$
\psi_{12.3=0} = \frac{m_{123}(0,0,0)\,m_{123}(1,1,0)}{m_{123}(0,1,0)\,m_{123}(1,0,0)} = \ldots = \frac{(1+\exp\{\lambda_L\})\,(1+\exp\{\lambda_L + \lambda_{1L} + \lambda_{2L}\})}{(1+\exp\{\lambda_L + \lambda_{1L}\})\,(1+\exp\{\lambda_L + \lambda_{2L}\})}
$$

and

$$
\psi_{12.3=1} = \frac{m_{123}(0,0,1)\,m_{123}(1,1,1)}{m_{123}(0,1,1)\,m_{123}(1,0,1)} = \ldots = \frac{(1+\exp\{\lambda_L + \lambda_{3L}\})\,(1+\exp\{\lambda_L + \lambda_{1L} + \lambda_{2L} + \lambda_{3L}\})}{(1+\exp\{\lambda_L + \lambda_{1L} + \lambda_{3L}\})\,(1+\exp\{\lambda_L + \lambda_{2L} + \lambda_{3L}\})}\,.
$$

Recall that $\log(\psi_{12.3=0}) = \lambda_{12}$ and $\log(\psi_{12.3=1}) = \lambda_{12} + \lambda_{123}$. Similar reasoning can be followed to obtain the four remaining conditional odds ratios. More generally, if i, j and k are three binary manifest variables, and corner point constraints are used,

$$
\psi_{ij.k=0} = \frac{(1+\exp\{\lambda_L\})\,(1+\exp\{\lambda_L + \lambda_{iL} + \lambda_{jL}\})}{(1+\exp\{\lambda_L + \lambda_{iL}\})\,(1+\exp\{\lambda_L + \lambda_{jL}\})}, \tag{5.27}
$$

$$\psi_{ij.k=1} = \frac{(1+\exp\{\lambda_L+\lambda_{kL}\})\ (1+\exp\{\lambda_L+\lambda_{iL}+\lambda_{jL}+\lambda_{kL}\})}{(1+\exp\{\lambda_L+\lambda_{iL}+\lambda_{kL}\})\ (1+\exp\{\lambda_L+\lambda_{jL}+\lambda_{kL}\})}. \tag{5.28}$$

Also,

$$\log(\psi_{ij.k=0}) = \lambda_{ij} \text{ and } \log(\psi_{ij.k=1}) = \lambda_{ij} + \lambda_{ijk}. \tag{5.29}$$

Consequently, $\lambda_{ijk} =$

$$\log\left[\frac{(1+\exp\{\lambda_L+\lambda_{iL}\})\ (1+\exp\{\lambda_L+\lambda_{jL}\})\ (1+\exp\{\lambda_L+\lambda_{kL}\})\ (1+\exp\{\lambda_L+\lambda_{iL}+\lambda_{jL}+\lambda_{kL}\})}{(1+\exp\{\lambda_L\})\ (1+\exp\{\lambda_L+\lambda_{iL}+\lambda_{jL}\})\ (1+\exp\{\lambda_L+\lambda_{iL}+\lambda_{kL}\})\ (1+\exp\{\lambda_L+\lambda_{jL}+\lambda_{kL}\})}\right]. \tag{5.30}$$

Since all cell probabilities are positive, only sensible latent class models are considered and $\lambda_1 \neq 0$, $\lambda_2 \neq 0$, $\lambda_3 \neq 0$, $\lambda_L \neq 0$, $\lambda_{1L} \neq 0$, $\lambda_{2L} \neq 0$ and $\lambda_{3L} \neq 0$, all six conditional odds ratios defined by Equations 5.27 and 5.28 have to be different from one. From Equation 5.29, the three conditional odds ratios defined by Equation 5.27 equal the three two-way interaction terms λ_{ij}. These three conditional odds ratios being different from one imply all $\lambda_{ij} \neq 0$, i.e., $\lambda_{12} \neq 0$, $\lambda_{13} \neq 0$ and $\lambda_{23} \neq 0$. Since the three conditional odds ratios defined by Equation 5.28 are also different from one, and all $\lambda_{ij} \neq 0$, λ_{ijk}, given by Equation 5.30, is also different from zero and uniquely determined. Therefore, there are no conditional independencies between the three manifest variables and the saturated model is obtained.

Note that, since both the latent class model and the marginal model of the manifest variables have zero degrees of freedom, parameters are uniquely determined. Indeed, Equations 5.27 and 5.30 establish the relationship between the parameters of the latent class model and the parameters of the saturated model of the manifest variables.

In conclusion: if the three binary manifest variables latent class model is sensible and holds, marginalising over the latent variable induces the saturated model for the manifest variables and a complete conditional independence graph.

The four manifest variables latent class model

In a latent class GLL model with four binary manifest variables there are fifteen knowns ($= 2^4 - 1$) and nine parameters to be estimated: λ_1, λ_2, λ_3, λ_4, λ_L, λ_{1L}, λ_{2L}, λ_{3L} and λ_{4L}. The number of degrees of freedom is positive and there is no guarantee that each of the λ can be uniquely determined as a function of the observed values (cell probabilities or conditional odds ratios between manifest variables). If the sensible latent class GLL model holds, all nine λ parameters have to be different from zero.

It is proposed that, if the four manifest variables latent class model is sensible and holds, marginalising over the latent variable L induces a complete independence graph

for the four manifest variables (with six edges: 12, 13, 14, 23, 24 and 34), implying no conditional independencies between them. There are now 24 conditional odds ratios between the four manifest variables, which can be expressed as a function of the λ terms of the log-linear expansion of the latent class model. As before, it is proved that all these conditional odds ratios are different from one, guaranteeing no conditional independencies between manifest variables.

Generally, if i, j, k and t are four binary manifest variables, and corner point constraints are used, the 24 conditional odds ratios can be obtained using the following four equations

$$\psi_{ij.k=0,\,t=0} = \frac{(1+\exp\{\lambda_L\})\,(1+\exp\{\lambda_L+\lambda_{iL}+\lambda_{jL}\})}{(1+\exp\{\lambda_L+\lambda_{iL}\})\,(1+\exp\{\lambda_L+\lambda_{jL}\})}, \tag{5.31}$$

$$\psi_{ij.k=0,\,t=1} = \frac{(1+\exp\{\lambda_L+\lambda_{tL}\})\,(1+\exp\{\lambda_L+\lambda_{iL}+\lambda_{jL}+\lambda_{tL}\})}{(1+\exp\{\lambda_L+\lambda_{iL}+\lambda_{tL}\})\,(1+\exp\{\lambda_L+\lambda_{jL}+\lambda_{tL}\})}, \tag{5.32}$$

$$\psi_{ij.k=1,\,t=0} = \frac{(1+\exp\{\lambda_L+\lambda_{kL}\})\,(1+\exp\{\lambda_L+\lambda_{iL}+\lambda_{jL}+\lambda_{kL}\})}{(1+\exp\{\lambda_L+\lambda_{iL}+\lambda_{kL}\})\,(1+\exp\{\lambda_L+\lambda_{jL}+\lambda_{kL}\})}, \tag{5.33}$$

$$\psi_{ij.k=1,\,t=1} = \frac{(1+\exp\{\lambda_L+\lambda_{kL}+\lambda_{tL}\})\,(1+\exp\{\lambda_L+\lambda_{iL}+\lambda_{jL}+\lambda_{kL}+\lambda_{tL}\})}{(1+\exp\{\lambda_L+\lambda_{iL}+\lambda_{kL}+\lambda_{tL}\})\,(1+\exp\{\lambda_L+\lambda_{jL}+\lambda_{kL}+\lambda_{tL}\})}. \tag{5.34}$$

Also,

$$\begin{aligned} \log(\psi_{ij.k=0,\,t=0}) &= \lambda_{ij} & \log(\psi_{ij.k=0,\,t=1}) &= \lambda_{ij}+\lambda_{ijt} \\ \log(\psi_{ij.k=1,\,t=0}) &= \lambda_{ij}+\lambda_{ijk} & \log(\psi_{ij.k=1,\,t=1}) &= \lambda_{ij}+\lambda_{ijk}+\lambda_{ijt}+\lambda_{ijkt}. \end{aligned} \tag{5.35}$$

Consequently, λ_{ijk} is still given by Equation 5.30, λ_{ijt} equals

$$\log\left[\frac{(1+\exp\{\lambda_L+\lambda_{iL}\})\,(1+\exp\{\lambda_L+\lambda_{jL}\})\,(1+\exp\{\lambda_L+\lambda_{tL}\})\,(1+\exp\{\lambda_L+\lambda_{iL}+\lambda_{jL}+\lambda_{tL}\})}{(1+\exp\{\lambda_L\})\,(1+\exp\{\lambda_L+\lambda_{iL}+\lambda_{jL}\})\,(1+\exp\{\lambda_L+\lambda_{iL}+\lambda_{tL}\})\,(1+\exp\{\lambda_L+\lambda_{jL}+\lambda_{tL}\})}\right]. \tag{5.36}$$

and

$$\lambda_{ijkt} = \log\left(\frac{\psi_{ij|k=0,\,t=0}\;\psi_{ij|k=1,\,t=1}}{\psi_{ij|k=0,\,t=1}\;\psi_{ij|k=1,\,t=0}}\right). \tag{5.37}$$

Since all cells are structurally non-empty, only sensible models are considered and all nine λ parameters are different from zero, all 24 conditional odds ratios defined by Equations 5.31 to 5.34 have to be different from one. From Equation 5.35, the six conditional odds ratios defined by Equation 5.31 equal λ_{ij}. These conditional odds ratios being different from one imply all six two-way interactions between the manifest

variables are different from zero, i.e., $\lambda_{12} \neq 0$, $\lambda_{13} \neq 0$, $\lambda_{14} \neq 0$, $\lambda_{23} \neq 0$, $\lambda_{24} \neq 0$ and $\lambda_{34} \neq 0$. Consequently, there are no conditional independencies between the four manifest variables, and a complete independence graph is obtained. Furthermore, the fact that the twelve conditional odds ratios defined by Equations 5.32 and 5.33 are different from one, and all λ terms are non-zero, guarantees that the four three-way interactions terms are different from zero, i.e., $\lambda_{123} \neq 0$, $\lambda_{124} \neq 0$, $\lambda_{134} \neq 0$ and $\lambda_{234} \neq 0$. Finally, the four-way interaction term, given by Equation 5.37, is also non zero: $\lambda_{1234} \neq 0$.

Nevertheless, these eleven interaction terms in the log-linear expansion of the model of the manifest variables are expressed as a function of just five λ terms in the log-linear expansion of the latent class model: λ_L, λ_{1L}, λ_{2L}, λ_{3L} and λ_{3L}. Indeed, Equations 5.31 to 5.37 establish the relationship between the parameters of the latent class model and the parameters of the model with no conditional independencies between the four manifest variables. It is not possible to obtain these five λ terms as a unique function of the interactions between the manifest variables (recall the latent class GLL model with four binary manifest variables has six degrees of freedom). Therefore, in this case marginalising over the latent variable implies no conditional independencies between manifest variables, but not necessarily the saturated model.

In conclusion: if the four binary manifest variables latent class model is sensible and holds, marginalising over the latent variable induces an independence graph for the manifest variables that is complete, with no conditional independencies between manifest variables.

One should note that in the latent class GLL model (with all variables binary) it is not possible to determine the conditions and combinations of parameters that lead to a unique solution for the model. Recall that, in the single-factor GG model, the fulfillment of the tetrad conditions by the scaled inverse variance matrix of the manifest variables guaranteed a unique solution (up to rotation). Due to the fact that, besides two-way interaction terms, higher order terms exist in the log-linear expansion of the latent class GLL model, the conditional log-odds ratios do not satisfy a tetrad structure. As mentioned by Cox and Wermuth (2002, page 464), *'The tetrad condition is a consequence of the linear structure underlying the Gaussian distribution ... and so it will not apply to binary variables'*.

Although the five and more manifest variables cases are not studied, it is possible to generalise the results of the current section in the following proposition: if the p binary manifest variables latent class model is sensible and holds, marginalising over the binary latent variable induces the complete independence graph for the manifest variables, with no conditional independencies between them. If $p = 3$ the induced model is the saturated model.

The results proposed so far require the sensible latent class GLL model to hold. In practice, even if there are no structurally empty cells, some of the observed cell probabilities may be very small and, consequently, one or more conditional odds ratios may be very close to one. That being the case, starting with a model with a complete independence graph between the manifest variables and performing backwards elimination, in order to detect the presence of a binary latent variable, may lead to excluding one or more edges. As in the Gaussian case, this suggests investigating the *power* of selecting a model consistent with a latent class GLL model. This was done in Section 4.2 of the thesis, where a theoretical normal approximation was proposed, as well as a non-central chi-square approximation for the two binary manifest variables case. The main conclusions drawn then, regarding the case of two binary manifest variables, are that power increases as the sample size increases and as the odds ratio deviates from one, faster for more balanced combinations of marginal probabilities. In the presence of three binary manifest variables it is recommended that the data analyst calculates the power of selecting the saturated model for the $2 \times 2 \times 2$ contingency table under analysis, as proposed in Section 4.2.4.

In conclusion: this chapter proposes parameterising the single-factor model as a graphical Gaussian model and the latent class model as a graphical log-linear model. Conditional independence graphs are used to represent both models, with associations between each of the manifest variables and the latent variable measured either by partial correlation coefficients (in the single-factor GG model) or by conditional log odds ratios (in the latent class GLL model). Detailed formulae are given to relate the proposed parameterisation to the classical parameterisation.

The main contributions of the chapter, regarding the single-factor GG model are

now summarised. The tetrad conditions are expressed in terms of the partial correlations and a new insight is given regarding admissible regions and compatible sign patterns for the partial correlation coefficients arising from a single-factor GG model. Some recommendations are given to the data analyst concerning detecting the presence of a Gaussian latent variable by inspecting the scaled inverse variance matrix of the manifest variables (rather then the variance or correlation matrix) and taking into account the power of selecting a model consistent with a single-factor model.

As far as the latent class GLL model is concerned, (conditional) odds ratios between manifest variables and, therefore, the independence structure of the latter, are related to the conditional log odds ratios between each of the manifest variables and the latent variable (given the remaining manifest variables are at level 1). After defining the concept of sensible latent class model, it is shown that if such a model holds, marginalising over the binary latent variable induces no conditional independencies between the binary manifest variables. In the case of four (or more) manifest variables the obtained model is not necessarily the saturated model.

Chapter 6

Conclusions

The aim of this last chapter is to summarise the main conclusions of the work undertaken in the research project and to highlight the main contributions of this thesis. Suggestions are given, regarding possible areas of further research. The work is developed within the framework defined by *graphical models*, in particular graphical Gaussian (GG) and graphical log-linear (GLL) models are considered. A review of the main concepts and definitions required for the understanding of the subsequent work is presented. Although such material was taken from existing literature, for consistency, it was required to establish an unifying notation.

The core of the research consists in investigating three main topics:

i) the distributions of the tests statistics for single edge exclusion from the saturated model, in particular under the alternative hypothesis that the saturated model holds;

ii) the power of the tests statistics for single edge exclusion from the saturated model;

iii) the parameterisation of the single-factor model as a graphical Gaussian model and of the latent class model as a graphical log-linear model (with all variables binary).

The main conclusions and contributions associated with each of this three topics are now presented in detail.

Distributions of the test statistics for single edge exclusion from the saturated model

Three test statistics for single edge exclusion from the saturated model are considered: the likelihood ratio, the Wald and the score test statistics. These test statistics, in particular their distributions under the null hypothesis of (conditional) independence, were investigated by Smith (1990). Extending Smith's work, this thesis investigates the distributions of the three test statistics under the alternative hypothesis that the saturated model holds. Non-signed and signed square-root (one-sided and two-sided) versions of the test statistics are considered.

Using the delta method, approximating normal distributions are derived. Smith (1990, pages 21 and 73) presented the asymptotic variance matrices of the m.l.e. of the ω (the elements of the inverse variance matrix, in GG models) and of the λ (the terms of the log-linear expansion of the cross-classified multinomial distribution density function, in GLL models). Because the test statistics for single edge exclusion from the saturated model can be written as functions of the $\hat{\omega}$, in GG models, and of the $\hat{\lambda}$, in GLL models, the delta method is used to derive approximating normal distributions to the three test statistics. As an intermediate step, it was required to express the test statistics for single edge exclusion from saturated GLL models as a function of the cell probabilities and of the m.l.e. of the λ terms in the log-linear expansion of f_{12} and of f_{123} (the density functions of cross-classified multinomial distributions of size one, respectively with two and three binary variables). Although Smith (1990) gave formulae for the LRT, the Wald and the score test statistics for a 2×2 contingency table, it was necessary to extend results to the $2 \times 2 \times 2$ case and to derive simplified formulae for the Wald and the score test statistics, both in terms of cell probabilities and $\hat{\lambda}$ terms. Formulae for the means, variances and covariances of the LRT, Wald and score test statistics, in the asymptotic distribution, are derived as a function of the partial correlation coefficients (for GG models, with p variables) and of the cell probabilities (for GLL models with two and three binary variables). All these formulae are of particular use when calculating power functions, as explained later.

In the two variables case, a non-central chi-square approximation to the distribution of the LRT statistic for single edge exclusion is also proposed, at a local alternative. In both GG and GLL models the non-centrality parameter is shown to be equal to

the expected value of the Wald test statistic in the corresponding asymptotic normal distribution, previously derived.

The quality of the proposed normal and non-central chi-square approximations to the distribution of the LRT statistic for single edge exclusion from a saturated model with two variables is assessed by simulation. Results show that the normal approximation holds asymptotically, i.e., for large sample sizes, and for values of the measure of association between the two variables, the (partial) correlation coefficient ρ_{12} in GG models and the log odds ratio $\log \psi_{12}$ in GLL models, not close to zero. The normal approximation is poor for small sample sizes and values of the measure of association close to independence. The non-central chi-square approximation performs better for values of ρ_{12} or $\log \psi_{12}$ close to zero, in particular if the sample size is not large. It becomes worse as the distance from the null increases. Since the distance from the null is measured as $\sqrt{n}\,(\theta - 0)$, where θ is the measure of association of interest (either ρ_{12} or $\log \psi_{12}$), as n increases the non-central chi-square approximation performs better for values of θ closer to zero. In GG models, 1 000 observations can be considered a large sample size, whereas in GLL models with two binary variables, a sample size of around 10 000 observations seems to be required for the normal approximation to hold.

Additionally, it is suggested that in GLL models the balance of the contingency table probabilities has also to be taken into account. The asymptotic normal and the non-central chi-square approximations to the distribution of the LRT statistic for single edge exclusion perform better for balanced combinations of marginal probabilities $\pi_1(0)$ and $\pi_2(0)$. A measure of balance in a 2×2 contingency table is proposed, relating to the information matrix. Some of its properties are analysed. The derived approximations to the distribution of the LRT statistic perform better for smaller values of the balance index. Yet, this measure is not fully developed; further research may include a better understanding of the relationship between the value of the balance index, the value of the odds ratio and the quality of the approximation and extending the measure to higher dimensional contingency tables.

Possible extensions of the work undertaken regarding the distributions of the test statistics for single edge exclusion from the saturated model include considering categorical variables with more than two categories, generalising the non-central chi-square approximation to the three variables case (possibly by using a non-central Wishart

distribution) and considering GLL models with more than three variables. This last task is a complex one because the notation becomes very messy, due to the number of parameters involved.

Power of the test statistics for single edge exclusion from the saturated model

The power of a backwards elimination model selection procedure for selecting the true (saturated) model is defined as the probability of selecting the true (saturated) model given the specified true model parameters. For a given size level (5%), the power of selecting the saturated model when using the likelihood ratio, the Wald and the score test statistics for single edge exclusion is investigated. Theoretical asymptotic power functions are derived, based on the asymptotic distributions of the test statistics for single edge exclusion previously obtained. Non-signed and signed square-root (one-sided and two-sided) versions of the test statistics are used.

Graphical Gaussian models are first considered. Starting with the simplest case of two variables, the power of selecting the saturated model using the LRT statistic is estimated by simulation and calculated using both the asymptotic normal and the non-central chi-square approximations to the distribution of the LRT statistic, already derived. The conclusions are that power increases as the absolute value of the correlation coefficient increases and as the sample size increases, being symmetric about zero correlation for two-sided hypothesis tests. For large sample sizes and values of ρ_{12} not close to zero, the asymptotic normal approximation to the power of the non-signed version of the LRT statistic performs well. The non-central chi-square approximation should be preferred for small sample sizes and correlation coefficients close to zero, i.e., small distances from the null. Asymptotic normal approximations to the power of selecting the saturated GG model with two variables, using the signed square-root versions of the three test statistics for single edge exclusion, are also derived. The quality of such approximations is good, even for small sample sizes and small correlation coefficients. Results confirm one-sided hypothesis tests are more powerful than corresponding two-sided tests.

The three Gaussian variables case is then analysed. The study of the power functions in the presence of three, or more, variables cases requires taking into account the

constraint that the scaled inverse variance matrix is positive definite. The shape of correlation matrices had been studied by Rousseeuw and Molenberghs (1994), who established a condition for positive definiteness in terms of correlation coefficients. Since the focus now is on partial correlations, such result is extended and a condition for positive definiteness is proposed, in terms of partial correlation coefficients.

A trivariate normal approximation to the power of the three hypothesis tests of excluding edge ij, edge ik and edge jk, for simplicity called the power of selecting the saturated model (in the three variables case) is derived. The approximation holds asymptotically, 1 000 observations being a large sample size. Yet, it is shown that a truncated normal approximation performs better for large sample sizes and values of the partial correlation coefficients close to zero. Power increases as partial correlations increase, although some non-symmetry and some non-monotonicity exist for small partial correlation values. In the case of the test of excluding edge ij from the saturated model, power just depends on the partial correlation coefficient of interest, $\rho_{ij.rest}$, whereas in the case of the two tests of excluding edge ij from the saturated model and excluding edge ik from the saturated model, as well as in the case of selecting the saturated model, power depends on the three partial correlation coefficients. Hence, it is possible to conclude that the correlation structure of the test statistics justifies some non-symmetry and non-monotonicity of the power functions. Asymptotic power functions for the signed square-root (one-sided and two-sided) versions of the test statistics are also derived using trivariate normal approximations. These hold even for sample sizes of 200 observations and partial correlation coefficients close to zero.

Results can be generalised to the cases of four or more variables, by using the multivariate normal distribution. However, the complexity of the problem increases considerably with the number of variables in the model: obtaining the power of selecting the saturated model in the five variables case requires calculating a ten dimensional integral, since there are ten test statistics for single edge exclusion. For this reason, in the four and the five variables cases, attention is restricted to models with equal partial correlations between all variables. Results for an equicorrelation matrix, existing in the literature, were adapted to establish the positive definiteness of a scaled inverse variance matrix with all off-diagonal elements (the negatives of the partial correlation coefficients) equal. Simulation results show that, for a certain value of $\rho_{ij.rest}$, the smaller the number of variables, the larger power is. Also, for large sample sizes,

differences in power values between models with three, four or five variables can only be detected for small $|\rho_{ij.rest}|$.

For graphical log-linear models, the power of the test statistics for single edge exclusion from the saturated model is analysed in detail for the two binary variables case and some insight is obtained regarding three binary variables. Although one parameter ($\rho_{ij.rest}$) is enough to measure the association between two normal variables, in the case of two or three binary variables, cross-classified in a contingency table, respectively three and seven parameters are required. The number of parameters to be considered increases substantially as the number of variables increases, and so does the complexity of the notation. For this reason only GLL models with two and three binary variables are used in this thesis.

An asymptotic normal and a non-central chi-square approximation are proposed to calculate the power of the LRT statistic for single edge exclusion from a saturated GLL model with two binary variables. Power increases as the sample size increases and as the value of the odds ratio deviates from one, faster for more balanced combinations of marginal probabilities. The non-central chi-square approximation performs better at a small distance from the null. For large sample sizes ($n_\emptyset = 10\,000$) and odds ratio values not close to one the normal approximation performs better, particularly for unbalanced contingency tables. When signed square-root versions of the test statistics are used, the proposed asymptotic normal approximation performs well, even for more moderate sample sizes ($n_\emptyset = 1\,000$), less balanced combinations of marginal probabilities and odds ratios close to one.

As far as the three binary variables case is concerned, it is suggested that, for a specific contingency table, asymptotic power functions can be obtained using a trivariate normal approximation. Since the parameter space is now seven dimensional, a comprehensive investigation of the power functions, for the different combinations of values of the seven parameters, is not possible.

When comparing the normal and the log-linear binary cases it is worth taking into account that, in both cases, power varies as a function of the sample size and of the degree of association between variables. However, in GLL models the balance of the contingency table has also to be considered, which suggests further investigating the proposed measure of balance, aiming at better quantifying its contribution to the variation in power. Extending results for the case of categorical variables with three

or more levels is also a possible area of further research.

The parameterisation of the single-factor model as a GG model and of the latent class model as a GLL model

The classical parameterisations of the single-factor model and of the latent class model are reviewed. The allowable region of the parameter space of the classical single-factor model is studied in detail for the three and four manifest variables cases and, since GG models require partial correlations, the tetrad conditions are expressed in terms of partial correlation coefficients.

The single-factor model is parameterised as a GG model, allowing GG models to incorporate a normal distributed latent variable. The classical and the proposed parameterisations are related. It is suggested that the single-factor GG model is represented by a conditional independence graph, each edge representing the partial correlation between one manifest variable and the latent variable, given the remaining manifest variables.

The conditional independence structure between manifest variables, arising from marginalising the single-factor GG model over the latent variable, is investigated. It is proved that marginalising over the latent variable in a single-factor GG model induces no conditional independencies between manifest variables and, consequently, an independence graph that is complete. In the three manifest variables case, the saturated model is obtained, whereas when four or more manifest variables are present a complete independence graph does not necessarily correspond to the saturated model. Indeed, there are constraints on the partial correlation coefficients between manifest variables compatible with a single-factor model. In the three variables case, the scaled inverse variance matrix is constrained to be positive definite. It is proved, however, that when there are four or more manifest variables, in order to have a model consistent with a single-factor model, the four or more partial correlation coefficients also have to satisfy the tetrad conditions. Hence, the conditions that guarantee that a scaled inverse variance matrix is compatible with a model arising from a single-factor model are derived. These conditions imply certain patterns of signs for the partial correlation coefficients. A rule is proposed for defining the patterns of signs for the partial correlations between manifest variables that are compatible with a single-factor GG model.

The power of selecting a model consistent with a single-factor GG model is derived, based on the asymptotic normal approximation previously proposed, with the additional constraint that the conditions that guarantee the scaled inverse variance matrix of the manifest variables is compatible with a single-factor GG model are satisfied. Some recommendations are then given to the data analyst, regarding detecting the presence of a Gaussian latent variable, i.e., identifying a model consistent with a single-factor GG model. In brief, consider partial correlation coefficients and the compatibility of their sign patterns, choose the test statistic for single edge exclusion to be used and calculate the power of selecting the saturated model. This can be done prior to collecting data, based on values the data analyst considers plausible for the partial correlation coefficients, or after data collection, treating sample partial correlations as if they were the 'true' population values. Test for each edge exclusion between manifest variables. When power values are not high, the data analyst must be prepared to fit a single-factor GG model even if there are conditional independencies between manifest variables and the conditional independence graph is not complete.

The latent class model is parameterised as a GLL model, allowing GLL models to include a binary latent variable. The single-factor GLL model is represented by a conditional independence graph, each edge representing the conditional log odds ratio between each of the manifest variables and the latent variable, given that the remaining manifest variables are at level 1. The concept of sensible model is defined, ensuring all cell probabilities, in all possible cross-tabulations between manifest variables and latent variable, are positive.

The conditional independence structure between manifest variables arising from a latent class GLL model is investigated and it is shown that, if the sensible latent class model holds, marginalising over the latent variable implies all conditional odds ratios between manifest variables different from one and, consequently, no conditional independencies and a complete independence graph between manifest variables. When four, or more, binary manifest variables are present, the fact that all interaction terms (two-way and higher order) in the log-linear expansion of the manifest variables are different from zero does not guarantee the obtained model is the saturated model. The justification being that it is not possible to obtain a one to one relationship between the λ terms of the log-linear expansion of the latent class GLL model and the interaction

terms between manifest variables. Recall that, in the Gaussian case, it is possible to establish a one to one relationship between the parameters of the single-factor GG model and the partial correlations between manifest variables, provided the tetrad conditions are fulfilled. As in the Gaussian case, it is recommended that the data analyst takes into account the power of selecting the saturated model when trying to detect the presence of a binary latent variable.

Possible areas of further research include investigating the parameterisation of models that relax the assumption of local independence, i.e., models in which the latent variable does not account for all associations among manifest variables, as well as the parameterisation of models with more than one latent variable.

A more challenging area of further research is to consider *mixed models* and *conditional-Gaussian distributions* and try to extend the main results of the thesis to this type of models. As far as latent variable models are concerned, this would allow models parallel to latent trait and latent profile analysis to be included in the unified framework of graphical models.

Bibliography

Agresti, A. (1996), *An Introduction to Categorical Data Analysis*, John Wiley & Sons.

Anderson, J. and Gerbing, D. (1984), 'The effect of sampling error on convergence, improper solutions and goodness-of-fit indices for maximum likelihood confirmatory factor analysis', *Psychometrika* **49**(2), 155–173.

Anderson, T. and Rubin, H. (1956), 'Statistical inference in factor analysis', *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* **V**, 111–150.

Bartholomew, D. and Knott, M. (1999), *Latent Variable Models and Factor Analysis*, Kendall's Library of Statistics - 7, second edn, Arnold Publishers.

Bartholomew, D., Steele, F., Moustaki, I. and Galbraith, J. (2002), *The analysis and interpretation of multivariate data for social scientists*, Chapman & Hall.

Birch, M. (1963), 'Maximum likelihood in three-way contingency tables', *Journal of the Royal Statistical Society. Series B* **25**(1), 220–233.

Birch, M. (1964), 'The detection of partial association, I: The 2 × 2 case', *Journal of the Royal Statistical Society. Series B* **26**(2), 313–324.

Birch, M. (1965), 'The detection of partial association, II: The general case', *Journal of the Royal Statistical Society. Series B* **27**(1), 111–124.

Bishop, Y. (1969), 'Full contingency tables, logits, and split contingency tables', *Biometrics* **25**(2), 383–399.

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

Blalock, H. (1971), *Causal Models in the Social Sciences*, Aldine - Atheston, Chicago.

Bollen, K. (1987), 'Outliers and improper solutions - a confirmatory factor analysis example', *Sociological Methods and Research* **15**(4), 375–384.

Bollen, K. (1989), *Structural Equations with Latent Variables*, John Wiley & Sons.

Boomsma, A. (1985), 'Nonconvergence, improper solutions and starting values in LISREL maximum likelihood estimation', *Psychometrika* **50**(2), 229–242.

Boomsma, A. and Hoogland, J. (2001), *The robustness of LISREL modeling revisited*, Vol. Structural Equation Modeling: present and future, Cudeck, Du Toit and Sörbom (ed.) - Scientific Software International, chapter 8, pp. 139–168.

Buse, A. (1982), 'The likelihood ratio, Wald and Lagrange multiplier tests: an expository note', *The American Statistician* **36**(3), 153–157.

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

Cox, D. and Wermuth, N. (1990), 'An approximation to maximum likelihood estimates in reduced models', *Biometrika* **77**(4), 747–761.

Cox, D. and Wermuth, N. (1993), 'Linear dependencies represented by chain graphs (with discussion)', *Statistical Science* **8**(3), 204–283.

Cox, D. and Wermuth, N. (1996), *Multivariate Dependencies - Models, Analysis and Interpretation*, Monographs on Statistics and Applied Probability - 67, Chapman & Hall.

Cox, D. and Wermuth, N. (2002), 'On some models for multivariate binary variables parallel in complexity with the multivariate gaussian distribution', *Biometrika* **89**(2), 462–469.

Darroch, J., Lauritzen, S. and Speed, T. (1980), 'Markov fields and log-linear interaction models for contingency tables', *Annals of Statistics* **8**(3), 522–539.

Davison, A., Smith, P. and Whittaker, J. (1991), 'An exact conditional test for covariance selection models', *Australian Journal of Statistics* **33**, 313–318.

Dawid, A. (1979), 'Conditional independence in statistical theory (with discussion)', *Journal of the Royal Statistical Society. Series B* **41**(1), 1–31.

Dawid, A. (1980), 'Conditional independence for statistical operations', *Annals of Statistics* **8**(3), 598–617.

Dempster, A. (1972), 'Covariance selection', *Biometrics* **28**(1), 157–175.

Dempster, A., Laird, N. and Rubin, D. (1977), 'Maximum likelihood from incomplete data vie the *EM* algorithm', *Journal of the Royal Statistical Society. Series B* **39**(1), 1–38.

Didelez, V. (1999), 'Local independence graphs for composable Markov processes', *Discussion paper 158* (SFB 386 - University of Munich).

Didelez, V. (2000), Graphical Models for Event History Data based on Local Independence, PhD thesis, Logos, Berlin.

Dijkstra, T. (1992), 'On statistical inference with parameter estimates on the boundary of the parameter space', *British Journal of Mathematical and Statistical Psychology* **45**, 289–309.

Duncan, O. (1975), *Introduction to Structural Equation Models*, New York: Academic Press.

Edwards, D. (2000), *Introduction to Graphical Modelling*, second edn, Springer-Verlag.

Edwards, D. and Havánek, T. (1985), 'A fast procedure for model search in multidimensional contingency tables', *Biometrika* **72**(2), 339–351.

Edwards, D. and Havánek, T. (1987), 'A fast model selection procedure for large families of models', *Journal of the American Statistical Association* **82**(397), 205–213.

Edwards, D. and Kreiner, S. (1983), 'The analysis of contingency tables by graphical models', *Biometrika* **70**(3), 553–565.

Ekholm, A., Smith, P. and McDonald, J. (1995), 'Marginal regression analysis of a multivariate binary response', *Biometrika* **82**(4), 847–854.

Ferguson, T. (1996), *A Course in Large Sample Theory*, Chapman & Hall.

Frydenberg, M. (1990), 'Marginalization and collapsibility in graphical interaction models', *Annals of Statistics* **18**(2), 790–805.

Frydenberg, M. and Lauritzen, S. (1989), 'Decomposition of maximum likelihood in mixed graphical interaction models', *Biometrika* **76**(3), 539–555.

Gabriel, K. (1969), 'Simultaneous test procedures - some theory of multiple comparisons', *Annals of Mathematical Statistics* **40**, 224–250.

Giudici, P. and Stanghellini, E. (2001), 'Bayesian inference for graphical factor analysis models', *Psychometrika* **66**(4), 577–592.

Goodman, L. (1969), 'On partitioning χ^2 and detecting partial association in three-way contingency tables', *Journal of the Royal Statistical Society. Series B* **31**(3), 486–498.

Goodman, L. (1970), 'The multivariate analysis of qualitative data: Interactions among multiple classifications', *Journal of the American Statistical Association* **65**(329), 226–256.

Goodman, L. (1971), 'Partitioning of chi-square, analysis of marginal contingency tables and estimation of expected frequencies in multidimensional contingency tables', *Journal of the American Statistical Association* **66**(334), 339–344.

Goodman, L. (1974), 'Exploratory latent structure analysis using both identifiable and unidentifiable models', *Biometrika* **61**(2), 215–231.

Haberman, S. (1974), *The Analysis of Frequency Data*, University of Chicago Press, Chicago.

Haberman, S. (1979), *Analysis of Qualitative Data*, Vol. 2, New York: Academic Press, chapter 10, pp. 541–569.

Harman, H. (1967), *Modern Factor Analysis*, second edn, The University of Chicago Press.

Heinen, T. (1996), *Latent Class and Discrete Latent Trait Models*, Advanced Quantitative Techniques in the Social Sciences - 6, SAGE Publications.

Hodapp, V. and Wermuth, N. (1983), 'Decomposable models: a new look at interdependence and dependence structures in psychological research', *Multivariate Behavioral Research* **18**, 361–390.

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**(1/2), 134–139.

Jöreskog, K. (1967), 'Some contributions to maximum likelihood factor analysis', *Psychometrika* **32**(4), 443–482.

Kiiveri, H. and Speed, T. (1982), *Structural Analysis of Multivariate Data: a Review*, Vol. Sociological Methodology, Samuel Leinhardt (ed.), chapter 6, pp. 209–289.

Koster, J. (1996), 'Markov properties of nonrecursive causal models', *The Annals of Statistics* **24**(5), 2148–2177.

Lauritzen, S. (1989), 'Mixed graphical association models (with discussion)', *Scandinavian Journal of Statistics* **16**, 273–306.

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

Lauritzen, S. (2001), *Some Modern Applications of Graphical Models*. Research Report R-01-2015, Department of Mathematical Sciences, Aalborg University.

Lauritzen, S. and Spiegelhalter, D. (1988), 'Local computations with probabilities on graphical structures and their application to expert systems', *Journal of the Royal Statistical Society. Series B* **50**(2), 157–224.

Lauritzen, S. and Wermuth, N. (1989), 'Graphical models for associations between variables, some of which are qualitative and some quantitative', *Annals of Statistics* **17**(1), 31–57.

Lazarsfeld, P. and Henry, N. (1968), *Latent Structure Analysis*, Boston: Houghton Mifflin Company.

Mardia, K., Kent, J. and Bibby, J. (1979), *Multivariate Analysis*, Academic Press.

Moustaki, I. (1996), 'A latent trait and a latent class model for mixed observed variables', *British Journal of Mathematical and Statistical Psychology* **49**, 313–334.

Porteous, B. (1985), 'Improved likelihood ratio statistics for covariance selection models', *Biometrika* **72**(1), 97–101.

Rousseeuw, P. and Molenberghs, G. (1994), 'The shape of correlation matrices', *The American Statistician* **48**(4), 276–279.

Roverato, A. and Whittaker, J. (1998), 'The isserlis matrix and its applications to non-decomposable graphical gaussian models', *Biometrika* **85**(3), 711–725.

Salgueiro, M., McDonald, J. and Smith, P. (2001), *The Observed Association Structure from Graphical Gaussian Models with a Single Latent Variable*, Vol. New Trends in Statistical Modelling - Proceedings of the 16th International Workshop on Statistical Modelling, Klein, B. and Korsholm, L. (ed.), pp. 347–354.

Sen, P. and Singer, J. (1993), *Large Sample Methods in Statistics - An Introduction with Applications*, Chapman & Hall.

Severini, T. (2000), *Likelihood Methods in Statistics*, Oxford University Press.

Smith, P. (1990), Edge Exclusion and Model Selection in Graphical Models, PhD thesis, Lancaster University.

Smith, P. (1992), *Assessing the Power of Model Selection Procedures used when Graphical Modelling*, Vol. Proceedings of COMSTAT, vol. 1, Dodge, Y. and Whittaker, J. - Heidelberg: Psysica-Verlag, pp. 275–280.

Smith, P. and Whittaker, J. (1998), *Learning and Inference in Graphical Models*, M.I. Jordan (ed.) - Dordrecht: Kluwer Academic Press, chapter Edge exclusion tests for Graphical Gaussian models, pp. 555–574.

Spearman, C. (1904), "General Intelligence' objectivelly determined and measured', *American Journal of Psychology* **15**, 201–293.

Speed, T. and Kiiveri, H. (1986), 'Gaussian Markov distributions over finite graphs', *Annals of Statistics* **14**(1), 138–150.

Spiegelhalter, D., Dawid, A., Lauritzen, S. and Cowell, R. (1993), 'Bayesian analysis in expert systems (with discussion)', *Statistical Science* **8**(3), 219–283.

Stanghellini, E. (1997), 'Identification of a single-factor model using graphical gaussian rules', *Biometrika* **84**(1), 241–244.

van Driel, O. (1978), 'On various causes of improper solutions in maximum likelihood factor analysis', *Psychometrika* **43**(2), 225–243.

Vicard, P. (2000), 'On the identification of a single-factor model with correlated residuals', *Biometrika* **87**(1), 199–205.

Wermuth, N. (1976*a*), 'Analogies between multiplicative models in contingency tables and covariance selection', *Biometrics* **32**, 95–108.

Wermuth, N. (1976*b*), 'Model search among multiplicative models', *Biometrics* **32**, 253–263.

Wermuth, N. (1980), 'Linear recursive equations, covariance selection and path analysis', *Journal of the American Statistical Association* **75**(372), 963–972.

Wermuth, N. (1985), *Data Analysis and Conditional Independence Structures*, Vol. 1 - Proceedings 45th Session - invited papers, Bulletin of the International Statistical Institute, pp. 1–13.

Wermuth, N. (1993), *Association Structures with Few Variables: Characteristics and Examples*, Vol. Population Health Research, SAGE Publications, chapter 9, pp. 181–203.

Wermuth, N. and Cox, D. (1992), *Graphical Models for Dependencies and Associations*, Vol. 1 - Computational Statistics, Y. Dodge & J. Whittaker (eds.) - Heidelberg: Psysica, pp. 235–249.

Wermuth, N. and Lauritzen, S. (1983), 'Graphical and recursive models for contingency tables', *Biometrika* **70**(3), 537–552.

Wermuth, N. and Lauritzen, S. (1990), 'On substantive research hypotheses, conditional independence graphs and graphical chain models', *Journal of the Royal Statistical Society. Series B* **52**(1), 21–50.

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

Whittaker, J. (1993), *Graphical Interaction Models: a New Approach for Statistical Modelling*, Vol. Population Health Research, SAGE Publications, chapter 8, pp. 160–180.

Wold, H. (1954), 'Causality and econometrics', *Econometrica* **22**(2), 162–177.

Wright, S. (1934), 'The method of path coefficients', *Annals of Mathematical Statistics* **5**(3), 161–215.

Appendix A

Partial Derivatives Used for Calculating the Variance of the Likelihood Ratio Test Statistic

$\frac{\partial\lambda_\emptyset}{\partial\lambda_1} = -\pi_1(1)$	$\frac{\partial\lambda_\emptyset}{\partial\lambda_2} = -\pi_2(1)$	$\frac{\partial\lambda_\emptyset}{\partial\lambda_{12}} = -\pi(1,1)$
$\frac{\partial\pi(0,0)}{\partial\lambda_1} = -\pi(0,0)\pi_1(1)$	$\frac{\partial\pi(0,0)}{\partial\lambda_2} = -\pi(0,0)\pi_2(1)$	$\frac{\partial\pi(0,0)}{\partial\lambda_{12}} = -\pi(0,0)\pi(1,1)$
$\frac{\partial\pi(0,1)}{\partial\lambda_1} = -\pi(0,1)\pi_1(1)$	$\frac{\partial\pi(0,1)}{\partial\lambda_2} = -\pi(0,1)(\pi_2(1)-1)$	$\frac{\partial\pi(0,1)}{\partial\lambda_{12}} = -\pi(0,1)\pi(1,1)$
$\frac{\partial\pi(1,0)}{\partial\lambda_1} = -\pi(1,0)(\pi_1(1)-1)$	$\frac{\partial\pi(1,0)}{\partial\lambda_2} = -\pi(1,0)\pi_2(1)$	$\frac{\partial\pi(1,0)}{\partial\lambda_{12}} = -\pi(1,0)\pi(1,1)$
$\frac{\partial\pi(1,1)}{\partial\lambda_1} = -\pi(1,1)(\pi_1(1)-1)$	$\frac{\partial\pi(1,1)}{\partial\lambda_2} = -\pi(1,1)(\pi_2(1)-1)$	$\frac{\partial\pi(1,1)}{\partial\lambda_{12}} = -\pi(1,1)(\pi(1,1)-1)$

Table A.1: Partial derivatives of the cell probabilities and of $\lambda_\emptyset$ with respect to the three λ.

$\frac{\partial \log\left(\frac{\pi(0,0)}{\pi_1(0)\pi_2(0)}\right)}{\partial \lambda_1} = \frac{\partial \log\left(\frac{\pi(1,0)}{\pi_1(1)\pi_2(0)}\right)}{\partial \lambda_1} = \pi_1(1) - \frac{\pi(1,0)}{\pi_2(0)}$	
$\frac{\partial \log\left(\frac{\pi(0,1)}{\pi_1(0)\pi_2(1)}\right)}{\partial \lambda_1} = \frac{\partial \log\left(\frac{\pi(1,1)}{\pi_1(1)\pi_2(1)}\right)}{\partial \lambda_1} = \pi_1(1) - \frac{\pi(1,1)}{\pi_2(1)}$	
$\frac{\partial \log\left(\frac{\pi(0,0)}{\pi_1(0)\pi_2(0)}\right)}{\partial \lambda_2} = \frac{\partial \log\left(\frac{\pi(0,1)}{\pi_1(0)\pi_2(1)}\right)}{\partial \lambda_2} = \pi_2(1) - \frac{\pi(0,1)}{\pi_1(0)}$	
$\frac{\partial \log\left(\frac{\pi(1,0)}{\pi_1(1)\pi_2(0)}\right)}{\partial \lambda_2} = \frac{\partial \log\left(\frac{\pi(1,1)}{\pi_1(1)\pi_2(1)}\right)}{\partial \lambda_2} = \pi_2(1) - \frac{\pi(1,1)}{\pi_1(1)}$	
$\frac{\partial \log\left(\frac{\pi(0,0)}{\pi_1(0)\pi_2(0)}\right)}{\partial \lambda_{12}} = \pi(1,1)$	$\frac{\partial \log\left(\frac{\pi(0,1)}{\pi_1(0)\pi_2(1)}\right)}{\partial \lambda_{12}} = \pi(1,1) - \frac{\pi(1,1)}{\pi_2(1)}$
$\frac{\partial \log\left(\frac{\pi(1,0)}{\pi_1(1)\pi_2(0)}\right)}{\partial \lambda_{12}} = \pi(1,1) - \frac{\pi(1,1)}{\pi_1(1)}$	$\frac{\partial \log\left(\frac{\pi(1,1)}{\pi_1(1)\pi_2(1)}\right)}{\partial \lambda_{12}} = 1 + \pi(1,1) - \frac{\pi(1,1)}{\pi_1(1)} - \frac{\pi(1,1)}{\pi_2(1)}$

Table A.2: Partial derivatives of the logarithmic terms with respect to the three λ.

$\frac{\partial \lambda_\emptyset}{\partial \lambda_1} = -\pi_1(1)$	$\frac{\partial \lambda_\emptyset}{\partial \lambda_2} = -\pi_2(1)$	$\frac{\partial \lambda_\emptyset}{\partial \lambda_{12}} = -\pi_{12}(1,1)$	$\frac{\partial \lambda_\emptyset}{\partial \lambda_3} = -\pi_3(1)$
$\frac{\partial \lambda_\emptyset}{\partial \lambda_{13}} = -\pi_{13}(1,1)$	$\frac{\partial \lambda_\emptyset}{\partial \lambda_{23}} = -\pi_{23}(1,1)$	$\frac{\partial \lambda_\emptyset}{\partial \lambda_{123}} = -\pi(1,1,1)$	

Table A.3: Partial derivatives of $\lambda_\emptyset$ with respect to the seven λ.

$\frac{\partial\pi(0,0,0)}{\partial\lambda_1}$	$= -\pi(0,0,0)\pi_1(1)$	$\frac{\partial\pi(0,1,0)}{\partial\lambda_1}$	$= -\pi(0,1,0)\pi_1(1)$	$\frac{\partial\pi(1,0,0)}{\partial\lambda_1}$	$= -\pi(1,0,0)(\pi_1(1)-1)$
$\frac{\partial\pi(1,1,0)}{\partial\lambda_1}$	$= -\pi(1,1,0)(\pi_1(1)-1)$	$\frac{\partial\pi(0,0,1)}{\partial\lambda_1}$	$= -\pi(0,0,1)\pi_1(1)$	$\frac{\partial\pi(0,1,1)}{\partial\lambda_1}$	$= -\pi(0,1,1)\pi_1(1)$
$\frac{\partial\pi(1,0,1)}{\partial\lambda_1}$	$= -\pi(1,0,1)(\pi_1(1)-1)$	$\frac{\partial\pi(1,1,1)}{\partial\lambda_1}$	$= -\pi(1,1,1)(\pi_1(1)-1)$		
$\frac{\partial\pi(0,0,0)}{\partial\lambda_2}$	$= -\pi(0,0,0)\pi_2(1)$	$\frac{\partial\pi(0,1,0)}{\partial\lambda_2}$	$= -\pi(0,1,0)(\pi_2(1)-1)$	$\frac{\partial\pi(1,0,0)}{\partial\lambda_2}$	$= -\pi(1,0,0)\pi_2(1)$
$\frac{\partial\pi(1,1,0)}{\partial\lambda_2}$	$= -\pi(1,1,0)(\pi_2(1)-1)$	$\frac{\partial\pi(0,0,1)}{\partial\lambda_2}$	$= -\pi(0,0,1)\pi_2(1)$	$\frac{\partial\pi(0,1,1)}{\partial\lambda_2}$	$= -\pi(0,1,1)(\pi_2(1)-1)$
$\frac{\partial\pi(1,0,1)}{\partial\lambda_2}$	$= -\pi(1,0,1)\pi_2(1)$	$\frac{\partial\pi(1,1,1)}{\partial\lambda_2}$	$= -\pi(1,1,1)(\pi_2(1)-1)$		
$\frac{\partial\pi(0,0,0)}{\partial\lambda_3}$	$= -\pi(0,0,0)\pi_3(1)$	$\frac{\partial\pi(0,1,0)}{\partial\lambda_3}$	$= -\pi(0,1,0)\pi_3(1)$	$\frac{\partial\pi(1,0,0)}{\partial\lambda_3}$	$= -\pi(1,0,0)\pi_3(1)$
$\frac{\partial\pi(1,1,0)}{\partial\lambda_3}$	$= -\pi(1,1,0)\pi_3(1)$	$\frac{\partial\pi(0,0,1)}{\partial\lambda_3}$	$= -\pi(0,0,1)(\pi_3(1)-1)$	$\frac{\partial\pi(0,1,1)}{\partial\lambda_3}$	$= -\pi(0,1,1)(\pi_3(1)-1)$
$\frac{\partial\pi(1,0,1)}{\partial\lambda_3}$	$= -\pi(1,0,1)(\pi_3(1)-1)$	$\frac{\partial\pi(1,1,1)}{\partial\lambda_3}$	$= -\pi(1,1,1)(\pi_3(1)-1)$		
$\frac{\partial\pi(0,0,0)}{\partial\lambda_{12}}$	$= -\pi(0,0,0)\pi_{12}(1,1)$	$\frac{\partial\pi(0,1,0)}{\partial\lambda_{12}}$	$= -\pi(0,1,0)\pi_{12}(1,1)$	$\frac{\partial\pi(1,0,0)}{\partial\lambda_{12}}$	$= -\pi(1,0,0)\pi_{12}(1,1)$
$\frac{\partial\pi(1,1,0)}{\partial\lambda_{12}}$	$= -\pi(1,1,0)(\pi_{12}(1,1)-1)$	$\frac{\partial\pi(0,0,1)}{\partial\lambda_{12}}$	$= -\pi(0,0,1)\pi_{12}(1,1)$	$\frac{\partial\pi(0,1,1)}{\partial\lambda_{12}}$	$= -\pi(0,1,1)\pi_{12}(1,1)$
$\frac{\partial\pi(1,0,1)}{\partial\lambda_{12}}$	$= -\pi(1,0,1)\pi_{12}(1,1)$	$\frac{\partial\pi(1,1,1)}{\partial\lambda_{12}}$	$= -\pi(1,1,1)(\pi_{12}(1,1)-1)$		
$\frac{\partial\pi(0,0,0)}{\partial\lambda_{13}}$	$= -\pi(0,0,0)\pi_{13}(1,1)$	$\frac{\partial\pi(0,1,0)}{\partial\lambda_{13}}$	$= -\pi(0,1,0)\pi_{13}(1,1)$	$\frac{\partial\pi(1,0,0)}{\partial\lambda_{13}}$	$= -\pi(1,0,0)\pi_{13}(1,1)$
$\frac{\partial\pi(1,1,0)}{\partial\lambda_{13}}$	$= -\pi(1,1,0)\pi_{13}(1,1)$	$\frac{\partial\pi(0,0,1)}{\partial\lambda_{13}}$	$= -\pi(0,0,1)\pi_{13}(1,1)$	$\frac{\partial\pi(0,1,1)}{\partial\lambda_{13}}$	$= -\pi(0,1,1)\pi_{13}(1,1)$
$\frac{\partial\pi(1,0,1)}{\partial\lambda_{13}}$	$= -\pi(1,0,1)(\pi_{13}(1,1)-1)$	$\frac{\partial\pi(1,1,1)}{\partial\lambda_{13}}$	$= -\pi(1,1,1)(\pi_{13}(1,1)-1)$		
$\frac{\partial\pi(0,0,0)}{\partial\lambda_{23}}$	$= -\pi(0,0,0)\pi_{23}(1,1)$	$\frac{\partial\pi(0,1,0)}{\partial\lambda_{23}}$	$= -\pi(0,1,0)\pi_{23}(1,1)$	$\frac{\partial\pi(1,0,0)}{\partial\lambda_{23}}$	$= -\pi(1,0,0)\pi_{23}(1,1)$
$\frac{\partial\pi(1,1,0)}{\partial\lambda_{23}}$	$= -\pi(1,1,0)\pi_{23}(1,1)$	$\frac{\partial\pi(0,0,1)}{\partial\lambda_{23}}$	$= -\pi(0,0,1)\pi_{23}(1,1)$	$\frac{\partial\pi(0,1,1)}{\partial\lambda_{23}}$	$= -\pi(0,1,1)(\pi_{23}(1,1)-1)$
$\frac{\partial\pi(1,0,1)}{\partial\lambda_{23}}$	$= -\pi(1,0,1)\pi_{23}(1,1)$	$\frac{\partial\pi(1,1,1)}{\partial\lambda_{23}}$	$= -\pi(1,1,1)(\pi_{23}(1,1)-1)$		
$\frac{\partial\pi(0,0,0)}{\partial\lambda_{123}}$	$= -\pi(0,0,0)\pi(1,1,1)$	$\frac{\partial\pi(0,1,0)}{\partial\lambda_{123}}$	$= -\pi(0,1,0)\pi(1,1,1)$	$\frac{\partial\pi(1,0,0)}{\partial\lambda_{123}}$	$= -\pi(1,0,0)\pi(1,1,1)$
$\frac{\partial\pi(1,1,0)}{\partial\lambda_{123}}$	$= -\pi(1,1,0)\pi(1,1,1)$	$\frac{\partial\pi(0,0,1)}{\partial\lambda_{123}}$	$= -\pi(0,0,1)\pi(1,1,1)$	$\frac{\partial\pi(0,1,1)}{\partial\lambda_{123}}$	$= -\pi(0,1,1)\pi(1,1,1)$
$\frac{\partial\pi(1,0,1)}{\partial\lambda_{123}}$	$= -\pi(1,0,1)\pi(1,1,1)$	$\frac{\partial\pi(1,1,1)}{\partial\lambda_{123}}$	$= -\pi(1,1,1)(\pi(1,1,1)-1)$		

Table A.4: Partial derivatives of the cell probabilities with respect to the seven λ.

$$\frac{\partial \log\left(\frac{\pi(0,0,0)\pi_3(0)}{\pi_{13}(0,0)\pi_{23}(0,0)}\right)}{\partial \lambda_1} = \frac{\partial \log\left(\frac{\pi(1,0,0)\pi_3(0)}{\pi_{13}(1,0)\pi_{23}(0,0)}\right)}{\partial \lambda_1} = \frac{\pi_{13}(1,0)}{\pi_3(0)} - \frac{\pi(1,0,0)}{\pi_{23}(0,0)}$$

$$\frac{\partial \log\left(\frac{\pi(0,1,0)\pi_3(0)}{\pi_{13}(0,0)\pi_{23}(1,0)}\right)}{\partial \lambda_1} = \frac{\partial \log\left(\frac{\pi(1,1,0)\pi_3(0)}{\pi_{13}(1,0)\pi_{23}(1,0)}\right)}{\partial \lambda_1} = \frac{\pi_{13}(1,0)}{\pi_3(0)} - \frac{\pi(1,1,0)}{\pi_{23}(1,0)}$$

$$\frac{\partial \log\left(\frac{\pi(0,0,1)\pi_3(1)}{\pi_{13}(0,1)\pi_{23}(0,1)}\right)}{\partial \lambda_1} = \frac{\partial \log\left(\frac{\pi(1,0,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(0,1)}\right)}{\partial \lambda_1} = \frac{\partial \log\left(\frac{\pi(0,0,1)\pi_3(1)}{\pi_{13}(0,1)\pi_{23}(0,1)}\right)}{\partial \lambda_{13}} = \frac{\partial \log\left(\frac{\pi(1,0,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(0,1)}\right)}{\partial \lambda_{13}} = \frac{\pi_{13}(1,1)}{\pi_3(1)} - \frac{\pi(1,0,1)}{\pi_{23}(0,1)}$$

$$\frac{\partial \log\left(\frac{\pi(0,1,1)\pi_3(1)}{\pi_{13}(0,1)\pi_{23}(1,1)}\right)}{\partial \lambda_1} = \frac{\partial \log\left(\frac{\pi(1,1,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(1,1)}\right)}{\partial \lambda_1} = \frac{\partial \log\left(\frac{\pi(0,1,1)\pi_3(1)}{\pi_{13}(0,1)\pi_{23}(1,1)}\right)}{\partial \lambda_{13}} = \frac{\partial \log\left(\frac{\pi(1,1,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(1,1)}\right)}{\partial \lambda_{13}} = \frac{\pi_{13}(1,1)}{\pi_3(1)} - \frac{\pi(1,1,1)}{\pi_{23}(1,1)}$$

$$\frac{\partial \log\left(\frac{\pi(0,0,0)\pi_3(0)}{\pi_{13}(0,0)\pi_{23}(0,0)}\right)}{\partial \lambda_2} = \frac{\partial \log\left(\frac{\pi(0,1,0)\pi_3(0)}{\pi_{13}(0,0)\pi_{23}(1,0)}\right)}{\partial \lambda_2} = \frac{\pi_{23}(1,0)}{\pi_3(0)} - \frac{\pi(0,1,0)}{\pi_{13}(0,0)}$$

$$\frac{\partial \log\left(\frac{\pi(1,0,0)\pi_3(0)}{\pi_{13}(1,0)\pi_{23}(0,0)}\right)}{\partial \lambda_2} = \frac{\partial \log\left(\frac{\pi(1,1,0)\pi_3(0)}{\pi_{13}(1,0)\pi_{23}(1,0)}\right)}{\partial \lambda_2} = \frac{\pi_{23}(1,0)}{\pi_3(0)} - \frac{\pi(1,1,0)}{\pi_{13}(1,0)}$$

$$\frac{\partial \log\left(\frac{\pi(0,0,1)\pi_3(1)}{\pi_{13}(0,1)\pi_{23}(0,1)}\right)}{\partial \lambda_2} = \frac{\partial \log\left(\frac{\pi(0,1,1)\pi_3(1)}{\pi_{13}(0,1)\pi_{23}(1,1)}\right)}{\partial \lambda_2} = \frac{\partial \log\left(\frac{\pi(0,0,1)\pi_3(1)}{\pi_{13}(0,1)\pi_{23}(0,1)}\right)}{\partial \lambda_{23}} = \frac{\partial \log\left(\frac{\pi(0,1,1)\pi_3(1)}{\pi_{13}(0,1)\pi_{23}(1,1)}\right)}{\partial \lambda_{23}} = \frac{\pi_{23}(1,1)}{\pi_3(1)} - \frac{\pi(0,1,1)}{\pi_{13}(0,1)}$$

$$\frac{\partial \log\left(\frac{\pi(1,0,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(0,1)}\right)}{\partial \lambda_2} = \frac{\partial \log\left(\frac{\pi(1,1,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(1,1)}\right)}{\partial \lambda_2} = \frac{\partial \log\left(\frac{\pi(1,0,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(0,1)}\right)}{\partial \lambda_{23}} = \frac{\partial \log\left(\frac{\pi(1,1,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(1,1)}\right)}{\partial \lambda_{23}} = \frac{\pi_{23}(1,1)}{\pi_3(1)} - \frac{\pi(1,1,1)}{\pi_{13}(1,1)}$$

$$\frac{\partial \log\left(\frac{\pi(0,0,0)\pi_3(0)}{\pi_{13}(0,0)\pi_{23}(0,0)}\right)}{\partial \lambda_{12}} = \frac{\pi(1,1,0)}{\pi_3(0)}$$

$$\frac{\partial \log\left(\frac{\pi(0,1,0)\pi_3(0)}{\pi_{13}(0,0)\pi_{23}(1,0)}\right)}{\partial \lambda_{12}} = \frac{\pi(1,1,0)}{\pi_3(0)} - \frac{\pi(1,1,1)}{\pi_{23}(1,0)}$$

$$\frac{\partial \log\left(\frac{\pi(1,0,0)\pi_3(0)}{\pi_{13}(1,0)\pi_{23}(0,0)}\right)}{\partial \lambda_{12}} = \frac{\pi(1,1,0)}{\pi_3(0)} - \frac{\pi(1,1,1)}{\pi_{13}(1,0)}$$

$$\frac{\partial \log\left(\frac{\pi(1,1,0)\pi_3(0)}{\pi_{13}(1,0)\pi_{23}(1,0)}\right)}{\partial \lambda_{12}} = 1 + \frac{\pi(1,1,0)}{\pi_3(0)} - \frac{\pi(1,1,1)}{\pi_{13}(1,0)} - \frac{\pi(1,1,1)}{\pi_{23}(1,0)}$$

$$\frac{\partial \log\left(\frac{\pi(0,0,1)\pi_3(1)}{\pi_{13}(0,1)\pi_{23}(0,1)}\right)}{\partial \lambda_{12}} = \frac{\partial \log\left(\frac{\pi(0,0,1)\pi_3(1)}{\pi_{13}(0,1)\pi_{23}(0,1)}\right)}{\partial \lambda_{123}} = \frac{\pi(1,1,1)}{\pi_3(1)}$$

$$\frac{\partial \log\left(\frac{\pi(0,1,1)\pi_3(1)}{\pi_{13}(0,1)\pi_{23}(1,1)}\right)}{\partial \lambda_{12}} = \frac{\partial \log\left(\frac{\pi(0,1,1)\pi_3(1)}{\pi_{13}(0,1)\pi_{23}(1,1)}\right)}{\partial \lambda_{123}} = \frac{\pi(1,1,1)}{\pi_3(1)} - \frac{\pi(1,1,1)}{\pi_{23}(1,1)}$$

$$\frac{\partial \log\left(\frac{\pi(1,0,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(0,1)}\right)}{\partial \lambda_{12}} = \frac{\partial \log\left(\frac{\pi(1,0,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(0,1)}\right)}{\partial \lambda_{123}} = \frac{\pi(1,1,1)}{\pi_3(1)} - \frac{\pi(1,1,1)}{\pi_{13}(1,1)}$$

$$\frac{\partial \log\left(\frac{\pi(1,1,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(1,1)}\right)}{\partial \lambda_{12}} = \frac{\partial \log\left(\frac{\pi(1,1,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(1,1)}\right)}{\partial \lambda_{123}} = 1 + \frac{\pi(1,1,1)}{\pi_3(1)} - \frac{\pi(1,1,1)}{\pi_{13}(1,1)} - \frac{\pi(1,1,1)}{\pi_{23}(1,1)}$$

Table A.5: Partial derivatives of f_{12}^{LRT} with respect to the seven λ.

The remaining partial derivatives not presented here equal zero.

$$\Delta[1,1] = \frac{-\pi_1(1)}{n_\emptyset} AE[LRT_{12}] + 2\left[\pi(1,0,0)\log\left(\frac{\pi(1,0,0)\pi_3(0)}{\pi_{13}(1,0)\pi_{23}(0,0))}\right) + \pi(1,1,0)\log\left(\frac{\pi(1,1,0)\pi_3(0)}{\pi_{13}(1,0)\pi_{23}(1,0)}\right) + \pi(1,0,1)\log\left(\frac{\pi(1,0,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(0,1))}\right) + \pi(1,1,1)\log\left(\frac{\pi(1,1,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(1,1)}\right)\right];$$

$$\Delta[2,1] = \frac{-\pi_2(1)}{n_\emptyset} AE[LRT_{12}] + 2\left[\pi(0,1,0)\log\left(\frac{\pi(0,1,0)\pi_3(0)}{\pi_{13}(0,0)\pi_{23}(1,0))}\right) + \pi(1,1,0)\log\left(\frac{\pi(1,1,0)\pi_3(0)}{\pi_{13}(1,0)\pi_{23}(1,0)}\right) + \pi(0,1,1)\log\left(\frac{\pi(0,1,1)\pi_3(1)}{\pi_{13}(0,1)\pi_{23}(1,1))}\right) + \pi(1,1,1)\log\left(\frac{\pi(1,1,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(1,1)}\right)\right];$$

$$\Delta[3,1] = \frac{-\pi_{12}(1,1)}{n_\emptyset} AE[LRT_{12}] + 2\left[\pi(1,1,0)\log\left(\frac{\pi(1,1,0)\pi_3(0)}{\pi_{13}(1,0)\pi_{23}(1,0)}\right) + \pi(1,1,1)\log\left(\frac{\pi(1,1,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(1,1)}\right)\right];$$

$$\Delta[4,1] = \frac{-\pi_3(1)}{n_\emptyset} AE[LRT_{12}] + 2\left[\pi(0,0,1)\log\left(\frac{\pi(0,0,1)\pi_3(1)}{\pi_{13}(0,1)\pi_{23}(0,1))}\right) + \pi(0,1,1)\log\left(\frac{\pi(0,1,1)\pi_3(1)}{\pi_{13}(0,1)\pi_{23}(1,1)}\right) + \pi(1,0,1)\log\left(\frac{\pi(1,0,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(0,1))}\right) + \pi(1,1,1)\log\left(\frac{\pi(1,1,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(1,1)}\right)\right];$$

$$\Delta[5,1] = \frac{-\pi_{13}(1,1)}{n_\emptyset} AE[LRT_{12}] + 2\left[\pi(1,0,1)\log\left(\frac{\pi(1,0,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(0,1))}\right) + \pi(1,1,1)\log\left(\frac{\pi(1,1,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(1,1)}\right)\right];$$

$$\Delta[6,1] = \frac{-\pi_{23}(1,1)}{n_\emptyset} AE[LRT_{12}] + 2\left[\pi(0,1,1)\log\left(\frac{\pi(0,1,1)\pi_3(1)}{\pi_{13}(0,1)\pi_{23}(1,1))}\right) + \pi(1,1,1)\log\left(\frac{\pi(1,1,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(1,1)}\right)\right];$$

$$\Delta[7,1] = \frac{-\pi(1,1,1)}{n_\emptyset} AE[LRT_{12}] + 2\left[\pi(1,1,1)\log\left(\frac{\pi(1,1,1)\pi_3(1)}{\pi_{13}(1,1)\pi_{23}(1,1)}\right)\right];$$

$$\Delta[1,2] = \frac{-\pi_1(1)}{n_\emptyset} AE[LRT_{13}] + 2\left[\pi(1,0,0)\log\left(\frac{\pi(1,0,0)\pi_2(0)}{\pi_{12}(1,0)\pi_{23}(0,0))}\right) + \pi(1,1,0)\log\left(\frac{\pi(1,1,0)\pi_2(1)}{\pi_{12}(1,1)\pi_{23}(1,0)}\right) + \pi(1,0,1)\log\left(\frac{\pi(1,0,1)\pi_2(0)}{\pi_{12}(1,0)\pi_{23}(0,1))}\right) + \pi(1,1,1)\log\left(\frac{\pi(1,1,1)\pi_2(1)}{\pi_{12}(1,1)\pi_{23}(1,1)}\right)\right];$$

$$\Delta[2,2] = \frac{-\pi_2(1)}{n_\emptyset} AE[LRT_{13}] + 2\left[\pi(0,1,0)\log\left(\frac{\pi(0,1,0)\pi_2(1)}{\pi_{12}(0,1)\pi_{23}(1,0))}\right) + \pi(1,1,0)\log\left(\frac{\pi(1,1,0)\pi_2(1)}{\pi_{12}(1,1)\pi_{23}(1,0)}\right) + \pi(0,1,1)\log\left(\frac{\pi(0,1,1)\pi_2(1)}{\pi_{12}(0,1)\pi_{23}(1,1))}\right) + \pi(1,1,1)\log\left(\frac{\pi(1,1,1)\pi_2(1)}{\pi_{12}(1,1)\pi_{23}(1,1)}\right)\right];$$

$$\Delta[3,2] = \frac{-\pi_{12}(1,1)}{n_\emptyset} AE[LRT_{13}] + 2\left[\pi(1,1,0)\log\left(\frac{\pi(1,1,0)\pi_2(1)}{\pi_{12}(1,1)\pi_{23}(1,0)}\right) + \pi(1,1,1)\log\left(\frac{\pi(1,1,1)\pi_2(1)}{\pi_{12}(1,1)\pi_{23}(1,1)}\right)\right];$$

$$\Delta[4,2] = \frac{-\pi_3(1)}{n_\emptyset} AE[LRT_{13}] + 2\left[\pi(0,0,1)\log\left(\frac{\pi(0,0,1)\pi_2(0)}{\pi_{12}(0,0)\pi_{23}(0,1))}\right) + \pi(0,1,1)\log\left(\frac{\pi(0,1,1)\pi_2(1)}{\pi_{12}(0,1)\pi_{23}(1,1)}\right) + \pi(1,0,1)\log\left(\frac{\pi(1,0,1)\pi_2(0)}{\pi_{12}(1,0)\pi_{23}(0,1))}\right) + \pi(1,1,1)\log\left(\frac{\pi(1,1,1)\pi_2(1)}{\pi_{12}(1,1)\pi_{23}(1,1)}\right)\right];$$

$$\Delta[5,2] = \frac{-\pi_{13}(1,1)}{n_\emptyset} AE[LRT_{13}] + 2\left[\pi(1,0,1)\log\left(\frac{\pi(1,0,1)\pi_2(0)}{\pi_{12}(1,0)\pi_{23}(0,1))}\right) + \pi(1,1,1)\log\left(\frac{\pi(1,1,1)\pi_2(1)}{\pi_{12}(1,1)\pi_{23}(1,1)}\right)\right];$$

$$\Delta[6,2] = \frac{-\pi_{23}(1,1)}{n_\emptyset} AE[LRT_{13}] + 2\left[\pi(0,1,1)\log\left(\frac{\pi(0,1,1)\pi_2(1)}{\pi_{12}(0,1)\pi_{23}(1,1))}\right) + \pi(1,1,1)\log\left(\frac{\pi(1,1,1)\pi_2(1)}{\pi_{12}(1,1)\pi_{23}(1,1)}\right)\right];$$

$$\Delta[7,2] = \frac{-\pi(1,1,1)}{n_\emptyset} AE[LRT_{13}] + 2\left[\pi(1,1,1)\log\left(\frac{\pi(1,1,1)\pi_2(1)}{\pi_{12}(1,1)\pi_{23}(1,1)}\right)\right].$$

Table A.6: Derivatives of f_{12}^{LRT} and f_{13}^{LRT} with respect to the seven λ expressed in cell probabilities.

$$\Delta[1,3] = \frac{-\pi_1(1)}{n_\emptyset} AE[LRT_{23}] + 2\left[\pi(1,0,0)\log\left(\frac{\pi(1,0,0)\pi_1(1)}{\pi_{12}(1,0)\pi_{13}(1,0))}\right) + \pi(1,1,0)\log\left(\frac{\pi(1,1,0)\pi_1(1)}{\pi_{12}(1,1)\pi_{13}(1,0)}\right) + \pi(1,0,1)\log\left(\frac{\pi(1,0,1)\pi_1(1)}{\pi_{12}(1,0)\pi_{13}(1,1))}\right) + \pi(1,1,1)\log\left(\frac{\pi(1,1,1)\pi_1(1)}{\pi_{12}(1,1)\pi_{13}(1,1)}\right)\right];$$

$$\Delta[2,3] = \frac{-\pi_2(1)}{n_\emptyset} AE[LRT_{23}] + 2\left[\pi(0,1,0)\log\left(\frac{\pi(0,1,0)\pi_1(0)}{\pi_{12}(0,1)\pi_{13}(0,0))}\right) + \pi(1,1,0)\log\left(\frac{\pi(1,1,0)\pi_1(1)}{\pi_{12}(1,1)\pi_{13}(1,0)}\right) + \pi(0,1,1)\log\left(\frac{\pi(0,1,1)\pi_1(0)}{\pi_{12}(0,1)\pi_{13}(0,1))}\right) + \pi(1,1,1)\log\left(\frac{\pi(1,1,1)\pi_1(1)}{\pi_{12}(1,1)\pi_{13}(1,1)}\right)\right];$$

$$\Delta[3,3] = \frac{-\pi_{12}(1,1)}{n_\emptyset} AE[LRT_{23}] + 2\left[\pi(1,1,0)\log\left(\frac{\pi(1,1,0)\pi_1(1)}{\pi_{12}(1,1)\pi_{13}(1,0)}\right) + \pi(1,1,1)\log\left(\frac{\pi(1,1,1)\pi_1(1)}{\pi_{12}(1,1)\pi_{13}(1,1)}\right)\right];$$

$$\Delta[4,3] = \frac{-\pi_3(1)}{n_\emptyset} AE[LRT_{23}] + 2\left[\pi(0,0,1)\log\left(\frac{\pi(0,0,1)\pi_1(0)}{\pi_{12}(0,0)\pi_{13}(0,1))}\right) + \pi(0,1,1)\log\left(\frac{\pi(0,1,1)\pi_1(0)}{\pi_{12}(0,1)\pi_{13}(0,1)}\right) + \pi(1,0,1)\log\left(\frac{\pi(1,0,1)\pi_1(1)}{\pi_{12}(1,0)\pi_{13}(1,1))}\right) + \pi(1,1,1)\log\left(\frac{\pi(1,1,1)\pi_1(1)}{\pi_{12}(1,1)\pi_{13}(1,1)}\right)\right];$$

$$\Delta[5,3] = \frac{-\pi_{13}(1,1)}{n_\emptyset} AE[LRT_{23}] + 2\left[\pi(1,0,1)\log\left(\frac{\pi(1,0,1)\pi_1(1)}{\pi_{12}(1,0)\pi_{13}(1,1))}\right) + \pi(1,1,1)\log\left(\frac{\pi(1,1,1)\pi_1(1)}{\pi_{12}(1,1)\pi_{13}(1,1)}\right)\right];$$

$$\Delta[6,3] = \frac{-\pi_{23}(1,1)}{n_\emptyset} AE[LRT_{23}] + 2\left[\pi(0,1,1)\log\left(\frac{\pi(0,1,1)\pi_1(0)}{\pi_{12}(0,1)\pi_{13}(0,1))}\right) + \pi(1,1,1)\log\left(\frac{\pi(1,1,1)\pi_1(1)}{\pi_{12}(1,1)\pi_{13}(1,1)}\right)\right];$$

$$\Delta[7,3] = \frac{-\pi(1,1,1)}{n_\emptyset} AE[LRT_{23}] + 2\left[\pi(1,1,1)\log\left(\frac{\pi(1,1,1)\pi_1(1)}{\pi_{12}(1,1)\pi_{13}(1,1)}\right)\right].$$

Table A.7: Derivatives of f_{23}^{LRT} with respect to the seven λ expressed in cell probabilities.

$$
\begin{aligned}
\Delta[1,1] &= \frac{(\lambda_{12})^2}{\frac{1}{\pi(0,0,0)}+\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,1,0)}}\left[-\pi_1(1)+\frac{\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,1,0)}}{\frac{1}{\pi(0,0,0)}+\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,1,0)}}\right] \\
&+\frac{(\lambda_{12}+\lambda_{123})^2}{\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,1)}}\left[-\pi_1(1)+\frac{\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,1)}}{\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,1)}}\right]; \\
\Delta[2,1] &= \frac{(\lambda_{12})^2}{\frac{1}{\pi(0,0,0)}+\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,1,0)}}\left[-\pi_2(1)+\frac{\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(1,1,0)}}{\frac{1}{\pi(0,0,0)}+\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,1,0)}}\right] \\
&+\frac{(\lambda_{12}+\lambda_{123})^2}{\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,1)}}\left[-\pi_2(1)+\frac{\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,1,1)}}{\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,1)}}\right]; \\
\Delta[3,1] &= \frac{(\lambda_{12})^2}{\frac{1}{\pi(0,0,0)}+\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,1,0)}}\left[-\pi_{12}(1,1)+\frac{\frac{1}{\pi(1,1,0)}}{\frac{1}{\pi(0,0,0)}+\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,1,0)}}\right] \\
&+\frac{(\lambda_{12}+\lambda_{123})^2}{\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,1)}}\left[-\pi_{12}(1,1)+\frac{\frac{1}{\pi(1,1,1)}}{\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,1)}}\right] \\
&+2\left[\frac{\lambda_{12}}{\frac{1}{\pi(0,0,0)}+\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,1,0)}}+\frac{\lambda_{12}+\lambda_{123}}{\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,1)}}\right]; \\
\Delta[4,1] &= \frac{(\lambda_{12}+\lambda_{123})^2}{\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,1)}}\left[1-\pi_3(1)\right]; \\
\Delta[5,1] &= \frac{(\lambda_{12}+\lambda_{123})^2}{\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,1)}}\left[-\pi_{13}(1,1)+\frac{\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,1)}}{\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,1)}}\right]; \\
\Delta[6,1] &= \frac{(\lambda_{12}+\lambda_{123})^2}{\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,1)}}\left[-\pi_{23}(1,1)+\frac{\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,1,1)}}{\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,1)}}\right]; \\
\Delta[7,1] &= \frac{(\lambda_{12}+\lambda_{123})^2}{\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,1)}}\left[-\pi(1,1,1)+\frac{\frac{1}{\pi(1,1,1)}}{\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,1)}}\right] \\
&+2\left[\frac{\lambda_{12}+\lambda_{123}}{\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,1)}}\right];
\end{aligned}
$$

Table A.8: Derivatives of f_{12}^{Wald} with respect to the seven λ expressed in cell probabilities.

$$\Delta[1,2] = \frac{(\lambda_{13})^2}{\frac{1}{\pi(0,0,0)}+\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}}\left[-\pi_1(1)+\frac{\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}}{\frac{1}{\pi(0,0,0)}+\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}}\right]$$

$$+\frac{(\lambda_{13}+\lambda_{123})^2}{\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\left[-\pi_1(1)+\frac{\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}{\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\right];$$

$$\Delta[2,2] = \frac{(\lambda_{13}+\lambda_{123})^2}{\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\left[1-\pi_2(1)\right];$$

$$\Delta[3,2] = \frac{(\lambda_{13}+\lambda_{123})^2}{\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\left[-\pi_{12}(1,1)+\frac{\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}{\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\right];$$

$$\Delta[4,2] = \frac{(\lambda_{13})^2}{\frac{1}{\pi(0,0,0)}+\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}}\left[-\pi_3(1)+\frac{\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(1,0,1)}}{\frac{1}{\pi(0,0,0)}+\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}}\right]$$

$$+\frac{(\lambda_{13}+\lambda_{123})^2}{\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\left[-\pi_3(1)+\frac{\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,1,1)}}{\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\right];$$

$$\Delta[5,2] = \frac{(\lambda_{13})^2}{\frac{1}{\pi(0,0,0)}+\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}}\left[-\pi_{13}(1,1)+\frac{\frac{1}{\pi(1,0,1)}}{\frac{1}{\pi(0,0,0)}+\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}}\right]$$

$$+\frac{(\lambda_{13}+\lambda_{123})^2}{\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\left[-\pi_{13}(1,1)+\frac{\frac{1}{\pi(1,1,1)}}{\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\right]$$

$$+2\left[\frac{\lambda_{13}}{\frac{1}{\pi(0,0,0)}+\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}}+\frac{\lambda_{13}+\lambda_{123}}{\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\right];$$

$$\Delta[6,2] = \frac{(\lambda_{13}+\lambda_{123})^2}{\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\left[-\pi_{23}(1,1)+\frac{\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,1,1)}}{\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\right];$$

$$\Delta[7,2] = \frac{(\lambda_{13}+\lambda_{123})^2}{\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\left[-\pi(1,1,1)+\frac{\frac{1}{\pi(1,1,1)}}{\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\right]$$

$$+2\left[\frac{\lambda_{13}+\lambda_{123}}{\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\right];$$

Table A.9: Derivatives of f_{13}^{Wald} with respect to the seven λ expressed in cell probabilities.

$$\Delta[1,3] = \frac{(\lambda_{23}+\lambda_{123})^2}{\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}} \left[1-\pi_1(1)\right];$$

$$\Delta[2,3] = \frac{(\lambda_{23})^2}{\frac{1}{\pi(0,0,0)}+\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}} \left[-\pi_2(1)+\frac{\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}}{\frac{1}{\pi(0,0,0)}+\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}}\right]$$
$$+\frac{(\lambda_{23}+\lambda_{123})^2}{\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}} \left[-\pi_2(1)+\frac{\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}{\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\right];$$

$$\Delta[3,3] = \frac{(\lambda_{23}+\lambda_{123})^2}{\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}} \left[-\pi_{12}(1,1)+\frac{\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}{\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\right];$$

$$\Delta[4,3] = \frac{(\lambda_{23})^2}{\frac{1}{\pi(0,0,0)}+\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}} \left[-\pi_3(1)+\frac{\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,1)}}{\frac{1}{\pi(0,0,0)}+\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}}\right]$$
$$+\frac{(\lambda_{23}+\lambda_{123})^2}{\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}} \left[-\pi_3(1)+\frac{\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,1)}}{\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\right];$$

$$\Delta[5,3] = \frac{(\lambda_{23}+\lambda_{123})^2}{\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}} \left[-\pi_{13}(1,1)+\frac{\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,1)}}{\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\right];$$

$$\Delta[6,3] = \frac{(\lambda_{23})^2}{\frac{1}{\pi(0,0,0)}+\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}} \left[-\pi_{23}(1,1)+\frac{\frac{1}{\pi(0,1,1)}}{\frac{1}{\pi(0,0,0)}+\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}}\right]$$
$$+\frac{(\lambda_{23}+\lambda_{123})^2}{\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}} \left[-\pi_{23}(1,1)+\frac{\frac{1}{\pi(1,1,1)}}{\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\right]$$
$$+2\left[\frac{\lambda_{23}}{\frac{1}{\pi(0,0,0)}+\frac{1}{\pi(0,0,1)}+\frac{1}{\pi(0,1,0)}+\frac{1}{\pi(0,1,1)}}+\frac{\lambda_{23}+\lambda_{123}}{\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\right];$$

$$\Delta[7,3] = \frac{(\lambda_{23}+\lambda_{123})^2}{\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}} \left[-\pi(1,1,1)+\frac{\frac{1}{\pi(1,1,1)}}{\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\right]$$
$$+2\left[\frac{\lambda_{23}+\lambda_{123}}{\frac{1}{\pi(1,0,0)}+\frac{1}{\pi(1,0,1)}+\frac{1}{\pi(1,1,0)}+\frac{1}{\pi(1,1,1)}}\right];$$

Table A.10: Derivatives of f_{23}^{Wald} with respect to the seven λ expressed in cell probabilities.

Appendix B

Expected cell counts

$\pi_1(0)$	$\pi_2(0)$	$\psi_{12}=1$		$\psi_{12}=1.5$		$\psi_{12}=2$		$\psi_{12}=2.5$		$\psi_{12}=3$		$\psi_{12}=4$	
0.1	0.1	**10**	90	**13.7**	86.3	16.9	83.1	19.7	80.3	22.1	77.9	26.3	73.7
		90	810	86.3	813.7	83.1	916.9	80.3	819.7	77.9	822.1	73.7	826.3
0.3	0.1	30	270	38.1	261.9	44.3	255.7	49.1	250.9	53.1	246.9	59.3	240.7
		70	630	61.9	638.1	55.7	644.3	50.9	649.1	46.9	653.1	40.7	659.3
0.5	0.1	50	450	59	441	65.1	434.9	69.6	430.4	73	427	77.9	422.1
		50	450	41	459	34.9	465.1	30.4	469.6	27	473	22.1	477.9
0.7	0.1	70	630	77.1	622.9	81.5	618.5	84.4	615.6	86.5	613.5	89.4	610.6
		30	270	22.9	277.1	18.5	281.5	15.6	284.4	**13.5**	286.5	**10.6**	289.4
0.9	0.1	90	810	92.9	807.1	94.5	805.5	95.5	804.5	96.2	803.8	97.1	802.9
		10	90	**7.1**	92.9	**5.5**	94.5	**4.5**	95.5	**3.8**	96.2	**2.9**	97.1
0.3	0.2	60	240	74.2	225.8	84.8	215.2	93.1	206.9	100	200	110.7	189.3
		140	560	125.8	574.2	115.2	584.8	106.9	593.1	100	600	89.3	610.7
0.5	0.2	100	400	116.1	383.9	127.2	372.8	135.4	364.6	141.7	358.3	151.1	348.9
		100	400	83.9	416.1	72.8	427.2	64.6	435.4	58.3	441.7	48.9	451.1
0.7	0.2	140	560	152.9	547.1	161	539	166.7	533.3	170.9	529.1	176.7	523.3
		60	240	47.1	252.9	39	261	33.3	266.7	29.2	270.9	23.3	276.7

Table B.1: Expected cell counts for different combinations of marginal probabilities ($\pi_1(0)$ and $\pi_2(0)$) and odds ratio values (ψ_{12}), when $n_{\emptyset} = 1\,000$.

Appendix C

Total Non-Admissible Region

	$n = 200$						
	$\rho_{12.3}$						
$\rho_{13.2}$; $\rho_{23.1}$	−0.9	−0.5	−0.1	0	0.1	0.5	0.9
0.1; 0.1	0.294	0.365	0.538	0.691	0.582	0.462	0.478
0.1; 0.2	0.241	0.269	0.439	0.621	0.488	0.311	0.313
0.2; 0.2	0.099	0.128	0.331	0.551	0.365	0.149	0.149
0.2; 0.3	0.075	0.082	0.288	0.524	0.318	0.088	0.088

	$n = 500$						
	$\rho_{12.3}$						
$\rho_{13.2}$; $\rho_{23.1}$	−0.9	−0.5	−0.1	0	0.1	0.5	0.9
0.1; 0.1	0.169	0.214	0.331	0.607	0.355	0.259	0.261
0.1; 0.2	0.131	0.135	0.242	0.552	0.262	0.142	0.142
0.2; 0.2	0.016	0.021	0.147	0.506	0.152	0.022	0.022
0.2; 0.3	0.011	0.011	0.138	0.502	0.142	0.012	0.012

	$n = 1\,000$						
	$\rho_{12.3}$						
$\rho_{13.2}$; $\rho_{23.1}$	−0.9	−0.5	−0.1	0	0.1	0.5	0.9
0.1; 0.1	0.076	0.098	0.155	0.545	0.162	0.112	0.112
0.1; 0.2	0.056	0.056	0.106	0.519	0.111	0.056	0.056
0.2; 0.2	0.001	0.001	0.057	0.499	0.057	0.001	0.001
0.2; 0.3	0.001	0.001	0.056	0.499	0.056	0.001	0.001

Table C.1: Values of the total non-admissible region (NAR), using non-signed LRT statistic