UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS MATEMÁTICAS

FACULTAD DE CIENCIAS

RESTRICTED OR COLOURED GRAPHICAL LOG-LINEAR MODELS

T E S I S

QUE PARA OBTENER EL GRADO ACADÉMICO DE

DOCTOR EN CIENCIAS

P R E S E N T A

RICARDO RAMÍREZ ALDANA

DIRECTOR DE TESIS: DRA. GUILLERMINA ESLAVA GÓMEZ

MÉXICO, D.F. **DICIEMBRE, 2010**

Acknowledgments

First of all, I would like to thank my dissertation committee for their helpful critical comments and for sharing their expertise and knowledge with me: Dr. Johan Van Horebeek, Dr. Ignacio Méndez-Ramírez, and especially Dr. Guillermina Eslava-Gómez, my main supervisor, to whom I owe a very special thank you. I would also like to thank Dr. Ruth Fuentes-García, Dr. Luis Antonio Rincón-Solís, and Dr. Eliane Rodrígues-Caloni, three examiners who despite not being part of my dissertation committee read this long dissertation and provided valuable suggestions that improved this manuscript.

During my stay at the University of Oxford, U.K. from January 2009 to July 2009 I continued my research work in the Department of Statistics. I wish to thank Professor Steffen Lauritzen for receiving me there, for sharing some of his valuable time with me, and for discussing with me concepts which were the basis of the main ideas presented in Chapter 4. My gratitude also goes to Helene Gehrmann for reading drafts containing those ideas and for sharing her knowledge in graphical Gaussian models with me; as well as to all the people I met at the Department and in Oxford who helped make my stay a memorable one.

I am also very grateful for the financial support I received from the National Council of Science and Technology (CONACYT); it would have been impossible for me to be a full-time student without this support.

I want to thank all my colleagues and friends who were also working on their respective theses at the National Autonomous University of Mexico (UNAM) and other institutions. We shared ideas as well as the same trials and tribulations of being a Ph. D. student.

I want to acknowledge all the professors who teach at my *alma mater*, the National Autonomous University of Mexico; they make this institution one of the greatest universities, and one that I am very proud to be a part of.

I would also like to thank all my friends who have contributed directly or indirectly to this dissertation. I want to to thank those friends whom I met long ago but our friendships are still enduring, those whom I met while I was trying to learn and improve my English, and also those who share one big passion in my life, art, which is a big part of who I am now.

I am especially grateful to my family. I want to thank my sister Laura and my parents, Ricardo and Laura, their support and endless encouragement mean a lot to me and they can not be expressed in words. I would like to thank my father for teaching me to be a hard worker and for showing me that it is possible to improve who you are and become a better version of yourself; and my mother, whose unconditional love and support is always there for me, thanks for everything.

Agradecimientos

Primeramente, quisiera agradecer a mi comité tutor por sus comentarios y críticas constructivas y por compartir algo de su experiencia profesional y conocimientos conmigo: Dr. Johan Van Horebeek, Dr. Ignacio Méndez Ramírez y especialmente a mi tutora principal, la Dra. Guillermina Eslava Gómez, a quien doy un agradecimiento todavía mayor. Quisiera agradecerles también a los tres sinodales quienes no formaron parte del comité tutor, Dra. Ruth Fuentes García, Dr. Luis Antonio Rincón Solís y a la Dra. Eliane Rodrígues Caloni, por haber leído esta tesis tan larga y por sus sugerencias las cuales mejoraron este trabajo.

De Enero a Julio del 2009 hice una estancia de investigación en el Departamento de Estadística de la Universidad de Oxford, R. U., quisiera agradecerle al Professor Steffen Lauritzen el haberme recibido y el tiempo valioso que me brindó, así como por las reuniones que tuvimos de las cuales derivaron muchas de las ideas que más adelante formarían el Capítulo 4 de esta tesis. También quisiera agradecerle a Helene Gehrmann el haber leído y comentado conmigo borradores que contenían esas ideas y el haber compartido su conocimiento respecto a modelos gráficos Gaussianos; así como a todas aquellas personas que conocí en el Departamento y en Oxford, quienes hicieron esa estancia memorable.

También le agradezco al Consejo Nacional de Ciencia y Tecnología (CONACYT) el apoyo financiero que me brindó a lo largo del Doctorado; hubiese sido imposible dedicarme solamente al estudio sin la beca que me otorgaron.

Mi agradecimiento a todos los colegas y amigos quienes también estaban trabajando en sus respectivos temas de investigación en la Universidad Nacional Autónoma de México (UNAM) o en otras instituciones. Compartimos juntos muchas ideas, así como los mismos problemas y adversidades por las cuales pasan en donde sea los estudiantes de Doctorado.

Agradezco también a todos los maestros que enseñan y han dado clases en mi *alma mater*, la Universidad Nacional Autónoma de México, quienes hacen de esta institución una de las mejores universidades y de la cual yo estoy muy orgulloso de ser parte.

También quisiera agradecerles su apoyo a todos mis amigos, de alguna manera todos ellos han contribuido en este trabajo. Quisiera agradecer a aquellos viejos amigos que conocí ya de tiempo atrás y todavía están conmigo, a aquellos que conocí mientras trataba de aprender y mejorar mi inglés y a todos aquellos con los cuales comparto una gran pasión, el arte, la cual se ha vuelto una parte muy importante de mí y de quien soy ahora.

Finalmente, agradezco mucho a mi familia, a mi hermana Laura y a mis padres, Ricardo y Laura, por su apoyo infinito, el cual está más allá de lo que simples palabras pueden expresar. Especialmente, gracias a mi padre, quien me ha enseñado a trabajar duro y a que uno puede siempre mejorar y volverse una versión nueva y mejor de uno mismo y a mi madre, quien siempre me ha dado su amor grande e incondicional y quien me ha apoyado a lo largo de toda mi vida sin importar qué, muchísimas gracias por todo.

Contents

List of Tables

List of Figures

Abstract

The main goals of this dissertation corresponded to define discrete models analogous to the ones defined and studied for the continuous case by Højsgaard and Lauritzen (2008) and to relate these discrete models to different concepts of *symmetry*. Some of these concepts have been already studied in the discrete case in models with only a certain number of variables, and in a few ones considering any number of variables, and others have been studied in the continuous case. As a consequence, new symmetry concepts and generalizations of other known symmetry definitions are obtained. These goals were achieved and they are discussed throughout this dissertation.

Log-linear models are used to determine the way in which a set of categorical variables are associated in a contingency table. A subset of them are graphical log-linear models in which conditional and marginal independences between variables are expressed in the model, additionally they use mathematical concepts from Graph theory. They can be represented through graphs formed by dots called vertices corresponding to variables and lines called edges, which depend on which parameters are contained in the model. Given the hierarchical nature of graphical log-linear models, if there is an interaction, there must be all interactions of lower order, so that the edges in the model are determined by the first-order interactions and consequently the edges are identified with those interactions.

We introduce new types of hierarchical log-linear models, restricted graphical log-linear (RGLL), in which certain main effects and first-order interactions are restricted to be equal. This equality restrictions depend on which elements are contained in classes obtained once the vertex and edge sets are partitioned. Additionally, RGLL models are defined according to mathematical concepts obtained from Graph theory. They can be represented by graphs with multiple edges between vertices, corresponding to all different permutations among the categories of each pair of associated variables, including both vertex and edge colourings. Having vertices with the same colour means that the main effects for their corresponding variables are the same in all their categories. Having multiple edges with the same colour means that the first-order interactions identified with such edges are equal. RGLL are more parsimonious than graphical log-linear models and they can be used to identify relationships between the expected frequencies in some cells of the table, and they also reflect independences between variables.

Additionally, they can be used to analyze variations of the concept of symmetry.

We study properties of RGLL models, obtain the associated likelihood equations, and derive an algorithm to solve the likelihood equations based on the *iterative proportional fitting* algorithm used on log-linear models. This is implemented on routines we use to fit or to search RGLL models that fit as good as possible to some data.

The classic concept of symmetry for two-way contingency tables corresponds to having that the expected frequency for any cell $m(i, j)$ above the diagonal is equal to the expected frequency for the corresponding cell $m(j, i)$ below the diagonal. This concept has been generalized for higher dimensions; however, there is not a unique way to do this, but in general these generalizations correspond to models in which the expected frequencies for certain cells are equal. These generalizations do not include the possibility of combining symmetry with the independences derived from a graphical log-linear model. Another classic model defined for two-way contingency tables is quasi-symmetry. Both the symmetry and quasi-symmetry models can be expressed as RGLL models, so that the latter can be considered as a generalization of the former.

Graphical Gaussian models are analogous to graphical log-linear models. Its definition is determined by which elements in its concentration matrix are zero. Højsgaard and Lauritzen (2008) define a particular type of graphical Gaussian models in which elements in the concentration matrix are restricted to be equal and whose graphical representation is based on coloured graphs. In this sense, these are similar to RGLL models. A second type of symmetry used in the case of continuous variables with Gaussian distribution corresponds to models in which the concentration or covariance matrix is preserved in spite of applying a transformation in the matrix, which implies that if the mean is zero, then the distribution is preserved. A particular instance of the models defined by Højsgard and Lauritzen are called *RCOP*, they use this second type of symmetry by permuting certain vertices in graphical Gaussian models. The kind of permutation is related to the concept of symmetry group as defined in Algebra.

Additionally, we define two particular cases of RGLL models, *label* and *level invariant* models, which are graphical log-linear and whose associated graphs are triangle-free. Label invariant models preserve the model, and consequently the distribution, after permuting certain vertices according to a set of permutations that preserve the graph. Additionally, they are scale invariant and preserve the independences obtained from the graph. Then, they are analogous to RCOP models since they combine the type of symmetry that preserves the distribution after applying vertex permutations with graphical log-linear models. Level invariant models preserve certain expected frequencies in the table after permuting the values taken by some variables, so that they combine the type of symmetry that equates the expected frequencies for certain cells with the independences derived from a graphical log-linear model.

We present illustrative examples and applications of *label* and *level invariant* and RGLL models in general. RGLL models are selected and fitted to three data sets, in these data the fit is improved when they are used. By examples of label and level invariant models we understand that we define particular models and symmetry conditions that can be represented using them. In this way we illustrate for label and level invariant models the following aspects: theoretical concepts, in which cases they can be used, their possible meaning, the parameter restrictions that should be satisfied, and how to represent them as RGLL models. Applications are concerned with data sets in which some of the models defined in the examples section can be fitted because there is symmetry, which can be confirmed through them.

Resumen

Los objetivos principales de esta tesis son definir modelos discretos análogos a aquellos estudiados en el caso continuo por Hojsgaard y Lauritzen (2008) y relacionar estos con diferentes conceptos de *simetría*. Algunos de estos conceptos ya han sido estudiados en el caso discreto en modelos definidos solo para un cierto número de variables, y en unos cuantos modelos que incluso consideran cualquier número de variables, y otros han sido estudiados en el caso continuo. Como consecuencia, nuevos conceptos de simetría y generalizaciones de otras definiciones de simetría conocidas son obtenidos. Estos objetivos fueron satisfechos y son discutidos a los largo de esta tesis.

Los modelos loglineales son usados para determinar la forma en que se asocian un conjunto de variables categóricas en una tabla de contingencia. Un subconjunto de ellos son los modelos gráficos loglineales en los cuales se expresan independencias condicionales y marginales entre variables, además estos usan conceptos matemáticos de teoría de Gráficas. Estos modelos pueden representarse mediante gráficas compuestas por puntos llamados vértices que corresponden a las variables y líneas llamadas aristas las cuales dependen de los parámetros contenidos en el modelo. Estos parámetros corresponden a efectos principales e interacciones entre las variables de distinto orden. Dada la naturaleza jerárquica de los modelos gráficos loglineales cuando hay una interacción deben existir todas las interacciones de orden inferior, así que las aristas en el modelo están determinadas por las interacciones de primer orden, por lo cual se identifican las aristas con estas interacciones.

Introducimos nuevos tipos de modelos loglineales jerárquicos, los modelos gráficos loglineales restringidos (RGLL), para los cuales ciertos efectos principales e interacciones de primer orden están restringidos a ser iguales. Estas igualdades dependen de cuales elementos están presentes en clases obtenidas mediante particiones del conjunto de vértices y del conjunto de aristas. Adicionalmente, los modelos RGLL están definidos de acuerdo a conceptos matemáticos de teoría de Gráficas. Estos modelos pueden representarse mediante gráficas con aristas múltiples entre los vértices, las cuales corresponden a todas las permutaciones posibles entre las categorías de cada par de variables asociadas, incluyendo además coloraciones por vértices y por aristas. El tener un conjunto de vértices con un mismo color significa que los efectos principales de las variables en esa clase son iguales en todos sus niveles. El tener aristas múltiples con

un mismo color significa que las interacciones de primer orden que se identifican con tales aristas son las mismas. Los modelos RGLL son más parsimoniosos que los modelos gráficos loglineales y pueden utilizarse para identificar relaciones existentes entre los valores esperados de algunas celdas de la tabla, además de reflejar independencias entre variables. Adicionalmente, pueden usarse para analizar variaciones del concepto de *simetría*.

Se estudian propiedades de los modelos RGLL, se obtienen las ecuaciones de verosimilitud asociadas y se deriva un algoritmo para resolverlas basado en el método de ajuste iterativo (*iterative proportional fitting*) usado en modelos loglineales. Este es implementado en rutinas creadas para ajustar o para buscar modelos RGLL que se ajusten bien a unos datos.

El concepto clásico de simetría en tablas de contingencia de dos dimensiones se refiere a que el valor esperado en cualquier celda $m(i, j)$ arriba de la diagonal es igual al valor esperado para la correspondiente celda $m(j, i)$ abajo de la diagonal. Este concepto de simetría se ha generalizado para dimensiones mayores; sin embargo, no hay una única manera de hacerlo aunque en general corresponde a modelos en los cuales los valores esperados de ciertas celdas se igualan. Estas generalizaciones no incluyen la posibilidad de combinar simetría con las independencias derivadas de un modelo gráfico loglineal. Otro modelo clásico definido para tablas de dos dimensiones es el de cuasisimetría. Tanto el modelo de simetría como el de cuasisimetría en dos dimensiones pueden ser expresados como modelos RGLL, así que estos últimos pueden considerarse como una generalización de los primeros.

Los modelos gráficos Gaussianos son análogos a los modelos gráficos loglineales. Su definición depende de cuales elementos de su matriz de concentración son cero. Højsgard y Lauritzen (2008) definen una clase particular de modelos gráficos Gaussianos en los cuales elementos en la matriz de concentración son iguales entre si y cuya representación está basada en gráficas coloreadas. En este sentido estos modelos son similares a los modelos RGLL. Un segundo tipo de simetría usado cuando se tienen variables aleatorias continuas con distribución Gaussiana se refiere a modelos en los cuales la matriz de concentración o covarianza se preserva a pesar de aplicar una transformación en la matriz, lo cual implica que si la media es cero entonces la distribución se conserva. Un caso particular de los modelos definidos por Højsgard y Lauritzen son los llamados *RCOP*, los cuales implementan este segundo tipo de simetría al permutar ciertos vértices en modelos gráficos Gaussianos. El tipo de permutación utilizada está relacionado con el concepto de grupos de simetría tal como se define en Álgebra.

En este trabajo además se definen dos casos particulares de modelos RGLL, *label invariant* y *level invariant*, que son modelos gráficos loglineales cuyas gráficas asociadas están libres de triángulos. Aquellos denominados *label invariant* preservan el modelo

y como consecuencia la distribución a pesar de permutar los vértices de acuerdo a un conjunto de permutaciones que conservan la gráfica. Adicionalmente, son invariantes a la escala y preservan las independencias dadas por el modelo gráfico loglineal. Esto quiere decir que son análogos a los *RCOP* ya que combinan el tipo de simetría que preserva distribuciones después de aplicar permutaciones de los vértices con modelos gráficos loglineales. Aquellos denominados *level invariant* conservan ciertos valores esperados en la tabla después de permutar los valores que toman ciertas variables. Así que combinan el tipo de simetría en la cual los valores esperados para ciertas celdas son iguales con las independencias derivadas de los modelos gráficos loglineales.

Se presentan ejemplos ilustrativos y aplicaciones de los modelos *label* y *level invariant*, así como de los RGLL en general. Varios modelos RGLL son seleccionados y ajustados en tres conjuntos de datos en los cuales el ajuste mejora cuando estos son usados. Por ejemplos de modelos *label* y *level invariant* nos referimos a definir modelos particulares y condiciones especiales de simetría que pueden representarse usándolos. Con estos ejemplos se ilustran para los modelos *label* y *level invariant* los siguientes aspectos: conceptos teóricos, algunos casos en los cuales pueden usarse, su posible significado, las restricciones en los parámetros que deben ser satisfechas y la manera como estos se representan como modelos RGLL. Por aplicaciones se entiende el ajustar algunos de los modelos definidos en la sección de ejemplos en bases de datos específicas en las cuales hay simetría, la cual se confirma a través de ellos.

Chapter 1

Introduction

On many problems we model the association between variables. Such variables can be continuous or discrete, or even a mix of both kinds. In this work we study models for discrete variables whose levels or categories are unordered. Techniques to model association have been developed for the continuous case and in general they can not be applied to the discrete case. In particular, the models we present here have been deeper studied in the continuous case.

To model association between variables we use log-linear models in the discrete case. Using them for any contingency table, we try to explain the underlying association between variables based on the expected frequencies for each cell in the table. These expected frequencies could be considered as response variables on ANOVA-type models in which the explanatory variables are given by parameters corresponding to the effects of different variables at different levels, or main effects, and the parameters corresponding to different-order interactions between those variables, for example first-order interactions are interactions formed by only two variables.

A particular kind of log-linear models are hierarchical log-linear models, in which, given any interaction in the model, we should have all corresponding lower-order interactions; *i.e.* all interactions whose variables are a subset of the set of variables forming the interaction we are interested in. These models contain graphical log-linear models, which are models in which association between variables is represented through graphs that could be easy to see and interpret and allow us to identify conditional and marginal independences between variables in the model. One additional advantage of graphical models is that many concepts given in probabilistic terms can be expressed through basic graph theory concepts, which can be used to interpret a model. For example, we can determine marginal and conditional independences between variables using the *separator set* concept defined in graph theory. They are even more useful as we increase the number of variables because in these cases the interpretation of log-linear models in terms of marginal or conditional independences could get too complex. We give a

formal definition of graphical log-linear models in Chapter 2.

Graphical log-linear models have associated graphs in which variables are represented here as circles called vertices and the presence of associations between variables with lines, called edges, which join the associated vertices. This association is determined according to which first-order interaction terms are in the model. When there is not an edge between two vertices, these variables are conditionally independent given the other variables.

For example, suppose that we have three variables, labeled as A, B, and C, and a model containing a constant term, main effects, and the first-order interactions formed by A and B and by B and C, which means that A is associated with B and at the same time B is associated with C. These associations can be graphically represented with the graph shown in figure 1.1. We see that A and C are conditionally independent given B, because there are no edges joining A and C.

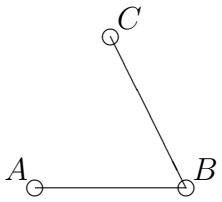


Figure 1.1: Graphical log-linear model formed by three variables labeled as A, B, and C.

In the continuous case similar graphical models are called graphical Gaussian models (Whittaker, 1990, ch. 6) or covariance selection models (Dempster, 1972) and they are also used to represent association between continuous variables. These models are defined by setting to zero some elements in the concentration matrix, inverse of the covariance matrix, on multivariate Gaussian observations in such a way that we connect two vertices with an edge on a graph if the entries corresponding to those variables are not zero. We observe then that the entries on the matrix have a role similar to that of first-order interactions on graphical log-linear models; however, discrete and continuous graphical models are not completely analogous because on the discrete case we consider interactions of higher order and the levels taken by the variables as part of the parameters, which does not apply in the continuous case. As before, if there is no edge between two variables, those variables are independent given the other variables.

When we model association between variables, our main goal is to get models that fit as good as possible to some data, but at the same time we seek simple models according to the parsimony principle; *i.e.* models with as few parameters as possible. Obviously, we also want models that reflect as much as possible the true association between variables. To get a better fit, we sometimes need models with a specific structure,

for example, models whose parameters are constrained in some way. The restrictions could be for example certain equalities between some parameters. Such are the kind of restrictions we will impose in this work to some parameters in certain hierarchical log-linear models with a graphical structure.

For example, consider the study presented by Agresti (2002a, p. 446) to determine the relation between purchase choice of different instant decaffeinated coffee brands at two different times (table 1.1).

First purchase	**Second purchase**				
	High Point	Taster's Choice	Sanka	Nescafe	Brim
High Point	93	17	44	7	10
Taster's Coice	9	46	11	0	9
Sanka	17	11	155	9	12
Nescafe	6	4	9	15	2
Brim	10	4	12	2	27

Table 1.1: Table corresponding to subjects' purchase choice of instant decaffeinated coffee at two times to show the meaning of possible log-linear models with two variables that may be fitted to these data.

A first log-linear model we could try to fit to those data is one model indicating independence between both variables. This model represents that the first purchase is not related with the second one. The graphical representation for this model is shown in figure 1.2(a). This representation is used because the vertices associated to each variable are not joined through an edge, which means that we have marginal independence. If it fitted well, it would mean that the first purchase is independent to the second purchase.

If the independence model did not fit well, it would mean that there is some kind of association between both variables. It would be important to figure out how this association is given. To determine this, we could try to use a graphical log-linear model different to the independence one. As we only have two variables, the only other possible graphical log-linear model is the one that allows association between both variables (figure 1.2(b)).

(a) Independence model (b) Saturated model

Figure 1.2: Graphical log-linear models for first purchase P and second purchase S of different coffee brands: (a) independence between purchases; (b) association between purchases.

This is a model that contains all possible parameters between both variables known as a saturated model. This model has a perfect fit to the table; *i.e.*, the estimated expected frequencies under the model are equal to the observed counts. Because of this fact, it would seem as a good model; however, it is not since it does not tell anything about existence of marginal independence or about the cells structure. Additionally, it is an over parametrized model, in the sense that we could possibly get a good fit using fewer parameters, consequently getting more precise estimators.

This means that in contingency tables besides determining marginal or conditional independences, it is also important to determine the cells structure, in the sense that we would like to know if there is certain relation among certain cells, if association between some vertices is similar, or if for any level some variables have similar effects. Additionally, it could be of interest to determine if under certain vertex interchange the distribution is preserved.

For the example given above, table 1.1, considering that there is association between both variables, we could have the following interests. Firstly, we could be interested in knowing if the number of people choosing any brand, for example Nescafe, at their first purchase is the same number of people who choose the same brand at their second purchase. This is known as marginal homogeneity. Secondly, we could be interested in knowing if there is symmetric association, *i.e.* if the association between purchasing brand i at the first purchase and brand j at the second purchase is similar to the one between purchasing brand j at the first purchase and brand i at the second purchase.

A model in which both conditions, marginal homogeneity and symmetric association, are jointly satisfied is a symmetry model, and it is interpreted as a model under which the expected frequencies for any cell (i, j) are equal to the expected frequencies for a cell (j, i). In this case, it means that the number of people who changed from a first brand at their first purchase to a second brand at their second purchase is the same as the number of people who changed from that second brand at their initial purchase to the first brand at the second purchase. So that, for a specific brand, we have that the number of consumers were not lost against a rival brand because the number that chose the rival brand is the same number that changed from that rival brand to the interest brand. A model in which marginal homogeneity is not satisfied but there is

symmetric association corresponds to a table that is not completely symmetric because of lack of marginal homogeneity. This model is known as a quasi-symmetry model. For this example, it means that even though the number of people who at the first purchase chose a brand is different to the number of people who chose the same brand at the second purchase, the association between pairs of brands is similar, so that if there were marginal homogeneity, we would have a symmetric table.

The research topic presented in this dissertation corresponds to propose and analyze a new kind of hierarchical log-linear models, which are similar to graphical log-linear models. These models are one of many ways of generalizing symmetry and quasi-symmetry models as defined for two-way contingency tables, and we call them restricted graphical log-linear models. They satisfy all conditions that characterize graphical log-linear models, but unlike graphical models, some parameters can be restricted to be equal. A group of parameters corresponding to first-order interactions could be equal for any permutation of the values taken simultaneously by both variables forming those interactions, indicating that associations between certain pairs of variables for certain levels are equal. Parameters corresponding to main effects could also be restricted. These restrictions correspond to equate main effects on all levels for all variables in the same subset of variables. Unlike the coffee brand example, in this work we could have tables with any number of variables.

Restricted graphical log-linear models have not been previously analyzed, except of course when they coincide with symmetry and quasi-symmetry models for two-way contingency tables. We notice that there is the possibility of all parameters being distinct, which means that graphical log-linear models are a particular case of restricted graphical log-linear models. We notice also that restricted graphical log-linear models could be more parsimonious because some parameters are restricted to be equal. Additionally, they could help us to get a better understanding of the cells structure in the table.

Graphs representing restricted graphical log-linear models need to include all different level permutations for any pair of variables. Because of this, instead of joining variables with one line only, we join them with multiple lines representing those permutations. All restrictions mentioned above can be represented on these graphs using vertex and edges colourings. For the main effects restrictions, we use vertex colourings and for the first-order interactions restrictions, we use edge colourings. For example, the quasi-symmetry model for the coffee brand selection data is represented as the graph given in figure 1.3.

This graph contains 25 edges, corresponding to 5^2 permutations, which are all possible ways of ordering 5 levels for the 2 variables. The 20 edges corresponding to first-order interactions of the form (i, j), $i \neq j$, are coloured in such a way that every

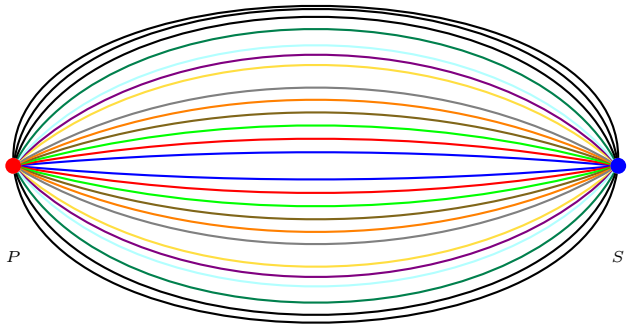


Figure 1.3: Colourings for the quasi-symmetry model for the coffee brand selection data. Edge colourings correspond to pairs of edges representing first-order interactions with level values (i, j) and (j, i), $i \neq j$ and other colourings correspond to edges of the form (i, i). Nodes S and P have a different colour to represent that main effects are not equal for all levels.

pair of edges (i, j) and (j, i), $i \neq j$, have the same colour because the association is the same on those level permutations. Then, we have 10 groups of pairs of edges of this kind, every group with the same colour, these colours are: blue, red, green, brown, orange, gray, yellow, violet, light blue, and dark green. All 5 edges of the form (i, i) should have a different colour; however, because to each edge corresponds a different colour all 5 edges are coloured in black to avoid using too many colours. The 2 vertices have different colours: red and blue, because the main effects of both variables are not the same for all levels of those variables.

By using restricted graphical log-linear models, we sometimes could get better approximations to the observed counts in the table when the data have a structure in which certain effects of some variables in all their levels are equal or when the association between some variables is similar. In some cases, we also simplify and get a better understanding of the relation between variables and their levels. We also get more parsimonious models because the number of parameters decreases. Additionally, there are particular cases of these models with interesting interpretations, for example models in which the distribution is preserved once interchanging some vertices or models in which the expected frequencies for certain cells are the same after restricting parameters.

Similar restrictions for the continuous case have been already applied by Højsgaard and Lauritzen (2005, 2007, 2008). They introduce new types of graphical Gaussian models by placing symmetry restrictions on the concentration or correlation matrices. They represent such restrictions with both vertex and edge colourings in the corresponding graphs. The diagonal elements of the matrices are equal if they are in the same vertex colouring and the elements outside of the diagonal are equal when they are in the same edge colouring. This means that they get matrices in which some parameters are restricted to be equal. Edge restrictions are similar to those in the discrete case because they indicate that the association between certain variables is similar; however, in the discrete case we also need to consider which levels are taken by those variables.

On the other hand, main effects take the place of the diagonal elements, which are the inverse of the variances in the concentration matrix.

Once defined restricted graphical log-linear models, we can determine and solve the corresponding likelihood equations. As a consequence we get the expected frequencies and deviance, the statistic used to determine how well the model fits, under any restricted model. Those likelihood equations can be solved using a numerical algorithm we specifically created for these models that converges to the solution. We also created comparison and selection methods to see if it is convenient to join colour classes and to choose a restricted model that fits as good as possible to some data. Finally, we define particular cases of restricted graphical log-linear models called *label* and *level invariant models*, which are models with a symmetry interpretation in terms of Graph theory and Statistics. These models extent in some way some similar models defined for the continuous case.

This work is divided into five chapters, including this Introduction as Chapter 1 and a chapter corresponding to related work, discussion, and conclusions as Chapter 5, and two Appendices.

In Appendix A we present basic Graph theory concepts based on Bondy and Murty (1976). We use these concepts throughout the dissertation.

In Appendix B we prove convergence to the maximum likelihood estimators of the values obtained with a numerical procedure defined in Chapter 3 similar to iterative proportional fitting that solves the likelihood equations associated to any restricted graphical log-linear model.

Chapter 2 corresponds to the theory on log-linear models already presented by other authors. We present log-linear models and we define the notation which will be also used in graphical log-linear models. We also present hierarchical log-linear models. Additionally, we determine likelihood equations and maximum likelihood estimators for log-linear models. We also explain how log-linear models can be expressed as generalized linear models. We show some particular log-linear models for two-way contingency tables, symmetry and quasi-symmetry models, and some generalizations for them. After that, we present the deviance concept, which is a statistic useful to evaluate how well a model fits some data and to generate model selection processes. We also describe graphical log-linear models in this chapter. We define them and we analyze the independence properties between variables, which can be inferred from their associated graphs. Finally, we present the deviance concept on the graphical model context, as well as some model selection methods that have been used to choose a graphical log-linear model.

In Chapter 3 we introduce new type of models corresponding to the main research topic in this project and they are what we innovated here. These are restricted graphical log-linear models, which can be represented through colourings on the corresponding associated graphs. In this chapter we define them in general and for different kinds of colourings: vertex, edge, and vertex and edge colourings. We also present examples for each case, in particular we present symmetry and quasi-symmetry models as special cases. Additionally, we derive the likelihood equations for each case corresponding to the model represented as log-linear models and as generalized log-linear models. In this last case, we can sometimes easily get reparametrized models with analogous restrictions on the reparametrized parameters or with restrictions on the reparametrized parameters that represent exactly the same model. We show here the restricted models which are of this kind.

We also introduce a numerical method we use to approximately solve the likelihood equations. This is a modified version of the iterative proportional fitting method used in log-linear models. We present a program in *Fortran*, based on some subroutines provided by Haberman (1972, 1976), in which we apply the method mentioned above to fit those models, join colour classes, and select models. We explain how it works and we give examples of some restricted graphical log-linear models fitted with it. It is important to note that model selection in this context refers only to select a colouring for a model, if we want to select the generating class that determines this model we can use other available software used for graphical models, in particular we use *MIM*.

Finally, in Chapter 4 we introduce two particular kinds of restricted graphical log-linear models, label and level invariant models. The first ones are models with a specific graph structure defined in such a way that the model is preserved in spite of interchanging some vertices on the graph and the level order. This means that they are scale invariant models whose distribution is preserved after such interchange. These models generalize symmetry from a graphical point of view. They are analogous to some models called *RCOP* defined on the Gaussian case that also preserve the distribution once applied certain symmetry groups to the vertex set. Level invariant models also have a specific graph structure and preserve expected frequencies after interchanging the values or levels taken by some groups of variables. Both models preserve the independences implied by the graphical model. We present various theoretical examples of those models as well as possible applications. We also fit some of them to specific data. Lastly, we outline how these models can be generalized to cover any graphical model and present one example to show how this generalization can be done.

Chapter 2

Log-linear and graphical log-linear models

2.1 Notation

The following notation is based on the one presented in section **4.2.1** by Lauritzen (1996) and used in both log-linear and graphical log-linear models.

$\|Z\|$	number of elements in a set Z.
Δ=V	set of vertices or set of variable names or labels.
E	set of edges corresponding to the associated graph.
$I = (I_\delta)_{\delta \in \Delta}$	discrete random variables associated to the set of vertices.
I_δ	categories or level set for I_δ.
$\|\mathrm{I}_\delta\|$	total number of categories for the variable I_δ.
I: $\mathrm{I} = \times_{\delta \in \Delta} \mathrm{I}_\delta$	variable value combinations.
i	cell or particular variable value combination, $i \in \mathrm{I}$. If we had a two-way contingency table, a particular cell could be denoted as (i_1, i_2). Similarly for q-way contingency tables, $i = (i_1, i_2, ..., i_q)$.
$p(i)$	probability than an object belongs to a cell i, $i \in \mathrm{I}$.
$m(i)$	expected frequency in cell i, $i \in \mathrm{I}$. In a two-way contingency table, the expected frequency for a specific cell is denoted as $m(i_1, i_2)$.
m	expected frequency vector, $m' = (m(i))_{i \in \mathrm{I}}$.
$\widehat{m}(i)$	estimated expected frequency in cell i, $i \in \mathrm{I}$.
$n(i)$	observed count in cell i, $i \in \mathrm{I}$. In a two-way contingency table, the observed count for a specific cell is denoted as $n(i_1, i_2)$.
n	observed count vector, $n' = (n(i))_{i \in \mathrm{I}}$.
I_a	for $a \subseteq \Delta$, marginal subvector of I with marginal cells i_a,

	$i_a \in \mathrm{I}_a = \times_{\delta \in a} \mathrm{I}_\delta$.
$\lvert \mathrm{I}_a \rvert$	marginal cells total.
A	generating class= cliques set for the associated graph.
K	set of all subsets of the elements in the generating class A.
$u_a(i_a)$	parameters for a log-linear model depending only on the values of the variables in subset a, i_a, for $a \subseteq \Delta$. $u_a(i_a)$ = constant, when $a = \emptyset$; $u_a(i_a)$ = main effect, when $\lvert a \rvert = 1$; $u_a(i_a)$ = interaction, when $\lvert a \rvert > 1$.
$n_a(i_a)$	$n_a(i_a) = \sum_{j:j_a=i_a} n(j)$, marginal count for i_a.
$m_a(i_a)$	$m_a(i_a) = \sum_{j:j_a=i_a} m(j)$, marginal expected frequency for i_a.
$\widehat{m}_a(i_a)$	$\widehat{m}_a(i_a) = \sum_{j:j_a=i_a} \widehat{m}(j)$, marginal estimated expected frequency for i_a.
$\lvert n \rvert$	Total observed count, $\sum_{i \in \mathrm{I}} n(i)$.
$\perp$	Symbol representing independence. For example $I_X \perp I_Y \lvert I_Z$ or its short notation $X \perp Y \lvert Z$ means that I_X is conditionally independent of I_Y given I_Z.

Observations: Δ is the set of variable names or labels assigned to the vertices of the graph representing the model. When these names or labels are not specified we use Greek letters, *i.e.* when we are analyzing theoretical aspects, but in particular examples we could use numerical values or capital letters. For each $\delta \in \Delta$ we have a random variable, I_δ, so that I is the set of random variables associated to the vertex set. We use I_δ, instead of only δ because once we have an specific cell i, the value i_δ taken by I_δ identifies that the entry for δ we are interested in is for i. If we did not do that, we would not know which cell we are trying to identify. Observe that the level set is denoted as I_δ,where I is not a slanted letter. Observe that a cell, i, is an ordered set whose elements are represented with lower-case letters, for example (j, k, l). In some cases the first subscript a in $u_a(i_a)$, $n_a(i_a)$, $m_a(i_a)$, and $\widehat{m}_a(i_a)$ can be suppressed when it is obvious to which elements in Δ we are referring to.

To illustrate some of the concepts discussed so far, consider the following contingency table, table 2.1, taken from Agresti (2002a, p. 322) based on a survey by the Wright University School of Medicine and the United Health Services in Dayton, Ohio.

Alcohol	**Cigarette**	**Marijuana**	
		Yes	No
Yes	Yes	911	538
	No	44	456
No	Yes	3	43
	No	2	279

Table 2.1: Alcohol, cigarette, and marijuana use for a sample of 2276 high school seniors in a non-urban area near Dayton, Ohio in a 1992 survey.

In this example

$$\Delta = \{A, C, M\},$$

with $A =$ Alcohol, $C =$ Cigarette, and $M =$ Marijuana.

The random variable corresponding to A is I_A and similarly for the other variables. Then

$$I = \{I_A, I_C, I_M\}.$$

All variables have two levels, *Yes* and *No*, then

$$\mathrm{I}_A = \mathrm{I}_C = \mathrm{I}_M = \{\, Yes, No\},$$

and

$$|\mathrm{I}_A| = |\mathrm{I}_C| = |\mathrm{I}_M| = 2.$$

Term I is formed by all cells in the table, it has 8 elements corresponding to all ordered triads. For instance, a cell $i = (I_A = Yes, I_C = Yes, I_M = No)$ where $n(i) = 538$. Similarly, we get $n(i)$ for the other cells, in consequence $n' = (911, 538, 44, 456, 3, 43, 2, 279)$ and the total observed count $|n| = 2276$. Observe that table 2.1 can also be represented as a list in which the cells together with their associated observed counts $n(i)$ are represented as rows as shown in table 2.2.

Alcohol	**Cigarette**	**Marijuana**	$n(i)$
Yes	Yes	Yes	911
Yes	Yes	No	538
Yes	No	Yes	44
Yes	No	No	456
No	Yes	Yes	3
No	Yes	No	43
No	No	Yes	2
No	No	No	279

Table 2.2: Alternative representation of the alcohol, cigarette, and marijuana data presented in contingency table 2.1.

If we take $a = \{A, C\}$, then

$$\mathrm{I}_a = \{(\, Yes, No), (\, Yes,\, Yes), (No,\, Yes), (No, No)\},$$

where the first entry in each ordered pair corresponds to values for I_A and the second to values for I_C. Each entry in I_a is one of the elements i_a mentioned above.

Suppose that we take $i_a = (Yes, Yes)$, then $n_a(i_a) = 911 + 538 = 1449$, because according to the notation we sum the observed values only for those cells in which both variables, alcohol and cigarette, take the value *Yes*. Finally, $p(i)$ and $m(i)$, and as a consequence m, are theoretical values depending on a model.

2.2 Log-linear models

Generally data for contingency tables are collected under three different sampling schemes, see for instance Lauritzen (1996, p. 69-70).

1. Poisson. Under the Poisson sampling scheme the cell counts are realizations of independent Poisson distributed random variables $\{N(i)\}_{i\in \mathrm{I}}$. Therefore $E(N(i)) = m(i)$ and the joint distribution of the counts becomes

$$P(N(i) = n(i), i \in \mathrm{I}) = \prod_{i\in \mathrm{I}} \frac{m(i)^{n(i)}}{n(i)!} \exp(-m(i)), \tag{2.1}$$

   where the sampling scheme restriction is that $m(i) \geq 0$.

2. Multinomial. If the total observed count $|n|$ is fixed, but the counts in each cell are random, and each of these observations is supposed independently to belong to a given cell i with probability $p(i)$, $i \in \mathrm{I}$, with $p(i) \geq 0$ and $\sum_{i\in \mathrm{I}} p(i) = 1$, then the counts follow a multinomial distribution

$$P(N(i) = n(i), i \in \mathrm{I}) = \frac{|n|!}{\prod_{i\in \mathrm{I}} n(i)!} \prod_{i\in \mathrm{I}} p(i)^{n(i)}, \tag{2.2}$$

   where $m(i) = |n|\, p(i)$. Then the sampling scheme restrictions are $m(i) \geq 0$ and $\sum_{i\in \mathrm{I}} m(i) = |n|$.

3. Restricted multinomial distribution with fixed margins. In the case where marginal counts $n(i_b)$ are fixed for each slice i_b, $b \subset \Delta$, we assume the counts in the slices to be independent and multinomially distributed with cell probabilities for cells i_a, obtained using the variables a not contained in b, given a slice i_b equal to $p(i_a|i_b)$, with $p(i_a|i_b) \geq 0$ and $\sum_{i_a} p(i_a|i_b) = 1$. Thus the joint distribution of the table becomes

$$P(N(i) = n(i), i \in \mathrm{I}) = \prod_{i_b\in \mathrm{I}_b} \left\{ \frac{n(i_b)!}{\prod_{i_a\in \mathrm{I}_a} n(i)!} \prod_{i_a\in \mathrm{I}_a} p(i_a|i_b)^{n(i)} \right\}, \tag{2.3}$$

where $m(i) = n(i_b)p(i_a|i_b)$. Then the sampling scheme restrictions are $m(i) \geq 0$ and $m(i_b) = n(i_b)$, for all $i_b \in \mathrm{I}_b$.

To measure the association between variables in an $|\mathrm{I}_1| \times |\mathrm{I}_2|$ contingency table we could use cross-product ratios, *e.g.*, Plackett (1981, p. 28)

$$\psi_{ab} = \left(p(a,b)p(|\mathrm{I}_1|,|\mathrm{I}_2|)\right) / \left(p(a,|\mathrm{I}_2|)p(|\mathrm{I}_1|,b)\right),\ a = 1,2,...,|\mathrm{I}_1|-1; b = 1,2,...,|\mathrm{I}_2|-1;$$

or its logarithm

$$u_{12}(ab) = \log \psi_{ab}.$$

Other useful association measures are the odds:

$$p(a,|\mathrm{I}_2|)/p(|\mathrm{I}_1|,|\mathrm{I}_2|),\ p(|\mathrm{I}_1|,b)/p(|\mathrm{I}_1|,|\mathrm{I}_2|),$$

or their logarithms.

To analyze the association between variables it is useful to have a transformation of the probabilities $p(i,j)$ as a function of other parameters besides $u_{12}(ab)$ as presented by Plackett (1981, p. 30). These parameters are functions of the odds

$$\begin{aligned} u &= \log p(|\mathrm{I}_1|,|\mathrm{I}_2|), \\ u_1(a) &= \log\left[p(a,|\mathrm{I}_2|)/p(|\mathrm{I}_1|,|\mathrm{I}_2|)\right], \\ u_2(b) &= \log\left[p(|\mathrm{I}_1|,b)/p(|\mathrm{I}_1|,|\mathrm{I}_2|)\right]. \end{aligned}$$

Therefore, we have the following relations between probabilities and parameters

$$\begin{aligned} \log p(|\mathrm{I}_1|,|\mathrm{I}_2|) &= u, \\ \log p(a,|\mathrm{I}_2|) &= u + u_1(a), \\ \log p(|\mathrm{I}_1|,b) &= u + u_2(b), \\ \log p(a,b) &= u + u_1(a) + u_2(b) + u_{12}(ab), \end{aligned}$$

with a=1,2,...,$|\mathrm{I}_1| - 1$; b=1,2,...,$|\mathrm{I}_2| - 1$.

These equations correspond to a linear model for the logarithm of the probabilities and it is called a log-linear model. We can see that there is an analogy with ANOVA models, in which the main effects and interactions are defined according to the probabilities in each cell and these probabilities are expressed as linear combinations of these

effects. This can be generalized to any number of variables as we see next, but before note that log-linear models are multiplicative because the linear expansion is with respect to the logarithm of the probabilities. There are other type of models called additive in which the linear expansion is for the probabilities, *e.g.*, Darroch and Speed (1983).

The log-linear expansion of a multinomial distribution is

$$\log p(i) = \sum_{a \subseteq \Delta} u_a(i_a).$$

This is a saturated model and it can also be expressed as a function of the expected frequencies, $m(i)$,

$$\log m(i) = \sum_{a \subseteq \Delta} u_a(i_a). \tag{2.4}$$

This way of writing log-linear models can be applied to any of the three sampling schemes.

Example 2.1. Suppose that we have three variables 1, 2, and 3, the associated saturated log-linear model is

$$\log m(i,j,k) = u + u_1(i) + u_2(j) + u_3(k) + u_{12}(ij) + u_{13}(ik) + u_{23}(jk) + u_{123}(ijk),$$

where i is a level of variable 1, j of variable 2, and k of variable 3, so that $i = 1, 2, ..., |\mathrm{I}_1|$, $j = 1, 2, ..., |\mathrm{I}_2|$, and $k = 1, 2, ..., |\mathrm{I}_3|$.

We can assign the value zero to any parameter $u_a, a \subseteq \Delta$ in (2.4) obtaining different models, which could represent different types of independence, conditional or marginal, between variables according to these assignations. We could also impose additional restrictions on the parameters, for instance letting some of them equal or setting specific parameters to zero to allow a graphical representation of the model.

As in ANOVA models, in log-linear models we have redundant parameters. This means we could use identifiability conditions applied to the parameters by adding parameter constraints. Usually we use the following two parameter constraints.

1. Dummy variables. We choose a reference category for each variable and if these categories are found in i_a for $u_a(i_a)$, then $u_a(i_a)$ takes the value zero.

2. Effect coding. These are ANOVA type restrictions and correspond to restrict the parameters sum over all levels to be zero. For example, in a three variables saturated model, *e.g.*, Fienberg (1980, p. 29), we have

$$\sum_i u_1(i) = \sum_j u_2(j) = \sum_k u_3(k) = 0,$$

$$\sum_i u_{12}(ij) = \sum_j u_{12}(ij) = \sum_i u_{13}(ik) = \sum_k u_{13}(ik) = \sum_j u_{23}(jk) = \sum_k u_{23}(jk) = 0,$$

$$\sum_i u_{123}(ijk) = \sum_j u_{123}(ijk) = \sum_k u_{123}(ijk) = 0,$$

for all i, j, and k.

2.2.1 Hierarchical log-linear models

Given any parameter in a log-linear model, we define lower-order parameters as terms whose set of associated variables is a subset of the set of variables contained in the first parameter. A log-linear model is hierarchical if the model includes all lower-order parameters for each parameter in the model. This is equivalent to say that a log-linear model is hierarchical if $u_a = 0$, then $u_t = 0$, for all $a \subseteq t$.

We can deduce that there is a set A of subsets of Δ used to get all parameters in the model; *i.e.*, the parameters in the model are those formed by variables in b, with b a subset of a for some $a \in A$. The set A is known as the generating class of a model. Denoting as K the set formed by all subsets obtained from A, the hierarchical log-linear model can be written as

$$\log m(i) = \sum_{a \in K} u_a(i_a). \tag{2.5}$$

Graphical log-linear models are particular cases of these models.

2.3 Maximum likelihood estimators in log-linear models

The likelihood function, $L(m)$, under Poisson sampling scheme, expression 2.1, is

$$L(m) = \prod_{i \in \mathrm{I}} \frac{m(i)^{n(i)}}{n(i)!} \exp(-m(i)).$$

Using that $m(i) = |n|\, p(i)$, the likelihood function, $L(m)$, under multinomial sampling scheme, expression 2.2, is

$$L(m) = \frac{|n|!}{\prod_{i\in \mathrm{I}} n(i)!} \prod_{i\in \mathrm{I}} \left(\frac{m(i)}{|n|}\right)^{n(i)}.$$

Finally, under the restricted multinomial sampling scheme, expression 2.3, once fixing the observed count for each configuration i_b of values in $b \subset \Delta$, we have

$$L(m) = \prod_{i_b\in \mathrm{I}_b} \left\{ \frac{n(i_b)!}{\prod_{i_a\in \mathrm{I}_a} n(i)!} \prod_{i_a\in \mathrm{I}_a} \left(\frac{m(i)}{n(i_b)}\right)^{n(i)} \right\}.$$

Observe that the restriction $\sum_{i\in \mathrm{I}} m(i) = |n|$ for a multinomial sampling scheme, means that

$$\exp(-\sum_{i\in \mathrm{I}} m(i)) = \exp(-|n|),$$

so that this value is a constant for any function of m. Something similar happens for the restricted multinomial with fixed margins sampling scheme.

Then, in any of the three sampling schemes once including the different constraints for the expected frequencies, we have (see for instance Lauritzen, 1996, p. 71)

$$L(m) \propto \prod_{i\in \mathrm{I}} m(i)^{n(i)} \exp(-m(i)). \tag{2.6}$$

To get the likelihood equations, we maximize function (2.6) or its logarithm

$$\sum_{i\in \mathrm{I}} n(i) \log m(i) - \sum_{i\in \mathrm{I}} m(i). \tag{2.7}$$

Under any log-linear model, we replace (2.4) in (2.7) getting an expression that depends on the parameters

$$\sum_{a\subseteq\Delta} \sum_{i_a} n_a(i_a) u_a(i_a) - \sum_{i\in \mathrm{I}} \exp(\sum_{a\subseteq\Delta} u_a(i_a)), \tag{2.8}$$

in which some u-parameters could be set to zero corresponding to specific models. For example, for a hierarchical model, equation (2.5), expression (2.7) becomes

$$\sum_{a\in K} \sum_{i_a} n_a(i_a) u_a(i_a) - \sum_{i\in \mathrm{I}} \exp(\sum_{a\in K} u_a(i_a)). \tag{2.9}$$

Deriving (2.9) with respect to each parameter and equating to zero, we get the likelihood equations

$$n_a(i_a) = m_a(i_a), \text{ for all } a \in K; \tag{2.10}$$

where $i_a \in \mathrm{I}_a$. To reduce the number of equations, we may eliminate the redundant ones which could or could not exist depending on the model. To obtain this reduction, we only use the equations corresponding to $a \in A$; for instance, Lauritzen (1996, p. 82) presents in equation (4.22) of theorem **4.11** the equation system

$$n_a(i_a) = m_a(i_a), \text{ for all } a \in A. \tag{2.11}$$

He also states and proves in the same theorem uniqueness of the corresponding maximum likelihood estimates and in Corollary 4.9 he proves existence of such estimates on models without structural zeros, which are entries in a contingency table that are certain to be zero because they correspond to an impossible outcome. He works with an enlarged parameter space under mean value parametrization in which the maximum likelihood estimators are always defined, at least in an extended way, to avoid the existence problems that sampling zeros can create. Using the ordinary parameter space, Haberman (1973) proved that the maximum likelihood estimators always exist when all observed counts are greater than zero and provided some theoretical sufficient and necessary conditions for the existence of the maximum likelihood estimators. For the cases in which there are sampling zeros, existence of maximum likelihood estimates including algorithms for computing extended maximum likelihood estimates, is discussed by Rinaldo (2006a,b) where he determines polyhedral conditions for existence of extended estimates in log-linear models not necessarily hierarchical as well as numerical procedures to calculate such estimates.

For a non-hierarchical model, the a's and K in (2.10) depend on the model and can be obtained by deriving expression (2.8) considering that some u-parameters are zero.

Example 2.2. The model

$$\log m(i,j,k) = u + u_X(i) + u_Y(j) + u_Z(k) + u_{XY}(ij) + u_{XZ}(ik)$$

is an example of a hierarchical log-linear model with generating class $\{\{X,Y\},\{X,Z\}\}$.

Using concepts given for graphical models, we will see in section **2.10** that this model implies that Y is conditionally independent to Z given X, $Y \perp Z|X$. Using conditional independence properties and the classic notation in which a dot in an entry represents a sum over all possible values of the variable whose values are replaced with a dot, we have

$$p(i,j,k) = \frac{p(i,.,k)p(i,j,.)}{p(i,.,.)}$$

or that

$$m(i,j,k) = \frac{m(i,.,k)m(i,j,.)}{m(i,.,.)}.$$

The likelihood equations for the elements in the generating class are

$$\widehat{m}(i,.,k) = n(i,.,k),\ \widehat{m}(i,j,.) = n(i,j,.).$$

From any of these equations we get that

$$\widehat{m}(i,.,.) = n(i,.,.).$$

Then, the estimated expected frequencies for any cell (i,j,k) are

$$\widehat{m}(i,j,k) = \frac{n(i,.,k)n(i,j,.)}{n(i,.,.)}.$$

Eight different hierarchical log-linear models with interesting interpretations for three-way contingency tables including their estimators are presented by Goodman (1970). He also presents 27 interpretable hierarchical log-linear models for four-way contingency tables and one example of a hierarchical log-linear model fitted for a 5-way contingency table.

2.4 Log-linear models expressed as GLM

Log-linear models can be expressed as generalized linear models (GLM), *e.g.*, Agresti (2002a, ch. 4 and ch. 8) or McCullagh and Nelder (1989, ch. 6). Readers not familiar with linear models may refer for instance to Graybill (1961). For log-linear models the GLM takes the form

$$\log(m) = \boldsymbol{X}\boldsymbol{\beta}, \tag{2.12}$$

where:

$N(i)$	random variable corresponding to the observed cell counts for every cell i. $N(i) \sim Poisson(m(i))$.
m	expectation of N, the random vector with entries $N(i)$.
$\boldsymbol{X\beta}$	systematic component.
$\boldsymbol{X}$	design matrix. Every row corresponds to a cell and it has as many columns as parameters included in the model. The entries are 0 or 1 depending on the presence or not of a parameter effect in the model.
$\boldsymbol{\beta}$	parameter vector $(\beta_1, ..., \beta_q)$.
log()	link function connecting the random and systematic components.

The kernel of the likelihood function, the part of the likelihood depending on the parameter m we are interested in, (2.6), written in matrix form and in a way that we identify the exponential family with its elements is

$$L(m) \propto \exp(\sum_{i\in \mathrm{I}} n(i) \sum_j x_{ij}\beta_j - \sum_{i\in \mathrm{I}} \exp(\sum_j x_{ij}\beta_j))$$

$$= \exp(\left\langle (\beta_1, ..., \beta_p), (\sum_{i\in \mathrm{I}} n(i)x_{i1}, ..., \sum_{i\in \mathrm{I}} n(i)x_{ip}) \right\rangle - \sum_{i\in \mathrm{I}} \exp(\sum_j x_{ij}\beta_j))$$

$$= \exp(\langle \boldsymbol{\beta}, t(n)\rangle - \psi(\boldsymbol{\beta})), \tag{2.13}$$

where

$\boldsymbol{\beta} = (\beta_1, ..., \beta_p)$	canonical parameter
$t(n) = (\sum_{i\in \mathrm{I}} n(i)x_{i1}, ..., \sum_{i\in \mathrm{I}} n(i)x_{ip})$	canonical statistics vector.
$\psi(\boldsymbol{\beta}) = \sum_{i\in \mathrm{I}} \exp(\sum_j x_{ij}\beta_j)$	cumulant function.

Using some exponential family properties presented by Lauritzen (1996, Apendix D), who summarizes and presents exact results available for regular exponential models and asymptotic theory of curved exponential families, the maximum likelihood statistic is obtained by equating the canonical statistics with their expectation. The expectation is the gradient vector for the cumulant function. In this case it is formed by the entries $\sum_{i\in \mathrm{I}} m(i)x_{ik}$, for $k = 1, 2, ..., p$. Then, the likelihood equations are

$$\sum_{i\in \mathrm{I}} n(i)x_{ik} = \sum_{i\in \mathrm{I}} m(i)x_{ik}, k = 1, 2, ...p$$

or

$$\boldsymbol{X}'n = \boldsymbol{X}'m.$$

Example 2.3. Suppose that we have a saturated model with two tricotomic variables

$$\log m(i, j) = u + u_1(i) + u_2(j) + u_{12}(ij),\ i, j = 1, 2, 3$$

The associated design matrix is

$$\boldsymbol{X} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The parameter vector $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta}' = (u, u_1(1), u_1(2), u_1(3), u_2(1), u_2(2), u_2(3), u_3(1), u_3(2), u_3(3), u_{12}(11), u_{12}(12), u_{12}(13),$$
$$u_{12}(21), u_{12}(22), u_{12}(23), u_{12}(31), u_{12}(32), u_{12}(33)).$$

The expected frequencies vector, m is

$$m' = (m(1,1), m(1,2), m(1,3), m(2,1), m(2,2), m(2,3), m(3,1), m(3,2), m(3,3)).$$

The canonical statistics vector, $t(n)$, is given by

$$t(n)' = (n(.,.), n(1,.), n(2,.), n(3,.), n(.,1), n(.,2), n(.,3), n(1,1), n(1,2), n(1,3), n(2,1),$$
$$n(2,2), n(2,3), n(3,1), n(3,2), n(3,3)).$$

The cumulant function, $\psi(\boldsymbol{\beta})$, is

$$\psi(\boldsymbol{\beta}) = \sum_{i,j=1,2,3} m(i,j).$$

This equality is because $\sum_j x_{ij}\beta_j$, for $i = 1, ...9$, corresponds to each row of $\log m$ using $\log(m) = \boldsymbol{X}\boldsymbol{\beta}$.

Then, the likelihood equation system for this model is

$$n(.,.) = m(.,.),\ n(1,.) = m(1,.),\ n(2,.) = m(2,.),\ n(3,.) = m(3,.),\ n(.,1) = m(.,1),$$
$$n(.,2) = m(.,2),\ n(.,3) = m(.,3),\ n(1,1) = m(1,1),\ n(1,2) = m(1,2),\ n(1,3) = m(1,3),$$
$$n(2,1) = m(2,1),\ n(2,2) = m(2,2),\ n(2,3) = m(2,3),\ n(3,1) = m(3,1),\ n(3,2) = m(3,2),$$
$$n(3,3) = m(3,3).$$

The matrix $\boldsymbol{X}$ is not invertible. However, it is always possible to get an adequate parametrization of the model whose associated design matrix is invertible. According to the parameter constraints defined in section **2.2**, there are at least two ways of getting invertible matrices:

- Using dummy variables. We create dummy variables for all columns in $\boldsymbol{X}$ corresponding to non-zero parameters. For each column, we assign the value one to all cells (rows) which take the value or values taken by the parameter associated to that column and zero otherwise.

- Effect coding. If we had a variable with k levels, we would get $x_1, x_2, ..., x_{k-1}$ columns corresponding to the main effects, one for each of the first $k-1$ levels. We assign 1 to the level i in x_i, -1 to the reference category, for example the last category k, and zero to the other categories, *e.g.*, Vermunt (2005). We do not need a column for the reference category because its effect is given through the other effects. The process is done for each variable. Interactions are the product of each entry of some x_i´s according to the interaction we are interested in.

For example, using the saturated model presented in example 2.3, the design matriz for the parametrized model using dummy variables is

	u	$u_1(1)$	$u_1(2)$	$u_2(1)$	$u_2(2)$	$u_{12}(11)$	$u_{12}(12)$	$u_{12}(21)$	$u_{12}(22)$
11	1	1	0	1	0	1	0	0	0
12	1	1	0	0	1	0	1	0	0
13	1	1	0	0	0	0	0	0	0
21	1	0	1	1	0	0	0	1	0
22	1	0	1	0	1	0	0	0	1
23	1	0	1	0	0	0	0	0	0
31	1	0	0	1	0	0	0	0	0
32	1	0	0	0	1	0	0	0	0
33	1	0	0	0	0	0	0	0	0

,

and under effect coding the design matrix is

	u	$u_1(1)$	$u_1(2)$	$u_2(1)$	$u_2(2)$	$u_{12}(11)$	$u_{12}(12)$	$u_{12}(21)$	$u_{12}(22)$
11	1	1	0	1	0	1	0	0	0
12	1	1	0	0	1	0	1	0	0
13	1	1	0	−1	−1	−1	−1	0	0
21	1	0	1	1	0	0	0	1	0
22	1	0	1	0	1	0	0	0	1
23	1	0	1	−1	−1	0	0	−1	−1
31	1	−1	−1	1	0	−1	0	−1	0
32	1	−1	−1	0	1	0	−1	0	−1
33	1	−1	−1	−1	−1	1	1	1	1

.

We note that the u-terms can be estimated, for example using the Newton-Raphson method; however, the interpretation of such parameters for assessing the magnitud of effects is fraught with difficulties and confusion because it depends on the parametrization used. To avoid this problem functions which do not vary with different constraints should be used as discussed by Long (1984), these functions are called *estimable*, among those estimable functions we find the expected frequencies, odds ratios and their functions, and the deviance. Some authors, *e.g.*, Agresti (2002a, p.321-324), pretend to interpret parameters by writing them in terms of odds ratios and functions of the odds ratios. Such odds and their functions can be obtained using the estimated expected frequencies; however, these expected frequencies may be estimated without using the u-terms estimates, which means that parameters estimators are not really useful and that they can only be interpreted only in terms of the estimable functions they are estimating. While it is possible to determine what estimable function the parameters are estimating, this task can be extremely difficult in complex models. For all these reasons, interpreting parameters can be misleading and that is why log-linear models are mostly used for testing specific statistical hypothesis related with association, independence, or even symmetry. We emphasize that such statistical hypothesis analysis depend on functions that are estimable: the deviance and the expected frequencies, so that the parametrization used is irrelevant. In graphical log-linear models interpretation of u-terms is not even considered and in all the new models we define in this dissertation the same happens.

2.5 Some useful cases of log-linear models

We present five useful models for two-way contingency tables. We consider tables with entries (i, j) whose counts have one of the distributions: Poisson, multinomial, or restricted multinomial. In the first three subsections we additionally suppose that we have square contingency tables, *i.e.* $|\mathrm{I}| \times |\mathrm{I}|$ tables, and we present three well known models: 1) marginal homogeneity, 2) symmetry, and 3) quasi-symmetry. Finally, in the last subsection we consider $|\mathrm{I}| \times |\mathrm{J}|$ tables and present two additional models: 4) independence and 5) quasi-independence.

2.5.1 Marginal homogeneity model

An $|\mathrm{I}| \times |\mathrm{I}|$ table satisfies a marginal homogeneity model (MH) if

$$m(i, .) = m(., i),\ i = 1, 2, ..., |\mathrm{I}|\,.$$

For a 2×2 table it is equivalent to

$$m(1, 2) = m(2, 1).$$

This model is not a log-linear model. However, we will see in subsection **2.5.3** that we can test an hypothesis to see if a marginal homogeneity model fits some data. It is important to note that, even though this model is not log-linear, we can get a design matrix representing it as Lipsitz *et al.* (1990) have discussed for this case and for the general case in which they have more than two variables.

2.5.2 Symmetry model

Bowker (1948) was the first to propose a test for two-way contingency tables with the same number of categories to see if a table was symmetric. A two-way table satisfies symmetry if

$$m(i,j) = m(j,i), \text{ for all } i \neq j.$$

This equality means the expected frequency in a cell (i,j) above the diagonal is equal to the expected frequency in the cell (j,i) under the diagonal. Clearly, this implies marginal homogeneity. This condition can also be represented using a log-linear model in the following way

$$\log m(i,j) = u + u_1(i) + u_2(j) + u_{12}(ij),$$

with restrictions

$$u_1(i) = u_2(i), \ i = 1, 2, ..., |\mathrm{I}|\,;$$
$$u_{12}(ij) = u_{12}(ji), \ i, j = 1, 2, ..., |\mathrm{I}|\,.$$

Identifiability restrictions are added to the previous restrictions. For example, under effect coding

$$\sum_i u_1(i) = \sum_i u_{12}(ij) = 0.$$

Using the log-linear representation, it is obvious that

$$\log m(i,j) = \log m(j,i).$$

The likelihood equations, which are also obtained in section **3.8**, are

$$\widehat{m}(i,j) + \widehat{m}(j,i) = n(i,j) + n(j,i), \text{ for all } i < j;$$
$$\widehat{m}(i,i) = n(i,i), i = 1, 2, ..., |\mathrm{I}|\,.$$

Then, the expected frequency estimators can be obtained using the exact formula

$$\widehat{m}(i,j) = \frac{n(i,j) + n(j,i)}{2}, \ i, j = 1, 2, ..., |\mathrm{I}|\,.$$

One statistic used to test goodness of fit of a log-linear model is the deviance (section **2.7**). If we denote as H_S to the symmetry model, then in this case the loglikelihood ratio test or deviance $d(H_S)$ has the closed form

$$d(H_S) = 2\sum_{i,j} n(i,j) \log \frac{2n(i,j)}{n(i,j)+n(j,i)}.$$

The Pearson chi-squared statistic (X^2), which is another goodness of fit statistic, for the symmetry model, *e.g.*, Bowker (1948), is

$$X^2(H_S) = \sum_{i>j} \frac{(n(i,j)-n(j,i))^2}{n(i,j)+n(j,i)}.$$

Asymptotically, both statistics tend to a χ^2 distribution with degrees of freedom

$$gl(H_S) = \frac{(|\mathrm{I}|)(|\mathrm{I}|-1)}{2}.$$

2.5.3 Quasi-symmetry model

Quasi-symmetry was first introduced by Caussinus (1965) and it is used to explain cases where there is no symmetry due to marginal heterogeneity, which means that the main effects in the symmetry model differ. A quasi-symmetry model is the log-linear model

$$\log m(i,j) = u + u_1(i) + u_2(j) + u_{12}(ij);$$

with restrictions

$$u_{12}(ij) = u_{12}(ji),\ i,j = 1,2,...,|\mathrm{I}|\,.$$

Under identifiability conditions there are additional restrictions. For example, under effect coding these restrictions are

$$\sum_i u_1(i) = \sum_j u_2(j) = \sum_i u_{12}(ij) = 0.$$

Observe that the symmetry model is a particular case in which we add the restriction

$$u_1(i) = u_2(i),\ i = 1,2,...,|\mathrm{I}|\,.$$

Bishop *et al.* (1975, p. 286-287) and Agresti (2002a, p. 428) show that quasi-symmetry and marginal homogeneity together imply symmetry. Denoting as H_S the symmetry model, as H_{QS} the quasi-symmetry model, and as H_{MH} the marginal homogeneity model, the following relation holds

$$H_S = H_{QS} \cap H_{MH}.$$

Observe that a quasi-symmetry model can be interpreted as a model in which if there was marginal homogeneity, then we would have a symmetric table. However, without this assumption we are not able to say anything about the presence of equalities between certain expected frequencies for any cells as in the symmetry model.

The corresponding likelihood equations for the quasi-symmetry model are

$$\widehat{m}(i,.) = n(i,.),\ i = 1,2,...,|\mathrm{I}|\,;$$
$$\widehat{m}(.,j) = n(.,j),\ j = 1,2,...,|\mathrm{I}|\,;$$
$$\widehat{m}(i,j) + \widehat{m}(j,i) = n(i,j) + n(j,i),\ \text{for all } i \leq j.$$

Observe that

$$\widehat{m}(i,i) = n(i,i),\ i = 1,2,...,|\mathrm{I}|\,.$$

Then, we have a perfect fit for the diagonal cells. For the other cells there is no exact solution so that to solve the equations involving them we use iterative methods. For instance, Haberman (1979, p. 491-493) proposed an iterative method for these models.

The deviance, $d(H_{QS})$, and Pearson statistic, $X^2(H_{QS})$, do not have a closed form in terms of $n(i,j)$ for this model, and they are obtained using the general formula for the deviance (equation (2.14)), using the estimated frequencies, $\widehat{m}(i,j)$. The number of degrees of freedom associated to both statistics is

$$gl(H_{QS}) = \frac{(|\mathrm{I}|-1)(|\mathrm{I}|-2)}{2}.$$

As a symmetry model is equivalent to simultaneously having quasi-symmetry and marginal homogeneity, we have that when a quasi-symmetry model fits well then marginal homogeneity is equivalent to symmetry. Then, conditioned to having quasi-symmetry, testing marginal homogeneity is equivalent to testing symmetry. This means that the statistic for testing marginal homogeneity, $d(H_{MH})$, *e.g.*, Bishop *et al.* (1975, p. 293-294), is

$$d(H_{MH}) = d(H_S|H_{QS}) = d(H_S) - d(H_{QS}),$$

with degrees of freedom

$$gl(H_{MH}) = gl(H_S) - gl(H_{QS}) = \frac{|\mathrm{I}|\,(|\mathrm{I}|-1)}{2} - \frac{(|\mathrm{I}|-1)(|\mathrm{I}|-2)}{2} = |\mathrm{I}| - 1.$$

Quasi-symmetry models have been linked with Rasch models. Tjur (1982) showed that for binary variables estimates of main effects parameters in the quasi-symmetry model are also conditional maximum likelihood estimates of item parameters when

fixed effects corresponding to the subjects in the Rasch item response model are used, conditioning on the item parameters sufficient statistics. More general Rasch models and their association with quasi-symmetry models have been studied by Agresti (2002b). Bradley-Terry model for paired preferences, *e.g.*, Agresti (2002a, p. 436-439) and Fienberg and Larntz (1976), can also be fitted using quasi-symmetry models. Quasi-symmetry applications to particular matched pairs models and connexions with reversibility of Markov chains are discussed by McCullagh (1982). Finally, we observe that the quasi-symmetry model shown here is based on the concept defined by Caussinus (1965); however, Goodman (2002) proposes an alternative Pearson's quasi-symmetry, but at the end he shows it is less acceptable than Caussinus' concept.

2.5.4 Independence and quasi-independence models

For an $|\mathrm{I}|\times|\mathrm{J}|$ table and for any of the three distributions, under statistical independence the expected frequencies $m(i,j)$ can be written as

$$m(i,j) = \mu\alpha_i\beta_j,\ i = 1,2,...,|\mathrm{I}|\,,\ j = 1,2,...,|\mathrm{J}|\,,$$

with α_i, β_j, and μ positive constants satisfying $\sum_i \alpha_i = \sum_j \beta_j = 1$. Writing the expected frequencies as a log-linear model, the following holds

$$\log m(i,j) = u + u_1(i) + u_2(j),\ i = 1,2,...,|\mathrm{I}|\,,\ j = 1,2,...,|\mathrm{J}|\,.$$

The likelihood equations for this model are

$$\widehat{m}(i,.) = n(i,.),\ i = 1,2,...,|\mathrm{I}|\,;$$
$$\widehat{m}(.,j) = n(.,j),\ j = 1,2,...,|\mathrm{J}|\,.$$

As in example 2.2, using the definition of independence we have

$$\widehat{m}(i,j) = \frac{n(i,.)n(.,j)}{|n|}.$$

The deviance can be obtained by replacing the values $\widehat{m}(i,j)$ in the general formula given in equation (2.14). The degrees of freedom for this model are $(|\mathrm{I}|-1)(|\mathrm{J}|-1)$.

In $|\mathrm{I}|\times|\mathrm{I}|$ square tables, we sometimes have certain dependence expressed by the fact that the observed counts in the diagonal cells are greater than the values predicted by the independence model. This means that the relation between variables for ordered pairs outside the diagonal can be expressed using an independence simple structure. This is called a quasi-independence model, denoted as H_{QI}, and can be written as the log-linear model

$$\log m(i,j) = u + u_1(i) + u_2(j) + \delta_i I(i=j),$$

with $I(i=j)=1$ if $i=j$, and 0 if $i \neq j$.

Then, we are adding a parameter to the independence model for each cell in the main diagonal. If $\delta_i > 0$, then $m(i,i)$ is greater than under the independence model.

The likelihood equations for this model, *e.g.*, Agresti (2002a, p. 426), are

$$\widehat{m}(i,.) = n(i,.),\ i = 1,2,...,|\mathrm{I}|\,,$$
$$\widehat{m}(.,j) = n(.,j),\ j = 1,2,...,|\mathrm{I}|\,,$$
$$\widehat{m}(i,i) = n(i,i),\ i = 1,2,...,|\mathrm{I}|\,.$$

Observe that there is a perfect fit for the diagonal cells and independence for the other cells. This model has |I| more parameters than the independence model and, as we will see in section **2.7**, the degrees of freedom are the number of non-zero independent parameters. Then to get the degrees of freedom, we subtract to the independence model degrees of freedom, |I| parameters that are not zero

$$gl(H_{QI}) = (|\mathrm{I}|-1)^2 - |\mathrm{I}|\,.$$

Example 2.4. To illustrate the fit of all five models given in this section to some data, we use a table corresponding to region of residence change between 1980 and 1985 of U.S. residents (table 2.3) presented and studied by Agresti (2002a, p. 423) and based on data presented by the U.S. Bureau of the Census.

Residence in 1980	**Residence in 1985**			
	Northeast	Midwest	South	West
Northeast	11607	100	366	124
Midwest	87	13677	515	302
South	172	225	17819	270
West	63	176	286	10192

Table 2.3: Region of residence, 1980 and 1985, for a sample of 55,981 U.S. residents.

We fitted the independence, quasi-independence, marginal homogeneity, quasi-symmetry, and symmetry models for this table using *S-Plus 6.1*. The estimated fitted expected frequencies are shown in table 2.4.

The quality of the fit is measured using the deviance (table 2.5). This statistic is compared with a χ^2 quantile whose degrees of freedom depend on the model and using a given significance level α (for example, α= 0.05). If the deviance is greater than this

Residence in 1980	**Residence in 1985**			
	Northeast	Midwest	South	West
Northeast	2599.09[1]	3089.07	4136.62	2372.25
	11607.00[2]	126.56	312.91	150.53
	11607.00[3]	98.08	265.69	93.97
	11607.00[4]	93.50	269.00	93.50
	11607.00[5]	95.79	370.44	123.77
Midwest	3107.10	3692.85	4945.15	2835.92
	117.39	13677.00	531.11	255.50
	88.73	13677.00	379.07	232.35
	93.50	13677.00	370.00	239.00
	91.21	13677.00	501.68	311.11
South	3939.23	4681.86	6269.55	3595.43
	133.23	243.80	17819.00	289.97
	276.47	350.80	17819.00	287.26
	269.00	370.00	17819.00	278.00
	167.56	238.32	17819.00	261.12
West	2283.72	2714.25	3634.69	2084.41
	71.39	130.64	322.98	10192.00
	92.53	251.27	269.77	10192.00
	93.50	239.00	278.00	10192.00
	63.23	166.89	294.88	10192.00

Table 2.4: Fitted expected frequencies under independence, quasi-independence, marginal homogeneity, symmetry, and quasi-symmetry for the data corresponding to U.S. migration between 1980 and 1985. From top to bottom we have the fitted values under [1]*independence,* [2]*quasi- independence,* [3]*marginal homogeneity,* [4]*symmetry, and* [5]*quasi-symmetry.*

quantile, we reject the null hypothesis that the model fits the data.

If we label the variable corresponding to region of residence in 1980 as R_1 and the one corresponding to 1985 as R_2, then the independence model is

$$\log m(i,j) = u + u_{R_1}(i) + u_{R_2}(j),\ i,j = 1,...,4.$$

According to the deviance, we see that the independence model fits bad because we reject the null hypothesis stating that the independence model fits the data. This means that there is not independence between region of residence in 1980 and 1985, so that the variables are in some way associated.

The quasi-independence model is

Model	Deviance	Degrees of freedom	Quantile (α=0.05)
Independence	125923.30	9	16.92
Quasi-independence	69.51	5	11.07
Marginal homogeneity	240.75	3	7.81
Symmetry	243.55	6	12.59
Quasi-symmetry	2.99	3	7.81

Table 2.5: Deviance, degrees of freedom, and quantiles for different models fitted in table 2.3.

$$\log m(i,j) = u + u_{R_1}(i) + u_{R_2}(j) + \delta_i I(i=j),$$

with $I(i=j)=1$ if $i=j$, and 0 if $i \neq j$, i,j=1,...,4.

This model indicates that the variables are independent, except for the diagonal cells of the table, and means that for the people that moved out, the region of residence in 1985 is independent of the region of residence in 1980. This model has a better fit to the data; however, we still reject the null hypothesis stating that the model fits the data.

The marginal homogeneity model is given by the conditions

$$m(i,.) = m(.,i),\ i = 1,2,....,4,$$

and indicates that the proportions or expected frequencies in each category for both variables are equal. In this case, it means that the expected number of residents in each region is the same for both years. We also reject this model.

The symmetry model is as follows:

$$\log m(i,j) = u + u_{R_1}(i) + u_{R_2}(j) + u_{R_1R_2}(ij),\ i,j = 1,...,4;$$

with restrictions

$$u_{R_1}(i) = u_{R_2}(i),\ i = 1,...,4,$$
$$u_{R_1R_2}(ij) = u_{R_1R_2}(ji)\ i,j = 1,...,4.$$

It means that the expected frequency in a cell (i,j) is equal to the one in cell (j,i). In this case, it would mean that the number of people emigrating from i in 1980 to j in 1985 is the same number of people who emigrate from j in 1980 to i in 1985. This means that between regions we have a closed population, so that the number of people moving in is the same number of people moving out. However, this model does not fit the data.

Finally, the quasi-symmetry model is

$$\log m(i,j) = u + u_{R_1}(i) + u_{R_2}(j) + u_{R_1R_2}(ij),\ i,j = 1,...,4;$$

with additional restriction

$$u_{R_1R_2}(ij) = u_{R_1R_2}(ji),\ i,j = 1,...,4.$$

This model fits well the data. It means that the lack of symmetry in the table is caused by some marginal heterogeneity, *i.e.* that the number of people emigrating from region i in 1980 to region j in 1985 would be similar to the one entering to i in 1985 from j in 1980 if the number of people in each region for 1980 were the same that the number of people in the same region in 1985.

We have said that a way of proving the convenience of using marginal homogeneity given that a quasi-symmetry model fits some data is by getting the difference in the deviances between the symmetry and quasi-symmetry models, and comparing it with a χ^2 quantile with degrees of freedom given by the difference between the corresponding degrees of freedom. In this case, we have that this statistic is $243.55 - 2.99 = 240.56$ with $6 - 3 = 3$ degrees of freedom and the corresponding χ^2 quantile using $\alpha = 0.05$ is 7.81. Then we confirm that the marginal homogeneity model does not fit well.

2.6 Some symmetry, quasi-symmetry, and marginal homogeneity generalizations

There are several ways of generalizing the models given in the previous section for contingency tables with more than two dimensions. We discuss some of them based on the work presented by Andersen (1991, p. 328-329), Bishop *et al.* (1975, p. 299-309), and Haberman (1979, p. 503-509)).

In a d-dimensional table we can examine the c-dimensional ($c = 1,...,d-1$) marginal values to get different kinds of marginal homogeneity. For example, for a four-dimensional table with |I| levels on each variable, Bishop *et al.* (1975, p. 299) based on the work presented by Kullback (1971) consider the following three types of marginal homogeneity models:

1. $m(i,.,.,.)$=$m(.,i,.,.)$=$m(.,.,i,.)$=$m(.,.,.,i)$; $i = 1,...,|\mathrm{I}|$.
2. $m(i,j,.,.)$=$m(i,.,j,.)$=$m(i,.,.,j)$=$m(.,i,j,.)$=$m(.,i,.,j)$=$m(.,.,i,j)$; $i,j = 1,...,|\mathrm{I}|$.
3. $m(i,j,k,.)$=$m(i,j,.,k)$=$m(i,.,j,k)$=$m(.,i,j,k)$; $i,j,k = 1,...,|\mathrm{I}|$.

Model 3) implies model 2) and model 2) implies model 1).

In three-dimensional tables Bishop *et al.* (1975, p. 299-309) and Andersen (1991, p. 328-329) studied different marginal homogeneity, symmetry, and quasi-symmetry generalizations.

Suppose that we have a three dimensional table $|\mathrm{I}| \times |\mathrm{I}| \times |\mathrm{K}|$, Andersen (1991, p. 328) and Bishop *et al.* (1975, p. 299-300) define a symmetry model conditional to the levels of variable 3 as a model in which

$$m(i,j,k) = m(j,i,k),\ i,j = 1,...,|\mathrm{I}|\,;\ k = 1,...,|\mathrm{K}|\,.$$

This means that for each k we have a bi-dimensional symmetry model. Then, the analysis of this model can be done by partitioning the model in $|\mathrm{K}|$ symmetry models. However, the model can also be written as

$$\log m(i,j,k) = u + u_1(i) + u_2(j) + u_3(k) + u_{12}(ij) + u_{13}(ik) + u_{23}(jk) + u_{123}(ijk)$$

with restrictions

$$u_{123}(ijk) = u_{123}(jik),\ u_{13}(ik) = u_{23}(jk),\ u_{12}(ij) = u_{12}(ji),\ u_1(i) = u_2(i);$$

$$i,j = 1,...,|\mathrm{I}|\,;\ k = 1,...,|\mathrm{K}|\,.$$

A *conditional marginal homogeneity model* is defined by Bishop *et al.* (1975, p. 300) as a model satisfying

$$m(i,.,k) = m(.,i,k)\ i = 1,...,I;\ k = 1,...,|\mathrm{K}|\,.$$

This model can be fitted using $|\mathrm{K}|$ marginal homogeneity models, but does not have an expression in terms of a log-linear model.

Other models proposed by Andersen (1991, p. 329) are those which constrain marginal values. The *complete marginal symmetry* model between variables 1 and 2 assumes that

$$m(i,j,.) = m(j,i,.),\ i,j = 1,...,|\mathrm{I}|\,.$$

On the other hand, the *marginal quasi-symmetry model* between variables 1 and 2 supposes that

$$\log m(i,j,k) = u + u_1(i) + u_2(j) + u_3(k) + u_{12}(ij) + u_{13}(ik) + u_{23}(jk),$$

adding the restrictions

$$u_{12}(ij) = u_{12}(ji),\ i, j = 1, ..., |\mathrm{I}|\,.$$

For three dimensional tables $|\mathrm{I}| \times |\mathrm{I}| \times |\mathrm{I}|$ other types of general models were proposed and studied by Bishop *et al.* (1975, p. 299-306) and they are presented in the rest of the section.

Complete symmetry models

Bishop *et al.* (1975, p. 299-309) defined complete symmetry based on an *interchangeability* concept provided by Madansky (1963). Complete symmetry models are those in which the expected frequencies are invariant to any permutation of the indices characterizing the levels taken by the variables, as follows:

$$m(i,j,k) = m(i,k,j) = m(j,i,k) = m(j,k,i) = m(k,j,i) = m(k,i,j),\ i, j, k = 1, ..., |\mathrm{I}|\,.$$

Written as a log-linear model

$$\log m(i,j,k) = u + u_1(i) + u_2(j) + u_3(k) + u_{12}(ij) + u_{13}(ik) + u_{23}(jk) + u_{123}(ijk),$$

with restrictions

$$u_1(i) = u_2(i) = u_3(i),\ i = 1, ..., |\mathrm{I}|\,;$$
$$u_{12}(ij) = u_{12}(ji) = u_{13}(ij) = u_{23}(ij),\ i, j = 1, ..., |\mathrm{I}|\,;$$
$$u_{123}(ijk) = u_{123}(ikj) = u_{123}(jik) = u_{123}(jki) = u_{123}(kji) = u_{123}(kij),\ i, j, k = 1, ..., |\mathrm{I}|\,.$$

Marginal homogeneity models

A complete symmetry model implies the following types of marginal homogeneity models, see Bishop *et al.* (1975, p-303):

1. $m(i,.,.) = m(.,i,.) = m(.,.,i);\ \ i = 1, ..., |\mathrm{I}|.$
2. $m(i,j,.) = m(j,i,.) = m(i,.,j) = m(.,i,j);\ \ i, j = 1, ..., |\mathrm{I}|.$

Observe that model 2) implies model 1).

Quasi-symmetry models

To get general quasi-symmetry models similar to the one defined in section **2.5.3**, we could use models analogous to complete symmetry, but without satisfying some of the marginal homogeneity conditions derived from this model. Bishop *et al.* (1975, p. 303) proposed two types of quasi-symmetry models for $|\mathrm{I}| \times |\mathrm{I}| \times |\mathrm{I}|$ tables. The first type, called complete quasi-symmetry, is a model in which neither the marginal homogeneity condition 1) nor the condition 2) are satisfied. In the second kind, the marginal homogeneity condition 1) does not hold.

1. Complete quasi-symmetry. This model preserves the two-dimensional margins and as a consequence the one-dimensional margins too. This means that there is symmetry only for the $u_{123}(ijk)$ parameters. Then, we have the model

$$\log m(i,j,k) = u + u_1(i) + u_2(j) + u_3(k) + u_{12}(ij) + u_{13}(ik) + u_{23}(jk) + u_{123}(ijk),$$

with restrictions

$$\begin{aligned} u_{123}(ijk) &= u_{123}(ikj) = u_{123}(jik) &&= u_{123}(jki) \\ &= u_{123}(kji) = u_{123}(kij) && i,j,k = 1,...,|\mathrm{I}|\,. \end{aligned}$$

2. Quasi-symmetry preserving only one-dimensional margins. In these models there is symmetry in both $u_{12}(ij)$ and $u_{123}(ijk)$ parameters. Then, we have the model

$$\log m(i,j,k) = u + u_1(i) + u_2(j) + u_3(k) + u_{12}(ij) + u_{13}(ik) + u_{23}(jk) + u_{123}(ijk),$$

adding the restrictions

$$\begin{aligned} u_{123}(ijk) &= u_{123}(ikj) = u_{123}(jik) &&= u_{123}(jki) \\ &= u_{123}(kji) = u_{123}(kij) && i,j,k = 1,...,|\mathrm{I}|\,; \end{aligned}$$

and

$$u_{12}(ij) = u_{12}(ji) = u_{13}(ij) = u_{23}(ij),\ i,j = 1,...,|\mathrm{I}|\,.$$

For tables with more than three dimensions Darroch and Bhapkar (1990) extended complete symmetry by defining models in which $m(i) = m(\rho i)$, where i is a cell in a specific order, for example $(i_1, ..., i_s)$ if we have s variables, and ρi denotes all possible permutations of i. They also extended quasi-symmetry models by defining a generalized interaction operation from where they get log-linear generalized A-interactions, defining models Q_k^s, where $k < s$ in which they equate all interactions of order greater or equal than k accordingly to a set of permutations ρ, so that they are restricting interactions in a saturated model. Complete quasi-symmetry and quasi-symmetry preserving one dimensional margins can be recovered as the models Q_2^3 and Q_1^3, respectively.

There are other extensions of symmetry models, for instance Haberman (1979, p. 503-509) presents an extension for a specific problem with four variables in which it is convenient to fit a hierarchical model equating some specific first-order interactions. This kind of model could be a way of extending the symmetry concept to more dimensions. Goodman (1985) suggests other symmetry and independence generalizations for two-dimensional tables. These are for example models in which we would have symmetry except for the fact that the elements above the diagonal contain one additional parameter and the same happens for the elements below the diagonal.

2.7 Some goodness of fit statistics

Suppose that we have a hierarchical log-linear model M_1 contained or nested in another hierarchical log-linear model M_0. The deviance between these models is the likelihood ratio test statistic between model M_1 and M_0. Denoting the maximum likelihood estimators for the expected frequencies as $\widehat{m}_1$ and $\widehat{m}_0$, respectively, and the saturated model estimator as $\widehat{m}$, the estimated deviance is given by

$$d_{01} = -2 \log \frac{L(\widehat{m}_1)}{L(\widehat{m}_0)}.$$

In particular, the deviance between model k, with $k = 0, 1$, and the saturated model is called residual deviance and it is denoted as d_k

$$d_k = -2 \log \frac{L(\widehat{m}_k)}{L(\widehat{m})}.$$

We saw in section **2.3**, expression (2.7), that the logarithm of the kernel of the likelihood, which is a function of the expected frequencies m such that $L(m)$ is proportional to it, is

$$\sum_{i \in \mathrm{I}} n(i) \log m(i) - \sum_{i \in \mathrm{I}} m(i),$$

and we also saw that the maximum likelihood estimators satisfy $n(i_a) = \widehat{m}(i_a)$, where $i_a \in \mathrm{I}_a$, for all $a \in A$, the generating class (equation (2.11)). In particular, for a saturated model $n(i) = \widehat{m}(i)$. Using these results the residual deviance is

$$d_k = -2\log\frac{L(\widehat{m}_k)}{L(\widehat{m})} = 2\sum_{\{i|n(i)>0\}} n(i)\log\frac{n(i)}{\widehat{m}_k(i)} + 2\sum_{\{i|n(i)>0\}}\{\widehat{m}_k(i) - n(i)\}$$

$$= 2\sum_{\{i|n(i)>0\}} n(i)\log\frac{n(i)}{\widehat{m}_k(i)}. \tag{2.14}$$

For the last equality we use that $\sum_i \widehat{m}_k(i) = \sum_i n(i)$, which is a consequence of the fact that $\sum_i \widehat{m}_k(i) = \sum_{i_a} \widehat{m}_k(i_a)$, $n(i_a) = \widehat{m}_k(i_a)$, and $\sum_i n(i) = \sum_{i_a} n(i_a)$. In all previous results i takes its values over the set $\{i|n(i) > 0\}$, because in this way all logarithms are feasible.

The deviance between two models is

$$d_{01} = d_1 - d_0 = 2\sum_i n(i)\log\frac{\widehat{m}_0(i)}{\widehat{m}_1(i)}.$$

The residual deviance can be approximated using the Pearson X^2 statistic. To prove this fact we use the Taylor expansion of the function

$$f(x) = x\log(x/a)$$

around a

$$f(x) = x - a + (x-a)^2/(2a) + o((x-a)^2).$$

We get the Taylor expansion of

$$n(i)\log\frac{n(i)}{\widehat{m}_k(i)}$$

around $\widehat{m}_k(i)$ to approximate d_k

$$d_k \approx 2\sum_i\{n(i) - \widehat{m}_k(i)\} + 2\sum_i \frac{(n(i) - \widehat{m}_k(i))^2}{2\widehat{m}_k(i)},$$

and using that $\sum_i \widehat{m}_k(i) = \sum_i n(i)$, we have that

$$d_k \approx \sum_i \frac{(n(i) - \widehat{m}_k(i))^2}{\widehat{m}_k(i)} = X^2.$$

Using the deviance, we test hypothesis for the models M_k, $k = 0, 1$, in which the null hypothesis H_0 indicates that the parameters not present in the hierarchical model

M_k but found in the saturated one are zero; *i.e.*, the model fits the data, against the alternative hypothesis (H_1) indicating that the parameters not found in the hierarchical model M_k but present in the saturated one are not zero.

Under the null hypothesis that the model M_k, with $k = 0, 1$ fits the data, the residual deviance has an asymptotic χ^2 distribution with degrees of freedom (df) given by the number of independent parameters equal to zero with respect to the saturated model, see *e.g.* Bishop *et al.* (1975, sec. 3.8) or specifically for graphical models see Whittaker (1990, p. 216).

Example 2.5. Consider a log-linear model with generating class $\{\{1,3\}, \{2,3\}\}$

$$\log m(i,j,k) = u + u_1(i) + u_2(j) + u_3(k) + u_{13}(ik) + u_{23}(jk),$$
$$i = 1, ..., |\mathrm{I}_1|\,;\ j = 1, ..., |\mathrm{I}_2|\,;\ k = 1, ..., |\mathrm{I}_3|\,.$$

In this example $u_{12} = u_{123} = 0$, with $(|\mathrm{I}_1| - 1)(|\mathrm{I}_2| - 1)$ independent parameters for u_{12} and $(|\mathrm{I}_1| - 1)(|\mathrm{I}_2| - 1)(|\mathrm{I}_3| - 1)$ independent parameters for u_{123}. As a consequence the total number of degrees of freedom is $(|\mathrm{I}_1| - 1)(|\mathrm{I}_2| - 1)(|\mathrm{I}_3|)$.

Another way of getting the number of degrees of freedom is by subtracting the number of cells minus the number of independent parameters in the model; for example, the number of parameters that under a parametrization using dummy variables are not set to zero. This is discussed by Agresti (2002a, p. 338) for three dimensional tables and by Bishop *et al.* (1975, p. 114-115, p. 119-122) for tables with even more dimensions.

Considering the log-linear model as a generalized linear model with design matrix X, the degrees of freedom can also be calculated as the total number of cells in the contingency table minus the rank of $\boldsymbol{X}$. This is because the number of linearly independent columns in $\boldsymbol{X}$ is equal to the number of independent parameters in the model. Then

$$df = |\mathrm{I}| - Rank(\boldsymbol{X}).$$

Observe that the rank of $\boldsymbol{X}$ is the same regardless of the parametrization used. Then, even though we do not parametrize a model using effect coding or dummy variables, we still calculate the number of degrees of freedom using the rank of the design matrix.

To get an adequate approximation to the χ^2 distribution some regularity conditions are required. For multinomial distributions, we need the parameters to be outside the boundary of the parametral space (Whittaker, 1990, p. 216), which means that we ideally require the expected frequencies $m(i)$ to be greater than zero, that is, we do

not want structural zeros. This condition does not always happens, but we will assume throughout this dissertation that we do not have structural zeros. When we have them, the contingency table needs to satisfy additional conditions to guarantee the existence of maximum likelihood estimates and we have to do changes in the iterative procedures used to fit the models and in the calculation of number of degrees of freedom as presented by Bishop *et al.* (1975, ch. 5) and recently discussed by Fienberg and Rinaldo (2007).

However, not even having structural zeros we might have sampling zeros, that is $n(i) = 0$ for some i. In this case it may or may not exist a maximum likelihood estimator, Haberman (1973) gave necessary and sufficient conditions for the existence; however, he did not give numerical procedures to know when the estimator exists. He proved that the estimators exist when $n(i) > 0$ for all i and also justified in some extent that in practice people tend to add small positive quantities to zero cells to obtain estimators. Bishop *et al.* (1975, p. 69) and Glonek *et al.* (1988) also discussed conditions ensuring the existence of maximum likelihood estimators, in general terms they correspond to say that positivity of the margins defining the likelihood equations is a necessary and sufficient condition for existence of maximum likelihood estimators if and only if the model is decomposable, decomposable models are defined in section **2.10**. Bishop *et al.* (1975, p. 115) and Christensen (1997, p. 286-293) suggest that even having zero margins we could fit a model by dropping all cells causing zero margins, then some kind of extended estimators are fitted for the other cells but we have to adjust the degrees of freedom. By adjusting the degrees of freedom, the approximation to the χ^2 distribution is improved. In Chapter 5 we discuss further existence of estimators.

To adjust the degrees of freedom, we get the number of cells with zero estimates z_e and the number of parameters z_p that can not be estimated. The adjusted degrees of freedom df' as presented by Bishop *et al.* (1975, p. 115) are

$$df' = df - z_e + z_p.$$

For example, suppose that we have a $2 \times 3 \times 2$ table with $n(1,1,1) = n(1,1,2) = 0$ and $n(i,j,k) > 0$ for all other i, j, and k. If we have a model in which $\{1,2\}$ is part of the generating class, then using the likelihood equations

$$\widehat{m}_{\{1,2\}}(1,1) = \sum_{j:(j_1,j_2)=(1,1)} \widehat{m}(j) = \widehat{m}(1,1,\cdot) = n(1,1,\cdot) = 0,$$

so that $\widehat{m}(1,1,1) = \widehat{m}(1,1,2) = 0$, which means we have two cells with zero estimates, *i.e.* $z_e = 2$. On the other hand, there is one u_{12} parameter corresponding to the categories $(1,1)$ which can not be estimated because this configuration for the first two variables is an empty cell in the corresponding marginal table, then $z_p = 1$ and $df' = df - 2 + 1$.

The deviance between M_0 and M_1, d_{01}, has asymptotically a χ^2 sampling distribution whose degrees of freedom are given by the difference in number of free parameters between M_0 and M_1 (Edwards, 2000, p. 21).

In conclusion, the residual deviance is used to prove the null hypothesis that the parameters not contained in a model are zero, so that if we do not reject the null hypothesis, then the model fits the data. To test it, we compare the residual deviance with a quantile from a χ^2 distribution using a significance level α. If the residual deviance is small with respect to this quantile, then we have that the saturated and proposed model are not too different in terms of what they explain or that the parameters which are not in the model are zero, not rejecting the null hypothesis that the model fits the data. This means that we reject the null hypothesis if

$$d_k \geq \chi^{2\,(1-\alpha)}_{df}.$$

Where $k = 0, 1$, α is a significance level, $\chi^{2\,(1-\alpha)}_{df}$ is a $1-\alpha$ quantile from the χ^2 distribution with degrees of freedom df.

Example 2.6. To illustrate how we calculate the number of degrees of freedom, consider a symmetry model for two variables

$$\log m(i,j) = u + u_1(i) + u_2(j) + u_{12}(ij), i, j = 1, 2, ..., |\mathrm{I}|,$$

with restrictions

$$u_1(i) = u_2(i),\ i = 1, 2, ..., |\mathrm{I}|,$$
$$u_{12}(ij) = u_{12}(ji),\ i, j = 1, 2, ..., |\mathrm{I}|.$$

Consider a parametrization of the model with dummy variables in such a way that every time category 1 appears on a u parameter the parameter is zero. The parameters in the model are: A constant u, $|\mathrm{I}| - 1$ parameters corresponding to the main effects, they are obtained by taking on account the parametrization and that $u_1(i) = u_2(i), i = 1, 2, ..., |\mathrm{I}|$, $\binom{|\mathrm{I}|-1}{2}$ $u_{12}(ij)$ parameters, with $i \neq j$, which are obtained using $u_{12}(ij) = u_{12}(ji)$ and $u_{12}(1j) = u_{12}(j1) = 0$. Finally, we have $|\mathrm{I}| - 1$ $u_{12}(ii)$ parameters. Then, the total number of parameters is

$$1 + |\mathrm{I}| - 1 + \binom{|\mathrm{I}| - 1}{2} + |\mathrm{I}| - 1 = 2\,|\mathrm{I}| - 1 + (|\mathrm{I}| - 1)(|\mathrm{I}| - 2)/2.$$

From where the number of degrees of freedom for the χ^2 distribution is the number of cells, $|\mathrm{I}|^2$, minus the number of independent parameters in the model

$$df = |\mathrm{I}|^2 - (2\,|\mathrm{I}| - 1 + (|\mathrm{I}| - 1)(|\mathrm{I}| - 2)/2) = (|\mathrm{I}|)(|\mathrm{I}| - 1)/2.$$

2.8 Graphical log-linear models

2.8.1 Definition and representation

The use of graphs to represent complex models is a useful tool whose origin can be traced back to many scientific areas like genetics with the so-called path analysis or physics. However, graphical models were formally introduced and studied by Darroch *et al.* (1980) and by Lauritzen and Wermuth (1989). These models are characterized by simultaneously using concepts from both graph theory and statistics. On the graphs corresponding to these models, we represent variables with vertices or nodes and the association between them with different variations of edges, which are lines joining the vertices, according to the model. One variation could be for example using directed edges or arcs. Here we are interested in graphical models with non-directed graphs and they are used to identify and represent conditional and marginal independences and associations between the variables involved in a problem. Graphical models with directed graphs are called *Bayesian networks* and are further discussed for instance in Cowell *et al.* (1999) and we studied them, including one application in medical research, in Ramírez-Aldana (2005).

There are three kinds of graphical models for non-directed graphs according to the data type: a) graphical log-linear models, they are used for discrete data and are the models we are interested in and study in the rest of this chapter, b) graphical Gaussian models, they are used for continuous data, and c) mixed models, they are a family of models that combine and generalize the two previous cases and were introduced by Lauritzen and Wermuth (1989) and extended to a broader class, the hierarchical interaction models by Edwards (2000, ch. 4).

We give a brief introduction to graphical Gaussian models (Whittaker, 1990, ch. 6) or covariance selection models (Dempster, 1972) because we will see in this document that there are similar aspects between the discrete and continuous case. These are models based on multivariate Gaussian observations defined by setting specified elements of the concentration matrix, the inverse of the covariance matrix, to zero. We join two nodes of the graph with an edge if the entries corresponding to those two variables in the concentration matrix are not set to zero. This means that when there is not an edge between two variables, the variables are independent given the others, see for instance Edwards (2000, p. 36).

Graphical log-linear models have been further studied and discussed by Christensen (1997, ch. 5.2), Edwards (2000, ch. 2), Lauritzen (1996, section **4.3**), and recently by Agresti and Gottard (2007), as part of hierarchical log-linear models. The associated contingency tables for these models have a multinomial or Poisson distribution, and, in a similar way than in the continuous case, when there is not an edge between two

variables they are independent given the other variables. Interaction graphs are defined here as a way to represent hierarchical log-linear models with generating class A, this definition is based on the ones given by Lauritzen (1996, p. 88) and Whittaker (1990, p. 209). However, we have that many hierarchical log-linear models could have the same graph so that the graph is not unique and it does not give a complete picture of the structure of the model. To solve these problems, a subclass of hierarchical log-linear models called graphical is defined. These are models defined solely in terms of conditional independence relationships depending on which two-factor interaction terms are set to zero.

Consider a hierarchical log-linear model

$$\log m(i) = \sum_{a \in K} u_a(i_a)$$

Definition 2.1. Given a set of labels contained in a set Δ, such that for every label $\delta \in \Delta$ there is an associated random variable I_δ, an interaction graph is a graph in which we assign to each variable I_δ a vertex δ. Additionally, two variables I_i and I_j are joined with an edge $\{i, j\}$ if there is an interaction term u_a in the log-linear model in which the set a contains both i and j.

In terms of graph theory, we have a graph $G(V, E)$, where the set of vertices V is equal to the set of labels Δ, $V = \Delta$, and the set of edges E is determined by the interactions in the model.

For example, we show in figure 2.1 the interaction graph for the hierarchical log-linear model generated by $\{\{A, B\}, \{B, D\}, \{A, C, D\}\}$,

$$\log m(i,j,k,l) = u + u_A(i) + u_B(j) + u_C(k) + u_D(l) + u_{AB}(ij) + u_{BD}(jl) + u_{AC}(ik)$$
$$+u_{AD}(il) + u_{CD}(kl) + u_{ACD}(ikl).$$

The set of vertices for this model is $V = \Delta = \{A, B, C, D\}$ and the set of edges is E= $\{\{A, C\}, \{C, D\}, \{A, D\}, \{A, B\}, \{B, D\}\}$.

An independence graph as defined by Whittaker (1990, p. 207) is as follows.

Definition 2.2. An independence graph is a graph in which each vertex represents a variable and the set of edges $E = \{e_r\}_{r \in \mathbb{N}}$ is formed by elements $\{i, j\}$, such that when $\{i, j\}$ is not contained in E, the variables I_i and I_j are conditionally independent given the other variables.

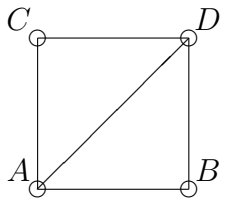


Figure 2.1: Interaction graph for the log-linear model with generating class $\{\{A, B\}, \{B, D\}, \{A, C, D\}\}$, $\log m(i, j, k, l)$=u+ $u_A(i)$+ $u_B(j)$+ $u_C(k)$+ $u_D(l)$+ $u_{AB}(ij)$+ $u_{BD}(jl)$+ $u_{AC}(ik)$+ $u_{AD}(il)$+ $u_{CD}(kl)$+ $u_{ACD}(ikl)$.

Whittaker (1990, p. 209) shows that every interaction graph is an independence graph and viceversa. This means that we can use any of the two terms without distinction.

According to Lauritzen (1996, p. 89), graphical log-linear models can be defined in the following way:

Definition 2.3. A log-linear model is *graphical* with associated graph $G = (V, E)$ if its generating class $A = C$ where C is the cliques set of the interaction graph.

Observe that, even though all hierarchical models can be represented with an interaction graph, not all of them are graphical because we need the generating class to be formed by all cliques of the graph. As we previously noticed, what these models have in common with graphical Gaussian models is that the lack of edges between variables represents conditional independence of two variables given the others and that these are models represented with graphs which provide the complete structure of the model.

Example 2.7. The log-linear model with generating class $\{\{A, B\}, \{B, D\}, \{ACD\}\}$ shown in figure 2.1 is not graphical because the clique $\{A, B, D\}$ contained in the corresponding cliques set $\{\{A, B, D\}, \{A, C, D\}\}$ is not contained in the generating class. However, the model with generating class $\{\{A, B, D\}, \{A, C, D\}\}$,

$$\log m(i, j, k, l) = u + u_A(i) + u_B(j) + u_C(k) + u_D(l) + u_{AB}(ij) + u_{BD}(jl) + u_{AC}(ik)$$

$$+u_{AD}(il) + u_{CD}(kl) + u_{ABD}(ijl) + u_{ACD}(ikl),$$

whose graph is the same, is graphical because the cliques set of the graph is equal to the generating class.

2.9 Marginal and conditional independence in log-linear expansions and their relation with graphical log-linear models

To introduce and analyze the relation between independence and graphical log-linear models, we will use an alternative definition of graphical log-linear models given by Whittaker (1990, p. 207). To do this, we need the following proposition also given by Whittaker (1990, p. 207), which applies for multinomial distributions; however, all results work whichever distribution, Poisson or restricted multinomial, is used.

Proposition 2.1. *If (I_a, I_b, I_c) is a partitioned multinomial random vector, then $I_b \perp I_c | I_a$ if and only if all u-terms in the log-linear expansion with one or more coordinate in b and with one or more coordinate in c, are zero.*

If a were empty, we would simply have marginal independence.

When b is formed by one variable I_j, c is formed also by one variable I_k, and a corresponds to the remaining variables, proposition 2.1 means that in order to have $I_j \perp I_k | I_{\Delta/\{j,k\}}$ we need $u_{\{j,k\}\cup t} = 0$ for $t \subseteq a$. From this fact, we derive that in graphical (as well as in hierarchical) log-linear models when there is not an edge between two vertices it is because these variables are independent given the other variables. The reason for this independence is as follows, when we do not have an edge in the graph is because all u-terms containing both variables in the edge are zero and, according to what we just said at the beginning of this paragraph, it means that both variables are independent given the other variables. For example, in the graphical model with generating class $\{\{A, B, D\}, \{A, C, D\}\}$, whose graph is the same given in figure 2.1, there is not an edge between B and C which means that $B \perp C | A, D$.

An alternative way of defining graphical log-linear models (Whittaker, 1990, p. 207), which will be useful to clarify the relation between conditional independence and graphical log-linear models, is:

Definition 2.4. Given an independence graph, the distribution of vector $I = (I_\delta)_{\delta \in \Delta}$ is a graphical model, or the log-linear model is graphical, if the distribution of I is arbitrary, apart from constraints in the log-linear representation $\log p(i) = \sum_{a \subseteq \Delta} u_a(i_a)$ of the form $u_a = 0$ if $\{i, j\} \subseteq a$ and $\{i, j\}$ is not in the edge set E.

Example 2.8. Suppose that we have the independence graph given in figure 2.2. We have that $X \perp Z | Y$ because the graph is an independence graph. This independence implies, according to proposition 2.1, that $u_{XZ} = u_{XYZ} = 0$, obtaining a model with generating class $\{\{X, Y\}, \{Y, Z\}\}$,

$$\log m(i, j, k) = u + u_X(i) + u_Y(j) + u_Z(k) + u_{XY}(ij) + u_{YZ}(jk).$$

This is a graphical model, according to definition 2.4, because $\{X, Z\}$ is the only edge not contained in E and $u_a = 0$ if $\{X, Z\} \subseteq a$ using that $u_{XZ} = 0$ and $u_{XYZ} = 0$.

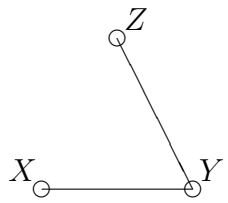


Figure 2.2: Independence graph for the graphical model with generating class $\{\{X, Y\}, \{Y, Z\}\}$, $\log m(i, j, k) = u + u_X(i) + u_Y(j) + u_Z(k) + u_{XY}(ij) + u_{YZ}(jk)$.

Observe from definition 2.4, using proposition 2.1, that the restrictions given to the graphical model parameters correspond to having conditional independence between variables not having edges joining them given the other variables and that these are the only zero restrictions that the parameters should have. Observe also that graphical models are hierarchical models, because according to what we saw in section **2.2.1**, hierarchical are models in which if $u_a = 0$ then $u_t = 0$, for all $a \subseteq t$, which is what graphical models satisfy in this alternative definition.

Example 2.9. An example of a non-graphical log-linear model is

$$\log m(i, j, k) = u + u_1(i) + u_2(j) + u_3(k) + u_{12}(ij) + u_{13}(ik) + u_{23}(jk);$$

i.e., a model with generating class

$$A = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\},$$

whose interaction graph is shown in figure 2.3.

There is no conditional independence between any pair of variables given the other variables because all parameters, u_{12}, u_{13}, and u_{23} are not set to zero and according to proposition 2.1 in order to have that kind of independence any of them should be zero. Using this and definition 2.4 we deduce that in order to have a graphical model we should not have restrictions for the parameters in the log-linear representation; however, we have the restriction $u_{123} = 0$ and as a consequence the model is not graphical. Another way of seeing this fact is by examining the corresponding interaction graph (figure 2.3) in which all edges between all variables are contained in the graph, and from this we see again that if the model were graphical it should not happen that $u_{123} = 0$.

In terms of definition 2.3, the graph given in figure 2.3 contains only one clique, the set $\{1, 2, 3\}$, which is not in the generating class A of the model, hence the model is not

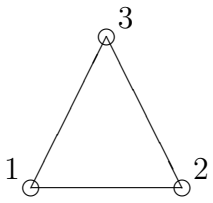


Figure 2.3: Graph for the non-graphical log-linear model with generating class $\{\{1,2\}, \{1,3\}, \{2,3\}\}$, $\log m(i,j,k) = u + u_1(i) + u_2(j) + u_3(k) + u_{12}(ij) + u_{13}(ik) + u_{23}(jk)$.

graphical.

Both definitions of graphical models, 2.3 and 2.4, are equivalent. This equivalence is because Whittaker (1990, p. 209) proved that a model is graphical according to definition 2.4 if and only if its generating class, the maximal u-terms, corresponds to the cliques of the associated graph. What is important to understand is that graphical models were built to represent conditional independences between two variables given the other variables, and at first these are the kinds of independences we can deduce from the corresponding graphs.

We notice that there is a one-to-one correspondence between graphs and graphical log-linear models, *e.g.*, Christensen (1997, p. 184). If we have any graph, we can get its cliques and from them build the generating class for a graphical log-linear model. On the other hand, if we have a hierarchical log-linear model, we can build the corresponding interaction graph, in particular if the model is graphical.

It is also important to notice that the two mentioned approaches to characterize graphical log-linear models are not the only ones. Darroch *et al.* (1980) get the same results using the so-called Markov fields theory, getting that models with the properties for their generating class given above also satisfy other group of conditional independences (Markov properties), among them conditional independence of two not adjacent variables given the other variables.

2.10 Graphical log-linear models interpretation

We present the Markov properties for a probability measure P in X, the probability space for the random variables $(X_v)_{v \in V}$ associated to the vertices on a graph $G = (V, E)$, where $V = \Delta$, the set of labels given to the variables, based on the work by both Darroch *et al.* (1980, p. 525) and Lauritzen (1996, p. 32).

- Pairwise Markov property. The probability measure obeys this property if for any non-adjacent vertices α and β, $X_\alpha \perp X_\beta | X_{\Delta \setminus \{\alpha,\beta\}}$.

- Local Markov property: Let a be a set of vertices, we define $bd(a)$ as the set of vertices in $\Delta \setminus a$ adjacent to some vertex in a and $cl(a)$ as the set $a \cup bd(a)$. The probability measure obeys the local Markov property if for any vertex, $\alpha \in V$ $X_\alpha \perp X_{\Delta \setminus cl(\alpha)} | X_{bd(\alpha)}$.
- Global Markov property. The probability measure obeys this property if for any (A, B, S) of disjoint subsets of V such that S separates A from B; *i.e.*, any path from A to B goes through some vertex in S, $X_A \perp X_B | X_S$.

According to what we mentioned in section **2.9**, graphical log-linear models are built following the pairwise Markov property. Additionally, in proposition **3.4** Lauritzen (1996, p. 33) proved that for any graph G and for any distribution, the global Markov property implies the local one and this last one implies the pairwise property.

For some particular distributions, for example for distributions in contingency tables with $p(i) > 0$, Lauritzen (1989, p. 18-19) proved that in graphical log-linear models the three properties are equivalent. He also outlines a proof (Lauritzen, 1996, p. 35) of the fact that in models whose distribution is factorized according to the set of cliques, like in graphical models, and if the probability density is positive and continuous, then all Markov properties are equivalent, this result is based on the one attributed to Hammersley and Clifford (1971) in the discrete case. In fact, having a positive density is enough to have equivalence between the three properties as discussed by both Edwards (2000, p. 8) and Whittaker (1990, p. 57). This can be obtained using the separation theorem presented and proved by Whittaker (1990, p. 67) and the implications between the three Markov properties. The separation theorem says that if we have random variables with positive density function, then the pairwise Markov property implies the global Markov property.

The previous statement implies that for any graphical model, we can get conditional independences between different sets of variables using Markov properties, not only getting conditional independences between pairs of variables given the rest, but also conditional independences involving other sets; for example, if we have a separator set of vertices S separating the vertices in A from the ones in B, where all sets are disjoint, then, using the global Markov property, we have that the variables in A are independent from the ones in B given the variables in S. As a particular case, if we have a set of vertices A in a connected component and another set of vertices B in another component, then the variables associated to A are independent to the ones associated to B, *i.e.* $A \perp B$. Conditional independences involving others sets of variables can be obtained using the local Markov property; in fact, this property is a particular case of the global Markov property because $bd(\alpha)$ separates α from $\Delta \setminus cl(\alpha)$.

Heckerman *et al.* (2004) proposed a property called *conditionally specified distribution* that the probability space should satisfy given an undirected graph $G(V, E)$ and

they tried to relate it with all properties defining graphical models, including Markov properties, to give another definition of graphical model or to obtain additional information from the graph associated to the model; however, there is not equivalence between all properties.

Example 2.10. Consider a log-linear model with generating class $\{\{A, B, D\}, \{A, C, D\}\}$,

$$\log m(i,j,k,l) = u + u_A(i) + u_B(j) + u_C(k) + u_D(l) + u_{AB}(ij) + u_{BD}(jl) + u_{AC}(ik)$$
$$+u_{AD}(il) + u_{CD}(kl) + u_{ABD}(ijl) + u_{ACD}(ikl),$$

whose graph corresponds to the one in figure 2.1. Observe that $\{A,\ D\}$ separates B from C, which means that $B \perp C | \{A, D\}$.

Example 2.11. Consider a log-linear model generated by $\{\{A, B, C\}, \{C, D, E\}\}$,

$$\log m(i,j,k,l,o) = u + u_A(i) + u_B(j) + u_C(k) + u_D(l) + u_E(o) + u_{AB}(ij) + u_{BC}(jk) + u_{AC}(ik)$$
$$+u_{CD}(kl) + u_{CE}(ko) + u_{DE}(lo) + u_{ABC}(ijk) + u_{CDE}(klo),$$

whose graph, figure 2.4, has cliques $\{A, B, C\}$ and $\{C, D, E\}$. The only way to go from one clique to the other is through C, which is the intersection between the elements in the generating class, then the vertex partition is $(\{A, B\},\{D, E\},\{C\})$, with $\{C\}$ the separating set. This means that $A, B \perp D, E | C$.

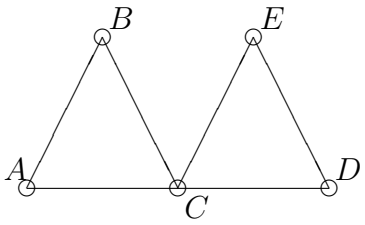


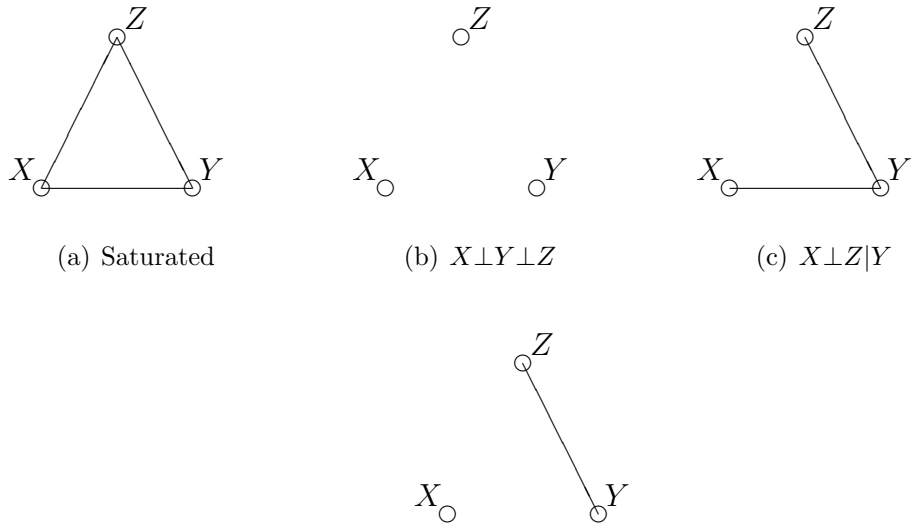
Figure 2.4: Graphical log-linear model with generating class $\{\{A, B, C\},\ \{C, D, E\}\}$ with cliques $\{A, B, C\}$ and $\{C, D, E\}$ and separating set $\{C\}$.

We list all kinds of graphical and non-graphical models for three-way contingency tables, see for instance Haberman (1978, p. 197-207). The graphical models are as follows:

a) Saturated model. It is the model generated by $A = \{\{X, Y, Z\}\}$,

$$\log m(i,j,k) = u + u_X(i) + u_Y(j) + u_Z(k) + u_{XY}(ij) + u_{XZ}(ik) + u_{YZ}(jk) + u_{XYZ}(ijk).$$

We observe from figure 2.5(a) that there are no separator sets between sets of variables, then there is not any kind of independence between variables.



(a) Saturated (b) $X \perp Y \perp Z$ (c) $X \perp Z|Y$

(d) $X \perp (Y, Z)$

Figure 2.5: Graphical log-linear models with different generating class: a) $\{\{X, Y, Z\}\}$, $\log m(i,j,k) = u + u_X(i) + u_Y(j) + u_Z(k) + u_{XY}(ij) + u_{XZ}(ik) + u_{YZ}(jk) + u_{XYZ}(ijk)$; *b)* $\{\{X\}, \{Y\}, \{Z\}\}$, $\log m(i,j,k) = u + u_X(i) + u_Y(j) + u_Z(k)$; *c)* $\{\{X, Y\}, \{Y, Z\}\}$, $\log m(i,j,k) = u + u_X(i) + u_Y(j) + u_Z(k) + u_{XY}(ij) + u_{YZ}(jk)$; *d)* $\{\{X\}, \{Y, Z\}\}$, $\log m(i,j,k) = u + u_X(i) + u_Y(j) + u_Z(k) + u_{YZ}(jk)$.

b) Model without first and second-order interactions. This is a model with generating class $A = \{\{X\}, \{Y\}, \{Z\}\}$,

$$\log m(i,j,k) = u + u_X(i) + u_Y(j) + u_Z(k).$$

It is represented as three isolated points, figure 2.5(b), so that there is no path joining the variables and the variables are mutually independent.

c) Models without second-order interaction and without one of the three first-order interaction terms. These are the models generated by sets of the form $A = \{\{X, Y\}, \{Y, Z\}\}$,

$$\log m(i,j,k) = u + u_X(i) + u_Y(j) + u_Z(k) + u_{XY}(ij) + u_{YZ}(jk).$$

The corresponding graph, figure 2.5(c), is a path of length two, implying that $X \perp Z|Y$. This independence relation can be also easily seen from proposition 2.1 using that all parameters containing X and Z at the same time are zero, $u_{XZ} = u_{XYZ} = 0$.

d) Models without second-order interaction and without two first-order interaction terms. They are generated by sets of the form $A = \{\{X\}, \{Y, Z\}\}$,

$$\log m(i, j, k) = u + u_X(i) + u_Y(j) + u_Z(k) + u_{YZ}(jk).$$

The graph, figure 2.5(d), consists on an isolated vertex and an edge between the remaining vertices, so that $X \perp (Y, Z)$, *i.e.* X is jointly independent of Y and Z.

The non-graphical log-linear models when there are three variables are the following:

a) The model not including second-order interaction, *i.e.* the model generated by $A = \{\{X, Y\}, \{Y, Z\}, \{X, Z\}\}$,

$$\log m(i, j, k) = u + u_X(i) + u_Y(j) + u_Z(k) + u_{XY}(ij) + u_{XZ}(ik) + u_{YZ}(jk).$$

As we saw before, example 2.9, this is a non-graphical model and does not have an interpretation in terms of marginal or conditional independence; however, it has been studied by Birch (1963) as a model for *partial association* between each pair of variables.

b) Non-hierarchical models. These are obviously non-graphical models because all graphical models are hierarchical. One example of this kind of model is

$$\log m(i, j, k) = u + u_X(i) + u_Y(j) + u_Z(k) + u_{XYZ}(ijk).$$

c) Non-comprehensive models. These were defined by Bishop *et al.* (1975, p. 38) as models depending on one or two variables in spite of having three variables on the contingency table. For example, the hierarchical model $\log m(i, j, k) = u + u_X(i) + u_Y(j)$, which indicates independence between X and Y without considering Z. This model can not be graphically represented with three vertices; however, if we considered the model without taking into account the dimension corresponding to Z, we would have a graph with only two vertices indicating independence between X and Y. These are models with generating classes of the form $A = \{\{X\}, \{Y\}\}$. We also have models generated by sets of the form $A = \{\{X, Y\}\}$, which according to Whittaker (1990, p. 212) indicate conditional equi-probability, *i.e.* the same probability without taking into account Z, models with generating classes of the form $A = \{\{X\}\}$ that do not indicate any dependence and that indicate that given one variable, all combinations of categories of the other two variables are equally probable, or models with only constant

term indicating that all combinations of the three variables are equally probable.

Darroch *et al.* (1980, p. 537) calculated the number of log-linear models as 2^{2^k-1}, where k is the number of variables, and the number of graphical log-linear models as $\sum_{i=0}^{k} \binom{k}{i} 2^{\binom{i}{2}}$. For the last formula they consider the order of the labeling in all variables, the number of graphical log-linear models obtained by taking subsets of the k variables ignoring the other variables, *i.e.* non-comprehensive models, and the empty graph. For example, with 3 variables, there are 128 log-linear and 18 graphical models; with 4 variables, there are 32,768 log-linear and 113 graphical models; and with 5 variables, there are 2,147,483,648 log-linear and 1,450 graphical models.

As the dimension increases, it could be more difficult to interpret a non-graphical model in terms of independences by only considering the linear expansion or basic association measures as odds ratios. This difficulty motivated the development of graphical log-linear models, because they are defined, for instance by Christensen (1997, p. 182), as models with interpretations in terms of conditional independence. One property of any hierarchical log-linear model is that it can be nested into a graphical one, at least in the saturated model, which is the largest graphical model. Then, to interpret a non-graphical hierarchical log-linear model, it could be useful to get the smallest graphical log-linear model containing it.

Example 2.12. The model with generating class

$$\{\{1,2\},\{1,3\},\{1,4\},\{2,3\}\},$$

$$\log m(i,j,k,l) = u + u_1(i) + u_2(j) + u_3(k) + u_4(l) + u_{12}(ij) + u_{23}(jk) + u_{13}(ik) + u_{14}(il),$$

with graph given in figure 2.6, is not graphical; however, it is contained in the graphical log-linear model generated by $\{\{1,2,3\},\{1,4\}\}$,

$$\log m(i,j,k,l) = u + u_1(i) + u_2(j) + u_3(k) + u_4(l) + u_{12}(ij) + u_{23}(jk) + u_{13}(ik)$$
$$+u_{14}(il) + u_{123}(ijk),$$

whose graph is also given in figure 2.6. This means that the non-graphical model keeps the independence $(2,3) \perp 4|1$ that can be deduced from the graphical one.

The smallest graphical model containing a hierarchical non-graphical model is obtained by getting the interaction graph associated to the non-graphical model. From

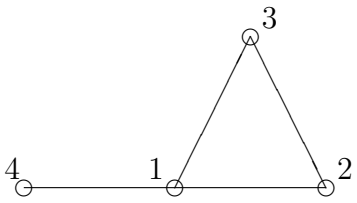


Figure 2.6: Interaction graph for two models: a) non-graphical hierarchical log-linear model with generating class $\{\{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}\}$, $\log m(i,j,k,l) = u + u_1(i) + u_2(j) + u_3(k) + u_4(l) + u_{12}(ij) + u_{23}(jk) + u_{13}(ik) + u_{14}(il)$*; b) graphical log-linear model with generating class* $\{\{1,2,3\}, \{1,4\}\}$, $\log m(i,j,k,l) = u + u_1(i) + u_2(j) + u_3(k) + u_4(l) + u_{12}(ij) + u_{23}(jk) + u_{13}(ik) + u_{14}(il) + u_{123}(ijk)$.

this graph, we get the set of cliques C, which corresponds to the generating class for the smallest graphical model containing the non-graphical hierarchical model, *e.g.*, Lauritzen (1996, p. 89). This implies that from non-graphical hierarchical log-linear models, we could get some interpretations in terms of conditional independence; however, from certain conditional independences, we do not necessarily get the same model. On the other hand, for graphical log-linear models, we have a one-to-one relation between conditional independences and the model. Then, to interpret a non-graphical hierarchical log-linear model, we use the smallest graphical log-linear model containing it or we could also interpret it by using proposition 2.1. For example, the model with generating class $\{\{1,2,3\}, \{1,4\}, \{2,4\}\}$ satisfies that all interaction terms containing 3 and 4 at the same time are zero, which implies that $3 \perp 4 | 1, 2$. We get the same result by using the smallest graphical model containing the model.

A particular kind of graphical log-linear models are decomposable models, see for instance Darroch *et al.* (1980, p. 22), Lauritzen (1989, p. 35), or Eriksen (2005) who also proposes a new model class, decomposable log-linear models, covering a wider range of models even some non-hierarchical which obviously are also not graphical. Decomposable models can be defined in terms of graph theory using the definition of decomposable graphs applied to the corresponding interaction graph, as discussed by Lauritzen (1996, p. 8, 90), but this definition is equivalent to say that the associated graphs are chordal or triangulated whose definition can be found on Appendix A. Decomposable models are characterized by having exact analytic formulas for their maximum likelihood estimators, *e.g.*, Edwards (2000, p. 22), such explicit formulas are presented for instance by Lauritzen (1996, p. 91) in proposition 4.18.

We have the following relations between different kinds of log-linear models: Given a group of variables, decomposable models are a subset of those graphical, which are a subset of hierarchical log-linear models, which at the same time are a subset of all possible log-linear models for that group of variables; *i.e.*

Log-linear models $\supseteq$ Hierarchical log-linear models $\supseteq$ Graphical log-linear models $\supseteq$ Decomposable models.

As mentioned above decomposable models have exact formulas for their maximum likelihood estimators, and, because they are graphical, they also have interpretations in terms of conditional independences. Hierarchical models are the log-linear models generally analyzed, because when we have a non-hierarchical model the statistical significance and the parameters interpretation depends on how variables are coded. This is what happens when in a model we include a term without including some terms of inferior order as discussed by Agresti (2002a, p. 316). Even so, some methods have been proposed to interpret non-hierarchical models, which are part of the so-called non-standard log-linear models discussed by Rindskopf (1990) and Mair (2007).

To illustrate decomposable models and to continue with the interpretation of graphical models presented so far, consider log-linear models having four variables. We consider as many hierarchical models as generating classes can be obtained using four variables because if the model does not contain all the variables, each one in at least one element in the generating class, we fall in the case of having non-comprehensive models which we are not considering. Without loss of generality, suppose that we label the variables as $1, 2, 3, 4$. Given that order for the variables, all possible graphical models are generated by the sets presented in the first column of table 2.6. We present in the same table the corresponding graphs and interpretations based on table 5.1 presented by Christensen (1997, p. 181). Observe also from table 2.6 that the total number of graphical models considering different labellings of the variables is 64 corresponding to $\binom{k}{i}2^{\binom{i}{2}}$, with $k = 4$ and $i = 4$, which is the part of the formula seen above used to calculate the number of graphical log-linear models that provides us with the number of graphical models with exactly four variables.

Observe that for example, the model with generating class $\{\{1, 2, 3\}, \{1, 2, 4\}\}$ is decomposable because there is only one cycle of length four for the corresponding graph and it contains the chord $\{1, 2\}$. In fact, there is only one non-decomposable model, the model generated by

$$\{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 1\}\}.$$

The corresponding interaction graph is a cycle of length four, so that there are not chords and this is why the model is not decomposable. There are not chords because there are not edges in the cycle joining non-consecutive vertices, neither the edge $\{1, 3\}$, nor the edge $\{2, 4\}$ are contained in the graph.

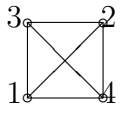
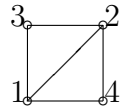
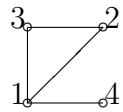
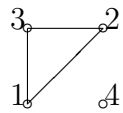
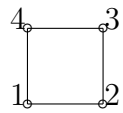
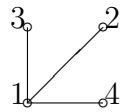
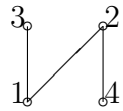
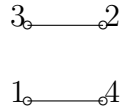
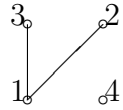
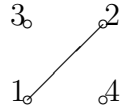
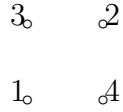
Generating class	Associated graph	Interpretation	# models with different labellings
$\{\{1,2,3,4\}\}$		without restrictions	1
$\{\{1,2,3\},\{1,2,4\}\}$		$3\perp 4\|1,2$	6
$\{\{1,2,3\},\{1,4\}\}$		$2,3\perp 4\|1$	12
$\{\{1,2,3\},\{4\}\}$		$1,2,3\perp 4$	4
$\{\{1,2\},\{2,3\},\{3,4\},\{4,1\}\}$		$[2\perp 4\|1,3]\cap[1\perp 3\|2,4]$	3
$\{\{1,2\},\{1,3\},\{1,4\}\}$		$2\perp 3\perp 4\|1$	4
$\{\{1,2\},\{1,3\},\{2,4\}\}$		$[3\perp 2,4\|1]\cap[1,3\perp 4\|2]$	12
$\{\{1,2\},\{3,4\}\}$		$2,3\perp 1,4$	3
$\{\{1,2\},\{1,3\},\{4\}\}$		$[1,2,3\perp 4]\cap[2\perp 3\|1]$	12
$\{\{1,2\},\{3\},\{4\}\}$		$1,2\perp 3\perp 4$	6
$\{\{1\},\{2\},\{3\},\{4\}\}$		$1\perp 2\perp 3\perp 4$	1

Table 2.6: Graphical log-linear models with four variables for an specific labelling. The only non-decomposable model is generated by $\{\{1,2\},\{2,3\},\{3,4\},\{4,1\}\}$.

Darroch *et al.* (1980) present tables with graphs and their interpretation for decomposable models of dimension fewer or equal to 5 whose graph is connected similar to table 2.6 presented here. Unlike the tables presented by them, we do not assume that the graphs are connected or that the model is decomposable.

2.11 Deviance in graphical log-linear models

The goodness of fit statistic for graphical models that we discuss in this section is the deviance. The reason for this discussion is that the deviance has some qualities that in general, as we will see at the end of this section, make it an adequate goodness of fit statistic and a statistic useful in model selection. In fact, most of the software uses this statistic in their analysis, in particular *MIM*, the free software for graphical models that we use throughout this work, and *Splus* use it.

Let I_a be a random sub-vector of I and denote as p_a to $p_{I_a}(i_a)$, *i.e.* the marginal probability for the sub-vector I_a evaluated in i_a. We have the following theorem based on proposition 7.1.1 presented by Whittaker (1990, p.203) without formal proof.

Theorem 2.1. *A multinomial vector, and in fact a vector with any distribution for the cells, Poisson or restricted multinomial with fixed margins, partitioned into* (I_a, I_b, I_c) *satisfies*

$$I_b \perp I_c | I_a \Leftrightarrow p_{abc} = p_{ac}p_{ab}/p_a$$

Proof: Observe that $p_{abc} = p(i)$, this is only a notational change. If $I_b \perp I_c | I_a$, then, using the definition of conditional independence, we have that $p_{I_b,I_c|I_a}(i_b, i_c|i_a) = p_{I_b|I_a}(i_b|i_a)p_{I_c|I_a}(i_c|i_a)$, from this expression we get the desired result. Conversely, using the factorization criterion that in this case indicates that $I_b \perp I_c | I_a$ if and only if $p_{abc} = f(i_a, i_b)g(i_a, i_c)$, where the function f can be replaced by p_{ab} and the function g by p_{ac}/p_a, we get that $p_{abc} = p_{ac}p_{ab}/p_a$ implies $I_b \perp I_c | I_a$.

Using that $|n| = \sum_i n(i)$ is the total observed count and that under a multinomial distribution $m(i) = |n|\, p(i)$, and supposing that we have multinomial variables with $I_b \perp I_c | I_a$, we have by theorem 2.1 that $p(i) = p_{ab}p_{ac}/p_a$. Then $\widehat{m}(i) = |n|\, \widehat{p}(i) = n(i_{ab})n(i_{ac})/n(i_a)$, using that $\widehat{p}_{ab} = n(i_{ab})/\,|n|$.

By using this result, the residual deviance formula, and that the number of cells in marginal distributions for I_a, I_b, and I_c are $|\mathrm{I}_a|$, $|\mathrm{I}_b|$, and $|\mathrm{I}_c|$, respectively, we have the following result given by Whittaker (1990, p. 223).

Proposition 2.2. *Under a multinomial distribution, the residual deviance for a conditional independence model* $I_b \perp I_c | I_a$ *is*

$$d(I_b \perp I_c | I_a) = 2 \sum_{i \in \mathrm{I}} n(i_{abc}) \log \frac{n(i_{abc}) n(i_a)}{n(i_{ab}) n(i_{ac})}.$$

The number of degrees of freedom for the deviance is $|\mathrm{I}_a| \, (|\mathrm{I}_b| - 1)(|\mathrm{I}_c| - 1)$.

The number of degrees of freedom given in proposition 2.2 can be derived like in example 2.6 or can be obtained by using the following method: If we have an independence model between I_b and I_c, we can easily see that the associated number of degrees of freedom is $(|\mathrm{I}_b| - 1)(|\mathrm{I}_c| - 1)$. This value corresponds to the number of cells minus the number of independent parameters $|\mathrm{I}_b| \, |\mathrm{I}_c| - (1 + (|\mathrm{I}_b| - 1) + (|\mathrm{I}_c| - 1))$. The model $I_b \perp I_c | I_a$ can be seen as an independence model of I_b and I_c at each value of I_a. Using this fact and the number of degrees of freedom in an independence model we get the desired number of degrees of freedom.

Obviously, the independence model, $I_b \perp I_c$, can be seen as a particular case of the model $I_b \perp I_c | I_a$, with $a = \emptyset$, where we define $n(i_\emptyset) = |n|$. The number of degrees of freedom in this case is $(|\mathrm{I}_b| - 1)(|\mathrm{I}_c| - 1)$.

Using proposition 2.2, we can get the residual deviance for a model without a specific edge $\{j, k\}$, *i.e.* for the model $I_j \perp I_k | I_{\Delta / \{j,k\}}$

$$d(I_j \perp I_k | I_{\Delta / \{j,k\}}) = 2 \sum_{i \in \mathrm{I}} n(i) \log \frac{n(i) n(i_{\Delta / \{j,k\}})}{n(i_{\Delta / \{j\}}) n(i_{\Delta / \{k\}})}, \tag{2.15}$$

whose asymptotic distribution is χ^2 with $\left|\mathrm{I}_{\Delta / \{j,k\}}\right| (|\mathrm{I}_j| - 1)(|\mathrm{I}_k| - 1)$ degrees of freedom.

Through the residual deviance given in equation (2.15), we can test which edges are non-significant to exclude them from a graphical model. In terms of graph theory, we are eliminating from the complete graph an edge each time. When the deviance is small with respect to a χ^2 distribution with its corresponding degrees of freedom we have that the model without the edge explains the data as well as when the edge is included in the complete graph, which implies that the edge should not be included. Formally, it means that we are not rejecting the null hypothesis that the parameters in the saturated model not forming part of the model without the edge are zero. By eliminating non-significant edges we can define a method that can be used to get a graphical model that fits some data.

The deviance is a useful statistic in graphical models because of its following properties:

- It is invariant to different parametrizations, *i.e.* it is an estimable function as discussed by Long (1984).

- It can be used to get selection criteria to choose graphical models that fit some data. It can be used not only for log-linear models, but also for Gaussian and mixed models.
- It can be interpreted as a divergence measure.
- It has a known asymptotic distribution. This property is inherited from the fact that it is a likelihood ratio test. For two nested models, under the null hypothesis that the parameters not contained in the smaller model are zero, the difference between the deviances has an asymptotic χ^2 distribution.

We note that if the sample size is relatively small or the number of variables in the model increases the asymptotic distributions could not be adequate. In this case we would need to get the exact distribution; however, this is in general intractable as discussed by Lauritzen (1996, p. 80). Another possibility is to correct the asymptotic distribution, for example using the so-called Bartlett corrections (Bartlett, 1937), although research developed by Frydenberg and Jensen (1989) indicate that there is no real improvement in the discrete case, or using another test statistics like the *power diverge statistics* discussed by Read and Cressie (1988).

2.12 Graphical log-linear models selection

Selection methods for log-linear models have been developed and further studied for instance by Goodman (1971). For graphical models there are three selection methods, the first one being similar to those methods available in log-linear models,

1. Stepwise methods: We start from an initial model and we add or remove edges until some criterion is fulfilled.
2. More global search techniques: We seek simple models consistent with the data.
3. Selection of a model that optimizes some information criterion; for example the AIC or BIC criteria.

All these methods should be used carefully because even though the methods identify models consistent with the data, this does not mean that they are reflecting reality. There are some problems: i) It might happen that we are not using all variables necessary to represent the phenomenon we want to describe. ii) The model selection is based on significance tests, for example, to know if we have to add or remove an edge, and we control the error levels for each of these tests, but we are not controlling the overall error. iii) All inferences are made under a theoretical fixed model; by fitting a model using some data an infering something from it, we have estimators biased according to the data.

The forward and backward methods are included in the stepwise methods class. In the backward method, edges are successively removed from an initial or current model, *i.e.* we have as possible contenders to all models in which an edge is removed, and we get the deviance, called exclusion deviance, between the models without an edge and the initial or current model. This process is similar to the one seen at the end of section **2.11** in which we removed edges from the saturated model. The deviance has a χ^2 sampling distribution under the null hypothesis, and we remove the least significant edge or edges, so that we delete the edges whose deviance between the models with and without the edges is small with respect to the quantiles for a χ^2 distribution with its corresponding degrees of freedom for a given significance level. This means that we do not reject the null hypothesis that the parameters in the model with the edge and not included in the model without the edge are zero. A usual stopping rule is to continue excluding edges until there are no *p-values* greater than 5% or the desired significance level.

One way to implement this method suggested by Whittaker (1990, p. 252) uses two steps, even though we could only use the first one, is:

- Obtain an independence graph. To do this, we exclude from the saturated model one edge each time, *i.e.* we get the exclusion deviances with respect to the saturated model with their corresponding p-values. We remove all non-significant edges, which are those edges in which the corresponding p-value is greater than the significance level, and we get the model G_1. Under this method we could remove more than one edge.

- We get the so-called inclusion deviances for all edges removed from the saturated model to get G_1; *i.e.*, we get the deviances for the model obtained from G_1 adding each time a missing edge. We include all significant edges, that is we add all edges whose deviances are greater than their corresponding quantiles, rejecting the null hypothesis that the model without the edge fits the data as well as when the edge is included. That is, we add all edges whose associated p-value is less than the significance level.

Edwards (2000, p. 158-160), proposes a slightly different backward method in which in each step we can only delete one edge. We take all models in which one edge is removed from a current model and compare them with the original model by using the deviance. We delete the least significant edge for that model; *i.e*, the edge with the largest p-value provided it were greater than the significance level α. We get a new graphical model in which we apply again the same method, and so on until all p-values $p < \alpha$.

In model selection, in general, we would like to satisfy the *coherence principle* defined by Gabriel (1989) and explained in particular for graphical models by Edwards

(2000, p. 165) and Whittaker (1990, p. 253), which says that if for two nested models $M_0 \subset M_1$, M_1 is rejected, *i.e.* M_1 is inconsistent with the data, then M_0 is also rejected. On the other hand, if M_0 is accepted, *i.e.* M_0 is consistent with the data, then M_1 is also accepted. In the previous algorithm, this principle translates to the following: if we reject removing an edge in one step, then that edge can not be chosen to be removed in any following step.

Another restriction that can be added for model selection is considering at each step only decomposable models. That is, if in one step once removing an edge we have a non-decomposable model, then in that step we can not delete such edge.

In forward methods, we add the most significant edges instead of removing the least significant edges. Then, the previous algorithm is modified. Now, we add in each step the edge with least p-value provided that it were less than the significance level α.

There are some variations in the way these methods are applied; for example, by examining which edges to add or remove in a random order so that the edge added or removed is the first edge satisfying the required condition. For the backward method we remove the first non-significant edge and for the forward method we add the first significant edge. This method is faster than the standard method and has been implemented in the computer software *MIM*. In this software the described algorithms for forward and backward methods have been also implemented.

The second kind of selection methods are global search techniques which examine the simplest models consistent with the data. To do this, we need to obtain the so-called adequate models, defined for instance by Whittaker (1990, p. 253), which are models whose residual deviance is small enough so that the model fits as well to the data as the saturated model. Minimally adequate models also defined by Whittaker (1990, p. 253) are required, they are adequate models, M, such that there is no other model N contained in M which is also adequate. To get them, algorithms as the EH-Procedure are used, these algorithms are presented by Edwards and Havránek (1985, 1987) and also discussed in Edwards (2000, p. 167).

The last kind of selection methods are the ones using information criteria as the Akaike (AIC) or Bayesian information criteria (BIC). These methods are computationally more challenging. In *MIM*, all submodels for an specified class of models are fitted and the corresponding information criterion is calculated. Edwards (2000, p. 172) emphasizes that this strategy of model search is feasible only for models with a small to moderate dimension, he illustrates this fact by presenting an example with only 5 variables and a specific model, for this case there are 1024 possible submodels to examine. He concludes saying that this number increases greatly as the number of variables increases. Recently, Edwards *et al.* (2010) developed a method to select a graphical

model whose associated graph is a tree or forest, a disconnected graph whose connected components are trees, that optimizes a penalized likelihood criterion, for example AIC or BIC, for high-dimensional data typical of genomic studies in which model selection methods that perform well for moderate dimensions may be intractable. Even though these model classes can be too simple to represent complex biological systems, they still can give a preliminary understanding of the dependence structure, can be used as a start model in a search algorithm based on richer model classes, or may allow dimension reduction.

Selection methods are a useful tool to explore data because even though we could not get a simple or satisfying model, the results could be the basis to get better models that we had not considered before. For example, we could think of using latent variables, or according to the selected models, we could try to fit models with a more specific structure, for example, models in which some parameters are restricted to be equal between them. In this work, these are precisely the kind of restrictions we impose to hierarchical log-linear models with a graphical representation.

That implies that first we use model selection with the purpose of discovering the structures in a data set finding conditional independences and after that we formulate a model with additional restrictions based on the model previously found. We chose in general to use stepwise methods, both forward and backward, because these methods are available on *MIM* and, as we discussed above, they are quicker; however, as we analyzed several data besides the ones that finally were included in this work to see if the models defined in the following chapters were adequate, we actually used, in some cases, the other methods available in *MIM* to see if there were graphical models that some methods did not identified but others did. After selecting a model, we also checked its fit using the deviance.

Chapter 3

Restricted or coloured graphical log-linear models

3.1 Generalities

In this chapter we define and study new type of hierarchical log-linear models corresponding to the original work for this dissertation. They are named restricted graphical log-linear models and as we will analyze later they have many features in common with graphical log-linear models. Firstly, we introduce in section **3.2** the notation that we will be using throughout this and the following chapters, many of the terms defined are just introduced, but others were introduced in section **2.1**; however, we chose to list all of them so that the reader could consult them in a single section. Secondly, we define the models, state their properties, give some examples, get the corresponding likelihood equations, and discuss how these models can be expressed as GLM models. Thirdly, we present a numerical method to approximately solve the likelihood equations, and finally, we provide examples of how these models can be fitted for particular contingency tables. We have already reported parts of this material in Eslava and Ramírez (2008).

3.2 Notation and terminology

$\|Z\|$	number of elements in a set Z.
Δ=V	set of vertices or set of variable names or labels.
E	set of multiple edges corresponding to the associated graph.
$(V_1, V_2, ..., V_T)$	vertex partition of V into T classes.
$(E_1, E_2, ..., E_S)$	edge partition of E into S classes.
v_k^i	k th vertex in the colour class V_i, i=1,...,T; k=1,...,$kver(i)$; $kver$ is the vector of dimension T of number of vertices in each vertex class.

$I = (I_\delta)_{\delta\in\Delta}$	discrete random variables associated to the set of vertices.
I_δ	categories or level set for I_δ.
$\lvert\mathrm{I}_\delta\rvert$	total number of categories for the variable I_δ.
I: $\mathrm{I} = \times_{\delta\in\Delta}\mathrm{I}_\delta$	variable value combinations.
i	cell or particular variable value combination, $i \in \mathrm{I}$. If we had a two-way contingency table, a particular cell could be denoted as (i_1, i_2). Similarly for q-way contingency tables, $i = (i_1, i_2, ..., i_q)$.
$I_{V_i} = (I_\delta)_{\delta\in V_i}$	discrete random variables associated to V_i, i=1,...,T.
I^{V_i}	Categories or level set for all variables in V_i, i=1,...,T.
$p(i)$	probability than an object belongs to a cell i, $i \in \mathrm{I}$.
$m(i)$	expected frequency in cell i, $i \in \mathrm{I}$. In a two-way contingency table, the expected frequency for a specific cell is denoted as $m(i_1, i_2)$.
m	expected frequency vector, $m' = (m(i))_{i\in\mathrm{I}}$.
$\widehat{m}(i)$	estimated expected frequency in cell i, $i \in \mathrm{I}$.
$n(i)$	observed count in cell i, $i \in \mathrm{I}$. In a two-way contingency table, the observed count for a specific cell is denoted as $n(i_1, i_2)$.
n	observed count vector, $n' = (n(i))_{i\in\mathrm{I}}$.
I_a	for $a \subseteq \Delta$, marginal subvector of I with marginal cells i_a, $i_a \in \mathrm{I}_a = \times_{\delta\in a}\mathrm{I}_\delta$.
$\lvert\mathrm{I}_a\rvert$	marginal cells total.
A	generating class= cliques set for the associated graph.
K	set of all subsets of the elements in the generating class A.
$u_a(i_a)$	parameters for a log-linear model depending only on the values of the variables in subset a, i_a, for $a \subseteq \Delta$. $u_a(i_a)$ = constant, when $a = \emptyset$; $u_a(i_a)$ = main effect, when $\lvert a\rvert = 1$; $u_a(i_a)$ = interaction, when $\lvert a\rvert > 1$.
$u_{l_r^t m_r^t}(i_r^t j_r^t)$	$r = 1, 2, ..., ked(t)$; t=1,2,...,S. First-order interaction. It can be identified with the edge e_r^t joining variable l_r^t to variable m_r^t at the value combination (i_r^t, j_r^t), where e_r^t is the r-th element in the colour class E_t. ked is the vector of dimension S of number of edges in each edge class.
$n_a(i_a)$	$n_a(i_a) = \sum_{j:j_a=i_a} n(j)$, marginal count for i_a.
$m_a(i_a)$	$m_a(i_a) = \sum_{j:j_a=i_a} m(j)$, marginal expected frequency for i_a.
$\widehat{m}_a(i_a)$	$\widehat{m}_a(i_a) = \sum_{j:j_a=i_a} \widehat{m}(j)$, marginal estimated expected frequency for i_a.
$\lvert n\rvert$	Total observed count, $\sum_{i\in\mathrm{I}} n(i)$.

3.3 Introduction and definition

In graphical models for discrete variables it could be useful to restrict some of their parameters, for example by equating a subset of parameters, to get a better fit or understanding of some data. For instance, both symmetry and quasi-symmetry models

for two-way contingency tables defined in section **2.5** are graphical models in which some of their parameters are equated. Those restrictions are generally imposed by the characteristics of the problem from which the information came from. Once the models are restricted, more parsimonious models are obtained and as a consequence we get more precise estimators. This is because the overall variability of the estimators in a model with fewer parameters about the true values for the cells is less than the overall variability for the model containing more parameters, as discussed by Altham (1984) in general, by Altham (1994) for generalized linear models, and by Bishop *et al.* (1975, p. 313) for models for counted data.

In this chapter we define new types of graphical log-linear models, restricted or coloured graphical log-linear models, which are particular cases of hierarchical log-linear models, and for simplicity we will call them *RGLL models*. RGLL models can be represented graphically including equalities between some parameters. We consider two kinds of restrictions: restrictions generated by classes in which the main effects in the same restriction class are identical in all their categories and restrictions generated by classes in which the first-order interactions in the same class are identical. Such restrictions and the corresponding models can be represented by colouring the associated graph.

RGLL models are analogous to graphical Gaussian models with edge and vertex symmetries introduced by Højsgaard and Lauritzen (2005, 2007, 2008). In the continuous case due to the characteristics of the Gaussian distribution, the symmetry restrictions or vertex and edge colourings are placed on the concentration matrix (inverse covariance matrix) which is formed by the parameters that determine the graphical models, whereas in the discrete case using vertex and edge colourings imposes restrictions only on parameters associated to main effects and first-order interactions, but does not imply restrictions on parameters associated to interactions of second or higher order, in this sense RGLL models can be still further generalized by restricting parameters associated to interactions of two or higher order.

Symmetry and quasi-symmetry models, which were presented in section **2.5** based on the work developed by Caussinus (1965) and later discussed for instance by Agresti (2002a, p. 423-431) or Bishop *et al.* (1975, p. 282-293), can be seen as particular cases of RGLL models with two variables. In this sense, RGLL models could be considered as a generalization of these models as suggested by Højsgaard and Lauritzen (2008, p. 1025). Other generalizations of symmetry and quasi-symmetry were discussed in section **2.6**. Graphical log-linear models may be seen as a particular case of RGLL models because we are able to define a RGLL model with no restrictions imposed on the parameters and this corresponds to a graphical log-linear model.

The following relations are observed between RGLL models and other log-linear

models:

Log-linear models$\supseteq$ Hierarchical log-linear models $\supseteq$ Restricted graphical log-linear models $\supseteq$ Graphical log-linear models.

Because we define RGLL models as hierarchical log-linear models whose generating class is equal to the cliques set of the corresponding graph, only adding some equalities between some parameters without setting to zero additional parameters, then we have models that inherit all properties that we originally had on graphical log-linear models. That is, we can obtain all conditional and marginal independences that the model represents by analyzing the corresponding graphs.

Advantages of RGLL models: a) they can give a better fit to the data when they reflect better the underlying structure, b) they can help to get a better understanding of the relations among cells on the table, c) they can be used to identify marginal and conditional independences among variables using the graphical concept of separator sets as in graphical log-linear models, d) they can be more parsimonious than the corresponding models without restrictions and as a consequence the estimators are more precise, e) they can present the information in a visual, intuitive, and accessible way, mainly when the number of variables is not so large, for example fewer than 10 variables, and finally, f) some models that generalize symmetry in discrete models in terms of both graph theory and tables of contingency can be seen as RGLL models, as we will see in Chapter 4.

RGLL models assume that we do not have structural zeros, *i.e.* $p(i) > 0$ or $m(i) > 0$ if the cells counts are realizations of multinomial or Poisson distributed random variables, respectively, although we could have sampling zeros, in this case the estimators might not exist as discussed in section **2.7**. When there are sampling zeros the model could provide zero and nonzero estimates, zero estimates are obtained when cells are arranged so that some of the configuration cells are empty. In this case we could adjust the corresponding degrees of freedom associated to the deviance as discussed in section **2.7**.

In this chapter we define restricted graphical log-linear models. We also present the necessary theory to understand and estimate them, as well as their relation with symmetry and quasi-symmetry models for two-way contingency tables.

Definition 3.1. A hierarchical log-linear model

$$\log m(i) = \sum_{a \in K} u_a(i_a),$$

where K is the set of all subsets of the elements in the generating class, is a *restricted graphical log-linear model (RGLL model)* with associated graph $G = (V, E)$ if it satisfies two properties: a) its generating class is C, the cliques set in the associated graph,

and b) the set of variables V and the set of first-order interactions E are partitioned as follows. The set V is partitioned into $V_1, ..., V_T$, with $V_j \neq \emptyset$, for j=1,...,T, with $T \in \{1, 2, ..., |V|\}$, such that the main effects of the variables in V_j are equal in all their levels in I^{V_j}. The set E is partitioned into $E_1, ..., E_S$, with $E_j \neq \emptyset$, for j=1,...,S, with $S \in \{1, 2, ..., |E|\}$, such that the interactions in every E_j are equal.

Example 3.1. An example of a RGLL model is the model with associated graph $G(V, E)$ generated by $\{\{W, X\}, \{X, Y\}, \{Y, Z\}\}$ in which $V = \{W, X, Y, Z\}$, where all variables are binary, and $V = (V_1, V_2)$, with $V_1 = \{W, X\}$ and $V_2 = \{Y, Z\}$, and set of first-order interactions $E = (E_1, E_2)$, where $E_1 = \{u_{WX}(01), u_{WX}(10), u_{XY}(01), u_{XY}(10), u_{YZ}(01), u_{YZ}(10)\}$ and $E_2 = \{u_{WX}(00), u_{WX}(11), u_{XY}(00), u_{XY}(11), u_{YZ}(00), u_{YZ}(11)\}$. The vertex partition by definition is equivalent to having

$$u_W(i) = u_X(i);\ i = 0, 1;$$
$$u_Y(i) = u_Z(i);\ i = 0, 1;$$

and the edge partition is equivalent to having

$$\begin{aligned} u_{WX}(01) = u_{WX}(10) &= u_{XY}(01) = u_{XY}(10) &= u_{YZ}(01) = u_{YZ}(10); \\ u_{WX}(00) = u_{WX}(11) &= u_{XY}(00) = u_{XY}(11) &= u_{YZ}(00) = u_{YZ}(11). \end{aligned}$$

We have some observations about the definition:

a) The graphs associated to RGLL models are similar to the independence or interaction graphs defined for graphical log-linear models (section **2.8.1**), in the sense that the underlying simple graph, which we denote as $G^u(V^u, E^u)$, associated to the RGLL model is an independence graph associated to a graphical log-linear model. This means that every variable is represented with a circle or dot and that two variables are joined with edges if the first-order interaction that contains both variables is included in the model. There are as many edges between two variables as different permutations among the categories of both variables. On the other hand, in graphical log-linear models there is only one edge between two variables included in a first-order interaction. It can be noted that the first-order interactions are parameters that determine the graph and they are also some kind of measure of the association between two variables in a model.

b) The clique concept defined in graph theory, Appendix A, is the same concept used in graphical models. In RGLL models we must consider that there are multiple edges between variables, so that if there is a clique between certain variables then it is not important which edges are used to obtain it, the important thing is knowing that

there is a clique.

c) When $|V_i| = 1$ or $|E_j| = 1$ for some i or j, $i = 1, ..., T$, $j = 1, ..., S$ we have vertex or edge atomic classes, respectively, otherwise we have composite classes, as defined for the continuous case in Højsgaard and Lauritzen, 2008, p. 1009. Atomic classes are those in which all the corresponding parameters are not restricted. A RGLL model with only atomic classes is a graphical log-linear model.

d) It is assumed that all variables have the same categories, *i.e.* I_δ is the same for all $\delta \in V$ or I^{V_i} is the same for all V_i, $i = 1, ..., T$, which implies that $|\mathrm{I}_\delta| = J$, for all $\delta \in V$. This condition could be relaxed assuming that only those variables in the same composite vertex classes should have the same categories. However, we observe that in order to select a model we need to join vertex colour classes, which makes necessary that all variables have the same categories, we will also define in the following chapters some models called label invariant that only make sense if we have the same categories for all variables, additionally, the programming is simplified and symmetry interpretations are possible under this assumption. Because of these reasons, we will assume that we have the same categories for all variables and when we do not we will let the reader know.

From the definition of RGLL model, we see that the number of vertex classes T and the number of edge classes S satisfy $1 \leq T \leq |V|$ and $1 \leq S \leq |E|$. There are particular cases according to which values take S and T:

1. If $S = |E|$ and $T = |V|$, we have a graphical log-linear model because all parameters are unrestricted.

2. If $1 \leq T \leq |V|$ and $S = |E|$, we have a RGLL model with restrictions only on main effects.

3. If $1 \leq S \leq |E|$ and $T = |V|$, we have a RGLL model with restrictions only on first-order interactions.

RGLL models are represented using colourings, understanding colourings as in graph theory (Appendix A). This means that we can represent the model using coloured graphs with multiple edges without loops, also known as multigraphs for instance by Diestel (2005, p.28) who also discusses that when we usually work with multigraphs we could call them just *graphs*, and given a coloured multigraph we can determine the restrictions on the parameters for its corresponding induced RGLL model. We have three particular cases of colourings: vertex, edges, and, vertex and edges colourings.

In the following sections we give different instances of RGLL models and some examples. Firstly, in section **3.4** we discuss the case in which $1 \leq T \leq |V|$ and $S = |E|$ called vertex colouring. Secondly, in section **3.5** we discuss the case in which

$1 \leq S \leq |E|$ and $T = |V|$ called edge colouring. Finally, in section **3.6** we present the general case in which $1 \leq T \leq |V|$ and $1 \leq S \leq |E|$ called vertex and edge colouring. We present the models in this order so that we can gradually introduce new notation and concepts.

3.4 Vertex colouring

Consider a RGLL model with restrictions only on the main effects, *i.e.* we have the RGLL model with graph $G = (V, E)$, where V is partitioned into $V_1, ..., V_T$, with $V_i \neq \emptyset$, for i=1,...,T, where $T \in \{1, 2, ..., |V|\}$ and the set of first-order interactions E is partitioned into $E_1, ..., E_S$, with $S = |E|$, so that the main effects for the variables in every set V_i are equal in all their levels in I^{V_i} and the first-order interaction terms are unrestricted.

Definition 3.2. Two vertices, X and Y, are in the same vertex colour class if and only if $u_X(i) = u_Y(i)$, for every $i = 1, ..., |\mathrm{I}_X| = |\mathrm{I}_Y|$, where
u_X, u_Y are main effects; and
$|\mathrm{I}_X| = |\mathrm{I}_Y|$ are the number of categories for the variables X and Y.

Suppose that we have a partition of V into $V_1, V_2, ..., V_T$ vertex colour classes; *i.e.* $V = (V_1, ..., V_T)$, in which every variable has J levels, and where $V_i = \left\{v_1^i, ..., v_{kver(i)}^i\right\}$, $i = 1, ..., T$, remember that v_k^i is the kth vertex in the colour class V_i, i=1,...,T; k=1,...,$kver(i)$; and $kver$ is the vector of dimension T of number of vertices in each vertex class. We define

$u_{v_k^i}(j)$: main effect for the variable v_k^i in the category j, for i=1,...,T; k=1,...,$kver(i)$, and j=1, 2,..., J.

Then, we have the following equality restrictions

$$u_{v_1^i}(j) = u_{v_2^i}(j) = ... = u_{v_{kver(i)}^i}(j) = u_i(j),\ j = 1, ..., J;\ i = 1, ..., T,$$

where $u_i(j)$ represents all parameters that are restricted to be equal.

Similarly, if we have a partition of the variables set in such a way that the main effects of the variables in some class are equal in all their levels, then we can assign the same colour to all the vertices in the class. This means that there is a correspondence between the restrictions in a RGLL model restricting the main effects only with a vertex colouring.

A final observation is that for simplicity we colour in black all edges because in these models all edges are in atomic classes, and if we did not make this assumption, we would need a different colour for each edge, which could make difficult the graphical representation.

Example 3.2. Suppose that we have four variables, X, Y, Z, W, and suppose that every variable has the same number of categories, $|\mathrm{I}_X| = |\mathrm{I}_Y| = |\mathrm{I}_Z| = |\mathrm{I}_W|$. A RGLL model restricted only on its main effects is the model with graph $G(V, E)$, with generating class $A = \{\{X, Y\}, \{Y, Z\}, \{Y, W\}\}$, with $V = \{X, Y, Z, W\}$ in which $V = (V_1, V_2)$, with $V_1 = \{X, Y\}$ and $V_2 = \{Z, W\}$ and $E = (E_1, ..., E_S)$, with $S = |E|$.

We show next the log-linear expansion corresponding to the generating class to identify which are the parameters of the model without restrictions

$$\log m(i, j, k, l) = u + u_X(i) + u_Y(j) + u_Z(k) + u_W(l) + u_{XY}(ij) + u_{YZ}(jk) + u_{YW}(jl).$$

In the given notation, $kver(1) = 2$, $v_1^1 = X$, $v_2^1 = Y$ and $kver(2) = 2$, $v_1^2 = Z$, $v_2^2 = W$. This colouring is represented by the following equalities

$$u_{v_1^1}(i) = u_X(i) = u_Y(i) = u_{v_2^1}(i) = u_1(i),\ i = 1, ..., |\mathrm{I}_X|\,;$$
$$u_{v_1^2}(i) = u_Z(i) = u_W(i) = u_{v_2^2}(i) = u_2(i)\ i = 1, ..., |\mathrm{I}_Z|\,.$$

Then, we have the parameter $u_1(i)$ instead of the parameters $u_X(i)$ and $u_Y(i)$, $i = 1, ..., |\mathrm{I}_X|$, and the parameter $u_2(i)$ instead of the parameters $u_Z(i)$ and $u_W(i)$, $i = 1, ..., |\mathrm{I}_Z|$. The model is

$$\log m(i, j, k, l) = u + u_1(i) + u_1(j) + u_2(k) + u_2(l) + u_{XY}(ij) + u_{YZ}(jk) + u_{YW}(jl).$$

And if for example $|\mathrm{I}_X| = 2$, the model has the associated graph shown in figure 3.1, the graphical log-linear model or unrestricted model has 21 parameters while the RGLL model has 17.

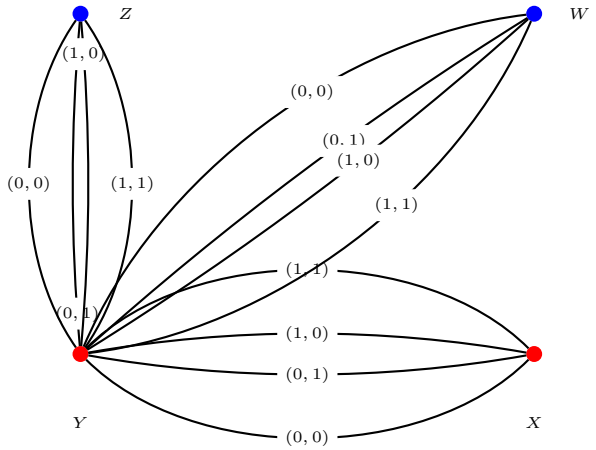


Figure 3.1: Vertex colouring for the RGLL model with generating class $\{\{X,Y\}, \{Y,Z\}, \{Y,W\}\}$*, where all variables are binary, with vertices* $V = (V_1, V_2)$*,* $V_1 = \{X,Y\}$ *and* $V_2 = \{Z,W\}$*, and with first-order interactions set* $E = (E_1, ..., E_S)$*, with* $S = |E| = 16$*.*

3.5 Edge colouring

Suppose that we have a RGLL model with restrictions only on first-order interactions, *i.e.* we have a RGLL model with graph $G = (V, E)$, where the first-order interaction set E is partitioned into $E_1, ..., E_S$, $E_i \neq \emptyset$, for i=1,...,S, with $S \in \{1, 2, ..., |E|\}$ and the set of variables V is partitioned into $V_1, ..., V_T$, with $T = |V|$, so that the first-order interactions in every set E_i are identical and the main effects are unrestricted.

Definition 3.3. Two edges, one joining X with Y at the variable values (i_1, j_1) and other joining Z with W at the variable values (k_1, l_1), are in the same edge colour class if and only if $u_{XY}(i_1, j_1) = u_{ZW}(k_1, l_1)$, where
$X, Y, Z, W \in V$, the set of variables;
$|\mathrm{I}_X|$, $|\mathrm{I}_Y|$, $|\mathrm{I}_Z|$, $|\mathrm{I}_W|$ are the number of categories for the variables X, Y, Z, and W, respectively;
$i_1 \in \mathrm{I}_X = \{1, ..., |\mathrm{I}_X|\}$, $j_1 \in \mathrm{I}_Y = \{1, ..., |\mathrm{I}_Y|\}$, $k_1 \in \mathrm{I}_Z = \{1, ..., |\mathrm{I}_Z|\}$, and $l_1 \in \mathrm{I}_W = \{1, ..., |\mathrm{I}_W|\}$; and
$u_{XY}(i_1, j_1)$, $u_{ZW}(k_1, l_1)$ are first-order interactions.

Suppose that we have a partition of the edge set E into $E_1, E_2, ..., E_S$ edge colour classes or $E = (E_1, E_2, ..., E_S)$, with $E_t = \left\{e^t_1, ..., e^t_{ked(t)}\right\}$, t=1,2,...,S, where the following notation is used.

e_r^t: r-th element in the colour class E_t.
$u_{l_r^t m_r^t}(i_r^t j_r^t)$, $r = 1, 2, ..., ked(t)$; t=1,2,...,S : first-order interaction between l_r^t and m_r^t at the values $(i_r^t,\ j_r^t)$. It can be identified with the edge e_r^t joining variable l_r^t to variable m_r^t for the value combination (i_r^t, j_r^t).
l_r^t: r-th variable for the t class in the first entry of the first-order interaction $u_{l_r^t m_r^t}(i_r^t j_r^t)$.
m_r^t: r-th variable for the t class in the second entry of the first-order interaction $u_{l_r^t m_r^t}(i_r^t j_r^t)$.
i_r^t: category of l_r^t.
j_r^t: category of m_r^t.

Observe that partitioning the edge set or the first-order interactions set is equivalent because there is a one-to-one correspondence between the $u_{l_r^t m_r^t}(i_r^t j_r^t)$ parameters and the corresponding edges.

Then, under an edge colouring we have the restrictions

$$u_{l_1^t m_1^t}(i_1^t j_1^t) = ... = u_{l_{ked(t)}^t m_{ked(t)}^t}(i_{ked(t)}^t j_{ked(t)}^t) = u_{E_t},$$

where u_{E_t} is the parameter that represents all parameters in the same colour class E_t, t=1,2,...,S.

Similarly, if we have a partition of the set of first-order interactions in which all interactions in the same class are identical, then we can assign to every class a different colour. This means that there is a correspondence between all restrictions in a RGLL model with restrictions only on first-order interactions with an edge colouring.

As in vertex colouring, for simplicity we colour in black all vertices because all vertices are in atomic classes, and if we did not make this assumption, we would need a different colour for each vertex, which could make difficult the graphical representation.

Example 3.3. Suppose that we have four binary variables, X, Y, Z, and W, a RGLL model with restrictions only on first-order interactions is the model with graph $G(V, E)$, with generating class $\{\{X, Y\}, \{W, Y\}, \{W, Z\}, \{X, Z\}\}$, where $V = \{W, X, Y, Z\}$ and $V = (V_1, ..., V_T)$, $T = |V| = 4$, and where the partition of the set of first-order interactions E, and as a consequence of the set of edges obtained by identifying each parameter with an edge, is $E = (E_1, E_2, E_3, E_4)$. We give explicitly each class below.

Firstly, we show the log-linear expansion corresponding to the generating class to identify which are the parameters of the model without restrictions

$$\log m(i, j, k, l) = u + u_W(i) + u_X(j) + u_Y(k) + u_Z(l) + u_{XY}(jk) + u_{WY}(ik) + u_{WZ}(il) + u_{XZ}(jl),$$

with i, j, k, l=0, 1. Then, the first-order interaction set or edge set, E, is

$$E = \{u_{XY}(00), u_{XY}(01), u_{XY}(10), u_{XY}(11), u_{WY}(00), u_{WY}(01), u_{WY}(10), u_{WY}(11)$$
$$u_{WZ}(00), u_{WZ}(01), u_{WZ}(10), u_{WZ}(11), u_{XZ}(00), u_{XZ}(01), u_{XZ}(10), u_{XZ}(11)\}.$$

We define the four classes E_1, E_2, E_3, E_4 as follows:

$$E_1 = \{u_{XY}(00), u_{XY}(11), u_{WZ}(00), u_{WZ}(11)\},$$
$$E_2 = \{u_{XY}(01), u_{XY}(10), u_{WZ}(01), u_{WZ}(10)\},$$
$$E_3 = \{u_{WY}(00), u_{WY}(11), u_{XZ}(00), u_{XZ}(11)\},$$
$$E_4 = \{u_{WY}(01), u_{WY}(10), u_{XZ}(01), u_{XZ}(10)\}.$$

In this case, $ked(i) = 4$, i=1,...,4. Although an edge is an element that can not be equated to a parameter, for simplicity we will use the sign $\approx$ to establish correspondence between edges and parameters. Then, we have

$$e_1^1 \approx u_{l_1^1 m_1^1}(i_1^1 j_1^1) = u_{XY}(00),\ e_2^1 \approx u_{l_2^1 m_2^1}(i_2^1 j_2^1) = u_{XY}(11),$$
$$e_3^1 \approx u_{l_3^1 m_3^1}(i_3^1 j_3^1) = u_{WZ}(00),\ e_4^1 \approx u_{l_4^1 m_4^1}(i_4^1 j_4^1) = u_{WZ}(11).$$
$$e_1^2 \approx u_{l_1^2 m_1^2}(i_1^2 j_1^2) = u_{XY}(01),\ e_2^2 \approx u_{l_2^2 m_2^2}(i_2^2 j_2^2) = u_{XY}(10),$$
$$e_3^2 \approx u_{l_3^2 m_3^2}(i_3^2 j_3^2) = u_{WZ}(01),\ e_4^2 \approx u_{l_4^2 m_4^2}(i_4^2 j_4^2) = u_{WZ}(10).$$
$$e_1^3 \approx u_{l_1^3 m_1^3}(i_1^3 j_1^3) = u_{WY}(00),\ e_2^3 \approx u_{l_2^3 m_2^3}(i_2^3 j_2^3) = u_{WY}(11),$$
$$e_3^3 \approx u_{l_3^3 m_3^3}(i_3^3 j_3^3) = u_{XZ}(00),\ e_4^3 \approx u_{l_4^3 m_4^3}(i_4^3 j_4^3) = u_{XZ}(11).$$
$$e_1^4 \approx u_{l_1^4 m_1^4}(i_1^4 j_1^4) = u_{WY}(01),\ e_2^4 \approx u_{l_2^4 m_2^4}(i_2^4 j_2^4) = u_{WY}(10),$$
$$e_3^4 \approx u_{l_3^4 m_3^4}(i_3^4 j_3^4) = u_{XZ}(01),\ e_4^4 \approx u_{l_4^4 m_4^4}(i_4^4 j_4^4) = u_{XZ}(10).$$

$$l_i^j = X,\ m_i^j = Y;\ i = 1, 2,\ j = 1, 2.$$
$$l_i^j = W,\ m_i^j = Z;\ i = 3, 4,\ j = 1, 2.$$
$$l_i^j = W,\ m_i^j = Y;\ i = 1, 2,\ j = 3, 4.$$
$$l_i^j = X,\ m_i^j = Z;\ i = 3, 4,\ j = 3, 4.$$
$$i_r^t = 0, j_r^t = 0;\ r = 1, 3,\ t = 1, 3.$$
$$i_r^t = 1, j_r^t = 1;\ r = 2, 4,\ t = 1, 3.$$
$$i_r^t = 0, j_r^t = 1;\ r = 1, 3,\ t = 2, 4.$$
$$i_r^t = 1, j_r^t = 0;\ r = 2, 4,\ t = 2, 4.$$

Given this colouring, we have the following restrictions

$$u_{XY}(00) = u_{XY}(11) = u_{WZ}(00) = u_{WZ}(11) = u_{E_1},$$
$$u_{XY}(01) = u_{XY}(10) = u_{WZ}(01) = u_{WZ}(10) = u_{E_2},$$
$$u_{WY}(00) = u_{WY}(11) = u_{XZ}(00) = u_{XZ}(11) = u_{E_3},$$
$$u_{WY}(01) = u_{WY}(10) = u_{XZ}(01) = u_{XZ}(10) = u_{E_4}.$$

Under this model there are only four parameters u_{E_1}, u_{E_2}, u_{E_3}, and u_{E_4}, instead of the 16 parameters contained in E_1, E_2, E_3, and E_4.

The corresponding coloured graph is shown in figure 3.2.

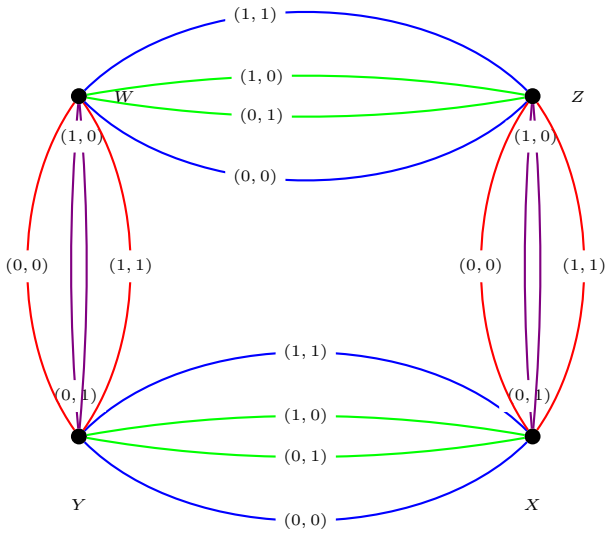


Figure 3.2: Edge colouring for the RGLL model with generating class $\{\{X,Y\}, \{W,Y\}, \{W,Z\}, \{X,Z\}\}$ *where all variables are binary, with first-order interactions set* $E = (E_1, E_2, E_3, E_4)$, $E_1 = \{u_{XY}(00), u_{XY}(11), u_{WZ}(00), u_{WZ}(11)\}$, $E_2 = \{u_{XY}(01), u_{XY}(10), u_{WZ}(01), u_{WZ}(10)\}$, $E_3 = \{u_{WY}(00), u_{WY}(11), u_{XZ}(00), u_{XZ}(11)\}$, $E_4 = \{u_{WY}(01), u_{WY}(10), u_{XZ}(01), u_{XZ}(10)\}$ *and* $V = (V_1, ..., V_T)$, $T = |V| = 4$.

3.6 Vertex and Edge colouring

In this case we use the definitions of vertex and edge colourings given in 3.2 and 3.3.

Consider a RGLL model with associated graph $G = (V, E)$, with $(V_1, V_2, ..., V_T)$ a partition of V, $V_i \neq \emptyset$, for i=1,...,T, with $T \in \{1, 2, ..., |V|\}$, and $(E_1, E_2, ..., E_S)$ a partition of E, $E_i \neq \emptyset$, for i=1,...,S, with $S \in \{1, 2, ..., |E|\}$. Suppose that every variable

belonging to a non-atomic class has J levels. We have:

$$V_i = \left\{v_1^i, ..., v_{kver(i)}^i\right\} \Leftrightarrow u_{v_1^i}(j) = u_{v_2^i}(j) = ... = u_{v_{kver(i)}^i}(j) = u_i(j),\ j = 1, ..., J;\ i = 1, ..., T.$$

$$E_t = \left\{e_1^t, ..., e_{ked(t)}^t\right\} \Leftrightarrow u_{l_1^t m_1^t}(i_1^t j_1^t) = ... = u_{l_{ked(t)}^t m_{ked(t)}^t}(i_{ked(t)}^t j_{ked(t)}^t) = u_{E_t};\ t = 1, ..., S.$$

Then, given a colouring, we get a number of restrictions on the parameters and vice versa. This means that there is a correspondence between the restrictions that define a RGLL model with vertex and edge colourings.

As we have discussed at the end of section **3.3** vertex colouring could be seen as a particular case of these models if we let every edge in an atomic class, *i.e.* if $S = |E|$, and similarly, edge colouring could be seen as a particular case of these models if every vertex conforms an atomic class, *i.e.* if $T = |V|$.

Example 3.4. Suppose that we have three binary variables A, C, and M, corresponding to alcohol, cigarette, and marijuana consumption, respectively. Every variable has two categories, *Yes* and *No*, represented with 0 and 1, respectively. We present a graphical log-linear model with these variables, and give an example of a RGLL model.

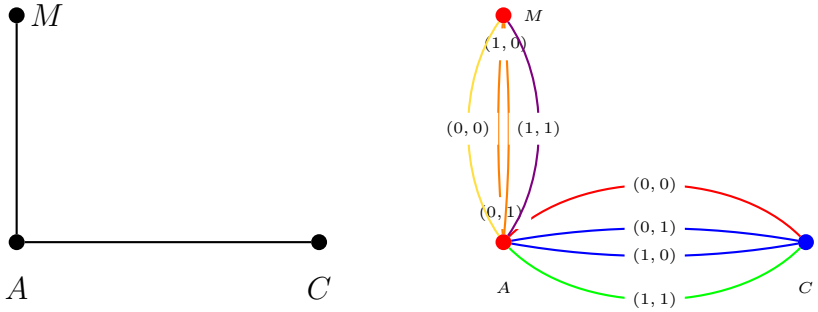
Suppose that we have the graphical log-linear model with generating class $\{\{A, C\}, \{A, M\}\}$ with associated graph $G^u(V^u, E^u)$ shown in figure 3.3(a). This model can be expressed as follows,

$$\log m(i, j, k) = u + u_A(i) + u_C(j) + u_M(k) + u_{AC}(ij) + u_{AM}(ik),\ i, j, k = 0, 1. \quad (3.1)$$

Now, consider we add some restriction to the graphical log-linear model (3.1) as follows,

$$\begin{aligned} u_A(0) &= u_M(0); \\ u_A(1) &= u_M(1); \\ u_{AM}(01) &= u_{AM}(10); \\ u_{AC}(01) &= u_{AC}(10). \end{aligned}$$

We get the graph given in figure 3.3(b). We observe that $V = \{A, C, M\}$, $V = (V_1, V_2)$, with $V_1 = \{A, M\}$ and $V_2 = \{C\}$. On the other hand, the set of first-order interactions or their corresponding edge set $E = (E_1, E_2, E_3, E_4, E_5, E_6)$, where



(a) Graphical model (b) RGLL model

Figure 3.3: Log-linear models with generating class $\{\{A, C\}, \{A, M\}\}$, $\log m(i, j, k) = u + u_A(i) + u_C(j) + u_M(k) + u_{AC}(ij) + u_{AM}(ik)$*: (a) graphical model; (b) RGLL model with vertex set* $V = (V_1, V_2)$, $V_1 = \{A, M\}$ *and* $V_2 = \{C\}$*, and first-order interactions set partitioned into* $E = (E_1, E_2, E_3, E_4, E_5, E_6)$, $E_1 = \{u_{AM}(01), u_{AM}(10)\}$, $E_2 = \{u_{AC}(01), u_{AC}(10)\}$, $E_3 = \{u_{AM}(00)\}$, $E_4 = \{u_{AM}(11)\}$, $E_5 = \{u_{AC}(00)\}$, $E_6 = \{u_{AC}(11)\}$.

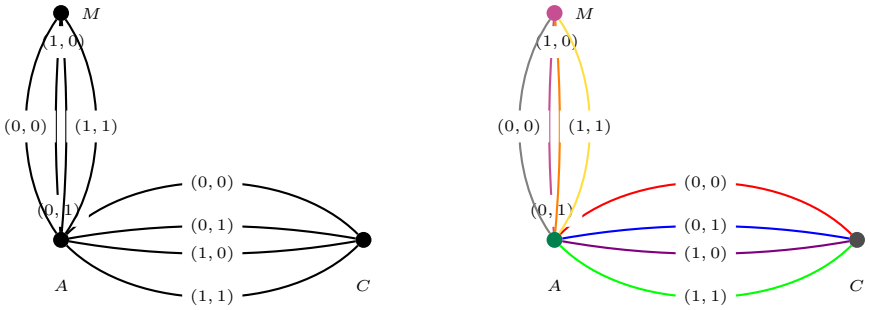
$$E = \{u_{AM}(00), u_{AM}(01), u_{AM}(10), u_{AM}(11), u_{AC}(00), u_{AC}(01), u_{AC}(10), u_{AC}(11)\},$$

and $E_1 = \{u_{AM}(01), u_{AM}(10)\}$, $E_2 = \{u_{AC}(01), u_{AC}(10)\}$, $E_3 = \{u_{AM}(00)\}$, $E_4 = \{u_{AM}(11)\}$, $E_5 = \{u_{AC}(00)\}$, and $E_6 = \{u_{AC}(11)\}$.

Observe that the underlying simple graph G^u is the one given in figure 3.3(a). Using any of the graphs in figure 3.3 we see that the generating class is identical to the corresponding cliques set, which together with the restrictions applied to parameters of the model, means that we have a RGLL model.

Observe that both the restricted and non-restricted graphical models have the same generating class; however, the restrictions given to the graphical model, model (3.1), appearing in the RGLL model reduce the number of parameters from 15 in the graphical model to 11 in the RGLL model. It is also important to notice that from the coloured graph 3.3(b), we could identify the associated RGLL model.

The graphical log-linear model (3.1) expressed as a particular case of a RGLL model is shown in figure 3.4, in this case each vertex and each edge are in atomic classes; *i.e.* $S = |E|$ and $T = |V|$. In figure 3.4 we show two available ways of representing colouring for atomic classes, the first one, figure 3.4(a), uses black to represent every atomic class, and the second one, figure 3.4(b), is a representation in which each vertex and edge atomic class has a different colour. Both conventions are useful, but it is not difficult to see that the representation in figure 3.4(b) becomes unfeasible when there are a lot of variables or high-dimensional problems.



(a) (b)

Figure 3.4: Two colourings for the graphical log-linear model with generating class $\{\{A, C\}, \{A, M\}\}$, $\log m(i, j, k) = u + u_A(i) + u_C(j) + u_M(k) + u_{AC}(ij) + u_{AM}(ik)$, *expressed as a particular case of a RGLL model. In this model* $V = (V_1, V_2, V_3)$, $V_1 = \{A\}$, $V_2 = \{C\}$, *and* $V_3 = \{M\}$. *The first-order interactions set* $E = (E_1, E_2, E_3, E_4, E_5, E_6, E_7, E_8)$, $E_1 = \{u_{AM}(10)\}$, $E_2 = \{u_{AM}(01)\}$, $E_3 = \{u_{AM}(00)\}$, $E_4 = \{u_{AM}(11)\}$, $E_5 = \{u_{AC}(10)\}$, $E_6 = \{u_{AC}(01)\}$, $E_7 = \{u_{AC}(00)\}$, $E_8 = \{u_{AC}(11)\}$*: a) atomic classes in black; b) each atomic class with a different colour.*

3.7 Symmetry and quasi-symmetry models as RGLL models

Quasi-symmetry and symmetry models, already discussed in section **2.5**, can be expressed as special cases of RGLL models, in particular as edge colourings, and vertex and edge colourings, respectively. Every vertex in the quasi-symmetry model belongs to a different atomic class and there are $|\mathrm{I}|$ atomic edge colour classes for every $u_{XY}(ii)$ interaction and $\binom{|\mathrm{I}|}{2}$ different edge colour classes for the interactions $u_{XY}(ij) = u_{XY}(ji), i \neq j$.

In symmetry models we have the same edge colour classes, but there is only one vertex colour class formed by both vertices.

Then, RGLL models can be considered as one way of extending the definition of symmetry and quasi-symmetry models as given by for example Agresti (2002a, p. 423-426) for two dimensions to the case of dimension higher than two.

Example 3.5. Consider two binary variables labeled as C and A. The variable C corresponds to cigarette consumption, with categories: 0, for fewer than two packets

of cigarettes, and 1, for two or more packets of cigarettes. The variable A corresponds to alcohol consumption with categories: 0, for fewer than one drink, and 1, for two or more drinks. The data are presented in a contingency table (table 3.1).

Cigarette	**Alcohol**	
	0	1
0	540	53
1	28	386

Table 3.1: Hypothetical square contingency table corresponding to alcohol and cigarette consumption for a sample of 1007 students.

Consider the saturated log-linear model,

$$\log m(i,j) = u + u_C(i) + u_A(j) + u_{CA}(ij),\ i,j = 0,1,$$

which is also a graphical log-linear model with generating class $\{\{C,\ A\}\}$ and graph $G^u(V^u, E^u)$, $V^u = \{C, A\}$ and edge set $E^u = \{\{C, A\}\}$. Now consider the RGLL model shown in the graph $G = (V, E)$ given in figure 3.5(a) where V=$\{C, A\}$ and E is the set of first-order interactions $E = \{u_{CA}(01), u_{CA}(10), u_{CA}(00), u_{CA}(11)\}$. Observe that this is a RGLL model because the generating class $\{\{C,\ A\}\}$ coincides with the corresponding cliques. From $G = (V, E)$, we observe that the set of first-order interactions $E = (E_1, E_2, E_3)$, where $E_1 = \{u_{CA}(00)\}$, $E_2 = \{u_{CA}(11)\}$, and $E_3 = \{u_{CA}(01), u_{CA}(10)\}$. The edges in E_3 belong to the same colour class indicating that the corresponding interactions are identical. The remaining edges belong to different atomic colour classes. As we discussed before, if we do not want to use different colours for different atomic classes, we could agree to represent all atomic classes with black lines (figure 3.5(b)). Additionally, $V = (V_1, V_2)$, with $V_1 = \{C\}$ and $V_2 = \{A\}$, indicating that the main effects are not restricted.

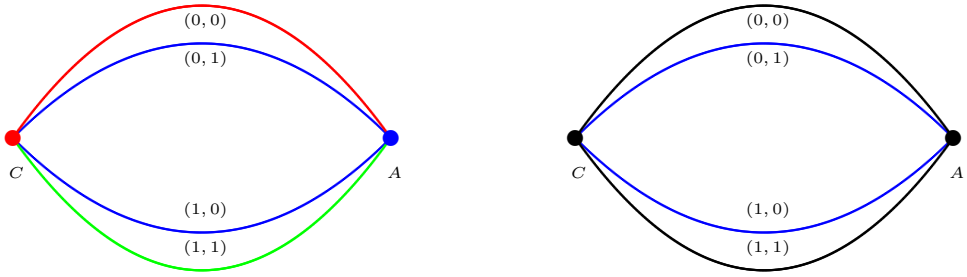
Then, the model associated to the graph shown in figure 3.5(a) can be expressed as

$$\log m(i,j) = u + u_C(i) + u_A(j) + u_{CA}(ij),\ i,j = 0,1;$$

in which

$$u_{CA}(ij) = u_{CA}(ji),\ i,j = 0,1.$$

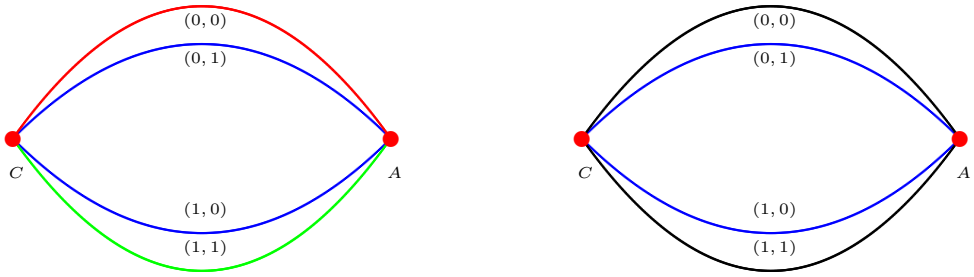
Then, the quasi-symmetry model for V=$\{C, A\}$ is the RGLL model generated by $\{C, A\}$ with vertex colour classes $V = (V_1, V_2)$, with $V_1 = \{C\}$ and $V_2 = \{A\}$, and first-order interaction terms or their corresponding edges $E = (E_1, E_2, E_3)$ with $E_1 = \{u_{CA}(00)\}$, $E_2 = \{u_{CA}(11)\}$, and $E_3 = \{u_{CA}(01), u_{CA}(10)\}$, whose associated graph $G(V, E)$ is given in figure 3.5(a) or figure 3.5(b).



(a) Graph colouring 1 (b) Graph colouring 2

Figure 3.5: Two alternative graph colourings for a quasi-symmetry model with two binary variables. $V = \{C, A\}$ is partitioned into V_1 and V_2, with $V_1 = \{C\}$ and $V_2 = \{A\}$. The first-order interactions set E is partitioned into $E_1 = \{u_{CA}(00)\}$, $E_2 = \{u_{CA}(11)\}$, and $E_3 = \{u_{CA}(01), u_{CA}(10)\}$: a) atomic classes in different colour; b) atomic classes in black.

Now, suppose that we have the RGLL model shown in the graph $G(V, E)$ given in figure 3.6(a), where V=$\{C, A\}$ and $E = \{u_{CA}(01), u_{CA}(10), u_{CA}(00), u_{CA}(11)\}$. From $G = (V, E)$, we observe that we have the same edge or first-order interactions partition as in the previous model, $E = (E_1, E_2, E_3)$, where $E_1 = \{u_{CA}(00)\}$, $E_2 = \{u_{CA}(11)\}$, and $E_3 = \{u_{CA}(01), u_{CA}(10)\}$. The vertex set, $V = \{C, A\}$, is partitioned into a single element, V indicating that the main effects are the same for all categories. Once again we could represent all atomic classes with black lines (figure 3.6(b))



(a) Graph colouring 1 (b) Graph colouring 2

Figure 3.6: Two alternative graph colourings for a symmetry model with two binary variables. The first-order interactions set E is partitioned into $E_1 = \{u_{CA}(00)\}$, $E_2 = \{u_{CA}(11)\}$, and $E_3 = \{u_{CA}(01), u_{CA}(10)\}$ and $V = \{C, A\}$: a) atomic classes in different colour; b) atomic classes in black.

The model in figure 3.6(a) can be expressed as

$$\log m(i,j) = u + u_C(i) + u_A(j) + u_{CA}(ij),\ i,j = 0,1;$$

in which

$$u_{CA}(ij) = u_{CA}(ji),\ i,j = 0,1;$$
$$u_A(i) = u_C(i),\ i = 0,1.$$

Then, the symmetry model for $V=\{C,A\}$ is the RGLL model generated by $\{C,A\}$ with vertex colour class V and first-order interaction set or their corresponding edge set $E = (E_1, E_2, E_3)$ with $E_1 = \{u_{CA}(00)\}$, $E_2 = \{u_{CA}(11)\}$, and $E_3 = \{u_{CA}(01), u_{CA}(10)\}$, whose associated graph $G(V,E)$ is given in figure 3.6(a) or figure 3.6(b).

Notice that, as there are explicit expressions for symmetry and quasi-symmetry models, the previous example could have been given in the other direction, *i.e.* we could draw the graph for the model according to the models restrictions by assigning the same colour to equal main effects and first-order interactions, obtaining a model whose cliques coincide with the generating class, which is a RGLL model.

We now discuss which of the symmetry generalizations already given in section **2.6** are RGLL models.

i) The conditional symmetry model,

$$m_{123}(i,j,k) = m_{123}(j,i,k),\ i,j = 1,...,|\mathrm{I}|\,;\ k = 1,...,|\mathrm{K}|\,,$$

can be represented as a group of k RGLL models by considering that it can be seen as a symmetry model between two variables for each level k of a third variable. For every k, there is a RGLL model with generating class $\{\{1,2\}\}$, where the vertex and edge set are similar to the ones in the symmetry model. The vertex set $V = \{1,2\}$ is not partitioned and the edge set E is partitioned as follows. It is partitioned into $|\mathrm{I}|$ atomic edge colour classes for every $u_{12}(ii)$ interaction and into $\binom{|\mathrm{I}|}{2}$ different edge colour classes for the interactions $u_{12}(ij) = u_{12}(ji), i \neq j$. For instance, if 1 and 2 are binary, then $E = (E_1, E_2, E_3)$, where $E_1 = \{u_{12}(00)\}$, $E_2 = \{u_{12}(11)\}$, and $E_3 = \{u_{12}(01), u_{12}(10)\}$.

ii) Marginal homogeneity models and their generalizations can not be seen as RGLL models because they do not have a log-linear representation.

iii) The complete marginal symmetry model,

$$m_{123}(i,j,.) = m_{123}(j,i,.),\ i,j = 1,...,|\mathrm{I}|\,,$$

can be represented as a RGLL model if we work with the corresponding marginal table. That is, we use a RGLL model for the table formed by the marginal counts $n_{123}(i,j,.)$ in which the vertex set $V=\{1,2\}$ is not partitioned and the edge set E is partitioned as follows. It is partitioned into $|\mathrm{I}|$ atomic edge colour classes for every $u_{12}(ii)$ interaction and $\binom{|\mathrm{I}|}{2}$ different edge colour classes for the interactions $u_{12}(ij)=u_{12}(ji), i\neq j$. As before, if the variables 1 and 2 are binary, then $E=(E_1,E_2,E_3)$, where $E_1=\{u_{12}(00)\}$, $E_2=\{u_{12}(11)\}$, and $E_3=\{u_{12}(01),u_{12}(10)\}$.

iv) The marginal quasi-symmetry model,

$$\log m(i,j,k)=u+u_1(i)+u_2(j)+u_3(k)+u_{12}(ij)+u_{13}(ik)+u_{23}(jk)$$

with restrictions

$$u_{12}(ij)=u_{12}(ji),\ i,j,k=1,...,|\mathrm{I}|\,,$$

is not a RGLL model because its generating class is $\{\{1,2\},\{1,3\},\{2,3\}\}$ and the cliques set associated to the corresponding graph is the set $\{\{1,2,3\}\}$. Therefore, the cliques set is not the same as the generating class (something analogous was already discussed in example 2.9), and as a consequence, the model is not RGLL, even though it has parameter restrictions like the ones used in RGLL models. However, we can define a similar model, a marginal quasi-symmetry model together with the independence $1\perp 3|2$, which is a model

$$\log m(i,j,k)=u+u_1(i)+u_2(j)+u_3(k)+u_{12}(ij)+u_{23}(jk)$$

adding the restrictions

$$u_{12}(ij)=u_{12}(ji),\ i,j,k=1,...,|\mathrm{I}|\,.$$

From the generating class $\{\{1,2\},\{2,3\}\}$ and its corresponding graph, one can see that variable 1 is conditionally independent of variable 3 given variable 2. Additionally, for a fixed k, there is quasi-symmetry, similar to what happened in the marginal quasi-symmetry model, because the log-linear expansions for $\log m(i,j,k)$ and $\log m(j,i,k)$

$$\log m(i,j,k)=u+u_1(i)+u_2(j)+u_3(k)+u_{12}(ij)+u_{23}(jk)=u+u_3(k)+u_1(i)+f_1(j)+u_{12}(ij),$$

$$\log m(j,i,k)=u+u_1(j)+u_2(i)+u_3(k)+u_{12}(ji)+u_{23}(ik)=u+u_3(k)+f_2(i)+u_1(j)+u_{12}(ij),$$

where $f_1(j)=u_2(j)+u_{23}(jk)$ and $f_2(i)=u_2(i)+u_{23}(ik)$, only differ on the marginal effects. In other words, $m(i,j,k)$ and $m(j,i,k)$ are not equal because the marginal effects are different. The corresponding RGLL model is the one with generating class $\{\{1,2\},\{2,3\}\}$, vertex set $V=(V_1,V_2,V_3)$, where $V_1=\{1\}$, $V_2=\{2\}$ and $V_3=\{3\}$, and edge set E partitioned into $|\mathrm{I}|$ atomic edge colour classes for every $u_{12}(ii)$ interaction, $\binom{|\mathrm{I}|}{2}$ different edge colour classes for the interactions $u_{12}(ij)=u_{12}(ji), i\neq j$, and

the remaining interactions $u_{23}(jk)$ are in different atomic classes.

We can even define a similar model in which additionally for every fixed k, one has $u_{23}(ik) = u_{12}(jk),\ i, j = 1, ..., |\mathrm{I}|$. Under this models we have the same properties than before, but the log-linear expansions are

$$\log m(i,j,k) = u+u_1(i)+u_2(j)+u_3(k)+u_{12}(ij)+u_{23}(jk) = u+u_3(k)+u_{23}(jk)+u_1(i)+u_2(j)+u_{12}(ij),$$

$$\log m(j,i,k) = u+u_1(j)+u_2(i)+u_3(k)+u_{12}(ji)+u_{23}(ik) = u+u_3(k)+u_{23}(jk)+u_2(i)+u_1(j)+u_{12}(ij),$$

so that the marginal effects are given by the parameters u_1 and u_2. The model is RGLL with generating class $\{\{1,2\},\ \{2,3\}\}$, vertex set $V = (V_1, V_2, V_3)$, where $V_1 = \{1\}$, $V_2 = \{2\}$ and $V_3 = \{3\}$, and edge set E partitioned into $|\mathrm{I}|$ atomic edge colour classes for every $u_{12}(ii)$ interaction, $\binom{|\mathrm{I}|}{2}$ different edge colour classes for the interactions $u_{12}(ij) = u_{12}(ji), i \neq j$, and into the classes $\{u_{23}(1k),\ u_{23}(2k),....,\ u_{23}(|\mathrm{I}|\, k)\}$, for $k = 1, ..., |\mathrm{I}|$.

v) Complete symmetry models and quasi-symmetry generalizations, complete and preserving one-dimensional margins, are not RGLL models because, even though they are graphical, they have restrictions for the interactions of order higher than one.

In Chapter **4** we will define specific classes of RGLL models that can be seen as symmetry generalizations in terms of graph theory and statistics, similar to the work done by Højsgaard and Lauritzen (2007, 2008) in the continuous case, mainly to what they call RCOP models.

We emphasize that the fact of restricting first-order interactions only is because we wanted to define models analogous to the ones defined by Højsgaard and Lauritzen (2007, 2008) for continuous data that could be represented graphically using colourings and it is not possible to represent in such a simple way restrictions for interactions of higher order. Generalization of RGLL models restricting additional parameters besides main effects and first-order interactions is possible; however, this generalization requires a lot of additional work, starting from a new definition, new associated likelihood equations, another algorithm to solve them together with the corresponding proof that the values obtained with it converge to the maximum likelihood estimators, etc. This work may correspond to future research.

3.8 Maximum likelihood estimators

Before proceeding to derive the likelihood equations for RGLL models in sections **3.8.1**, **3.8.2**, and **3.8.3**, we define the following notation that will be used to express the likelihood equations in a simple way.

Definition 3.4. Let $\{v_k^i\}_{k=1,...,kver(i),i=1,...,T}$ be a vertex or variable labels set, each variable with J categories, partitioned into $V_1, V_2, ..., V_T$, in such a way that $V_i = \left\{v_1^i, ..., v_{kver(i)}^i\right\}$, i=1,..., T. The marginal total $n(v_k^i = l)$ and marginal expected frequency $m(v_k^i = l)$ for the k-th variable in the colour class i for the category l are defined as

$$n(v_k^i = l) = \sum_{j:j_{v_k^i}=l} n(j),\ l = 1, 2, ..., J;\ k = 1, ..., kver(i);\ i = 1, 2, ..., T. \tag{3.2}$$

$$m(v_k^i = l) = \sum_{j:j_{v_k^i}=l} m(j),\ l = 1, 2, ..., J;\ k = 1, ..., kver(i);\ i = 1, 2, ..., T. \tag{3.3}$$

Example 3.6. Consider the data given in table 3.1, and suppose that A and C are in the same colour class V_1, *i.e.* V_1=$\{v_1^1, v_2^1\}$, with v_1^1=C and v_2^1=A, then

$$n(v_1^1 = 0) = n(C = 0) = \sum_{j:j_C=0} n(j) = n(0,0) + n(0,1) = n(0,.) = 593.$$
$$n(v_1^1 = 1) = n(C = 1) = \sum_{j:j_C=1} n(j) = n(1,0) + n(1,1) = n(1,.) = 414.$$
$$n(v_2^1 = 0) = n(A = 0) = \sum_{j:j_A=0} n(j) = n(0,0) + n(1,0) = n(.,0) = 568.$$
$$n(v_2^1 = 1) = n(A = 1) = \sum_{j:j_A=1} n(j) = n(0,1) + n(1,1) = n(.,1) = 439.$$

The dot in these expressions, for example $n(1,.)$, indicates that we are summing the observed counts associated to the cells over all possible values of the entry where the dot appears. The process is similar for the expected frequencies, simply replacing $n(,)$ with $m(,)$.

Definition 3.5. Let $\{e_k^t\}_{k=1,...,ked(t),t=1,...,S}$ be an edge set, with e_k^t identified with the parameter $u_{l_k^t m_k^t}(i_k^t, j_k^t)$, the first-order interaction between variable l_k^t and m_k^t at variable values (i_k^t, j_k^t), partitioned into $E_1, E_2, ...E_S$ colour classes, with $E_t = \left\{e_1^t, ..., e_{ked(t)}^t\right\}$, t=1,..., S. The marginal total $n(l_k^t = i_k^t, m_k^t = j_k^t)$ and the marginal expected frequency $m(l_k^t = i_k^t, m_k^t = j_k^t)$ for the k-th pair of variables, with their corresponding levels, in the colour class t are defined as

$$n(l_k^t = i_k^t, m_k^t = j_k^t) = \sum_{s:(s_{l_k^t}, s_{m_k^t})=(i_k^t, j_k^t)} n(s),\ k = 1, ..., ked(t);\ t = 1, ..., S.$$

$$m(l_k^t = i_k^t, m_k^t = j_k^t) = \sum_{s:(s_{l_k^t}, s_{m_k^t})=(i_k^t, j_k^t)} m(s),\ k = 1, ..., ked(t);\ t = 1, ..., S.$$

Example 3.7. Using the data given in table 3.1, suppose that the first-order interaction set or its corresponding associated edge set E, is partitioned into $E_1 = \{u_{CA}(00), u_{CA}(11)\}$ and $E_2 = \{u_{CA}(01), u_{CA}(10)\}$, which means that $ked(1) = ked(2) = 2$; $l_k^t = C$, k, t= 1,2; $m_k^t = A$, k, t= 1,2; $i_k^t = 0$, $t = 1, 2$, $k = 1$; $i_k^t = 1$, $t = 1, 2$, $k = 2$; $j_k^t = 0$, $t = k = 1, 2$; $j_k^t = 1$, $k \neq t$, then

$$n(l_1^1 = i_1^1, m_1^1 = j_1^1) = n(0, 0) = 540.$$
$$n(l_2^1 = i_2^1, m_2^1 = j_2^1) = n(1, 1) = 386.$$
$$n(l_1^2 = i_1^2, m_1^2 = j_1^2) = n(0, 1) = 53.$$
$$n(l_2^2 = i_2^2, m_2^2 = j_2^2) = n(1, 0) = 28.$$

Using expression (2.7), including the model with the different kinds of restrictions according to the colourings, and the notation given above, we obtain the corresponding likelihood equations.

3.8.1 Vertex colouring

Consider a RGLL model with graph $G = (V, E)$, where V is partitioned into $V_1, ..., V_T$, with $V_i \neq \emptyset$, for i=1,...,T, with $T \in \{1, 2, ..., |V|\}$ and the set of first-order interactions E is partitioned into $E_1, ..., E_S$, with $S = |E|$. In this case, the logarithm of the kernel of the likelihood function, the part of the likelihood function depending on the expected frequencies m considering any sampling scheme: Poisson, multinomial, or restricted multinomial (expression (2.7)), including the restrictions takes the form

$$\sum_{i=1}^{T}\sum_{l=1}^{J}\sum_{k=1}^{kver(i)} n(v_k^i = l)u_i(l) + \sum_{a\in K, |a|\neq 1}\sum_{i_a} n_a(i_a)u_a(i_a) - \sum_{i\in \mathrm{I}} \exp(\sum_{a\in K} u_a(i_a)), \quad (3.4)$$

where K is the set of subsets of the elements in the generating class A, formed by all the cliques of the corresponding associated graph $G(V, E)$.

Deriving (3.4) with respect to each parameter, including the $u_i(l)$ parameters corresponding to V_i, i=1,2,...,T for a category l, l=1,2,...,J, and equating to zero, we obtain the following likelihood equations:

$$\begin{aligned}
&\sum_{j=1}^{kver(i)} n(v_j^i = 1) = \sum_{j=1}^{kver(i)} m(v_j^i = 1),\ i = 1, ..., T;\\
&\sum_{j=1}^{kver(i)} n(v_j^i = 2) = \sum_{j=1}^{kver(i)} m(v_j^i = 2),\ i = 1, ..., T;\\
&\qquad\qquad\vdots\\
&\sum_{j=1}^{kver(i)} n(v_j^i = J) = \sum_{j=1}^{kver(i)} m(v_j^i = J),\ i = 1, ..., T;\\
&n_a(i_a) = m_a(i_a),\ \text{for all } a \in K,\ |a| \neq 1.
\end{aligned} \tag{3.5}$$

As we have discussed in section **2.3**, there could be redundant equations for hierarchical log-linear models if we consider all elements in K for the last row in equation system (3.5). Then, to eliminate redundant equations and as a consequence to decrease the number of equations, we substitute the last row in equation system (3.5) with

$$n_a(i_a) = m_a(i_a),\ \text{for all } a \in A,\ |a| \neq 1$$

Remember that $A =$ generating class, whereas $K =$ set of all subsets of the elements in the generating class A, $|A| \leq |K|$.

All the previous discussion about likelihood equations for a vertex colouring can be summarized in the following theorem.

Theorem 3.1. *The likelihood equations for a RGLL model under a vertex colouring, which is a RGLL model with generating class A and associated graph $G = (V, E)$, where V is partitioned into $V_1, ..., V_T$, with $V_i \neq \emptyset$, where i=1,...,T, for any $T \in \{1, 2, ..., |V|\}$ and the set of first-order interactions E is partitioned into $E_1, ..., E_S$, with $S = |E|$, are as follows.*

$$\begin{aligned}
&\sum_{j=1}^{kver(i)} n(v_j^i = 1) = \sum_{j=1}^{kver(i)} m(v_j^i = 1),\ i = 1, ..., T;\\
&\sum_{j=1}^{kver(i)} n(v_j^i = 2) = \sum_{j=1}^{kver(i)} m(v_j^i = 2),\ i = 1, ..., T;\\
&\qquad\qquad\vdots
\end{aligned}$$

$$\sum_{j=1}^{kver(i)} n(v_j^i = J) = \sum_{j=1}^{kver(i)} m(v_j^i = J),\ i = 1, ..., T;$$
$$n_a(i_a) = m_a(i_a),\ \textit{for all } a \in A,\ |a| \neq 1.$$

Example 3.8. Suppose that $V = \{W,\ X,\ Y,\ Z\}$, where all variables have the same number of levels J, and consider the model given in example 3.2, which is the RGLL model with graph $G(V, E)$, with $V = \{W,\ X,\ Y,\ Z\}$ and first-order interaction set E, with generating class $\{\{X, Y\},\ \{Y, Z\},\ \{Y, W\}\}$, with colour classes $V_1 = \{X, Y\}$ and $V_2 = \{Z, W\}$ and $E = (E_1, ..., E_S)$; $S = |E|$, and that we have a contingency table whose observed values are given by $n(w, x, y, z)$, where w, x, y, and z are levels of W, X, Y, and Z, respectively, then equations (3.2) become

$$n(v_1^1 = k) = n(X = k) = \sum_{j:j_X=k} n(j) = \sum_{w,y,z=1}^{J} n(w, k, y, z) = n(., k, ., .),\ k = 1, 2, ..., J;$$

$$n(v_2^1 = k) = n(Y = k) = \sum_{j:j_Y=k} n(j) = \sum_{w,x,z=1}^{J} n(w, x, k, z) = n(., ., k, .),\ k = 1, 2, ..., J;$$

$$n(v_1^2 = k) = n(Z = k) = \sum_{j:j_Z=k} n(j) = \sum_{w,x,y=1}^{J} n(w, x, y, k) = n(., ., ., k),\ k = 1, 2, ..., J;$$

$$n(v_2^2 = k) = n(W = k) = \sum_{j:j_W=k} n(j) = \sum_{x,y,z=1}^{J} n(k, x, y, z) = n(k, ., ., .),\ k = 1, 2, ..., J.$$

The likelihood equations corresponding to the main effects are

$$n(., k, ., .) + n(., ., k, .) = m(., k, ., .) + m(., ., k, .),\ k = 1, 2, ..., J;$$
$$n(., ., ., k) + n(k, ., ., .) = m(., ., ., k) + m(k, ., ., .),\ k = 1, 2, ..., J.$$

We add to these the equations

$$n_a(i_a) = m_a(i_a),\ a = \{X, Y\}, \{Y, Z\}, \{Y, W\}.$$

If $i = (w, x, y, z)$, these equations are

$$n(., x, y, .) = m(., x, y, .),\ x, y = 1, ..., J;$$
$$n(., ., y, z) = m(., ., y, z),\ y, z = 1, ..., J;$$
$$n(w, ., y, .) = m(w, ., y, .),\ w, y = 1, ..., J.$$

Therefore, the simultaneous equation system is

$$\begin{aligned} n(.,k,.,.)+n(.,.,k,.) &= m(.,k,.,.)+m(.,.,k,.),\ k=1,2,...,J;\\ n(.,.,.,k)+n(k,.,.,.) &= m(.,.,.,k)+m(k,.,.,.),\ k=1,2,...,J;\\ n(.,x,y,.) &= m(.,x,y,.),\ x,y=1,...,J;\\ n(.,.,y,z) &= m(.,.,y,z),\ y,z=1,...,J;\\ n(w,.,y,.) &= m(w,.,y,.),\ w,y=1,...,J. \end{aligned}$$

The maximum likelihood estimators correspond to a vector of estimated expected frequencies $\widehat{m}$ satisfying the previous system of equations.

3.8.2 Edge colouring

Consider a RGLL model with graph $G=(V,E)$, where the first-order interaction set E is partitioned into $E_1,...,E_S$, $E_i\neq\emptyset$, for i=1,...,S, with $S\in\{1,2,...,|E|\}$ and the set of variables V is partitioned into $V_1,...,V_T$, with $T=|V|$. In this case, the logarithm of the kernel of the likelihood function considering any sampling scheme: Poisson, multinomial, or restricted multinomial, expression (2.7), including the restrictions takes the form

$$\sum_{t=1}^{S}\sum_{k=1}^{ked(t)} n(l_k^t=i_k^t,m_k^t=j_k^t)u_{E_t}+\sum_{a\in K,|a|\neq 2}\sum_{i_a} n_a(i_a)u_a(i_a)-\sum_{i\in \mathrm{I}}\exp(\sum_{a\in K}u_a(i_a)). \quad (3.6)$$

Deriving (3.6) with respect to each parameter and equating to zero, we get the following equation system

$$\begin{aligned} \sum_{k=1}^{ked(t)} n(l_k^t=i_k^t,m_k^t=j_k^t) &= \sum_{k=1}^{ked(t)} m(l_k^t=i_k^t,m_k^t=j_k^t),\ t=1,2,....,S;\\ n_a(i_a) &= m_a(i_a), \text{for all } a\in K,|a|\neq 2. \end{aligned} \quad (3.7)$$

Similarly to the vertex colouring, we can eliminate some redundant equations corresponding to the elements in K no contained in the generating class and formed by a number of variables different to two. To accomplish this reduction, we eliminate the last part of the equation system and instead we use the equations for the main effects and the equations corresponding to the generating class, *i.e.*, we substitute the last part in equation system (3.7) with

$$\begin{aligned} n_a(i_a) &= m_a(i_a),\ \text{for all } a\in A,\ |a|\neq 2,\\ n_a(i_a) &= m_a(i_a),\ |a|=1. \end{aligned} \quad (3.8)$$

All the previous discussion about likelihood equations for an edge colouring can be summarized in the following theorem.

Theorem 3.2. *The likelihood equations for a RGLL model under an edge colouring, which is a RGLL model with generating class A and associated graph $G = (V, E)$, where the first-order interaction set E is partitioned into $E_1, ..., E_S$, with $E_i \neq \emptyset$, where i=1,...,S, for any $S \in \{1, 2, ..., |E|\}$ and the set of variables V is partitioned into $V_1, ..., V_T$, with $T = |V|$, are as follows.*

$$\sum_{k=1}^{ked(t)} n(l_k^t = i_k^t, m_k^t = j_k^t) = \sum_{k=1}^{ked(t)} m(l_k^t = i_k^t, m_k^t = j_k^t),\ t = 1, 2, ..., S;$$
$$n_a(i_a) = m_a(i_a),\ \textit{for all } a \in A,\ |a| \neq 2;$$
$$n_a(i_a) = m_a(i_a),\ |a| = 1.$$

Example 3.9. Suppose that we have a set V= $\{W,\ X,\ Y,\ Z\}$ of binary variables included in the same model given in example 3.3. This is a RGLL model with graph $G(V, E)$, generating class $\{\{X, Y\}, \{W, Y\}, \{W, Z\}, \{X, Z\}\}$, and edge set E= (E_1, E_2, E_3, E_4); where E_1 =$\{u_{XY}(00),\ u_{XY}(11),\ u_{WZ}(00),\ u_{WZ}(11)\}$; E_2 =$\{u_{XY}(01),\ u_{XY}(10),\ u_{WZ}(01),\ u_{WZ}(10)\}$; E_3 = $\{u_{WY}(00),\ u_{WY}(11),\ u_{XZ}(00),\ u_{XZ}(11)\}$; E_4 =$\{u_{WY}(01),\ u_{WY}(10),\ u_{XZ}(01),\ u_{XZ}(10)\}$, with $V = (V_1, ..., V_T)$, $T = |V| = 4$, and suppose that we have a contingency table whose observed values are given by $n(w, x, y, z)$, where w, x, y, and z are levels of W, X, Y, and Z, respectively; w, x, y, z= 0, 1. We have the following equalities for the E_1 elements.

$$n(l_1^1 = i_1^1, m_1^1 = j_1^1) = n(X = 0, Y = 0) = \sum_{s:(s_X, s_Y)=(0,0)} n(s) = \sum_{w,z} n(w, 0, 0, z) = n(., 0, 0, .),$$

$$n(l_2^1 = i_2^1, m_2^1 = j_2^1) = n(X = 1, Y = 1) = \sum_{s:(s_X, s_Y)=(1,1)} n(s) = \sum_{w,z} n(w, 1, 1, z) = n(., 1, 1, .),$$

$$n(l_3^1 = i_3^1, m_3^1 = j_3^1) = n(W = 0, Z = 0) = \sum_{s:(s_W, s_Z)=(0,0)} n(s) = \sum_{x,y} n(0, x, y, 0) = n(0, ., ., 0),$$

$$n(l_4^1 = i_4^1, m_4^1 = j_4^1) = n(W = 1, Z = 1) = \sum_{s:(s_W, s_Z)=(1,1)} n(s) = \sum_{x,y} n(1, x, y, 1) = n(1, ., ., 1).$$

For E_2 we have

$$n(l_1^2 = i_1^2, m_1^2 = j_1^2) = n(., 0, 1, .),$$
$$n(l_2^2 = i_2^2, m_2^2 = j_2^2) = n(., 1, 0, .),$$
$$n(l_3^2 = i_3^2, m_3^2 = j_3^2) = n(0, ., ., 1),$$

$$n(l_4^2 = i_4^2, m_4^2 = j_4^2) = n(1, ., ., 0);$$

for E_3

$$\begin{aligned} n(l_1^3 = i_1^3, m_1^3 = j_1^3) &= n(0, ., 0, .), \\ n(l_2^3 = i_2^3, m_2^3 = j_2^3) &= n(1, ., 1, .), \\ n(l_3^3 = i_3^3, m_3^3 = j_3^3) &= n(., 0, ., 0), \\ n(l_4^3 = i_4^3, m_4^3 = j_4^3) &= n(., 1, ., 1); \end{aligned}$$

and finally for E_4

$$\begin{aligned} n(l_1^4 = i_1^4, m_1^4 = j_1^4) &= n(0, ., 1, .), \\ n(l_2^4 = i_2^4, m_2^4 = j_2^4) &= n(1, ., 0, .), \\ n(l_3^4 = i_3^4, m_3^4 = j_3^4) &= n(., 0, ., 1), \\ n(l_4^4 = i_4^4, m_4^4 = j_4^4) &= n(., 1, ., 0). \end{aligned}$$

We have something similar for the expected frequencies m. The resulting likelihood equations for the first-order interactions, first part of equations (3.7), are

$$\begin{aligned} n(., 0, 0, .) + n(., 1, 1, .) + n(0, ., ., 0) + n(1, ., ., 1) &= \\ m(., 0, 0, .) + m(., 1, 1, .) + m(0, ., ., 0) + m(1, ., ., 1), & \\ n(., 0, 1, .) + n(., 1, 0, .) + n(0, ., ., 1) + n(1, ., ., 0) &= \\ m(., 0, 1, .) + m(., 1, 0, .) + m(0, ., ., 1) + m(1, ., ., 0), & \\ n(0, ., 0, .) + n(1, ., 1, .) + n(., 0, ., 0) + n(., 1, ., 1) &= \\ m(0, ., 0, .) + m(1, ., 1, .) + m(., 0, ., 0) + m(., 1, ., 1), & \\ n(0, ., 1, .) + n(1, ., 0, .) + n(., 0, ., 1) + n(., 1, ., 0) &= \\ m(0, ., 1, .) + m(1, ., 0, .) + m(., 0, ., 1) + m(., 1, ., 0). & \end{aligned} \tag{3.9}$$

We should add to this equations system the equations (3.8); however, in this model the generating class A is formed by $\{X, Y\}$, $\{W, Y\}$, $\{W, Z\}$, $\{X, Z\}$, for which $|a| = 2$, but (3.8) does not consider the corresponding likelihood equations, in fact the likelihood equations for first-order interactions are already contained in (3.9), so that the only missing equations are the corresponding to main effects

$$n_a(i_a) = m_a(i_a), \ a = \{X\}, \{Y\}, \{Z\}, \{W\}.$$

Explicitly, they are

$$\begin{aligned} n(i,.,.,.) &= m(i,.,.,.),\ i=0,1; \\ n(.,i,.,.) &= m(.,i,.,.),\ i=0,1; \\ n(.,.,i,.) &= m(.,.,i,.),\ i=0,1; \\ n(.,.,.,i) &= m(.,.,.,i),\ i=0,1. \end{aligned} \tag{3.10}$$

Then, the system of equations to be solved is the one conformed by equations (3.9) and (3.10).

Notice that the main effects are not being forced to satisfy any equalities between them, so that we could think in a partition of the vertex set into four subsets, each formed by one variable, W, X, Y, and Z, which is a vertex colouring. In fact, the equations we obtain here for the main effects are the same equations we would get using equations (3.5).

3.8.3 Vertex and edge colouring

Consider a RGLL model with graph $G=(V,E)$, where V is partitioned into $V_1,...,V_T$, with $V_i \neq \emptyset$, for i=1,...,T, with $T \in \{1,2,...,|V|\}$ and the set of first-order interactions E is partitioned into $E_1,...,E_S$, $E_i \neq \emptyset$, for i=1,...,S, with $S \in \{1,2,...,|E|\}$. In this case, the logarithm of the kernel of the likelihood function considering any of the three sampling schemes: Poisson, multinomial, or restricted multinomial, expression (2.7), including both types of parameter restrictions takes the form

$$\begin{aligned} &\sum_{i=1}^{T}\sum_{l=1}^{J}\sum_{k=1}^{kver(i)} n(v_k^i=l)u_i(l) + \sum_{t=1}^{S}\sum_{k=1}^{ked(t)} n(l_k^t=i_k^t, m_k^t=j_k^t)u_{E_t} + \\ &\quad \sum_{a\in K,|a|\neq 1,2}\sum_{i_a} n_a(i_a)u_a(i_a) - \sum_{i\in \mathrm{I}} \exp\Big(\sum_{a\in K} u_a(i_a)\Big). \end{aligned} \tag{3.11}$$

Deriving this expression with respect to each parameter and equating to zero we obtain the equation systems (3.5) and (3.7) simultaneously. The equation system is

$$\begin{aligned}
\sum_{j=1}^{kver(i)} n(v_j^i = 1) &= \sum_{j=1}^{kver(i)} m(v_j^i = 1), i = 1, ..., T; \\
\sum_{j=1}^{kver(i)} n(v_j^i = 2) &= \sum_{j=1}^{kver(i)} m(v_j^i = 2), i = 1, ..., T; \\
&\vdots \\
\sum_{j=1}^{kver(i)} n(v_j^i = J) &= \sum_{j=1}^{kver(i)} m(v_j^i = J), i = 1, ..., T; \\
\sum_{k=1}^{ked(t)} n(l_k^t = i_k^t, m_k^t = j_k^t) &= \sum_{k=1}^{ked(t)} m(l_k^t = i_k^t, m_k^t = j_k^t),\ t = 1, 2, ..., S; \\
n_a(i_a) &= m_a(i_a),\ \text{for all } a \in K,\ |a| \neq 1, 2.
\end{aligned} \tag{3.12}$$

As before, the last set of equations in the equation system could be replaced by

$$n_a(i_a) = m_a(i_a),\ \text{for all } a \in A,\ |a| \neq 1, 2.$$

All the previous discussion about likelihood equations for a vertex and edge colouring can be summarized in the following theorem.

Theorem 3.3. *The likelihood equations for a RGLL model, which is a model including both vertex and edge colourings where the generating class is A and the associated graph is $G = (V, E)$, in which V is partitioned into $V_1, ..., V_T$, with $V_i \neq \emptyset$, where i=1,...,T, for any $T \in \{1, 2, ..., |V|\}$ and the set of first-order interactions E is partitioned into $E_1, ..., E_S$, with $E_i \neq \emptyset$, where i=1,...,S, for any $S \in \{1, 2, ..., |E|\}$, are as follows.*

$$\begin{aligned}
\sum_{j=1}^{kver(i)} n(v_j^i = 1) &= \sum_{j=1}^{kver(i)} m(v_j^i = 1),\ i = 1, ..., T; \\
\sum_{j=1}^{kver(i)} n(v_j^i = 2) &= \sum_{j=1}^{kver(i)} m(v_j^i = 2),\ i = 1, ..., T; \\
&\vdots
\end{aligned}$$

$$\sum_{j=1}^{kver(i)} n(v_j^i = J) = \sum_{j=1}^{kver(i)} m(v_j^i = J), i = 1, ..., T;$$

$$\sum_{k=1}^{ked(t)} n(l_k^t = i_k^t, m_k^t = j_k^t) = \sum_{k=1}^{ked(t)} m(l_k^t = i_k^t, m_k^t = j_k^t),\ t = 1, 2, ..., S;$$

$$n_a(i_a) = m_a(i_a),\ \textit{for all } a \in A,\ |a| \neq 1, 2.$$

Example 3.10. To illustrate the derivation of the likelihood equations, consider the i) quasi-symmetry and ii) symmetry models given in example 3.5. As we saw before, these are RGLL models with generating class $\{\{C, A\}\}$ and associated graphs $G(V, E)$, $V = \{C, A\}$ and $E = \{u_{CA}(01),\ u_{CA}(10),\ u_{CA}(00),\ u_{CA}(11)\}$.

i) The quasi-symmetry model additionally has three edge colour classes or partitions of the first-order interactions set: $E_1 = \{u_{CA}(00)\}$, $E_2 = \{u_{CA}(01),\ u_{CA}(10)\}$, and $E_3 = \{u_{CA}(11)\}$. There is one equation associated to each edge class as follows,

$$E_1 :\ n(0,0) = m(0,0),$$
$$E_2 :\ n(0,1) + n(1,0) = m(0,1) + m(1,0),$$
$$E_3 :\ n(1,1) = m(1,1),$$

which can be rewritten as

$$m(i,j) + m(j,i) = n(i,j) + n(j,i),\ \text{for all } i \leq j,\ i = j = 0, 1. \tag{3.13}$$

We also have two vertex colour classes $V = (V_1, V_2)$, with $V_1 = \{C\} = \{v_1^1\}$ and $V_2 = \{A\} = \{v_1^2\}$. Then we have

$$n(v_1^1 = 0) = n(0, .),$$
$$n(v_1^1 = 1) = n(1, .),$$
$$n(v_1^2 = 0) = n(., 0),$$
$$n(v_1^2 = 1) = n(., 1).$$

The equations corresponding to class V_1 are

$$n(1, .) = m(1, .),$$
$$n(0, .) = m(0, .);$$

and to V_2

$$n(.,1) = m(.,1),$$
$$n(.,0) = m(.,0).$$

These equations can be rewritten as

$$m(i,.) = n(i,.),\ i = 0,1;\ m(.,j) = n(.,j),\ j = 0,1. \tag{3.14}$$

Equations systems (3.13) and (3.14) have to be solved simultaneously.

ii) In the symmetry model, there are also three edge colour classes or partitions of the first-order interactions set: $E_1 = \{u_{CA}(00)\}$, $E_2 = \{u_{CA}(01),\ u_{CA}(10)\}$, and $E_3 = \{u_{CA}(11)\}$. There is one equation associated to each edge class as follows,

$$E_1 :\ n(0,0) = m(0,0),$$
$$E_2 :\ n(0,1) + n(1,0) = m(0,1) + m(1,0),$$
$$E_3 :\ n(1,1) = m(1,1),$$

which can be rewritten as

$$m(i,j) + m(j,i) = n(i,j) + n(j,i),\ \text{for all } i \leq j,\ i = j = 0,1. \tag{3.15}$$

Additionally, both variables are in the same colour class, $V = \{C, A\}$, the likelihood equations corresponding to this class are

$$n(0,.) + n(.,0) = m(0,.) + m(.,0), \tag{3.16}$$
$$n(1,.) + n(.,1) = m(1,.) + m(.,1). \tag{3.17}$$

Then, (3.15), (3.16), and (3.17) have to be solved simultaneously; however, we notice that equations in (3.15), imply equations (3.16) and (3.17), so that it is only necessary to solve (3.15).

3.9 RGLL models expressed as GLM

Log-linear models can be expressed as generalized linear models as shown in section **2.4**. There is no much literature discussing design matrices for log-linear models in which parameters are equated in some way, there are some examples and brief discussion in Rindskopf (1984) and Vermunt (1996, 2005). All RGLL models can be represented as a generalized linear model (GLM) with design matrix $\boldsymbol{X}$, usually $\boldsymbol{X}$ is a non-invertible matrix, but sometimes it is possible to give an adequate parametrization of the model so that its associated design matrix is invertible. However, the design matrix $\boldsymbol{X}$ can always be used to obtain the degrees of freedom associated to the asymptotic distribution of the deviance, the statistic used to evaluate the goodness of fit of a model.

3.9.1 Vertex colouring

In order to get the design matrix $\boldsymbol{X}$ including the vertex colouring, we need the following notation.

$\boldsymbol{\beta}$	$\boldsymbol{\beta}$=$(\beta_1, ..., \beta_p)$, vector of parameters, it includes all parameters obtained due to vertex colouring.
$\boldsymbol{X}$	design matrix for the RGLL model.
$\boldsymbol{X}^{v_l^i(j)}$	matrix in which for every row corresponding to the parameter $u_i(j)$, it takes value one when the variable v_l^i takes the value j and zero otherwise; $l = 1, 2, ..., kver(i)$; $i = 1, ..., T$, and j a fixed level, j = 1,2,...,J. The matrix has as many rows as cells in the contingency table and it has p columns; *i.e.* its dimension is $\|\mathrm{I}\| \times p$. $\boldsymbol{X}^{v_l^i(j)}$ indicates the variable v_l^i effect over the $u_i(j)$ parameter for a specific colouring i and a category j.

Let

$$\boldsymbol{X}^{V_i(j)} = \sum_{l=1}^{kver(i)} \boldsymbol{X}^{v_l^i(j)},$$

for $i = 1, 2, ...T$ and j = 1,2,...,J. The matrix $\boldsymbol{X}^{V_i(j)}$ indicates the effect of the colouring V_i for a category j.

We denote $x^{V_i(j)}$ the column vector in $\boldsymbol{X}^{V_i(j)}$ associated to $u_i(j)$, $i = 1, 2, ...T$, j = 1,2,...,J.

Then, we have that

$$\sum_{i=1}^{T} \sum_{j} \boldsymbol{X}^{V_i(j)}$$

is the matrix for the main effects considering the vertex colouring.

Additionally, taking into account

$\boldsymbol{X}^{u,int.}$	matrix including the values corresponding to the constant term, the interactions of any order, and assigning zero to the remaining terms, these remaining terms are the vertex colouring parameters.

We observe that

$$\boldsymbol{X} = \boldsymbol{X}^{u,int.} + \sum_{i=1}^{T} \sum_{j} \boldsymbol{X}^{V_i(j)} = \boldsymbol{X}^{u,int.} + \sum_{i=1}^{T} \sum_{j} \sum_{l=1}^{kver(i)} \boldsymbol{X}^{v_l^i(j)}. \tag{3.18}$$

Example 3.11. To illustrate how we obtain a design matrix in a RGLL model restricted only in its main effects or vertex colouring, consider the RGLL model with associated graph $G(V, E)$, with vertex set V= $\{X, Y, Z, W\}$, where all variables are binary, and generating class $\{\{X\}, \{Z\}, \{Y, W\}\}$. Suppose also that V is partitioned into V_1= $\{X, Y\}$= $\{v_1^1, v_2^1\}$ and V_2= $\{Z, W\}$= $\{v_1^2, v_2^2\}$, so that $v_1^1 = X$, $v_2^1 = Y$, $v_1^2 = Z$, and $v_2^2 = W$, and that the first-order interactions set $E = (E_1, ..., E_S)$, with $S = |E|$, figure 3.7.

To identify which parameters correspond to the generating class, we give the model associated to that generating class without considering restrictions.

$$\log m(i, j, k, l) = u + u_W(i) + u_X(j) + u_Y(k) + u_Z(l) + u_{WY}(ik), \;\; i, j, k, l = 0, 1.$$

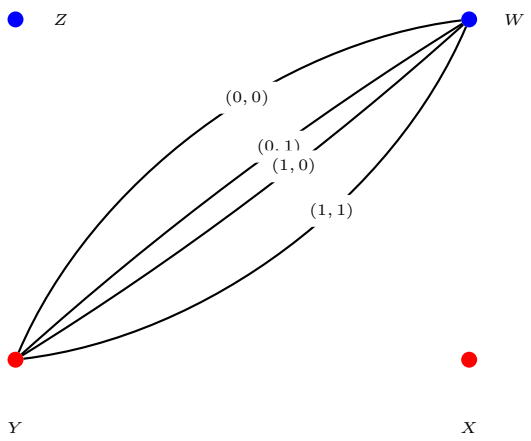


Figure 3.7: Vertex colouring for the RGLL model with generating class $\{\{X\}, \{Z\}, \{Y, W\}\}$, $V = (V_1, V_2)$, with V_1= $\{X, Y\}$ and V_2= $\{Z, W\}$ and $E = (E_1, ..., E_S)$, with $S = |E|$=4.

Denoting (w, x, y, z) as a cell, where w, x, y, and z are levels for W, X, Y, and Z, respectively, the cells can be ordered as folllows.

$$((0,0,0,0),(0,0,0,1),(0,0,1,0),(0,0,1,1),(0,1,0,0),(0,1,0,1),(0,1,1,0),(0,1,1,1),$$

$$(1,0,0,0),(1,0,0,1),(1,0,1,0),(1,0,1,1),(1,1,0,0),(1,1,0,1),(1,1,1,0),(1,1,1,1))'$$

The vector of parameters is

$$\boldsymbol{\beta}' = (u, u_1(0), u_1(1), u_2(0), u_2(1), u_{WY}(00), u_{WY}(01), u_{WY}(10), u_{WY}(11)).$$

Then, we obtain the following matrices for the vertex colour class $V_1 = \{X, Y\} = \{v_1^1, v_2^1\}$:

$$\boldsymbol{X}^{v_1^1(0)} = \begin{array}{c} \\ 0000 \\ 0001 \\ 0010 \\ 0011 \\ 0100 \\ 0101 \\ 0110 \\ 0111 \\ 1000 \\ 1001 \\ 1010 \\ 1011 \\ 1100 \\ 1101 \\ 1110 \\ 1111 \end{array} \begin{array}{c} \begin{array}{ccccccccc} u & u_1(0) & u_1(1) & u_2(0) & u_2(1) & u_{WY}(00) & u_{WY}(01) & u_{WY}(10) & u_{WY}(11) \end{array} \\ \left[\begin{array}{ccccccccc} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \end{array}.$$

The first column corresponds to u, the second one to $u_1(0)$, etc., using the same order given in $\boldsymbol{\beta}'$. The first row corresponds to cell $(0, 0, 0, 0)$, the second one to cell $(0, 0, 0, 1)$, etc., using the order given for the cells vector. In all the following matrices the columns and rows represent exactly the same parameters and same cells, respectively, as in $\boldsymbol{X}^{v_1^1(0)}$ shown above. Then, we have

$$\boldsymbol{X}^{v_1^1(1)} = \begin{array}{c} \\ 0000 \\ 0001 \\ 0010 \\ 0011 \\ 0100 \\ 0101 \\ 0110 \\ 0111 \\ 1000 \\ 1001 \\ 1010 \\ 1011 \\ 1100 \\ 1101 \\ 1110 \\ 1111 \end{array} \begin{array}{c} \begin{array}{ccccccccc} u & u_1(0) & u_1(1) & u_2(0) & u_2(1) & u_{WY}(00) & u_{WY}(01) & u_{WY}(10) & u_{WY}(11) \end{array} \\ \left[\begin{array}{ccccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \end{array},$$

$$
\boldsymbol{X}^{v_2^1(0)} = \begin{array}{c} \\ 0000 \\ 0001 \\ 0010 \\ 0011 \\ 0100 \\ 0101 \\ 0110 \\ 0111 \\ 1000 \\ 1001 \\ 1010 \\ 1011 \\ 1100 \\ 1101 \\ 1110 \\ 1111 \end{array}
\left[\begin{array}{ccccccccc}
u & u_1(0) & u_1(1) & u_2(0) & u_2(1) & u_{WY}(00) & u_{WY}(01) & u_{WY}(10) & u_{WY}(11) \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right],
$$

$$
\boldsymbol{X}^{v_2^1(1)} = \begin{array}{c} \\ 0000 \\ 0001 \\ 0010 \\ 0011 \\ 0100 \\ 0101 \\ 0110 \\ 0111 \\ 1000 \\ 1001 \\ 1010 \\ 1011 \\ 1100 \\ 1101 \\ 1110 \\ 1111 \end{array}
\left[\begin{array}{ccccccccc}
u & u_1(0) & u_1(1) & u_2(0) & u_2(1) & u_{WY}(00) & u_{WY}(01) & u_{WY}(10) & u_{WY}(11) \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right].
$$

Then

$$
\boldsymbol{X}^{V_1(0)} = \boldsymbol{X}^{v_1^1(0)} + \boldsymbol{X}^{v_2^1(0)} = \begin{array}{c} \\ 0000 \\ 0001 \\ 0010 \\ 0011 \\ 0100 \\ 0101 \\ 0110 \\ 0111 \\ 1000 \\ 1001 \\ 1010 \\ 1011 \\ 1100 \\ 1101 \\ 1110 \\ 1111 \end{array}
\left[\begin{array}{ccccccccc}
u & u_1(0) & u_1(1) & u_2(0) & u_2(1) & u_{WY}(00) & u_{WY}(01) & u_{WY}(10) & u_{WY}(11) \\
0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right],
$$

$$
\boldsymbol{X}^{V_1(1)} = \boldsymbol{X}^{v_1^1(1)} + \boldsymbol{X}^{v_2^1(1)} = \begin{array}{c} \\ 0000 \\ 0001 \\ 0010 \\ 0011 \\ 0100 \\ 0101 \\ 0110 \\ 0111 \\ 1000 \\ 1001 \\ 1010 \\ 1011 \\ 1100 \\ 1101 \\ 1110 \\ 1111 \end{array}
\left[\begin{array}{ccccccccc}
u & u_1(0) & u_1(1) & u_2(0) & u_2(1) & u_{WY}(00) & u_{WY}(01) & u_{WY}(10) & u_{WY}(11) \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right].
$$

The same is done for $V_2 = \{Z, W\} = \{v_1^2, v_2^2\}$

$$
\boldsymbol{X}^{V_2(0)} = \begin{array}{c} \\ 0000 \\ 0001 \\ 0010 \\ 0011 \\ 0100 \\ 0101 \\ 0110 \\ 0111 \\ 1000 \\ 1001 \\ 1010 \\ 1011 \\ 1100 \\ 1101 \\ 1110 \\ 1111 \end{array}
\left[\begin{array}{ccccccccc}
u & u_1(0) & u_1(1) & u_2(0) & u_2(1) & u_{WY}(00) & u_{WY}(01) & u_{WY}(10) & u_{WY}(11) \\
0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right],
$$

$$
\boldsymbol{X}^{V_2(1)} = \begin{array}{c} \\ 0000 \\ 0001 \\ 0010 \\ 0011 \\ 0100 \\ 0101 \\ 0110 \\ 0111 \\ 1000 \\ 1001 \\ 1010 \\ 1011 \\ 1100 \\ 1101 \\ 1110 \\ 1111 \end{array}
\left[\begin{array}{ccccccccc}
u & u_1(0) & u_1(1) & u_2(0) & u_2(1) & u_{WY}(00) & u_{WY}(01) & u_{WY}(10) & u_{WY}(11) \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0
\end{array}\right].
$$

Finally

$$\boldsymbol{X}^{u,int} = \begin{array}{c} \\ 0000 \\ 0001 \\ 0010 \\ 0011 \\ 0100 \\ 0101 \\ 0110 \\ 0111 \\ 1000 \\ 1001 \\ 1010 \\ 1011 \\ 1100 \\ 1101 \\ 1110 \\ 1111 \end{array} \begin{bmatrix} u & u_1(0) & u_1(1) & u_2(0) & u_2(1) & u_{WY}(00) & u_{WY}(01) & u_{WY}(10) & u_{WY}(11) \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Then, $\boldsymbol{X}$, the matrix including the colouring is

$$\boldsymbol{X} = \boldsymbol{X}^{u,int.} + \sum_{i=1}^{2} \sum_{j=0}^{1} \boldsymbol{X}^{V_i(j)} =$$

$$\begin{array}{c} \\ 0000 \\ 0001 \\ 0010 \\ 0011 \\ 0100 \\ 0101 \\ 0110 \\ 0111 \\ 1000 \\ 1001 \\ 1010 \\ 1011 \\ 1100 \\ 1101 \\ 1110 \\ 1111 \end{array} \begin{bmatrix} u & u_1(0) & u_1(1) & u_2(0) & u_2(1) & u_{WY}(00) & u_{WY}(01) & u_{WY}(10) & u_{WY}(11) \\ 1 & 2 & 0 & 2 & 0 & 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 2 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 2 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 2 & 2 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 2 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 2 & 0 & 0 & 2 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 2 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 2 & 0 & 0 & 1 & 0 \\ 1 & 0 & 2 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 2 & 0 & 2 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The logarithm of the kernel of the likelihood function, expression (2.7), for a model expressed as a GLM under the vertex colouring, which is obtained by replacing equation (2.12) in expression (2.7) considering that $\boldsymbol{X}$ is a matrix including the colouring, is

$$\log(kernel(L(m))) = \sum_{i \in \mathrm{I}} n(i) \sum_{l=1}^{T} \sum_{k} x_i^{V_l(k)} u_l(k) + \sum_{i \in \mathrm{I}} n(i) \sum_{j} x_{ij} \beta_j$$

$$- \sum_{i \in \mathrm{I}} \exp\left(\sum_{l=1}^{T} \sum_{k} x_i^{V_l(k)} u_l(k) + \sum_{j} x_{ij} \beta_j\right), \tag{3.19}$$

where x_{ij} are entries of $X^{u,int.}$.

Based on expression (3.19), we obtain the corresponding likelihood equations analogously of what we did to get the likelihood equations for a GLM model in section **2.4** where we equated in (2.13) the canonical statistics to their expectation. These equations are

$$\sum_{i\in \mathrm{I}} n(i)x_i^{V_l(k)} = \sum_{i\in \mathrm{I}} m(i)x_i^{V_l(k)},\ l = 1, 2, ..., T; k = 1, ..., J. \tag{3.20}$$

$$(\boldsymbol{X}^{u,int.})'n = (\boldsymbol{X}^{u,int.})'m. \tag{3.21}$$

These equation systems, (3.20) and (3.21), can be expressed also as

$$\boldsymbol{X}'n = \boldsymbol{X}'m.$$

These results can be summarized in the following theorem.

Theorem 3.4. *The likelihood equations for a RGLL model expressed as a GLM under a vertex colouring are*

$$\boldsymbol{X}'n = \boldsymbol{X}'m,$$

where $\boldsymbol{X}$ *is the matrix* $\boldsymbol{X} = \boldsymbol{X}^{u,int.} + \sum_{i=1}^{T}\sum_j \boldsymbol{X}^{V_i(j)} = \boldsymbol{X}^{u,int.} + \sum_{i=1}^{T}\sum_j \sum_{l=1}^{kver(i)} \boldsymbol{X}^{v_l^i(j)}$, *which includes the vertex colouring.*

3.9.2 Edge colouring

Similar to the case of vertex colouring, we have to add some notation.

$\boldsymbol{\beta}$	$\boldsymbol{\beta}$=$(\beta_1, ..., \beta_p)$, vector of parameters. This includes all parameters obtained due to edge colouring.
$\boldsymbol{X}$	design matrix for the RGLL model.
$\boldsymbol{X}^{u_{l_r^t m_r^t}(i_r^t, j_r^t)}$	$r = 1, 2, ..., ked(t)$, $t = 1, 2, ...S$. Matrix in which for every row of the column corresponding to u_{E_t}, an entry takes value one when the first-order interaction $u_{l_r^t m_r^t}(i_r^t, j_r^t)$ is present and zero otherwise. The presence of an interaction term $u_{l_r^t m_r^t}(i_r^t, j_r^t)$ means that l_r^t takes the value i_r^t and m_r^t takes the value j_r^t.

We define the following matrix

$\boldsymbol{X}^{E_t} = \sum_{r=1}^{ked(t)} \boldsymbol{X}^{u_{l_r^t m_r^t}(i_r^t, j_r^t)},\ t = 1, 2, ..., S$, matrix corresponding to the u_{E_t} parameters.

Additionally, we denote

x^{E_t}	column vector of $\boldsymbol{X}^{E_t}$ containing the coefficients associated to u_{E_t}.
$\boldsymbol{X}^{u,main\ ef.,int.}$	matrix including the coefficients corresponding to the constant, main effects, and interactions of two or higher order. Additionally, this matrix includes zeros in the columns associated to the edge colouring parameters.

Using this notation, we have that

$$\boldsymbol{X} = \boldsymbol{X}^{u,main\ ef.,int.} + \sum_{t=1}^{S} \boldsymbol{X}^{E_t} = \boldsymbol{X}^{u,main\ ef.,int.} + \sum_{t=1}^{S} \sum_{r=1}^{ked(t)} \boldsymbol{X}^{u_{l_r^t m_r^t}(i_r^t, j_r^t)}. \tag{3.22}$$

Example 3.12. Suppose that we have the same model given in example 3.3. Denoting (w, x, y, z) as a cell, where w, x, y, and z are levels for W, X, Y, and Z, respectively, the cells can be ordered in the following way.

$$((0,0,0,0),(0,0,0,1),(0,0,1,0),(0,0,1,1),(0,1,0,0),(0,1,0,1),(0,1,1,0),(0,1,1,1),$$
$$(1,0,0,0),(1,0,0,1),(1,0,1,0),(1,0,1,1),(1,1,0,0),(1,1,0,1),(1,1,1,0),(1,1,1,1))'$$

The vector of parameters is

$$\boldsymbol{\beta}' = (u, u_W(0), u_W(1), u_X(0), u_X(1), u_Y(0), u_Y(1), u_Z(0), u_Z(1), u_{E_1}, u_{E_2}, u_{E_3}, u_{E_4}).$$

For E_1, we obtain the matrices shown below. In all matrices of this example the first column corresponds to u, the second one to $u_W(0)$, etc., using the same order given for $\boldsymbol{\beta}'$. Additionally, the first row corresponds to cell $(0,0,0,0)$, the second one to $(0,0,0,1)$, etc., using the same order given for the cells vector.

$$\boldsymbol{X}^{u_{l_1^1 m_1^1}(i_1^1, j_1^1)} = \begin{array}{c} \\ 0000 \\ 0001 \\ 0010 \\ 0011 \\ 0100 \\ 0101 \\ 0110 \\ 0111 \\ 1000 \\ 1001 \\ 1010 \\ 1011 \\ 1100 \\ 1101 \\ 1110 \\ 1111 \end{array} \left[\begin{array}{ccccccccccccc} u & u_W(0) & u_W(1) & u_X(0) & u_X(1) & u_Y(0) & u_Y(1) & u_Z(0) & u_Z(1) & u_{E_1} & u_{E_2} & u_{E_3} & u_{E_4} \\ 0&0&0&0&0&0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0 \end{array} \right],$$

$$\boldsymbol{X}^{u_{l_2^1 m_2^1}(i_2^1, j_2^1)} = \begin{array}{c} \\ 0000 \\ 0001 \\ 0010 \\ 0011 \\ 0100 \\ 0101 \\ 0110 \\ 0111 \\ 1000 \\ 1001 \\ 1010 \\ 1011 \\ 1100 \\ 1101 \\ 1110 \\ 1111 \end{array} \left[\begin{array}{ccccccccccccc} u & u_W(0) & u_W(1) & u_X(0) & u_X(1) & u_Y(0) & u_Y(1) & u_Z(0) & u_Z(1) & u_{E_1} & u_{E_2} & u_{E_3} & u_{E_4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{array} \right],$$

$$\boldsymbol{X}^{u_{l_3^1 m_3^1}(i_3^1, j_3^1)} = \begin{array}{c} \\ 0000 \\ 0001 \\ 0010 \\ 0011 \\ 0100 \\ 0101 \\ 0110 \\ 0111 \\ 1000 \\ 1001 \\ 1010 \\ 1011 \\ 1100 \\ 1101 \\ 1110 \\ 1111 \end{array} \left[\begin{array}{ccccccccccccc} u & u_W(0) & u_W(1) & u_X(0) & u_X(1) & u_Y(0) & u_Y(1) & u_Z(0) & u_Z(1) & u_{E_1} & u_{E_2} & u_{E_3} & u_{E_4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right],$$

$$\boldsymbol{X}^{u_{l_4^1 m_4^1}(i_4^1, j_4^1)} = \begin{array}{c} \\ 0000 \\ 0001 \\ 0010 \\ 0011 \\ 0100 \\ 0101 \\ 0110 \\ 0111 \\ 1000 \\ 1001 \\ 1010 \\ 1011 \\ 1100 \\ 1101 \\ 1110 \\ 1111 \end{array} \left[\begin{array}{ccccccccccccc} u & u_W(0) & u_W(1) & u_X(0) & u_X(1) & u_Y(0) & u_Y(1) & u_Z(0) & u_Z(1) & u_{E_1} & u_{E_2} & u_{E_3} & u_{E_4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{array} \right].$$

Then

$$\boldsymbol{X}^{E_1} = \sum_{r=1}^{4} \boldsymbol{X}^{u_{l_r^1 m_r^1}(i_r^1, j_r^1)} =$$

	u	$u_W(0)$	$u_W(1)$	$u_X(0)$	$u_X(1)$	$u_Y(0)$	$u_Y(1)$	$u_Z(0)$	$u_Z(1)$	u_{E_1}	u_{E_2}	u_{E_3}	u_{E_4}
0000	0	0	0	0	0	0	0	0	0	2	0	0	0
0001	0	0	0	0	0	0	0	0	0	1	0	0	0
0010	0	0	0	0	0	0	0	0	0	1	0	0	0
0011	0	0	0	0	0	0	0	0	0	0	0	0	0
0100	0	0	0	0	0	0	0	0	0	1	0	0	0
0101	0	0	0	0	0	0	0	0	0	0	0	0	0
0110	0	0	0	0	0	0	0	0	0	2	0	0	0
0111	0	0	0	0	0	0	0	0	0	1	0	0	0
1000	0	0	0	0	0	0	0	0	0	1	0	0	0
1001	0	0	0	0	0	0	0	0	0	2	0	0	0
1010	0	0	0	0	0	0	0	0	0	0	0	0	0
1011	0	0	0	0	0	0	0	0	0	1	0	0	0
1100	0	0	0	0	0	0	0	0	0	0	0	0	0
1101	0	0	0	0	0	0	0	0	0	1	0	0	0
1110	0	0	0	0	0	0	0	0	0	1	0	0	0
1111	0	0	0	0	0	0	0	0	0	2	0	0	0

.

The same is done for E_2, E_3, and E_4

$\boldsymbol{X}^{E_2} =$

	u	$u_W(0)$	$u_W(1)$	$u_X(0)$	$u_X(1)$	$u_Y(0)$	$u_Y(1)$	$u_Z(0)$	$u_Z(1)$	u_{E_1}	u_{E_2}	u_{E_3}	u_{E_4}
0000	0	0	0	0	0	0	0	0	0	0	0	0	0
0001	0	0	0	0	0	0	0	0	0	0	1	0	0
0010	0	0	0	0	0	0	0	0	0	0	1	0	0
0011	0	0	0	0	0	0	0	0	0	0	2	0	0
0100	0	0	0	0	0	0	0	0	0	0	1	0	0
0101	0	0	0	0	0	0	0	0	0	0	2	0	0
0110	0	0	0	0	0	0	0	0	0	0	0	0	0
0111	0	0	0	0	0	0	0	0	0	0	1	0	0
1000	0	0	0	0	0	0	0	0	0	0	1	0	0
1001	0	0	0	0	0	0	0	0	0	0	0	0	0
1010	0	0	0	0	0	0	0	0	0	0	2	0	0
1011	0	0	0	0	0	0	0	0	0	0	1	0	0
1100	0	0	0	0	0	0	0	0	0	0	2	0	0
1101	0	0	0	0	0	0	0	0	0	0	1	0	0
1110	0	0	0	0	0	0	0	0	0	0	1	0	0
1111	0	0	0	0	0	0	0	0	0	0	0	0	0

,

$\boldsymbol{X}^{E_3} =$

	u	$u_W(0)$	$u_W(1)$	$u_X(0)$	$u_X(1)$	$u_Y(0)$	$u_Y(1)$	$u_Z(0)$	$u_Z(1)$	u_{E_1}	u_{E_2}	u_{E_3}	u_{E_4}
0000	0	0	0	0	0	0	0	0	0	0	0	2	0
0001	0	0	0	0	0	0	0	0	0	0	0	1	0
0010	0	0	0	0	0	0	0	0	0	0	0	1	0
0011	0	0	0	0	0	0	0	0	0	0	0	0	0
0100	0	0	0	0	0	0	0	0	0	0	0	1	0
0101	0	0	0	0	0	0	0	0	0	0	0	2	0
0110	0	0	0	0	0	0	0	0	0	0	0	0	0
0111	0	0	0	0	0	0	0	0	0	0	0	1	0
1000	0	0	0	0	0	0	0	0	0	0	0	1	0
1001	0	0	0	0	0	0	0	0	0	0	0	0	0
1010	0	0	0	0	0	0	0	0	0	0	0	2	0
1011	0	0	0	0	0	0	0	0	0	0	0	1	0
1100	0	0	0	0	0	0	0	0	0	0	0	0	0
1101	0	0	0	0	0	0	0	0	0	0	0	1	0
1110	0	0	0	0	0	0	0	0	0	0	0	1	0
1111	0	0	0	0	0	0	0	0	0	0	0	2	0

,

$$\boldsymbol{X}^{E_4} = \begin{array}{c} \\ 0000 \\ 0001 \\ 0010 \\ 0011 \\ 0100 \\ 0101 \\ 0110 \\ 0111 \\ 1000 \\ 1001 \\ 1010 \\ 1011 \\ 1100 \\ 1101 \\ 1110 \\ 1111 \end{array} \begin{bmatrix} u & u_W(0) & u_W(1) & u_X(0) & u_X(1) & u_Y(0) & u_Y(1) & u_Z(0) & u_Z(1) & u_{E_1} & u_{E_2} & u_{E_3} & u_{E_4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

On the other hand,

$$\boldsymbol{X}^{u,main\ ef.,int.} =$$

$$\begin{array}{c} \\ 0000 \\ 0001 \\ 0010 \\ 0011 \\ 0100 \\ 0101 \\ 0110 \\ 0111 \\ 1000 \\ 1001 \\ 1010 \\ 1011 \\ 1100 \\ 1101 \\ 1110 \\ 1111 \end{array} \begin{bmatrix} u & u_W(0) & u_W(1) & u_X(0) & u_X(1) & u_Y(0) & u_Y(1) & u_Z(0) & u_Z(1) & u_{E_1} & u_{E_2} & u_{E_3} & u_{E_4} \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Then, the matrix $\boldsymbol{X}$, including the edge colouring is

$$\boldsymbol{X} = \boldsymbol{X}^{u,main\ ef.,int.} + \sum_{l=1}^{4} \boldsymbol{X}^{E_l} =$$

$$\begin{array}{c} \\ 0000 \\ 0001 \\ 0010 \\ 0011 \\ 0100 \\ 0101 \\ 0110 \\ 0111 \\ 1000 \\ 1001 \\ 1010 \\ 1011 \\ 1100 \\ 1101 \\ 1110 \\ 1111 \end{array} \begin{bmatrix} u & u_W(0) & u_W(1) & u_X(0) & u_X(1) & u_Y(0) & u_Y(1) & u_Z(0) & u_Z(1) & u_{E_1} & u_{E_2} & u_{E_3} & u_{E_4} \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 2 & 0 & 2 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 2 & 0 & 2 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 2 & 2 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 2 & 0 & 0 & 2 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 2 & 0 & 0 & 2 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 2 & 2 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 2 & 0 & 2 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 2 & 0 & 2 & 0 \end{bmatrix}.$$

The logarithm of the kernel of the likelihood function considering the edge colouring is

$$\log(kernel(L(m))) = \sum_{i\in \mathrm{I}} n(i) \sum_{l=1}^{S} x_i^{E_l} u_{E_l} + \sum_{i\in \mathrm{I}} n(i) \sum_j x_{ij}\beta_j$$
$$- \sum_{i\in \mathrm{I}} \exp(\sum_{l=1}^{S} x_i^{E_l} u_{E_l} + \sum_j x_{ij}\beta_j), \tag{3.23}$$

where x_{ij} are entries of $X^{u,main\ ef.,int.}$.

The likelihood equations obtained from (3.23) are

$$\sum_{i\in \mathrm{I}} n(i) x_i^{E_l} = \sum_{i\in \mathrm{I}} m(i) x_i^{E_l}, l = 1, 2, ..., S; \tag{3.24}$$
$$(\boldsymbol{X}^{u,main\ ef.,int.})' n = (\boldsymbol{X}^{u,main\ ef.,int.})' m. \tag{3.25}$$

These equations, (3.24) and (3.25), can be expressed as

$$\boldsymbol{X}' n = \boldsymbol{X}' m.$$

These results can be summarized in the following theorem.

Theorem 3.5. *The likelihood equations for a RGLL model expressed as a GLM under an edge colouring are*

$$\boldsymbol{X}' n = \boldsymbol{X}' m,$$

where $\boldsymbol{X}$ *is the following matrix,* $\boldsymbol{X} = \boldsymbol{X}^{u,main\ ef.,int.} + \sum_{t=1}^{S} \boldsymbol{X}^{E_t} = \boldsymbol{X}^{u,main\ ef.,int.} + \sum_{t=1}^{S} \sum_{r=1}^{ked(t)} \boldsymbol{X}^{u_{l_r^t m_r^t}(i_r^t, j_r^t)}$*, which includes the edge colouring.*

3.9.3 Vertex and edge colouring

We use the same notation defined before for vertex and for edge colourings. However, we have to consider that $\boldsymbol{\beta}$ contains parameters for both kinds of colourings, and that the design matrix $\boldsymbol{X}$ corresponds to a model whose vertices and edges are simultaneously coloured. We add also the following notation.

$\boldsymbol{X}^{u,int\geq 2}$ matrix including the coefficients corresponding to the constant, and the interactions of two or higher order. The matrix includes zeros in the remaining columns.

We have that

$$\boldsymbol{X} = \boldsymbol{X}^{u,int.\geq 2} + \sum_{l=1}^{S} \boldsymbol{X}^{E_l} + \sum_{i=1}^{T} \sum_{j} \boldsymbol{X}^{V_i(j)}. \tag{3.26}$$

Example 3.13. Consider a model similar to the one in example 3.3, the RGLL model with graph $G(V, E)$, V= $\{W,\ X,\ Y,\ Z\}$ a set of binary variables such that $V = (V_1, V_2)$, $V_1 = \{X, Y\}$ and $V_2 = \{Z, W\}$, with generating class $\{\{X, Y\},\ \{W, Y\},\ \{W, Z\},\ \{X, Z\}\}$, and edge set $E = (E_1, E_2, E_3, E_4)$; E_1 =$\{u_{XY}(00),\ u_{XY}(11),\ u_{WZ}(00),\ u_{WZ}(11)\}$; E_2 =$\{u_{XY}(01),\ u_{XY}(10),\ u_{WZ}(01),\ u_{WZ}(10)\}$; E_3 = $\{u_{WY}(00),\ u_{WY}(11),\ u_{XZ}(00),\ u_{XZ}(11)\}$; E_4 =$\{u_{WY}(01),\ u_{WY}(10),\ u_{XZ}(01),\ u_{XZ}(10)\}$.

The cells are ordered in the following way:

$$((0,0,0,0),(0,0,0,1),(0,0,1,0),(0,0,1,1),(0,1,0,0),(0,1,0,1),(0,1,1,0),(0,1,1,1),$$

$$(1,0,0,0),(1,0,0,1),(1,0,1,0),(1,0,1,1),(1,1,0,0),(1,1,0,1),(1,1,1,0),(1,1,1,1))'$$

The vector of parameters is

$$\boldsymbol{\beta}' = (u, u_1(0), u_1(1), u_2(0), u_2(1), u_{E_1}, u_{E_2}, u_{E_3}, u_{E_4}).$$

We obtain $\sum_{i=1}^{2} \sum_{j=0}^{1} \boldsymbol{X}^{V_i(j)}$ in the same way done in example 3.11.

$$\sum_{i=1}^{2} \sum_{j=0}^{1} \boldsymbol{X}^{V_i(j)} = \begin{array}{c} \\ 0000 \\ 0001 \\ 0010 \\ 0011 \\ 0100 \\ 0101 \\ 0110 \\ 0111 \\ 1000 \\ 1001 \\ 1010 \\ 1011 \\ 1100 \\ 1101 \\ 1110 \\ 1111 \end{array} \begin{bmatrix} u & u_1(0) & u_1(1) & u_2(0) & u_2(1) & u_{E_1} & u_{E_2} & u_{E_3} & u_{E_4} \\ 0 & 2 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 2 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The matrices $\boldsymbol{X}^{E_1}$, $\boldsymbol{X}^{E_2}$, $\boldsymbol{X}^{E_3}$, and $\boldsymbol{X}^{E_4}$ have the same columns for parameters u_{E_1}, u_{E_2}, u_{E_3}, u_{E_4}, respectively, as in example 3.12, we only add zero columns corresponding to the other parameters u, $u_1(0)$, $u_1(1)$, $u_2(0)$, and $u_2(1)$. The matrix $\boldsymbol{X}^{u,int.\geq 2}$ takes value one in the column corresponding to u and zeros in the remaining columns. Then, $\boldsymbol{X}$ is

$$\boldsymbol{X} = \boldsymbol{X}^{u,int.\geq 2} + \sum_{l=1}^{4} \boldsymbol{X}^{E_l} + \sum_{i=1}^{2}\sum_{j=0}^{1} \boldsymbol{X}^{V_i(j)} =$$

	u	$u_1(0)$	$u_1(1)$	$u_2(0)$	$u_2(1)$	u_{E_1}	u_{E_2}	u_{E_3}	u_{E_4}
0000	1	2	0	2	0	2	0	2	0
0001	1	2	0	1	1	1	1	1	1
0010	1	1	1	2	0	1	1	1	1
0011	1	1	1	1	1	0	2	0	2
0100	1	1	1	2	0	1	1	1	1
0101	1	1	1	1	1	0	2	2	0
0110	1	0	2	2	0	2	0	0	2
0111	1	0	2	1	1	1	1	1	1
1000	1	2	0	1	1	1	1	1	1
1001	1	2	0	0	2	2	0	0	2
1010	1	1	1	1	1	0	2	2	0
1011	1	1	1	0	2	1	1	1	1
1100	1	1	1	1	1	0	0	0	2
1101	1	1	1	0	2	1	1	1	1
1110	1	0	2	1	1	1	1	1	1
1111	1	0	2	0	2	2	2	2	0

.

The resulting likelihood equations in vertex and edge colouring models are

$$\sum_{i\in \mathrm{I}} n(i) x_i^{V_l(k)} = \sum_{i\in \mathrm{I}} m(i) x_i^{V_l(k)},\ l = 1, 2, ..., T, k = 1, 2, ..., J; \tag{3.27}$$

$$\sum_{i\in \mathrm{I}} n(i) x_i^{E_l} = \sum_{i\in \mathrm{I}} m(i) x_i^{E_l},\ l = 1, 2, ...S; \tag{3.28}$$

$$(\boldsymbol{X}^{u,int\geq 2})' n = (\boldsymbol{X}^{u,int\geq 2})' m. \tag{3.29}$$

These systems of equations, (3.27), (3.28), and (3.29), can be expressed as

$$\boldsymbol{X}' n = \boldsymbol{X}' m.$$

These results can be summarized in the following theorem.

Theorem 3.6. *The likelihood equations for a RGLL model, which includes both vertex and edge colourings, expressed as a GLM are*

$$\boldsymbol{X}' n = \boldsymbol{X}' m,$$

where $\boldsymbol{X}$ *is the matrix* $\boldsymbol{X} = \boldsymbol{X}^{u,int.\geq 2} + \sum_{l=1}^{S} \boldsymbol{X}^{E_l} + \sum_{i=1}^{T}\sum_{j} \boldsymbol{X}^{V_i(j)}$ *including the vertex and edge colourings.*

In section **3.9.5** we give an example of this way of writing the likelihood equations for symmetry and quasi-symmetry models.

3.9.4 Reparametrized RGLL models

The matrices $\boldsymbol{X}$ obtained for vertex, edge, and vertex and edge colourings, expressions 3.18, 3.22, and 3.26, respectively, do not have full rank. It would be interesting to get

models keeping the same restrictions we have but with reparametrized parameters in such a way that their design matrices $\boldsymbol{Y}$ had full rank for all possible models. Full-rank matrices would allow to easily apply methods based on linear algebra and numerical analysis, like Newton-Raphson method, to approximately solve the corresponding likelihood equations. However, to keep the same restrictions and to get full-rank matrices is not always possible for all models. For vertex colouring we can always get these reparametrized models; but, for edge colouring we can only get them in some particular models. Then, for vertex and edge colouring models we can only get these reparametrized models in some cases.

Vertex colouring

If we have a vertex colouring, then we use the same procedure used for non-reparametrized models, understanding these as the models where the parametrization is the one we have used throughout this dissertation, but we define $\boldsymbol{Y}^{v_l^i(j)}$, l=1,2,..., $kver(i)$, i=1, 2,..., T, for a category $j = 1, ..., J-1$, using some reparametrization; *i.e.*, we have to assign to the column corresponding to $u_i(j)$ in $\boldsymbol{Y}^{v_l^i(j)}$ the column associated to the corresponding reparametrized parameter $u_{v_l^i}(j)$, instead of using ones and zeros as before. Similarly, the $\boldsymbol{Y}^{u,int.}$ matrix depends on the reparametrization. The reparametrizations to be used are any of the commonly seen in most literature, for example, effect coding, which is an ANOVA type reparametrization, or those using dummy variables.

The reason behind this procedure is that equating the reparametrized parameters is equivalent to equating the original parameters, as we will see next. Suppose that we have a variable labeled as X and that we use an effect coding reparametrization with last category, J, as reference category. Suppose also that $u_X(i)$ is the main effect of X for the parametrized model in a level i, $i = 1, ..., J-1$, and that $\lambda_X(i)$ is the main effect for the non-parametrized model in a level i, $i = 1, , ..., J$. The effect coding parametrization implies that

$$\lambda_X(i) = u_X(i),\ i = 1, ..., J-1; \tag{3.30}$$

$$\lambda_X(J) = -u_X(1) - u_X(2) - .... - u_X(J-1). \tag{3.31}$$

If we additionally suppose that we have two variables labeled as X and Y that are in the same colour class, then by definition $\lambda_X(i)$=$\lambda_Y(i)$ for all i, in particular $\lambda_X(i)$=$\lambda_Y(i)$ for $i = 1, , ..., J-1$ and this means according to equation (3.30), that $u_X(i)$=$u_Y(i)$ for $i = 1, , ..., J-1$. Similarly, if $u_X(i)$=$u_Y(i)$ for $i = 1, , ..., J-1$, then $\lambda_X(i)$=$\lambda_Y(i)$ for $i = 1, , ..., J-1$. Using (3.31) and $u_X(i)$=$u_Y(i)$, $i = 1, ..., J-1$, it is immediate that $\lambda_X(J)$=$\lambda_Y(J)$. Then, the equalities over the $J-1$ parameters for the reparametrized model, $u_X(i)$=$u_Y(i)$; $i = 1, , ..., J-1$, are equivalent to the corresponding equalities for the J original parameters, $\lambda_X(i)$=$\lambda_Y(i)$; $i = 1, , ..., J$.

As a consequence, it can be seen that the resulting matrix $\boldsymbol{Y}$ associated to a reparametrized restricted model is obtained by summing the reparametrized main effects of the variables in the same vertex colour class for every category obtained using the reparametrization.

These results can be summarized as follows.

Theorem 3.7. *Under a RGLL model including a vertex colouring only, we can indistinctly use the parameters obtained under effect coding $u_X(i)$, $i = 1, ..., J-1$, or the terms $\lambda_X(i)$, $i = 1, , ..., J$, the terms using the parametrization we have used throughout all this dissertation, to obtain the equality restrictions for the corresponding main effects associated to the colouring. As a consequence, the process to obtain the design matrix associated to the model is the same independently of the parametrization used.*

Example 3.14. To ilustrate how we get the design matrix for a reparametrized RGLL model corresponding to a vertex colouring, suppose that we have two tricotomic variables labeled as X and Y and the RGLL model with graph $G(V, E)$, generating class $A = \{\{X, Y\}\}$, in which both vertices are in the same vertex colour class $V = V_1 = \{v_1^1, v_2^1\}$ with $v_1^1 = X$ and $v_2^1 = Y$ (figure 3.8), and $E = (E_1, ..., E_S)$, with $S = |E|$.

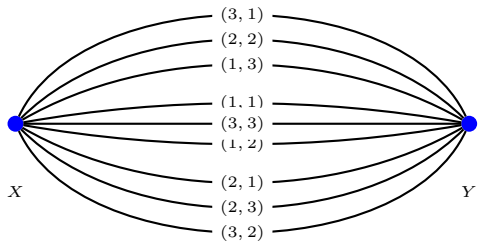


Figure 3.8: RGLL model generated by $\{X, Y\}$, with colour class $V = V_1 = \{X, Y\}$ and $E = (E_1, ..., E_S)$, with $S = |E| = 9$.

The design matrix Z associated to the model reparametrized under effect coding without including the colouring was shown in section **2.4** and it is

$$Z = \begin{array}{c} \\ 11 \\ 12 \\ 13 \\ 21 \\ 22 \\ 23 \\ 31 \\ 32 \\ 33 \end{array} \begin{array}{c} \begin{array}{cccccccc} u & u_X(1) & u_X(2) & u_Y(1) & u_Y(2) & u_{XY}(11) & u_{XY}(12) & u_{XY}(21) & u_{XY}(22) \end{array} \\ \left[\begin{array}{cccccccc} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & -1 & -1 & -1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & -1 & -1 & 0 & 0 & -1 & -1 \\ 1 & -1 & -1 & 1 & 0 & -1 & 0 & -1 & 0 \\ 1 & -1 & -1 & 0 & 1 & 0 & -1 & 0 & -1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \end{array}\right] \end{array}.$$

The first column corresponds to the constant u, the next four to main effects $u_X(i)$, $u_Y(i)$, $i = 1, 2$, and the last four correspond to $u_{XY}(ij)$, $i, j = 1, 2$. The rows correspond to cells, from top to bottom $(1, 1)$, $(1, 2)$, $(1, 3)$, $(2, 1)$, $(2, 2)$, $(2, 3)$, $(3, 1)$, $(3, 2)$, $(3, 3)$.

Under the vertex colouring, we have two parameters $u_1(1)$ and $u_1(2)$ not depending on the variables but on the colouring. Therefore, $\boldsymbol{Y}^{v_1^1(1)}$ has two fewer columns than the previous matrix Z and zeros in all entries except for the ones in the column corresponding to $u_1(1)$ in which we have the values corresponding to $u_X(1)$. The matrix is

$$\boldsymbol{Y}^{v_1^1(1)} = \begin{array}{c} \\ 11 \\ 12 \\ 13 \\ 21 \\ 22 \\ 23 \\ 31 \\ 32 \\ 33 \end{array} \begin{array}{c} \begin{array}{ccccccc} u & u_1(1) & u_1(2) & u_{XY}(11) & u_{XY}(12) & u_{XY}(21) & u_{XY}(22) \end{array} \\ \left[\begin{array}{ccccccc} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \end{array}.$$

Similarly, for $\boldsymbol{Y}^{v_1^2(1)}$ and considering that in all matrices for this example the same labeling for the rows and columns is used because the rows and columns represent the same cells and parameters in each case, we have

$$\boldsymbol{Y}^{v_1^2(1)} = \begin{array}{c} \\ 11 \\ 12 \\ 13 \\ 21 \\ 22 \\ 23 \\ 31 \\ 32 \\ 33 \end{array} \begin{bmatrix} u & u_1(1) & u_1(2) & u_{XY}(11) & u_{XY}(12) & u_{XY}(21) & u_{XY}(22) \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The matrix $\boldsymbol{Y}^{V_1(1)}$ is the sum of $\boldsymbol{Y}^{v_1^1(1)}$ and $\boldsymbol{Y}^{v_1^2(1)}$,

$$\boldsymbol{Y}^{V_1(1)} = \begin{array}{c} \\ 11 \\ 12 \\ 13 \\ 21 \\ 22 \\ 23 \\ 31 \\ 32 \\ 33 \end{array} \begin{bmatrix} u & u_1(1) & u_1(2) & u_{XY}(11) & u_{XY}(12) & u_{XY}(21) & u_{XY}(22) \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Similarly, we get $\boldsymbol{Y}^{V_1(2)}$,

$$\boldsymbol{Y}^{V_1(2)} = \begin{array}{c} \\ 11 \\ 12 \\ 13 \\ 21 \\ 22 \\ 23 \\ 31 \\ 32 \\ 33 \end{array} \begin{bmatrix} u & u_1(1) & u_1(2) & u_{XY}(11) & u_{XY}(12) & u_{XY}(21) & u_{XY}(22) \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Finally, $\boldsymbol{Y}$ is the sum of the previous two matrices with $\boldsymbol{Y}^{u,int.}$, which includes the constant term and the interactions. The matrix $\boldsymbol{Y}$ has full rank and includes the colouring, and it is

$$\boldsymbol{Y} = \boldsymbol{Y}^{u,int.} + \sum_{i=1}^{1}\sum_{j=1}^{2} \boldsymbol{Y}^{V_i(j)} =$$

$$\begin{array}{c} \\ 11 \\ 12 \\ 13 \\ 21 \\ 22 \\ 23 \\ 31 \\ 32 \\ 33 \end{array} \begin{bmatrix} u & u_1(1) & u_1(2) & u_{XY}(11) & u_{XY}(12) & u_{XY}(21) & u_{XY}(22) \\ 1 & 2 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & -1 & -1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 2 & 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 & 0 & -1 & -1 \\ 1 & 0 & -1 & -1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 0 & -1 & 0 & -1 \\ 1 & 2 & -2 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Edge colouring

If we have an edge colouring, then the procedure given for vertex colouring can not be in general applied. This is because the restrictions given to the reparametrized first-order interaction parameters do not always coincide with the equivalent restrictions for the original parameters. In fact, we have fewer reparametrized parameters than original parameters, which means that there could be restrictions on the original parameters that can not be directly seen in the reparametrized parameters. However, there are some particular cases where we can obtain reparametrized models with the same meaning that the original model has, for example, quasi-symmetry models.

Another example of an edge colouring model where we can get a reparametrized model keeping the same restrictions that the original model has is a model with restrictions of the kind

$$\lambda_{XY}(ij) = \lambda_{ZR}(ij),\ i,j = 1,...,J;$$

where λ represents non-reparametrized parameters. Every restriction of this kind for each permutation (i,j) is a different colour class. Additionally, if we had other parameters, λ_{RS}, then we would have the same kind of restrictions or we would have $\lambda_{RS}(ij)$ parameters not restricted. In this case the colour classes are of the kind $E_t = \{e_1^t,...,e_{ked(t)}^t\}$, $t = 1,2,...,S$, where

$$e_k^t = \lambda_{l_k^t m_k^t}(i,j),\ t = 1,2,...,S; k = 1,...,ked(t),$$

for a permutation (i,j). The number of colour classes of this kind for the same group of l_k^t's and m_k^t's is J^2. The total number of colour classes S is determined by the groups of variables whose first-order interactions we are equating.

In this specific edge colouring it is always possible to equate the reparametrized parameters to get the same restrictions given for the original parameters. As in vertex colouring, we can get the design matrix by summing the columns corresponding to the reparametrized parameters in the same class. Then, we obtain $\boldsymbol{Y}^{ul_r^t m_r^t(i_r^t, j_r^t)}$ substituting the column corresponding to the reparametrized parameter $u_{l_r^t m_r^t}(i_r^t, j_r^t)$ in the column corresponding to u_{E_t}, and then we proceed the same as when we did not use any reparametrization.

The reason why in this case is the same equating the reparametrized parameters than equating the other parameters is the following. Suppose that we have two variables X and Y with J levels each one. Under effect coding where the last category, J, is the reference category, we have $J-1$ main effects and $J-1 \times J-1$ first-order interactions. Naming as $\lambda_{XY}(ij)$, $i, j \leq J$ to the original parameters whose values are over all possible permutation of the levels, we have

$$\lambda_{XY}(ij) = u_{XY}(ij),\ i, j \leq J-1, \tag{3.32}$$

$$\lambda_{XY}(iJ) = -\sum_{j=1}^{J-1} u_{XY}(ij),\ i \leq J-1, \tag{3.33}$$

$$\lambda_{XY}(Jj) = -\sum_{i=1}^{J-1} u_{XY}(ij),\ j \leq J-1, \tag{3.34}$$

$$\lambda_{XY}(JJ) = \sum_{i,j=1}^{J-1} u_{XY}(ij). \tag{3.35}$$

Considering two other variables, R and S, with the same characteristics of X and Y and using (3.32), we have

$$u_{XY}(ij) = u_{RS}(ij),\ i, j < J \Leftrightarrow \lambda_{XY}(ij) = \lambda_{RS}(ij),\ i, j < J.$$

If $u_{XY}(ij) = u_{RS}(ij)$, $i, j < J$ then $\lambda_{XY}(iJ) = \lambda_{RS}(iJ)$, $i < J$, because according to (3.33), $\lambda_{XY}(iJ) = -\sum_{j=1}^{J-1} u_{XY}(ij)$, $i \leq J-1$ and $\lambda_{RS}(iJ) = -\sum_{i=1}^{J-1} u_{RS}(ij)$, $i \leq J-1$.

Similarly, if $u_{XY}(ij) = u_{RS}(ij)$, $i, j < J$, using (3.34) and (3.35), we have $\lambda_{XY}(Jj) = \lambda_{RS}(Jj)$, $j \leq J-1$, and $\lambda_{XY}(JJ) = \lambda_{RS}(JJ)$.

This means that by equating the reparametrized parameters $u_{XY}(ij)$ to the reparametrized parameters $u_{RS}(ij)$, for $i, j < J$, we get the restrictions $\lambda_{XY}(ij) = \lambda_{RS}(ij)$, for any i y j, and then we can work with any of the two kinds of parameters.

We can also use indistinctly reparametrized or non-reparametrized terms in RGLL models with restrictions of the following type:

a) $\lambda_{XY}(ij) = \lambda_{XY}(ji)$, $i < j$. This is because as in the previous case we can see from (3.32), (3.33), (3.34), and (3.35) that $u_{XY}(ij) = u_{XY}(ji)$, $i < j$, $i, j = 1, ..., J-1$ $\Leftrightarrow \lambda_{XY}(ij) = \lambda_{XY}(ji)$, $i < j$, $i, j = 1, ..., J$.

b) $\lambda_{XY}(ij) = \lambda_{ZR}(ji)$, $i \leq j$. This is because we can see from (3.32), (3.33), (3.34), and (3.35) that $u_{XY}(ij) = u_{ZR}(ji)$, $i \leq j$, $i, j = 1, ..., J-1 \Leftrightarrow \lambda_{XY}(ij) = \lambda_{ZR}(ji)$, $i \leq j$, $i, j = 1, ..., J$.

All these results can be summarized as follows.

Theorem 3.8. *Under a RGLL model including an edge colouring only, we can indistinctly use the parameters obtained under effect coding $u_{XY}(ij)$, $i, j < J$, or the terms $\lambda_{XY}(ij)$, $i, j \leq J$, the terms using the parametrization we have used throughout all this dissertation, to obtain the equality restrictions for the corresponding first-order interactions associated to the colourings in which we have restrictions of the following type:*

a) $\lambda_{XY}(ij) = \lambda_{ZR}(ij)$, $i, j = 1, ..., J$;

b) $\lambda_{XY}(ij) = \lambda_{XY}(ji)$, $i < j$, $i, j = 1, ..., J$; and

c) $\lambda_{XY}(ij) = \lambda_{ZR}(ji)$, $i \leq j$, $i, j = 1, ..., J$.

As a consequence, the process to obtain the design matrix associated to the model is the same independently of the parametrization used.

Vertex and edge colouring

Vertex and edge colouring models can not always be represented with a reparametrized model in which once restricting in some way its parameters we obtain the restrictions we had with the original parameters. However, for particular cases like symmetry models, reparametrized models can be obtained. Other examples are models with restrictions for the first-order interactions of the form $\lambda_{XY}(ij) = \lambda_{ZR}(ij)$, $\lambda_{XY}(ij) = \lambda_{XY}(ji)$, $i < j$, or $\lambda_{XY}(ij) = \lambda_{ZR}(ji)$, $i \leq j$ adding whichever colouring to the vertices. One particular example of this type of model is obviously the edge colouring model described above; however, in those models all vertices were in different atomic classes, and in a more general example, they could be in different vertex classes.

3.9.5 Example of a RGLL model expressed as a GLM: symmetry and quasi-symmetry

Symmetry and quasi-symmetry models are simple examples in which it can be shown reparametrized and non-reparametrized RGLL models and their associated design matrices. These are particular models where we can get the reparametrized models, although we have vertex and edge colourings. We show these models with their corresponding likelihood equations for binary variables as follows:

i) Quasi-symmetry

Consider two binary variables labeled as C and A, as in example 3.5. A quasi-symmetry model is the RGLL model with graph $G(V, E)$ generated by $\{\{C, A\}\}$ with vertex colour classes $V = (V_1, V_2)$, with $V_1 = \{C\}$ and $V_2 = \{A\}$, and first-order interaction set $E = (E_1, E_2, E_3)$ with $E_1 = \{\lambda_{CA}(00)\}$, $E_2 = \{\lambda_{CA}(11)\}$, and $E_3 = \{\lambda_{CA}(01), \lambda_{CA}(10)\}$. The design matrix for this model is

$$\boldsymbol{X} = \begin{array}{c} \\ 00 \\ 01 \\ 10 \\ 11 \end{array} \begin{array}{c} \begin{array}{cccccccc} \lambda & \lambda_C(0) & \lambda_C(1) & \lambda_A(0) & \lambda_A(1) & \lambda_{E_1} & \lambda_{E_2} & \lambda_{E_3} \end{array} \\ \left[\begin{array}{cccccccc} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \end{array}.$$

The first column corresponds to the constant term λ, the next two columns to $\lambda_C(i)$, $i = 0, 1$, the following two to $\lambda_A(i)$, $i = 0, 1$, and the last three columns to the three edge classes: one including $\lambda_{CA}(00)$, another including the equality $\lambda_{CA}(01) = \lambda_{CA}(10)$, and the last one including $\lambda_{CA}(11)$. The rows represent, from top to bottom, the four cells $(0, 0)$, $(0, 1)$, $(1, 0)$ and $(1, 1)$, where the first entry corresponds to C and the second corresponds to A. Observe that the column corresponding to the equality $\lambda_{CA}(01) = \lambda_{CA}(10)$ is obtained by summing the $\lambda_{CA}(01)$ and $\lambda_{CA}(10)$ effects. Formally, $\boldsymbol{X}$ is obtained using the $\boldsymbol{X}^{v_l^i(j)}$ and the other matrices defined above.

The associated reparametrized model has design matrix

$$\boldsymbol{Y} = \begin{array}{c} \\ 00 \\ 01 \\ 10 \\ 11 \end{array} \begin{array}{c} \begin{array}{cccc} u & u_C(0) & u_A(0) & u_{CA} \end{array} \\ \left[\begin{array}{cccc} 1 & 1 & 1 & 0 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 0 \end{array}\right] \end{array}.$$

As before, the first column corresponds to the constant term, the following two columns correspond to the reparametrized parameters $u_C(0)$ and $u_A(0)$. The last column correspond to a reparametrized parameter u_{CA} representing the equality $\lambda_{CA}(01) = \lambda_{CA}(10)$ associated to the edge colouring. This reparametrization is obtained by eliminating the

columns corresponding to the $\lambda_{CA}(ii)$ parameters and leaving the columns corresponding to the $\lambda_{CA}(ij) = \lambda_{CA}(ji), i \neq j$ restrictions.

Using either matrix, $\boldsymbol{X}$ or $\boldsymbol{Y}$, with corresponding equation systems $\boldsymbol{X}'n = \boldsymbol{X}'m$ or $\boldsymbol{Y}'n = \boldsymbol{Y}'m$, respectively, and eliminating redundant equations, we obtain the set of likelihood equations

$$\begin{aligned} \widehat{m}(i,.) &= n(i,.),\ i = 0,1; \\ \widehat{m}(.,i) &= n(.,i),\ i = 0,1; \\ \widehat{m}(i,j) + \widehat{m}(j,i) &= n(i,j) + n(j,i),\ i \leq j, i, j = 0,1; \end{aligned}$$

which are the likelihood equations corresponding to a quasi-symmetry model, *e.g.*, Agresti (2002a, p. 425).

ii) Symmetry

The symmetry model for V=$\{C, A\}$ is the RGLL model with graph $G(V, E)$ generated by $\{\{C, A\}\}$ with vertex colour class V and first-order interaction set $E = (E_1, E_2, E_3)$ with $E_1 = \{\lambda_{CA}(00)\}$, $E_2 = \{\lambda_{CA}(11)\}$, and $E_3 = \{\lambda_{CA}(01), \lambda_{CA}(10)\}$. The design matrix $\boldsymbol{X}$ for this model is as follows:

$$\boldsymbol{X} = \begin{array}{c} \\ 00 \\ 01 \\ 10 \\ 11 \end{array} \begin{array}{c} \begin{array}{cccccc} \lambda & \lambda_1(0) & \lambda_1(1) & \lambda_{E_1} & \lambda_{E_2} & \lambda_{E_3} \end{array} \\ \left[\begin{array}{cccccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 2 & 0 & 0 & 1 \end{array}\right] \end{array}.$$

As before, the first column corresponds to λ. The second column represents the $\lambda_1(0)$ parameter, that is, the common effect of both variables, which are in the same vertex colour class, in the first category. The third column corresponds to the $\lambda_1(1)$ parameter, the common effect of both variables in the second category. The last three columns correspond to the edge colour classes formed by $\lambda_{CA}(00)$, the class in which $\lambda_{CA}(01) = \lambda_{CA}(10)$, and the class formed by $\lambda_{CA}(11)$, respectively.

The corresponding reparametrized model has design matrix

$$\boldsymbol{Y} = \begin{array}{c} \\ 00 \\ 01 \\ 10 \\ 11 \end{array} \begin{array}{c} \begin{array}{ccc} u & u_1(0) & u_{CA} \end{array} \\ \left[\begin{array}{ccc} 1 & 2 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -2 & 0 \end{array}\right] \end{array}.$$

The first column corresponds to u. The second column represents $u_1(0)$, which is the reparametrization for the unique vertex colour class, and it is obtained summing

the corresponding reparametrized main effects. The last column corresponds to a reparametrized parameter representing the equality $\lambda_{CA}(01) = \lambda_{CA}(10)$, *i.e.* the edge colour class E_3, and it is obtained in the same way as in the quasi-symmetry model.

Using either matrix, $\boldsymbol{X}$ or $\boldsymbol{Y}$, and eliminating redundant equations we obtain the following likelihood equations

$$\begin{aligned} \widehat{m}(i,i) &= n(i,i),\ i = 0,1; \\ \widehat{m}(i,j) + \widehat{m}(j,i) &= n(i,j) + n(j,i),\ i < j, i, j = 0,1; \end{aligned}$$

which are the likelihood equations obtained for a symmetry model, *e.g.*, Agresti (2002a, p. 424).

The reparametrizations for symmetry and quasi-symmetry models already explained can be used for variables with any number of levels. For example, if we had two tricotomic variables, X and Y, the matrix associated to the corresponding reparametrized quasi-symmetry model is

$$\boldsymbol{Y} = \begin{array}{c} \\ 11 \\ 12 \\ 13 \\ 21 \\ 22 \\ 23 \\ 31 \\ 32 \\ 33 \end{array} \begin{array}{c} \begin{array}{cccccccc} u & u_X(1) & u_X(2) & u_Y(1) & u_Y(2) & u_{XY}(12) & u_{XY}(13) & u_{XY}(23) \end{array} \\ \left[\begin{array}{cccccccc} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & -1 & -1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & -1 & -1 & 0 & 0 & 1 \\ 1 & -1 & -1 & 1 & 0 & 0 & 1 & 0 \\ 1 & -1 & -1 & 0 & 1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 \end{array}\right] \end{array}.$$

The rows correspond, from top to bottom, to cells $(1,1)$, $(1,2)$, $(1,3)$, $(2,1)$, $(2,2)$, $(2,3)$, $(3,1)$, $(3,2)$, and $(3,3)$. The first column corresponds to the constant term u, the following two to $u_X(1)$ and $u_X(2)$, the main effects of X for the first two levels of the reparametrized model, the next two to $u_Y(1)$ and $u_Y(2)$, the main effects of Y for the first two levels of the reparametrized model. The three last columns correspond to the restrictions $\lambda_{XY}(12) = \lambda_{XY}(21)$, $\lambda_{XY}(13) = \lambda_{XY}(31)$, $\lambda_{XY}(23) = \lambda_{XY}(32)$, respectively, corresponding to the edge colouring.

And for the symmetry model we have

$$\boldsymbol{Y} = \begin{array}{c} \\ 11 \\ 12 \\ 13 \\ 21 \\ 22 \\ 23 \\ 31 \\ 32 \\ 33 \end{array} \begin{bmatrix} u & u_1(1) & u_1(2) & u_{XY}(12) & u_{XY}(13) & u_{XY}(23) \\ 1 & 2 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 1 \\ 1 & 0 & -1 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 & 1 \\ 1 & -2 & -2 & 0 & 0 & 0 \end{bmatrix}.$$

The rows represent the same cells as before. The second column corresponds to $u_1(1)$, a parameter for the reparametrized model representing the effect of the only vertex colour class at level one and it is obtained by summing in the previous matrix the columns corresponding to the main effects of both variables at level one, the second column corresponds to $u_1(2)$, a parameter for the reparametrized model representing the effect of the only vertex colour class at level two and it is obtained by summing in the previous matrix the columns corresponding to the main effects of both variables at level two. The remaining columns represent the same restrictions discussed above for the symmetry model.

There are alternative ways of reparametrizing symmetry models. Meiser *et al.* (1997) based on the work developed by Rindskopf (1990) and von Eye and Spiel (1996) write the model as

$$\log m(i,j) = \beta + \beta_{ij},\ \beta_{ij} = \beta_{ji}.$$

where β_{ii} is a parameter representing at the same time the main effects and interactions when the levels are equal for both variables. β_{ij} is similar, but it is used when the levels differ. They also use the constraint

$$\sum_i \beta_{ii} = 0,$$

which is an ANOVA type constraint. This constraint makes necessary to use a permutation of the categories of the kind (i,i) as a reference category. Under this parametrization, the symmetry model for two tricotomic variables is

$$\boldsymbol{X} = \begin{array}{c} \\ 11 \\ 12 \\ 13 \\ 21 \\ 22 \\ 23 \\ 31 \\ 32 \\ 33 \end{array} \begin{bmatrix} \beta & \beta_{11} & \beta_{22} & \beta_{12} & \beta_{13} & \beta_{23} \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & -1 & 0 & 0 & 0 \end{bmatrix}$$

The first column corresponds to the constant term β, the second to the reparametrized parameter β_{11} considering β_{33} as a reference category, so that we assign 1 to the entry corresponding to cell $(1,1)$, -1 to $(3,3)$ and zero to the other cells. In a similar way we get the column for β_{22}. The last columns are the same as in the previous reparametrization and belong to β_{ij} for $i < j$ with $\beta_{ij} = \beta_{ji}$.

We notice that all these reparametrized models can be classified as non-standard log-linear models as defined by Rindskopf (1990) and discussed by Mair (2007), which are models whose design matrix has a coding system different to the usual systems, effect coding and using dummy variables.

3.10 Solution of the likelihood equations

The system of likelihood equations for RGLL models can sometimes be solved using a closed expression, for example, for a symmetry model (section **2.5.2**); however, in general, numerical methods are used to get an approximated solution. There are two commonly used methods to fit log-linear models: a) iterative proportional fitting or IPF. This method was introduced by Deming and Stephan (1940), discussed for instance by Bishop *et al.* (1975, p. 83-102), Christensen (1997, p. 87-89), Fienberg (1970), and Lauritzen (1996, p. 82-84), and generalized by expanding the list of probability distributions in product form that is possible to estimate by maximum likelihood by Darroch and Ratcliff (1972); and b) Newton-Raphson method or Fisher scoring method, discussed for instance by Agresti (2002a, p. 143-146 and p. 342-343) and Christensen (1997, p. 346-347).

We adapt these methods to solve the likelihood equations corresponding to RGLL models. The Newton-Raphson method is only used when we have a reparametrized model representing the RGLL model, which means that a full-rank matrix is associated to the model. The Newton-Raphson method used to fit RGLL models is the same method presented in categorical data analysis books, for instance in Agresti (2002a,

p. 342-343), but using the matrix $\boldsymbol{Y}$ associated to the reparametrized model including the colouring, instead of the commonly used design matrix associated to a model. This means that if we have the matrix $\boldsymbol{Y}$, then we can fit the model using any available software that fits log-linear models using the Newton-Raphson method, for example *SPlus*, *SPSS*, *SAS*, *R*, etc.

A method that can always be used is a modification of the IPF method. The commonly used IPF method described for example by Lauritzen (1996, p. 82-84) consists of the following steps applied to every cell i, $i \in \mathrm{I}$:

1. Assign a value to $m_0(i)$, $i \in \mathrm{I}$. For instance, $m_0(i) = 1$.

2. Take all the elements in the generating class A, and order them in a set $(b_1, b_2, ..., b_k)$. Define $T_v = T_{b_v}$, $v = 1, 2, ..., k$, as

$$(T_{b_v} m)(i) = m(i) \frac{n(i_{b_v})}{m(i_{b_v})}, \ i \in \mathrm{I}.$$

   Define recursively

$$m_{r+1}(i) = (T_1 T_2 ... T_k) m_r(i), \ r = 0, 1, 2, ...$$

   In every step, we make adjustments for all the elements in the generating class, which means that we have k sub-steps for every step and every sub-step implies adjustments in such a way that the marginal count is equal to the marginal adjusted expected frequency for every b_k, which is exactly what the corresponding likelihood equations say ($n_a(i_a) = m_a(i_a)$, for all $a \in A$). For example, for the first step, we have the following sub-steps:

$$m_1^1(i) = m_0(i) \frac{n(i_{b_k})}{m_0(i_{b_k})}.$$
$$m_1^{l+1}(i) = m_1^l(i) \frac{n(i_{b_{k-l}})}{m_1^l(i_{b_{k-l}})}, l = 1, 2, ..., k-1.$$

3. The process continues until the maximal difference between the marginal counts and the marginal adjusted expected frequencies reaches a predetermined error δ.

For RGLL models, there are likelihood equations that have to be solved additional to the equations obtained for the elements in the generating class A, equations (3.5) and (3.7), corresponding to first-order interactions and main effects. Therefore, we have to add other transformations. Supposing that all variables have the same number of

categories J, we define for every vertex colour class V_k, $k = 1, 2, ..., T$, the following transformations

$$(T_{V_k(1)}m)(i) = m(i)\frac{\sum_{v_j^k \in V_k} n(v_j^k = 1)}{\sum_{v_j^k \in V_k} m(v_j^k = 1)},\ i \in \{(i_1, i_2, ..., i_\Delta) \in \mathrm{I} | i_{v_j^k} = 1, \text{ for some } v_j^k \in V_k\}$$

$$(T_{V_k(2)}m)(i) = m(i)\frac{\sum_{v_j^k \in V_k} n(v_j^k = 2)}{\sum_{v_j^k \in V_k} m(v_j^k = 2)},\ i \in \{(i_1, i_2, ..., i_\Delta) \in \mathrm{I} | i_{v_j^k} = 2, \text{ for some } v_j^k \in V_k\}$$

$$\vdots$$

$$(T_{V_k(J)}m)(i) = m(i)\frac{\sum_{v_j^k \in V_k} n(v_j^k = J)}{\sum_{v_j^k \in V_k} m(v_j^k = J)},\ i \in \{(i_1, i_2, ..., i_\Delta) \in \mathrm{I} | i_{v_j^k} = J, \text{ for some } v_j^k \in V_k\}.$$

For the edge colour class, E_l, with $l = 1, 2, ..., S$, we define the following transformations for $i \in \{(s_1, s_2, ..., s_\Delta) \in \mathrm{I} | s_{l_v} = i_v,\ s_{r_v} = j_v, \text{ for some } \{l_v = i_v, r_v = j_v\} \in E_l\}$

$$(T_{E_l}m)(i) = m(i)\frac{\sum_{\{l_v = i_v, r_v = j_v\} \in E_l} n(l_v = i_v, r_v = j_v)}{\sum_{\{l_v = i_v, r_v = j_v\} \in E_l} m(l_v = i_v, r_v = j_v)}.$$

In all cases we define $0/0 = 0$. The set $\{l_v = i_v, r_v = j_v\}$ denotes the edge in E_l joining variable l_v to r_v for the value combination (i_v, j_v), $v = 1, ..., ked(l)$, $l = 1, ..., S$.

The algorithm to solve RGLL models is similar to the one explained above; however, we have to add on step 2 these last transformations. That is, we have to get for $r = 0, 1, 2, ...$

$$m_{r+1}(i) = (T_1 T_2 ... T_k T_{V_1(1)} T_{V_1(2)} ... T_{V_1(J)} ... T_{V_T(1)} T_{V_T(2)} ... T_{V_T(J)} T_{E_1} T_{E_2} .... T_{E_S}) m_r(i)$$

We have to consider that the transformation $T_{b_i} = T_i$ is applied for $|b_i| \neq 1, 2$ and that not all transformations associated to the colourings are applied to every cell, it depends on the class and cell. For example, if we have a cell whose entries corresponding to all the variables in a colouring r are all different to j, then there is no sense to apply to this cell the transformation corresponding to $T_{V_r(j)}$.

The proof of the convergence of this method to the maximum likelihood estimators is shown in Appendix B.

We have written a Fortran program to fit RGLL models included in the group of programs called REGRAPH. It fits models using the modified IPF method. It is written in *Fortran 6.5*, using some subroutines from Haberman (1972, 1976) to fit log-linear models, but most of the subroutines were specifically created for these models. For a

RGLL model, REGRAPH calculates the fitted expected frequencies, deviance, associated design matrix for the non-reparametrized model, and degrees of freedom for the asymptotic chi square distribution associated to the deviance using the design matrix. Notice that in quasi-symmetry models, exact p-value calculations have been obtained by numerical computation (Booth *et al.*, 2005).

Numerical results obtained with REGRAPH have been compared with those obtained by using other software like *MIM 3.2.0.6* or *Splus 6.1* for models in which the comparison is possible: log-linear models without restrictions, including graphical models, and symmetry and quasi-symmetry models (section **3.11**). In particular, we made that comparison for the data we present in section **3.11.5** corresponding to risk factors for coronary heart disease, observing that under the graphical log-linear model presented there we obtained the same fitted expected frequencies, deviance, and degrees of freedom using *REGRAPH* or *Splus*. We also made comparisons for the symmetry and quasi-symmetry models applied to the change of residence data introduced in table 2.3 and discussed in section **2.5**, and we observed that under the symmetry model the fitted expected frequencies, deviance, and degrees of freedom were exactly the same using *REGRAPH* or *Splus* and under the quasi-symmetry model the deviance and degrees of freedom were the same and the few cells whose expected frequencies differed corresponded to differences of only one hundredth. We do not present all these tables here because they are tables in which the same values are constantly repeated; however, the quasi-symmetry model and RGLL models obtained from it are further discussed in section **3.11.5** so that the fitted expected frequencies under the quasi-symmetry model can be found in table 3.3 using *REGRAPH* and table 2.4 using *S-Plus*.

In REGRAPH, the user provides the data, including the number of variables and number of categories for each variable, as well as the generating class, and the vertex and edges colour classes of the model. In the following section we show in detail how this process is done.

The generating class has to be known, if we do not know it, we could use a model search procedure to look for a graphical log-linear model that fits the data. This process can be done in any available software for graphical models, for example, *MIM*.

3.11 Fit and selection of RGLL models with REGRAPH

In this section we describe the set of six independent programs that conform REGRAPH that allow to fit, join colour classes, and select RGLL models for some data.

REGRAPH was written in *Fortran 6.5*. This language was chosen because of its speed, specially for algorithms that involve a lot of nested cycles, which is what happens in this problem. In *Fortran* each of the independent programs in REGRAPH forms a *Workspace* which is composed of different subroutines found in different independent files and a main program that calls all these subroutines, all of these files are opened once we open the Workspace. All programs and their corresponding subroutines are commented and explained in their corresponding files, so that the user is able to understand each step and if he wants to, he can modify any subroutine or process.

The different processes that can be done with the different Workspaces in REGRAPH are:

1. Fit a RGLL model. The user provides the data, the generating class, and the vertex and edge colour classes. The name of this Workspace is *ajustemodloglinealrestingles8102008.dsw*.

2. Join two vertex colour classes. In this case the user has a RGLL model and what he needs to know is if joining two vertex colour classes, he gets a model with a better fit to the data. This Workspace is called *unirclasesporvertingles10102008.dsw*.

3. Join two edge colour classes. In this case the user has a RGLL model and what he needs to know is if joining two edge colour classes, he gets a model with a better fit to the data. This Workspace is called *unirclasesporaringles10102008.dsw*.

4. Selection of a RGLL model restricted only in its main effects or vertex colouring. This Workspace, called *aprenderclasesverting10102008.dsw*, allows us to get a RGLL model restricted in its main effects only that fits the data.

5. Selection of a RGLL model restricted only in its first-order interactions or edge colouring. This Workspace, called *aprenderclasesaring10102008.dsw*, allows us to get a RGLL model restricted in its first-order interactions only that fits the data.

6. Selection of a RGLL model restricted in both its main effects and first-order interactions. This Workspace, called *aprenderclasveryaring10102008.dsw*, allows us to get a RGLL model restricted in both its main effects and first-order interactions that fits as good as possible to some data.

3.11.1 Reading data and other information

There are three input files that the user provides: the data, the generating class, and a file with the edge colour classes.

Data should be given as a table in a file with extension .dat. We consider that the variables are labeled as variable 1, variable 2, etc. This input file must include: a) the number of variables $|V|$ as an integer with maximum length 5, b) the number of categories or levels of each variable $(|\mathrm{I}_\delta|)_{\delta \in V}$, and c) a table with all possible values that can be taken by all variables or cells i followed by the observed count $n(i)$ and by the algorithm initial values for each cell $m_0(i)$, it has to be a positive number and by convention $m_0(i) = 1$ is generally used. In both, levels and cells, we consider the same order previously given for the labeling; *i.e.* the first column corresponds to variable 1, the second one to variable 2, and so on. For simplicity, all values except the number of variables, are introduced as real values with two decimal digits and a maximum length of 8 including the decimal point. This eliminates the need of having to distinguish which values are real and which are not.

The generating class should be provided in a file with extension .dat. This file includes: a) the number of elements in the generating class $|A|$ and b) in different rows the variables forming each of these elements. All these values are introduced as integers with a maximum extension of 5 digits. This restriction can be modified in the main program. We suppose that the generating class is provided by the user, that is, the program does not select generating classes for RGLL models. If the user does not know which generating class fits the data, then he can get it using any available software that allows model selection for graphical models, we chose to use *MIM* which is freely available in `http://www.hypergraph.dk`, and with a manual describing implementation-specific aspects of MIM in Edwards (2000, Appendix B).

REGRAPH provides different options with the goal that, according to the user desires, the process could become more automatized or that we could have more control over some parts. One of these options is the possibility of fitting models in which each vertex and edge are in different atomic classes, *i.e.* unrestricted models, which we call *default models*. In this case, the program obtains all simple edges as if the model were graphical using the generating class, these edges are of the form $\{i, j\}$, with $i < j$, where i and j are labels for the variables, and from them it obtains the multiple edges considering all the level permutations. Additionally, the program gets from the generating class all subsets containing three or more elements eliminating all repeated subsets, these subsets are important because we use them to get the degrees of freedom associated to the model. Other options available in REGRAPH are changing the names for the output files or for the labeling in the graph.

When the user provides colour classes, *i.e.* when not all classes are atomic, we observed that everything was simplified if the user provided a file with the elements forming each edge colour class. Then, for the edge classes, the program requires the number of colour classes S and the number of elements $ked(t)$ in each class t, $t = 1, ..., S$. The user provides a .dat file containing E_1, E_2, ..., E_S in the following way: the first

$ked(1)$ rows are the edges in the class 1, E_1, the following $ked(2)$ represent the edges in class 2, E_2, and so on. The edges are given using four columns, the first two correspond to the variables that are joined by an edge, l_r^t and m_r^t, $r = 1, 2, ..., ked(t)$; t=1,2,...,S, and the last two columns correspond to a particular permutation of the levels of the two variables that are joined, (i_r^t, j_r^t), $r = 1, 2, ..., ked(t)$; t=1,2,...,S. The file is formed by integer values with a length of maximum 5 digits. For the vertex colouring is quite simple for the user to provide the classes V_1, ..., V_T that the program needs without providing an additional file.

A final observation is that we have fitted and analyzed RGLL models with at most 6 variables and 4 levels; however, all algorithms were built considering that we could have greater number of variables and levels. Of course, if the number of variables and levels increases too much, then we could start having problems with REGRAPH because the input files become longer and more difficult to write, for instance providing the data containing the edge colour classes could be difficult considering that the number of first-order interactions also increases. On the other hand, obtaining the design matrix X requires a lot of nested cycles and if the number of variables or levels increased too much, these processes would get too slow, in this case the routines used to obtain the rank of X would also start to fail. As we increase the number of variables or levels we probably would start having some problems with the routines used to obtain marginal counts and to get proportional adjustment because the number of cells in the corresponding marginal tables also increases, then a lot of margins should be obtained and consequently the routines for proportional adjustment will require longer processes because they depend on different margins and sum of these margins updated in each step of the iterations and even though not all margins are stored in new vectors in each step, they still are obtained using many nested cycles. There are even more cycles when we consider all steps required to obtain convergence. Finally, increasing the number of variables may produce graphs with too many edges and colours with no simple interpretation.

3.11.2 Fitting a RGLL model

Once the user provides the data, generating class, and colouring, the likelihood equations can be approximately solved using the modified IPF algorithm defined in section **3.10**. Firstly, the program provides the estimated parameters for a graphical log-linear model without considering colour classes according to the generating class. These estimates are obtained using a Newton-Raphson algorithm available in one library in *Fortran*. These values could be useful if the user wanted to identify if some parameters are similar to see if there are some colour classes he could use for his data; however, if he does not have any idea of which classes to use, it is advisable to use a model selection method as we describe below in section **3.11.4**.

The most important routines used in REGRAPH to fit a RGLL model are the following:

1. *collapver* and *collapar*. Haberman (1972, 1976) proposed three subroutines used to fit log-linear models. One of these routines is called *COLLAP* and allows us to get the marginal counts from a contingency table according to a generating class. In log-linear models these margins are used in the original table and in the different tables generated when we fit the expected frequencies in each iteration. We modified this routine in subroutines that allow us to get the sum of margins required both for the vertex and edge colourings; *i.e.*, we get values of the type $n(v_k^i = l) = \sum_{j:j_{v_k^i}=l} n(j),\ \ l = 1, 2, ..., J;\ k = 1, ..., kver(i);\ i = 1, 2, ..., T$ and $n(l_k^t = i_k^t, m_k^t = j_k^t) = \sum_{s:(s_{l_k^t}, s_{m_k^t})=(i_k^t, j_k^t)} n(s),\ k = 1, ..., ked(t);\ t = 1, ..., S$. We called these subroutines *collapver* and *collapar*, respectively.

2. *sumver* and *sumaris*. These subroutines allow us to sum some margins according to the elements conforming the classes; *i.e.*, they are used to obtain values of the kind $\sum_{v_j^k \in V_k} m(v_j^k = l),\ l = 1, 2, ..., J;\ k = 1, 2, ..., T$ or $\sum_{\{l_v = i_v, r_v = j_v\} \in E_t} m(l_v = i_v, r_v = j_v),\ t = 1, ..., S$; respectively.

3. *LOGLINREST*. Haberman (1972, 1976) proposes a routine called *ADJUST* that allows to get a proportional adjustment for each cell in a table according to the elements in the generating class. This routine was used to fit the elements in the generating class with a cardinality greater than three and parts of it were the basis to create processes that allow the fit of vertex and edge colour classes. These processes were included in a bigger routine, subroutine *LOGLINREST*, which is used to iteratively solve the likelihood equations corresponding to a RGLL model according to the modified IPF method. Here and in general, we considered posible errors we could have because the data provided by the user were wrong, for example when a set in the generating class has the same variable repeated twice or more times, with something missing, or if there were a problem in the fitting process. These errors are reported to the user. Some of these possibilities made necessary to create new subroutines or modify the existing ones, for example the lack of edge colour classes, like when we do not have edges, which happens in an independence model, made necessary to create a new routine called *LOGLINRESTScero* which is a modification of subroutine *LOGLINREST*.

4. *Xinterac*, *efprinc*, and *Xaris*. We created algorithms and their corresponding associated subroutines, that allow us to obtain for any RGLL model the parts of the design matrix $\boldsymbol{X}$ for interactions of order greater than one $\boldsymbol{X}^{int.\geq 2}$, for vertex colourings $\sum_{i=1}^{T} \sum_j \boldsymbol{X}^{V_i(j)}$, and for edge colourings $\sum_{l=1}^{S} \boldsymbol{X}^{E_l}$; subroutines *Xinterac*, *efprinc*, and *Xaris*, respectively, as described in sections **3.9.1**, **3.9.2**, and **3.9.3**. Parts of these algorithms were inspired on the method presented by

Murray and Speed (1978). The different parts conforming the design matrix $\boldsymbol{X}$, *i.e.*, the matrices for the vertex colouring, edge colouring, and interactions of order greater than one are saved with the names *matrizefprinc.dat*, *matrizaristas.dat*, and *matrizinteracciones.dat*, repectively.

5. *gradoslibcol* and *gradoslibScero*. We get the design matrix $\boldsymbol{X}$, which is available for the user with the name *matrizdiseno.dat*. We also get the rank of the design matrix and as a consequence the number of degrees of freedom associated to the model. In the main program the number of degrees of freedom are corrected when there are estimated zeros or if some of the margins for the elements we want to fit are zero.

6. *pvalue*. Using the number of degrees of freedom and the deviance value we get the corresponding p-value for testing the hypothesis that a RGLL model fits the data.

7. *graficascoloreadas* and *graficascoloreadasdef*. These subroutines generate a file that can be used to get the corresponding associated graph according to the options selected by the user, these can be default options or the user could provide the width, height, direction, and labels used on the graph and the name of the file.

All these subroutines are joined in a main program. This program reads the data and orders them according to the order used in *Fortran* to write matrices. In fact, most of the information, like the generating class, the elements conforming the vertex colour classes, and each of the elements conforming the edge colour classes (labels and levels) are rewritten in a matricial form or in some cases in a vectorial form, for example the observed counts and fitted values. After this process, we call all the different subroutines including the ones described above to finally fit the model.

We have as output the deviance, number of degrees of freedom, and the p-value associated to the model. Additionally, the fitted expected frequencies $\widehat{m}(i)$ are saved in a file with extension .dat containing the fitted values in the same order as in the table provided by the user. We also get as output a file with extension .dot containing instructions that can be easily run in the free software *Graphviz* (`http://www.grapviz.org`) by simply opening it and pressing the run button. By doing this, we get an image file containing the coloured graph associated to the model.

By default, the output files for the fitted values are named by concatenating the word *fit* with the name of the data and the number of vertex and edge colour classes. For the graph output, the word *graph* is concatenated instead of the word *fit*. In addition, by default the maximum number of iterations for the modified IPF algorithm is 5000 and the maximum allowed difference between the estimated margins or the sum of

estimated margins that solve the different likelihood equations and the corresponding observed values is 0.05.

A final observation is that in this Workspace, as well as in all others, all input and output files are contained in the same folder in which the Workspace is located.

3.11.3 Joining colour classes

In this case the user has a RGLL model and he provides two colour classes he wants to join, such classes are chosen using information that the user somehow knows or suspects. What the user wants to know is if joining those colour classes, he gets a model with a better fit to the data. When he does not know which classes to join, then a selection method may be used instead as discussed in section **3.11.4**, the selection methods automatically join colour classes according to a computational procedure that iteratively joins colour classes.

Vertex colour classes

The subroutines used in the corresponding Workspace are the same already seen in section **3.11.2** plus two additional subroutines:

unirvert and *maxkverprima*. These subroutines generate matrices used internally in REGRAPH to represent vertex colour classes updated according to the new vertex colour classes obtained once joining two classes. Such matrices will be used in the subroutine that fits the new model and to get a new design matrix in which we represent the new colour classes.

The main program is basically the same as before, as well as the required input files, but we add a part for the class joint. The user provides the classes he wants to join and a significance level. This significance level is used to test the null hypothesis that both models explain in a similar way the data, which happens when the deviance difference is small. If we do not reject the null hypothesis, we keep the model joining the classes because it is a more parsimonious model.

The process followed in the main program is the following. We fit the new model obtained by joining colour classes and take the degrees of freedom and deviance differences between the new and original models. If the degrees of freedom difference is zero, we choose the model with the largest p-value. If the degrees of freedom difference is positive, we have two cases: when the deviance difference is negative, we prefer the new reduced model because it has a smaller deviance and more degrees of freedom than the larger model. When the deviance difference is non-negative, we need to do an hypothesis test to test that both models explain in a similar way the data using the

deviance and degrees of freedom differences. If the p-value for this test is greater than the significance level, we do not reject the null hypothesis and we keep the reduced model.

REGRAPH provides the user for both, the original and reduced model, with the deviances, degrees of freedom, p-values, and the fitted values under the original model and under the reduced model when it is convenient to join the colour classes. By default the fitted values are saved in files named by concatenating the word *fit1*, for the original model, or *fit2*, for the new model, with the name of the file containing the data, and the number of vertex and edge colour classes in the model. The program also produces as output the necessary files to create the graph of the model we chose as discussed in section **3.11.2**. The name of these files are composed by default by the concatenation of the word *graph* with the name of the data and the corresponding vertex and edge colour classes.

Edge colour classes

The subroutines, processes, hypothesis tests, and outputs are similar to the ones explained in the previous case only that using edge colour classes instead of vertex colour classes. In this case, the corresponding Workspace requires two additional subroutines:

unirar and *maxkarprima*. These subroutines generate matrices used internally in REGRAPH to represent edge colour classes updated according to the new colour classes obtained once joining two edge colour classes. Such matrices will be used in the subroutine that fits the new model and to get a new design matrix in which we represent the new colour classes.

3.11.4 Model selection

In this section we discuss selection methods that can be used to obtain from an initial RGLL model, or even a graphical log-linear model or default model when the user does not have a RGLL model, a RGLL model that fits, according to tests based on the deviance, as good as possible to the data. This process is obtained by iteratively joining colour classes.

Selection of a RGLL model restricted only in its main effects

The subroutines contained in the corresponding Workspace, *aprenderclasesverting1010 2008.dsw*, are the same used in *unirclasesporvertingles10102008.dsw*, the Workspace that joins vertex colour classes, but now we iteratively apply the joint of distinct classes.

We start with an initial RGLL model provided by the user, the input files are the same explained in section **3.11.1**. If we do not have an initial RGLL model, *i.e.* we

know the generating class but not the colouring, we could tell the program that we are using a default model, *i.e.* the model in which each vertex and edge is in a different atomic class. Using the initial model and a significance level provided by the user, for example $\alpha = 0.05$, the program iteratively joins, when it is convenient, different vertex colour classes.

In general terms, the iterative process consists in testing if it is convenient to join the first and last class according to the criteria seen above in **3.11.3**, if it is so, we get a new class and we update all matrices and vectors representing this new colouring. Otherwise, we test if it is convenient to join the second and last class, if it is not the case, we continue the process in a similar way until we test if it is convenient to join the the last but one and last class, if this joint were not convenient, then we would start the same comparison but using the last but one class instead of the last class. The process continues by trying to join different colour classes until we cover all possible joints. When two any classes are joined, the new class is considered as a new class, so that we apply the same process but considering the new class. The process ends when it is not convenient to join more classes or when all classes are joined into one.

REGRAPH allows the user to modify options along the process, including those in which default values are defined, for example the name of the output files, graph labels, maximum number of iterations, etc. We get for both, the original and selected model, the same statistics as before: deviance, degrees of freedom, and p-values. We also get the matrix for the elements forming each colour class. This matrix has as many columns as classes and each column represents the vertices forming each class; *i.e.* column i contains the elements v_k^i, k=1,...,$kver(i)$. This information is also represented in the graph that REGRAPH generates.

Finally, we get for the initial and selected model as output the fitted expected frequencies and a file that allows to obtain the graph for the selected model as described in **3.11.2**. The fitted values are saved by default in a file whose name is formed by concatenating the word *fit1*, for the original model, or *fit2*, for the chosen model, with the name of the data and the vertex and edge colour classes that each model contains. The graph is named by default by concatenating the word *graph* with the name of the data and the vertex and edge colour classes.

Selection of a RGLL model restricted only in its first-order interactions

The process to select a model is the same used in selection of a RGLL model restricted only in its main effects, the only difference is that we use edge colour classes instead of vertex colour classes.

As before, from an initial RGLL model provided by the user or a default model, a

RGLL model that fits as good as possible to the data according to the goodness of fit statistic corresponding to the deviance is obtained with the corresponding value of its deviance and fitted cell counts. These values are obtained for both, the original and selected model.

We also get four matrices that indicate which edges form each colour class. These matrices have as many columns as colour classes. A first matrix has columns whose rows for the t column represent the variables that conform the first element of which will become the edges in the colour class t, *i.e.* l_r^t, $r = 1, 2, ..., ked(t)$. For the second matrix and the same column t, the rows represent the variables conforming the second element for the edges in the colour class t, *i.e.* m_r^t, $r = 1, 2, ..., ked(t)$. The last two matrices are similar but correspond to the levels conforming the edges, *i.e.* i_r^t and j_r^t, $r = 1, 2, ..., ked(t)$. So that if we take the row r and the column t in each of these matrices, then we get the edge r in the colour class t or $u_{l_r^t m_r^t}(i_r^t j_r^t)$ because we have the variables and levels forming it. This information is also represented in the graph REGRAPH generates. The default names for the output files are similar to the ones described above for RGLL models restricted only in their main effects.

Selection of a RGLL model restricted in its main effects and first-order interactions

This Workspace first performs the process described for selection of a RGLL model restricted only in its main effects and after that the process described for selection of a RGLL model restricted only in its first-order interactions. Then, from a given RGLL model, possibly a default model, and a given significance level, we get a RGLL model colouring both the vertices and edges that fits as good as possible to the data. As before, the program displays goodness of fit statistics as well as the fitted expected frequencies for both the original and selected models. Additionally, we obtain a coloured graph corresponding to the selected model and matrices that show which elements are in which colour class. This information is also contained in the graph generated by REGRAPH. The default names for the output files are similar to the ones described for the two other cases.

3.11.5 Examples

The first example corresponds to a data set analyzed by Edwards (2000, p. 22-26) and Edwards and Havránek (1985) corresponding to risk factors for coronary heart disease for a sample of 1,841 car-workers in Czechoslovakia to whom the following questions were asked: if they had family with history of coronary heart disease (F), if the ratio of beta to alpha lipoproteins was less than three (E), if their systolic blood pressure was less than 140 mm. (D), if their work was strenuous physically (C), if their work was strenuous mentally (B), and if they smoked (A).

All variables have the same two categories labeled as 1 and 2 for Yes and No, respectively. Edwards presents the data in the format used in *MIM* and according to the way in which Edwards (2000, p. 255) explains how the format is defined, we organized the data in a table (table 3.2). Observe that we present all values as real numbers with two decimals because as we explained in section **3.11.1** this is the way the data are read in REGRAPH.

Edwards (2000, p. 24-26) selected an adequate graphical log-linear model for these data by using in *MIM* a backward selection method starting with the saturated model as initial model. We recreated the same process using *MIM 3.2.0.6* and got the same graphical log-linear model he obtained, see Edwards (2000, p. 24-25), such model has generating class $\{\{A, D, E\}, \{A, C, E\}, \{B, F\}, \{A, B, C\}\}$, Model 1.

F	E	D	C	B	A	Obs. count	Initial val.	Exp. freq. mod. 1	Exp. freq. mod. 2	Dev. cont. mod. 1	Dev. cont. mod. 2
1.00	1.00	1.00	1.00	1.00	1.00	44.00	1.00	41.05	49.16	6.11	-9.76
1.00	1.00	1.00	1.00	1.00	2.00	40.00	1.00	33.23	37.42	14.83	5.33
1.00	1.00	1.00	1.00	2.00	1.00	112.00	1.00	106.24	109.05	11.83	5.98
1.00	1.00	1.00	1.00	2.00	2.00	67.00	1.00	69.22	63.34	-4.37	7.53
1.00	1.00	1.00	2.00	1.00	1.00	129.00	1.00	124.31	125.95	9.55	6.17
1.00	1.00	1.00	2.00	1.00	2.00	145.00	1.00	142.15	146.66	5.76	-3.30
1.00	1.00	1.00	2.00	2.00	1.00	12.00	1.00	14.27	10.60	-4.16	2.98
1.00	1.00	1.00	2.00	2.00	2.00	23.00	1.00	24.07	20.64	-2.09	4.98
1.00	1.00	2.00	1.00	1.00	1.00	35.00	1.00	32.67	34.57	4.82	0.87
1.00	1.00	2.00	1.00	1.00	2.00	12.00	1.00	16.08	16.33	-7.02	-7.39
1.00	1.00	2.00	1.00	2.00	1.00	80.00	1.00	84.54	76.70	-8.83	6.74
1.00	1.00	2.00	1.00	2.00	2.00	33.00	1.00	33.50	34.98	-0.99	-3.85
1.00	1.00	2.00	2.00	1.00	1.00	109.00	1.00	98.92	112.45	21.15	-6.79
1.00	1.00	2.00	2.00	1.00	2.00	67.00	1.00	68.80	64.02	-3.55	6.10
1.00	1.00	2.00	2.00	2.00	1.00	7.00	1.00	11.35	9.43	-6.77	-4.17
1.00	1.00	2.00	2.00	2.00	2.00	9.00	1.00	11.65	11.40	-4.65	-4.25
1.00	2.00	1.00	1.00	1.00	1.00	23.00	1.00	27.01	22.42	-7.39	1.17
1.00	2.00	1.00	1.00	1.00	2.00	32.00	1.00	31.19	29.01	1.64	6.28
1.00	2.00	1.00	1.00	2.00	1.00	70.00	1.00	69.90	74.01	0.20	-7.80
1.00	2.00	1.00	1.00	2.00	2.00	66.00	1.00	64.98	63.56	2.06	4.97
1.00	2.00	1.00	2.00	1.00	1.00	50.00	1.00	52.83	50.10	-5.51	-0.20
1.00	2.00	1.00	2.00	1.00	2.00	80.00	1.00	84.51	84.72	-8.77	-9.17
1.00	2.00	1.00	2.00	2.00	1.00	7.00	1.00	6.06	7.70	2.02	-1.33
1.00	2.00	1.00	2.00	2.00	2.00	13.00	1.00	14.31	15.43	-2.50	-4.46
1.00	2.00	2.00	1.00	1.00	1.00	24.00	1.00	26.86	22.30	-5.40	3.53
1.00	2.00	2.00	1.00	1.00	2.00	25.00	1.00	26.11	21.45	-2.17	7.66
1.00	2.00	2.00	1.00	2.00	1.00	73.00	1.00	69.51	73.61	7.15	-1.21

Continues in the following page

Continues from the previous page

F	E	D	C	B	A	Obs. count	Initial val.	Exp. freq. mod. 1	Exp. freq. mod. 2	Dev. cont. mod. 1	Dev. cont. mod. 2
1.00	2.00	2.00	1.00	2.00	2.00	57.00	1.00	54.39	59.44	5.34	-4.78
1.00	2.00	2.00	2.00	1.00	1.00	51.00	1.00	52.54	49.82	-3.03	2.39
1.00	2.00	2.00	2.00	1.00	2.00	63.00	1.00	70.73	62.62	-14.58	0.76
1.00	2.00	2.00	2.00	2.00	1.00	7.00	1.00	6.03	7.65	2.09	-1.24
1.00	2.00	2.00	2.00	2.00	2.00	16.00	1.00	11.98	14.43	9.26	3.30
2.00	1.00	1.00	1.00	1.00	1.00	5.00	1.00	5.92	5.13	-1.69	-0.26
2.00	1.00	1.00	1.00	1.00	2.00	7.00	1.00	4.79	5.83	5.31	2.56
2.00	1.00	1.00	1.00	2.00	1.00	21.00	1.00	20.53	19.84	0.95	2.39
2.00	1.00	1.00	1.00	2.00	2.00	9.00	1.00	13.38	11.52	-7.14	-4.44
2.00	1.00	1.00	2.00	1.00	1.00	9.00	1.00	17.93	10.38	-12.41	-2.57
2.00	1.00	1.00	2.00	1.00	2.00	17.00	1.00	20.50	22.84	-6.37	-10.04
2.00	1.00	1.00	2.00	2.00	1.00	1.00	1.00	2.76	2.88	-2.03	-2.12
2.00	1.00	1.00	2.00	2.00	2.00	4.00	1.00	4.65	3.76	-1.20	0.50
2.00	1.00	2.00	1.00	1.00	1.00	4.00	1.00	4.71	3.61	-1.31	0.82
2.00	1.00	2.00	1.00	1.00	2.00	3.00	1.00	2.32	3.22	1.54	-0.42
2.00	1.00	2.00	1.00	2.00	1.00	11.00	1.00	16.34	13.95	-8.71	-5.23
2.00	1.00	2.00	1.00	2.00	2.00	8.00	1.00	6.47	6.36	3.40	3.67
2.00	1.00	2.00	2.00	1.00	1.00	14.00	1.00	14.27	11.73	-0.53	4.95
2.00	1.00	2.00	2.00	1.00	2.00	17.00	1.00	9.92	12.61	18.31	10.16
2.00	1.00	2.00	2.00	2.00	1.00	5.00	1.00	2.19	2.56	8.26	6.69
2.00	1.00	2.00	2.00	2.00	2.00	2.00	1.00	2.25	2.07	-0.47	-0.14
2.00	2.00	1.00	1.00	1.00	1.00	7.00	1.00	3.90	4.42	8.19	6.44
2.00	2.00	1.00	1.00	1.00	2.00	3.00	1.00	4.50	4.52	-2.43	-2.46
2.00	2.00	1.00	1.00	2.00	1.00	14.00	1.00	13.51	13.46	1.00	1.10
2.00	2.00	1.00	1.00	2.00	2.00	14.00	1.00	12.56	13.33	3.04	1.37
2.00	2.00	1.00	2.00	1.00	1.00	9.00	1.00	7.62	7.80	3.00	2.58
2.00	2.00	1.00	2.00	1.00	2.00	16.00	1.00	12.19	13.19	8.70	6.18
2.00	2.00	1.00	2.00	2.00	1.00	2.00	1.00	1.17	2.09	2.14	-0.18
2.00	2.00	1.00	2.00	2.00	2.00	3.00	1.00	2.77	3.24	0.48	-0.46
2.00	2.00	2.00	1.00	1.00	1.00	4.00	1.00	3.87	4.39	0.26	-0.74
2.00	2.00	2.00	1.00	1.00	2.00	0.00	1.00	3.77	4.23	0.00	0.00
2.00	2.00	2.00	1.00	2.00	1.00	13.00	1.00	13.43	13.39	-0.85	-0.77
2.00	2.00	2.00	1.00	2.00	2.00	11.00	1.00	10.51	12.46	1.00	-2.74
2.00	2.00	2.00	2.00	1.00	1.00	5.00	1.00	7.58	7.76	-4.16	-4.40
2.00	2.00	2.00	2.00	1.00	2.00	14.00	1.00	10.20	12.34	8.87	3.53
2.00	2.00	2.00	2.00	2.00	1.00	4.00	1.00	1.17	2.08	9.83	5.23
2.00	2.00	2.00	2.00	2.00	2.00	4.00	1.00	2.31	3.03	4.39	2.22

Table 3.2: Observed counts, fitted values, and deviance contribution for a study of coronary heart disease in a sample of 1,841 car-workers under two models: a) Model 1, graphical log-linear model with generating class $\{\{A, D, E\}, \{A, C, E\}, \{B, F\}, \{A, B, C\}\}$ and 2) Model 2, a RGLL model with the same generating class.

In terms of REGRAPH, we should include in one input file: a) the number of variables, $|V| = 6$, b) the total number of categories for each variable as real values with

two decimals, $(|\mathrm{I}_\delta|)_{\delta\in\Delta}$= (2.00, 2.00, 2.00, 2.00, 2.00, 2.00, 2.00), and c) all columns in the data presented in table 3.2 before the column (including it) corresponding to the initial values for the algorithm, $m_0(i)$. Additionally, as all these programs work with numeric labels instead of with characters, we have that the variable associated to the first column, F, corresponds to variable 1, so that we label it as 1, the variable associated to the next column corresponds to variable 2, so that we assign to it the label 2, and so on. As a consequence, the generating class for this model is $\{\{6,3,2\}, \{6,4,2\}, \{5,1\}, \{6,5,4\}\}$. The input file for the generating class should include: a) the number of elements in the generating class, $|A| = 4$, and b) in different rows we should add each element conforming the generating class written according to the labeling.

The estimated expected frequencies under Model 1 using the program in REGRAPH that fits RGLL models, considering that each vertex and edge colour class is atomic, that is a graphical log-linear model, is presented in the ninth column of table 3.2. The deviance for this model is 51.36 with 46 degrees of freedom and a p-value of 0.27, so that in general, for example for a significance level of 0.05, we do not reject the null hypothesis that this model fits the data well. The Pearson X^2 statistic is 51.10. If we use *Splus* to fit the model, we get exactly the same expected estimated frequencies, deviance, and degrees of freedom as the corresponding ones computed by using REGRAPH. If we do all calculations using *MIM*, we get also a deviance of 51.36 with the same degrees of freedom. *MIM* gets the estimated frequencies using a methodology based on the fact that this is a decomposable model so that it gets the estimators by using closed formulas.

If we use the program in REGRAPH that selects a model, for both vertex and edge classes, that fits as good as possible to some data in the sense that the deviance and tests based on it are used in the selection process to see if it is convenient to join classes to obtain a new RGLL model, considering besides that the initial model is the one in which all vertices and edges are in different atomic classes, that we use the default convergence options, a significance level of 0.05, and the same generating class as in Model 1, then we get a new model, Model 2, with a deviance of 30.40 with 46 degrees of freedom, a p-value of 0.96, and Pearson X^2 statistic of 26.55. This means that Model 2 has a much better fit to the data than Model 1. The estimated expected frequencies are shown in the tenth column of table 3.2. The eleventh and twelfth columns of this table correspond to the deviance contribution of each cell for both models, *i.e.* twice the observed count multiplied by the logarithm of the observed count divided by the expected frequencies under each model, $2n(i)\log(n(i)/\widehat{m}_k(i))$. Obviously, the smaller these terms are, the better fit to the data. In this case, 39 cells fitted under Model 2 have a less deviance contribution than the corresponding cells under Model 1, in 16 cells we have the opposite, and with the remaining 9 the contribution is almost the same for both models.

Model 2 has three vertex colour classes and 13 edge colour classes as follows:

$$V_1 = \{1, 2, 5, 6\},\ V_2 = \{3\},\ V_3 = \{4\}.$$

$$E_1 = \{u_{36}(11)\},\ E_2 = \{u_{36}(12)\},\ E_3 = \{u_{36}(21)\},\ E_4 = \{u_{24}(21)\}, E_5 = \{u_{46}(12)\},$$
$$E_6 = \{u_{15}(12)\},\ E_7 = \{u_{15}(21)\},\ E_8 = \{u_{26}(22), u_{46}(21), u_{24}(22), u_{15}(22)\},$$
$$E_9 = \{u_{36}(22), u_{23}(12), u_{46}(11), u_{15}(11)\},\ E_{10} = \{u_{26}(21), u_{26}(11), u_{23}(22), u_{23}(21), u_{56}(22)\},$$
$$E_{11} = \{u_{24}(11), u_{26}(12), u_{23}(11), u_{56}(21)\},\ E_{12} = \{u_{45}(12), u_{56}(11), u_{45}(11), u_{45}(21), u_{56}(12)\},$$
$$E_{13} = \{u_{46}(22), u_{24}(12), u_{45}(22)\}.$$

The graph for this model obtained using *Graphviz* is shown in figure 3.9, this graph was organized and reordered in such a way that it matches up with the one shown by Edwards (2000, p. 26) when he fits a usual graphical log-linear model.

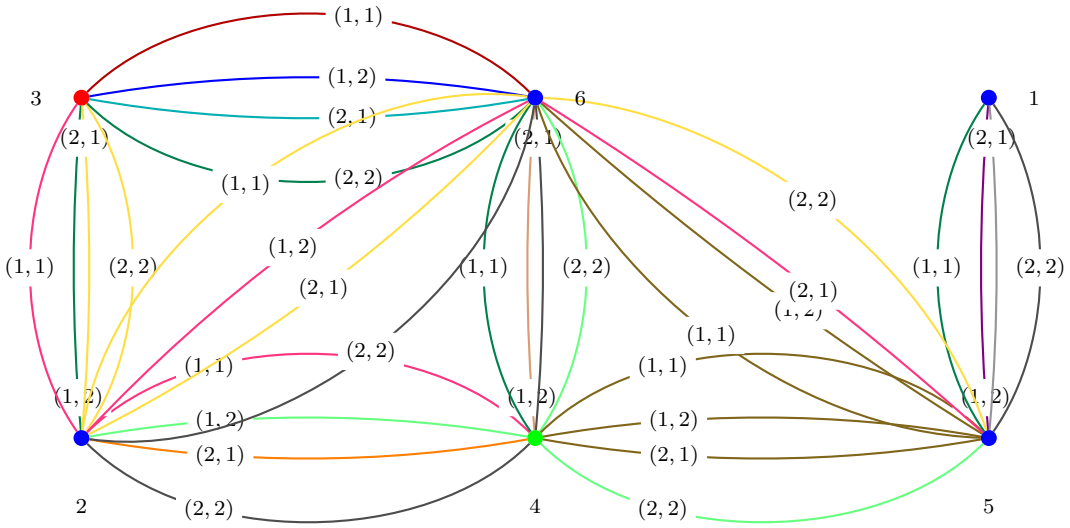


Figure 3.9: Vertex and edge colouring for a RGLL model generated by $\{\{2, 3, 6\}, \{2, 4, 6\}, \{1, 5\}, \{4, 5, 6\}\}$, *with vertex set* $V = (V_1, V_2, V_3)$, $V_1 = \{1, 2, 5, 6\}$, $V_2 = \{3\}$, *and* $V_3 = \{4\}$, *and first-order interactions set* $E = (E_1, E_2, \ldots, E_{13})$, $E_1 = \{u_{36}(11)\}$, $E_2 = \{u_{36}(12)\}$, $E_3 = \{u_{36}(21)\}$, $E_4 = \{u_{24}(21)\}$, $E_5 = \{u_{46}(12)\}$, $E_6 = \{u_{15}(12)\}$, $E_7 = \{u_{15}(21)\}$, $E_8 = \{u_{26}(22), u_{46}(21), u_{24}(22), u_{15}(22)\}$, $E_9 = \{u_{36}(22), u_{23}(12), u_{46}(11), u_{15}(11)\}$, $E_{10} = \{u_{26}(21), u_{26}(11), u_{23}(22), u_{23}(21), u_{56}(22)\}$, $E_{11} = \{u_{24}(11), u_{26}(12), u_{23}(11), u_{56}(21)\}$, $E_{12} = \{u_{45}(12), u_{56}(11), u_{45}(11), u_{45}(21), u_{56}(12)\}$, $E_{13} = \{u_{46}(22), u_{24}(12), u_{45}(22)\}$.

The second example corresponds to RGLL models for the data on change of region of residence, table 2.3, presented in Chapter 2, section **2.5**. In this case, the best fit

was achieved by using a quasi-symmetry model. As we saw before in section **3.7**, these models are a particular case of RGLL models. In terms of REGRAPH, we label the vertex associated to region of residence in 1980 as 1 and the one associated to region of residence in 1985 as 2. Similarly, the levels corresponding to regions Northeast, Midwest, South, and West are labeled 1, 2, 3, and 4, respectively. As in **3.7**, the quasi-symmetry model is associated to a graph with two vertex colour classes and ten edge colour classes as follows:

$$V = (V_1, V_2);\ V_1 = \{1\},\ V_2 = \{2\}.$$

$$E = (E_1, ..., E_{10});\ E_i = \{u_{12}(ii)\}, i = 1, .., 4,$$
$$E_5 = \{u_{12}(12), u_{12}(21)\},\ E_6 = \{u_{12}(13), u_{12}(31)\},\ E_7 = \{u_{12}(14), u_{12}(41)\}$$
$$E_8 = \{u_{12}(23), u_{12}(32)\},\ E_9 = \{u_{12}(24), u_{12}(42)\},\ E_{10} = \{u_{12}(34), u_{12}(43)\}.$$

The graph corresponding to this model is presented in figure 3.10.

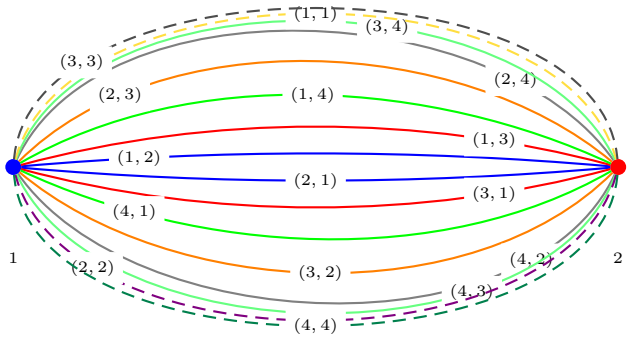


Figure 3.10: Quasi-symmetry model presented as a RGLL model with generating class {1,2}*, vertex set* $V = (V_1, V_2)$*,* V_1={1} *and* V_2={2}*, and with first-order interaction set* E=(E_1,E_2,...,E_{10})*,* E_i=$\{u_{12}(ii)\}$*,* $i = 1, .., 4$*,* E_5=$\{u_{12}(12), u_{12}(21)\}$*,* E_6=$\{u_{12}(13), u_{12}(31)\}$*,* E_7= $\{u_{12}(14), u_{12}(41)\}$*,* E_8=$\{u_{12}(23), u_{12}(32)\}$*,* E_9=$\{u_{12}(24), u_{12}(42)\}$*,* E_{10}=$\{u_{12}(34), u_{12}(43)\}$*.*

The estimated expected frequencies under this model using the program in REGRAPH that fits RGLL models (table 3.3) are almost the same obtained in table 2.4 shown in Chapter 2, section **2.5**, where *SPlus* was used to fit the model. The few cell counts that change, differ only by a hundredth. The deviance is the same value got before, 2.99, with 3 degrees of freedom with a p-value of 0.39 and Pearson statistic value of 2.98. This means that in general we do not reject the null hypothesis that the model fits the data.

The input files to fit the previous model are two, a file containing: a) the number of variables, $|V| = 2$, b) the number of levels for each variable as real values with two

Residence in 1980	**Residence in 1985**			
	Northeast	Midwest	South	West
Northeast	11607.00 (11607)	95.78 (100)	370.42 (366)	123.77 (124)
Midwest	91.22 (87)	13677.00 (13677)	501.69 (515)	311.11 (302)
South	167.57 (172)	238.31 (225)	17819.00 (17819)	261.12 (270)
West	63.23 (63)	166.89 (176)	294.88 (286)	10192.00 (10192)

Table 3.3: Estimated expected frequencies under the RGLL model with generating class {1, 2}*, with vertex set* $V = (V_1, V_2)$*,* V_1={1} *and* V_2={2}*, and with first-order interaction set* E=(E_1,E_2,...,E_{10})*,* E_i=$\{u_{12}(ii)\}$*,* $i = 1, .., 4$*,* E_5=$\{u_{12}(12), u_{12}(21)\}$*,* E_6=$\{u_{12}(13), u_{12}(31)\}$*,* E_7= $\{u_{12}(14), u_{12}(41)\}$*,* E_8=$\{u_{12}(23), u_{12}(32)\}$*,* E_9=$\{u_{12}(24), u_{12}(42)\}$*,* E_{10}=$\{u_{12}(34), u_{12}(43)\}$*, and observed values in parenthesis.*

decimals, $(|\mathrm{I}_\delta|)_{\delta\in\Delta}$= (4.00, 4.00, 4.00, 4.00), and c) the contingency table as real values with two decimals, in this case we introduce all cells i, first the values related to the variable associated with vertex 1 and then those related to the variable associated to vertex 2, followed by the observed count $n(i)$ and initial values $m_0(i) = 1$. We also need an input file for the generating class, this file contains: a) the number of elements in the generating class, $|A| = 1$, which represents in this case that the generating class has only one element and b) the elements of this class, 1 and 2 or $\{1, 2\}$.

Starting from the quasi-symmetry model, we want to select a RGLL model using the selection by only edges method that fits the data. Using the corresponding program in REGRAPH, the convergence default options, and a significance level of 0.05 for all tests that allow joining colour classes, we get a model whose fit is shown in table 3.4. This model has a deviance of 2.98 with 3 degrees of freedom and a p-value of 0.39, which indicates that we do not reject the null hypothesis that the model fits the data. The value of the X^2 statistic is 2.98.

Obviously, this model has two vertex colour classes, because these were the ones included in the initial model, and we got through the selection process seven edge colour classes. Of these classes, there are only two that differ with respect to the initial model. The class that contained $u_{12}(34)$ and $u_{12}(43)$ now also contains the elements $u_{12}(11)$ and $u_{12}(33)$, and the class that contained $u_{12}(24)$ and $u_{12}(42)$ now also contains $u_{12}(44)$ (figure 3.11); *i.e.* $E_1 \cup E_3 \cup E_{10}$ in the quasi-symmetry model is equal to E_6 in the RGLL model and $E_4 \cup E_9$ in the quasi-symmetry model is equal to E_7 in the RGLL model.

Residence in 1980	**Residence in 1985**			
	Northeast	Midwest	South	West
Northeast	11606.99 (11607)	95.79 (100)	370.44 (366)	123.77 (124)
Midwest	91.21 (87)	13677.00 (13677)	501.67 (515)	311.14 (302)
South	167.56 (172)	238.33 (225)	17818.94 (17819)	261.16 (270)
West	63.23 (63)	166.92 (176)	294.92 (286)	10191.93 (10192)

Table 3.4: Estimated expected frequencies under the RGLL model generated by {1, 2}*, with edge set* E=(E_1,E_2,...,E_7)*,* E_1= {$u_{12}(22)$}*,* E_2= {$u_{12}(12)$, $u_{12}(21)$}*,* E_3={$u_{12}(13)$, $u_{12}(31)$}*,* E_4= {$u_{12}(14)$, $u_{12}(41)$}*,* E_5= {$u_{12}(23)$, $u_{12}(32)$}*,* E_6= {$u_{12}(33)$, $u_{12}(11)$, $u_{12}(34)$, $u_{12}(43)$}*,* E_7= {$u_{12}(44)$, $u_{12}(24)$, $u_{12}(42)$} *and vertex set* $V = (V_1, V_2)$*,* V_1={1}*,* V_2={2} *compared with the observed values, in parenthesis.*

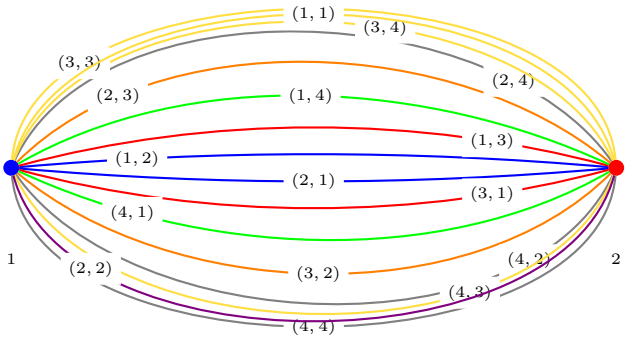


Figure 3.11: RGLL model generated by {1, 2}*, with first interaction order set* E = (E_1, E_2, ..., E_7)*,* E_1= {$u_{12}(22)$}*,* E_2={$u_{12}(12)$, $u_{12}(21)$}*,* E_3={$u_{12}(13)$, $u_{12}(31)$}*,* E_4={$u_{12}(14)$, $u_{12}(41)$}*,* E_5={$u_{12}(23)$, $u_{12}(32)$}*,* E_6= {$u_{12}(33)$, $u_{12}(11)$, $u_{12}(34)$, $u_{12}(43)$}*,* E_7= {$u_{12}(44)$, $u_{12}(24)$, $u_{12}(42)$} *and vertex set* $V = (V_1, V_2)$*,* V_1={1}*,* V_2={2}*.*

A third and last example corresponds to data similar to the one given in the first example, but where the data have not been published before. We analyzed it in Ramírez-Aldana (2005) using Bayesian networks. The data correspond to approximately 50 variables measured for each of 861 patients who were admitted to an Intensive Care Unit (ICU) in two main hospitals in Mexico City: The "Siglo XXI" National Medical Center and the "La Raza" Medical Center between 2002 and 2004. In Ramírez-Aldana (2005), the quality of life was analyzed by using models that according to medical experience and selection methods based on logistic regressions were adequate for the data. The models included the following variables: initial and posterior quality of life (*iql* and *pql*, respectively), both with categories coded as 1 and 2 for good and bad, respectively; emergency surgery (*qxurgent*) with categories coded as 1 and 2 for No and Yes, respectively; and age (*age*) with categories coded as 1 and 2 for less or equal than 60 years and 61 and more, respectively. We also considered three failures: respiratory (*resf*), neurological (*neuf*), and cardiac (*cardf*), with categories coded as 1 and 2 for No and Yes, respectively.

In this work, we consider three possible sets of models that include the previous variables, but where not all failures are included at the same time: 1) A first set of models including only *cardf*, 2) A second set of models including only *resf*, 3) A third set of models including only *neuf*. We present a summary of the statistics and information about the three sets of models in table 3.5. We present below with more detail one of the RGLL models that includes *resf* because it fitted better than the others according to the deviance; however, for the other models the process is analogous.

Observe that in this example we consider variables with the same number of categories, but the categories are not the same, which was an assumption considered for RGLL models. However, if we do not consider symmetry interpretations, we still can fit the models as RGLL, and what we get is a more parsimonious log-linear model than the unrestricted graphical log-linear model.

For 1), models containing *iql*, *pql*, *cardf*, *qxurgent*, and *age*, using the forward, backward, and bidirectional selection methods available in *MIM*, we obtain the graphical log-linear model with generating class $\{\{qxurgent, pql\}, \{age, pql, iql\}, \{age, cardf\}\}$ and a p-value of 0.11. Considering the same generating class, we selected using REGRAPH a RGLL model with one vertex class and eight edge classes and a p-value of 0.66.

For 2), models containing *iql*, *pql*, *resf*, *qxurgent*, and *age*, using all selection methods in *MIM*, we obtained the generating class: $\{\{qxurgent, resf, pql\}, \{age, pql, iql\}\}$. All interactions in the generating class are of order higher than one, which implies that many colourings can fit the data in the same way because the likelihood equations corresponding to the interactions of order higher than two automatically imply the likelihood

	Models	deviance	p-value	X^2	d.f.	Vertex classes	Edge classes
1)	Graphical log-linear model generated by {{4,2}, {5, 2, 1}, {5, 6}}	28.06	0.11	25.62	20	5	20
	RGLL model generated by {{4,2}, {5, 2, 1}, {5, 6}}	16.92	0.66	15.64	20	1	8
2)	Graphical log-linear model generated by {{4, 3, 2}, {5, 2, 1}}	18.41	0.43	18.54	18	5	24
	RGLL model generated by {{4, 3, 2}, {5, 2, 1}}	18.41	0.43	18.54	18	1	7
	Graphical log-linear model generated by {{4, 3}, {3, 2}, {5, 2, 1}}	29.29	0.08	27.96	20	5	20
	RGLL model generated by {{4, 3}, {3, 2}, {5, 2, 1}}	**11.81**	**0.92**	**9.30**	**20**	**2**	**4**
3)	Graphical log-linear model generated by {{4, 2}, {5, 2, 1}, {7, 2}}	19.70	0.48	17.27	20	5	20
	RGLL model generated by {{4, 2}, {5, 2, 1}, {7, 2}}	12.42	0.90	9.58	20	2	5

Table 3.5: Values for the deviance and X^2 statistics, p-values, degrees of freedom, and number of vertex and edge colour classes associated to three possible sets of RGLL models for the Intensive Care Unit data: 1) models including cardf, 2) models including resf, and 3) models including neuf, where 1 = iql, 2 = pql, 3 = resf, 4 = qxurgent, 5 = age, 6 = cardf, and 7 = neuf.

equations for various colourings. As a consequence, the deviance, degrees of freedom, and fitted expected frequencies for the RGLL model obtained using REGRAPH and the graphical log-linear model generated by {{*qxurgent*, *resf*, *pql*}, {*age*, *pql*, *iql*}} are the same, which means that there are no real benefits from fitting a RGLL model.

Using the generating class {{*qxurgent*, *resf*, *pql*}, {*age*, *pql*, *iql*}} in 2), we analyzed with *MIM* if an edge could be removed and still obtain a good graphical log-linear model. We deleted the edge with largest p-value, {*qxurgent*, *pql*}, obtaining the model generated by {{*qxurgent*, *resf*}, {*resf*, *pql*}, {*age*, *pql*, *iql*}} with a p-value of 0.08. Considering this generating class, we obtained using REGRAPH a RGLL model with a much better fit to the data. This RGLL model contains two vertex and four edge colour classes and a p-value of 0.92. We discuss further this model in the following paragraphs.

In terms of REGRAPH there is one input file including: a) the number of variables, $|V| = 5$, b) the total number of categories for each variable, $(|\mathrm{I}_\delta|)_{\delta\in\Delta}$= (2, 2, 2, 2, 2) as real values with two decimals, and c) all columns in the data presented in table 3.6 before the column (including it) corresponding to the initial values for the algorithm, $m_0(i)$, considering all values as real numbers with two decimals. Additionally, the variable associated to the first column, *iql*, is labeled as 1, the variable associated to the next column, *pql* is labeled as 2, and so on. As a consequence, the generating class for this model is $\{\{4, 3\}, \{3, 2\}, \{5, 2, 1\}\}$. The input file for the generating class should include: a) the number of elements in the generating class, $|A| = 3$, and b) in different rows we add each element conforming the generating class written according to the labeling.

The estimated expected frequencies under the graphical log-linear model generated by $\{\{qxurgent, resf\}, \{resf, pql\}, \{age, pql, iql\}\}$, Model A, obtained using REGRAPH considering that each vertex and edge colour class is atomic, is presented in the eighth column of table 3.6. The deviance for this model is 29.29 with 20 degrees of freedom and a p-value of 0.08 as mentioned before, so that for a significance level of 0.05, we do not reject the null hypothesis that this model fits the data. The value of X^2 is 27.96.

If we use the program in REGRAPH that selects a model, for both vertex and edge colour classes, considering as initial model the one in which all vertices and edges are in different atomic classes and using the default convergence options, a significance level of 0.05, and the same generating class $\{\{qxurgent, resf\}, \{resf, pql\}, \{age, pql, iql\}\}$, then we get a new RGLL model, Model B. This model has a deviance of 11.81 with 20 degrees of freedom, a p-value of 0.92, and Pearson X^2 statistic of 9.30. Model B has a much better fit to the data than Model A. The estimated expected frequencies are shown in the ninth column of table 3.6. The tenth and eleventh columns of this table correspond to the deviance contribution of each cell for both models, $2n(i)\log(n(i)/\widehat{m}_k(i))$. In this case, 25 cells fitted under Model B have a less deviance contribution than the corresponding cells under Model A, in 5 cells it is the opposite, and with the remaining 2 the contribution is the same for both models.

Model B has two vertex colour classes and 4 edge colour classes as follows:

$$V_1 = \{3\},\ V_2 = \{1, 2, 4, 5\}.$$
$$E_1 = \{u_{34}(12)\},\ E_2 = \{u_{23}(11)\},$$
$$E_3 = \{u_{25}(11), u_{12}(21), u_{12}(12), u_{23}(21), u_{34}(21), u_{34}(11), u_{25}(22)\},$$
$$E_4 = \{u_{15}(21), u_{25}(12), u_{15}(22), u_{12}(11), u_{34}(22), u_{15}(12), u_{15}(11),$$
$$u_{12}(22), u_{23}(22), u_{23}(12), u_{25}(21)\}.$$

iql	*pql*	*resf*	*qxurgent*	*age*	Obs. count	Initial val.	Exp. freq. mod. A	Exp. freq. mod. B	Dev. cont. mod. A	Dev. cont. mod. B
1	1	1	1	1	52.00	1.00	48.77	53.08	6.67	-2.14
1	1	1	1	2	15.00	1.00	12.96	13.43	4.39	3.32
1	1	1	2	1	6.00	1.00	8.29	4.95	-3.88	2.31
1	1	1	2	2	0.00	1.00	2.20	1.64	0.00	0.00
1	1	2	1	1	66.00	1.00	55.27	66.61	23.42	-1.21
1	1	2	1	2	14.00	1.00	14.69	16.85	-1.35	-5.19
1	1	2	2	1	19.00	1.00	30.67	18.36	-18.20	1.30
1	1	2	2	2	9.00	1.00	8.15	6.08	1.79	7.06
1	2	1	1	1	10.00	1.00	11.57	9.42	-2.92	1.20
1	2	1	1	2	7.00	1.00	8.64	7.03	-2.95	-0.06
1	2	1	2	1	4.00	1.00	1.97	3.18	5.67	1.84
1	2	1	2	2	1.00	1.00	1.47	2.38	-0.77	-1.73
1	2	2	1	1	29.00	1.00	34.38	27.20	-9.87	3.72
1	2	2	1	2	21.00	1.00	25.66	20.30	-8.42	1.42
1	2	2	2	1	24.00	1.00	19.08	27.20	11.01	-6.01
1	2	2	2	2	21.00	1.00	14.24	20.30	16.32	1.42
2	1	1	1	1	1.00	1.00	1.36	1.04	-0.61	-0.08
2	1	1	1	2	3.00	1.00	4.43	3.38	-2.34	-0.72
2	1	1	2	1	1.00	1.00	0.23	0.35	2.94	2.10
2	1	1	2	2	1.00	1.00	0.75	1.14	0.58	-0.26
2	1	2	1	1	2.00	1.00	1.55	1.30	1.02	1.72
2	1	2	1	2	5.00	1.00	5.02	4.24	-0.04	1.65
2	1	2	2	1	0.00	1.00	0.86	1.30	0.00	0.00
2	1	2	2	2	4.00	1.00	2.79	4.24	2.88	-0.47
2	2	1	1	1	6.00	1.00	5.01	5.50	2.16	1.04
2	2	1	1	2	6.00	1.00	7.26	7.13	-2.29	-2.07
2	2	1	2	1	2.00	1.00	0.85	1.86	3.42	0.29
2	2	1	2	2	2.00	1.00	1.23	1.50	1.94	1.15
2	2	2	1	1	16.00	1.00	14.88	15.89	2.32	0.22
2	2	2	1	2	20.00	1.00	21.55	20.58	-2.99	-1.14
2	2	2	2	1	5.00	1.00	8.26	5.74	-5.02	-1.38
2	2	2	2	2	14.00	1.00	11.96	12.79	4.41	2.53

Table 3.6: Observed counts, fitted values, and deviance contribution for a study of 861 patients admitted to an Intensive Care Unit under two models: a) Model A, graphical log-linear model with generating class {{qxurgent, resf}, {resf, pql}, {age, pql, iql}} and 2) Model B, a RGLL model with the same generating class.

The corresponding graph is shown in figure 3.12.

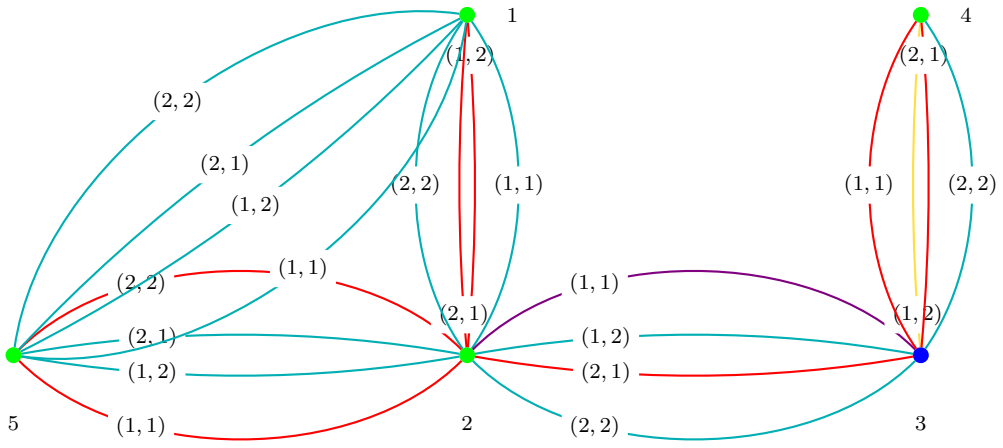


Figure 3.12: RGLL model generated by $\{\{4,3\}, \{3,2\}, \{5,2,1\}\}$*, with vertex set* $V = (V_1, V_2)$*,* $V_1 = \{3\}$ *and* $V_2 = \{1, 2, 4, 5\}$*, and first-order interactions set* $E = (E_1, E_2, E_3, E_4)$*,* $E_1 = \{u_{34}(12)\}$*,* $E_2 = \{u_{23}(11)\}$*,* $E_3 = \{u_{25}(11), u_{12}(21), u_{12}(12), u_{23}(21), u_{34}(21), u_{34}(11), u_{25}(22)\}$*,* $E_4 = \{u_{15}(21), u_{25}(12), u_{15}(22), u_{12}(11), u_{34}(22), u_{15}(12), u_{15}(11), u_{12}(22), u_{23}(22), u_{23}(12), u_{25}(21)\}$*.*

For 3), models containing *iql*, *pql*, *neuf*, *qxurgent*, and *age*, we obtained using any of the selection methods available in *MIM* a graphical log-linear model with generating class $\{\{qxurgent, pql\}, \{age, pql, iql\}, \{age, neuf, pql\}\}$. In this case, the model is decomposable, there are 6 sampling zeros in the corresponding table and REGRAPH warns us that some marginal counts defining the likelihood equations are zero, which implies that the maximum likelihood estimators do not exist as we discuss in chapter 5. Then, we should not use this model. We deleted the edge with largest p-value, $\{age, neuf\}$, obtaining the model generated by $\{\{qxurgent, pql\}, \{age, pql, iql\}, \{neuf, pql\}\}$ with a p-value of 0.48. From this model, we selected using REGRAPH a RGLL model with two vertex and five edge classes and a p-value of 0.90.

There is a connection between log-linear and logit models, e.g., Bishop *et al.* (1975, section 8.5), in the following paragraphs we identify the logit model corresponding to a RGLL model and we identify how the equality restrictions in the latter translate to new restrictions in the former.

In these data, the posterior quality of life, *pql*, can be considered as a dependent

or response variable and the other variables can be seen as explanatory variables. The RGLL model denoted as B, the model with generating class $\{\{qxurgent, resf\}, \{resf, pql\}, \{age, pql, iql\}\}$, was the RGLL model involving the dependent variable with better fit to the data according to the deviance and using the selection method given in this work. However, any RGLL model (in fact, any log-linear model) treats all variables symmetrically, focusing on association between variables instead of describing how a single response variable depends on the other variables.

Define the probability that the posterior quality of life is bad given particular values for the other variables, p, as

$$p = P(pql = 2|iql = x, age = y, resf = z, qxurgent = v),$$

where x,y,z,v= 1, 2, and write the corresponding expected frequencies when $pql = 1$ and when $pql = 2$ as $m_{pql\ iql\ age\ resf\ qxurgent}(1, x, y, z, v)$ and $m_{pql\ iql\ age\ resf\ qxurgent}(2, x, y, z, v)$, respectively. One can obtain the associated logit model as follows

$$\log\left(\frac{m_{pql\ iql\ age\ resf\ qxurgent}(2, x, y, z, v)}{m_{pql\ iql\ age\ resf\ qxurgent}(1, x, y, z, v)}\right) = \log\left(\frac{p}{1-p}\right) =$$
$$\beta + \beta_{resf}(z) + \beta_{age}(y) + \beta_{iql}(x) + \beta_{age\ iql}(yx),$$

where, using the same labeling we have used throughout this example, we have

$$\beta = u_2(2) - u_2(1),\ \beta_{resf}(z) = u_{23}(2z) - u_{23}(1z),\ \beta_{age}(y) = u_{25}(2y) - u_{25}(1y),$$

$$\beta_{iql}(x) = u_{12}(x2) - u_{12}(x1),\ \beta_{age\ iql}(yx) = u_{521}(y2x) - u_{521}(y1x).$$

The log-linear model treats the 32 cells as independent Poisson random variables. The logit model treats the corresponding counts, those without considering the response variable, as 16 independent binomial random variables.

For a general case, the method to obtain the associated logit model from a RGLL model corresponds to: 1) eliminate any interactions not involving the response variable, 2) transform all first-order interactions that include a certain variable Z and the dependent variable, in this example pql, into main effects for Z, and 3) transform all interactions of order t, $t > 2$, including a set of variables and the response variable into interactions of order $t - 1$ formed by that set of variables.

The constant term and all main effects in the RGLL models are transformed into the constant term in the logistic regression. Then, a vertex colouring in a RGLL model does not become a parameter constraint in the logistic regression, in fact it disappears. The only equality restrictions in the RGLL model that remain are those corresponding to first-order interactions, *i.e.* edge colourings, which become restrictions for the main

effects in the logistic regression. In this example, the parameter restrictions given by the edge colouring (E_1, E_2, E_3, E_4) seen above for the RGLL model B imply the following parameter constraints in the corresponding logistic regression:

$$\beta_{resf}(1) = u_{E_3} - u_{E_2},\ \beta_{resf}(2) = 0;$$
$$\beta_{iql}(1) = u_{E_3} - u_{E_4},\ \beta_{iql}(2) = u_{E_4} - u_{E_3};$$
$$\beta_{age}(1) = u_{E_4} - u_{E_3},\ \beta_{age}(2) = u_{E_3} - u_{E_4}.$$

One of these parameter restrictions that can be interpreted is $\beta_{resf}(2) = 0$. It means that when there is respiratory failure, the corresponding logit or the odd, depends only on the age, initial quality of life, and their interaction. Observe that the restriction $u_{23}(12) = u_{23}(22)$ (where remember that pql is labeled as 2 and $resf$ as 3) obtained from the RGLL model states something similar: when there is respiratory failure the association with posterior quality of life is the same for any value of the posterior quality of life, so that any differences in the expected frequencies are determined by the other variables and their associations.

Clearly, the fitted RGLL models are more parsimonious than the unrestricted models and also provide a better fit to the data according to goodness of fit tests based on the deviance, in Model 1 for example 1, a usual graphical log-linear model, we have 68 parameters: 12 for the main effects, 32 for the first-order interactions, and 24 for the second-order interactions; whereas in Model 2, a RGLL model, we have 43 parameters in total: 6 for the vertex colouring, 13 for the edge colouring, and 24 for the second-order interactions. On the other hand, under a saturated model for two variables with 4 levels each one as in example 2 we have 24 parameters: 8 for the main effects and 16 for the first-order interactions; whereas in the corresponding quasi-symmetry model, which is also a RGLL model, we have 18 parameters: 8 corresponding to the main effects and 10 to the first-order interactions; and in the RGLL model obtained from the quasi-symmetry model, we have 15 parameters: 8 for the vertex colouring and 7 for the edge colouring. Finally, in Model A for example 3, we have 38 parameters: 10 for the main effects, 20 for the first-order interactions, and 8 for the second-order interactions; whereas in the corresponding RGLL model, Model B, there are 16 parameters: 4 for the vertex colouring, 4 for the edge colouring, and 8 for the second-order interactions. The number of parameters for all models in this section discussed in this paragraph are shown in table 3.7

Additionally, we observe that using RGLL models we can say something about which association between pairs of variables is similar according to the colourings, and sometimes we can say something about the data, for example, the model in figure 3.11 indicates, according to the edge colour classes, which values for the cells would be the same considering only the interactions even though in reality the values differ because of the difference in the marginal values between levels as in quasi-symmetry models. In

Model	Main effects	1st order int.	2nd order int.	Total
Graphical log-linear model heart disease data (Model 1)	12	32	24	68
RGLL model heart disease data (Model 2)	6	13	24	43
Saturated model change of residence data	8	16	0	24
Quasi-symmetry (RGLL model) change of residence data	8	10	0	18
RGLL model change of residence data	8	7	0	15
Graphical log-linear model Intensive Care data (Model A)	10	20	8	38
RGLL model Intensive Care data (Model B)	4	4	8	16

Table 3.7: Number of parameters for all the models discussed in section 3.11.5.

this case we have also presence of certain symmetry inherited from the quasi-symmetry model, which is that the cells above the diagonal are similar to the ones under the diagonal, again they are not equal because of the difference between the marginal values. However, the interpretation of a RGLL models in terms of some kind of symmetry between the cells of the table or the variables is not always clear, it depends on the parameters restrictions we choose and if according to those restrictions the expected frequencies for some cells are equal or the model is preserved, such interpretation is even less clear when we introduce interactions of order greater than one.

In the following chapter, we introduce particular classes of RGLL models with interpretation in terms of symmetry in two senses: first in the sense that we can change some levels associated to some variables according to a permutation set and after that we ask that the estimated frequencies are preserved, and second in the sense that we can interchange vertices in the graph according to sets of permutations and also permute the levels for the variables and still preserve the same model. These last kind of models are equivalent to the RCOP models introduced by Højsgaard and Lauritzen (2008) for the continuous case, and in fact concepts like vertex and edge orbits, graph automorphisms, permutation symmetry, and scale invariance used for these models are also used in the discrete case.

Chapter 4

Label and level invariant models

In this chapter, we consider two particular cases of RGLL models: *label and level invariant models*, whose importance is given by their symmetry interpretation. These models can be considered as a type of a symmetry model generalization for discrete data in the following sense. Symmetry for two-way contingency tables, which is a model where $m(i,j) = m(j,i)$, for all $i \neq j$ (see section **2.5.2**), is a particular case of these models, additionally they can be used for any number of variables so that there is not a constraint of only having two variables as in symmetry for two-way contingency tables or in the two-dimensional generalized symmetry models defined by Goodman (1985); however, label and level invariant models are only defined for graphical models with a triangle-free structure where there are no interaction effects of order higher than one.

To motivate label and level invariant models, we present one instance which we further discuss in example 4.13. We use the data presented and analyzed by Drton and Richardson (2008) based on a study from Kendler *et al.* (1992). The data, presented in table 4.1, correspond to 597 observations for female monozygotic twins, indicating whether or not each twin is alcohol dependent, variables A_i, for i=1, 2; and whether or not the twins suffer from major depression, variables D_i, for i=1, 2. The values taken by each variable are 0 and 1, corresponding to *no* and *yes*, respectively.

Symmetry between twins corresponds to a model unaltered in spite of permuting the variables for each twin, A_1 with A_2 and D_1 with D_2, and the category order. In particular, the distribution is preserved after such permutation. A model with these properties is the one with generating class $\{\{A_1, A_2\}, \{A_2, D_2\}, \{D_2, D_1\}, \{D_1, A_1\}\}$ and graph given in figure 4.1. In this model A_1 and D_2 are conditionally independent given D_1 and A_2, $A_1 \perp D_2 | D_1, A_2$, and $A_2 \perp D_1 | A_1, D_2$. As we shall see later, it is a *label invariant model* and is a particular case of a RGLL model.

After interchanging vertices according to a set of permutations of the vertices, label invariant models are defined to preserve their associated graph, so that we can say that

A_1	D_1	A_2			
		0		1	
		D_2			
		0	1	0	1
0	0	288	80	15	9
	1	92	51	7	10
1	0	8	4	3	2
	1	8	9	4	7

Table 4.1: Alcohol dependence and major depression for 597 pairs of female twins.

there is symmetry in a graphical sense, and they also preserve the distribution so that the expected frequencies after such permutations remain the same, which can as well be considered as some kind of symmetry. Additionally, they preserve the conditional and marginal independences in the graph and are scale invariant as discussed below.

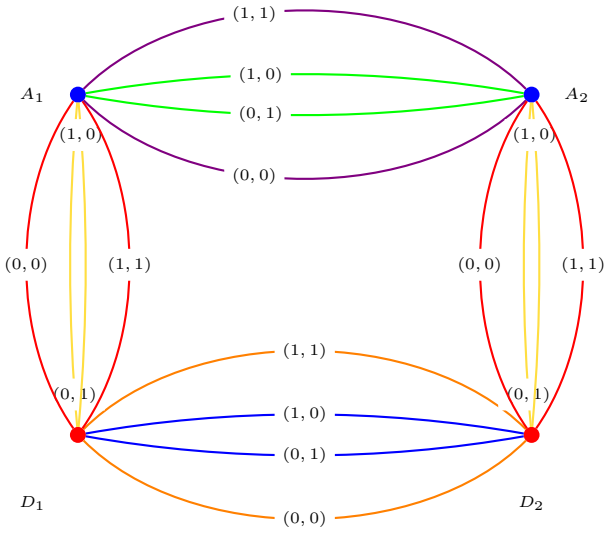


Figure 4.1: A RGLL model with generating class $\{\{A_1, A_2\}, \{A_2, D_2\}, \{D_2, D_1\}, \{D_1, A_1\}\}$, *vertex set* $V = (V_1, V_2)$, *where* $V_1 = \{A_1, A_2\}$ *and* $V_2 = \{D_2, D_1\}$, *and first-order interactions set* $E = (E_1, E_2, E_3, E_4, E_5, E_6)$, *where* $E_1 = \{u_{A_1D_1}(00),\ u_{A_1D_1}(11),\ u_{A_2D_2}(00),\ u_{A_2D_2}(11)\}$, $E_2 = \{u_{A_1D_1}(01),\ u_{A_1D_1}(10),\ u_{A_2D_2}(01),\ u_{A_2D_2}(10)\}$, $E_3 = \{u_{A_1A_2}(00),\ u_{A_1A_2}(11)\}$, $E_4 = \{u_{A_1A_2}(01),\ u_{A_1A_2}(10)\}$, $E_5 = \{u_{D_2D_1}(00),\ u_{D_2D_1}(11)\}$, *and* $E_6 = \{u_{D_2D_1}(01),\ u_{D_2D_1}(10)\}$.

On the other hand, level invariant models equate the expected frequencies for certain cells according to a set of permutations, preserving at the same time the marginal and conditional independences obtained from the graph associated to the model. Other symmetry generalizations for three-way contingency tables, *e.g.*, Andersen (1991, p.

328-329) and Bishop *et al.* (1975, p. 299-309), and even the generalizations defined by Darroch and Bhapkar (1990) for any number of variables, consider equalities between expected frequencies of some cells but they do not consider the simultaneous inclusion of marginal independences and conditional independences because they are based on using a saturated model, which is equivalent to say that these generalizations consider only graphical models whose graph is complete. As a consequence, they might be considered as less *general*. Level and label invariant models consider such independences and even though they are defined only for a subclass of graphical models, those with triangle-free graphs, they could be considered as one small step in another direction to continue the study of the symmetry, a concept which as we can see is difficult to define in a general sense. Later on in section **4.3** we discuss how these models can be generalized to include all graphical models.

We consider graphical models whose highest-order interaction is one, that is models with only main effects and first-order interactions, under this consideration all the models defined in this chapter are RGLL models. Graphically, this means that we are restricting to consider triangle-free graphs. This is equivalent to graphs with clique number, the number of vertices in the largest clique of a graph, less or equal than two, or graphs with no induced 3-cycles. Some of the graphs included in this kind of graphs are: a) Trees, which obviously include graphs forming lines, which are similar to the graphs for Markov chains, b) Cycles with four or more vertices, and c) Bipartite graphs. The definition of trees, cycles, and bipartite graphs as well as figures illustrating them (figure A.2) can be found in Appendix A.

We consider also that all variables have the same categories, *i.e.* I_δ is the same for all $\delta \in V$.

For label invariant models we use the concept of *automorphism* of a graph defined for instance by Bondy and Murty (1976, p. 6-7) or Rotman (1995, p. 174) which is as follows.

Definition 4.1. An automorphism of a graph G is a bijection mapping the vertices of G back to vertices of G, *i.e.* a permutation on the vertices, such that adjacent vertices remain being adjacent.

Definition 4.1 is equivalent to say that an automorphism of a graph G is a graph isomorphism with itself, where an isomorphism between two graphs G and H is a bijection between the vertices of G and H, $\theta : V(G) \to V(H)$, such that two vertices u and v are adjacent in G if and only if $\theta(u)$ and $\theta(v)$ are adjacent in H.

The group of all automorphisms of G is denoted as $Aut(G)$ and it can be seen as a permutation group formed by all permutations that preserve the graph. Then, $Aut(G)$

is a subset of the symmetric group formed by all permutations of V, $S(V)$.

In this dissertation, we denote permutations in *cycle notation* as defined by for instance Rotman (1995, p. 3), meaning that $(i_1 i_2 ... i_r)$ denotes a permutation in which i_j is mapped to i_{j+1}, $1 \leq j \leq r-1$ and i_r is mapped to i_1. For $r = 2$, the permutation $(i_1 i_2)$ is equivalent to interchanging i_1 with i_2. By definition, under this notation it is understood that all variables not included in $(i_1 i_2 ... i_r)$ do not change after the permutation, for example if we have three variables, W, X, and Y, we write (WY) to denote a permutation that interchanges W with Y and where X is fixed. The trivial permutation or identity which is the mapping that fixes each element is denoted as Id. Using cycle notation, multiplication or composition defines a new permutation, for example if we have a set of variables $\{W, X, Y, Z\}$ and $\gamma = \alpha \circ \beta$ or $\gamma = \alpha\beta$ where $\alpha = (WX)$ and $\beta = (YZ)$, then $\gamma W = (\alpha \circ \beta)W = \alpha(\beta W) = \alpha W = X$ and similarly $\gamma X = W$, $\gamma Y = Z$, and $\gamma Z = Y$; *i.e.* γ is a permutation where W and X are interchanged and simultaneously Y and Z are also interchanged.

Example 4.1. Consider a graph G corresponding to a cycle of length four with vertex set $V = \{A, B, C, D\}$ and edge set $E = \{\{A, B\}, \{B, C\}, \{C, D\}, \{D, A\}\}$, figure 4.2. In terms of a graphical log-linear model, it corresponds to a model with generating class $\{\{A, B\}, \{B, C\}, \{C, D\}, \{D, A\}\}$.

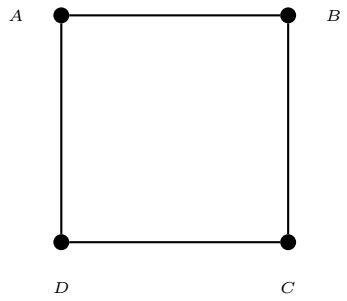


Figure 4.2: Graphical log-linear model with generating class $\{\{A, B\}, \{B, C\}, \{C, D\}, \{D, A\}\}$ *and associated graph* G, $\log m(i, j, k, l) = u + u_A(i) + u_B(j) + u_C(k) + u_D(l) + u_{AB}(ij) + u_{BC}(jk) + u_{CD}(kl) + u_{DA}(li)$.

The permutation group formed by all automorphisms of G is

$$Aut(G) = \{Id, (AC), (BD), (AC)(BD), (AB)(CD), (AD)(BC), (ABCD), (ADCB)\}.$$

Here, for example $(AC)(BD)$ represents a permutation in which we simultaneously interchange A with C and B with D. In table 4.2 we present each element in $Aut(G)$ and the graph obtained from each of its elements, for example the graph in the second

row corresponds to the graph obtained from interchanging A with C in G. Observe that to each element in $Aut(G)$ corresponds a new graph isomorphic to G, where all vertices that where adjacent in G are still adjacent after the mapping and there are not adjacent vertices after the mapping that were not adjacent before applying it.

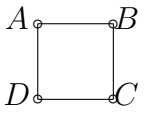
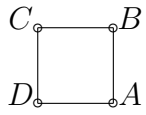
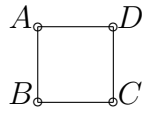
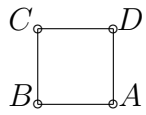
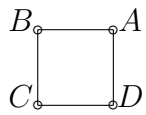
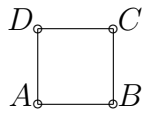
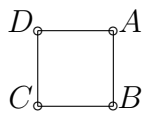
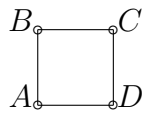
Permutation	**Graph after the mapping**
Id	
(AC)	
(BD)	
$(AC)(BD)$	
$(AB)(CD)$	
$(AD)(BC)$	
$(ABCD)$	
$(ADCB)$	

Table 4.2: Group of all automorphisms of G, $Aut(G)$, and graphs associated after each mapping, where G is the graph corresponding to a graphical log-linear model with generating class $\{\{A, B\}, \{B, C\}, \{C, D\}, \{D, A\}\}$.

Before giving the formal definition of the two kind of models we study in this chapter: label and level invariant models, we discuss further our motivation to define them as well as their general properties.

In the continuous case Perlman (1987) discusses the so-called group symmetry covariance models based on the models introduced by Eaton (1983, Sec. 9.4) and Andersson (1975). Let Y be a multivariate observation from a normal population and Γ a finite group of orthogonal matrices. The symmetry model is the family of covariance matrices $S_\Gamma^+ = \{\Sigma | \Sigma \text{ is definite positive, } g\Sigma g' = \Sigma \text{ for all } g \in \Gamma\}$, which is equivalent to say that if $Cov(Y) \in S_\Gamma^+$ then $Cov(Y) = Cov(gY)$ for all $g \in \Gamma$, that is the distribution is preserved assuming that the mean of Y is zero, hence the term symmetry model. Many covariance structures for instance sphericity, that is $\Sigma = \sigma^2 I$, $\sigma^2 > 0$, may be considered as particular cases of those models. There is no one method or approach to explicitly describe S_Γ^+ for all closed groups Γ; however, Andersson (1975) presented a general structure theory that tell us what S_Γ^+ should look like but does not tell us how to derive the particular form of S_Γ^+. Group symmetry covariance models are Gaussian models that do not consider graphical structure and consequently marginal and conditional independences derived from them; however, Andersson and Madsen (1998) introduced and Andersson *et al.* (1999) discussed *group symmetry conditional lattice independence (GS-LCI) models*, which allow both group symmetry restrictions and a specific class of conditional independences given by a distributive lattice, allowing for example covariance models where certain blocks satisfy conditional independence restrictions across blocks and group symmetry restrictions within blocks, or viceversa.

Højsgaard and Lauritzen (2008) consider more specific symmetry covariance models called RCOP in which the Gaussian model is graphical with associated graph G, additionally: a) Γ is a particular group $\Gamma \subseteq Aut(G)$, the set of all automorphisms of G, b) the concentration matrix K is used instead of Σ, and c) $K \in S_G^+$, where S_G^+ is the set of positive definite matrices with specific elements set to zero according to the graphical model. Then $K \in S_G^+ \cap S_\Gamma^+$. These are graphical Gaussian models in which the joint distribution is unaltered after permuting some elements of the vertex set V according to a permutation group that preserves the zero elements of the concentration matrix, *i.e.* it maps edges to edges.

The first kind of model we define, label invariant, is analogous to the RCOP models but defined for the discrete case. We consider a model in which a) we can change the labels (vertices or variable names) according to a set of permutations Γ that preserve the graph, *i.e.* neighbour vertices remain being neighbours, so that these permutations are automorphisms of the graph, and b) that we can change the scale of the variables according to the symmetry group of all permutations applied to the level set, *i.e.* the model is scale invariant. The graphical model is preserved after applying these permutations, so that we can change the level order (or rename them) and the labels according to Γ and we obtain the same model. In terms of contingency tables, what is done by permuting labels according to Γ is replacing some variables with others in the table and, after these changes, we should have the same expected frequencies, *i.e.* the distribution should remain. For example, for a two-way contingency table,

the only available automorphism besides Id consists in interchanging both labels, and in terms of the contingency table the corresponding label invariant model equates the expected frequencies for the elements below the diagonal with those above the diagonal $m(i,j) = m(j,i)$, $i \neq j$; but additionally, there is scale invariance, something that the usual symmetry model does not take into account.

The second kind of model, level invariant, allows us to permute some of the levels taken by some variables, so that the levels taken by some variables replace the levels of others according to a set formed by label permutations. Then, we equate the expected frequencies corresponding to the different permutations. These models are a generalization of the symmetry defined in two-way contingency tables (section **2.5.2**) and also, as said at the beginning of this chapter, are a different way of study generalized symmetry compared to what Bishop *et al.* (1975, p. 300-303) or Darroch and Bhapkar (1990) did, mainly because instead of using a saturated model to equate some of the corresponding expected frequencies we use a graphical model whose associated graph is triangle-free (saturated models are graphical models with a complete graph), by doing this, we are preserving the different conditional independences implied by the model. In level invariant models, we are not really replacing some variables with others, only the levels taken by those variables, so that it is not necessary that the permutations preserve the graph, *i.e.* we do not necessarily require the set of permutations to be a subset of the group of graph automorphisms.

We observe that the sets of permutations defined for both label and level invariant models are subsets of groups, but they are not necessarily groups themselves.

Both label and level invariant models, can be expressed as instances of a RGLL model. We observe that RGLL models in particular contain those RGLL models whose associated graphs are triangle-free and these triangle-free graph models contain both label and level invariant models.

For any label invariant model, we can get its corresponding RGLL model representing it in a general way. For level invariant models, the RGLL model representing it is obtained according to each particular model; in fact, there could be more than one RGLL model representing it. This means that, for level invariant models, having a RGLL model that fits well the data gives a necessary but not sufficient condition to have the expected frequencies equalities or symmetry we are looking for, *i.e.* if the RGLL model fits the data, then we have the equalities we were looking for, but the equalities do not imply a unique model, so that if the model does not fit the data, we can not say that there are not the equalities or symmetry we defined.

Consider $\mathbb{M}$ the vector of expected frequencies $(m(i))_{i\in \mathrm{I}}$, $\mathrm{I} = \times_{\delta\in\Delta}\mathrm{I}_\delta$, including only the constraints corresponding to the sampling scheme: Poisson, multinomial, or

restricted multinomial and $S(\mathrm{I}_\delta)$, the symmetry group including all the permutations of the level set. Additionally, we understand that a graphical log-linear model $\boldsymbol{M}$ is a set of expected frequencies written as log-linear expansions depending on the model.

Definition 4.2. A graphical log-linear model $\boldsymbol{M}$, $\boldsymbol{M} \subseteq \mathbb{M}$ is scale invariant under $S(\mathrm{I}_\delta)$ if when $\mathrm{M} \in \boldsymbol{M}$ then $\sigma^*(\mathrm{M}) \in \boldsymbol{M}$, for all $\sigma^* \in S(\mathrm{I}_\delta)$.

The transformation $\sigma^*(\mathrm{M})$ corresponds to a new model using the new order for I_δ for each σ^*. This means that the models are preserved in spite of changing the level set. For example, in a binary case with values 0 and 1, we could interchange both values, $0 \leftrightarrow 1$, and the model should be preserved or in the case of variables with three categories where I_δ= {0, 1, 2}, we have six permutations which map I_δ to {0, 1, 2}, {1, 0, 2}, {0, 2, 1}, {1, 2, 0}, {2, 0, 1}, and {2, 1, 0}, and for all of them the model should be preserved.

Suppose that there are two binary variables X and Y with an interaction between them, scale invariance means that the interactions $u_{XY}(01) = u_{XY}(10)$ and $u_{XY}(00) = u_{XY}(11)$. In the case of variables with three categories, scale invariance means that $u_{XY}(01)= u_{XY}(10)= u_{XY}(02)= u_{XY}(20)= u_{XY}(12)= u_{XY}(21)$ and $u_{XY}(00)= u_{XY}(11)= u_{XY}(22)$. Observe that by doing this, we are preserving the model and that we get one parameter for the case in which the levels are the same and another for the case when the levels differ. To illustrate the reason for using such restrictions, we present the following example.

Example 4.2. Consider a graph of a line representing three binary variables, I_W, I_X, and I_Y, with corresponding labels or vertex set V={W, X, Y}. The corresponding graphical log-linear model $\boldsymbol{M}$ has generating class A= $\{\{W, X\}, \{X, Y\}\}$, figure 4.3. In this example $\mathrm{I}_\delta = \{0, 1\}$ and $S(I_\delta) = \{Id, (01)\}$, $\delta = W, X, Y$, where Id indicates that the first level is coded as 0 and the other as 1 and (01) indicates that the first level is coded as 1 and the second as 0. In terms of definition 4.2, σ^* is Id or (01). Considering the coding given by Id, the set $\mathbb{M}$ is the vector of expected frequencies

$$(m_{WXY}(0,0,0), m_{WXY}(0,0,1), m_{WXY}(0,1,0), m_{WXY}(0,1,1),$$
$$m_{WXY}(1,0,0), m_{WXY}(1,0,1), m_{WXY}(1,1,0), m_{WXY}(1,1,1)),$$

adding the constraints corresponding to a sampling scheme: Poisson, multinomial, or restricted multinomial.

Coding the first level as 0 and the other as 1, *i.e.* using $\sigma^* = Id$, which is considered as the coding for the level set in the original model $\boldsymbol{M}$, we get the following log-linear expansion for each expected frequency M, $\mathrm{M} \in \boldsymbol{M}$, in each cell

$$\log m_{WXY}(0,0,0) = u + u_W(0) + u_X(0) + u_Y(0) + u_{WX}(00) + u_{YX}(00),$$

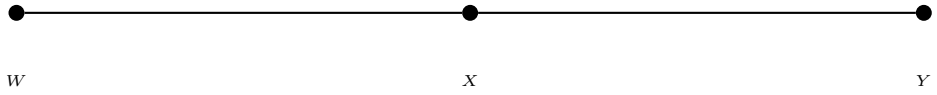


Figure 4.3: Graphical model with generating class $\{\{W,X\},\ \{X,Y\}\}$, $\log m(i,j,k) = u + u_W(i) + u_X(j) + u_Y(k) + u_{WX}(ij) + u_{YX}(kj)$, $i,j,k = 0,1$.

$$\begin{aligned}
\log m_{WXY}(0,0,1) &= u + u_W(0) + u_X(0) + u_Y(1) + u_{WX}(00) + u_{YX}(10),\\
\log m_{WXY}(0,1,0) &= u + u_W(0) + u_X(1) + u_Y(0) + u_{WX}(01) + u_{YX}(01),\\
\log m_{WXY}(0,1,1) &= u + u_W(0) + u_X(1) + u_Y(1) + u_{WX}(01) + u_{YX}(11),\\
\log m_{WXY}(1,0,0) &= u + u_W(1) + u_X(0) + u_Y(0) + u_{WX}(10) + u_{YX}(00),\\
\log m_{WXY}(1,0,1) &= u + u_W(1) + u_X(0) + u_Y(1) + u_{WX}(10) + u_{YX}(10),\\
\log m_{WXY}(1,1,0) &= u + u_W(1) + u_X(1) + u_Y(0) + u_{WX}(11) + u_{YX}(01),\\
\log m_{WXY}(1,1,1) &= u + u_W(1) + u_X(1) + u_Y(1) + u_{WX}(11) + u_{YX}(11).
\end{aligned}$$

Now, recoding the first level as 1 and the other as 0 by using $\sigma^* = (01)$, and using the same cell order as before, the model has the following log-linear expansion for each expected frequency $\sigma^*(\mathrm{M})$ in each cell

$$\begin{aligned}
\log m_{WXY}(1,1,1) &= u + u_W(1) + u_X(1) + u_Y(1) + u_{WX}(11) + u_{YX}(11),\\
\log m_{WXY}(1,1,0) &= u + u_W(1) + u_X(1) + u_Y(0) + u_{WX}(11) + u_{YX}(01),\\
\log m_{WXY}(1,0,1) &= u + u_W(1) + u_X(0) + u_Y(1) + u_{WX}(10) + u_{YX}(10),\\
\log m_{WXY}(1,0,0) &= u + u_W(1) + u_X(0) + u_Y(0) + u_{WX}(10) + u_{YX}(00),\\
\log m_{WXY}(0,1,1) &= u + u_W(0) + u_X(1) + u_Y(1) + u_{WX}(01) + u_{YX}(11),\\
\log m_{WXY}(0,1,0) &= u + u_W(0) + u_X(1) + u_Y(0) + u_{WX}(01) + u_{YX}(01),\\
\log m_{WXY}(0,0,1) &= u + u_W(0) + u_X(0) + u_Y(1) + u_{WX}(00) + u_{YX}(10),\\
\log m_{WXY}(0,0,0) &= u + u_W(0) + u_X(0) + u_Y(0) + u_{WX}(00) + u_{YX}(00).
\end{aligned}$$

A scale invariant model is one satisfying the following equalities:

$$u_{WX}(00) = u_{WX}(11),\ u_{YX}(00) = u_{YX}(11); \tag{4.1}$$

$$u_{WX}(01) = u_{WX}(10),\ u_{YX}(01) = u_{YX}(10). \tag{4.2}$$

Observe that under these equalities and using any of the two codings for the levels, we get the same interaction terms for any log-linear expansion of the expected frequency associated to a cell. For example, for the fourth cell we have the expansion

$$u + u_W(0) + u_X(1) + u_Y(1) + u_{WX}(01) + u_{YX}(11),$$

using the first coding or original model $\boldsymbol{M}$, and

$$u + u_W(1) + u_X(0) + u_Y(0) + u_{WX}(10) + u_{YX}(00),$$

using the second coding; however, since $u_{WX}(01)$ $=u_{WX}(10)$ and $u_{YX}(00)=$ $u_{YX}(11)$, we have that the interaction terms in the log-linear expansion are the same independently of the coding, that once interchanging levels all terms become part of the same group of terms, main effects for a variable become main effects for the same variable and first-order interactions become part of the same group of terms representing interactions of the same kind, and also that the model defined according to the generating class is preserved.

If we had J categories in each variable instead of two, in order to get a scale invariant model, we would get two sets of interactions that should be equal. One for the first-order interactions in which the levels of the corresponding variables differ and another for the first-order interactions whose levels are the same. That is, for two variables X and Y with an interaction between them, we get the following sets whose elements should be equal.

$$\{u_{XY}(ij), i \neq j; i, j = 1, ..., J\},$$
$$\{u_{XY}(ii), i = 1, ..., J\};$$

that is $u_{XY}(12)= u_{XY}(13)= ...= u_{XY}(1J)= u_{XY}(21)= u_{XY}(23)= u_{XY}(2J)= ... =u_{XY}(J-1J)$ for the first set and $u_{XY}(11)= u_{XY}(22)= ...= u_{XY}(JJ)$ for the second set.

We observe that having a scale invariant model does not imply that the expected frequencies are the same after permuting the levels. For example, if we have three binary variables associated to W, X, and Y, in general it is not true that $m_{WXY}(001)$ is equal to $m_{WXY}(110)$, the expected frequency after permuting $0 \leftrightarrow 1$, this inequality can be seen in example 4.2 where $\log m_{WXY}(0,0,1)= u+ u_W(0)+ u_X(0)+ u_Y(1)+ u_{WX}(00)+ u_{YX}(10)$ and $\log m_{WXY}(1,1,0)= u+ u_W(1)+ u_X(1)+ u_Y(0)+ u_{WX}(11)+ u_{YX}(01)$ so that even considering equalities (4.1) and (4.2) the expected frequencies are not the same; however; all terms related to the association between the variables are

the same after permuting levels, this means that the scale is irrelevant in measuring the association between variables and that after any permutation in the level order all terms belong to the same group of terms, preserving the model.

Notice that the expected frequencies not being the same is caused by the fact that the main effects for the same vertex class differ after permuting levels; however, each main effect belongs to the same group of terms and as they do not involve associations between variables the model is preserved in this sense.

If we wanted the expected frequencies to be the same because we needed to define that a model is preserved like this, then we would need to equate the main effects in all their levels. If we had J levels, for each variable Q we would need to have $u_Q(1)=$ $u_Q(2)= ...=u_Q(J)$, but this would mean that the main effects are always the same, so that we would have constant terms instead of main effects and we would be changing the model by eliminating main effects, getting a different model that is not even hierarchichal, much less graphical. So that we should not talk about preserving the model in this sense.

This way of defining scale invariance is analogous to the one given by Højsgaard and Lauritzen (2008) in the Gaussian case, where the concentration matrix K can be seen as $K = ACA$, with A a diagonal matrix with entries equal to the inverse partial standard deviations, terms that depend on only one variable, and C a matrix in which all diagonal entries are equal to one and all off-diagonal entries are the negative partial correlations, which are terms that depend on the association between variables. Scale invariance in Gaussian models means that lineal transformations for the scale in the same vertex class affect only the matrix A and not C. In the discrete case it is analogous, changes in the scale affect the main effects in the same vertex class, but not the first-order interactions, which are the terms related with the association between pairs of variables.

Consider a graphical log-linear model with generating class $C(G)$, the set of cliques of G, and associated graph $G = (V, E')$, where G is a triangle-free graph. Consider also $Aut(G)$ a subset of the symmetric group formed by all permutations of the vertex set V, $S(V)$, consisting on permutations that preserve the graph, $S(\mathrm{I}_\delta)$ the symmetry group including all the permutations of the level set, and $\mathbb{M}$ the expected frequency vector $(m(i))_{i\in \mathrm{I}}$, $\mathrm{I} = \times_{\delta\in\Delta}\mathrm{I}_\delta$, including only the constraints corresponding to the sampling scheme: Poisson, multinomial, or restricted multinomial.

Definition 4.3. A label invariant model $\boldsymbol{M}(G, \Gamma)$ determined by a triangle-free graph G and by $\Gamma \subseteq Aut(G)$, is given by assuming that: i) we have a graphical log-linear model $\boldsymbol{M}$ with generating class $C(G)$ and $\boldsymbol{M} \subseteq \mathbb{M}$, and ii) if $\mathrm{M} \in \boldsymbol{M}$ then $\sigma(\sigma^*(\mathrm{M})) \in \boldsymbol{M}$, for all $\sigma^* \in S(\mathrm{I}_\delta)$ and for all $\sigma \in \Gamma$.

As before, σ^*(M) changes the level set order according to all possible permutations of I_δ. That is, we are asking to have a scale invariant model. As Γ is contained in $Aut(G)$, then Γ is a set of permutations preserving neighbours in the graph, which means that the labels can be changed without getting a new graph. Additionally, to get that $\sigma(M') \in \boldsymbol{M}$ for $\sigma \in \Gamma$, we need the main effects and first-order interactions to be restricted in such a way that after permuting the labels the model remains, this implies that the expected frequencies, and as a consequence the distribution are the same after the permutations. The restrictions correspond to equate main effects for some variables in each level and to equate first-order interactions for different pairs of variables. Then, we have restrictions like the ones defined for RGLL models.

These models are analogous to the RCOP models defined by Højsgaard and Lauritzen (2008) in the Gaussian case. RCOP models satisfy the following properties: a) they preserve the distribution after applying permutation groups which permute the labels on the graph, b) they are scale invariant, and c) they generate symmetric graphs according to the permutation group in the sense that the graph is preserved after permuting labels. The only difference with respect to label invariant models is that the condition to get scale invariance is not added because it is automatically obtained using that RCOP models are a particular case of RCOR models (Højsgaard and Lauritzen, 2008), and RCOR models are scale invariant models in which some of their concentration matrix elements and some partial correlations are restricted to be equal.

In that sense, RGLL models could be considered analogous to the RCON models defined in the Gaussian case (Højsgaard and Lauritzen, 2008) because they are models not necessarily scale invariant in which some of their concentration matrix elements are equal.

Example 4.3. Consider as in example 4.2 the graphical log-linear model $\boldsymbol{M}$ with generating class A= $\{\{W,X\},\{X,Y\}\}$ and graph G. The graph G is triangle-free. Additionally, suppose that $\Gamma = \{(WY)\}$, (WY) is an automorphism of G because the graph associated to that mapping, figure 4.4(b), is isomorphic to G, figure 4.4(a), as a consequence $\Gamma \subseteq Aut(G)$. For these G and Γ, we define a label invariant model $\boldsymbol{M}(G,\Gamma)$.

Restrictions for scale invariance were already obtained in example 4.2, equations (4.1) and (4.2). We additionally require restrictions implying that if M' $\in \boldsymbol{M}$ then σ(M') $\in \boldsymbol{M}$, for all $\sigma \in \Gamma$, *i.e.* that after permuting labels according to Γ the expected frequencies follow the same model. To get them, observe that the expected frequencies expanded as a log-linear model for each cell before applying the permutation set Γ to the vertex set V are

$$\log m_{WXY}(0,0,0) = u + u_W(0) + u_X(0) + u_Y(0) + u_{WX}(00) + u_{YX}(00),$$

(a) Graphical model (b) Model under (WY)

Figure 4.4: Graphical log-linear model with generating class $\{\{W,X\},\{X,Y\}\}$: (a) model representation without using any mapping, (b) model representation under the mapping (WY).

$$\log m_{WXY}(0,0,1) = u + u_W(0) + u_X(0) + u_Y(1) + u_{WX}(00) + u_{YX}(10),$$
$$\log m_{WXY}(0,1,0) = u + u_W(0) + u_X(1) + u_Y(0) + u_{WX}(01) + u_{YX}(01),$$
$$\log m_{WXY}(0,1,1) = u + u_W(0) + u_X(1) + u_Y(1) + u_{WX}(01) + u_{YX}(11),$$
$$\log m_{WXY}(1,0,0) = u + u_W(1) + u_X(0) + u_Y(0) + u_{WX}(10) + u_{YX}(00),$$
$$\log m_{WXY}(1,0,1) = u + u_W(1) + u_X(0) + u_Y(1) + u_{WX}(10) + u_{YX}(10),$$
$$\log m_{WXY}(1,1,0) = u + u_W(1) + u_X(1) + u_Y(0) + u_{WX}(11) + u_{YX}(01),$$
$$\log m_{WXY}(1,1,1) = u + u_W(1) + u_X(1) + u_Y(1) + u_{WX}(11) + u_{YX}(11).$$

After applying the permutation set Γ to the vertex set V, which in this case corresponds to permuting labels W and Y, the expected frequencies expanded as a log-linear model for each cell are

$$\log m_{YXW}(0,0,0) = u + u_Y(0) + u_X(0) + u_W(0) + u_{YX}(00) + u_{WX}(00),$$
$$\log m_{YXW}(0,0,1) = u + u_Y(0) + u_X(0) + u_W(1) + u_{YX}(00) + u_{WX}(10),$$
$$\log m_{YXW}(0,1,0) = u + u_Y(0) + u_X(1) + u_W(0) + u_{YX}(01) + u_{WX}(01),$$
$$\log m_{YXW}(0,1,1) = u + u_Y(0) + u_X(1) + u_W(1) + u_{YX}(01) + u_{WX}(11),$$
$$\log m_{YXW}(1,0,0) = u + u_Y(1) + u_X(0) + u_W(0) + u_{YX}(10) + u_{WX}(00),$$
$$\log m_{YXW}(1,0,1) = u + u_Y(1) + u_X(0) + u_W(1) + u_{YX}(10) + u_{WX}(10),$$
$$\log m_{YXW}(1,1,0) = u + u_Y(1) + u_X(1) + u_W(0) + u_{YX}(11) + u_{WX}(01),$$
$$\log m_{YXW}(1,1,1) = u + u_Y(1) + u_X(1) + u_W(1) + u_{YX}(11) + u_{WX}(11).$$

The log-linear expansion for the expected frequency of the second cell, $\log m_{WXY}(0,0,1)$, represents the same model after interchanging labels, that is $\log m_{YXW}(0,0,1)$ defines the same model than $\log m_{WXY}(0,0,1)$ if we set the following equalities

$$u_W(0) = u_Y(0),\ u_W(1) = u_Y(1);$$
$$u_{WX}(00) = u_{YX}(00);\ u_{WX}(10) = u_{YX}(10).$$

Observe that a consequence of these equalities is that

$$\log m_{WXY}(0,0,1) = \log m_{YXW}(0,0,1).$$

Considering all cells, the equalities that preserve the model after interchanging labels according to $\Gamma = \{(WY)\}$ are

$$u_W(0) = u_Y(0),\ u_W(1) = u_Y(1); \tag{4.3}$$

$$\begin{aligned} u_{WX}(00) = u_{YX}(00);\ u_{WX}(10) = u_{YX}(10);\\ u_{WX}(01) = u_{YX}(01);\ u_{WX}(11) = u_{YX}(11). \end{aligned} \tag{4.4}$$

Observe that a consequence of these restrictions, equations (4.3) and (4.4), is that the expected frequencies for a cell before and after the permutation are equal, for example $m_{WXY}(0,0,1) = m_{YXW}(0,0,1)$. Restrictions (4.3) and (4.4) together with the scale invariance restrictions, equations (4.1) and (4.2), imply the following equalities

$$u_W(0) = u_Y(0),\ u_W(1) = u_Y(1); \tag{4.5}$$

$$\begin{aligned} u_{WX}(00) = u_{YX}(00) = u_{WX}(11) = u_{YX}(11);\\ u_{WX}(01) = u_{YX}(01) = u_{WX}(10) = u_{YX}(10). \end{aligned} \tag{4.6}$$

The label invariant model $\boldsymbol{M}(G,\Gamma)$ is the graphical log-linear model with graph G satisfying (4.5) and (4.6).

We present the algebraic concept of orbits as defined in terms of permutation sets and graphs by Højsgaard and Lauritzen (2008), a general definition is provided for instance by Rotman (1995, p.56). This is a useful concept to get the restrictions needed in label invariant models so that we do not have to obtain for each case the log-linear expansions to know which terms to equate.

The *vertex orbits* of Γ are the classes of the relation

$$\alpha \equiv_\Gamma \beta \Leftrightarrow \beta = \sigma\alpha,\ \text{for some } \sigma \in \Gamma,$$

where α and β are vertices contained in the set of vertices of G, V.

The *edge orbits* of Γ are the classes of the relation

$$\{\alpha,\gamma\} \equiv_\Gamma \{\beta,\delta\} \Leftrightarrow \{\beta,\delta\} = \{\sigma\alpha,\sigma\gamma\}, \text{ for some } \sigma \in \Gamma.$$

where $\{\alpha,\gamma\}$ and $\{\beta,\delta\}$ are edges contained in the set of edges of G, E'.

Label invariant models can be represented with a simple graph, G^*, corresponding to G coloured according to the vertex and edge orbits, in which all vertices in the same orbit have the same colour and the same for the edges. The graph G^* represents the different parameters equalities that have to be satisfied in order to be able to change the labels according to Γ preserving the model; however, this colouring does not represent the conditions required to have scale invariance; although it helps us to determine the coloured multi-graph G^{**}, the graph with multiple edges, representing the RGLL model that has to be fitted in order to fit the label invariant model, obviously including scale invariance.

A one-to-one correspondence holds between the vertex orbits $V_1, V_2, ..., V_T$ and the vertex colour classes $V_1, V_2, ..., V_T$ in a RGLL model. This is because if under a permutation σ, $\beta = \sigma\alpha$, *i.e.* α and β are in the same vertex orbit, then we have that α takes the place of β once permuting them, so that in order to have the same model and as a consequence the same distribution after permuting variables, we should have that the main effects $u_\alpha(r) = u_\beta(r)$ for all r in the level set. Obviously, if we had more than two elements in a vertex orbit the corresponding equalities are applied to all the elements in the same orbit. To sum it up, vertex orbits are vertex classes.

For the edge orbits $\xi_1, ..., \xi_l$ the association with edge colour classes in a RGLL model and its associated coloured graph G^{**} is not a one-to-one correspondence. Suppose that we take the edge orbit ξ_1 and that $\{\alpha,\gamma\}$ and $\{\beta,\delta\}$ are edges in this orbit. To keep the same model, we should have that $u_{\beta\delta}(rs) = u_{\alpha\gamma}(rs)$ for all r and s in the level set. If all variables had J levels, this fact would mean that there are J^2 restrictions,

$$u_{\beta\delta}(11) = u_{\alpha\gamma}(11),\ ...,\ u_{\beta\delta}(1J) = u_{\alpha\gamma}(1J),\ ..., u_{\beta\delta}(J-11) = u_{\alpha\gamma}(J-11),$$

$$....,\ u_{\beta\delta}(J-1J) = u_{\alpha\gamma}(J-1J),\ u_{\beta\delta}(JJ) = u_{\alpha\gamma}(JJ).$$

As we also need scale invariance, we only have two restriction groups,

$$u_{\beta\delta}(12) = u_{\alpha\gamma}(12) = ... = u_{\beta\delta}(rs) = u_{\alpha\gamma}(rs) = ... = u_{\beta\delta}(J-1J) = u_{\alpha\gamma}(J-1J),\ r \neq s,$$

$$\text{and } u_{\beta\delta}(11) = u_{\alpha\gamma}(11) = ... = u_{\beta\delta}(JJ) = u_{\alpha\gamma}(JJ).$$

This means that the edges $\{\alpha,\gamma\}$ and $\{\beta,\delta\}$ belonging to the same orbit ξ_1 in the simple graph G^* induce a division of the edges in the graph G^{**} of a RGLL model into

two colour classes E_1 and E_2. The edges corresponding to different categories, *i.e.* the edges $u_{\beta\delta}(rs)$, $u_{\alpha\gamma}(rs)$, $r \neq s$ are in the same class E_1 and the edges corresponding to the same categories $u_{\beta\delta}(rr)$ and $u_{\alpha\gamma}(rr)$ are in other class E_2. Then, all elements having the same colour in G^* are divided into two, so that in G^{**} we have two colours instead of one, but it is in such a way that if we took the edges in E_1 (or E_2) and considered the corresponding underlying simple graph, *i.e.* taking only one edge instead of all, we would have the same colouring as in G^*.

We illustrate in the following example how we obtain the parameters restrictions using vertex and edge orbits. The example corresponds to the label invariant model $\boldsymbol{M}(G,\Gamma)$ studied in example 4.3, which is based on the graphical log-linear model presented in example 4.2.

Example 4.4. Consider the graphical log-linear model $\boldsymbol{M}$ with generating class A= $\{\{W,X\},\{X,Y\}\}$ and associated graph G and the label invariant model $\boldsymbol{M}(G,\Gamma)$ with $\Gamma = \{(WY)\}$. We got in equations (4.5) and (4.6) the restrictions that main effects and first-order interactions should satisfy to have a label invariant model. These restrictions can be obtained using the vertex and edge orbits concepts defined above as follows.

We have that $W = \sigma Y$ for $\sigma = (WY)$, this equality implies that W and Y are in the same class. As there are no other σ, X is in a class containing only X. Then, the vertex orbits are

$$V_1 = \{W,Y\},$$
$$V_2 = \{X\}.$$

On the other hand, $\{W,X\} = \{\sigma Y, \sigma X\}$ for $\sigma = (WY)$, which implies that both edges are in the same edge orbit as follows,

$$\xi_1 = \{\{W,X\},\{Y,X\}\}.$$

As a consequence, the orbits graph G^*, as defined above, is obtained colouring W and Y with the same colour and X with other colour and both edges are coloured the same, figure 4.5. This graph provides the restrictions that allow to preserve the model after permuting variable labels according to Γ, such restrictions are that the main effects for vertices in the same vertex orbit are equal in each level and that each pair of different interactions $u_{YX}(ij)$ and $u_{WX}(ij)$ is equated for all i and j according to the unique edge orbit, $\{\{W,X\},\{Y,X\}\}$, see equations (4.3) and (4.4).

Seen as a RGLL model, the vertex orbits are equivalent to vertex classes and we use the restrictions obtained from the edge orbits together with the scale invariance restrictions, equations (4.1) and (4.2), to obtain the edge colour classes

$$E_1 = \{u_{WX}(00), u_{YX}(00), u_{WX}(11), u_{YX}(11)\},$$

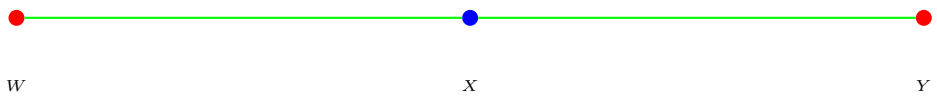


Figure 4.5: Orbits graph, G^*, for the label invariant model $\boldsymbol{M}(G,\Gamma)$ with generating class $\{\{W,X\},\{X,Y\}\}$, graph G, permutation set $\Gamma=\{(WY)\}$, and vertex orbits $\{W,Y\}$ and $\{X\}$ and edge orbit $\{\{W,X\},\{Y,X\}\}$.

$$E_2=\{u_{WX}(01),u_{YX}(01),u_{WX}(10),u_{YX}(10)\}.$$

As a consequence, we obtain a RGLL model with generating class $\{\{W,X\},\{X,Y\}\}$ with vertices $V=(V_1,V_2)$, $V_1=\{W,Y\}$ and $V_2=\{X\}$, and edges $E=(E_1,E_2)$, E_1= $\{u_{WX}(00),\ u_{YX}(00),\ u_{WX}(11),\ u_{YX}(11)\}$, E_2= $\{u_{WX}(01),\ u_{YX}(01),\ u_{WX}(10),\ u_{YX}(10)\}$. The coloured graph G^{**} (figure 4.6) for this RGLL model has the same vertex colouring than G^*, but as it is a multi-graph, the edge colouring is applied in 8 edges according to the edge classes.

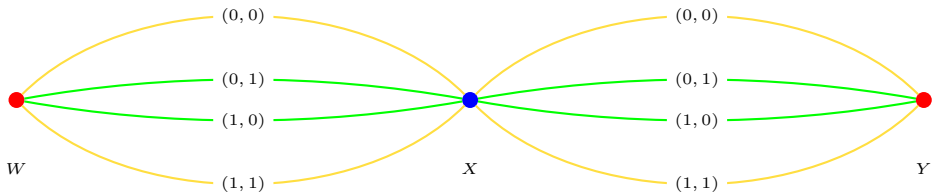


Figure 4.6: RGLL model with generating class $\{\{W,X\},\{X,Y\}\}$ and associated graph $G^{**}(V,E)$, where all variables are binary, with vertex set $V=(V_1,V_2)$, $V_1=\{W,Y\}$ and $V_2=\{X\}$, and edge set $E=(E_1,E_2)$, E_1= $\{u_{WX}(00),\ u_{YX}(00),\ u_{WX}(11),\ u_{YX}(11)\}$, E_2= $\{u_{WX}(01),\ u_{YX}(01),\ u_{WX}(10),\ u_{YX}(10)\}$.

We observe that using parameters reparametrized under effect coding, *i.e.* parameters in which the sum of the main effects or interactions over any index is zero, the restrictions that allow to preserve the model after interchanging labels (the restrictions obtained directly from the orbits) using reparametrized parameters are equivalent to the restrictions using the parameters defined for RGLL models because the kind of restrictions needed do not depend on the reparametrization used. The reason for this is that the restrictions that equate the main effects over all levels for some variables are the same using either type of parameter, as we saw in section **3.9.4**. Additionally, as we

also explained in section **3.9.4**, there are three kind of restrictions over the first-order interactions that are equivalent using either effect coding reparametrized parameters or the parameters used for RGLL models. Denoting as u' the reparametrized parameters and as u the others, the three type of parameter restrictions that are the same for first-order interactions independently of the parametrization used (see theorem 3.8) are

1. Restrictions of the kind $u_{XY}(ij) = u_{ZR}(ij)$. These restrictions are the same in the reparametrized parameters, *i.e.* $u_{XY}(ij) = u_{ZR}(ij) \Leftrightarrow u'_{XY}(ij) = u'_{ZR}(ij)$.

2. Restrictions of the kind $u_{XY}(ij) = u_{XY}(ji)$, $i < j$. These restrictions are the same in the reparametrized parameters, *i.e.* $u_{XY}(ij) = u_{XY}(ji)\ i < j \Leftrightarrow u'_{XY}(ij) = u'_{XY}(ji)$, $i < j$.

3. Restrictions of the kind $u_{XY}(ij) = u_{ZR}(ji)$, $i \leq j$. These restrictions are the same in the reparametrized parameters, *i.e.* $u_{XY}(ij) = u_{ZR}(ji)$, $i \leq j \Leftrightarrow u'_{XY}(ij) = u'_{ZR}(ji)$, $i \leq j$.

These are the kind of first-order interaction restrictions we have when we permute labels. In the binary case it means that instead of having four parameters representing all the interactions between a pair of variables we only have one reparametrized parameter and because in this case conditions to preserve the model after permuting labels can be obtained using either kind of parameter and scale invariance is also directly implied by the reparametrized parameter, the only parameter u'_{XY} implies $u_{XY}(01) = u_{XY}(10)$ and $u_{XY}(00) = u_{XY}(11)$, then we could use reparametrized parameters to obtain a label invariant model. As a consequence, in this case the vertex orbit graph G^* is the same as the RGLL model graph considering reparametrized parameters representing a label invariant model.

The graphs G^* obtained for label invariant models to represent vertex and edge orbits can always be used to get the graphs G^{**} corresponding to RGLL models, but in general both graphs are not the same. In the Gaussian case, the graphs obtained for RCOP models are the same that the ones obtained for the associated RCON or RCOR models because in this case vertex orbits are the model vertex classes and edge orbits are edge classes. This fact was used to determine that RCOP models could be solved using the same results and procedures obtained for RCOR or RCON models although there are simpler formulas that can be derived to be used only in RCOP models. Here we propose something similar, we have a label invariant model, we write it as a RGLL model, and we use the graph G^* obtained for the label invariant model to get the graph G^{**} for the RGLL model, and then we are able to use the theory obtained for RGLL models to fit it.

If we have a triangle-free coloured graph, we can obtain in some cases a label invariant model with the corresponding parameter restrictions that define it according to

the colouring. Then, given any coloured triangle-free graph, we could ask ourselves if this graph and the corresponding multi-graph represent a label invariant model. This affirmation is not always true as we discuss at the end of example 4.8. It is related with other research work corresponding to a Ph.D. research project still in process developed by Neufeld (2009) where the same question is formulated for Gaussian models.

Consider a graphical log-linear model with generating class $C(G)$, the set of cliques of G, $S(V)$ the symmetric group formed by all permutations of V, and $\mathbb{M}$ the expected frequency vector $(m(i))_{i\in \mathrm{I}}$, $\mathrm{I} = \times_{\delta\in\Delta}\mathrm{I}_\delta$, including only the constraints corresponding to the sampling scheme: Poisson, multinomial, or restricted multinomial.

Definition 4.4. A level invariant model $\boldsymbol{M'}(G,\Gamma')$ determined by a triangle-free graph G and by $\Gamma' \subseteq S(V)$ is given by assuming that: i) we have a graphical log-linear model $\boldsymbol{M'}$ with generating class $C(G)$ and $\boldsymbol{M'} \subseteq \mathbb{M}$ and ii) if $M \in \boldsymbol{M'}$, then $M(i) = M(\sigma' i)$ for all $\sigma' \in \Gamma'$ and $i \in \mathrm{I}$.

Observe that when $\Gamma' = S(V)$, *i.e.* Γ' corresponds to the set of all possible permutations of the labels, then we will be equating $m(i) = m(per(i))$, where $per(i)$ denotes all possible permutations of the elements forming the cells. This is the concept of symmetry as defined for instance by Bishop *et al.* (1975, p. 300-303) where the generating class they use corresponds to a saturated model; however, under level invariant models we use a generating class different to the one for saturated models, as long as the corresponding graphical log-linear model has an associated triangle-free graph, so that we are preserving the dependences implied by the model. In this sense, it can be considered as a different concept of symmetry. For example, if we have a graphical log-linear model generated by A= $\{\{W,X\},\{X,Y\}\}$ with associated graph G and we define a model $\boldsymbol{M'}(G,\Gamma')$ with $\Gamma' = S(V)$, then we have a model in which simultaneously $m(i) = m(per(i))$ and $W \perp Y|Z$, where as usual this conditional independence can be derived from the graphical log-linear model.

We have that the restrictions obtained for a label invariant model can be used to get a particular level invariant model, the model in which we change the levels of the vertices according to the label permutation set Γ, *i.e.* we use $\Gamma' = \Gamma$; however, by using the same model, we will be adding scale invariant restrictions that are not necessary in level invariant models. In fact, if we have a set Γ' that can be used for a label invariant model, *i.e.* Γ' is a subset of the group of automorphisms of G $Aut(G)$, we could use the vertex and edge orbits of Γ' to get restrictions that could be used in the level invariant model. This is done by using the vertex orbits as vertex classes and by equating the first-order interactions in the same edge orbit for all possible permutation of the levels.

Then, a graph representing a label invariant model is also one representation of a particular case of level invariant model, the model in which $\Gamma' = \Gamma$. In this sense, we have that the graphs for label invariant models are contained in the graphs for level

invariant models, level invariant model graphs are contained in triangle-free coloured graphs, and finally, these are contained in graphs for RGLL models.

Finally, we observe that there are permutations of the vertices or labels that can not be applied in label invariant models because they are not automorphisms of the graph; however, permuting the levels taken by those labels could make sense. For example, if we have a cycle of length four, a label invariant model only interchanging two neighbour vertices does not make sense because we would have new interactions, or edges, that did not exist before; however, it is possible to have a level invariant model in which we change the values taken by two neighbour variables for this cycle, we will discuss this model in example 4.8.

Example 4.5. Consider once again the graphical log-linear model with generating class A= $\{\{W,X\},\{X,Y\}\}$ and associated graph G as in example 4.2 and additionally suppose that the model satisfies

$$m_{WXY}(w,x,y) = m_{WXY}(y,x,w),\ w \in \mathrm{I}_W,\ x \in \mathrm{I}_X,\ y \in \mathrm{I}_Y,$$

$\mathrm{I}_W = \mathrm{I}_X = \mathrm{I}_Y = \{0,1\}$. This means that the values or levels taken by the variables are modified in such a way that the value of the last variable becomes the value of the first, the value of the second variable remains equal, and the value of the first one becomes the value of the last and the expected frequency after such change should be the same. This is a level invariant model $\boldsymbol{M'}(G,\Gamma')$ with $\Gamma' = \{(WY)\}$ since G is a triangle-free graph and since for a cell (w,x,y), $\sigma'(w,x,y) = (y,x,w)$ for $\sigma' = (WY)$, and the corresponding expected frequencies are equated.

To obtain restrictions that may be used to get such equalities between the expected frequencies, we write them as log-linear expansions according to the generating class

$$\log m_{WXY}(w,x,y) = u + u_W(w) + u_X(x) + u_Y(y) + u_{WX}(wx) + u_{YX}(yx),\ w,x,y = 0,1;$$

$$\log m_{WXY}(y,x,w) = u + u_W(y) + u_X(x) + u_Y(w) + u_{WX}(yx) + u_{YX}(wx),\ w,x,y = 0,1.$$

To equate both expressions we use the restrictions $u_{WX}(ij) = u_{YX}(ij)$ and $u_W(i) = u_Y(i)$, for all i,j. Then we have the equalities $u_{WX}(00) = u_{YX}(00)$, $u_{WX}(01) = u_{YX}(01)$, $u_{WX}(10) = u_{YX}(10)$, $u_{WX}(11) = u_{YX}(11)$, $u_W(0) = u_Y(0)$, and $u_W(1) = u_Y(1)$.

This is a RGLL model with generating class A= $\{\{W,X\},\{X,Y\}\}$, vertices $V = (V_1,V_2)$ with

$$V_1 = \{W,Y\};$$
$$V_2 = \{X\};$$

and edges $E = (E_1, E_2, E_3, E_4)$, where

$$E_1 = \{u_{WX}(00), u_{YX}(00)\};$$
$$E_2 = \{u_{WX}(01), u_{YX}(01)\};$$
$$E_3 = \{u_{WX}(10), u_{YX}(10)\};$$
$$E_4 = \{u_{WX}(11), u_{YX}(11)\};$$

whose graph is shown in figure 4.7.

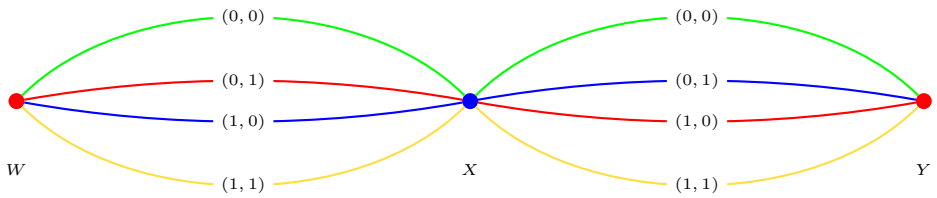


Figure 4.7: RGLL model with generating class $\{\{W, X\}, \{X, Y\}\}$*, where all variables are binary, with vertices* $V = (V_1, V_2)$*,* $V_1 = \{W, Y\}$ *and* $V_2 = \{X\}$*, and* $E = (E_1, E_2, E_3, E_4)$*,* $E_1 = \{u_{WX}(00), u_{YX}(00)\}$*,* $E_2 = \{u_{WX}(01), u_{YX}(01)\}$*,* $E_3 = \{u_{WX}(10), u_{YX}(10)\}$*, and* $E_4 = \{u_{WX}(11), u_{YX}(11)\}$*.*

Because $\Gamma' \subseteq Aut(G)$, which can be seen in the same way as in example 4.3, then we can get all restrictions from the vertex and edge orbits. As in example 4.4 the vertex orbits are $V_1 = \{W, Y\}$ and $V_2 = \{X\}$ and the edge orbit is $\xi_1 = \{\{W, X\}, \{Y, X\}\}$, which implies that the main effects for the the vertices in the same vertex orbit are equal in each level and that each pair of different interactions $u_{YX}(ij)$ and $u_{WX}(ij)$ is equated, equations (4.3) and (4.4). These are exactly the same equalities defining a RGLL model that we obtained before using the log-linear expansions.

Observe that the parameter restrictions and RGLL model used for the label invariant model $\boldsymbol{M}(G, \Gamma)$ with $\Gamma = \{(WY)\}$ presented in examples 4.3 and 4.4 preserve the expected frequency equalities as required for the level invariant model $\boldsymbol{M'}(G, \Gamma')$, although the parameters are even more restricted. This means that the coloured graph in figure 4.6 also represents the level invariant model $\boldsymbol{M'}(G, \Gamma')$ and that there are at least two RGLL models representing the same level invariant model.

4.1 Examples

In this section we present six examples of label and level invariant models for different graphical models to illustrate the concepts seen so far in this chapter. Remembering

that all label and level invariant models can be represented as RGLL models, we present in the examples involving label invariant models the associated RGLL model and for level invariant model we present at least one RGLL model associated to it.

Example 4.6. For a graphical log-linear model whose associated graph G is a line exists a general way to express a level invariant model in which we permute the order of the levels taken for the variables from front to back, *i.e.* the value of the last variable is interchanged with the value of the first one, the value of the previous to last variable is interchanged with the value of the second one, the value of the third from last variable is interchanged with the value of the third one, and so on. We want the expected frequencies to be the same after this change. This is equivalent to a level invariant model in which the first and last variable levels are interchanged, the previous to last and the second are interchanged, the third from last and the third are interchanged, and so on. Example 4.5 is a particular case of these models where the number of vertices $|V|$ is 3.

Supposing that the vertex set is V= $\{a_1,\ a_2,\ ...,\ a_{|V|}\}$, the generating class associated to G is A= $\{\{a_1,\ a_2\},\ ....,\ \{a_{|V-1|},\ a_{|V|}\}\}$. The restrictions and colour classes needed to obtain a RGLL model with generating class A representing such level invariant model $\boldsymbol{M'}(G,\Gamma')$, with Γ'=$\{(a_1a_{|V|})\ (a_2a_{|V|-1})\ ....\ (a_{\frac{|V|-1}{2}}a_{\frac{|V|+3}{2}})\}$ if $|V|$ is odd and Γ'=$\{(a_1a_{|V|})\ (a_2a_{|V|-1})\ ....\ (a_{\frac{|V|}{2}}a_{\frac{|V|}{2}+1})\}$ if $|V|$ is even, are given by

If $|V|$ is odd we use V=$(V_1,\ V_2,....,\ V_{\frac{|V|-1}{2}},\ V_{\frac{|V|-1}{2}+1})$ with $V_1 = \{a_1, a_{|V|}\}$, $V_2 = \{a_2, a_{|V|-1}\}$,...., $V_{\frac{|V|-1}{2}} = \{a_{\frac{|V|-1}{2}}, a_{\frac{|V|+3}{2}}\}$, $V_{\frac{|V|-1}{2}+1} = \{a_{\frac{|V|-1}{2}+1}\}$. We have $\frac{|V|-1}{2}$ groups of restrictions for each (i,j) permutation of the levels for the first-order interactions, $u_{a_1a_2}(ij) = u_{a_{|V|}a_{|V|-1}}(ij)$, ..., $u_{a_{\frac{|V|-1}{2}}a_{\frac{|V|-1}{2}+1}}(ij) = u_{a_{\frac{|V|+3}{2}}a_{\frac{|V|-1}{2}+1}}(ij)$. Then the edge colour classes are $\{u_{a_1a_2}(ij),\ u_{a_{|V|}a_{|V|-1}}(ij)\}$, ..., $\{u_{a_{\frac{|V|-1}{2}}a_{\frac{|V|-1}{2}+1}}(ij),\ u_{a_{\frac{|V|+3}{2}}a_{\frac{|V|-1}{2}+1}}(ij)\}$, for all i,j.

If $|V|$ is even we use V= $(V_1,\ V_2,....,\ V_{\frac{|V|}{2}})$ with $V_1 = \{a_1, a_{|V|}\}$, $V_2 = \{a_2, a_{|V|-1}\}$,...., $V_{\frac{|V|}{2}} = \{a_{\frac{|V|}{2}}, a_{\frac{|V|}{2}+1}\}$. We have $\frac{|V|}{2}$ groups of restrictions for each (i,j) permutation of the levels for the first-order interactions, $u_{a_1a_2}(ij) = u_{a_{|V|}a_{|V|-1}}(ij)$, ..., $u_{a_{\frac{|V|}{2}-1}a_{\frac{|V|}{2}}}(ij) = u_{a_{\frac{|V|}{2}+2}a_{\frac{|V|}{2}+1}}(ij)$, and $u_{a_{\frac{|V|}{2}}a_{\frac{|V|}{2}+1}}(ij) = u_{a_{\frac{|V|}{2}}a_{\frac{|V|}{2}+1}}(ji)$. Then the edge colour classes are $\{u_{a_1a_2}(ij),\ u_{a_{|V|}a_{|V|-1}}(ij)\}$, ..., $\{u_{a_{\frac{|V|}{2}-1}a_{\frac{|V|}{2}}}(ij),\ u_{a_{\frac{|V|}{2}+2}a_{\frac{|V|}{2}+1}}(ij)\}$, and $\{u_{a_{\frac{|V|}{2}}a_{\frac{|V|}{2}+1}}(ij),\ u_{a_{\frac{|V|}{2}}a_{\frac{|V|}{2}+1}}(ji)\}$ for all i,j.

For instance if $|V| = 5$, V= $\{a_1,\ a_2,\ a_3,\ a_4,\ a_5\}$, the level invariant model $\boldsymbol{M'}(G,\Gamma')$ with Γ'=$\{(a_1a_5)\ (a_2a_4)\}$ and whose generating class associated to G is A= $\{\{a_1,\ a_2\},\ \{a_2,\ a_3\},\ \{a_3,\ a_4\},\ \{a_4,\ a_5\}\}$ is represented as a RGLL model with generating class A, vertices V=$(V_1,\ V_2,\ V_3)$ with $V_1 = \{a_1, a_5\}$, $V_2 = \{a_2, a_4\}$, and $V_3 = \{a_3\}$, and edge

colour classes $\{u_{a_1a_2}(ij),\ u_{a_5a_4}(ij)\}$, $\{u_{a_2a_3}(ij),\ u_{a_4a_3}(ij)\}$ for all i, j.

We can also get a RGLL model for label invariant models in which a_1 is permuted with $a_{|V|}$, a_2 with $a_{|V|-1}$, etc, *i.e.* the model $\boldsymbol{M}(G, \Gamma)$ with $\Gamma = \Gamma'$. Example 4.3 is a particular case of these models where the number of vertices $|V| = 3$. The vertex colouring for the associated RGLL model whose generating class is A is the same as in the level invariant case. The edge colouring is the following.

If $|V|$ is odd we have the following sets whose interactions should be the same or edge colour classes: $\{u_{a_1a_2}(ij), u_{a_{|V|}a_{|V|-1}}(ij), i \neq j\}$, $\{u_{a_1a_2}(ii), u_{a_{|V|}a_{|V|-1}}(ii); i = 1, ..., J\}$, ..., $\{u_{a_{\frac{|V|-1}{2}}a_{\frac{|V|-1}{2}+1}}(ij), u_{a_{\frac{|V|+3}{2}}a_{\frac{|V|-1}{2}+1}}(ij), i \neq j\}$, $\{u_{a_{\frac{|V|-1}{2}}a_{\frac{|V|-1}{2}+1}}(ii), u_{a_{\frac{|V|+3}{2}}a_{\frac{|V|-1}{2}+1}}(ii); i = 1, ..., J\}$.

If $|V|$ is even we have the following sets whose interactions should be the same or edge colour classes: $\{u_{a_1a_2}(ij), u_{a_{|V|}a_{|V|-1}}(ij), i \neq j\}$, $\{u_{a_1a_2}(ii), u_{a_{|V|}a_{|V|-1}}(ii); i = 1, ..., J\}$, ..., $\{u_{a_{\frac{|V|}{2}-1}a_{\frac{|V|}{2}}}(ij),\ u_{a_{\frac{|V|}{2}+2}a_{\frac{|V|}{2}+1}}(ij),\ i \neq j\}$, $\{u_{a_{\frac{|V|}{2}-1}a_{\frac{|V|}{2}}}(ii),\ u_{a_{\frac{|V|}{2}+2}a_{\frac{|V|}{2}+1}}(ii); i = 1, ..., J\}$, $\{u_{a_{\frac{|V|}{2}}a_{\frac{|V|}{2}+1}}(ij),\ i \neq j\}$, and $\{u_{a_{\frac{|V|}{2}}a_{\frac{|V|}{2}+1}}(ii)\}$.

For instance if $|V| = 5$, $V = \{a_1,\ a_2,\ a_3,\ a_4,\ a_5\}$, the label invariant model $\boldsymbol{M}(G, \Gamma)$ with $\Gamma = \{(a_1a_5)\ (a_2a_4)\}$ and whose generating class associated to G is $A = \{\{a_1,\ a_2\},\ \{a_2,\ a_3\},\ \{a_3,\ a_4\},\ \{a_4,\ a_5\}\}$ is represented as a RGLL model with generating class A, vertices $V = (V_1,\ V_2,\ V_3)$ with $V_1 = \{a_1, a_5\}$, $V_2 = \{a_2, a_4\}$, and $V_3 = \{a_3\}$, and edge colour classes $\{u_{a_1a_2}(ij), u_{a_5a_4}(ij), i \neq j\}$, $\{u_{a_1a_2}(ii), u_{a_5a_4}(ii); i = 1, ..., J\}$, $\{u_{a_2a_3}(ij), u_{a_4a_3}(ij), i \neq j\}$, and $\{u_{a_2a_3}(ii), u_{a_4a_3}(ii); i = 1, ..., J\}$.

Example 4.7. (Cyclic permutation). Suppose that we have a graphical log-linear model corresponding to a cycle of length four or more and that we change the levels of the variables in a cyclic way, *i.e.* the value of one first variable becomes the value of the second, the value of the second variable becomes the value of the third, and so on until the value of the last variable takes the place of the first one forming a new configuration. We proceed in a similar way with this new configuration, and so on until we get the original configuration. We want the expected frequencies or probabilities to be the same for all configurations. This is called a cyclic permutation and it is a level invariant model.

For example, suppose that we have a model whose graph G is given by a four-cycle (figure 4.2) with vertices labeled as A, B, C, D and edges $\{A, B\}$, $\{B, C\}$, $\{C, D\}$, $\{D, A\}$. To get a cyclic permutation, we use a level invariant model $\boldsymbol{M'}(G, \Gamma')$, with $\Gamma' = \{(ADCB), (AC)(DB), (ABCD)\}$, because the graph G is triangle-free and because under the permutation set Γ' the following equality holds

$$m_{ABCD}(a,b,c,d) = m_{ABCD}(b,c,d,a) = m_{ABCD}(c,d,a,b) = m_{ABCD}(d,a,b,c),$$

which is the equality required for cyclic permutation. To obtain parameter restrictions that could be used to get such equalities, we write the expected frequencies as log-linear expansions according to the generating class. Once doing this process the equalities can be obtained if we let $u_A(i) = u_B(i) = u_C(i) = u_D(i)$, for all i, and if $u_{AB}(ij) = u_{BC}(ij) = u_{CD}(ij) = u_{DA}(ij)$, for all i and j, *i.e.* we obtain a RGLL model with vertex colour class $V = \{A, B, C, D\}$ and edge colour classes $\{u_{AB}(ij), u_{BC}(ij), u_{CD}(ij), u_{DA}(ij)\}$, $i, j = 1, ..., J$. The corresponding graph when all variables are binary is presented in figure 4.8. Note that in level invariant models we are not actually permuting variables and that the permutation set is not always used as in label invariant models to directly get the restrictions. These restrictions can be obtained once we write the model according to the parameters in such a way that we get the desired equalities.

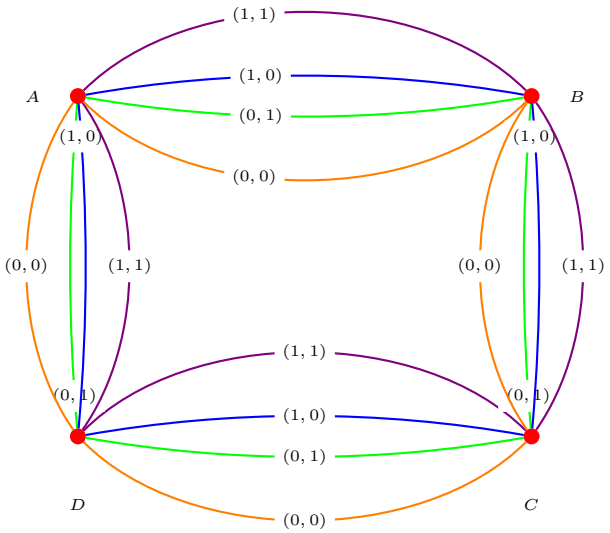


Figure 4.8: RGLL model with generating class $\{\{A, B\}, \{B, C\}, \{C, D\}, \{D, A\}\}$*, where all variables are binary, with vertices* $V = \{A, B, C, D\}$ *and first-order interactions set* $E = (E_1, E_2, E_3, E_4)$*,* $E_1 = \{u_{AB}(01), u_{BC}(01), u_{CD}(01), u_{DA}(01)\}$*,* $E_2 = \{u_{AB}(00), u_{BC}(00), u_{CD}(00), u_{DA}(00)\}$*,* $E_3 = \{u_{AB}(10), u_{BC}(10), u_{CD}(10), u_{DA}(10)\}$*, and* $E_4 = \{u_{AB}(11), u_{BC}(11), u_{CD}(11), u_{DA}(11)\}$*.*

For a cycle with $|V|$ vertices, $|V| \geq 4$, with the vertices representing variables with J levels each one, the cyclic permutation model can be represented as a RGLL model with generating class $A = \{\{a_1, a_2\}, \{a_2, a_3\}, ..., \{a_{|V|-1}, a_{|V|}\}, \{a_{|V|}, a_1\}\}$ in which all vertices

are in the same vertex colour class V and we have J^2 edge colour classes $\{u_{a_1a_2}(ij)$, $u_{a_2a_3}(ij)$, ..., $u_{a_{|V|-1}a_{|V|}}(ij)$, $u_{a_{|V|}a_1}(ij)\}$, $i,j= 1,...,J$.

Example 4.8. Consider a graphical log-linear model with generating class $\{\{A,B\}$, $\{B,C\}$, $\{C,D\}$, $\{D,A\}\}$ and corresponding graph G, figure 4.2. We analyze 8 different interesting label and level invariant models whose associated graph G is the same.

1) If we interchange at the same time two pair of neighbour vertices, for example A with B and C with D, *i.e.* using the label invariant model $\boldsymbol{M}(G,\Gamma)$ with $\Gamma= \{(AB)$ $(CD)\}$, we get that the graph of the orbits, G^*, corresponds to a graph in which the vertices in the first vertex orbit, A and B, have one colour and the vertices in the other orbit, C and D, have another colour. The edges not joining the neighbours we interchange, $\{A,D\}$ and $\{B,C\}$, which are in the same edge orbit, have the same colouring and the rest of the edges have different colours (figure 4.9). Observe that this graph is similar to the one presented in example 1.3 by Højsgaard and Lauritzen (2008, p. 1009) for some continuous data on a study of heredity of heads dimensions made by Frets (1921).

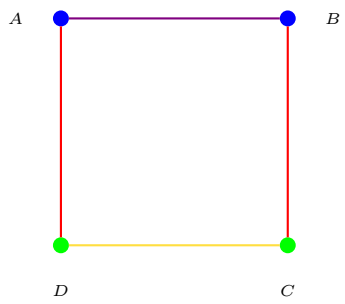


Figure 4.9: Vertex and edge orbits graph, G^*, for the label invariant model $\boldsymbol{M}(G,\Gamma)$ with generating class $\{\{A,B\}$, $\{B,C\}$, $\{C,D\}$, $\{D,A\}\}$, associated graph G, and permutation set $\Gamma= \{(AB)$, $(CD)\}$.

We represent the label invariant model with a RGLL model in which the main effects are equal for all levels of the vertices in the same orbit and the interactions of the same orbit are equal, first when the levels are equal, and then when they differ. So that the RGLL model has generating class $\{\{A,B\}$, $\{B,C\}$, $\{C,D\}$, $\{D,A\}\}$, vertex set $V = (V_1,V_2)$, with $V_1 = \{A,B\}$ and $V_2 = \{C,D\}$, and edge set $E= (E_1,E_2,E_3,E_4,E_5,E_6)$, with $E_1= \{u_{AD}(ii)$, $u_{BC}(ii)$, $i = 1,...,J\}$, $E_2= \{u_{AD}(ij)$, $u_{BC}(ij)$, $i \neq j$, $i,j = 1,...,J\}$, $E_3=\{u_{AB}(ii)$, $i = 1,...,J\}$, $E_4=\{u_{AB}(ij)$, $i \neq j$, $i,j = 1,...,J\}$, $E_5=\{u_{CD}(ii)$, $i = 1,...,J\}$, $E_6=\{u_{CD}(ij)$, $i \neq j$, $i,j = 1,...,J\}$. This RGLL model is given in figure 4.10 for the binary case.

The interpretation of this model is similar to the one in the continuous case. Here we can permute the variables A with B and C with D and the model is preserved, in the sense that it is scale invariant and that the distribution after the change is the same we had before the change was made. In the case of the Frets' heads data, which are continuous, A and D represent the measures for a first son and B and C the same measures for the other son, and the model presented by Højsgaard and Lauritzen (2008, p. 1008) means that there is symmetry between the two sons. If Frets' heads data were discrete, the label invariant model $\boldsymbol{M}(G,\Gamma)$ with $\Gamma = \{(AB)\ (CD)\}$ would have the same interpretation, symmetry between the sons.

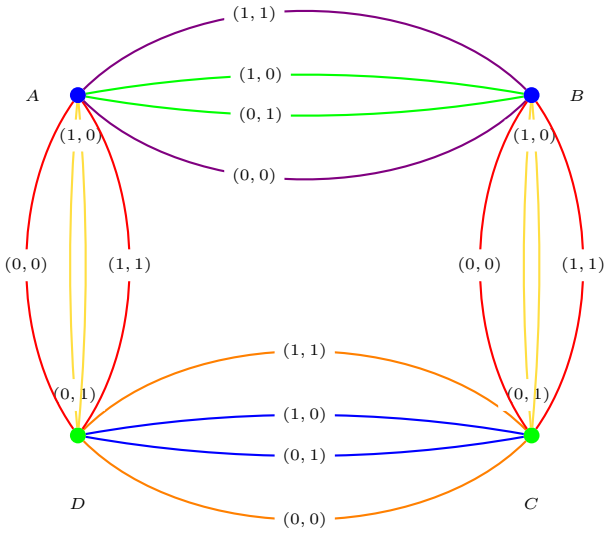


Figure 4.10: RGLL model with generating class $\{\{A,B\},\{B,C\},\{C,D\},\{D,A\}\}$*, where all variables are binary, with vertices* $V=(V_1,V_2)$*,* $V_1=\{A,B\}$ *and* $V_2=\{C,D\}$*, and first-order interactions set* $E=$ *(*E_1*,* E_2*,* E_3*,* E_4*,* E_5*,* E_6*),* $E_1=\{u_{AD}(00),\ u_{AD}(11),\ u_{BC}(00),\ u_{BC}(11)\}$*,* $E_2=\{u_{AD}(01),\ u_{AD}(10),\ u_{BC}(01),\ u_{BC}(10)\}$*,* $E_3=\{u_{AB}(00),\ u_{AB}(11)\}$*,* $E_4=\{u_{AB}(01),\ u_{AB}(10)\}$*,* $E_5=\{u_{CD}(00),\ u_{CD}(11)\}$*,* $E_6=\{u_{CD}(01),\ u_{CD}(10)\}$*.*

2) Assuming a level invariant model $\boldsymbol{M'}(G,\Gamma')$ with the same permutation set $\Gamma'=\{(AB)\ (CD)\}$, which is the model whose restrictions for the expected frequencies are

$$m_{ABCD}(a,b,c,d) = m_{ABCD}(b,a,d,c),$$

we may use the same RGLL or a less restricted model. In this case, we could use a RGLL model with the same vertex classes as in the model given in 1), $V=(V_1,V_2)$, with $V_1=\{A,B\}$ and $V_2=\{C,D\}$, but with fewer edge colour classes because we do not need to join the edges with the same levels in one class and the edges with different

levels in another. According to the model, we may use the following restrictions for the first-order interactions $u_{AB}(ij) = u_{AB}(ji)$, $i < j$, $u_{CD}(ij) = u_{CD}(ji)$, $i < j$, and $u_{BC}(ij) = u_{AD}(ij)$ or their corresponding edge colour classes $\{u_{AB}(ij), u_{AB}(ji)\}$, $i < j$; $\{u_{AB}(ii)\}$; $\{u_{CD}(ij), u_{CD}(ji)\}$, $i < j$; $\{u_{CD}(ii)\}$, and $\{u_{BC}(ij), u_{AD}(ij)\}$. Observe that the restrictions for $\{A, B\}$ and for $\{C, D\}$ are similar to the ones in symmetric models for two variables (Agresti, 2002a, p. 424). For example, in the binary case we get ten edge classes instead of six, which are: E_1= $\{u_{AD}(00), u_{BC}(00)\}$, E_2= $\{u_{AD}(01), u_{BC}(01)\}$, E_3= $\{u_{AD}(10), u_{BC}(10)\}$, E_4= $\{u_{AD}(11), u_{BC}(11)\}$, E_5=$\{u_{AB}(00)\}$, E_6= $\{u_{AB}(11)\}$, E_7= $\{u_{AB}(01), u_{AB}(10)\}$, E_8= $\{u_{CD}(00)\}$, E_9= $\{u_{CD}(11)\}$, and E_{10}= $\{u_{CD}(01), u_{CD}(10)\}$.

Observe that in these particular cases in which we have a label invariant model $\boldsymbol{M}(G, \Gamma)$ and we get the corresponding level invariant model $\boldsymbol{M'}(G, \Gamma')$ using $\Gamma = \Gamma'$, that is when Γ' is an automorphism, it can still be helpful to use the orbits because it can help to know which parameters to equate to obtain a RGLL model. The vertex orbits are the vertex classes and in this case once we get the edge orbits we observe that the edge $\{A, D\}$ was transformed into $\{B, C\}$, $\{A, B\}$ into $\{B, A\}$, and $\{C, D\}$ into $\{D, C\}$, and the corresponding parameters are equated for all level permutation (i, j), for example $u_{BC}(ij)$ is equated with $u_{AD}(ij)$, getting the RGLL model shown in the previous paragraph. It is important to remark that we always need to be careful in the order in which the vertices forming edges are transformed, which may (even though it should not) cause confusion because usually the order of the vertices is irrelevant when defining edges, for example in this case the edge $\{B, C\}$ turns into $\{A, D\}$ and not into $\{D, A\}$ because under the permutation $(AB)(CD)$ B is replaced with A and C with D. This order is important to get the right restrictions for the interactions.

3) Another example is a level invariant model for two neighbour vertices. Suppose that there are two neighbour vertices α and β, *i.e.* there is an edge $\{\alpha, \beta\}$, and that there are edges $\{\delta, \alpha\}$ and $\{\beta, \gamma\}$. We define a level invariant model $\boldsymbol{M'}(G, \Gamma')$ with Γ'= $\{(\alpha\beta)\}$. For example, if $\alpha = A$, $\beta = B$, $\gamma = C$, and $\delta = D$ the model satisfies

$$m_{ABCD}(a, b, c, d) = m_{ABCD}(b, a, c, d).$$

To obtain these equalities we could use a RGLL model with $V = (V_1, V_2, V_3)$ where V_1= $\{\alpha, \beta\}$, V_2= $\{\gamma\}$, and V_3= $\{\delta\}$, and the following first-order interactions restrictions

$$u_{\alpha\beta}(ij) = u_{\alpha\beta}(ji),\ i < j;\ u_{\beta\gamma}(1c) = u_{\beta\gamma}(2c) = ... = u_{\beta\gamma}(Jc),\ c = 1, ..., J;\ \text{and}$$

$$u_{\delta\alpha}(d1) = u_{\delta\alpha}(d2) = ... = u_{\delta\alpha}(dJ),\ d = 1, ..., J;$$

or their corresponding edge classes $\{u_{\alpha\beta}(ij), u_{\alpha\beta}(ji)\}$, $i < j$; $\{u_{\alpha\beta}(ii)\}$, $i = 1, ...J$; $\{u_{\beta\gamma}(1c), u_{\beta\gamma}(2c), ..., u_{\beta\gamma}(Jc)\}$, $c = 1, ..., J$; $\{u_{\delta\alpha}(d1), u_{\delta\alpha}(d2), ..., u_{\delta\alpha}(dJ)\}$, $d = 1, ..., J$; and the remaining first-order interactions in different atomic classes. The corresponding

graph when all variables are binary with categories 0 and 1; with $\alpha = A$, $\beta = B$, $\gamma = C$, and $\delta = D$; and where all atomic classes are coloured in black is presented in figure 4.11.

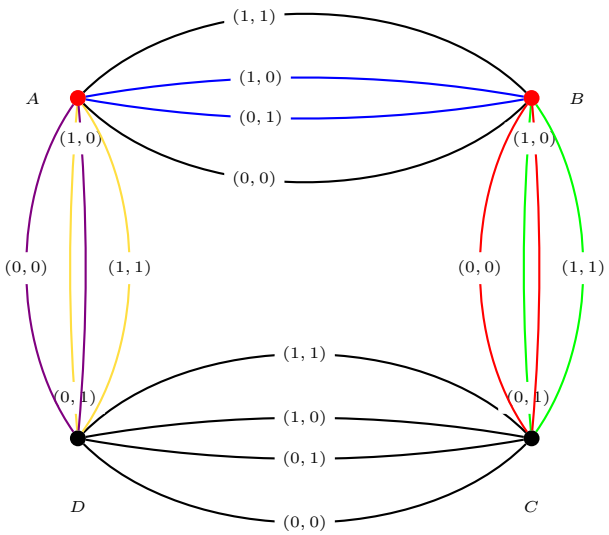


Figure 4.11: RGLL model with generating class $\{\{A, B\}, \{B, C\}, \{C, D\}, \{D, A\}\}$*, where all variables are binary, with vertices* $V = (V_1, V_2, V_3)$*,* $V_1 = \{A, B\}$*,* $V_2 = \{C\}$*, and* $V_3 = \{D\}$*, and first-order interactions set* $E = (E_1, E_2, E_3, E_4, E_5, E_6, E_7, E_8, E_9, E_{10}, E_{11})$*,* $E_1 = \{u_{AB}(01), u_{AB}(10)\}$*,* $E_2 = \{u_{AB}(00)\}$*,* $E_3 = \{u_{AB}(11)\}$*,* $E_4 = \{u_{BC}(00), u_{BC}(10)\}$*,* $E_5 = \{u_{BC}(01), u_{BC}(11)\}$*,* $E_6 = \{u_{DA}(00), u_{DA}(01)\}$*,* $E_7 = \{u_{DA}(10), u_{DA}(11)\}$*,* $E_8 = \{u_{CD}(00)\}$*,* $E_9 = \{u_{CD}(11)\}$*,* $E_{10} = \{u_{CD}(01)\}$*, and* $E_8 = \{u_{CD}(10)\}$*. All atomic classes in black.*

Observe that a model permuting neighbour vertices does not make sense for a label invariant case because when we change two neighbour vertices we do not get an automorphism of the graph, for example after the permutation $(\alpha\beta) = (AB)$ we obtain a new edge $\{\alpha, \gamma\} = \{A, C\}$ we did not have before (figure 4.12).

Observe also that this kind of model can be generalized to any cycle of length four or more. Given a n-cycle, $n \geq 4$, including the edges $\{\alpha, \beta\}$, $\{\beta, \gamma\}$, and $\{\delta, \alpha\}$, the level invariant model $\boldsymbol{M'}(G, \Gamma')$ with $\Gamma' = \{(\alpha\beta)\}$ can be represented as the RGLL model with the same restrictions given before for the four-cycle.

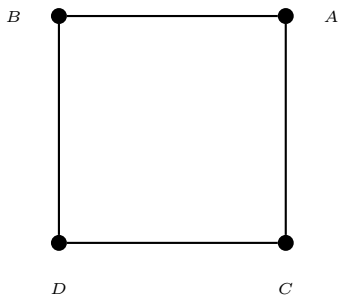


Figure 4.12: Graph obtained after applying the permutation $(\alpha\beta)=(AB)$. *This graph is not isomorphic to G, the four-cycle shown in figure 4.2.*

4) A fourth example corresponds to a label invariant model $\boldsymbol{M}(G,\Gamma)$ with $\Gamma=\{(AC)\ (BD)\}$, which is a model interchanging opposite vertices in the graph at the same time. For this case the vertex orbits are $\{A,C\}$ and $\{B,D\}$. In consequence we obtain a RGLL model in which V is partitioned into $V_1=\{A,C\}$ and $V_2=\{B,D\}$. The edges classes in the RGLL model, which can be obtained from the edges orbits, $\{\{A,B\},\ \{C,D\}\}$ and $\{\{B,C\},\ \{D,A\}\}$, are $E_1=\{u_{AB}(ii),\ u_{CD}(ii),\ i=1,...,J\}$, $E_2=\{u_{AB}(ij),\ u_{CD}(ij),\ i\neq j,\ i,j=1,...,J\}$, $E_3=\{u_{BC}(ii),\ u_{DA}(ii),\ i=1,...,J\}$, $E_4=\{u_{BC}(ij),\ u_{DA}(ij),\ i\neq j,\ i,j=1,...,J\}$. The graph for the binary case is given in figure 4.13.

5) A level invariant model $\boldsymbol{M'}(G,\Gamma')$ with the same permutation set $\Gamma'=\Gamma$ can be obtained using a less restricted RGLL model. As stated before, we could use the orbits to get the classes. The vertex classes are the same as before. Additionally, we need the following restrictions $u_{AB}(ij)=u_{CD}(ij)$ and $u_{BC}(ij)=u_{DA}(ij)$, $i,j=1,...,J$ with their corresponding edge colour classes $\{u_{AB}(ij),\ u_{CD}(ij)\}$ and $\{u_{BC}(ij),\ u_{DA}(ij)\}$, $i,j=1,...,J$.

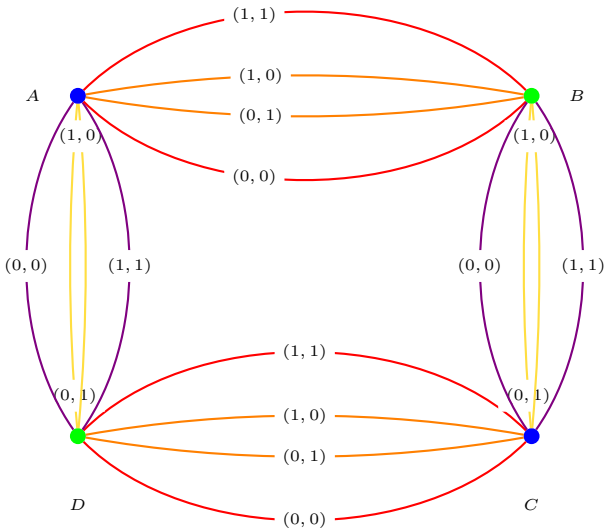


Figure 4.13: RGLL model with generating class $\{\{A,B\}, \{B,C\}, \{C,D\}, \{D,A\}\}$*, where all variables are binary, with vertices* $V = (V_1, V_2)$*,* $V_1 = \{A,C\}$ *and* $V_2 = \{B,D\}$*, and first-order interactions set* $E = (E_1, E_2, E_3, E_4)$*,* $E_1 = \{u_{AB}(00),\ u_{CD}(00),\ u_{AB}(11),\ u_{CD}(11)\}$*,* $E_2 = \{u_{AB}(01),\ u_{CD}(01),\ u_{AB}(10),\ u_{CD}(10)\}$ *,* $E_3 = \{u_{BC}(00),\ u_{DA}(00),\ u_{BC}(11),\ u_{DA}(11)\}$*,* $E_4 = \{u_{BC}(01),\ u_{DA}(01),\ u_{BC}(10),\ u_{DA}(10)\}$.

6) A sixth example corresponds to a label invariant model in which we only permute two opposite vertices, for example the label invariant model $\boldsymbol{M}(G,\Gamma)$ using the permutation set $\Gamma = \{(AC)\}$. In this case the corresponding associated RGLL model has three vertex classes $V_1 = \{A,C\}$, $V_2 = \{B\}$, and $V_3 = \{D\}$, and four edge classes $E_1 = \{u_{AB}(ii),\ u_{CB}(ii),\ i = 1,...,J\}$, $E_2 = \{u_{AB}(ij),\ u_{CB}(ij),\ i \neq j,\ i,j = 1,...,J\}$, $E_3 = \{u_{CD}(ii),\ u_{AD}(ii),\ i = 1,...,J\}$, $E_4 = \{u_{CD}(ij),\ u_{AD}(ij),\ i \neq j,\ i,j = 1,...,J\}$. The graph for the binary case is given in figure 4.14.

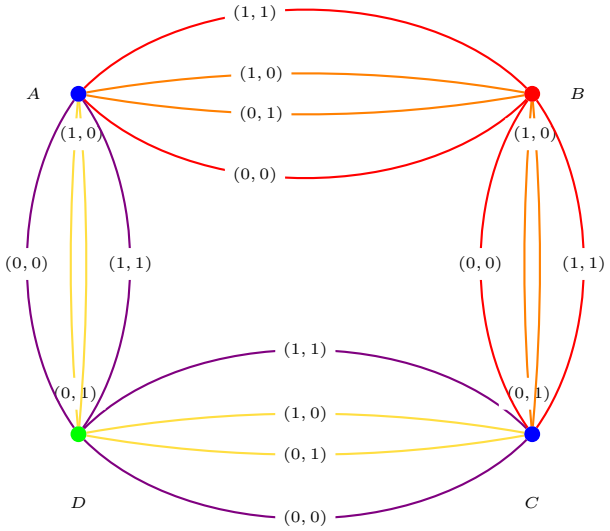


Figure 4.14: RGLL model with generating class $\{\{A,B\},\{B,C\},\{C,D\},\{D,A\}\}$*, where all variables are binary, with vertices* $V=(V_1,V_2,V_3)$*,* $V_1=\{A,C\}$*,* $V_2=\{B\}$*, and* $V_3=\{D\}$ *and first-order interactions set* $E=(E_1, E_2, E_3, E_4)$*,* $E_1=\{u_{AB}(00), u_{CB}(00), u_{AB}(11), u_{CB}(11)\}$*,* $E_2=\{u_{AB}(01), u_{AB}(10), u_{CB}(01), u_{CB}(10)\}$*,* $E_3=\{u_{CD}(00), u_{AD}(00), u_{CD}(11), u_{AD}(11)\}$*,* $E_4=\{u_{CD}(01), u_{CD}(10), u_{AD}(01), u_{AD}(10)\}$*.*

7) A level invariant model $\boldsymbol{M'}(G,\Gamma')$ with the same permutation set $\Gamma'=\Gamma$ can be obtained using a less restricted RGLL model. The vertex classes are the same. Additionally, we need the following restrictions $u_{AB}(ij)=u_{CB}(ij)$ and $u_{CD}(ij)=u_{AD}(ij)$, $i,j=1,...,J$ with their corresponding edge colour classes $\{u_{AB}(ij), u_{CB}(ij)\}$ and $\{u_{CD}(ij), u_{AD}(ij)\}$, $i,j=1,...,J$.

8) Finally, suppose that for the same graphical model with graph G we have a vertex and edge orbits graph, G^*, coloured as in figure 4.15. We want to know if it represents a label invariant model $\boldsymbol{M}(G,\Gamma)$ determined by Γ, $\Gamma\subseteq Aut(G)=\{Id, (AC), (BD), (AC)(BD), (AB)(CD), (AD)(BC), (ABCD), (ADCB)\}$. Without considering the empty set, there are $2^8-1=255$ subsets of $Aut(G)$, each of these subsets generates an orbits graph G^*. Different subsets can generate the same graph G^*. After analyzing all subsets Γ of $Aut(G)$ by getting the orbits and coloured graph G^* associated to these orbits, we derived 8 different colourings which are shown in table 4.3. On the other hand, the number of different vertex and edge colourings of G is 81, this means that there are colourings of G, colourings that possibly could be identified with vertex and edge orbits, that do not represent a label invariant model. In particular the graph shown in figure 4.15 corresponds to a vertex and edge colouring of G not contained in the set of different orbits graphs G^* that can be obtained from any $\Gamma\subseteq Aut(G)$, table 4.3, and in consequence the graph does not belong to a label invariant model.

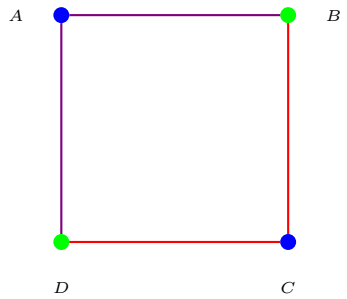


Figure 4.15: Vertex and edge orbits graph, G^*, for the model with generating class $\{\{A,B\}, \{B,C\}, \{C,D\}, \{D,A\}\}$, $\log m(i,j,k,l) = u + u_A(i) + u_B(j) + u_C(k) + u_D(l) + u_{AB}(ij) + u_{BC}(jk) + u_{CD}(kl) + u_{DA}(li)$ for which there is not an associated label invariant model, illustrating that not any coloured graph represents a label invariant model.

This example shows that there are coloured graphs, which can be seen as representations of vertex and edge orbits, that can not be used to derive a label invariant model. In the continuous case, a similar colouring for this graph is presented by Højsgaard and Lauritzen (2008) to show a graph whose associated model is not a RCOP model. Then in both cases, continuous and discrete, not all colourings represent invariant models, label invariant for the discrete case and RCOP for the continuous case. Properties that coloured graphs should satisfy to represent RCOP models are currently being studied and identified by Neufeld (2009).

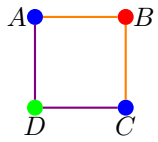
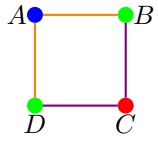
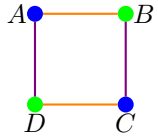
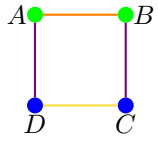
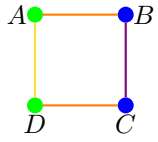
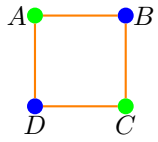
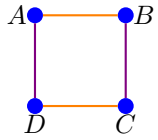
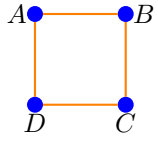
Permutation set Γ	Orbits graphs G^*
$\{(AC)\}$	
$\{(BD)\}$	
$\{(AC)(BD)\}$	
$\{(AB)(CD)\}$	
$\{(AD)(BC)\}$	
$\Gamma_1 = \{(AC),(BD)\}$, $\Gamma_2 = \{(AC),(AC)(BD)\}$, $\Gamma_3 = \{(BD),(AC)(BD)\}$ and all Γ, $\lvert\Gamma\rvert = 3$, with Γ_1, Γ_2, or Γ_3 and $Id \subset \Gamma$	
$\Gamma_4 = \{(AC)(BD),(AB)(CD)\}$, $\Gamma_5 = \{(AC)(BD),(AD)(BC)\}$, $\Gamma_6 = \{(AB)(CD),(AD)(BC)\}$ and all Γ, $\lvert\Gamma\rvert = 4$, with Γ_4, Γ_5, or Γ_6 and $Id \subset \Gamma$	
Any other Γ	

Table 4.3: Subsets Γ of $Aut(G)$ and their associated orbits graphs G^* for the graphical log-linear model with generating class $\{\{A,B\},\{B,C\},\{C,D\},\{D,A\}\}$ and associated graph G.

Example 4.9. Consider a model with generating class $\{\{A, D\}, \{B, D\}, \{C, D\}\}$ (figure 4.16). We can define a level invariant model $\boldsymbol{M'}(G, \Gamma')$ with $\Gamma' = \{(ABC), (ACB)\}$.

This is a model in which

$$m_{ABCD}(a, b, c, d) = m_{ABCD}(c, a, b, d) = m_{ABCD}(b, c, a, d).$$

To get these equalities, we can use a RGLL model in which the vertex set is partitioned into the vertex classes $V_1 = \{A, B, C\}$ and $V_2 = \{D\}$ and the edge set into J^2 edge classes given by $\{u_{AD}(ij),\ u_{BD}(ij),\ u_{CD}(ij)\},\ i, j = 1, ..., J$. This model is similar to cyclic permutation presented in example 4.7 for A, B, and C.

For the same graph, we define a label invariant model $\boldsymbol{M}(G, \Gamma)$ with $\Gamma = \{per(ABC)\}$, where $per(ABC)$ denotes all possible permutations of A, B, and C. In this case, the vertex orbits and vertex classes are given by $V_1 = \{A, B, C\}$ and $V_2 = \{D\}$, and the unique edge orbit is $\{\{A, D\}, \{B, D\}, \{C, D\}\}$, as a consequence the edge classes are $E_1 = \{u_{AD}(ij),\ u_{BD}(ij),\ u_{CD}(ij),\ i \neq j,\ i, j = 1, ..., J\}$ and $E_2 = \{u_{AD}(ii),\ u_{BD}(ii),\ u_{CD}(ii),\ i = 1, ..., J\}$, obtaining again a RGLL model.

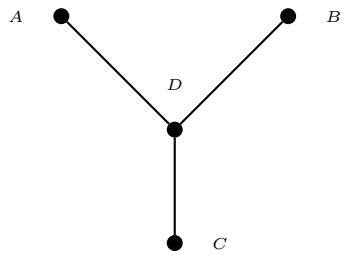


Figure 4.16: Graphical log-linear model with generating class $\{\{A, D\}, \{B, D\}, \{C, D\}\}$, $\log m(i, j, k, l) = u + u_A(i) + u_B(j) + u_C(k) + u_D(l) + u_{AD}(il) + u_{BD}(jl) + u_{CD}(kl)$.

Example 4.10. Consider a model with generating class $\{\{A, B\}, \{A, C\}, \{B, D\}, \{B, E\}, \{C, F\}, \{C, G\}\}$ (figure 4.17), and associated graph G' which is a tree. We define a level invariant model $\boldsymbol{M'}(G', \Gamma')$ with $\Gamma' = \{(DF)(EG)\}$.

This model allows to permute the values of D with the ones in F and the values of E with the ones in G. Under this model the following equality is satisfied

$$m_{ABCDEFG}(a, b, c, d, e, f, g) = m_{ABCDEFG}(a, b, c, f, g, d, e). \tag{4.7}$$

Writing the expected frequencies according to the model parameters

$$\log m_{ABCDEFG}(a, b, c, d, e, f, g) = u + u_A(a) + u_B(b) + u_C(c) + u_D(d) + u_E(e) + u_F(f) +$$

$$u_G(g) + u_{AB}(ab) + u_{AC}(ac) + u_{BD}(bd) + u_{BE}(be) + u_{CF}(cf) + u_{CG}(cg);$$

and taking into account the permutation set Γ', the expected frequencies can be expressed as

$$\log m_{ABCDEFG}(a,b,c,f,g,d,e) = u + u_A(a) + u_B(b) + u_C(c) + u_D(f) + u_E(g) + u_F(d) +$$

$$u_G(e) + u_{AB}(ab) + u_{AC}(ac) + u_{BD}(bf) + u_{BE}(bg) + u_{CF}(cd) + u_{CG}(ce);$$

we observe that one way to restrict the parameters to satisfy equality (4.7) is by partitioning the vertex set into the vertex classes $V_1 = \{D, F\}$, $V_2 = \{E, G\}$, $V_3 = \{A\}$, $V_4 = \{B\}$, and $V_5 = \{C\}$. Additionally, in order to satisfy (4.7) the first-order interactions should satisfy the following restrictions

$$u_{BD}(b1) = ... = u_{BD}(bJ),\ b = 1, ..., J;$$

$$u_{BE}(b1) = ... = u_{BE}(bJ),\ b = 1, ..., J;$$

$$u_{CF}(c1) = ... = u_{CF}(cJ),\ c = 1, ..., J;$$

$$u_{CG}(c1) = ... = u_{CG}(cJ),\ c = 1, ..., J.$$

This means that the parameters $u_{AB}(ij)$ and $u_{AC}(ij)$ are unrestricted since the variables corresponding to these interactions for this model are not permuted in any way, so that each parameter for all possible permutation of their levels must be in a different atomic class, and that we have J edge classes for each parameter u_{BD}, u_{BE}, u_{CF}, and u_{CG}. For example, for u_{BD} we get the classes $\{u_{BD}(11)$, $u_{BD}(12)$ ...., $u_{BD}(1J)\}$,...., $\{u_{BD}(J1)$, $u_{BD}(J2)$ ...., $u_{BD}(JJ)\}$. Then we have a RGLL model with V= $(V_1, V_2, V_3, V_4, V_5)$ and first-order interaction set partitioned into the edge classes $\{u_{BD}(b1)$,...., $u_{BD}(bJ)\}$, $b = 1, ..., J$; $\{u_{BE}(b1)$,...., $u_{BE}(bJ)\}$, $b = 1, ..., J$; $\{u_{CF}(c1)$,...., $u_{CF}(cJ)\}$, $c = 1, ..., J$; $\{u_{CG}(c1)$,...., $u_{CG}(cJ)\}$, $c = 1, ..., J$; $\{u_{AB}(ij)\}$, for all i, j; and $\{u_{AC}(ij)\}$ for all i, j.

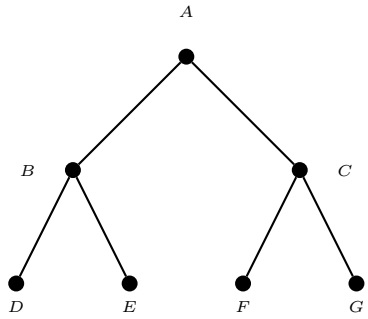


Figure 4.17: Graphical log-linear model with generating class $\{\{A, B\}$, $\{A, C\}$, $\{B, D\}$, $\{B, E\}$, $\{C, F\}$, $\{C, G\}\}$, $\log m(a, b, c, d, e, f, g) = u + u_A(a) + u_B(b) + u_C(c) + u_D(d) + u_E(e) + u_F(f) + u_G(g) + u_{AB}(ab) + u_{AC}(ac) + u_{BD}(bd) + u_{BE}(be) + u_{CF}(cf) + u_{CG}(cg)$.

This model is interesting because it may be interpreted in different ways, we show here three interpretations. a) If the variable associated to A corresponds to a categorical variable measured to someone one year, B and C correspond to measures for the same person of the same variable in the following year, which was divided into two semesters, so that B corresponds to the first semester and C corresponds to the second, and variables D, E, F, and G correspond to the same variable measured to the same person again in a third year divided accordingly into four trimesters, then we have a panel data, *i.e.* a time series unequally spaced for each person. The corresponding contingency table has in each cell the number of people with the same response patterns through time. The level invariant model means that the measures in one semester depend on the measures in the previous year, that the same happens with the trimester information which depends on the semesters of the previous year, the two first trimesters depend on the first semester and the two last trimesters depend on the second semester, and that we can interchange the values of the last two trimesters with the ones of the two first trimesters getting the same expected frequencies, so that the semester in which the trimester measures belong is not relevant.

b) A second interpretation could be if we had information concerning family members or any other kind of cluster. For example, A represents the value of a categorical variable measured for a grandfather, for example social class, B represents the same value for a first descendant and C for a second descendant. Finally, D and E are the values corresponding to the first descendant sons and F and G correspond to the values for the second descendant sons. Each cell in the corresponding contingency table represents the number of families with a particular response pattern. The meaning of this model is that the variables are associated with those of the ascendants and that the values taken by the corresponding variables for the first grandsons and the second grandsons can be interchanged, so that for a determined response pattern it does not matter which parents the second generation have.

c) Finally, we could think that the data correspond to an item-response data, in which each variable corresponds to a different item represented by the variables A, B, C, D, E, F, and G, all of them having the same options, for example true and false. Additionally, we suppose that all subjects respond all items. The corresponding contingency table has cells representing the number of subjects with an special response pattern. The meaning of the level invariant model is that even though the items associated to D and E and the items associated to F and G depend on different items, their values can be interchanged, respectively, so that the response pattern for this pair of items is similar.

Observe that we can not get a label invariant model $\boldsymbol{M}(G', \Gamma)$ using $\Gamma = \Gamma'$, because the only element contained in Γ' is not an automorphism of the graph. However, we could get a label invariant model $\boldsymbol{M}(G', \Gamma)$ by interchanging at the same time the labels

B with C, D with F, and E with G, *i. e.* interchanging the two different branches of the tree, so that we have $\Gamma = \{(BC)(DF)(EG)\}$. In this case, the vertex orbits and in consequence vertex classes are $V_1 = \{D, F\}$, $V_2 = \{E, G\}$, $V_3 = \{B, C\}$, and $V_4 = \{A\}$. The edge orbits are $\{\{A, B\},\ \{A, C\}\}$, $\{\{B, D\},\ \{C, F\}\}$, and $\{\{B, E\},\ \{C, G\}\}$, so that the corresponding edge classes are E_1= $\{u_{AB}(ij),\ u_{AC}(ij),\ i \neq j,\ i, j = 1, ..., J\}$, E_2= $\{u_{AB}(ii),\ u_{AC}(ii),\ i = 1, ..., J\}$, E_3= $\{u_{BD}(ij),\ u_{CF}(ij),\ i \neq j,\ i, j = 1, ..., J\}$, E_4= $\{u_{BD}(ii),\ u_{CF}(ii),\ i = 1, ..., J\}$, E_5= $\{u_{BE}(ij),\ u_{CG}(ij),\ i \neq j,\ i, j = 1, ..., J\}$, and E_6= $\{u_{BE}(ii),\ u_{CG}(ii),\ i = 1, ..., J\}$ and we obtain a RGLL model with those vertex and edge classes.

Example 4.11. Consider a level invariant model $\boldsymbol{M}'(G, \Gamma')$ with generating class $\{\{A, B\},\ \{A, C\},\ \{B, D\},\ \{C, D\},\ \{C, E\},\ \{D, F\},\ \{E, F\}\}$ and associated graph G (figure 4.18), with Γ'= $\{(AE)(BF)\}$. That is, we have a model in which

$$m_{ABCDEF}(a, b, c, d, e, f) = m_{ABCDEF}(e, f, c, d, a, b). \tag{4.8}$$

Writing the expected frequencies according to the model parameters

$$\log m(a, b, c, d, e, f) = u + u_A(a) + u_B(b) + u_C(c) + u_D(d) + u_E(e) + u_F(f) + u_{AB}(ab) +$$

$$u_{BD}(bd) + u_{CD}(cd) + u_{AC}(ac) + u_{DF}(df) + u_{EF}(ef) + u_{EC}(ec);$$

and taking into account the permutation set Γ', the expected frequencies can be expressed as

$$\log m(e, f, c, d, a, b) = u + u_A(e) + u_B(f) + u_C(c) + u_D(d) + u_E(a) + u_F(b) + u_{AB}(ef) +$$

$$u_{BD}(fd) + u_{CD}(cd) + u_{AC}(ec) + u_{DF}(db) + u_{EF}(ab) + u_{EC}(ac).$$

In order to satisfy equality (4.8), we have to equate the corresponding parameters and we may use a RGLL model with vertex classes $V_1 = \{A, E\}$, $V_2 = \{B, F\}$, $V_3 = \{C\}$, $V_4 = \{D\}$ and edge classes given by $\{u_{AC}(ij),\ u_{EC}(ij)\}$, $i, j = 1, ..., J$, $\{u_{BD}(ij),\ u_{FD}(ij)\}$, $i, j = 1, ..., J$, $\{u_{AB}(ij),\ u_{EF}(ij)\}$, $i, j = 1, ..., J$, and $\{u_{CD}(ij)\}$, $i, j = 1, ..., J$. Observe that if we considered the permutation set Γ' as a set interchanging labels, we would get a set formed by an automorphism of the graph, so that in this particular case, even though we are not talking about a label invariant model, we could use if we wanted, the vertex and edge orbits to obtain the restrictions necessary to get the required expected frequencies equalities for the RGLL model. In this case, the vertex orbits $\{A, E\}$, $\{B, F\}$, $\{C\}$, and $\{D\}$ correspond to the vertex classes V_1, V_2, V_3, and V_4, respectively, and the edge orbits $\{\{A, C\},\ \{E, C\}\}$, $\{\{B, D\},\ \{F, D\}\}$, $\{\{A, B\},\ \{E, F\}\}$, and $\{\{C, D\}\}$ are used to get the edge classes. These are formed by the first-order interactions corresponding to the edges in the same orbit for all possible permutation of the levels. At the end we obtain the edge classes already presented.

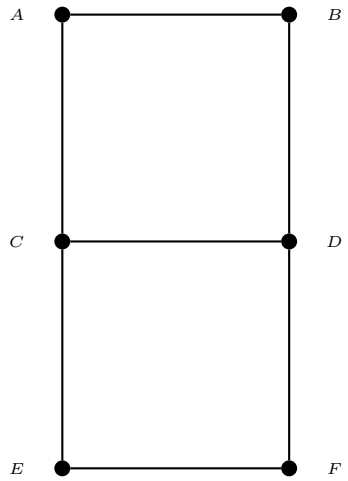


Figure 4.18: Graphical log-linear model with generating class $\{\{A,B\}, \{A,C\}, \{B,D\}, \{C,D\}, \{C,E\}, \{D,F\}, \{E,F\}\}$, $\log m(a,b,c,d,e,f) = u + u_A(a) + u_B(b) + u_C(c) + u_D(d) + u_E(e) + u_F(f) + u_{AB}(ab) + u_{BD}(bd) + u_{CD}(cd) + u_{AC}(ac) + u_{DF}(df) + u_{EF}(ef) + u_{EC}(ec)$.

It is evident, using the orbits already obtained and scale invariance restrictions, that the label invariant model using the same symmetry group, $\Gamma = \{(AE)(BF)\}$, is the RGLL model with vertex classes $V_1 = \{A, E\}$, $V_2 = \{B, F\}$, $V_3 = \{C\}$, and $V_4 = \{D\}$, and edge classes $E_1 = \{u_{AC}(ij),\ u_{EC}(ij),\ i \neq j,\ i,j = 1,...,J\}$, $E_2 = \{u_{AC}(ii),\ u_{EC}(ii),\ i = 1,...,J\}$, $E_3 = \{u_{BD}(ij),\ u_{FD}(ij),\ i \neq j,\ i,j = 1,...,J\}$, $E_4 = \{u_{BD}(ii),\ u_{FD}(ii),\ i = 1,...,J\}$, $E_5 = \{u_{AB}(ij),\ u_{EF}(ij),\ i \neq j,\ i,j = 1,...,J\}$, $E_6 = \{u_{AB}(ii),\ u_{EF}(ii),\ i = 1,...,J\}$, $E_7 = \{u_{CD}(ij),\ i \neq j,\ i,j = 1,...,J\}$, and $E_8 = \{u_{CD}(ii),\ i = 1,...,J\}$.

Suppose that the variables represent a categorical variable measured at different times, A and B representing measures in a first year, for example in two different semesters, C and D in a second year, and E and F in a third year, the label invariant model would mean that a) the first and third years are conditionally independent given the second one, this is due to the conditional independences of the graphical model alone, and b) that we could interchange the first and third years, permute the level set associated to the variables, or both and still preserve the model, which implies that $m_{ABCDEF}(a,b,c,d,e,f) = m_{ABCDEF}(e,f,c,d,a,b)$ and consequently that the distribution is preserved after the permutation.

4.2 Applications of label and level invariant models

In this section we analyze two different contingency tables by using label and level invariant models to illustrate how these models can be fitted as well as their interpretation for data in which such models make sense. The first example corresponds to a discretization of the data analyzed in the continuous case by Højsgaard and Lauritzen (2008) using RCOP models and the second example was analyzed by Drton and Richardson (2008) using what they call binary models for marginal independence.

Example 4.12. Frets' heads data (Frets, 1921) correspond to continuous observations containing the length and breadth of the heads of 25 pairs of first and second sons. We have L_1^*, L_2^*, B_1^*, and B_2^*, corresponding to the head length of the first and second son, and head breadth of the first and second son, respectively. These data were reported in Mardia *et al.* (1979, p. 121) and can be also found in Whittaker (1990, p. 255, 265). We derived discrete variables from the continuous variables, deriving a binary variable L_1, L_2, B_1, and B_2 from each continuous variable L_1^*, L_2^*, B_1^*, and B_2^*. These variables take the value 0 if the value taken by the corresponding continuous variable is less than the mean obtained from the corresponding variable values and 1 otherwise, we emphasize that the cut value used is somehow arbitrary.

Once the variables are split, we get the contingency table summarizing the values taken for each observation shown in table 4.4.

L_1	B_1	L_2			
		0		1	
		B_2			
		0	1	0	1
0	0	7	0	1	1
	1	0	0	2	0
1	0	0	0	1	3
	1	2	1	0	7

Table 4.4: Length and breadth of heads binary variables associated to first and second sons measures, L_1, L_2, B_1, and B_2, in Frets' heads data.

We used the selection methods available in *MIM* to find a generating class corresponding to a graphical log-linear model that fitted the data and that simultaneously had an associated triangle-free graph. Some selection methods obtained models whose graphs were not triangle-free; however, one of the models that fitted better according to the deviance corresponded to a graphical log-linear model with associated triangle-free graph. The selected model was obtained using an unrestricted Backward selection method. The term *unrestricted* included in the name of the selection method means that both decomposable and non-decomposable models are allowed in the selection process. Starting from the saturated model, we obtained a graphical log-linear model with generating class $\{\{L_1, B_2\}, \{B_2, L_2\}, \{L_2, B_1\}, \{L_1, B_1\}\}$, whose deviance is 11.17 with 7 degrees of freedom and a p-value of 0.13. So that, in general, for example for a significance level of 0.05, we do not reject the null hypothesis that the model fits the data. This model has the advantage that their associated label and level invariant models have logical interpretations.

The graphical Gaussian model that fitted well in the continuous case, the model with generating class $\{\{L_1,\ L_2\},\ \{B_2,\ L_2\},\ \{B_1,\ B_2\},\ \{L_1,\ B_1\}\}$(Whittaker, 1990, p. 254-259 and Højsgaard and Lauritzen, 2008), does not fit the data in the discrete case with this arbitrary discretization, which is not surprising given that we are using a specific discretization. Its deviance is 18.55 with 7 degrees of freedom and a p-value of 0.01.

We selected the model generated by $\{\{L_1,\ B_2\},\ \{B_2,\ L_2\},\ \{L_2,\ B_1\},\ \{L_1,\ B_1\}\}$, because it fits well the data and it serves as a way to illustrate both label and level invariant models. The graph associated to this model, G, is the four-cycle presented in figure 4.19.

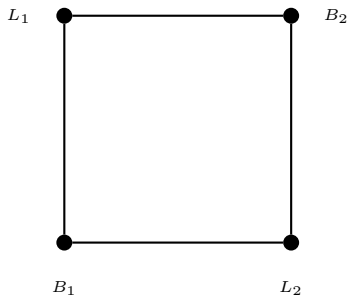


Figure 4.19: Graphical log-linear model with generating class $\{\{L_1, B_2\},\ \{B_2, L_2\},\ \{L_2, B_1\},\ \{L_1, B_1\}\}$ *and associated graph* G.

We examine four label and one level invariant models chosen because of their interesting interpretation or because, like cyclic permutation, have been previously defined in section **4.1**.

i) A first label invariant model with an interesting interpretation we examine is one in which we interchange L_1 and L_2, *i.e.* a label invariant model $M(G, \Gamma)$ in which $\Gamma = \{(L_1L_2)\}$ and G is the graph associated to the model generated by $A = \{\{L_1,\ B_2\},\ \{B_2,\ L_2\},\ \{L_2,\ B_1\},\ \{L_1,\ B_1\}\}$, with $A = C(G)$, the set of cliques of G. In this case, we get a RGLL model whose graph is the one given in figure 4.20. This is a RGLL model with vertex classes

$$V_1 = \{L_1, L_2\},\ V_2 = \{B_1\},\ V_3 = \{B_2\};$$

and edge classes

$$\begin{aligned}
E_1 &= \{u_{L_1B_2}(01), u_{L_1B_2}(10), u_{L_2B_2}(01), u_{L_2B_2}(10)\},\\
E_2 &= \{u_{L_1B_2}(00), u_{L_1B_2}(11), u_{L_2B_2}(00), u_{L_2B_2}(11)\},\\
E_3 &= \{u_{L_1B_1}(01), u_{L_1B_1}(10), u_{L_2B_1}(01), u_{L_2B_1}(10)\},\\
E_4 &= \{u_{L_1B_1}(00), u_{L_1B_1}(11), u_{L_2B_1}(00), u_{L_2B_1}(11)\}.
\end{aligned}$$

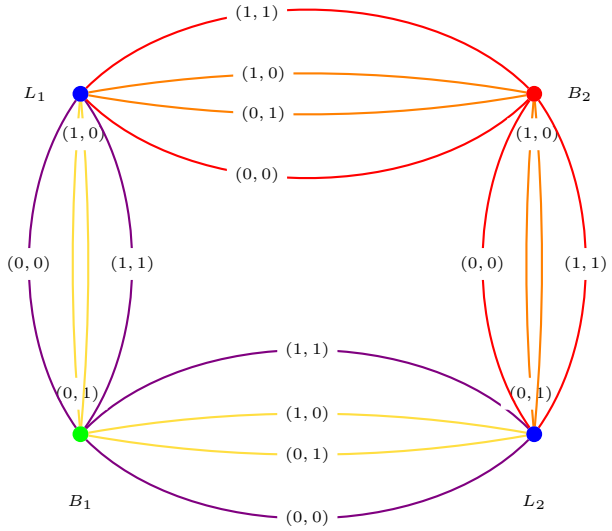


Figure 4.20: RGLL model with generating class $\{\{L_1, B_2\}, \{B_2, L_2\}, \{L_2, B_1\}, \{B_1, L_1\}\}$, *where all variables are binary, with vertices* $V = (V_1, V_2, V_3)$, $V_1 = \{L_1, L_2\}$, $V_2 = \{B_1\}$, *and* $V_3 = \{B_2\}$ *and first-order interactions set* $E = (E_1, E_2, E_3, E_4)$, $E_1 = \{u_{L_1B_2}(01), u_{L_1B_2}(10), u_{L_2B_2}(01), u_{L_2B_2}(10)\}$, $E_2 = \{u_{L_1B_2}(00), u_{L_1B_2}(11), u_{L_2B_2}(00), u_{L_2B_2}(11)\}$, $E_3 = \{u_{L_1B_1}(01), u_{L_1B_1}(10), u_{L_2B_1}(01), u_{L_2B_1}(10)\}$, $E_4 = \{u_{L_1B_1}(00), u_{L_1B_1}(11), u_{L_2B_1}(00), u_{L_2B_1}(11)\}$.

Using REGRAPH, we get a deviance of 13.54 with 10 degrees of freedom, and a p-value of 0.195. The value of X^2 is 10.81 with a p-value of 0.37. This means that this model fits the data well, even better than the graphical log-linear model with generating class $\{\{L_1, B_2\}, \{B_2, L_2\}, \{L_2, B_1\}, \{L_1, B_1\}\}$. Additionally, the model is preserved when we interchange the measures corresponding to the length of head for both brothers. This implies not only that the expected frequencies for the contingency table are preserved after permuting both variables, but also that there is scale invariance, so that which category is called zero and which category is called one is irrelevant.

ii) A similar label invariant model, interchanging B_1 with B_2 instead of L_1 with L_2 was also fitted; *i.e.*, a label invariant model $M(G, \Gamma)$ with $\Gamma = \{(B_1B_2)\}$. Its associated RGLL model has the graph given in figure 4.21. The vertex colour classes associated to this RGLL model are

$$V_1 = \{B_1, B_2\},\ V_2 = \{L_1\},\ V_3 = \{L_2\};$$

and the edge classes are

$$E_1 = \{u_{B_2L_2}(01), u_{B_2L_2}(10), u_{B_1L_2}(01), u_{B_1L_2}(10)\},$$
$$E_2 = \{u_{B_2L_2}(00), u_{B_2L_2}(11), u_{B_1L_2}(00), u_{B_1L_2}(11)\},$$

$$E_3 = \{u_{B_2L_1}(01), u_{B_2L_1}(10), u_{B_1L_1}(01), u_{B_1L_1}(10)\},$$
$$E_4 = \{u_{B_2L_1}(00), u_{B_2L_1}(11), u_{B_1L_1}(00), u_{B_1L_1}(11)\}.$$

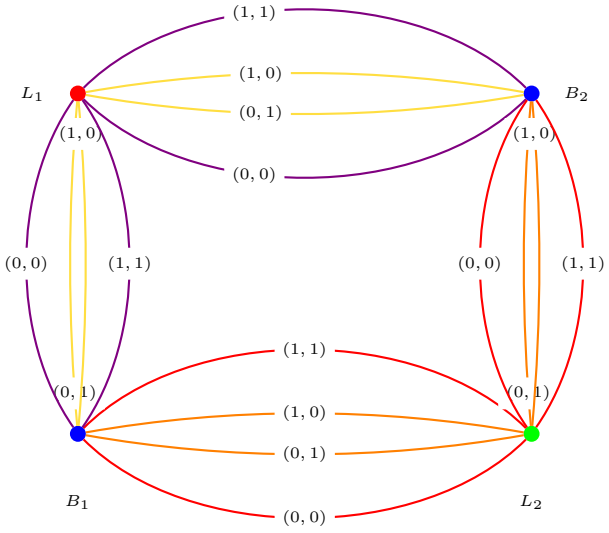


Figure 4.21: RGLL model with generating class $\{\{L_1, B_2\}, \{B_2, L_2\}, \{L_2, B_1\}, \{B_1, L_1\}\}$, *where all variables are binary, with vertices* $V = (V_1, V_2, V_3)$, $V_1 = \{B_1, B_2\}$, $V_2 = \{L_1\}$, *and* $V_3 = \{L_2\}$ *and first-order interactions set* $E = (E_1, E_2, E_3, E_4)$, $E_1 = \{u_{B_2L_2}(01), u_{B_2L_2}(10), u_{B_1L_2}(01), u_{B_1L_2}(10)\}$, $E_2 = \{u_{B_2L_2}(00), u_{B_2L_2}(11), u_{B_1L_2}(00), u_{B_1L_2}(11)\}$, $E_3 = \{u_{B_2L_1}(01), u_{B_2L_1}(10), u_{B_1L_1}(01), u_{B_1L_1}(10)\}$, $E_4 = \{u_{B_2L_1}(00), u_{B_2L_1}(11), u_{B_1L_1}(00), u_{B_1L_1}(11)\}$.

This model has an associated deviance of 15.73 with 10 degrees of freedom, a p-value of 0.107, and the value of X^2 is 14.00 with a p-value of 0.17. So that the model fits well, but not as well as the previous model nor even as well as the original graphical log-linear model with generating class $\{\{L_1, B_2\}, \{B_2, L_2\}, \{L_2, B_1\}, \{L_1, B_1\}\}$.

iii) Another label invariant model is one in which we interchange neighbour vertices for different brothers. That is, we permute L_1 with B_2 and B_1 with L_2, getting the label invariant model $M(G, \Gamma)$, where $\Gamma = \{(L_1B_2)(L_2B_1)\}$. We obtain a RGLL model associated to the label invariant model whose graph is the one given in figure 4.22. This RGLL model has vertex classes

$$V_1 = \{L_1, B_2\},\ V_2 = \{B_1, L_2\};$$

and edge classes

$$E_1 = \{u_{L_1B_1}(01), u_{L_1B_1}(10), u_{B_2L_2}(01), u_{B_2L_2}(10)\},$$

$$E_2 = \{u_{L_1B_1}(00), u_{L_1B_1}(11), u_{B_2L_2}(00), u_{B_2L_2}(11)\},$$
$$E_3 = \{u_{L_1B_2}(11), u_{L_1B_2}(00)\},$$
$$E_4 = \{u_{L_1B_2}(01), u_{L_1B_2}(10)\},$$
$$E_5 = \{u_{L_2B_1}(11), u_{L_2B_1}(00)\},$$
$$E_6 = \{u_{L_2B_1}(01), u_{L_2B_1}(10)\}.$$

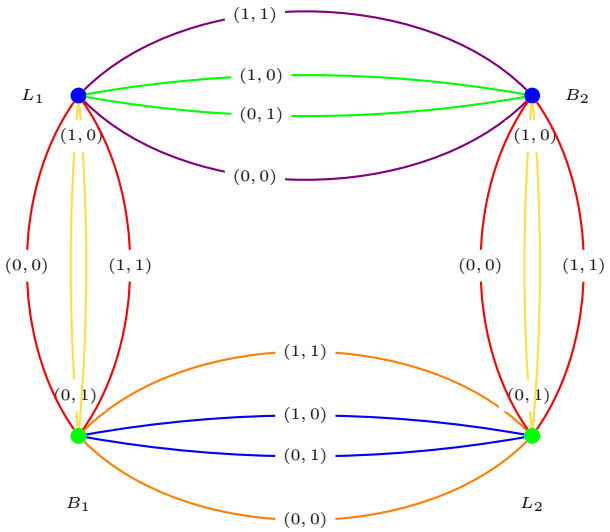


Figure 4.22: RGLL model with generating class $\{\{L_1, B_2\}, \{B_2, L_2\}, \{L_2, B_1\}, \{B_1, L_1\}\}$, *where all variables are binary, with vertices* $V = (V_1, V_2)$, $V_1 = \{L_1, B_2\}$, $V_2 = \{B_1, L_2\}$, *and first-order interactions set* $E = (E_1, E_2, E_3, E_4, E_5, E_6)$, $E_1 = \{u_{L_1B_1}(01), u_{L_1B_1}(10), u_{B_2L_2}(01), u_{B_2L_2}(10)\}$, $E_2 = \{u_{L_1B_1}(00), u_{L_1B_1}(11), u_{B_2L_2}(00), u_{B_2L_2}(11)\}$, $E_3 = \{u_{L_1B_2}(11), u_{L_1B_2}(00)\}$, $E_4 = \{u_{L_1B_2}(01), u_{L_1B_2}(10)\}$, $E_5 = \{u_{L_2B_1}(11), u_{L_2B_1}(00)\}$, $E_6 = \{u_{L_2B_1}(01), u_{L_2B_1}(10)\}$.

The deviance associated to this model is 13.61 with 10 degrees of freedom, a p-value of 0.191, and the value of X^2 is 14.70 with a p-value of 0.14, so that the model fits the data well. The meaning of this model is that we can interchange the measures that are directly associated for the two different brothers and still get the same model and that there is scale invariance.

iv) The last label invariant model we present is $M(G, \Gamma)$ with $\Gamma = \{(L_1L_2)\ (B_1B_2)\}$, which is a label invariant model in which we permute the measures corresponding to both brothers. In this sense, even though the graph is not the same as in the continuous case, this model has a similar meaning than the model presented by Højsgaard and

Lauritzen (2008) for the Gaussian case.

The resulting associated graph is given in figure 4.13, but with vertices L_1, B_2, L_2, B_1 instead of A, B, C, and D, respectively. Figure 4.23 shows the corresponding graph with its new labeling. The corresponding RGLL model has vertex classes

$$V_1 = \{L_1, L_2\},\ V_2 = \{B_1, B_2\};$$

and edge classes

$$\begin{aligned}
E_1 &= \{u_{L_1B_2}(01), u_{L_1B_2}(10), u_{L_2B_1}(01), u_{L_2B_1}(10)\},\\
E_2 &= \{u_{L_1B_2}(00), u_{L_1B_2}(11), u_{L_2B_1}(00), u_{L_2B_1}(11)\},\\
E_3 &= \{u_{B_2L_2}(01), u_{B_2L_2}(10), u_{B_1L_1}(01), u_{B_1L_1}(10)\},\\
E_4 &= \{u_{B_2L_2}(00), u_{B_2L_2}(11), u_{B_1L_1}(00), u_{B_1L_1}(11)\}.
\end{aligned}$$

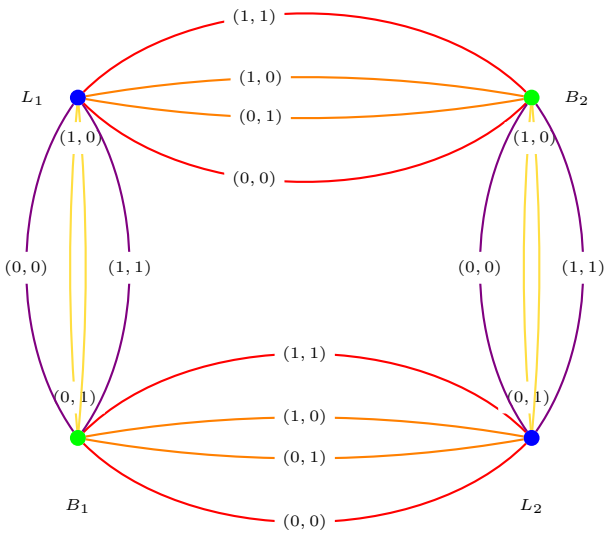


Figure 4.23: RGLL model with generating class $\{\{L_1, B_2\}, \{B_2, L_2\}, \{L_2, B_1\}, \{B_1, L_1\}\}$, *where all variables are binary, with vertices* $V = (V_1, V_2)$, $V_1 = \{L_1, L_2\}$ *and* $V_2 = \{B_1, B_2\}$, *and first-order interactions set* $E = (E_1, E_2, E_3, E_4)$, $E_1 = \{u_{L_1B_2}(01), u_{L_1B_2}(10), u_{L_2B_1}(01), u_{L_2B_1}(10)\}$, $E_2 = \{u_{L_1B_2}(00), u_{L_1B_2}(11), u_{L_2B_1}(00), u_{L_2B_1}(11)\}$, $E_3 = \{u_{B_2L_2}(01), u_{B_2L_2}(10), u_{B_1L_1}(01), u_{B_1L_1}(10)\}$, $E_4 = \{u_{B_2L_2}(00), u_{B_2L_2}(11), u_{B_1L_1}(00), u_{B_1L_1}(11)\}$.

We get a deviance of 14.40 with 11 degrees of freedom, and a p-value of 0.21, which means that this model fits well according to the deviance, even better than any of the other models presented. Its meaning is that there is symmetry in both measures between the two sons, so that the model is preserved after the change, implying that

the expected frequencies for the contingency table are preserved after permuting the variables, and that there is scale invariance, so that the name or order given to the levels is irrelevant. The value of the statistic X^2 is 11.66 with a p-value of 0.39.

The fitted values are presented in table 4.5. Observe that the fitted values obtained are such that if we interchanged L_1 with L_2 and B_1 with B_2, we would get exactly the same table, in which instead of having L_1, we would have L_2, and instead of B_1 we would have B_2. In fact, in this case, the model is such that the fitted values for the cells under the diagonal are equal to the values for the cells above the diagonal. We observe then that the concept of label invariant models implies symmetry, not only in terms of the graph, but also in terms of the contingency table, so that we are generalizing in some way the concept of symmetry.

L_1	B_1	L_2			
		0		1	
		B_2			
		0	1	0	1
0	0	6.09[1] 7[2]	0.38 0	1.42 1	1.11 1
	1	0.38 0	0.05 0	0.61 2	0.47 0
1	0	1.42 0	0.61 0	0.54 1	1.42 3
	1	1.11 2	0.47 1	1.42 0	7.50 7

Table 4.5: [1]Fitted expected frequencies for the discretized Frets' heads data under the label invariant model $M(G,\Gamma)$, with permutation set $\Gamma = \{(L_1L_2)\ (B_1B_2)\}$ and graph G with generating class $A = \{\{L_1,\ B_2\},\ \{B_2,\ L_2\},\ \{L_2,\ B_1\},\ \{L_1,\ B_1\}$, compared with the [2]observed counts.

v) A final model we present is the level invariant model known as cyclic permutation defined in example 4.7. This is a level invariant model $M'(G,\Gamma')$ with $\Gamma' = \{(L_1B_1L_2B_2),\ (L_1L_2)(B_1B_2),\ (L_1B_2L_2B_1)\}$. For the corresponding RGLL model whose graph is shown in figure 4.24, we get only one vertex class

$$V_1 = V = \{L_1, B_1, L_2, B_2\};$$

and four edge classes given by

$$E_1 = \{u_{L_1B_1}(00), u_{B_1L_2}(00), u_{L_2B_2}(00), u_{B_2L_1}(00)\},$$
$$E_2 = \{u_{L_1B_1}(01), u_{B_1L_2}(01), u_{L_2B_2}(01), u_{B_2L_1}(01)\},$$

$$E_3 = \{u_{L_1B_1}(10), u_{B_1L_2}(10), u_{L_2B_2}(10), u_{B_2L_1}(10)\},$$
$$E_4 = \{u_{L_1B_1}(11), u_{B_1L_2}(11), u_{L_2B_2}(11), u_{B_2L_1}(11)\}.$$

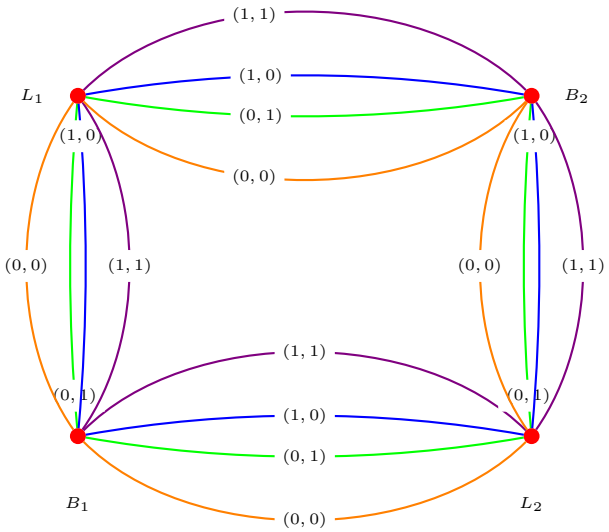


Figure 4.24: RGLL model with generating class $\{\{L_1, B_2\}, \{B_2, L_2\}, \{L_2, B_1\}, \{B_1, L_1\}\}$, *where all variables are binary, with vertices* $V = \{L_1, B_1, L_2, B_2\}$ *and first-order interactions set* $E = (E_1, E_2, E_3, E_4)$, $E_1 = \{u_{L_1B_1}(00), u_{B_1L_2}(00), u_{L_2B_2}(00), u_{B_2L_1}(00)\}$, $E_2 = \{u_{L_1B_1}(01), u_{B_1L_2}(01), u_{L_2B_2}(01), u_{B_2L_1}(01)\}$, $E_3 = \{u_{L_1B_1}(10), u_{B_1L_2}(10), u_{L_2B_2}(10), u_{B_2L_1}(10)\}$, $E_4 = \{u_{L_1B_1}(11), u_{B_1L_2}(11), u_{L_2B_2}(11), u_{B_2L_1}(11)\}$.

This model does not fit the data well. The deviance is 37.27 with 12 degrees of freedom and a p-value of 0.0001. The value of X^2 is 135.78 with a p-value less than 0.0001. If this model had fitted the data, it would have meant that the levels of the variables could have been changed in a cyclic way and the expected frequencies would have been the same.

In table 4.6 we present the values of the deviance and X^2 statistics and the number of degrees of freedom for the five RGLL models discussed here. The X^2 statistic is shown in this case because the number of observations is small, $|n| = 25$. Using any of the two statistics the same conclusions are obtained for all models, in the sense that when a model is not rejected using a statistic is also not rejected using the other, and for the cyclic permutation model we observe that with both statistics the model is rejected, these conclusions can be obtained by comparing p-values for both statistics. Observe also that using either the deviance or X^2 the model with greater p-value corresponds to the label invariant model $M(G, \Gamma)$ with $\Gamma = \{(L_1L_2)\ (B_1B_2)\}$. The estimated expected frequencies under each model are shown in table 4.7.

Model	**deviance**	**p-value**	X^2	**p-value**	**d.f.**
Unrestricted graphical model	11.17	0.13	9.61	0.21	7
i)$M(G,\Gamma)$, Γ= $\{(L_1L_2)\}$	13.54	0.20	10.81	0.37	10
ii)$M(G,\Gamma)$, Γ= $\{(B_1B_2)\}$	15.73	0.11	14.00	0.17	10
iii)$M(G,\Gamma)$, Γ= $\{(L_1B_2)(L_2B_1)\}$	13.61	0.19	14.70	0.14	10
iv)$M(G,\Gamma)$, Γ= $\{(L_1L_2)(B_1B_2)\}$	14.40	0.21	11.66	0.39	11
v) Cyclic permutation	37.27	>0.01	137.78	>0.01	12

Table 4.6: Values for the deviance and X^2 statistics, p-values, and degrees of freedom associated to all RGLL models for the discretized Frets' heads data analyzed in example 4.12.

L_1	B_1	L_2	B_2	$n(i)$	Model i)	Model ii)	Model iii)	Model iv)	Model v)
0	0	0	0	7	6.26	5.87	5.86	6.09	6.06
0	0	0	1	0	0.17	0.36	0.30	0.38	0.47
0	0	1	0	1	1.12	2.59	1.67	1.42	0.07
0	0	1	1	1	0.47	0.67	1.17	1.11	2.91
0	1	0	0	0	1.33	0.36	1.67	0.38	0.07
0	1	0	1	0	0.04	0.05	0.12	0.05	0.02
0	1	1	0	2	0.80	0.67	0.80	0.61	0.46
0	1	1	1	0	0.34	0.42	0.41	0.47	0.90
1	0	0	0	0	1.12	1.11	0.30	1.42	0.70
1	0	0	1	0	0.47	0.66	0.54	0.61	0.46
1	0	1	0	1	0.46	0.49	0.12	0.54	0.02
1	0	1	1	3	2.96	1.24	1.54	1.42	0.14
1	1	0	0	2	0.80	0.66	1.17	1.11	4.27
1	1	0	1	1	0.34	0.94	1.54	0.47	0.14
1	1	1	0	0	1.12	1.24	0.41	1.42	1.32
1	1	1	1	7	7.23	7.67	7.38	7.50	6.98

Table 4.7: Estimated expected frequencies under the five label and level invariant models fitted for the discretized Frets' heads data analyzed in example 4.12: i) $M(G,\Gamma)$, Γ= $\{(L_1L_2)\}$, ii) $M(G,\Gamma)$, Γ= $\{(B_1B_2)\}$, iii) $M(G,\Gamma)$, Γ= $\{(L_1B_2)(L_2B_1)\}$, iv) $M(G,\Gamma)$, Γ= $\{(L_1L_2)(B_1B_2)\}$, and v) cyclic permutation.

Example 4.13. As a second example, we fit a label invariant model for the data corresponding to alcohol dependence and depression in female monozygotic twins presented in table 4.1 at the beginning of this chapter. This table is similar to the one presented by Drton and Richardson (2008) with an ordering in the variables such that the variables corresponding to each twin are in the same side of the table.

We used the selection methods available in *MIM* to find a generating class corresponding to a graphical log-linear model that fits the data and that simultaneously has an associated triangle-free graph. We chose the model obtained using the unrestricted selection method available in *MIM* starting with the saturated model because this method provided a model that fitted better, according to the deviance, than any other graphical log-linear model whose associated graph was triangle-free selected using another selection method and the model even fitted better or equally as good as other selected models whose associated graphs were not triangle-free. We obtain a model that fits the data, satisfying at the same time the requirements needed for label invariant models. This model has generating class $\{\{A_1,\, A_2\},\, \{A_1,\, D_1\},\, \{A_2,\, D_2\},\, \{D_1,\, D_2\}\}$. Its corresponding deviance is 4.75 with 7 degrees of freedom and a p-value of 0.69. This means that, in general, for example for a significance level of 0.05, we do not reject the null hypothesis that the model fits the data.

The graph associated to this model, G, corresponds to a four-cycle as the one presented in figure 4.2, with A_1, A_2, D_2, and D_1 instead of A, B, C, and D, respectively, and it is also shown in figure 4.25(a).

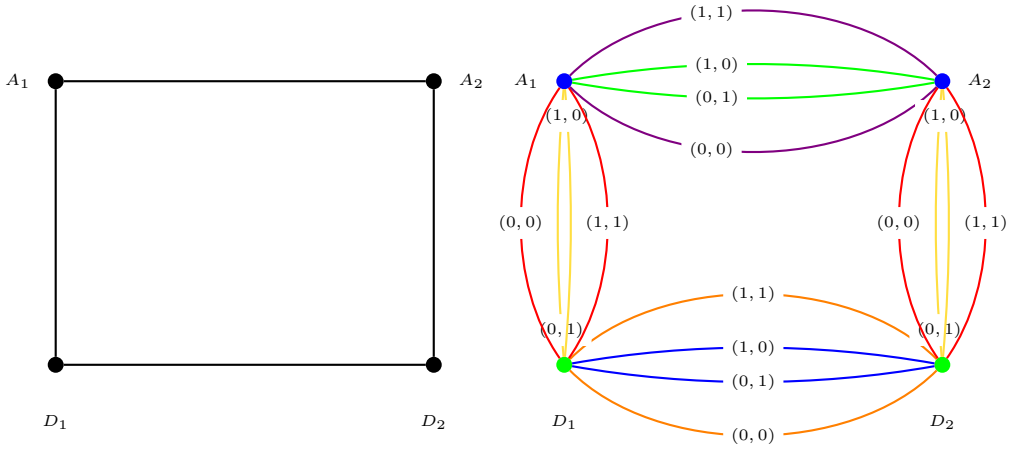
An interesting label invariant model that according to the data may fit well and that is also meaningful is one in which we interchange A_1 with A_2 and D_1 with D_2 at the same time. This is the label invariant model $M(G, \Gamma)$, where $\Gamma = \{(D_1D_2)\ (A_1A_2)\}$ and G is the graph associated to the model generated by $A = \{\{A_1,\, A_2\},\, \{A_1,\, D_1\},\, \{A_2,\, D_2\},\, \{D_1,\, D_2\}\}$, with $A = C(G)$, the set of cliques of G. The graph of the associated RGLL model is the same given in figure 4.10, but with A_1, A_2, D_2, and D_1 instead of A, B, C, and D, respectively. The corresponding graph is also shown in figure 4.25(b). This is a RGLL model, with vertex classes

$$V_1 = \{A_1, A_2\},\ V_2 = \{D_1, D_2\};$$

and six edge classes given by

$$\begin{aligned} E_1 &= \{u_{A_1D_1}(00), u_{A_1D_1}(11), u_{A_2D_2}(00), u_{A_2D_2}(11)\},\\ E_2 &= \{u_{A_1D_1}(01), u_{A_1D_1}(10), u_{A_2D_2}(01), u_{A_2D_2}(10)\},\\ E_3 &= \{u_{A_1A_2}(00), u_{A_1A_2}(11)\},\\ E_4 &= \{u_{A_1A_2}(01), u_{A_1A_2}(10)\}, \end{aligned}$$

$$E_5 = \{u_{D_1D_2}(00), u_{D_1D_2}(11)\},$$
$$E_6 = \{u_{D_1D_2}(01), u_{D_1D_2}(10)\}.$$



(a) Graphical model (b) RGLL model

Figure 4.25: Log-linear models with generating class $\{\{A_1, A_2\}, \{A_1, D_1\}, \{A_2, D_2\}, \{D_1, D_2\}\}$ *where all variables are binary: (a) graphical model; (b) RGLL model with vertices* $V = (V_1, V_2)$, $V_1 = \{A_1, A_2\}$ *and* $V_2 = \{D_1, D_2\}$, *and first-order interactions set* $E = (E_1, E_2, E_3, E_4, E_5, E_6)$, $E_1 = \{u_{A_1D_1}(00), u_{A_1D_1}(11), u_{A_2D_2}(00), u_{A_2D_2}(11)\}$, $E_2 = \{u_{A_1D_1}(01), u_{A_1D_1}(10), u_{A_2D_2}(01), u_{A_2D_2}(10)\}$, $E_3 = \{u_{A_1A_2}(00), u_{A_1A_2}(11)\}$, $E_4 = \{u_{A_1A_2}(01), u_{A_1A_2}(10)\}$, $E_5 = \{u_{D_1D_2}(00), u_{D_1D_2}(11)\}$, $E_6 = \{u_{D_1D_2}(01), u_{D_1D_2}(10)\}$.

Using REGRAPH, we get a deviance of 13.06 with 10 degrees of freedom and a p-value of 0.22. These values indicate that this model fits the data well according to the deviance because, in general, we do not reject the corresponding null hypothesis. The value of the Pearson X^2 statistic is 13.27 with a p-value of 0.21.

Additionally, the model has various interpretations. It corresponds to irrelevance of the labels given to the two twins and of the level names, so that we can interchange the measures corresponding to alcohol dependency and depression for both twins and preserve the model and besides, which category is coded zero and which category is coded one is irrelevant. In other words, the expected frequencies for the contingency table are preserved after permuting both variables and there is scale invariance.

The fitted values for this model are presented in table 4.8. Observe that if we interchange the variables corresponding to the first twin, A_1 and D_1, with the ones of the second twin, A_2 and D_2, we get a new table which is exactly the same table we had before. In this case, this symmetry is also reflected in the symmetry of the table where

the values above the diagonal are equal to the ones below the diagonal.

A_1	D_1	A_2							
		0				1			
		D_2							
		0		1		0		1	
0	0	287.60[1]	(0.80)	82.35	(-4.63)	11.22	(8.71)	12.83	(-6.38)
		288[2]		80		15		9	
		275.34[3]	(25.90)	88.71	(-16.54)	14.87	(0.27)	12.18	(-5.44)
	1	82.35	(20.39)	58.71	(-14.63)	4.75	(5.43)	6.19	(9.59)
		92		51		7		10	
		88.71	(6.69)	57.13	(-11.58)	3.40	(10.10)	6.21	(9.53)
1	0	11.22	(-5.41)	4.75	(-1.37)	3.97	(-1.68)	3.07	(-1.71)
		8		4		3		2	
		14.87	(-9.91)	3.40	(1.29)	5.97	(-4.13)	2.27	(-0.50)
	1	12.83	(-7.56)	6.19	(6.74)	3.07	(2.12)	5.90	(2.39)
		8		9		4		7	
		12.18	(-6.72)	6.21	(6.68)	2.27	(4.54)	3.22	(10.85)

Table 4.8: [1]*Fitted expected frequencies for the female twins data under the label invariant model* $M(G,\Gamma)$*, with permutation set* $\Gamma = \{(A_1A_2)\ (D_1D_2)\}$ *and graph* G *with generating class* $A = \{\{A_1,\ A_2\},\ \{A_1,\ D_1\},\ \{A_2,\ D_2\},\ \{D_1,\ D_2\}\}$ *representing* $A_1 \perp D_2 | D_1, A_2$ *and* $A_2 \perp D_1 | A_1, D_2$*, compared with the* [2]*observed values and the* [3]*expected frequencies under a model proposed by Drton and Richardson that assumes that* $A_1 \perp D_2$ *and* $A_2 \perp D_1$ *with the same symmetry given by* Γ*. In parenthesis, the deviance contribution of each cell.*

Using the separator sets associated to the graph and the global Markov property (section **2.10**), we have that $A_1 \perp D_2 | D_1, A_2$ and $A_2 \perp D_1 | A_1, D_2$. So that, we have a combined model that consists of these conditional independences together with the symmetry and scale invariance.

A similar model for the same data, but considering marginal instead of conditional independences, was proposed and fitted by Drton and Richardson (2008). They define *binary models for marginal independence* as an alternative to graphical models, these models consist on probability distributions for binary random vectors $X = (X_v)_{v \in V}$ associated to bi-directed graphs, graphs with arrows in both ends of an edge, in which a Markov property, analogous to the global Markov property used in graphical models, called the *connected set Markov* property should be satisfied. A random vector $X = (X_v)_{v \in V}$ satisfies the connected set Markov property if C is independent to $V \setminus Sp(C)$, where $Sp(C)$ is the set of all vertices adjacent to C, whenever $\emptyset \neq C \subseteq V$ is a connected set. This property implies that the lack of an edge between two vari-

ables indicates marginal independence between them, which is analogous to the pairwise Markov property for graphical models. For example, the marginal independence model indicating $A_1 \perp D_2$ and $A_2 \perp D_1$ is presented in figure 4.26.

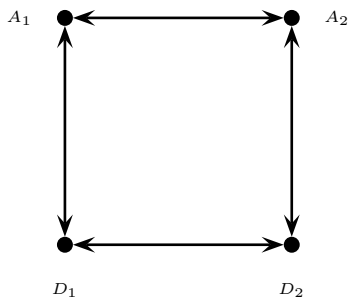


Figure 4.26: Marginal independence model indicating $A_1 \perp D_2$ and $A_2 \perp D_1$ whose graph corresponds to a bi-directed four-cycle.

To illustrate the use of marginal independence models together with symmetry, Drton and Richardson (2008) use the same symmetry given by the permutation set $\Gamma = \{(D_1 D_2)\ (A_1 A_2)\}$ for the twins data. A minor observation is that they included Id in Γ in their model to form a group; however, its inclusion in label and level invariant models is optional because it is a transformation that does not modify the models. Applying only this symmetry group without considering additional conditions for the distribution, *i.e.* considering only irrelevance of the labels given to the twins, they got a model with a deviance of 4.62 with 6 degrees of freedom and a p-value of 0.59. When they also considered the marginal independence $A_1 \perp D_2$ and $A_2 \perp D_1$, *i.e.* the marginal independence model together with the symmetry group, they got a deviance of 16.15 with 2 degrees of freedom, taking the model with only the symmetry group as the alternative, with a p-value of 0.0003, indicating a poor fit. On the other hand, taking the saturated model as alternative and using the probabilities provided by the authors, we got a residual deviance of 21.02 with 8 degrees of freedom and a p-value of 0.0071, which still indicates a poor fit.

The expected frequencies under the model combining marginal independence and symmetry are also presented in table 4.8. We also present there for both models the deviance contribution of each cell. When these values are small, we have a better fit for that cell. There are seven cells in which the label invariant model proposed here improves over the other model, six in which it is the opposite, and three in which both models fit in a similar way.

All this means that although a marginal independence model together with the corresponding symmetry group does not have a good fit according to the deviance and the tests obtained from it, there is a similar model, the one proposed here considering

conditional independence for the same variables instead of marginal independence, that fits the data well.

4.3 General label and level invariant models

In this section we generalize the concept of label and level invariant models, so that it includes any graphical log-linear model even those whose graph is not necessarily triangle-free. In this way, we define models that preserve the conditional independences derived from any graphical model together with symmetry, understanding this concept for the label invariant case as scale invariant models that preserve both the graph and model, and consequently the distribution, and for level invariant models as models that preserve certain equalities between expected frequencies. Both models are defined exactly in the same way as the non-general models, the only difference is that the associated graph G could be any graph. When G is not triangle-free, the resulting model is not a RGLL model because we are restricting interactions of order higher than one, in this case there is not a software implementation that fits such models. We consider also that all variables have the same categories, *i.e.* I_δ is the same for all $\delta \in V$, so that we have J levels for each variable.

Consider a graphical log-linear model with generating class $C(G)$, the set of cliques of G, and associated graph $G = (V, E)$. Consider also $Aut(G)$ a subset of the symmetric group formed by all permutations of V, $S(V)$, consisting on permutations that preserve the graph and $\mathbb{M}$ the expected frequency vector $(m(i))_{i\in \mathrm{I}}$, $I = (I_\delta)_{\delta\in V}$, including only the constraints corresponding to the sampling scheme: Poisson, multinomial, or restricted multinomial.

Definition 4.5. A general label invariant model $\boldsymbol{M}(G, \Gamma)$ determined by a graph G and by a permutation set $\Gamma \subseteq Aut(G)$ is given by assuming that: i) we have a graphical log-linear model $\boldsymbol{M}$ with generating class $C(G)$ and $\boldsymbol{M} \subseteq \mathbb{M}$, and ii) if $\mathrm{M} \in \boldsymbol{M}$ then $\sigma(\sigma^*(\mathrm{M})) \in \boldsymbol{M}$, for all $\sigma^* \in S(\mathrm{I}_\delta)$ and for all $\sigma \in \Gamma$.

Consider a graphical log-linear model with generating class $C(G)$, the set of cliques of G, $S(V)$ the symmetric group formed by all permutations of V, and $\mathbb{M}$ the expected frequency vector $(m(i))_{i\in \mathrm{I}}$, $I = (I_\delta)_{\delta\in V}$, including only the constraints corresponding to the sampling scheme: Poisson, multinomial, or restricted multinomial.

Definition 4.6. A general level invariant model $\boldsymbol{M'}(G, \Gamma')$ determined by a graph G and by a permutation set $\Gamma' \subseteq S(V)$ is given by assuming that: i) we have a graphical log-linear model $\boldsymbol{M'}$ with generating class $C(G)$ and $\boldsymbol{M'} \subseteq \mathbb{M}$, and ii) if $M \in \boldsymbol{M'}$ then $M(i) = M(\sigma' i)$, for all $\sigma' \in \Gamma'$ and $i \in \mathrm{I}$.

The concept of scale invariance satisfied by a general label invariant model is similar to the one corresponding to label invariant models for triangle-free graphs studied in

the previous sections. Namely, models that are preserved after permuting the level set. As a consequence we have that under scale invariance, after permuting the level set, the parameters should belong to the same group of parameters. Then, the association between variables reflected in the interactions of any order is going to be the same for any level permutation. However, this property does not imply that after permuting the levels the expected frequencies are necessarily the same.

As in label invariant models for triangle-free graphs, we observe that in general label invariant models once permuting labels the expected frequencies are the same, which means that the distribution is preserved.

In the Gaussian case there are no similar general models, because in that case the only existing concern is the association between two variables given by the concentration matrix or partial correlation matrix whereas in graphical log-linear models associations between three or more variables are represented by interactions of two or higher order.

To obtain a graphical log-linear model restricted in some of its parameters associated to these general models, we need to generalize some of the concepts used before for triangle-free label and level invariant models.

The n-term orbits of Γ, $n = 1, 2, ...$, are the classes of the relation

$$\{\alpha_1, ..., \alpha_n\} \equiv_\Gamma \{\beta_1, ..., \beta_n\} \Leftrightarrow \{\beta_1, ..., \beta_n\} = \{\sigma\alpha_1, ..., \sigma\alpha_n\}, \text{ for some } \sigma \in \Gamma.$$

If $n = 1$, we have vertex orbits; if $n = 2$, we have edge orbits as for triangle-free graphs.

To identify the restrictions on the parameters that allow us to permute the labels preserving the same model, we do something similar to what we did in the non-general case where we had associated triangle-free graphs. First, we get the n-term orbits associated to the model. When $n = 1$, the vertex orbits determine the restrictions on the main effects, which are that all main effects for the vertices in the same orbit are equal for all their levels, so that the vertex orbits are vertex classes as in RGLL models. When $n = 2$, the edge orbits determine the set of first-order interactions that need to be equal; however, as in the non-general case, we need to add scale invariance conditions to obtain the final edge classes used on the restricted model. Then, the 3-term orbits provide the partitions of the two-order interactions whose elements should be equal to be able to permute labels. To obtain the 3-term classes, we have to add restrictions to satisfy scale invariance. The process is the same for all n-terms orbits, with n taking values from 1 to the number of vertices of the element in the generating class with the largest cardinality, *i.e.* the number of elements in the largest clique on the graph.

Setting restrictions on the parameters that represent scale invariance for models with non triangular-free graphs is what differs the most with respect to the non-general case with triangle-free graphs. This is because in the general case we may have several restrictions for the interactions of two or higher order that allow having this property, so that there could be many different ways of restricting the parameters in the model and still getting scale invariance, consequently there is no uniqueness in the set of restrictions on the parameters. This is not a theoretical problem but a practical one, because we could fit different restricted models to some data and maybe we could choose one that does not fit as well as others making us think that the label invariant model does not fit, when it does for other ways of representing scale invariance. We did not have this problem in triangle-free graphs because to obtain scale invariance in those models we only had two type of restrictions for the one type of available interactions, first-order interactions, either the levels were the same for both variables and the corresponding interactions were equal or the levels were different and the corresponding interactions were equal.

For any possible way of representing scale invariance, we have that after permuting the level set the parameter is kept in the same group of parameters, preserving the model and implying that the association is the same in spite of the scale used.

For example, suppose that we have a model containing second-order interactions formed by the vertices X, Y, and Z. Three ways to get scale invariance are:

a) Using the classes $\{u_{XYZ}(iii),\ i = 1, ..., J\}$, $\{u_{XYZ}(ijk),\ i = j,\ j \neq k;\ i, j, k = 1, ..., J\}$, $\{u_{XYZ}(ijk),\ i \neq j,\ j = k;\ i, j, k = 1, ..., J\}$, $\{u_{XYZ}(ijk),\ j \neq i,\ i = k;\ i, j, k = 1, ..., J\}$, and $\{u_{XYZ}(ijk),\ i \neq j,\ i \neq k,\ j \neq k;\ i, j, k = 1, ..., J\}$.

b) Using the classes $\{u_{XYZ}(iii),\ i = 1, ..., J\}$, $\{u_{XYZ}(ijk),\ i = j,\ j \neq k;\ i \neq j,\ j = k;$ or $j \neq i,\ i = k;\ i, j, k = 1, ..., J\}$, and $\{u_{XYZ}(ijk),\ i \neq j,\ i \neq k,\ j \neq k;\ i, j, k = 1, ..., J\}$.

c) Using the classes $\{u_{XYZ}(iii),\ i = 1, ..., J\}$, $\{u_{XYZ}(ijk)$, with not all levels equal; $i, j, k = 1, ..., J\}$.

For instance, if $J = 2$ the classes given in a) are $\{u_{XYZ}(111),\ u_{XYZ}(222)\}$, $\{u_{XYZ}(112),\ u_{XYZ}(221)\}$, $\{u_{XYZ}(122),\ u_{XYZ}(211)\}$, and $\{u_{XYZ}(121),\ u_{XYZ}(212)\}$; and the classes given in b) and c) are $\{u_{XYZ}(111),\ u_{XYZ}(222)\}$ and $\{u_{XYZ}(112),\ u_{XYZ}(221),\ u_{XYZ}(122),\ u_{XYZ}(211),\ u_{XYZ}(121),\ u_{XYZ}(212)\}$. Observe that in this case b) and c) are formed by the same sets because $\{u_{XYZ}(ijk),\ i \neq j,\ i \neq k,\ j \neq k;\ i, j, k = 1, ..., J\}$ is an empty set.

As before, each class indicates that the terms in each set should be equal. The first case is the one that lets the parameters less restricted still conserving scale invariance,

as a consequence there are more classes. In the second case the parameters are more restricted. Finally, the last case is the most restrictive. In any case, a), b), or c), once we apply these restrictions and the restrictions for the first-order interactions that allow having scale invariance, we obtain the same model after permuting levels. It is the same model in the sense that each parameter after permuting levels belongs to the same group of parameters that the original parameter belonged to and that the association between variables is not depending on the scale, *i.e.* all terms related to interactions are the same without taking into account the scale.

As in the non-general case, the classes used to represent the the general label invariant model are the result of combining the classes that allow to permute labels preserving the model, *i.e.* the classes obtained from the orbits, and the ones that allow having scale invariance.

As the interaction order increases, getting all possible scale invariance conditions that work gets more and more complex. They could be any possible partition of the interactions that once equated make possible preserving the model. We could always use classes as the ones seen in c), this is a class formed by interactions whose levels are equal and another class formed by interactions in which not all of the levels are equal, but this could be a very restrictive way of expressing scale invariance. In general, for a $(r-1)$-order interaction, $r \geq 2$, *i.e.* an interaction formed by r elements, and supposing we have vertices A_1, A_2,..., A_r, a less restricted and more general way to express the classes or sets whose elements should be equal is given by

$$\{u_{A_1A_2...A_r}(i_1i_2...i_r),\ i_1 = i_2 = ... = i_r;\ i_1, ..., i_r = 1, ..., J\} \tag{4.9}$$

$$\{u_{A_1A_2...A_r}(i_1i_2...i_r),\ i_j \neq i_k,\ j \neq k;\ i_1, ..., i_r = 1, ..., J\} \tag{4.10}$$

For $r \geq 3$ and $q = 2, ..., r-1$

$$\begin{cases} \{u_{A_1A_2...A_r}(i_1i_2...i_r),\ \text{one } q-\text{group of levels are equal};\ i_1, ..., i_r = 1, ..., J\} \\ \{u_{A_1A_2...A_r}(i_1i_2...i_r),\ \text{two } q-\text{groups of levels are equal and different} \\ \text{between them};\ i_1, ..., i_r = 1, ..., J\} \\ \qquad\qquad \vdots \\ \{u_{A_1A_2...A_r}(i_1i_2...i_r),\ \left[\frac{r}{q}\right]\ \ q-\text{groups of levels are equal and different} \\ \text{between them};\ i_1, ..., i_r = 1, ..., J\},\ 1 \leq \left[\frac{r}{q}\right] \leq J \end{cases} \tag{4.11}$$

For example, for second-order interactions we obtain the classes seen in b). For first-order interaction we obtain the classes we already know, one for the case in which the levels are equal and another for the case in which the levels are not equal.

Then, for a graph G and a permutation set $\Gamma \subseteq Aut(G)$ we have a group of models, log-linear expansions in terms of parameters which some are restricted to be equal, that have the properties needed to represent a general label invariant model $\boldsymbol{M}(G, \Gamma)$. The process to obtain such restrictions that lead to that general label invariant model is the following.

1) Consider that the number of variables forming each possible parameter according to the generating class is n, with $1 \leq n \leq \Omega$, where Ω is the number of elements in the largest clique on the graph. Get all n-orbits, $1 \leq n \leq \Omega$. These orbits will generate sets of parameters that should be equated to preserve the model after permuting labels. For $n = 1$, we have vertex classes and the main effects for the vertices in the same class are equal in each level.

2) Use the sets defined above, equations (4.9), (4.10), and (4.11), for all interactions, $n \geq 2$, to obtain restrictions that allow having scale invariance. We could use any other sets indicating which parameters have to be equated as long as they generate scale invariant parameters.

3) Combine the sets got in step 1) and 2) for $n \geq 2$, *i.e.* join the elements $u_{A_1 A_2 \ldots A_n}(i_1 i_2 \ldots i_n)$, $n \geq 2$ in the same class for each set in the scale invariance condition sets according to the orbits.

After this process, we obtain a graphical log-linear model whose parameters are restricted to be equal according to a general label invariant model $\boldsymbol{M}(G, \Gamma)$.

If we have a general level invariant model $\boldsymbol{M'}(G, \Gamma')$ whose associated permutation set Γ' is formed by automorphisms of the graph G, then we can get the restrictions that make possible the expected frequencies to be equal after permuting levels according to Γ' using the n-terms orbits. The definition of vertex classes is similar to the one used in a label invariant model and the rest of n-terms orbits imply that the interactions in the same orbit are equal for all possible permutations of the levels that all vertices can take in each interaction. In fact, a general label invariant model could also be used to represent this model.

Observe that these general label and level invariant models are not particular cases of RGLL models. This is because even though in RGLL models we allow interactions of any order, we only restrict the main effects and first-order interactions; however, in these general invariant models we can restrict any parameter including interactions of any order.

Observe also that getting the graphical representation of these models as a colouring is not straightforward. This is because there is no way to colour the vertex or edges on

the graph indicating all restrictions, even in the graph with multiple edges, much less when we consider the levels taken by the variables forming the interactions.

One possibility could be using the concept of hypergraphs to obtain graphs representing the orbits or even some graphs associated to a model. An hypergraph $H = (V, E)$ is formed by a set of vertices, V, and a set of hyperedges, E, which are subsets of the vertex set, $E \subseteq P(V) \setminus \emptyset$, where $P(V)$ is the power set of V. In particular, when all hyperedges have the same cardinality k, we have a k-uniform hypergraph. For example, a 3-uniform hypergraph is formed by a collection of triples. We observe then that n-term orbits, $n \geq 2$ are sets of hyperedges in the associated n-uniform hypergraph. We could get a hypergraph, G^*, for each n, $n \geq 3$ and colour the hyperedges in the same orbit with the same colour. This means that we could use a graph to represent 1 and 2-term orbits, as in label and level invariant models, and for each of the remaining orbits we could use a coloured hypergraph indicating which parameters are in the same orbit. However; this seems unpractical, in the sense that the representation of the orbits is more complicated than explicitly giving the sets without graphs. In fact, this will get even more complex if we also want to represent scale invariance conditions.

Example 4.14. Suppose that we have five vertices, A, B, C, D, and E in a graphical log-linear model generated by $\{\{A, B, E\}, \{A, D, E\}, \{C, B, E\}, \{C, D, E\}\}$, figure 4.27. Suppose also that we have $\Gamma = \{(AC), (BD)\}$, where $\sigma_1 = (AC)$ and $\sigma_2 = (BD)$. This is a permutation set that permutes A with C on one side and B and D on the other. Both elements in Γ belong to the automorphism set $Aut(G)$ because they preserve the graph, G.

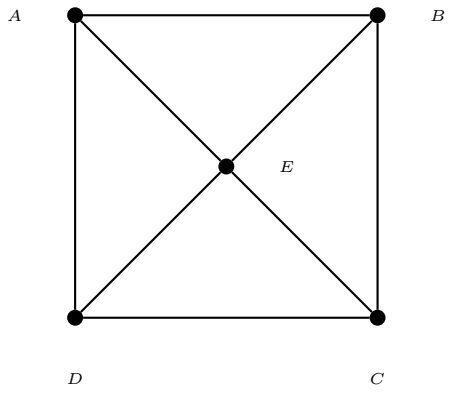


Figure 4.27: Graphical log-linear model with generating class $\{\{A, B, E\}, \{A, D, E\}, \{C, B, E\}, \{C, D, E\}\}$, $\log m(a, b, c, d, e) = u + u_A(a) + u_B(b) + u_C(c) + u_D(d) + u_E(e) + u_{AB}(ab) + u_{CB}(cb) + u_{CD}(cd) + u_{AD}(ad) + u_{AE}(ae) + u_{BE}(be) + u_{CE}(ce) + u_{DE}(de) + u_{ABE}(abe) + u_{CBE}(cbe) + u_{CDE}(cde) + u_{ADE}(ade)$.

We want a label invariant model $\boldsymbol{M}(G,\Gamma)$ associated to this graphical log-linear model. To do this, first we get all n-term orbits, n= 1, 2, 3. The 1-term orbits or vertex orbits are

$$\{A,C\},\ \{B,D\}, \text{and } \{E\}.$$

The 2-term orbits or edge orbits are determined using that

$$\{A,E\}=\{\sigma_1 C,\sigma_1 E\},\ \{A,D\}=\{\sigma_1 C,\sigma_1 D\},\ \{A,B\}=\{\sigma_1 C,\sigma_1 B\},$$

$$\{A,B\}=\{\sigma_2 A,\sigma_2 D\},\ \{C,B\}=\{\sigma_2 C,\sigma_2 D\},\ \{B,E\}=\{\sigma_2 D,\sigma_2 E\},$$

so that the classes or edge orbits are

$$\{\{A,E\},\ \{C,E\}\},\ \{\{B,E\},\ \{D,E\}\},\ \{\{A,B\},\ \{A,D\},\ \{C,B\},\ \{C,D\}\}.$$

The 3-term orbits are determined using that

$$\{A,B,E\}=\{\sigma_1 C,\sigma_1 B,\sigma_1 E\},\ \{A,D,E\}=\{\sigma_1 C,\sigma_1 D,\sigma_1 E\},$$

$$\{A,D,E\}=\{\sigma_2 A,\sigma_2 B,\sigma_2 E\},\ \{C,D,E\}=\{\sigma_2 C,\sigma_2 B,\sigma_2 E\},$$

so that the class or 3-term orbit is

$$\{\{A,B,E\},\ \{C,B,E\},\ \{C,D,E\},\ \{A,D,E\}\}.$$

These orbits allow us to obtain restrictions for the parameters in the same orbit such that if they are equated for all possible values that jointly can take all variables the model obtained is the same after permuting labels. This means that under these restrictions the expected frequencies after permuting A and C remain and also that the expected frequencies do not change after permuting B and D.

The only missing part is obtaining the conditions that allow having scale invariance. Then we can get the classes whose parameters should be equated to get a label invariant model. Using the classes suggested in a) above for second-order interactions, we have the following classes for the associated restricted model. The vertex or main effects classes are

$$V_1=\{A,C\},\ V_2=\{B,D\}, \text{and } V_3=\{E\}.$$

The edge or first-order interaction classes are

$$
\begin{aligned}
E_1 &= \{u_{AB}(ii), u_{AD}(ii), u_{CB}(ii), u_{CD}(ii);\ i = 1, ..., J\},\\
E_2 &= \{u_{AB}(ij), u_{AD}(ij), u_{CB}(ij), u_{CD}(ij),\ i \neq j;\ i = 1, ..., J\},\\
E_3 &= \{u_{AE}(ii), u_{CE}(ii);\ i = 1, ..., J\},\\
E_4 &= \{u_{AE}(ij), u_{CE}(ij),\ i \neq j;\ i = 1, ..., J\},\\
E_5 &= \{u_{BE}(ii), u_{DE}(ii);\ i = 1, ..., J\},\\
E_6 &= \{u_{BE}(ij), u_{DE}(ij),\ i \neq j;\ i = 1, ..., J\}.
\end{aligned}
$$

The second-order interaction classes are

$$
\begin{aligned}
A_1 &= \{u_{ABE}(iii), u_{CBE}(iii), u_{CDE}(iii), u_{ADE}(iii);\ i = 1, ..., J\},\\
A_2 &= \{u_{ABE}(ijk), u_{CBE}(ijk), u_{CDE}(ijk), u_{ADE}(ijk),\ i = j, j \neq k; i, j, k = 1, ..., J\},\\
A_3 &= \{u_{ABE}(ijk), u_{CBE}(ijk), u_{CDE}(ijk), u_{ADE}(ijk),\ i \neq j, j = k; i, j, k = 1, ..., J\},\\
A_4 &= \{u_{ABE}(ijk), u_{CBE}(ijk), u_{CDE}(ijk), u_{ADE}(ijk),\ j \neq i, i = k; i, j, k = 1, ..., J\},\\
A_5 &= \{u_{ABE}(ijk), u_{CBE}(ijk), u_{CDE}(ijk), u_{ADE}(ijk),\ i \neq j,\ i \neq k,\ j \neq k;\ i, j, k = 1, ..., J\}.
\end{aligned}
$$

If we used the process explained before in equations (4.9), (4.10), and (4.11), the second interactions classes are

$$
\begin{aligned}
A'_1 =& \{u_{ABE}(iii), u_{CBE}(iii), u_{CDE}(iii), u_{ADE}(iii);\ i = 1, ..., J\},\\
A'_2 =& \{u_{ABE}(ijk), u_{CBE}(ijk), u_{CDE}(ijk), u_{ADE}(ijk),\ i = j, j \neq k; i \neq j, j = k; \text{or } j \neq i, i = k;\\
& \quad i, j, k = 1, ..., J\},\\
A'_3 =& \{u_{ABE}(ijk), u_{CBE}(ijk), u_{CDE}(ijk), u_{ADE}(ijk),\ i \neq j,\ i \neq k,\ j \neq k;\ i, j, k = 1, ..., J\}.
\end{aligned}
$$

Observe that $A_2 \cap A_3 \cap A_4$ is equal to the set A'_2 and that $A_1 = A'_1$ and $A_5 = A'_3$. We could use even more restricted classes

$$
\begin{aligned}
A''_1 =& \{u_{ABE}(iii), u_{CBE}(iii), u_{CDE}(iii), u_{ADE}(iii);\ i = 1, ..., J\},\\
A''_2 =& \{u_{ABE}(ijk), u_{CBE}(ijk), u_{CDE}(ijk), u_{ADE}(ijk);\ \text{with not all levels equal;}\\
& \quad i, j, k = 1, ..., J\}.
\end{aligned}
$$

Observe that $A'_2 \cap A'_3$ is equal to the set A''_2 and that $A'_1 = A''_1$.

Supposing we were interested in a level invariant model $\boldsymbol{M'}(G, \Gamma')$ using the same graph G and permutation set, *i.e.* $\Gamma = \Gamma'$, then we still could use the orbits to get the corresponding restrictions. In fact, we could use the same representation used for the label invariant model, but as we do not need scale invariance, we could also use a less

restricted model with the following classes.

The vertex classes are the same, so that we equate the main effects for the elements in the same class for each level.

The edge classes are

J^2 classes $\{u_{AB}(ij), u_{AD}(ij), u_{CB}(ij), u_{CD}(ij)\},\ i, j = 1, ..., J.$
J^2 classes $\{u_{AE}(ij), u_{CE}(ij)\},\ i, j = 1, ..., J.$
J^2 classes $\{u_{BE}(ij), u_{DE}(ij)\},\ i, j = 1, ..., J.$

The second-order interaction classes are

J^3 classes $\{u_{ABE}(ijk), u_{CBE}(ijk), u_{CDE}(ijk), u_{ADE}(ijk)\},\ i, j, k = 1, ..., J.$

These classes determine the restrictions, which correspond to equate the elements in the same class.

We observe then that the necessary theory to obtain the restrictions that determine general label invariant models and general level invariant models is similar to the one seen before for the non-general label and level invariant models; however, if we wanted to fit these models for specific data, we would need to write programs to fit models of this kind because general label and level invariant models are no longer RGLL models.

Chapter 5

Related work, discussion, and conclusions

5.1 Related work

Symmetry for discrete data has a long history, it was first studied by McNemar (1947) who tested that $p(1,.) = p(.,1)$ for matched pairs with binary outcomes, this test corresponds to marginal homogeneity but it is also equivalent to test symmetry in two-dimensional contingency tables with binary variables. Bowker (1948) developed a test for two-way contingency tables with the same number of categories to see if a table was symmetric. Quasi-symmetry models were defined and studied by Caussinus (1965). For two-dimensional tables Goodman (1985) defined generalized symmetry models: triangle asymmetry, RC asymmetry, and diagonal asymmetry, which are models with additional parameters modifying the distribution. The diagonal asymmetry model satisfies $p(i,j) = \rho_{ij}\delta_k;\ k = i - j,\ i,j = 1,...,J$ where ρ_{ij} and δ_k are the parameters of the model, the RC asymmetry model corresponds to quasi-symmetry, and the triangle asymmetry model is equivalent to conditional symmetry as defined by McCullagh (1978). Conditional symmetry or triangle asymmetry are models with an additional parameter for the elements above the diagonal and another parameter for elements below the diagonal, even though the association between those elements is the same. McCullagh (1978) also studied symmetry models for ordinal variables using what he defined as *palindromic symmetry*.

Symmetry generalizations for three-dimensional tables were presented by Bishop *et al.* (1975, p. 299-309) who defined complete symmetry based on an *interchangeability* concept given by Madansky (1963), even though Madansky's work was more related to marginal homogeneity generalizations. Bishop *et al.* (1975, p. 303) also presented quasi-symmetry generalizations for three-way contingency tables. Symmetry and quasi-symmetry generalizations for higher dimension were discussed by Darroch and Bhapkar (1990). Other symmetry generalizations, for example marginal quasi-symmetry, were

presented by Andersen (1991, p. 328-329) and Bishop *et al.* (1975, p. 299-309) and we discussed them in section **2.6**, where in fact most of the generalizations mentioned in the last two paragraphs were presented and discussed.

Symmetry generalizations for more than two dimensions not only include generalizing symmetry as defined for two-way contingency tables, other generalizations correspond to marginal homogeneity extensions for any number of variables or restricting marginal distributions as studied by Bergsma and Rudas (2002), which can be applied according to Madansky (1963) to panel data in which a variable is followed through time and the researcher wants to know if the marginal distribution does not change.

Wermuth and Marchetti (2009) generalize symmetry for binary variables in the sense that the levels are equally probable and at the same time they assume the existence of something called a triangular system, which is a recursive process in terms of the main effects. Under these triangular systems for symmetric binary variables, the symmetry in all margins is carried over the joint probabilities, for example for three variables with levels 1 and -1, $p_{123}(1,-1,1) = p_{123}(-1,1,-1)$, *i.e.* the distribution is preserved after interchanging levels, which does not happen with RGLL models, even with label invariant models because in these models scale invariant conditions mean that what is preserved after that interchange is the model, not the distribution. Wermuth and Marchetti (2009) also discuss how certain graphical Markov models can be expressed using those symmetric models for triangular systems, which is interesting because they can be easily expressed in terms of simple correlation and concentration matrices defined for discrete data.

Equating parameters in log-linear models has been presented for instance by Haberman (1979, p. 503-509) in one specific example and also discussed by Vermunt (2005) and Rindskopf (1984). Models that combines symmetry by equating certain parameters and graphical log-linear models are presented by Gottard (2009) and Gottard *et al.* (2008, 2010), we will discuss more about this model below.

Graphical models for discrete data and undirected graphs or graphs representing only independences have been extended in different ways. For instance, Drton and Richardson (2008) defined a model class called binary models for marginal independence providing a framework for modeling marginal independence instead of conditional independence as in traditional graphical log-linear models. Also, van Horebeek and Teugels (1998) defined models in which the presence or absence of an edge between two variables depends on the values taken by the remaining variables so that there is conditional independence for particular values taken by the variables in the conditional part and for other values there is no such independence. RGLL models are another way of generalizing graphical log-linear models combining graphical log-linear models and restrictions on certain parameters of the model.

We have defined RGLL models as a way of getting something analogous to what was done by Højsgaard and Lauritzen (2005, 2007, 2008) in the Gaussian case with the so-called graphical Gaussian models with edge and vertex symmetries, which include RCON and RCOP models. RGLL models were also defined as a way of generalizing symmetry in graphical log-linear models. The analogy with graphical Gaussian models with edge and vertex symmetries is obtained once we impose constraints on the main effects and first-order interaction parameters of a graphical log-linear model. In RCON models restrictions are imposed on the concentration matrix, which determines association between variables and consequently the edges in the graph. The diagonal terms equated in that matrix determine vertex colourings and the non-diagonal terms equated determine edge colourings. The equivalent terms in the discrete case are main effects and first-order interactions, respectively, so that we chose to define models in which we equate these terms. RCOP models are models that preserve the distribution after permuting some vertices according to a permutation group that preserves the graph. We have defined label invariant models for the discrete case as models equivalent to RCOP.

Currently, Neufeld (2009) continues the study of graphical Gaussian models with edge and vertex symmetries, in particular her research is devoted to the identification of RCOP models, where her primary goal is to identify when a coloured graph represents a RCOP model and if not how to modify it so that it does.

Gottard (2009) and Gottard *et al.* (2008, 2010) defined models called *quasi-symmetric (QS) graphical models* and their general form *Symmetric and quasi-symmetric (SQS) graphical models* that, like RGLL models, preserve the independences given by a graph. SQS graphical models use a vertex colouring definition similar to the one presented here, but for edge colourings, QS and SQS graphical models use restrictions of the kind

$$\lambda_{XY}(ij) = \lambda_{RS}(ij), \text{ for all } i, j = 1, ..., J$$

for elements in the same colour class, including restrictions

$$\lambda_{XY}(ij) = \lambda_{XY}(ji), \text{ for all } i, j = 1, ..., J \text{ and } \lambda_{RS}(ij) = \lambda_{RS}(ji), \text{ for all } i, j = 1, ..., J;$$

and with restrictions for the second-order interactions of the type

$$\lambda_{V}(ijk) = \lambda_{V}(jik), \text{ for all } i, j = 1, ..., J;$$

where V is a set formed by three variables, or similar for different positions of k according to which first-order interactions are in the model. For example if we had the parameters $\lambda_{XY}(ij)$ and $\lambda_{XYR}(ijk)$ in the model, we would have the restrictions: $\lambda_{XYR}(ijk) = \lambda_{XYR}(jik)$ for all $i, j, k = 1, ..., J$. SQS graphical models are defined for

at most three-way contingency tables; however, Gottard (2009) says that it can be naturally extended to higher dimensions, she does not show how, but refers to Gottard *et al.* (2010) where only QS graphical models are defined for tables with any dimension by restricting all interactions as with second-order interactions.

Complete symmetry and quasi-symmetry preserving only one-dimensional margins defined in section **2.6** are particular cases of SQS graphical models. Observe that the first-order restrictions in these models are a particular case of the kind of restrictions defined for those terms in RGLL models, and as a consequence for triangle-free graphs QS and SQS graphical models are a particular case of RGLL models. On the other hand, label invariant models could not be seen as QS or SQS graphical models and viceversa. When we have interactions of second or higher order the models differ because of the definition used for second-order interactions restrictions in QS and SQS graphical models. Additionally, the kind of restrictions included in QS and SQS graphical models have some computational advantages because the restrictions are preserved in spite of the parametrization used, so that we can easily find the restrictions for the parameters in a parametrized model whose design matrix has full rank, sum the appropriate columns in the design matrix, and solve the model by using Newton-Raphson, which implies we could use any available statistical software that fits log-linear models without the necessity of programing a software as REGRAPH, like *SPSS*, *SAS*, *R*, or *SPlus*.

In Gottard *et al.* (2010), the authors fit QS graphical models for a data set corresponding to attitudes about legalized abortion and about death penalty. The data consist of four binary variables with the same scale, *yes* and *no*, which can be coded as 1 and 2, respectively. The first three variables response to whether abortion should be legal: a) when there is a strong chance of a serious defect in the baby, variable D, b) when the mother's health is endangered, variable H, and c) when the pregnancy is the result of a rape, variable R. The last variable is whether a subject favors death penalty, variable P. Variables D, H, P, and R are labeled as 1, 2, 3, and 4, respectively. The QS graphical model that fitted the data according to the deviance has generating class $\{\{1,2,4\}, \{3,4\}\}$ and two edge colour classes (E_0, E_1), where $E_0 = \{\{3,4\}\}$ and $E_1 = \{\{1,2\}, \{1,4\}, \{2,4\}\}$, which according to their definition of colouring implies the following parameter restrictions:

$$u_{12}(12) = u_{14}(12) = u_{24}(12) = u_{12}(21) = u_{14}(21) = u_{24}(21),$$
$$u_{12}(11) = u_{14}(11) = u_{24}(11),$$
$$u_{12}(22) = u_{14}(22) = u_{24}(22),$$
$$u_{124}(ijk) = u_{124}(jik);\ u_{124}(ijk) = u_{124}(kji);\ u_{124}(ijk) = u_{124}(ikj),\ \text{for } i,j,k = 1,2.$$

The last equalities are equivalent to $u_{124}(ijk) = u_{124}(per(ijk))$, where $per(ijk)$ denotes any permutation of the elements in the argument. The deviance has a value

of 8.80 with eight degrees of freedom and a p-value of 0.36, indicating a good fit. Note that this is not a RGLL model because there are equality restrictions for second-order interactions. However, we can search for a RGLL model with the same generating class that fits the data by using the selection method included in REGRAPH. The RGLL model obtained has three vertex and seven edge colour classes as follows.

$$V_1 = \{2\},\ V_2 = \{3\},\ V_3 = \{1, 4\};$$
$$E_1 = \{u_{12}(22)\},\ E_2 = \{u_{14}(12)\},\ E_3 = \{u_{14}(22)\},\ E_4 = \{u_{24}(21)\},\ E_5 = \{u_{12}(12), u_{34}(21)\},$$
$$E_6 = \{u_{24}(11), u_{34}(12), u_{14}(11), u_{34}(11), u_{24}(12), u_{12}(21), u_{12}(11), u_{34}(22)\},$$
$$E_7 = \{u_{14}(21), u_{24}(22)\}.$$

This model has an associated deviance of 3.49 with six degrees of freedom and a p-value of 0.75, so that it fits even better than the QS graphical model. The QS graphical model has an interpretation that the RGLL does not have, which is that there is a quasi-symmetric structure preserving one-dimensional margins for variables 1, 2 and 4 as seen in section **2.6**. The RGLL model that may be considered similar to the QS graphical log-linear model is one where the same first-order interaction restrictions are used; however, note that the second-order interactions are not restricted. That is, we have a RGLL model with the same generating class as before and four vertex and seven edge colour classes as follows.

$$V_1 = \{1\},\ V_2 = \{2\},\ V_3 = \{3\},\ V_4 = \{4\};$$
$$E_1 = \{u_{12}(12), u_{14}(12), u_{24}(12), u_{12}(21), u_{14}(21), u_{24}(21)\},$$
$$E_2 = \{u_{12}(11), u_{14}(11), u_{24}(11)\},\ E_3 = \{u_{12}(22), u_{14}(22), u_{24}(22)\},$$
$$E_4 = \{u_{34}(11)\},\ E_5 = \{u_{34}(12)\},\ E_6 = \{u_{34}(21)\},\ E_7 = \{u_{34}(22)\}.$$

Using REGRAPH, we obtained a deviance of 6.65 with six degrees of freedom and a p-value of 0.35, which is similar to the p-value corresponding to the QS graphical model; in fact, using the deviance and degrees of freedom differences between the two models, 2.15 and 2, respectively, we obtained a p-value of 0.34. Then, we do not reject, for instance using $\alpha = 0.05$, that both models explain in a similar way the data; however, as stated before, the QS graphical model fitted to these data can be somewhat interpreted.

5.2 Discussion

Regardless of having analogies between graphical Gaussian models with edge and vertex symmetries in the continuous case and RGLL models in the discrete case, as discussed above in section **5.1**, there are some differences. The most important differences are that in the discrete case we have terms indicating association between three or more

variables, the interactions of two or higher order, and that we take into account all level or category permutations between two variables to determine edges and parameter restrictions, whereas in the continuous case we only consider pairs of variables to determine edges, which depend on the concentration matrix or partial correlation terms. Another difference is that in the discrete case we could have different parameters according to the parametrization used. Additionally, observe that there are scale invariant conditions added to label invariant models not required for RCOP models because they are automatically obtained.

Symmetry generalizations that are particular cases of RGLL models not only correspond to quasi-symmetry models, but also to other models like conditional symmetry as discussed in section **2.6**. We have chosen to use both the parametrization and the kind of restrictions used by Agresti (2002a, p. 423-426) for symmetry and quasi-symmetry models, but we could have used and identify restrictions through any other kind of parameter available, for example parameters under effect coding. In fact, we saw in section **3.9.4** that for any vertex colouring and in three particular cases of edge colourings, the restrictions with any kind of parametrization are exactly the same. The disadvantage of restricting parameters under other parametrizations instead of the parameters of the kind $u_a(i_a)$ (the ones used for this dissertation) is that using parameters different to $u_a(i_a)$ we can not restrict any possible permutation of the levels taken by a pair of variables because, once we parametrize the model in other way, there are permutations not represented by any parameter.

We can obtain design matrices for RGLL models; however, they do not have associated full-rank matrices unless we get a parametrization of the model satisfying this property that represents the model with all its restrictions. There are cases in which we can easily get this parametrized model, for example symmetry and quasi-symmetry models or the cases we analyzed in section **3.9.4**, but in general it is not straightforward. This is the reason why we chose to fit the models using a modified iterative proportional fitting algorithm, which provides us with a numerical method that solves the likelihood equations converging to the maximum likelihood estimators and that does not use the design matrices. If we had such parametrized model, we would get full-rank matrices and we could use the Newton-Raphson or Fisher scoring method to solve the likelihood equations by numerical approximation. This means that we could use any available software that solves log-linear models, for example *SPlus*, *SPSS*, *SAS*, or *R*, to fit a RGLL model with such parametrization. Even if the design matrices for RGLL models are not full-rank matrices, they are important because they are used to obtain the number of degrees of freedom associated to a model.

It might be interesting to consider only parameter restrictions that not depend on the parametrization used, which as it was said above always happens for vertex colouring and for the particular edge colourings presented in section **3.9.4** and consequently

only for certain vertex and edge colourings; however, by only considering models of this kind we are restricting even more the number of models and not considering many interesting models, which is inconvenient.

We have chosen the likelihood ratio test statistic defined for log-linear models, the deviance, as a goodness of fit statistic to see if a model fits well and we also used it to verify the convenience of joining classes and to select a model. This statistic has an asymptotic chi square distribution. There are no general rules to determine if this is an adequate approximation, but in general Lauritzen (1996, p. 81) suggests that the approximation requires $m(i)$ rather large for all $i \in I_k^* = \{i|m_k(i) > 0\}$. The exact distribution is in general intractable as discussed by Lauritzen (1996, p. 80), and this is why we have to rely on asymptotic results. However, there are some particular cases, for example quasi-symmetry models which are also RGLL models, in which exact p-values were numerically calculated by Booth *et al.* (2005) and there are also other methods, *e.g.*, Whittaker (1990, p. 285-296), that pretend to obtain p-values that are closer to the exact values by conditioning the cell counts on the minimal sufficient statistics, and there is even a software developed by Caffo (2008) that implements them for log-linear models.

Another goodness of fit statistic we have used is the Pearson X^2 statistic, in fact, we saw in section **2.7** that a Taylor expansion can be used to approximate the deviance with X^2. Some properties of the deviance are the following: a) it can be used to compare models because the deviance can be decomposed into the sum of other deviances, which is useful in model selection, b) it can be seen as a divergence measure and it is minimized when we get the maximum likelihood estimators, c) it has an interpretation in terms of hypothesis testing, d) in general, for higher dimension but not sparse tables it is always advisable to use it because it provides a better approximation. The Pearson X^2 statistic has a better chi squared approximation for smaller sample sizes and more sparse tables than the deviance, in fact the distribution of the deviance is usually poorly approximated when the sample size divided by the number of cells is less than 5 as it is the case in one of the applications. We used the deviance throughout the work; however, we presented the statistic X^2 in all examples in Chapters **3** and **4**, the chapters where RGLL models were fitted, specially for the example 4.12 where there are sampling zeros and a sparse table.

In order to fit RGLL models, we needed a specific computer program, and we wrote REGRAPH in *Fortran 90*. We wrote subprograms in it that: fit RGLL models, join colour classes, and select RGLL models for some data. As discussed at the end of section **3.10**, for log-linear models without restrictions including graphical log-linear models and symmetry and quasi-symmetry models, numerical results were compared with those obtained using other available software like *MIM* or *Splus*. These models were chosen because in all these cases we know adequate parametrizations with asso-

ciated full-rank matrices that represent them. In REGRAPH, the user has to provide among other things the generating class, so that it has to be known. In practical applications, the generating class of the model is not known, so that we could use a model search procedure to look for a graphical log-linear model that fits the data. This search can be done through available software that fits graphical models, like *CoCo* or *MIM*. We have used the last one in this work.

We have developed methods that join colour classes, where we have used the deviance as a measure to decide whether a joint is convenient. We used these methods to get a model search procedure using any initial vertex and edge colour class or even a default model in which every vertex and every edge is in an atomic class. We generated a method that joins iteratively classes whose joint is statistically significant until we get a RGLL model that, with the information provided, better fits the data according to the deviance. That significance is given in the sense that using deviance differences we do not reject the null hypothesis that the parameters corresponding to the same colour class once joining classes are equal in the model without joining classes. This method could be considered as some kind of Forward algorithm, because we begin from an initial model and we iteratively join colour classes until we get a final model. It could be interesting to try to implement a similar Backward algorithm, in which starting from a model in which all edges and vertices are in the same class, we iteratively separate the classes until we get one model that fits as good as possible to the data.

As any automatized selection method, ours is useful as a guide to select a model, but it does not provide the best model or the model that best represents the relations among all cells and variables. In fact, depending on the model, we could suggest that more than one colouring is adequate even before fitting the model. This is because the equations corresponding to the elements in the generating class sometimes automatically imply the equations corresponding to various colourings. This means that given a generating class more than one colouring could fit the data according to the deviance, in the sense that the statistic is the same for any of the colourings. We emphasize that it is important to identify an adequate generating class as starting point to get a RGLL model that fits well.

We have considered models in which all variables have the same categories. The reasons for this assumption are the following: 1) If we want to join any vertex colour classes or vertices in the same class, we need all vertices to have the same number of categories, like in symmetry and quasi-symmetry models, because when we join vertices in one class we equate main effects for all levels for the variables in the same class. 2) It allows symmetry interpretations, which makes it a necessary assumption for label and level invariant models; for example, the former require having the same categories because we permute vertices, and this change only make sense when we have variables whose values are the same even though the variables differ. This difference could be for

example that a variable is measured in different periods of time or for different members in a family. 3) By using this assumption, computer programming is less difficult and more efficient and we also have the possibility to include model selection. 4) This assumption is similar to what is required in symmetry and quasi-symmetry models, in which we consider only square tables, or even in some symmetry generalizations, for example complete symmetry that requires all variables to have the same number of categories.

All programs in REGRAPH run under the assumption that all variables have the same number of levels or categories. The assumption that all variables have the same levels could be relaxed by asuming that only the variables in the same composite class have the same categories. We could fit RGLL models with these characteristics; however, we would not have the possibility to join any vertex classes if they have different categories and consequently we could not apply the automatized model selection method as before. If we wanted to include this kind of models, we would need to modify REGRAPH to fit them.

We also have supposed that we do not have structural zeros, *i.e* that $p(i) > 0$ or that the expected frequencies $m(i) > 0$ if the cells counts correspond to multinomial or Poisson distributed random variables, respectively, although we could have sampling zeros, *i.e.* observed counts $n(i) = 0$ for some i. When $n(i) = 0$, it might be necessary to adjust the corresponding degrees of freedom. However, the existence of maximum likelihood estimators even when we only have sampling zeros is a topic that has been deeply discussed with many questions still open, because the patterns of these random sampling zeros determine which log-linear models can be fitted.

Fienberg and Rinaldo (2007) discuss and give a historical review on the existence of maximum likelihood estimators in the presence of sampling zeros. Haberman (1973) was the first who gave necessary and sufficient conditions for their existence; however, as Rinaldo (2006b) points out, he gave a non-constructive characterization in the sense that it does not lead to numerical procedures to know when the maximum likelihood estimators exist. One of his results corresponds to existence of maximum likelihood estimators when the observed counts $n(i) > 0$ for all i and also justifies to some extent that sometimes in practice the presence of sampling zeros is dealt with adding small positive quantities to zero cells.

Bishop *et al.* (1975, p. 69) also discuss that when the cell estimates can be derived directly, then having positive values for the sufficient statistics ensures the existence of maximum likelihood estimators. That the cell estimates can be derived directly means that there are exact formulas to estimate the expected frequencies and having positive values for the sufficient statistics means that all marginal counts $n_a(i_a)$ that define the likelihood equations are positive. This result is equivalent, as discussed by Glonek

et al. (1988), to say that positivity of those marginal counts is a necessary and sufficient condition for the existence of maximum likelihood estimators if and only if the model is decomposable. In non-decomposable log-linear models positivity of such marginal counts is only a necessary condition for the existence of the estimators. That the condition is only necessary means that there are non-decomposable log-linear models and associated contingency tables whose marginal counts are positive and still there is no maximum likelihood estimator. Examples of this phenomenon are shown by Glonek *et al.* (1988), but it is something rare and the cases shown correspond to hypothetical tables specially built to prove this happens.

Some authors, *e.g.*, Bishop *et al.* (1975, p. 115) or Christensen (1997, p. 286-293), suggest that even with zero marginal counts, in either decomposable or non-decomponsable models, we still could fit a model by dropping all cells that cause such zero marginal counts obtaining estimators for the remaining cells and after that adjusting the degrees of freedom as discussed in section **2.7**. Those estimators would be some kind of extended maximum likelihood estimators. Using the IPF algorithm, such process is obtained by equating to zero all expected frequencies for those cells whose marginal counts are zero. REGRAPH does this process, it equates to zero all cells whose marginal counts are zero and adjusts the degrees of freedom, but warns the user that when at least one of the marginal counts is zero, the estimators do not exist and the estimated values do not correspond to maximum likelihood estimates.

In practice indicators which have been used to check that the estimators do not exist are given by the lack of convergence or errors of whatever algorithm is used to compute the estimators and by the fact that under IPF the fitted values are equal to the observed values. This is justified by a theorem presented by Glonek *et al.* (1988) stating that the maximum likelihood estimators exist if and only if there exists a table x of positive real values with $n_a(i_a) = x_a(i_a)$, for all $a \in A$. These equalities correspond to the likelihood equations, then the theorem is equivalent to say that the estimators exist if and only if the expected frequencies that solve the likelihood equations $\widehat{m}(i)$ are positive. In fact, Andersen (1974) shows that under IPF the estimators exist if and only if the expected frequencies to which the algorithm converges are positive. Lack of convergence means that the likelihood equations are not solved, therefore we do not find positive values satisfying the equations. On the other hand, if the estimated frequencies are equal to the observed values, then the likelihood equations are solved by expected frequencies with zero values because the original table had zeros, which implies that the estimators do not exist.

Lauritzen (1996, p. 71-72) does not deal with the problem of the existence of estimators because he works with models for which he defines an enlarged parameter space where the maximum likelihood estimators are always defined in an extended way.

Recent advances in the field of algebraic statistics have been used by Eriksson *et al.* (2006) to determine polyhedral conditions for the non-existence of the estimator in hierarchical log-linear models, these results were generalized by Rinaldo (2006a,b) to include any log-linear model.

All these results on the existence of maximum likelihood estimators when some observed counts are zero can be extrapolated to RGLL models. If 1) the graphical log-linear model defined by the generating class associated to a RGLL model is a decomposable model; and 2) the marginal counts associated to the elements in the generating class with cardinality greater than three as well as the sums of marginal counts associated to the colourings are strictly positive, then there are necessary and sufficient conditions for the estimators to exist. Condition 1), that the model is decomposable, is obvious since the RGLL model and the associated graphical log-linear model are both log-lineal models with the same generating class, and condition 2) is equivalent to say that the corresponding marginal counts associated to the RGLL model should be positive. In fact, if the marginal counts for the generating class are positive then the marginal counts for the RGLL model are also positive because the latter are obtained by summing the former marginal counts depending on the vertex and edge classes. When the model is not decomposable we fall on the case in which the positivity of the marginal counts is only a necessary condition for the existence of maximum likelihood estimators and then, even having such positivity, we still should check if the estimated expected frequencies are equal to the observed values in all cells, if the algorithm converges, and if $\widehat{m}(i) > 0$ in order to see if there is an existence problem.

We observe that in example 4.12 there are 7 zeros out of 16 cell values and in consequence the maximum likelihood estimators could not exist, additionally as we discussed in section **2.10** all models whose associated graph is a 4-cycle are not decomposable so that we can not use the result seen above involving positive margins and decomposable models; however, all margins associated to the model, $n_{L_1B_2L_2B_1}(i,j,.,.)$, $n_{L_1B_2L_2B_1}(.,i,j,.)$, $n_{L_1B_2L_2B_1}(.,.,i,j)$, $n_{L_1B_2L_2B_1}(j,.,.,i)$; $i,j,=0,1$, are positive and in consequence the necessary condition is satisfied but as it is not a sufficient condition we could still have problems. But, we observed that the fitted values were not equal to the observed counts for all cells, and more important there were not convergence problems, *i.e.* the maximum likelihood equations are solved, and all estimated expected frequencies $\widehat{m}(i)$ are positive so that sufficient conditions are satisfied and the maximum likelihood estimators exist.

If we considered structural zeros we would be fitting models for incomplete tables. The theory to fit models for incomplete tables is different to the one used for complete tables and it was presented for hierarchical log-linear models by Bishop *et al.* (1975, p. 177-228), who fitted models for incomplete two-way contingency tables, in particular for quasi-independent models, and for tables with different zero structures and presented

a general theory for incomplete multi-way tables establishing conditions for maximum likelihood estimators existence as well as for degrees of freedom adjustment.

Additionally, we supposed that we do not have missing values, *i.e.* we do not have incomplete data. If we wanted to include this kind of data, we would have to use other methods, for example some kind of EM algorithm (expectation and maximization), to fit the model. This method was presented by Edwards (2000, p. 103-117, 313-314) for mixed models, graphical models with continuous and discrete variables, and also by Lauritzen (1996, p. 233-236), who also explains how the method could be applied to hierarchical log-linear models.

A particular kind of RGLL models are label invariant models and they correspond to scale invariant models whose distribution is preserved after permuting some variables, so that symmetry is seen in terms of a model that is preserved after applying a permutation set formed by automorphisms to the vertices. They can also be considered as graphical models with certain graph symmetry, understanding this symmetry as a graph that is preserved after a permutation of the vertices, including all conditional independences that can be inferred from that graph. As we said before, the equivalence between label invariant models and RCOP models is completed once we include scale invariant conditions. Label invariant models can be represented as RGLL models considering that we restricted their graph structure to be triangle-free, which produces restrictions for the parameters as the ones defined for RGLL models. This restriction makes also possible to have a graphical representation of the model. Such restriction is not something rare, in fact, there are some graphs called *pairwise Markov graphs* where it is supposed that there is only second-order dependence, which is equivalent to that restriction. There are even some particular models as discussed by Hastie *et al.* (2009, p. 638) called *Ising models* in the statistical mechanics literature or *Boltzmann machines* in the machine learning literature for discrete binary variables that can be seen as first-order interaction Poisson log-linear models. All restrictions in label invariant models can be derived using concepts equivalent to the ones defined for the Gaussian case, basically using the concept of orbit.

Another type of RGLL models are level invariant models. They are defined as models including restrictions corresponding to equalities of the expected frequencies for different cells according to a permutation group, generalizing the concept of symmetry in contingency tables in the sense that we have a model that equates the expected frequencies in certain cells. This generalization does not only allow symmetry between some cells, but also preserves the conditional independences given by the model, which differs to what has been done before because in other symmetry generalizations the saturated model is used and this model does not consider any kind of independence.

We proposed ways of generalizing label and level invariant models in the sense that

these general models are defined for any graphical log-linear model and not only for graphical log-linear models whose associated graphs are triangle-free. In general label invariant models, we get the necessary restrictions for interactions of any order by defining generalized orbits. Scale invariance conditions differ because we can get not only one set, but many sets of restrictions for the interactions of two or higher order that could be used and we can choose any of them. We can also get general level invariant models. In these, when the set of permutations defining the models is formed by automorphisms, we can use the generalized orbits to get the corresponding restrictions. Otherwise, the restrictions are obtained once we obtain the log-linear expansions for all the expected frequencies that should be equal and we deduce which parameters should be the same to get such equalities. We obtain models in which we restrict main effects and all interactions. By restricting interactions of order higher than one we do not have a RGLL model anymore and we have models that can not be represented by graphs. This means that they can not be solved with any available software, nor even with REGRAPH.

RGLL models and in particular label and level invariant models can be fitted for different kinds of data, considering that the variables involved have the same categories. For example, we could use them in panel data, in particular when a discrete variable is being followed through time, in studies using clustering or matching, *e.g.*, social mobility studies for grandfather-father-son triads or twin or brothers studies, in agreement studies, or even in item-response studies. Under the condition that the number of units, which are individuals, clusters, objects rated, or subjects taking a test, respectively, do not change, we can represent these data as contingency tables as discussed by Lovinson (2000) and it could make sense to use RGLL models. In this document we presented, in addition to the data corresponding to a study of coronary heart disease and those for patients admitted to an intensive care unit presented in section **3.11.5**, panel data for the change of region of residence data presented in section **3.11.5** and studies using clustering for the brothers and twins data presented in examples 4.12 and 4.13.

5.3 Conclusions

RGLL models are more parsimonious than graphical log-linear models and they could help to get a better fit and understanding of the data. According to a model, we could know more about the distribution of the data, for example we can obtain all conditional independences the model represents as in graphical log-linear models, and the relation between the cells on a contingency table; however, we have to be careful with the interpretation of a fitted model, because in some cases it could be too complex to be easily interpreted in terms of the relation between cells, so that what we could obtain is a parsimonious model without gaining in terms of interpretability.

RGLL models are a way of establishing symmetry in terms of a log-linear model and they pretend to be analogous to the RCON models presented for the Gaussian case. They also pretend to extend classic models of symmetry for two-way contingency tables, though this is not the only way to do that. The definition of these models was somewhat arbitrary, in the sense that we have decided which parameters to restrict, and obviously there could be other ways. We chose restrictions that in some way were equivalent to the ones in the continuous case and such that we still could have a graphical representation; however, this representation might become too complex if the model has for example six or more variables and more than three levels or categories because there might be too many colours in the corresponding graph. In fact, problems of graphical representation arise in any graphical model with too many variables; so that a graphical representation for a graphical model is not strictly necessary and sometimes the term *graphical* used to identify the name of these models is more related to the use of concepts of graph theory than to the graphical representation. We could extend RGLL models by restricting additional parameters, but it gets more complicated, or by changing the kind of parameters we restrict as we discussed above in section **5.2**.

Generalizations of RGLL models by restricting additional parameters would mean that restrictions are not only in the main effects and first-order interactions parameters, but also in higher-order interactions. The graphical representation of such models, even with multi-graphs, is not possible, much less when we consider that the equalities involve not only variables but also their values. We could use hyper-graphs, but even so the representation is complex, though as we discussed in the previous paragraph such representation might not be necessary. However, if we obtained these generalizations, more symmetry generalizations, for example complete symmetry or quasi-symmetry presented by Bishop *et al.* (1975, p. 299-306) (see section **2.6**) and general label and level invariant models, would be particular cases. This generalization requires additional work, not only in computational terms but also in theoretical terms because we need to give a new definition, obtain the associated likelihood equations, and define a new algorithm to fit the model and analyze its convergence. It might even be useful to define all concepts in algebraic or even geometric terms analogously to what for instance Rinaldo (2006a) did for log-linear models. Drton *et al.* (2009) present in their book techniques in algebraic geometry, commutative algebra, and combinatronics, to address problems in statistics and its applications including graphical models.

Additionally, label invariant models are analogous to RCOP models defined in the Gaussian case. Amazingly, we found that many of the concepts used in the continuous case could still be used with certain modifications, even though we had to restrict the types of graphs in which the models could be applied. The symmetry generalization we have obtained through label invariant models are in terms of graph theory and the model. Additionally, as we saw in examples 4.12 and 4.13, we can see that we also have a symmetry generalization in terms of the contingency table because we got multi-way

contingency tables in which under certain arrangement of the variables in the table the elements above the diagonal are equal to the ones below. Level invariant models also generalize symmetry in terms of the contingency table and both, label and level invariant models, preserve the dependence structure given by the graph.

According to all those similarities between the discrete and continuous case, future work could consist on extending RGLL models to mixed graphical models. Other future work could be to determine if given any coloured graph and its corresponding multigraph satisfying scale invariance, we can identify if we have a label invariant model. As we stated before, additional work could consist in including the analysis of incomplete tables or tables with missing data. Alternative types of graphical models for discrete data were discussed above in section **5.1**, for example the binary model for marginal independence defined by Drton and Richardson (2008) in which marginal independence is what is represented using graphs with edges having arrows in both ends instead of using simply edges and representing conditional independence as in graphical log-linear models. Additional work could consist on using the concepts used for RGLL models and apply them for those alternative types of graphical models.

In conclusion, RGLL models are extensions for graphical log-linear models and discrete symmetry models that could be useful for describing the structure of a contingency table, and, even though they have limitations, they could be a starting point to understand symmetry in graphical log-linear models with more than two variables and to understand differences and similitudes of this concept between the continuous and discrete case. Additionally, even though we used some theory defined for graphical log-linear models and the models defined by Højsgaard and Lauritzen (2008) as a starting point, we mostly defined everything from scratch, so that doing this work was useful to understand how to build models and the problems, both theoretical and practical, when doing it.

Appendix A

Basic graph theory concepts

A *graph*, $G = (V, E)$, is an ordered pair consisting of a finite set $V(G)$, or V when there is not confusion in which graph we are referring to, of elements called *vertices* or *nodes*, so that $V = \{v_1, v_2, ..., v_k\}$, $k \in \mathbb{N}$, and of a finite set $E(G)$ or E, with $E = \{e_1, e_2, ..., e_l\}$, $l \in \mathbb{N}$, of elements called *edges*, where E is a subset of non-ordered pairs of $V \times V$ of the form $e_r = \{v_i, v_j\}$ for some $i, j = 1, 2, ..., k$ with $r = 1, 2, ..., l$.

The vertices v_i and v_j are the ends of the edge e_r. Two vertices in G are *adjacent* if they are the ends of some common edge; *i.e.*, if there is an edge joining them or if there is an edge in common. Two edges are *adjacent* if they have one end in common, *i.e.* if the edges share one same vertex. Graphically, we represent the vertices as dots or circles and the edges as lines joining the corresponding vertices. Edges are called *multiple* when there are more than one edge in E joining a vertex v_i with a vertex v_j. A *loop* is an edge of the form $\{v_i, v_i\}$. A graph G is *simple* if it does not have multiple edges or loops. A graph is *trivial* if there are not edges between its vertices, *i.e.* if the graph consists of isolated vertices.

The number of vertices in a graph, $|V|$, is denoted as ν and the number of edges, $|E|$, is denoted as ϵ.

A *subgraph* $G' = (V', E')$ of $G = (V, E)$, denoted as $G' \subseteq G$, is a graph in which $V' \subseteq V$ and $E' \subseteq E$. A *spanning subgraph* of G is a subgraph H with $V(H) = V(G)$. A *subgraph induced by a subset of vertices* V' of V, denoted as $G[V']$, is a subgraph with vertex set V' an whose edge set are all edges in G between the elements in V'; *i.e.*, the edge set is the set of those edges in G that have both ends in V'. A *subgraph induced by a set of edges* E', denoted as $G[E']$, is a subgraph with edge set E' and whose vertex set is the set of ends of edges in E'. If we consider the subset $V' = V - V''$, with $V'' \subseteq V$, then the induced subgraph $G[V']$ is a graph in which we delete all vertices in V'' and all edges with ends in that set, we denote it as $G - V''$. In asimilar way, we can get $G - E''$ with $E'' \subseteq E$. If $V' = \{v\}$, then for simplicity we write $G - v$ for $G - \{v\}$.

Similarly for the edges, if $E' = \{e\}$, we write $G - e$ for $G - \{e\}$.

A simple graph is *complete* if all pair of distinct vertices are joined by an edge. A complete graph with ν vertices is denoted as K_ν. The graph K_1 does not contain edges, *i.e.* we have a trivial graph. The graph K_2 is an edge. The graph K_3 corresponds to a graph in which all three vertices are joined between them. If we draw the corresponding graph by joining the vertices with straight lines, then we have a triangle. In a similar way we get all complete graphs.

A *clique* for G is an subgraph induced by a set of vertices $S \subseteq V$, $G[S]$, that is complete and maximal with respect to the contention; *i.e.*, a subgraph G is a clique not only if it is complete, but also if it is not contained in another complete induced subgraph of G. Using this definition, we get all cliques in any graph. For example, the cliques for the graph given in figure A.1 are $\{1, 2, 3\}, \{2, 4\}$ and $\{4, 5, 6\}$.

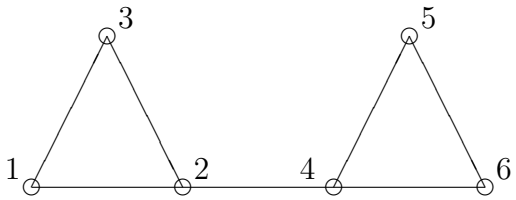


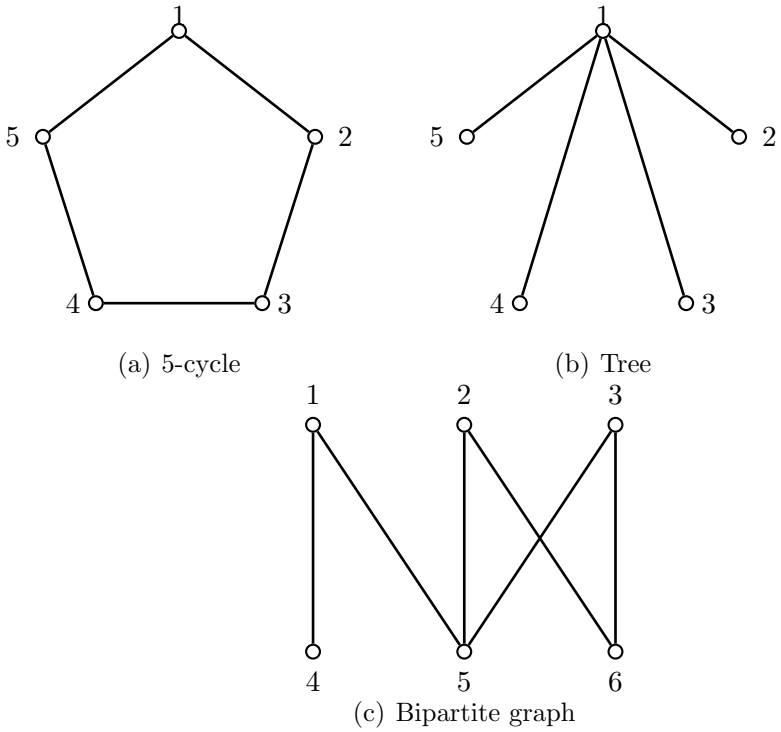
Figure A.1: Graph to illustrate the concept of clique.

A *walk* of length k is a finite non-null sequence of vertices and edges in G of the form $\{v_0, e_1, v_1, e_2, v_2, ..., e_k, v_k\}$, where v_i are elements in V not necessarily different and e_i are elements in E not necessarily distinct. In a simple graph, we could shorten the way of writing the sequence by using only the vertices, *i.e.* by writing $\{v_0, v_1, v_2, ..., v_k\}$. A *trial* is a walk in which an edge can not be used more than once, that is, all $e_1, e_2, ..., e_k$ are distinct. A *path* is a walk in which we can not use the same vertex more than once, that is, all $v_0, v_1, ..., v_k$ are distinct, as a consequence we can not repeat edges.

Two vertices u and v are *connected* if there is a path or (u, v)-path between them in G. *Connection* is an equivalence relation between vertices, therefore we get a partition of V into $V_1, V_2, ...V_\omega$ classes, in such a way that two vertices u and v are connected if and only if they are in the same equivalence class V_i. The subgraphs induced by each of these classes, $G[V_1], G[V_2], ..., G[V_\omega]$, are called the *connected components* of G. If G has exactly one component, then G is *connected*, otherwise G is *disconnected*.

A walk in which $v_0 = v_k$ is a *closed walk* of length k. A *closed trial* or *circuit* is a closed walk in which we can not use one edge more than once. Finally, a closed path in which we can not use a vertex more than once is a *cycle*, as it has length k, it is also called a k-cycle, one example is given in figure A.2(a). A graph without cycles

is an *acyclic* graph. A *tree* is a connected acyclic graph, one example is provided in figure A.2(b). A *bipartite graph* is a graph whose vertex set can be partitioned into two subsets X and Y so that each edge has one end in X and one end in Y, a graph is bipartite if and only if it contains no odd cycle. One example of a bipartite graph is given in figure A.2(c).



(a) 5-cycle

(b) Tree

(c) Bipartite graph

Figure A.2: Examples of a k-cycle, a tree, and a bipartite graph in which all edges have one end in $X = \{1, 2, 3\}$ *and one end in* $Y = \{4, 5, 6\}$.

To illustrate some of the concepts defined above we use the graph G given in figure A.3 that is shown below.

An example of a walk in G is given by the sequence

$$\{1, a, 2, b, 1, a, 2, f, 5, g, 3\}$$

and of a closed walk is

$$\{1, a, 2, b, 1, a, 2, f, 5, g, 3, c, 2, b, 1\}\,.$$

An example of a trial is

$$\{1, a, 2, c, 3, d, 4, h, 3\}$$

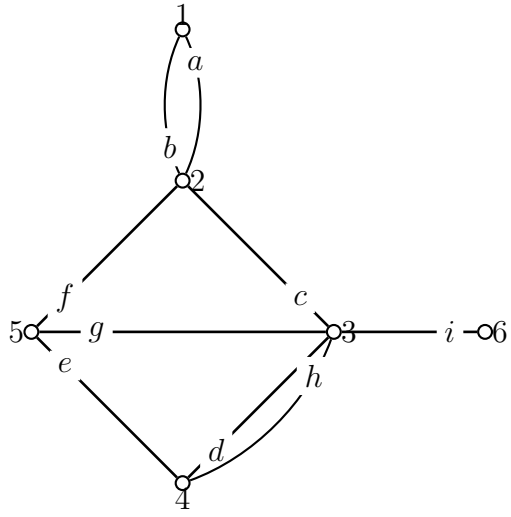


Figure A.3: Graph G used to illustrate the walk concept and its variants.

and of a closed trial is

$$\{1, a, 2, c, 3, d, 4, h, 3, g, 5, f, 2, b, 1\}.$$

Finally, an example of a path is

$$\{2, f, 5, e, 4, h, 3\}$$

and of a cycle is

$$\{2, f, 5, e, 4, h, 3, c, 2\}.$$

An $(\alpha, \beta) - separator$, with $\alpha, \beta \in V$, in a graph is a subset $C \subseteq V$ such that all paths from α to β go through C. The subset C separates A from B or C is a *separator set* of A from B if it is an (α, β)-separator for every $\alpha \in A$ and $\beta \in B$. Another way of seeing a separator set C is as a set of vertices C such that once deleted, *i.e.* once obtaining $G - C$, we have a graph in which there is not any path from the elements in A to the elements in B. In the graph given in figure A.3, $\{3, 5\}$ separates $\{1, 2\}$ from $\{4\}$ and it also separates $\{4\}$ from $\{6\}$.

Let σ be a cycle of length k, $\{v_0, v_1, ..., v_k\}$ with $v_0 = v_k$. A *chord* for this cycle is an edge $\{v_i, v_j\}$ of two non-consecutive vertices. A graph is *triangulated* or *chordal* if every cycle of length ≥ 4 posseses a chord (Diestel, 2005, p. 127).

A.1 Graph colourings

A *T-vertex colouring* of G is an assignment of T colours, $1, 2, ..., T$, to the vertices of G. That is, the vertex set V has a partition $(V_1, V_2, ...V_T)$ in such a way that all elements in the same colour class $V_i \subseteq V$ have the same colour i. Considering that G does not have loops, we have a *proper* colouring if two distinct adjacent vertices do not have the same colour i. In this case, each V_i forms an *independent set* which is a subset S of V in G in which no two vertices can be adjacent. G is *T-vertex colourable* if G has a proper T-vertex colouring.

It is important to remark that the vertex colouring definition is not restricted by the fact that the graph G is simple, in fact, in proper colourings we have that a graph G without loops is T-vertex colourable if and only if its underlying simple graph is T-vertex colourable, understanding as *underlying simple graph* to a simple spanning subgraph of a graph G in which we delete all loops and join with only one edge all vertices that were already joined in G, *i.e.* we considere only one edge between vertices even though we had multiple edges between them. We observe then that to define vertex colourings, proper or not, it does nor matter if there is only one edge or multiple edges between vertices. In the case of proper colouring, the fact of having one edge between two adjacent vertices means that we can not use the same colour for those vertices and then it becomes irrelevant if there were only one edge or more than one edge between them because in any case the vertices should have different colour.

For edge colourings, we have similar definitions. An *S-edge colouring* of a loop-less graph G is an assignment of S colours, $1, 2, ...S$, to the edges of G. That is, an S-edge colouring is a partition $(E_1, E_2, ...E_S)$ of the edge set E, where $E_i \subseteq E$ is the colour class i. The S-edge colouring is *proper* if no two adjacent edges have the same colour. Each E_i in the partition induced by the colouring forms a *matching*. A matching is a subset M of edges in E such that its elements are not adjacent in G. It is a concept analogous to independent sets, but applied to the edges of G. G is *S-edge colourable* if it has a proper S-edge colouring. We have that for proper edge colourings it is relevant the fact of having or no multiple edges; for example to have a proper colouring if two vertices are joined using different edges, we need a different colour for each edge.

In the graph given in figure A.4 we show a vertex and an edge colouring. Observe that both are proper colourings because neither adjacent vertices have the same colour for the vertex colouring, nor adjacent edges have the same colour for the edge colouring.

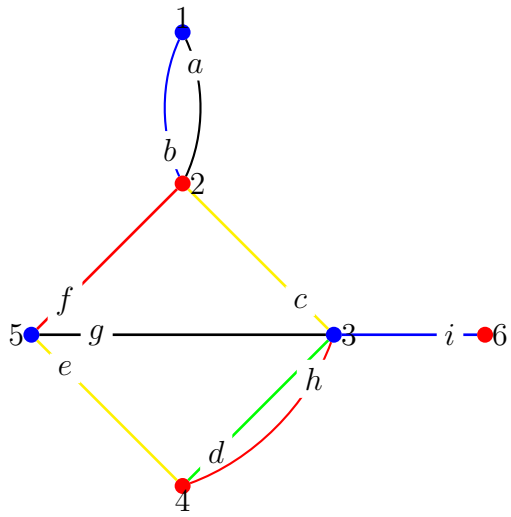


Figure A.4: Graph G to illustrate vertex and edge colourings. In this case E is partitioned into $E_1 = \{a,g\}$, $E_2 = \{b,i\}$, $E_3 = \{f,h\}$, $E_4 = \{c,e\}$, *and* $E_5 = \{d\}$ *and V is partitioned into* $V_1 = \{1,3,5\}$ *and* $V_2 = \{2,4,6\}$.

Appendix B

Convergence of a modified IPF method to the maximum likelihood estimators

In this appendix we present three lemmas that are used to prove theorem B.1 presented at the end of this appendix. This theorem states that the modified iterative proportional method presented in section **3.10** converges to the maximum likelihood estimators.

The notation used throughout this appendix is the same defined at the beginning of Chapters **3** and **4** as follows. A vertex colouring $(V_1, V_2, ..., V_T)$ is a partition of V into T classes, where v_j^r is the j-th vertex in the colour class V_r, r=1,...,T; j=1,...,$kver(r)$ with $kver$ a vector of dimension T of number of vertices in each vertex class. The marginal total $n(v_j^r = p)$ and marginal expected frequency $m(v_j^r = p)$ for the j-th variable in the colour class r for the category p are defined as

$$n(v_j^r = p) = \sum_{i:i_{v_j^r}=p} n(i),\ p = 1, 2, ..., J;\ j = 1, ..., kver(r);\ r = 1, 2, ..., T.$$

$$m(v_j^r = p) = \sum_{i:i_{v_j^r}=p} m(i),\ p = 1, 2, ..., J;\ j = 1, ..., kver(r);\ r = 1, 2, ..., T.$$

An edge colouring $(E_1, E_2, ..., E_S)$ is a partition of the edge set or first-order interaction set E into S classes, where $u_{l_v^z r_v^z}(i_v^z j_v^z)$, $v = 1, 2, ..., ked(z)$; z=1,2,...,S, is the first-order interaction parameter identified with the edge e_v^z joining variable l_v^z to variable r_v^z at the values (i_v^z, j_v^z) and e_v^z is the v-th element in the colour class E_z with ked the vector of dimension S of number of edges in each edge class. The marginal total

$n(l_v^z = i_v^z, r_v^z = j_v^z)$ and the marginal expected frequency $m(l_v^z = i_v^z, r_v^z = j_v^z)$ for the v-th pair of variables, with their corresponding levels, in the colour class z are defined as

$$n(l_v^z = i_v^z, r_v^z = j_v^z) = \sum_{s:(s_{l_v^z}, s_{r_v^z})=(i_v^z, j_v^z)} n(s),\ v = 1, ..., ked(z);\ z = 1, ..., S.$$

$$m(l_v^z = i_v^z, r_v^z = j_v^z) = \sum_{s:(s_{l_v^z}, s_{r_v^z})=(i_v^z, j_v^z)} m(s),\ v = 1, ..., ked(z);\ z = 1, ..., S.$$

For simplicity, $\{l_v = i_v, r_v = j_v\}$ with $\{l_v = i_v, r_v = j_v\} \in E_z$ denotes the edge joining variable l_v^z to r_v^z at the values (i_v^z, j_v^z), $v = 1, ..., ked(z)$, $z = 1, ...S$, and in this case $n(l_v = i_v, r_v = j_v)$ and $m(l_v = i_v, r_v = j_v)$ can be used instead of $n(l_v^z = i_v^z, r_v^z = j_v^z)$ and $m(l_v^z = i_v^z, r_v^z = j_v^z)$, respectively.

Additionally, we defined in section **3.10** for every vertex colour class V_r, $r = 1, 2, ..., T$, the following transformations

$$(T_{V_r(1)}m)(i) = m(i)\frac{\sum_{v_j^r \in V_r} n(v_j^r = 1)}{\sum_{v_j^r \in V_r} m(v_j^r = 1)},\ i \in \{(s_1, s_2, ..., s_\Delta) \in \mathrm{I} | s_{v_j^r} = 1, \text{ for some } v_j^r \in V_r\};$$

$$(T_{V_r(2)}m)(i) = m(i)\frac{\sum_{v_j^r \in V_r} n(v_j^r = 2)}{\sum_{v_j^r \in V_r} m(v_j^r = 2)},\ i \in \{(s_1, s_2, ..., s_\Delta) \in \mathrm{I} | s_{v_j^r} = 2, \text{ for some } v_j^r \in V_r\};$$

$$\vdots$$

$$(T_{V_r(J)}m)(i) = m(i)\frac{\sum_{v_j^r \in V_r} n(v_j^r = J)}{\sum_{v_j^r \in V_r} m(v_j^r = J)},\ i \in \{(s_1, s_2, ..., s_\Delta) \in \mathrm{I} | s_{v_j^r} = J, \text{ for some } v_j^r \in V_r\};$$

and for the edge colour class, E_z, with $z = 1, 2, ..., S$, we defined for $i \in \{(s_1, s_2, ..., s_\Delta) \in \mathrm{I} | s_{l_v} = i_v,\ s_{r_v} = j_v, \text{ for some } \{l_v = i_v, r_v = j_v\} \in E_z\}$ the following transformations

$$(T_{E_z}m)(i) = m(i)\frac{\sum_{\{l_v = i_v, r_v = j_v\} \in E_z} n(l_v = i_v, r_v = j_v)}{\sum_{\{l_v = i_v, r_v = j_v\} \in E_z} m(l_v = i_v, r_v = j_v)}.$$

Lemma B.1. *For a vertex colouring $(V_1, V_2, ..., V_T)$ the following equality holds*

$$\sum_{v_j^r \in V_r} n(v_j^r = p) = \sum_{v_j^r \in V_r} (T_{V_r(p)}m)(v_j^r = p), r = 1, ..., T;\ p = 1, 2, ..., J;$$

where

$$(T_{V_r(p)}m)(v_j^r = p) = \sum_{i: i_{v_j^r} = p} (T_{V_r(p)}m)(i),\ p = 1, 2, ..., J;\ j = 1, ..., kver(r);\ r = 1, 2, ..., T.$$

For an edge colouring $(E_1, E_2, ..., E_S)$, we have that for all $z = 1, ..., S$ the following equality holds

$$\sum_{\{l_v=i_v, r_v=j_v\}\in E_z} n(l_v = i_v, r_v = j_v) = \sum_{\{l_v=i_v, r_v=j_v\}\in E_z} (T_{E_z} m)(l_v = i_v, r_v = j_v);$$

where

$$(T_{E_z} m)(l_v = i_v, r_v = j_v) = \sum_{s:(s_{l_v}, s_{r_v})=(i_v, j_v)} (T_{E_z} m)(s),\ \{l_v = i_v, r_v = j_v\} \in E_z.$$

Proof. Using the definition of $T_{V_r(p)}$, we have

$$\sum_{i:i_{v_j^r}=p} (T_{V_r(p)} m)(i) = m(v_j^r = p)\frac{\sum_{v_j^r\in V_r} n(v_j^r = p)}{\sum_{v_j^r\in V_r} m(v_j^r = p)},$$

summing over all $v_j^r \in V_r$ we obtain

$$\sum_{v_j^r\in V_r} (T_{V_r(p)} m)(v_j^r = p) = \sum_{v_j^r\in V_r} m(v_j^r = p)\frac{\sum_{v_j^r\in V_r} n(v_j^r = p)}{\sum_{v_j^r\in V_r} m(v_j^r = p)},$$

the $\sum_{v_j^r\in V_r} m(v_j^r = p)$ terms are canceled out getting the result.

Using the definition of T_{E_z}, we have

$$\sum_{s:(s_{l_v}, s_{r_v})=(i_v, j_v)} (T_{E_z} m)(s) = m(l_v = i_v, r_v = j_v)\frac{\sum_{\{l_v=i_v, r_v=j_v\}\in E_z} n(l_v = i_v, r_v = j_v)}{\sum_{\{l_v=i_v, r_v=j_v\}\in E_z} m(l_v = i_v, r_v = j_v)},$$

summing over all $\{l_v = i_v, r_v = j_v\} \in E_z$ we obtain

$$\sum_{\{l_v=i_v, r_v=j_v\}\in E_z} (T_{E_z} m)(l_v = i_v, r_v = j_v) =$$

$$\sum_{\{l_v=i_v, r_v=j_v\}\in E_z} m(l_v = i_v, r_v = j_v)\frac{\sum_{\{l_v=i_v, r_v=j_v\}\in E_z} n(l_v = i_v, r_v = j_v)}{\sum_{\{l_v=i_v, r_v=j_v\}\in E_z} m(l_v = i_v, r_v = j_v)},$$

the $\sum_{\{l_v=i_v, r_v=j_v\}\in E_z} m(l_v = i_v, r_v = j_v)$ terms are canceled out getting the result.

□

Similar to what Lauritzen does (Lauritzen, 1996, see p. 71-72, 81) when he defines *log-affine models* and using the same notation used by him, we define the set of expected frequencies satisfying a RGLL model as $M(q, H)$. This is the set of expected frequencies satisfying a sampling scheme (section **2.2**) and that $m = q\exp(v)$, $v \in H$ with

$$H = \sum_{a\in A,|a|\neq 1,2} F_a + \sum_{r=1}^{T}\sum_{p=1}^{J} F_{V_r(p)} + \sum_{z=1}^{S} F_{E_z},$$

subspace of $\mathbb{R}^{\mathrm{I}}$, where:

For $|a| \neq 1, 2$, $x \in F_a \Leftrightarrow x(i) = x(j)$, for all i, j with $i_a = j_a$.

For $V_r = \{v_1^r, v_2^r, ..., v_{kver(r)}^r\}$, $r = 1, .., T$; $p = 1, ..., J$; $x \in F_{V_r(p)} \Leftrightarrow x(i) = x(j)$, for all i, j with $i_{v_k^r} = p = j_{v_k^r}$, $k = 1, ..., kver(r)$, $v_k^r \in V_r$. That is, $F_{V_r(p)}$ is the function space that depends on the cell i through $i_{v_k^r} = p$, $k = 1, ..., kver(r)$.

For $E_z = \{e_1^z, e_2^z, ..., e_{ked(z)}^z\}$, $z = 1, .., S$ with $e_v^z = \{l_v^z = i_v^z, r_v^z = j_v^z\}$, $v = 1, ..., ked(z)$; $x \in F_{E_z} \Leftrightarrow x(s) = x(o)$, for all s, o with $s_{l_v^z} = i_v^z = o_{l_v^z}$, $s_{r_v^z} = j_v^z = o_{r_v^z}$, $v = 1, ..., ked(z)$.

The vector q is fixed and not identically zero. If $q = 1$, we have that the log-affine model is a log-linear model, the cells where $q(i) = 0$ are structural zeros. As in Lauritzen (1996, see p. 72), we can define the corresponding extended models $\bar{M}$ that consist of all pointwise limits of vectors in the model. We emphasize that Lauritzen defined log-affine models in this form because it makes possible the existence of maximum likelihood estimators in an extended way.

In section **3.8** we derived the maximum likelihood estimators according to different sampling schemes and that the expected frequencies had the kinds of restrictions considered for elements in $M(q, H)$. Then, the estimators $\widehat{m}$ are elements in $M(q, H)$. As all the three distributions belong to the exponential family, the maximum likelihood estimators are unique. In consequence:

Lemma B.2. *The maximum likelihood estimate $\widehat{m}$ is the unique element in $\bar{M}(q, H)$ satisfying the likelihood equations (3.5), (3.7), and (3.12), corresponding to vertex, edge, and vertex and edge colourings, respectively.*

For the more general case, with vertex and edge colouring simultaneously, these equations are

$$\sum_{j=1}^{kver(r)} n(v_j^r = 1) = \sum_{j=1}^{kver(r)} m(v_j^r = 1),\ r = 1, ..., T;$$
$$\sum_{j=1}^{kver(r)} n(v_j^r = 2) = \sum_{j=1}^{kver(r)} m(v_j^r = 2),\ r = 1, ..., T;$$
$$\vdots$$
$$\sum_{j=1}^{kver(r)} n(v_j^r = J) = \sum_{j=1}^{kver(r)} m(v_j^r = J),\ r = 1, ..., T;$$
$$\sum_{v=1}^{ked(z)} n(l_v^z = i_v^z, m_v^z = j_v^z) = \sum_{v=1}^{ked(z)} m(l_v^z = i_v^z, m_v^z = j_v^z),\ z = 1, 2, ..., S;$$
$$n_a(i_a) = m_a(i_a),\ \text{for all } a \in A,\ |a| \neq 1, 2.$$

Some important properties satisfied by the transformations are presented in the following lemma B.3, which is analogous to lemma 4.12 presented by Lauritzen (1996, p. 83) but modified to include the transformations corresponding to the vertex classes $T_{V_r(p)}$, $r = 1, ..., T$, $p = 1, ..., J$ and to the edge classes T_{E_l}, $l = 1, ..., S$, besides the transformations corresponding to the elements in the generating class T_b, $|b| \neq 1, 2$.

Lemma B.3. *The transformation functions $T_{V_r(p)}$, $r = 1, ..., T$, $p = 1, ..., J$; T_{E_l}, $l = 1, ..., S$; and T_b, $|b| \neq 1, 2$ satisfy the following*

(i) *$T_{V_r(p)}$, $r = 1, ..., T$, $p = 1, ..., J$; T_{E_l}, $l = 1, ..., S$; and T_b, $|b| \neq 1, 2$ are continuous on $M^* = \{m | if\ n(i) > 0\ then\ m(i) > 0\}$;*

(ii) *$T_{V_r(p)}m$, $r = 1, ..., T$, $p = 1, ..., J$; $T_{E_l}m$, $l = 1, ..., S$; and T_b, $|b| \neq 1, 2$ are the uniquely determined maximum likelihood estimators in the models $\bar{M}(m, F_{V_r(p)})$, $\bar{M}(m, F_{E_l})$, and $\bar{M}(m, F_b)$, respectively;*

(iii) *$L(T_{V_r(p)}m) \geq L(m)$, $r = 1, ..., T$, $p = 1, ..., J$; $L(T_{E_l}m) \geq L(m)$, $l = 1, ..., S$; $L(T_bm) \geq L(m)$, $|b| \neq 1, 2$, with equality if and only if $\sum_{v_k^r \in V_r} n(v_k^r = p) = \sum_{v_k^r \in V_r} m(v_k^r = p)$, $\sum_{\{l_v = i_v, r_v = j_v\} \in E_l} n(l_v = i_v, r_v = j_v) = \sum_{\{l_v = i_v, r_v = j_v\} \in E_l} m(l_v = i_v, r_v = j_v)$, and $n(i_b) = m(i_b)$, $i_b \in \mathrm{I}_b$, respectively, which occurs if and only if $T_{V_r(p)}m = m$, $T_{E_l}m = m$, and $T_bm = m$, respectively;*

(iv) *$T_{V_r(p)}(M^* \cap \bar{M}) \subseteq M^* \cap \bar{M}$, $r = 1, ..., T$, $p = 1, ..., J$; $T_{E_l}(M^* \cap \bar{M}) \subseteq M^* \cap \bar{M}$, $l = 1, ..., T$; and $T_b(M^* \cap \bar{M}) \subseteq M^* \cap \bar{M}$, $|b| \neq 1, 2$.*

Proof. For (i) first we prove that $T_{V_r(p)}$, $r = 1, ..., T$, $p = 1, ..., J$ is continuous on M^*. We take $(m_q) \in M^*$ with $m_q \to m \in M^*$. If $\sum_{v_k^r \in V_r} m(v_k^r = p) \neq 0$ then $T_{V_r(p)} m_q \to T_{V_r(p)} m$. If $\sum_{v_k^r \in V_r} m(v_k^r = p) = 0$, then $m(j) = 0$ for all j with $j_{v_k^r} = p$, $k = 1, ..., kver(r)$, but as $m \in M^*$ then $n(j) = 0$ for all j with $j_{v_k^r} = p$, $k = 1, ..., kver(r)$, which means that $\sum_{v_k^r \in V_r} n(v_k^r = p) = 0$ then $(T_{V_r(p)} m_q)(i) = 0 = (T_{V_r(p)} m)(i)$ for all q and $T_{V_r(p)} m_q \to T_{V_r(p)} m$.

To prove that T_{E_l}, $l = 1, ..., S$ is continuous on M^*, we take $(m_q) \in M^*$ with $m_q \to m \in M^*$. If $\sum_{\{l_v = i_v, r_v = j_v\} \in E_l} m(l_v = i_v, r_v = j_v) \neq 0$ then $T_{E_l} m_q \to T_{E_l} m$. If $\sum_{\{l_v = i_v, r_v = j_v\} \in E_l} m(l_v = i_v, r_v = j_v) = 0$ then $m(s) = 0$ for all s with $s_{l_v} = i_v, s_{r_v} = j_v$, $\{l_v = i_v, r_v = j_v\} \in E_l$, but then $n(s) = 0$ for all s with $s_{l_v} = i_v, s_{r_v} = j_v$, which means that $\sum_{\{l_v = i_v, r_v = j_v\} \in E_l} n(l_v = i_v, r_v = j_v) = 0$ and then $(T_{E_l} m_q)(i) = 0 = (T_{E_l} m)(i)$ for all q, so that $T_{E_l} m_q \to T_{E_l} m$.

For T_b, $|b| \neq 1, 2$ the proof is the same given in Lauritzen (1996, p. 83) as follows. Let $(m_q) \in M^*$ with $m_q \to m \in M^*$. Then, if $m(i_b) \neq 0$ we trivially have that $T_b m_q \to T_b m$. If $m(i_b) = 0$, $m(j) = 0$ for all j with $j_b = i_b$. But then, since $m \in M^*$, we must have $n(j) = 0$ for all such j, whereby $n(i_b) = 0$. Thus $(T_b m_q)(i) = 0 = (T_b m)(i)$ for all q and $T_b m_q \to T_b m$.

To prove (ii), we use that $T_{V_r(p)} m \in \bar{M}(m, F_{V_r(p)})$ $r = 1, ..., T$, $p = 1, ..., J$. This is because $T_{V_r(p)} m$ is a function of m and elements that only depend on the fact that $i_{v_k^r} = p$, $k = 1, ..., kver(r)$, so that $T_{V_r(p)} m$ can be expressed as $m \exp(v)$ with $v \in F_{V_r(p)}$.

$$
\begin{aligned}
m \exp(v) &= \begin{cases} m(i) \exp(\log(\frac{\sum_{v_k^r \in V_r} n(v_k^r = p)}{\sum_{v_k^r \in V_r} m(v_k^r = p)})), & i \in \{(i_1, i_2, ..., i_\Delta) \in \mathrm{I} | i_{v_k^r} = p, \text{ for some } v_k^r \in V_r\} \\ m(i) \exp(0), & \text{otherwise.} \end{cases} \\
&= T_{V_r(p)} m.
\end{aligned}
$$

Using this last result, lemma B.1 and lemma B.2, we have that $T_{V_r(p)} m$, $r = 1, ..., T$, $p = 1, ..., J$ is the unique maximum likelihood estimator in $\bar{M}(m, F_{V_r(p)})$. Similarly, we have that $T_{E_l} m \in \bar{M}(m, F_{E_l})$, $l = 1, ..., S$, because

$$
\begin{aligned}
m \exp(v) &= \begin{cases} m(i) \exp(\log(\sum_{\{l_v = i_v, r_v = j_v\} \in E_l} n(l_v = i_v, r_v = j_v) / \sum_{\{l_v = i_v, r_v = j_v\} \in E_l} m(l_v = i_v, r_v = j_v))), \\ \qquad i \in \{(s_1, s_2, ..., s_\Delta) \in \mathrm{I} | s_{l_v} = i_v, \ s_{r_v} = j_v, \text{ for some } \{l_v = i_v, r_v = j_v\} \in E_l\} \\ m(i) \exp(0), \qquad \text{otherwise.} \end{cases} \\
&= T_{E_l} m.
\end{aligned}
$$

Using this result, lemma B.1, and lemma B.2, we have that $T_{E_l} m$ is the uniquely determined maximum likelihood estimator in $\bar{M}(m, F_{E_l})$. For T_b, $|b| \neq 1, 2$ the proof is

the same given by Lauritzen (1996, p. 83).

(*iii*) consists of three results that have to be individually proved. In the first result given in (*iii*), all the inequalities are direct consequence of (*ii*), because in each case we have unique maximum likelihood estimators in different model spaces.

Fo the second result in (*iii*), suppose $L(T_{V_r(p)}m) = L(m)$, $r = 1, ..., T$, $p = 1, ..., J$, then $T_{V_r(p)}m = m$ because $T_{V_r(p)}m$ is the unique maximum likelihood estimator in $\bar{M}(m, F_{V_r(p)})$ ((*ii*)). Then, the likelihood equations must be satisfied by m, that is $\sum_{v_k^r \in V_r} n(v_k^r = p) = \sum_{v_k^r \in V_r} m(v_k^r = p)$. If $\sum_{v_k^r \in V_r} n(v_k^r = p) = \sum_{v_k^r \in V_r} m(v_k^r = p)$, then m satisfies the likelihood equations, meaning that L is maximized in m, but as $T_{V_r(p)}m$ is the unique maximum likelihood estimator in $\bar{M}(m, F_{V_r(p)})$, then $L(T_{V_r(p)}m) = L(m)$. Similarly, $L(T_{E_l}m) = L(m)$, $l = 1, ..., S$ if and only if $\sum_{\{l_v=i_v, r_v=j_v\} \in E_l} n(l_v = i_v, r_v = j_v) = \sum_{\{l_v=i_v, r_v=j_v\} \in E_l} m(l_v = i_v, r_v = j_v)$ and $L(T_b m) = L(m)$ if and only if $n(i_b) = m(i_b)$, $i_b \in \mathrm{I}_b$.

Finally, we prove the third and last result in (*iii*). If $\sum_{v_k^r \in V_r} n(v_k^r = p) = \sum_{v_k^r \in V_r} m(v_k^r = p)$, then m satisfies the likelihood equations so that m is a maximum likelihood estimator, but since $T_{V_r(p)}m$ is a unique maximum likelihood estimator then $T_{V_r(p)}m = m$. If $T_{V_r(p)}m = m$ then $\sum_{v_k^r \in V_r} n(v_k^r = p) = \sum_{v_k^r \in V_r} m(v_k^r = p)$ because $T_{V_r(p)}m$ satisfies $\sum_{v_j^r \in V_r} n(v_j^r = l) = \sum_{v_j^r \in V_r} (T_{V_r(p)}m)(v_j^r = l)$ (lemma B.1). The results for the edges and for the elements b in the generating class, $|b| \neq 1,\ 2$, can be analogously proved.

To prove (*iv*), let $m \in M^* \cap M$. As $m \in M^*$ we have that if $n(i) > 0$ then $m(i) > 0$. Suppose that $n(i) > 0$ and that for the vertex classes we take $i \in \{(i_1, i_2, ..., i_\Delta) \in \mathrm{I} | i_{v_j^r} = p, \text{ for some } v_j^r \in V_r\}$. By definition $T_{V_r(p)}m(i) = m(i)\frac{\sum_{v_j^r \in V_r} n(v_j^r = p)}{\sum_{v_j^r \in V_r} m(v_j^r = p)}$, but if we take $\sum_{v_j^r \in V_r} m(v_j^r = p) = 0$, we have $m(i) = 0$, which contradicts that $m(i) > 0$, then exists $\epsilon > 0$ such that

$$T_{V_r(p)}m(i) = \lim_{\epsilon \to 0} m(i)\frac{\sum_{v_j^r \in V_r} n(v_j^r = p) + \epsilon}{\sum_{v_j^r \in V_r} m(v_j^r = p) + \epsilon}.$$

$m(i) > 0$ and $\sum_{v_j^r \in V_r} n(v_j^r = p) > 0$, the latter because at least $n(i) > 0$. This means that if $n(i) > 0$ then $(T_{V_r(p)}m)(i) > 0$, so that $(T_{V_r(p)}m)(i) \in M^*$. But $(T_{V_r(p)}m)(i) \in \bar{M}$, so that $(T_{V_r(p)}m)(i) \in M^* \cap \bar{M}$. This result is trivial for any other $i,\ i \notin \{(i_1, i_2, ..., i_\Delta) \in \mathrm{I} | i_{v_j^r} = p, \text{ for some } v_j^r \in V_r\}$. Then

$$T_{V_r(p)}(M^* \cap M) \subseteq M^* \cap \bar{M}.$$

The same can be done for the other two transformations. Denoting by T^* to any of

the three type of transformations $T_{V_r(p)}$, $r = 1, ..., T$, $p = 1, ..., J$; T_{E_l}, $l = 1, ..., S$; and T_b, $|b| \neq 1, 2$, we have

$$T^*(M^* \cap M) \subseteq M^* \cap \bar{M}$$

As in Lauritzen (1996, p. 84), using that $M^* \cap \bar{M} = \{m \in \bar{M} | L(m) > 0\}$, that all T^*´s are continuous on M^* (see (i)), and (iii) we have that

$$T^*(M^* \cap \bar{M}) \subseteq M^* \cap \bar{M}.$$

□

It is not always true that $|n| = \left|T_{V_r(p)} m\right|$, $r = 1, ..., T$, $p = 1, ..., J$ or that $|n| = |T_{E_l} m|$, $l = 1, ..., S$. However, we can see from the vertex colouring equations that $|n| = |\widehat{m}|$. This means that the estimators satisfy the appropriated restrictions due to the multinomial sampling scheme (section **2.2**). In graphical log-linear models it is always true that the corresponding transformations T_b for any b in the generating class A satisfy $|n| = |T_b m|$ independently of the sampling scheme, which means in the multinomial case that the transformed values satisfy the multinomial sampling scheme restrictions.

As the proof for the convergence of the algorithm in the multinomial case requires that the transformed values satisfy the multinomial sampling scheme restrictions, we suppose that the sampling scheme is Poisson because in this case the corresponding restriction, that the transformed expected frequencies are equal or more than zero, is satisfied; however, for the multinomial case it seems that the convergence is also true but we have still been unable to prove it.

The following theorem B.1 is analogous to theorem 4.13 given by Lauritzen (1996, p. 84); however, we additionally suppose a Poisson sampling scheme and that the starting values $m_0 \in M^* \cap \bar{M}$, and consider that the transformed expected frequencies on step q are m_q.

Theorem B.1. *Under Poisson sampling scheme, for any starting value $m_0 \in M^* \cap \bar{M}$ (for example $m_0 = 1$) it holds that*

$$\widehat{m} = \lim_{q \to \infty} m_q,$$

the maximum likelihood estimator.

Proof. As in theorem 4.13 given by Lauritzen (1996, p. 84), the proof is the direct application of iterative proportional maximization, so that we need to satisfy the required conditions.

We assume that $m_0 \in M^* \cap \bar{M}$ similarly to what Lauritzen does (Lauritzen, 1989, p. 24). Additionally, we define

$$\Theta = \{m \in \bar{M}(q, H) \mid L(m) \geq L(m_0)\}$$

Take $m \in \Theta$. For simplicity, we denote all transformations, $T_{V_r(p)}m$, $r = 1, ..., T$, $p = 1, ..., J$, $T_{E_l}m$, $l = 1, ..., S$, and $T_b m$, $|b| \neq 1, 2$, as T^*m. All T^*m are elements in $\bar{M}(q, H)$ that maximize the likelihood ((ii) lemma B.3), then $L(T^*m) \geq L(m) \geq L(m_0)$, so that $T^*m \in \Theta$.

Take $m_0 \in M^* \cap \bar{M}$, we know that $M^* \cap \bar{M} = \{m \in \bar{M} | L(m) > 0\}$ (Lauritzen, 1996, p. 83), so that $L(m_0) > 0$. If we take $m \in \Theta$, then $m \in M^* \cap \bar{M}$ because we have an $m \in \bar{M}(q, H)$ such that $L(m) \geq L(m_0) > 0$. This means that $\Theta \subseteq M^* \cap \bar{M}$. Additionally, $T^*(M^* \cap \bar{M}) \subseteq M^* \cap \bar{M}$ ((iv) in lemma B.3), then as all elements in Θ are part of $M^* \cap \bar{M}$ it should be satisfied that $T^*(\Theta) \subseteq \Theta$. Then we have that all the transformations T^* are continuous transformations from Θ to Θ.

According to (ii) in lemma B.3 the transformations for the vertex colouring, edge colouring, and interactions of order different to one, maximize L over $\bar{M}(m, F_{V_r(p)})$, $r = 1, ..., T$, $p = 1, ..., J$, $\bar{M}(m, F_{E_l})$, $l = 1, ..., S$, and $\bar{M}(m, F_b)$, $|b| \neq 1, 2$, respectively; in consequence they maximize L over the sections $\bar{M}(m, F_{V_r(p)}) \cap \Theta$, $\bar{M}(m, F_{E_l}) \cap \Theta$, and $\bar{M}(m, F_b) \cap \Theta$. Also we know by (iii) in lemma B.3 that $L(T^*m) > L(m)$ if $m \neq T^*m$, *i.e.* that T^*m are the uniquely determined points where L is maximized over each section.

We also know by lemma B.2 that the global maximum of L is uniquely determined in $\bar{M}(q, H)$, which means that by maximizing all sections we are maximizing L in general. Then, we can use iterative proportional maximization getting that

$$\widehat{m} = \lim_{q \to \infty} m_q,$$

where $\widehat{m}$ is the maximum likelihood estimator, the unique point where L attains its maximum. □

Bibliography

Agresti, A. (2002a). *Categorical data analysis*. New York: John Wiley and Sons, 2nd edition.

Agresti, A. (2002b). Links between binary and multicategory logit item response models and quasi-symmetric log-linear models. In *Annales de la Faculté des Sciences de Toulouse*, volume XI (4), pages 434–454.

Agresti, A. and Gottard, A. (2007). Independence in multiway contingency tables: S. N. Rays's breakthroughs and later developments. *Journal of Statistical Planning and Inference* **137**, 3216–3226. Available from `http://www.ds.unifi.it/ricerca/pubblicazioni/working_papers/2005/wp2005_08.pdf`.

Altham, P. M. E. (1984). Improving the precision of estimation by fitting a model. *Journal of the Royal Statistical Society Series B* **46**, 118–119.

Altham, P. M. E. (1994). Improving the precision of estimation by fitting a generalized linear model, and quasilikelihood. *GLIM newsletter* **23**. Available from `http://www.statslab.cam.ac.uk/~pat`.

Andersen, A. H. (1974). Multidimensional contingency tables. *Scandinavian Journal of Statistics* **1 (3)**, 115–127.

Andersen, E. (1991). *The statistical analysis of categorical data*. Heidelberg: Springer-Verlag, 2nd edition.

Andersson, S. A. (1975). Invariant normal models. *The Annals of Statistics* **3 (1)**, 132–154.

Andersson, S. A., Madigan, D., Perlman, M. D., and Richardson, T. S. (1999). Graphical Markov models in multivariate analysis. In *Multivariate analysis, design of experiments, and survey sampling*, pages 187–229.

Andersson, S. A. and Madsen, J. (1998). Symmetry and lattice conditional independence in a multivariate normal distribution. *The Annals of Statistics* **26 (2)**, 525–572.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **160 (901)**, 268–282.

Bergsma, W. P. and Rudas, T. (2002). Marginal models for categorical data. *Annals of Statistics* **30**, 140–159.

Birch, M. W. (1963). Maximum likelihood in three-way contingency tables. *Journal of the Royal Statistics Society Series B* **25 (1)**, 220–233.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete multivariate analysis*. Cambridge, Massachusetts: MIT Press. Reprinted by Springer, New York, 2007.

Bondy, J. A. and Murty, U. S. R. (1976). *Graph theory with applications*. London: MacMillan Press.

Booth, J. G., Capanu, M., and Heigenhauser, L. (2005). Exact conditional p-value calculation for the quasi-symmetry model. *Journal of computational and graphical statistics* **14 (3)**, 716–725.

Bowker, A. H. (1948). A test for symmetry in contingency tables. *Journal of the American Statistical Association* **43 (244)**, 572–574.

Caffo, B. (2008). Exact hypothesis tests for log-linear models with exactLoglinTest. *Journal of Statistical Software* **17 (7)**, 1–17.

Caussinus, H. (1965). Contribution à l′analyse statistique des tableux de corrélation. *Annales de la Faculté des Sciences de Toulouse* **29**, 77–182.

Christensen, R. (1997). *Log-linear models and logistic regression*. New York: Springer-Verlag, 2nd edition.

Cowell, R. G., Dawid, A. P., Lauritzen, S. L., and Spiegelhalter, D. J. (1999). *Probabilistic Networks and Expert Systems*. New York: Springer-Verlag.

Darroch, J. N. and Bhapkar, V. P. (1990). Marginal symmetry and quasi-symmetry models of general order. *Journal of Multivariate Analysis* **34 (2)**, 173–184.

Darroch, J. N., Lauritzen, S. L., and Speed, T. P. (1980). Markov fields and log-linear interaction models for contingency tables. *The Annals of Statistics* **8 (3)**, 522–539.

Darroch, J. N. and Ratcliff, D. (1972). Generalized iterative scaling for log-linear models. *The Annals of Mathematical Statistics* **43 (5)**, 1470–1480.

Darroch, J. N. and Speed, T. P. (1983). Additive and multiplicative models and interactions. *The Annals of Statistics* **11 (3)**, 724–738.

Deming, W. E. and Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics* **11 (4)**, 427–444.

Dempster, A. P. (1972). Covariance selection. *Biometrics* **28**, 157–175.

Diestel, R. (2005). *Graph theory*. New York: Springer-Verlag, 3rd edition.

Drton, M. and Richardson, T. S. (2008). Binary models for marginal independence. *Journal of the Royal Statistics Society Series B* **70 (2)**, 287–309.

Drton, M., Sturmfels, B., and Sullivant, S. (2009). *Lectures on Algebraic Statistics*. Berlin: Birkhäuser Verlag AG.

Eaton, M. L. (1983). *Multivariate statistics: a vector space approach*. New York: John Wiley and Sons.

Edwards, D. (2000). *Introduction to graphical modelling*. New York: Springer-Verlag, 2nd edition.

Edwards, D., de Abreu, G. C. G., and Labouriau, R. (2010). Selecting high-dimensional mixed graphical models using minimal AIC or BIC forests. *BMC Bioinformatics* **11 (18)**, 1–13. Available from `http://www.biomedcentral.com/1471-2105/11/18`.

Edwards, D. and Havránek, T. (1985). A fast procedure for model search in multidimensional contingency tables. *Biometrika* **72 (2)**, 339–351.

Edwards, D. and Havránek, T. (1987). A fast model selection procedure for large families of models. *Journal of the American Statistical Association* **83 (397)**, 205–213.

Eriksen, P. S. (2005). *Decomposable log-linear models*. Research report R-2005-16, Department of Mathematical Sciences, Aalborg University. Available from `http://vbn.aau.dk/research/decomposable_loglinear_models(112654)/`.

Eriksson, N., Fienberg, S. E., Rinaldo, A., and Sullivant, S. (2006). Polyhedral conditions for the nonexistence of the MLE for hierarchical log-linear models. *Journal of Symbolic Computation* **41**, 222–233.

Eslava, G. and Ramírez, R. (2008). *Restricted or coloured graphical log-linear models*. Research report 5-08, Department of Mathematics, Faculty of Sciences, National University of Mexico. Available from `http://www.matematicas.unam.mx/reportes/Eslava_Ramirez_reporte_12-10-2008.pdf`.

Fienberg, S. E. (1970). An iterative procedure for estimation in contingency tables. *Annals of Statistics* **41 (3)**, 907–917.

Fienberg, S. E. (1980). *The analysis of cross-classified categorical data*. Cambridge, Massachusetts: MIT Press, 2nd edition.

Fienberg, S. E. and Larntz, K. (1976). Log-linear representation for paired and multiple comparisons models. *Biometrika* **63 (2)**, 245–254.

Fienberg, S. E. and Rinaldo, A. (2007). Three centuries of categorical data analysis: log-linear models and maximum likelihood estimation. *Journal of Statistical Planning and Inference* **137**, 3430–3445.

Frets, G. P. (1921). Heredity of head form in man. *Genetica* **3**, 193–400.

Frydenberg, M. and Jensen, J. L. (1989). Is the "improved likelihood ratio statistic" really improved in the discrete case? *Biometrika* **76 (4)**, 655–661.

Gabriel, K. R. (1989). Simultaneous test procedures – some theory of multiple comparisons. *The Annals of Mathematical Statistics* **40 (1)**, 224–250.

Glonek, G. F. V., Darroch, J. N., and Speed, T. P. (1988). On the existence of maximum likelihood estimators for hierarchical log-linear models. *Scandinavian Journal of Statistics* **15 (3)**, 187–193.

Goodman, L. A. (1970). The multivariate analysis of qualitative data: interactions among multiple classifications. *Journal of the American Statistical Association* **65 (329)**, 226–256.

Goodman, L. A. (1971). The analysis of multidimensional contingency tables: stepwise procedures and direct estimation methods for building models for multiple classifications. *Technometrics* **13 (1)**, 33–61.

Goodman, L. A. (1985). The analysis of cross classified data having ordered and/or unordered categories: Association models, correlation models, and asymmetry models for contingency tables with or without missing entries. *The Annals of Statistics* **13 (1)**, 10–69.

Goodman, L. A. (2002). Contributions to the statistical analysis of contingency tables: Notes on quasi-symmetry, quasi- independence, log-linear models, log-bilinear models and correspondence analysis models. In *Annales de la Faculté des Sciences de Toulouse*, volume XI (4), pages 525–540.

Gottard, A. (2009). *Graphical log-linear models for homologous factors*. Paper presented at S.Co.2009, Politecnico di Milano, Milano, Italy. Available from `http://www2.mate.polimi.it/ocs/viewabstract.php?id=137&cf=7`.

Gottard, A., Marchetti, G. M., and Agresti, A. (2008). *Some considerations on graphical models with exchangeability structure*. Manuscript presented at the poster session of the XLIV Riunione Scientifica Università della Calabria, Calabria, Italy. Available from `http://www.sis-statistica.it/files/pdf/atti/rs08_poster_41.pdf`.

Gottard, A., Marchetti, G. M., and Agresti, A. (2010). Quasi-symmetric graphical log-linear models. Accepted by *The Scandinavian Journal of Statistics*, `doi:10.111/j.1467-9469.2010.00713.x`.

Graybill, F. A. (1961). *An Introduction to Statistical Linear Models Vol. 1*. New York: Mc Graw Hill Book Co.

Haberman, S. J. (1972). Algorithm AS 51: log-linear fit for contingency tables. *Applied Statistics* **21 (2)**, 218–225.

Haberman, S. J. (1973). Log-linear models for frequency data: sufficient statistics and likelihood equations. *The Annals of Statistics* **1 (4)**, 617–632.

Haberman, S. J. (1976). Correction to AS 51: log-linear fit for contingency tables. *Applied Statistics* **25 (2)**, 193.

Haberman, S. J. (1978). *Analysis of qualitative data, Vol. 1: Introductory topics*. New York: Academic Press.

Haberman, S. J. (1979). *Analysis of qualitative data, Vol. 2: New developments*. New York: Academic Press.

Hammersley, J. M. and Clifford, P. E. (1971). *Markov fields on finite graphs and lattices*. Unpublished manuscript. Available from `http://www.statslab.cam.ac.uk/~grg/books/hammfest/hamm-cliff.pdf`.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. New York: Springer, 2nd edition.

Heckerman, D., Meek, C., and Richardson, T. (2004). *Variations on undirected graphical models and their relationships*. Tech report MSR-TR-2004-94, Microsoft Research Group. Available from `http://research.microsoft.com/apps/pubs/default.aspx?id=70094`.

Højsgaard, S. and Lauritzen, S. L. (2005). Restricted concentration models – graphical Gaussian models with concentration parameters restricted to being equal. In *Proceedings of the tenth international workshop on Artificial Intelligence and Statistics*. Society for Artificial Intelligence and Statistics. Available from `http://www.gatsby.ucl.ac.uk/aistats/fullpapers/159.pdf`.

Højsgaard, S. and Lauritzen, S. L. (2007). Inference in graphical Gaussian models with edge and vertex symmetries with the gRc package for R. *Journal of Statistical Software* **23 (6)**. Available from `http://www.jstatsoft.org/v23/i06/paper`.

Højsgaard, S. and Lauritzen, S. L. (2008). Graphical Gaussian models with edge and vertex symmetries. *Journal of the Royal Statistics Society Series B* **70 (5)**, 1005–1027.

Kendler, K. S., Neale, M. C., Kessler, R. C., Kessler, A. C., and Eaves, L. J. (1992). A population-based twin study of major depression in women: the impact of varying definition of illness. *Archives of General Psychiatry* **49**, 257–266.

Kullback, S. (1971). Marginal homogeneity of multidimensional contingency tables. *The Annals of Mathematical Statistics* **42 (2)**, p. 594–606.

Lauritzen, S. L. (1989). *Lectures on contingency tables*. Aalborg, Denmark: Department of Mathematics, 3rd edition. Electronic version 2002 available from `http://www.stats.ox.ac.uk/~steffen/papers/cont.pdf`.

Lauritzen, S. L. (1996). *Graphical models*. Oxford, England: Clarendon Press.

Lauritzen, S. L. and Wermuth, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *The Annals of Statistics* **17 (1)**, 31–57.

Lipsitz, S. R., Laird, N. M., and Harrington, D. P. (1990). Finding the design matrix for the marginal homogeneity model. *Biometrika* **77 (2)**, 353–358.

Long, J. S. (1984). Estimable functions in log-linear models. *Sociological methods and research* **12 (4)**, 399–432.

Lovinson, G. (2000). Generalized symmetry models for hypercubic concordance tables. *International Statistical Review* **68 (3)**, 323–338.

Madansky, A. (1963). Tests of homogeneity for correlated samples. *Journal of the American Statistical Association* **58 (301)**, 97–119.

Mair, P. (2007). A Framework to interpret nonstandard log-linear models. *Austrian Journal of Statistics* **36 (2)**, 89–103.

Mardia, K. V., Kent, J. K., and Bibby, J. M. (1979). *Multivariate analysis*. New York: Academic Press.

McCullagh, P. (1978). A class of parametric models for the analysis of square contingency tables with ordered categories. *Biometrika* **65 (2)**, 413–418.

McCullagh, P. (1982). Some applications of quasi-symmetry. *Biometrika* **69 (2)**, 303–308.

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*. London: Chapmann and Hall, 2nd edition.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* **12**, 153–157.

Meiser, T., von Eye, A., and Spiel, C. (1997). Log-linear symmetry and quasi-symmetry models for the analysis of change. *Biometrical Journal* **39 (3)**, 351–368.

Murray, L. W. and Speed, F. M. (1978). An algorithm for efficiently generating the design matrix and the associated covariance matrix in linear models. In *Proceedings of the 16th annual ACM Southeast regional conference*, pages 208–209.

Neufeld, H. (2009). *Graphical Gaussian models with symmetries*. Dissertation submitted to the University of Oxford in support of an application for transfer to doctoral student status. Available from `http://www.stats.ox.ac.uk/~gehrmann/transfer_thesis.pdf`.

Perlman, M. D. (1987). Comment: Group symmetry covariance models. *Statistical Science* **2 (4)**, 421–425.

Plackett, R. L. (1981). *The analysis of categorical data*. London: Griffin, 2nd edition.

Ramírez-Aldana, R. (2005). *An application of probabilistic graphical models in medical research (In Spanish)*. Master's thesis, Graduate studies in Mathematics, National Autonomous University of Mexico, UNAM.

Read, T. R. C. and Cressie, N. A. C. (1988). *Goodness of fit statistics for discrete multivariate data*. New York: Springer-Verlag.

Rinaldo, A. (2006a). *Computing maximum likelihood estimates in log-linear models*. Preliminar manuscript, Department of Statistics, Carnegie Mellon University. Available from `http://www.stat.cmu.edu/~arinaldo/`.

Rinaldo, A. (2006b). *On maximum likelihood estimation in log-linear models*. Technical report 833, Department of Statistics, Carnegie Mellon University. Available from `http://www.stat.cmu.edu/~arinaldo/`.

Rindskopf, D. (1984). Linear equality restrictions in regression and log-linear models. *Psychological bulletin* **96**, 597–603.

Rindskopf, D. (1990). Non standard log-linear models. *Psychological bulletin* **108**, 150–162.

Rotman, J. J. (1995). *An introduction to the theory of groups*. New York: Springer-Verlag, 4th edition.

Tjur, T. (1982). A connection between Rasch's item response analysis model and a multiplicative Poisson model. *Scandinavian Journal of Statistics* **9**, 23–30.

van Horebeek, J. and Teugels, J. L. (1998). Generalized graphical models for discrete data. *Statistics and Probability Letters* **38 (1)**, 41–47.

Vermunt, J. K. (1996). *Log-linear event history analysis: A general approach with missing data, latent variables, and unobserved hetereogeneity*. Tilburg, Netherlands: Tilburg Univerity Press.

Vermunt, J. K. (2005). Log-linear models. In *Encyclopedia of Statistics in behavioral science*, pages 1082–1093. Chichester, England: John Wiley and Sons. Available from `http://spitswww.uvt.nl/~vermunt/`.

von Eye, A. and Spiel, C. (1996). Standard and non standard log-linear symmetry model for measuring change in categorical variables. *American Statistician* **50**, 300–305.

Wermuth, N. and Marchetti, G. M. (2009). Triangular systems for symmetric binary variables. *Electronic Journal of Statistics* **3**, 932–955.

Whittaker, J. (1990). *Graphical models in applied multivariate statistics*. Chichester, England: John Wiley and Sons. Paperback edition 2009 from Wiley.

Electronic pages

MIM can be downloaded from

`http://www.hypergraph.dk`

Graphviz can be downloaded from

`http://www.graphviz.org`