Restricted or coloured graphical log-linear models

Dr. Guillermina Eslava-Gómez [1]
M. Sc. Ricardo Ramírez-Aldana [2]
October 2008

Contents

1 Summary **2**

2 Introduction **2**

3 Notation and terminology **3**

4 Vertex colouring **6**

5 Edge colouring **8**

6 Vertex and Edge colouring **11**

7 Symmetry and quasi-symmetry models expressed as restricted graphical log-linear models **14**

8 Likelihood equations **17**
8.1 *Vertex colouring* 20
8.2 *Edge colouring* 20
8.3 *Vertex and edge colouring* 21

9 Restricted graphical log-linear models expressed as GLM **23**
9.1 *Vertex colouring* 23
9.2 *Edge colouring* 29
9.3 *Vertex and Edge colouring* 33

10 Reparametrized models **35**

11 Solution to the likelihood equations **38**

12 Discussion and perspectives **41**

13 Basic References **43**

[1]Dept. of Mathematics, Faculty of Sciences, UNAM, Ciudad Universitaria, 04510, D.F. Mexico. E-mail eslava@matematicas.unam.mx

[2]Ph. D. Candidate. This material is part of the work being done in order to get a Ph. D. in Statistics at the National Autonomous University of Mexico. Dr. Guillermina Eslava is supervisor of this project. E-mail richram_mx@yahoo.es

1 Summary

We introduce new types of graphical log-linear models (Edwards, 2000; Lauritzen, 1989, 1996; Whittaker, 1990) by placing restrictions on the main effects and first order interactions. These restrictions can be represented using colours in the corresponding associated graphs. The graphs include multiple edges between vertices to represent all the different permutations among the categories of the associated variables included according to the first order interaction terms in the model. The vertices or variables on the same colour class have the same main effects in all their categories, and the multiple edges in the same colour class imply that we have the same first order interaction for the elements in the same class. The symmetry and quasi-symmetry models introduced by Caussinus (1965) are particular cases of these models. We study the properties of the restricted models and calculate the associated likelihood equations. Also, we express restricted models as generalized linear models (GLM) and get the corresponding design matrix which can only be reparametrized as a full rank matrix in some particular cases . Finally, we derive an algorithm to solve the likelihood equations, which is a modification of the Iterative Proportional Fitting algorithm.

Keywords: Generalized linear models, graph colourings, graphical log-linear models, iterative proportional fitting, log-linear models, Newton Raphson method, symmetry and quasi-symmetry models.

2 Introduction

This report introduces a new type of graphical log-linear models, restricted or coloured graphical log-linear models, which are particular cases of hierarchical log-linear models. Log-linear models, specifically restricted graphical models, are used to understand the association between discrete variables for data which are usually presented as contingency tables. Log-linear models could be expressed as ANOVA models in which the response variable is the expected frequency and the explanatory variables are reflected in the main effects and interactions included in the model. Different models are represented according to the parameters included. Restricted graphical log-linear models can be represented graphically including equality restrictions between certain parameters. There are 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.

These models with discrete variables are analogous to graphical Gaussian models with edge and vertex symmetries introduced by Højsgaard and Lauritzen (2005, 2007a, and 2007b). 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 by using vertex and edge colourings imposes restrictions on parameters associated to main effects and first order interactions, but does not imply restrictions on parameters associated to second or higher order interactions, in this sense restricted graphical log-linear models can be still further generalized by restricting parameters associated to two or higher order interactions.

Some advantages of restricted graphical log-linear models are that they can: a) give a better fit to the data because they could adapt better to the underlying structure, b) help to get a better understanding of the relations among the cells on the table, c) be used to visualize the same information got in graphical log-linear models, for example we can get marginal and conditional independences among the variables using the graphical concept of separator sets, d) be more parsimonious than the corresponding models without restrictions, and finally, e) present the information in a visual, intuitive, and accessible way, mainly when the number of variables is not so large, for example less than 10 variables.

Symmetry and quasi-symmetry models (Caussinus, 1965, Agresti, 2002, p. 423-431, Bishop *et al.*, 1975, p. 282-308) can be seen as particular cases of restricted graphical models with two variables. In this sense, restricted graphical log-linear models could be considered as a generalization of these models as suggested by Højsgaard and Lauritzen (2007a, p. 21). For some other generalizations of symmetry and quasi-symmetry see Andersen, 1991, p. 328-329, Bishop *et al.*, 1975, p. 299-309, and Haberman, 1978 and 1979, p. 503-509. Graphical log-linear models are also a particular case of restricted graphical models in which there are no additional restrictions imposed on the parameters which is a possibility included in restricted graphical models.

The following relations are observed among the restricted graphical log-linear 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.

3 Notation and terminology

$\vert Z\vert$	number of elements in a set Z.
$\Delta{=}V$	set of vertices or set of variable names.
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,...,$k(i)$.
$I = (I_\delta)_{\delta\in\Delta}$	discrete random variables associated to the set of vertices.
I_δ	categories or level set for I_δ.
$\vert\mathrm{I}_\delta\vert$	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 for a given 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, $m' = (m(i))_{i \in \mathrm{I}}$.
$\hat{m}(i)$	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 sub vector of I with values 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 associated and depending only on the values of the variables in subset a, $a \subseteq \Delta$. $u_a(i_a)$ = constant, when $a = \emptyset$; $u_a(i_a)$ = main effect, when $\lvert a \rvert = 1$; $u_a(i_a)$ = interaction, when $\lvert a \rvert > 1$.
$u_{l_r^t m_r^t}(i_r^t j_r^t)$	$r = 1, 2, ..., k(t)$; t=1,2,...,S. First order interaction. It can be identified with the edge e_r^t joining the variable l_r^t to the variable m_r^t with the value combination (i_r^t, j_r^t), where e_r^t is the r-th element in the colour class E_t.
$n_a(i_a)$	$n_a(i_a) = \sum_{j: j_a = i_a} n(j)$, 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.
$\hat{m}_a(i_a)$	$\hat{m}_a(i_a) = \sum_{j: j_a = i_a} \hat{m}(j)$, marginal estimated expected frequency for i_a.
$\lvert n \rvert$	Total observed count, $\sum_{i \in \mathrm{I}} n(i)$.

A saturated model, which is a log-linear model that includes all possible effects for every variable, can be written as

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

Example 1. Suppose we have three variables, X, Y, and Z, *i.e.* $\Delta = \{X, Y, Z\}$, then the saturated model is

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

where, i, j, and k are categories or levels for X, Y, and Z, respectively.

According to zero assignations to some parameters, we can get different kinds of models. Hierarchical log-linear models are those models in which if an interaction is included, then all interactions whose variables are subsets of that interaction are also included. It can be defined a generating class A, which is a set of subsets used to determine all the parameters included in a model. Then, a hierarchical log-linear model takes the form

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

where K is the set of subsets of the elements in the generating class A.

Example 2. Suppose we have $\Delta = \{X, Y, Z\}$, and a hierarchical model with generating class A= $\{\{X, Y\}$, $\{Y, Z\}\}$. The corresponding model is

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

where, i, j, and k are categories or levels for X, Y, and Z, respectively. The model is obtained using the fact that the set of subsets of the elements in the generating class K, is K= $\{\{X, Y\}$, $\{Y, Z\}$, $\{X\}$, $\{Y\}$, $\{Z\}$, $\emptyset\}$.

Definition

A *restricted graphical log-linear model* with graph $G = (V, E)$ is a hierarchical log-linear model

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

with generating class C, the cliques set in the associated graph, where K is the set of all subsets of the elements in the generating class and where the set of variables V and the set of first order interactions E are partitioned as follows. V is partitioned into $(V_1, ..., V_T)$, with $V_i \neq \emptyset$, i=1,...,T, $T \in \{1, 2, ..., |V|\}$, such that the main effects of the variables in V_i are equal in all their levels, E is partitioned into $(E_1, ..., E_S)$, with $E_i \neq \emptyset$, i=1,...,S, $S \in \{1, 2, ..., |E|\}$, such that the interactions in every E_i are equal.

We have some observations about this definition:

a) The graphs associated to the restricted models are similar to the independence or interaction graphs defined in graphical log-linear models, in the sense that the underlying graph associated to the restricted graphical model is the same to the independence graph associated to the 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 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 is taken directly from graph theory and it is the same concept used in graphical models. The cliques are the complete maximal subgraphs of a graph. In restricted graphical log-linear 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 or E_i are formed by a single element, we call them atomic classes, otherwise they are composite classes as defined for the continuous case in Højsgaard and Lauritzen, 2007a, p. 5. Atomic classes are those in which all the corresponding parameters are not restricted. A restricted model with only atomic classes is a graphical model.

d) It is assumed that all variables have the same number of categories. This condition could be relaxed assuming that only those variables in composite vertex classes should have the same number of categories.

Restricted graphical models can be represented using colourings, understanding colourings as the concept defined in graph theory (Bondy, 1976, ch. 6 and 8). This means that we can represent the model using a graph in which some vertices and edges have different colours and also according to a coloured graph we can determine the restrictions in the parameters. We have three kinds of colourings: vertex, edges, and, vertex and edges colourings.

In the following sections we give different types of restricted graphical log-linear models and some examples.

4 Vertex colouring

Suppose we have a restricted graphical log-linear model restricting the main effects only, *i.e.* we have the model

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

with graph $G = (V, E)$, where V is partitioned into $(V_1, ..., V_T)$, with $V_i \neq \emptyset$, i=1,...,T, $T \in \{1, 2, ..., |V|\}$, in such a way that the main effects for the variables in every set V_i are equal in all their levels.

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: main effects.
$|\mathrm{I}_X| = |\mathrm{I}_Y|$: total number of categories for the variables X and Y.

Suppose we have a partition of V into $V_1, V_2, ..., V_T$ vertex colour classes, in which every variable has J levels, and where $V_i = \left\{v_1^i, ..., v_{k(i)}^i\right\}$, i = 1, ..., T. We define

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

Then, we have the restrictions

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

where $u_i(j)$ is the parameter that represents all the identical parameters.

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 we can identify restrictions in a restricted graphical log-linear model restricting the main effects only with a vertex colouring.

Example 3. Suppose we have four variables, V= $\{X, Y, Z, W\}$, and suppose that every variable has the same number of categories, $|\mathrm{I}_X| = |\mathrm{I}_Y| = |\mathrm{I}_Z| = |\mathrm{I}_W|$. A graphical log-linear model restricted in its vertices is the model with generating class $A = \{\{X, Y\}, \{Y, Z\}, \{Y, W\}\}$, *i.e.* the model

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

in which V has been partitioned into (V_1, V_2), with $V_1 = \{X, Y\}$ and $V_2 = \{Z, W\}$.

In the notation given, $k(1) = 2$, $v_1^1 = X$, $v_2^1 = Y$ and $k(2) = 2$, $v_1^2 = Z$, $v_2^2 = W$. 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 parameters $u_1(i)$ instead of the parameters $u_X(i)$ and $u_Y(i)$, and the parameters $u_2(i)$ instead of the parameters $u_Z(i)$ and $u_W(i)$, for all i.

Assuming that all the variables are binaries, $|\mathrm{I}_X| = 2$, the model has the associated graph presented in figure 1.

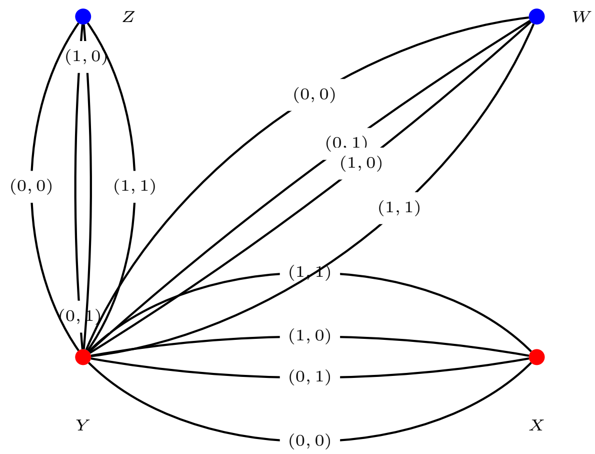

Figure 1. *Vertex colouring for the restricted graphical log-linear model with generating class* $\{\{X,Y\}, \{Y,Z\}, \{Y,W\}\}$, $\log m(i,j,k,l) = u + u_X(i) + u_Y(j) + u_Z(k) + u_W(l) + u_{XY}(ij) + u_{YZ}(jk) + u_{YW}(jl)$, $i,j,k,l = 0,1$; *with vertices* $V = (V_1, V_2)$, $V_1 = \{X,Y\}$ *and* $V_2 = \{Z,W\}$.

5 Edge colouring

Suppose we have a restricted graphical log-linear model restricting only the first order interactions, *i.e.*

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

with $G = (V, E)$ and where the first order interaction set E is partitioned into $(E_1, ..., E_S)$, $E_i \neq \emptyset$, i=1,...,S, $S \in \{1, 2, ..., |E|\}$, in such a way that the first order interactions in every set E_i are identical.

Two edges are in the same edge colour class if and only if $u_{XY}(i,j) = u_{ZR}(l,m)$, where

$X, Y, Z, R \in V$.
$|\mathrm{I}_X|$, $|\mathrm{I}_Y|$, $|\mathrm{I}_Z|$, $|\mathrm{I}_R|$: total number of categories for the variables X, Y, Z, and R, respectively.
$i \in \mathrm{I}_X = \{1, ..., |\mathrm{I}_X|\}$, $j \in \mathrm{I}_Y = \{1, ..., |\mathrm{I}_Y|\}$, $l \in \mathrm{I}_Z = \{1, ..., |\mathrm{I}_Z|\}$, and $m \in \mathrm{I}_R = \{1, ..., |\mathrm{I}_R|\}$.
$u_{XY}(i,j)$, $u_{ZR}(l,m)$: first order interactions.

Suppose we have a partition of the edges set E into $E_1, E_2, ..., E_S$ edge colour classes, with $E_t = \left\{e_1^t, ..., e_{k(t)}^t\right\}$, t=1,2,...,S, where:

e_r^t: r-th element in the colour class E_t.
$u_{l_r^t m_r^t}(i_r^t j_r^t)$, $r = 1, 2, ..., k(t)$; t=1,2,...,S : first order interaction. It can be identified with the edge e_r^t joining the variable l_r^t to the variable m_r^t joining the value combination

(i^t_r, j^t_r).
l^t_r: r-th variable for the t class in the first entry of the first order interaction $u_{l^t_r m^t_r}(i^t_r j^t_r)$.
m^t_r: r-th variable for the t class in the second entry of the first order interaction $u_{l^t_r m^t_r}(i^t_r j^t_r)$.
i^t_r: category for l^t_r.
j^t_r: category for m^t_r.

Observe that partitioning the edges set or the first order interactions set is equivalent. This is a consequence of the fact that the $u_{l^t_r m^t_r}(i^t_r j^t_r)$ parameters are identified with the corresponding edges.

Then, we have the restrictions

$$u_{l^t_1 m^t_1}(i^t_1 j^t_1) = ... = u_{l^t_{k(t)} m^t_{k(t)}}(i^t_{k(t)} j^t_{k(t)}) = u_{E_t},$$

where u_{E_t} is the parameter that represents all the identical parameters in the same colour class.

Similarly, if we have a partition of the set of first order interactions in such a way that all interactions in the same class are identical, then we can assign to every class a different colour. This means that we can identify restrictions in a restricted graphical model restricting the first order interactions only with an edge colouring.

Example 4. Suppose we have four binary variables, X, Y, Z, and W, and a restricted graphical log-linear model restricting some of its edges. The generating class of this model is $\{\{X, Y\}, \{W, Y\}, \{W, Z\}, \{X, Z\}\}$; *i.e.*,

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

with i, j, k, l=0, 1. The first order interaction set, and consequently the edges set, E, obtained by identifying each parameter with an edge, is

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

The partition of the set of first order interactions, and as a consequence of the set of edges, is a partition into four classes E_1, E_2, E_3, E_4,

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

In this case, $k(i) = 4$, 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

$$e_1^1 \approx u_{XY}(00),\ e_2^1 \approx u_{XY}(11),\ e_3^1 \approx u_{WZ}(00),\ e_4^1 \approx u_{WZ}(11).$$
$$e_1^2 \approx u_{XY}(01),\ e_2^2 \approx u_{XY}(10),\ e_3^2 \approx u_{WZ}(01),\ e_4^2 \approx u_{WZ}(10).$$
$$e_1^3 \approx u_{WY}(00),\ e_2^3 \approx u_{WY}(11),\ e_3^3 \approx u_{XZ}(00),\ e_4^3 \approx u_{XZ}(11).$$
$$e_1^4 \approx u_{WY}(01),\ e_2^4 \approx u_{WY}(10),\ e_3^4 \approx u_{XZ}(01),\ e_4^4 \approx u_{XZ}(10).$$
$$l_i^j = X,\ m_i^j = Y;\ i = 1,2,\ j = 1,2.$$
$$l_i^j = W,\ m_i^j = Z;\ i = 3,4,\ j = 1,2.$$
$$l_i^j = W,\ m_i^j = Y;\ i = 1,2,\ j = 3,4.$$
$$l_i^j = X,\ m_i^j = Z;\ i = 3,4,\ j = 3,4.$$
$$i_r^t = 0, j_r^t = 0;\ r = 1,3,\ t = 1,3.$$
$$i_r^t = 1, j_r^t = 1;\ r = 2,4,\ t = 1,3.$$
$$i_r^t = 0, j_r^t = 1;\ r = 1,3,\ t = 2,4.$$
$$i_r^t = 1, j_r^t = 0;\ r = 2,4,\ t = 2,4.$$

Given this colouring, we have the restrictions:

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

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, respectively.

The corresponding graph is shown in figure 2.

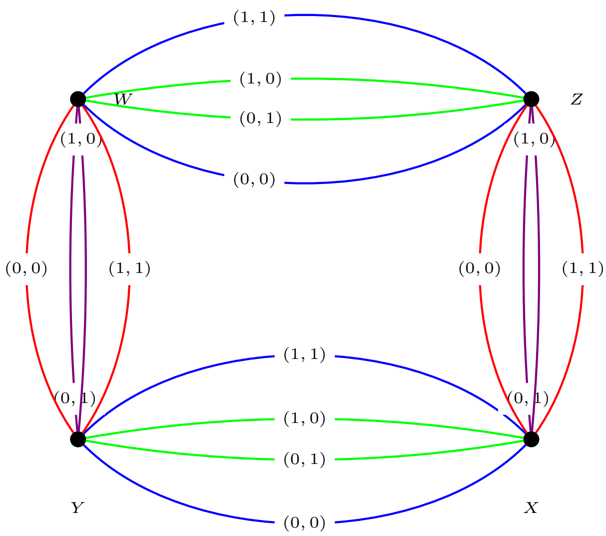

Figure 2. *Edge colouring for the restricted graphical log-linear model with generating class* $\{\{X,Y\}, \{W,Y\}, \{W,Z\}, \{X,Z\}\}$, $\log m(i,j,k,l) = u+ u_W(i)+ u_X(j)+ u_Y(k)+ u_Z(l)+ u_{XY}(jk)+ u_{WY}(ik)+ u_{WZ}(il)+ u_{XZ}(jl)$, *with first order interactions set partitioned into* $E=(E_1,E_2,E_3,E_4)$, $E_1=\{u_{XY}(00), u_{XY}(11), u_{WZ}(00), u_{WZ}(11)\}$, $E_2=\{u_{XY}(01), u_{XY}(10), u_{WZ}(01), u_{WZ}(10)\}$, $E_3=\{u_{WY}(00), u_{WY}(11), u_{XZ}(00), u_{XZ}(11)\}$, $E_4=\{u_{WY}(01), u_{WY}(10), u_{XZ}(01), u_{XZ}(10)\}$.

6 Vertex and Edge colouring

In this case we simultaneously use the same definitions of colourings given above. Let $G=(V,E)$ be a restricted graphical log-linear model, with $(V_1,V_2,...,V_T)$ a partition of V, $V_i \neq \emptyset$, i=1,...,T, $T \in \{1,2,...,|V|\}$, and $(E_1, E_2,...,E_S)$ a partition of E, $E_i \neq \emptyset$, i=1,...,S, $S \in \{1,2,...,|E|\}$. Suppose that every variable belonging to a non atomic class has J levels. We have

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

$$E_t = \left\{e_1^t, ..., e_{k(t)}^t\right\} \Leftrightarrow u_{l_1^t m_1^t}(i_1^t j_1^t) = u_{l_2^t m_2^t}(i_2^t j_2^t) = ... = u_{l_{k(t)}^t m_{k(t)}^t}(i_{k(t)}^t j_{k(t)}^t) = u_{E_t};\ t=1,...,S.$$

Then, given a colouring, we get a number of restrictions on the parameters and vice versa. This means that we can identify restrictions in a restricted graphical model with vertex and edge colourings. Vertex colouring could be seen as a particular case of these models if we let every edge in an atomic class, *i.e.* $T=|V|$. Similarly, edge colouring could be seen as a particular case of these models if every edge conforms an atomic class, *i.e.* $S=|E|$.

Example 5. Suppose we have three binary variables A, C, and M. A corresponds to alcohol consumption, C to cigarette consumption, and M corresponds to marijuana consumption. Every variable has two categories, *Yes* and *No*, represented with 0 and 1, respectively. The data are presented as a contingency table, table 1 (Agresti, 2002, p. 322).

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

Table 1. *Consumption of alcohol, cigarette and marijuana by high school seniors.*

We present a log-linear graphical model with these variables, and give an example of a restricted graphical log-linear model.

Suppose we have the hierarchical log-linear model with generating class $\{\{A, C\}, \{A, M\}\}$. This is a graphical model because the cliques of the graph are the edges $\{A, C\}$ and $\{A, M\}$, which form the generating class (figure 3), and can be expressed as follows,

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

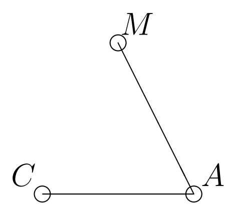

Figure 3. *Graph associated to the graphical model with generating class* $\{\{A, C\}, \{A, M\}\}$, $\log m(i, j, k) = u + u_A(i) + u_C(j) + u_M(k) + u_{AC}(ij) + u_{AM}(ik)$.

Suppose we add to (1) the restrictions

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

We get the graph given in figure 4.

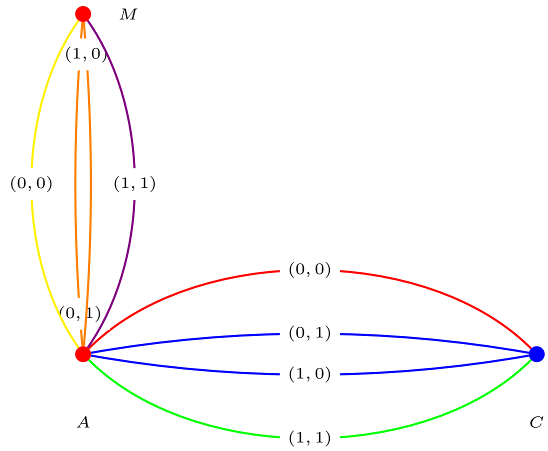

Figure 4. *Colourings for the restricted graphical log-linear model with generating class* $\{\{A,C\},\{A,M\}\}$, $\log m(i,j,k) = u+ u_A(i)+ u_C(j)+ u_M(k)+ u_{AC}(ij)+ u_{AM}(ik)$, *with vertices* $V = (V_1,V_2)$, $V_1 = \{A,M\}$ *and* $V_2 = \{C\}$, *and first order interactions set partitioned into* $E = (E_1,E_2,E_3,E_4,E_5,E_6)$, $E_1=\{u_{AM}(01),\ u_{AM}(10)\}$, $E_2=\{u_{AC}(01),\ u_{AC}(10)\}$, $E_3=\{u_{AM}(00)\}$, $E_4=\{u_{AM}(11)\}$, $E_5=\{u_{AC}(00)\}$, $E_6=\{u_{AC}(11)\}$.

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 edges set is partitioned into $E = (E_1,E_2,E_3,E_4,E_5,E_6)$, where

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

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

Observe that the underlying graph, *i.e.* the associated simple graph, is the one given in figure 3. Using this graph or carefully using the graph given in figure 4 we see that the generating class is identical to the corresponding cliques set. This, together with the restrictions applied to the model, means that we have a restricted graphical log-linear model.

The graphical log-linear model (1) expressed as a restricted graphical log-linear model is shown in figure 5. It is important to notice that given the coloured graph, we could get the associated restricted model.

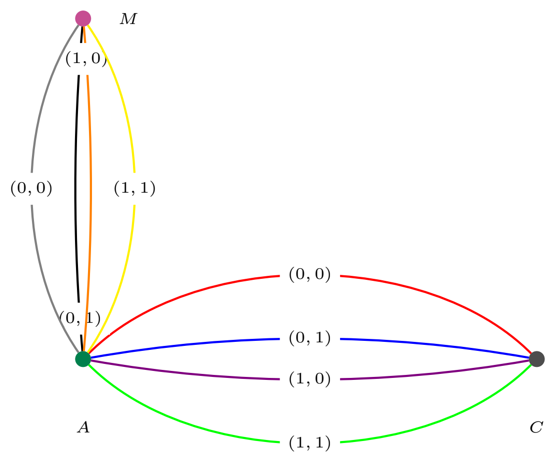

Figure 5. *Colourings for the graphical log-linear model with generating class* $\{\{A, C\}, \{A, M\}\}$, $\log m(i, j, k) = u+ u_A(i)+ u_C(j)+ u_M(k)+ u_{AC}(ij)+ u_{AM}(ik)$, *seen as a restricted graphical log-linear model. In this model* $V = (V_1, V_2, V_3)$, $V_1 = \{A\}$, $V_2 = \{C\}$, *and* $V_3 = \{M\}$. *The first order interactions set is partitioned into* $E = (E_1, E_2, E_3, E_4, E_5, E_6, E_7, E_8)$, $E_1=\{u_{AM}(10)\}$, $E_2=\{u_{AM}(01)\}$, $E_3=\{u_{AM}(00)\}$, $E_4=\{u_{AM}(11)\}$, $E_5=\{u_{AC}(10)\}$, $E_6=\{u_{AC}(01)\}$, $E_7=\{u_{AC}(00)\}$, $E_8=\{u_{AC}(11)\}$.

7 Symmetry and quasi-symmetry models expressed as restricted graphical log-linear models

Symmetry and quasi-symmetry models are defined to analyze square contingency tables, *i.e.* tables with the same number of row and column categories, $|\mathrm{I}|$. Denote $n(i, j)$ and $m(i, j)$ as the observed counts and expected frequency, respectively, for the cell (i, j); i, j $=1, 2, ..., |\mathrm{I}|$. We say that a square table satisfies symmetry if

$$m(i, j) = m(j, i), \ \forall \ i \neq j.$$

This means the expected frequency in a cell (i, j) above the diagonal is similar to the expected frequency in the cell (j, i) under the diagonal.

Supposing that the variables are X and Y, symmetry can be represented as the log-linear model

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

with the restrictions

$$u_X(i) = u_Y(i), \ i = 1, 2, ..., |\mathrm{I}| \, ;$$
$$u_{XY}(ij) = u_{XY}(ji), \ i, j = 1, 2, ..., |\mathrm{I}| \, .$$

Quasi-symmetry 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. This model can be written as

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

with the restrictions

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

Symmetry and quasi-symmetry models can be seen as edge colourings, and vertex and edge colourings, respectively. Every vertex in the quasi-symmetry model belongs to a different atomic class and we have |I| atomic edge color 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 we have only one vertex colour class formed by both vertices.

As an example, suppose we have two binary variables C and A. C corresponds to cigarette consumption, with categories: 0, which indicates less than two packets of cigarettes, and 1, which indicates two or more packets of cigarettes. Variable A corresponds to alcohol consumption with two categories: 0, which indicates less than one drink, and 1, which indicates two or more drinks. The data are presented in a contingency table (table 2).

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

Table 2. *Square contingency table corresponding to alcohol and cigarette consumption for a sample of 1007 students.*

Suppose that we have the saturated log-linear model,

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

As we have an interaction term, there is an edge between A and C. Additionally, we suppose that we have the restricted graphical log-linear model shown in the graph given in figure 6.

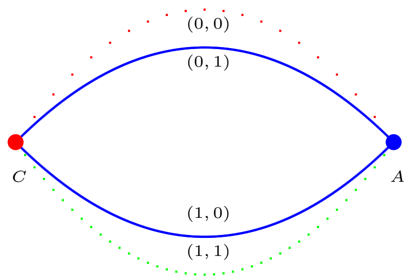

Figure 6. *Colourings for a quasi-symmetry model with two binary variables.* $V = \{C, A\}$ *is partitioned into* $V = (V_1, V_2)$*, with* $V_1 = \{C\}$ *and* $V_2 = \{A\}$*. The first order interactions set* E *is partitioned into* $E_1 = \{u_{CA}(00)\}$*,* $E_2 = \{u_{CA}(11)\}$*, and* $E_3 = \{u_{CA}(01), u_{CA}(10)\}$*.*

Using the previous graph $G = (V, E)$, 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 and this means that the corresponding interactions are identical. The remaining edges belong to different atomic colour classes. As an additional observation, if we do not want to use different colours for different atomic classes, we could agree to represent all atomic classes with dot or with black lines.

$V = \{C, A\}$ is partitioned in two classes (V_1, V_2), with $V_1 = \{C\}$ and $V_2 = \{A\}$. This means that the main effects are not restricted.

The model in figure 6 can be expressed as

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

with the restrictions

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

This is a quasi-symmetry model. This means that the quasi-symmetry model for V=$\{A, C\}$ is the restricted graphical model generated by $\{A, C\}$ 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 $E = (E_1, E_2, E_3)$ partitioned into $E_1 = \{u_{CA}(00)\}$, $E_2 = \{u_{CA}(11)\}$, and $E_3 = \{u_{CA}(01), u_{CA}(10)\}$, whose associated graph is given in figure 6.

On the other hand, suppose that we have the restricted graphical log-linear model shown in the graph given in figure 7, which is also a restricted graphical log-linear model.

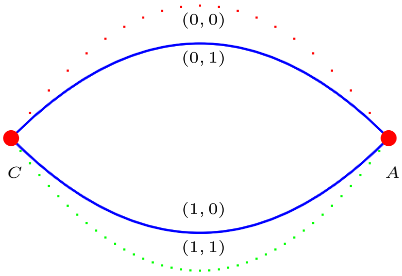

Figure 7. *Colourings for a symmetry model with two binary variables. The first order interactions set E is partitioned into $E_1 = \{u_{CA}(00)\}$, $E_2 = \{u_{CA}(11)\}$, and $E_3 = \{u_{CA}(01), u_{CA}(10)\}$.*

Using the previous graph, $G = (V, E)$, we observe that we have the same edge or first order interactions partition as before, $E = (E_1, E_2, E_3)$, 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 in a single element, V. This means that the main effects are the same for all the categories.

The model in figure 7 can be expressed as

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

with restrictions

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

which corresponds to a symmetry model. This means that the symmetry model for V=$\{A, C\}$ is the restricted graphical model generated by $\{A, C\}$ with vertex colour class V and and first order interaction set or their corresponding edge set $E = (E_1, E_2, E_3)$ partitioned into $E_1 = \{u_{CA}(00)\}$, $E_2 = \{u_{CA}(11)\}$, and $E_3 = \{u_{CA}(01), u_{CA}(10)\}$, whose associated graph is given in figure 7.

8 Likelihood equations

Data for contingency tables are collected under different sampling schemes. Cell counts could follow a Poisson, multinomial, or a restricted multinomial distribution with fixed marginals. Under the Poisson sampling scheme the cell counts are realizations of independent and Poisson distributed random variables $\{N(i)\}_{i\in \mathrm{I}}$. Thus $E(N(i)) = m(i)$ and the joint distribution of the counts becomes

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

with $m(i) \geq 0$.

If the total observed count $|n|$ is fixed, but the counts in each cell are random, and each of these 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

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

where $m(i) = |n|\, p(i)$.

Finally, 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 no 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

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

where $m(i) = n(i_b)p(i_a|i_b)$.

In any of the three sampling schemes, the logarithm of the kernel of the likelihood equation is

$$\sum_{i\in\mathrm{I}} n(i) \log m(i) - \sum_{i\in\mathrm{I}} m(i), \tag{2}$$

where $n(i)$ and $m(i)$ are the observed count and the expected frequency for a cell i, respectively, $i \in \mathrm{I}$.

If we consider a log-linear model for $m(i)$ in equation (2), we obtain an equation that depends on the parameters. For example, using a hierarchical log-linear model, equation (2) becomes

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

Before proceeding to obtain the likelihood equations in sections **8.1**, **8.2**, and **8.3** we define the following notation:

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

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

Example 6. Suppose we have the data given in table 2, and 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 cells over all possible values of the entry where the dot is. The process is similar for the expected frequencies, simply changing the $n(,)$ by $m(,)$.

Let $\{e_k^t\}_{k=1,...,k(t),t=1,...,S}$ be an edge set, with e_k^t identified with the parameter $u_{l_k^t m_k^t}(i_k^t, j_k^t)$, partitioned into $E_1, E_2, ...E_S$ colour classes, with $E_t = \left\{e_1^t, ..., e_{k(t)}^t\right\}$, t=1,..., S. The marginal total for the k-th variable pair, with their corresponding levels, in the colour class t is defined as

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

Example 7. Using the data given in table 2, suppose that the first order interaction set or their 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 $k(1) = k(2) = 2$; $l_k^t = C$, 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 equation(2), including the model with the different kinds of restrictions according to the colourings, and using the notation given above, we obtain the corresponding likelihood equations.

8.1 *Vertex colouring*

In this case, the logarithm of the kernel of the likelihood function, equation (2), including the restrictions takes the form

$$\sum_{i=1}^{T}\sum_{l=1}^{J}\sum_{k=1}^{k(i)} n(v_k^i = l)u_i(l) + \sum_{a\subseteq K, |a|\neq 1}\sum_{i_a} n_a(i_a)u_a(i_a) - \sum_{i\in \mathrm{I}} \exp(\sum_{a\subseteq K} u_a(i_a)), \quad (3)$$

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.

Deriving (3) 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

$$\begin{aligned}
\sum_{j=1}^{k(i)} n(v_j^i = 1) &= \sum_{j=1}^{k(i)} m(v_j^i = 1), i = 1, ..., T \\
\sum_{j=1}^{k(i)} n(v_j^i = 2) &= \sum_{j=1}^{k(i)} m(v_j^i = 2), i = 1, ..., T \\
&\vdots \\
\sum_{j=1}^{k(i)} n(v_j^i = J) &= \sum_{j=1}^{k(i)} m(v_j^i = J), i = 1, ..., T \\
n_a(i_a) = m_a(i_a),\ &\forall a \subseteq K,\ |a| \neq 1,
\end{aligned} \quad (4)$$

To reduce the number of equations, we could eliminate redundant equations. To do this, we substitute the last part in the previous equation system (4) with

$$n_a(i_a) = m_a(i_a),\ \forall a \in A,\ |a| \neq 1$$

8.2 *Edge colouring*

In this case, the logarithm of the kernel of the likelihood function, equation (2), including the restrictions takes the form

$$\sum_{t=1}^{S}\sum_{k=1}^{k(t)} n(l_k^t = i_k^t, m_k^t = j_k^t)u_{E_t} + \sum_{a\subseteq K, |a|\neq 2}\sum_{i_a} n_a(i_a)u_a(i_a) - \sum_{i\in \mathrm{I}} \exp(\sum_{a\subseteq K} u_a(i_a)). \quad (5)$$

Deriving (5) with respect to each parameter and equating to zero, we get the following equation system

$$\sum_{k=1}^{k(t)} n(l_k^t = i_k^t, m_k^t = j_k^t) = \sum_{k=1}^{k(t)} m(l_k^t = i_k^t, m_k^t = j_k^t),\ t = 1, 2, ..., S, \tag{6}$$
$$n_a(i_a) = m_a(i_a), \forall a \subseteq K, |a| \neq 2.$$

Similarly to the vertex colouring, we can eliminate some redundant equations. To accomplish this, we have to eliminate the last part of the equation system, and instead we have to use the equations for the main effects, and the equations corresponding to the generating class, *i.e.*, we substitute the last part in the equation system (6) with

$$n_a(i_a) = m_a(i_a),\ \forall a \in A,\ |a| \neq 2,$$
$$n_a(i_a) = m_a(i_a),\ |a| = 1.$$

8.3 *Vertex and edge colouring*

In this case, the logarithm of the kernel of the likelihood function, equation (2), including both types of parameter restrictions takes the form

$$\sum_{i=1}^{T}\sum_{l=1}^{J}\sum_{k=1}^{k(i)} n(v_k^i = l)u_i(l) + \sum_{t=1}^{S}\sum_{k=1}^{k(t)} n(l_k^t = i_k^t, m_k^t = j_k^t)u_{E_t} +$$
$$\sum_{a\subseteq K, |a|\neq 1,2}\ \sum_{i_a} n_a(i_a)u_a(i_a) - \sum_{i \in \mathrm{I}} \exp(\sum_{a \subseteq K} u_a(i_a)). \tag{7}$$

Deriving this equation with respect to each parameter and equating to zero we obtain the equation systems (4) and (6) simultaneously. The equation system is

$$\sum_{j=1}^{k(i)} n(v_j^i = 1) = \sum_{j=1}^{k(i)} m(v_j^i = 1), i = 1, ..., T$$
$$\sum_{j=1}^{k(i)} n(v_j^i = 2) = \sum_{j=1}^{k(i)} m(v_j^i = 2), i = 1, ..., T$$
$$\vdots$$
$$\sum_{j=1}^{k(i)} n(v_j^i = J) = \sum_{j=1}^{k(i)} m(v_j^i = J), i = 1, ..., T$$
$$\sum_{k=1}^{k(t)} n(l_k^t = i_k^t, m_k^t = j_k^t) = \sum_{k=1}^{k(t)} m(l_k^t = i_k^t, m_k^t = j_k^t),\ t = 1, 2, ..., S$$
$$n_a(i_a) = m_a(i_a),\ \forall a \subseteq K,\ |a| \neq 1, 2.$$

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

$$n_a(i_a) = m_a(i_a),\ \forall a \in A,\ |a| \neq 1, 2.$$

To exemplify the likelihood equations, consider the symmetry and quasi-symmetry models given in section **7**. As we saw before, there are three edge colour classes or partitions of the first order interactions set in both models: $E_1 = \{u_{CA}(00)\}$, $E_2 = \{u_{CA}(01),\ u_{CA}(10)\}$, and $E_3 = \{u_{CA}(11)\}$. There is an equation associated to each class. These are given by

$$\begin{aligned} E_1 :\ & n(0,0) = m(0,0), \\ E_2 :\ & n(0,1) + n(1,0) = m(0,1) + m(1,0), \\ E_3 :\ & n(1,1) = m(1,1), \end{aligned}$$

which can be rewritten as

$$\hat{m}(i,j) + \hat{m}(j,i) = n(i,j) + n(j,i),\ \forall i \leq j, i = j = 0, 1. \tag{8}$$

In the quasi-symmetry model, we also have two vertex colour classes, $V_1 = \{C\} = \{v_1^1\}$ and $V_2 = \{A\} = \{v_1^2\}$. Then we have

$$\begin{aligned} n(v_1^1 = 0) &= n(0, .), \\ n(v_1^1 = 1) &= n(1, .), \\ n(v_1^2 = 0) &= n(., 0), \\ n(v_1^2 = 1) &= n(., 1). \end{aligned}$$

The equations corresponding to class V_1 are

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

and to V_2

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

These equations can be rewritten as

$$\hat{m}(i, .) = n(i, .),\ i = 0, 1;\ \hat{m}(., j) = n(., j),\ j = 0, 1. \tag{9}$$

Equations systems (8) and (9) have to be solved simultaneously.

In the symmetry model, both variables are in the same colour class, $V = \{C, A\}$. The likelihood equations corresponding to this class are

$$\begin{aligned} n(0, .) + n(., 0) =& m(0, .) + m(., 0), \\ n(1, .) + n(., 1) =& m(1, .) + m(., 1). \end{aligned} \tag{10}$$

Then, we have to solve (8) and (10) simultaneously; however, we notice that equations in (8), imply equations in (10), so that it is only necessary to solve (8).

9 Restricted graphical log-linear models expressed as GLM

All restricted 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 get an adequate parametrization of the model whose 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 determine the good or bad adjustment of a model.

A log-linear model written as a generalized linear model (GLM) has the form (see *e.g.* Agresti, 2002, p. 116)

$$\log(m) = \boldsymbol{X}\boldsymbol{\beta}, \tag{11}$$

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 $N(i)$.
$\boldsymbol{X}\boldsymbol{\beta}$	systematic component.
$\boldsymbol{X}$	design matrix. Every row corresponds to a cell and it has as many columns as parameters included in the model.
$\boldsymbol{\beta}$	parameter vector $(\beta_1, ..., \beta_q)$.
$\log()$	link function connecting the random and systematic components.

9.1 *Vertex colouring*

In order to get the design matrix including the vertex or edge colouring, we need the following notation.

$\boldsymbol{\beta}$	$\boldsymbol{\beta}$=$(\beta_1, ..., \beta_p)$, vector of parameters, it includes all parameters got due to vertex colouring.
$\boldsymbol{X}$	design matrix for the restricted graphical log-linear 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; $l = 1, 2, ..., k(i)$; $i = 1, ..., T$, 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. $\boldsymbol{X}^{v_l^i(j)}$ indicates the variable v_l^i effect over the $u_i(j)$ parameter for a specific colouring i and a category j.

Let

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

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

We denote

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

Then, we have

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

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, and assigning zero to the remaining terms, these remaining terms are the vertex colouring parameters.

We observe that

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

Example 8. Suppose we have the vertex set $V=\{X,\ Y,\ Z,\ W\}$, where all variables are binary, and generating class $\{\{X\},\ \{Z\},\ \{Y, W\}\}$; *i.e.* we have the model

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

Suppose V is partitioned into $V_1=\{X,Y\}$ and $V_2=\{Z,W\}$, so that $v_1^1 = X$, $v_2^1 = Y$, $v_1^2 = Z$, and $v_2^2 = W$, figure 8.

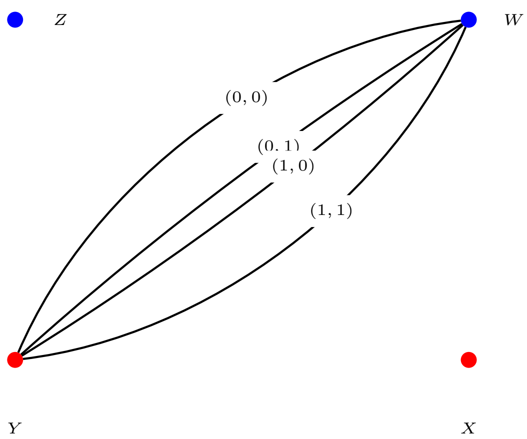

Figure 8. *Vertex colouring for the restricted graphical log-linear model with generating class* $\{\{X\}, \{Z\}, \{Y, W\}\}$, $\log m(i,j,k,l) = u + u_W(i) + u_X(j) + u_Y(k) + u_Z(l) + u_{WY}(ik)$, $i,j,k,l = 0,1$; $V = (V_1, V_2)$, *with* $V_1 = \{X, Y\}$ *and* $V_2 = \{Z, W\}$.

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 parameters vector 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 colour class V_1:

$\boldsymbol{X}^{v_1^1(0)} =$

	u	$u_1(0)$	$u_1(1)$	$u_2(0)$	$u_2(1)$	$u_{WY}(00)$	$u_{WY}(01)$	$u_{WY}(10)$	$u_{WY}(11)$
0000	0	1	0	0	0	0	0	0	0
0001	0	1	0	0	0	0	0	0	0
0010	0	1	0	0	0	0	0	0	0
0011	0	1	0	0	0	0	0	0	0
0100	0	0	0	0	0	0	0	0	0
0101	0	0	0	0	0	0	0	0	0
0110	0	0	0	0	0	0	0	0	0
0111	0	0	0	0	0	0	0	0	0
1000	0	1	0	0	0	0	0	0	0
1001	0	1	0	0	0	0	0	0	0
1010	0	1	0	0	0	0	0	0	0
1011	0	1	0	0	0	0	0	0	0
1100	0	0	0	0	0	0	0	0	0
1101	0	0	0	0	0	0	0	0	0
1110	0	0	0	0	0	0	0	0	0
1111	0	0	0	0	0	0	0	0	0

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 $(0,0,0,0)$, the second one to $(0,0,0,1)$, etc., using the order given for the cells vector. On the other hand

$$\boldsymbol{X}^{v_1^1(1)} = \begin{bmatrix} 0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0 \end{bmatrix}$$

$$\boldsymbol{X}^{v_2^1(0)} = \begin{bmatrix} 0&1&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \end{bmatrix}$$

$$\boldsymbol{X}^{v_2^1(1)} = \begin{bmatrix} 0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0 \end{bmatrix}.$$

Then

$$\boldsymbol{X}^{V_1(0)} = \boldsymbol{X}^{v_1^1(0)} + \boldsymbol{X}^{v_2^1(0)} = \begin{bmatrix} 0&2&0&0&0&0&0&0&0 \\ 0&2&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&2&0&0&0&0&0&0&0 \\ 0&2&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \end{bmatrix}$$

$$\boldsymbol{X}^{V_1(1)} = \boldsymbol{X}^{v_1^1(1)} + \boldsymbol{X}^{v_2^1(1)} = \begin{bmatrix} 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0&0\\ 0&0&2&0&0&0&0&0&0\\ 0&0&2&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0&0\\ 0&0&2&0&0&0&0&0&0\\ 0&0&2&0&0&0&0&0&0 \end{bmatrix}.$$

The same is done for V_2

$$\boldsymbol{X}^{V_2(0)} = \begin{bmatrix} 0&0&0&2&0&0&0&0&0\\ 0&0&0&1&0&0&0&0&0\\ 0&0&0&2&0&0&0&0&0\\ 0&0&0&1&0&0&0&0&0\\ 0&0&0&2&0&0&0&0&0\\ 0&0&0&1&0&0&0&0&0\\ 0&0&0&2&0&0&0&0&0\\ 0&0&0&1&0&0&0&0&0\\ 0&0&0&1&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0 \end{bmatrix}$$

$$\boldsymbol{X}^{V_2(1)} = \begin{bmatrix} 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&1&0&0&0&0\\ 0&0&0&0&1&0&0&0&0\\ 0&0&0&0&2&0&0&0&0\\ 0&0&0&0&1&0&0&0&0\\ 0&0&0&0&2&0&0&0&0\\ 0&0&0&0&1&0&0&0&0\\ 0&0&0&0&2&0&0&0&0\\ 0&0&0&0&1&0&0&0&0\\ 0&0&0&0&2&0&0&0&0 \end{bmatrix}.$$

Finally

$$\boldsymbol{X}^{u,int} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

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

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

The logarithm of the kernel of the likelihood function, equation (2), for a model expressed as a GLM under the vertex colouring is

$$\log(kernel(L(m))) = \sum_{i\in \mathrm{I}} n(i) \sum_{l=1}^{T}\sum_{k} x_i^{V_l(k)} u_l(k) + \sum_{i\in \mathrm{I}} n(i) \sum_{j} x_{ij}\beta_j$$
$$- \sum_{i\in \mathrm{I}} \exp(\sum_{l=1}^{T}\sum_{k} x_i^{V_l(k)} u_l(k) + \sum_{j} x_{ij}\beta_j), \quad (12)$$

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

Based on equation (12), we obtain the corresponding likelihood equations

$$\sum_{i\in \mathrm{I}} n(i) x_i^{V_l(k)} = \sum_{i\in \mathrm{I}} m(i) x_i^{V_l(k)},\ l = 1, 2, ..., T; k = 1, ..., J. \quad (13)$$

$$(\boldsymbol{X}^{u,int.})' n = (\boldsymbol{X}^{u,int.})' m. \quad (14)$$

The previous equation systems, (13) and (14), can be expressed as

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

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 got due to edge colouring.
$\boldsymbol{X}$	design matrix for the restricted graphical log-linear model.
$\boldsymbol{X}^{u_{l_r^t m_r^t}(i_r^t, j_r^t)}$	$r = 1, 2, ..., k(t)$, $t = 1, 2, ...S$. Matrix in which for every row of the column corresponding to u_{E_t}, an entry takes value one when the $u_{l_r^t m_r^t}(i_r^t, j_r^t)$ interaction 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.

It can be obtained

$\boldsymbol{X}^{E_t} = \sum_{r=1}^{k(t)} \boldsymbol{X}^{u_{l_r^t m_r^t}(i_r^t, j_r^t)}$, $t = 1, 2, ..., S$, the 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 the u_{E_t} parameter.
$\boldsymbol{X}^{u,main\ ef.,int.}$	matrix including the coefficients corresponding to the constant, main effects, and two or more order interactions. Additionally, this matrix includes zeros in the columns main associated to the edge colouring parameters.

Using this notation, we have that

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

Example 9. Suppose we have the vertex set $V = \{X, Y, Z, W\}$, where all variables are binaries, and the same model given in example 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 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 parameters vector 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. 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.

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

$$\boldsymbol{X}^{u_{l_2^1 m_2^1}(i_2^1,j_2^1)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{X}^{u_{l_3^1 m_3^1}(i_3^1,j_3^1)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{X}^{u_{l_4^1 m_4^1}(i_4^1, j_4^1)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

Then

$$\boldsymbol{X}^{E_1} = \sum_{r=1}^{4} \boldsymbol{X}^{u_{l_r^1 m_r^1}(i_r^1, j_r^1)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \end{bmatrix}.$$

The same is done for E_2, E_3, and E_4

$$\boldsymbol{X}^{E_2} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$
\boldsymbol{X}^{E_3} = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0
\end{bmatrix}
$$

$$
\boldsymbol{X}^{E_4} = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}.
$$

On the other hand,

$$
\boldsymbol{X}^{u,main\ ef.,int.} = \begin{bmatrix}
1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0
\end{bmatrix}.
$$

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

$$\boldsymbol{X} = \boldsymbol{X}^{u,main\ ef.,int.} + \sum_{l=1}^{4} \boldsymbol{X}^{E_l} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 2 & 0 & 2 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 2 & 0 & 2 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 2 & 2 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 2 & 0 & 0 & 2 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 2 & 0 & 0 & 2 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 2 & 2 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 2 & 0 & 2 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 2 & 0 & 2 & 0 \end{bmatrix}.$$

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

$$\begin{aligned} \log(kernel(L(m))) = & \sum_{i \in \mathrm{I}} n(i) \sum_{l=1}^{S} x_i^{E_l} u_{E_l} + \sum_{i \in \mathrm{I}} n(i) \sum_{j} x_{ij} \beta_j \\ & - \sum_{i \in \mathrm{I}} \exp(\sum_{l=1}^{S} x_i^{E_l} u_{E_l} + \sum_{j} x_{ij} \beta_j), \end{aligned} \tag{15}$$

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

The likelihood equations obtained from (15) are

$$\sum_{i \in \mathrm{I}} n(i) x_i^{E_l} = \sum_{i \in \mathrm{I}} m(i) x_i^{E_l}, l = 1, 2, ..., S; \tag{16}$$

$$(\boldsymbol{X}^{u,main\ ef.,int.})' n = (\boldsymbol{X}^{u,main\ ef.,int.})' m. \tag{17}$$

These equations, (16) and (17), can be expressed as

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

9.3 *Vertex and Edge colouring*

We use the same notation used for vertex and edge of colourings. However, we have to consider that $\boldsymbol{\beta}$ contains parameters for both kinds of colourings, and the design matrix $\boldsymbol{X}$ corresponds to a model whose vertices and edges are simultaneously coloured. We add the following notation.

$\boldsymbol{X}^{u,int \geq 2}$ matrix including the coefficients corresponding to the constant, and the two or more order interactions. The matrix includes zeros in the remaining columns.

We have that

$$\boldsymbol{X} = \boldsymbol{X}^{u,int.\geq 2} + \sum_{l=1}^{S} \boldsymbol{X}^{E_l} + \sum_{i=1}^{T} \sum_{j} \boldsymbol{X}^{V_i(j)}.$$

Example 10. Suppose we have the same model given in example 9, but additionally V is partitioned into (V_1, V_2) with $V_1 = \{X, Y\}$ and $V_2 = \{Z, W\}$.

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 parameters vector 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 8.

$\sum_{i=1}^{2} \sum_{j=0}^{1} \boldsymbol{X}^{V_i(j)} =$

	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	0	2	0	2	0	0	0	0	0
0001	0	2	0	1	1	0	0	0	0
0010	0	1	1	2	0	0	0	0	0
0011	0	1	1	1	1	0	0	0	0
0100	0	1	1	2	0	0	0	0	0
0101	0	1	1	1	1	0	0	0	0
0110	0	0	2	2	0	0	0	0	0
0111	0	0	2	1	1	0	0	0	0
1000	0	2	0	1	1	0	0	0	0
1001	0	2	0	0	2	0	0	0	0
1010	0	1	1	1	1	0	0	0	0
1011	0	1	1	0	2	0	0	0	0
1100	0	1	1	1	1	0	0	0	0
1101	0	1	1	0	2	0	0	0	0
1110	0	0	2	1	1	0	0	0	0
1111	0	0	2	0	2	0	0	0	0

The matrices $\boldsymbol{X}^{E_1}$, $\boldsymbol{X}^{E_2}$, $\boldsymbol{X}^{E_3}$,and $\boldsymbol{X}^{E_4}$ are the same given in example 9, only deleting some zero columns. $\boldsymbol{X}^{u,int.\geq 2}$ takes ones in the column corresponding to u and zeros in the remaining columns. Then, $\boldsymbol{X}$ is

$$\boldsymbol{X} = \boldsymbol{X}^{u,int.\geq 2} + \sum_{l=1}^{4} \boldsymbol{X}^{E_l} + \sum_{i=1}^{2} \sum_{j=0}^{1} \boldsymbol{X}^{V_i(j)} = \begin{bmatrix} 1 & 2 & 0 & 2 & 0 & 2 & 0 & 2 & 0 \\ 1 & 2 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 2 & 0 & 2 \\ 1 & 1 & 1 & 2 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 2 & 2 & 0 \\ 1 & 0 & 2 & 2 & 0 & 2 & 0 & 0 & 2 \\ 1 & 0 & 2 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 0 & 2 & 2 & 0 & 0 & 2 \\ 1 & 1 & 1 & 1 & 1 & 0 & 2 & 2 & 0 \\ 1 & 1 & 1 & 0 & 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 2 \\ 1 & 1 & 1 & 0 & 2 & 1 & 1 & 1 & 1 \\ 1 & 0 & 2 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 2 & 0 & 2 & 2 & 2 & 2 & 0 \end{bmatrix}.$$

The resulting likelihood equations in vertex and edge colouring models are

$$\sum_{i\in\mathrm{I}} n(i)x_i^{V_l(k)} = \sum_{i\in\mathrm{I}} m(i)x_i^{V_l(k)},\ l = 1, 2, ..., T, k = 1, 2, ..., J; \tag{18}$$

$$\sum_{i\in\mathrm{I}} n(i)x_i^{E_l} = \sum_{i\in\mathrm{I}} m(i)x_i^{E_l},\ l = 1, 2, ...S; \tag{19}$$

$$(\boldsymbol{X}^{u,int\geq 2})'n = (\boldsymbol{X}^{u,int\geq 2})'m. \tag{20}$$

These equations systems, (18), (19), and (20), can be expressed as

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

In section **10** we give an example of this way of writing the likelihood equations for the symmetry and quasi-symmetry models.

10 Reparametrized models

The matrices $\boldsymbol{X}$ obtained in section **9** do not have full rank. It would be interesting to get reparametrized models, whose design matrices $\boldsymbol{Y}$ are full rank matrices, for all the restricted models because it would allow to apply some methods based on linear algebra and numerical analysis, like Newton Raphson method, to solve the corresponding likelihood equations. However, this is not always possible for all models. We can always get reparametrized models for vertex colouring; but, we can only get reparametrized models for some particular edge colouring models. As a consequence, we can get reparametrized models for only some vertex and edge colouring models.

If we have a vertex colouring, then we use the same procedure used for non-reparametrized models, but we have to define $\boldsymbol{Y}^{v_l^i(j)}$, l=1,2,..., $k(i)$, i=1, 2,..., T, for a category $j = 1, ..., J$, 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 available are any of the commonly used, for example, effect coding, which is an ANOVA type reparametrization, or those using dummy variables. The reason behind this procedure is that it can be shown that equating the reparametrized parameters is the same that equating the original parameters.

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.

If we have an edge colouring, then the procedure given above can not be applied in general. This is because the restrictions given to the reparametrized first order interaction parameters do not always coincide with the equivalent restrictions in the original parameters. In fact, we have less reparametrized parameters than original parameters, which means that there could be restrictions in the original parameters that can not be seen in the reparametrized parameters. However, there are some particular cases where we can obtain reparametrized models, for example, quasi-symmetry models.

Another example of an edge colouring model where we can get a reparametrized model is a model with restrictions

$$\lambda_{XY}(ij) = \lambda_{ZR}(ij),$$

for any i and j, where non reparametrized parameters are represented by λ. Every restriction forms a different colour class. Additionally, if we had different parameters, λ_{RS}, then we would have the same kind of restrictions or we would have $\lambda_{RS}(ij)$ parameters free to take any value.

In this kind of colouring it is also possible to equate the reparametrized parameters to get the restrictions given for the original parameters. As in the vertex colouring, we can get the design matrix by summing the columns corresponding to the reparametrized parameters in the same class. This means that 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.

For models with edge and vertex colouring, it is not always possible to have a reparametrized model. However, for particular cases like the symmetry models, reparametrized models can be obtained. Another example is a model with restrictions of the form $\lambda_{XY}(ij) = \lambda_{ZR}(ij)$ adding whichever colouring to the vertices.

Example 11. Easy examples to show reparametrized and non-reparametrized restricted models and their associated design matrices are symmetry and quasi-symmetry models. These are particular models where we can get the reparametrized models, although we have a vertex and edge colouring.

Suppose we have the same symmetry and quasi-symmetry models shown in section **7** for the binary variables C and A. We obtain the following matrix for the corresponding quasi-symmetry model

$$\boldsymbol{X} = \begin{array}{c} \\ 00 \\ 01 \\ 10 \\ 11 \end{array} \left[\begin{array}{cccccccc} \lambda & \lambda_C(0) & \lambda_C(1) & \lambda_A(0) & \lambda_A(1) & \lambda_{CA}(00) & \lambda_{CA}(01) = \lambda_{CA}(10) & \lambda_{CA}(11) \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \end{array} \right].$$

The first column corresponds to the constant term λ, the next two columns correspond to the parameters $\lambda_C(i)$, $i = 0, 1$, the following two columns correspond to $\lambda_A(i)$, $i = 0, 1$, and the last three columns correspond 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 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 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

$$\boldsymbol{Y} = \begin{array}{c} \\ 00 \\ 01 \\ 10 \\ 11 \end{array} \left[\begin{array}{cccc} u & u_C(0) & u_A(0) & \lambda_{CA}(01) = \lambda_{CA}(10) \\ 1 & 1 & 1 & 0 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 0 \end{array} \right].$$

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 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 $\boldsymbol{Y}$, and eliminating redundant equations, we can obtain the likelihood equations based on the equation systems $\boldsymbol{X}'n = \boldsymbol{X}'m$ or $\boldsymbol{Y}'n = \boldsymbol{Y}'m$, respectively, corresponding to the following equations.

$$\begin{aligned} \hat{m}(i,.) &= n(i,.), \ \ i = 0,1; \\ \hat{m}(.,i) &= n(.,i), \ \ i = 0,1; \\ \hat{m}(i,j) + \hat{m}(j,i) &= n(i,j) + n(j,i), \ \ i \leq j, i, j = 0,1; \end{aligned}$$

which are the likelihood equations obtained for a quasi-symmetry model as presented for example in Agresti, 2002, p. 425.

In a similar way, we can get the matrix for the corresponding symmetry model

$$\boldsymbol{X} = \begin{array}{c} \\ 00 \\ 01 \\ 10 \\ 11 \end{array} \left[\begin{array}{cccccc} \lambda & \lambda_1(0) & \lambda_1(1) & \lambda_{CA}(00) & \lambda_{CA}(01) = \lambda_{CA}(10) & \lambda_{AC}(11) \\ 1 & 2 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 2 & 0 & 0 & 1 \end{array} \right].$$

As before, the first column corresponds to λ. The second column represents the $\lambda_1(0)$ parameter, this 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

$$
\boldsymbol{Y} = \begin{array}{c} \\ 00 \\ 01 \\ 10 \\ 11 \end{array} \left[\begin{array}{ccc} u & u_1(0) & \lambda_{CA}(01) = \lambda_{CA}(10) \\ 1 & 2 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -2 & 0 \end{array} \right].
$$

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, 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

$$
\begin{aligned}
\hat{m}(i,i) &= n(i,i), \ i = 0, 1; \\
\hat{m}(i,j) + \hat{m}(j,i) &= n(i,j) + n(j,i), \ i < j, i, j = 0, 1;
\end{aligned}
$$

which are the likelihood equations obtained for a symmetry model as presented for example in Agresti, 2002, p. 424.

11 Solution to the likelihood equations

The likelihood equations for restricted graphical log-linear models can sometimes be solved using a closed expression, for example, for a symmetry model; however, in general, numerical methods are used to get an approximated solution. There are two commonly used methods to fit log-linear models: iterative proportional fitting, IPF, (Bishop *et al.*, 1975, p. 83-102, Christensen, 1990, p. 78-81, Lauritzen, 1996, p. 83-84, and Fienberg *et al.*, 1976) and Newton Raphson method (Agresti, 2002, p. 143-145, p. 342-343 and Christensen, 1990, p. 216-217).

We need to adapt these methods to solve the likelihood equations corresponding to restricted graphical log-linear models. The Newton Raphson method is only used when it is possible to have a reparametrized model, which means that a full rank matrix is associated to the model. The Newton Raphson method for restricted models is the same method presented in categorical data analysis books, for example in Agresti, 2002, p. 342-343, but using the matriz $\boldsymbol{Y}$ associated to the reparametrized model including the colouring, instead of the commonly used design matrix associated to the model that does not include any restrictions or colourings. 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. One advantage of this method is that we can approximate variances and covariances for the parameters.

A method that can always be used is a modification of the IPF method. The commonly used IPF method (Lauritzen, 1996, p. 83) consists of the following steps applied to every cell i, $i \in \mathrm{I}$:

1. We assign a value to $m_0(i)$, $i \in \mathrm{I}$. $m_0(i) = 1$, for example.
2. We take all the elements in the generating class A, and we order them as desired in a set $(b_1, b_2, ..., b_k)$. We define $T_v = T_{b_v}$, $v = 1, 2, ..., k$, as

$$(T_{b_v} m)(i) = m(i) \frac{n(i_{b_v})}{m(i_{b_v})}, \ i \in \mathrm{I}.$$

   We recursively define

$$m_{r+1}(i) = (T_1 T_2 ... T_k) m_r(i), \ r = 0, 1, 2, ...$$

   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 ($n_a(i_a) = m_a(i_a)$, $\forall a \in A$). 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 pre determinated error δ.

For restricted graphical log-linear models, there are some likelihood equations that have to be solved additional to the equations obtained for the elements in the generating class A. These additional equations correspond to the first order interactions and main effects. Therefore, we have to add other transformations. Supposing that all variables have the same total categories J, we define for every vertex colouring V_k, $k = 1, 2, ..., T$, the following transformations

$$(T_{V_k(1)} m)(i) = m(i) \frac{\sum_{j \in V_k} n(v_j^k = 1)}{\sum_{j \in V_k} m(v_j^k = 1)},$$
$$(T_{V_k(2)} m)(i) = m(i) \frac{\sum_{j \in V_k} n(v_j^k = 2)}{\sum_{j \in V_k} m(v_j^k = 2)},$$

$$\vdots$$

$$(T_{V_k(J)}m)(i) = m(i)\frac{\sum_{j\in V_k} n(v_j^k = J)}{\sum_{j\in V_k} m(v_j^k = J)}.$$

For the edge colouring, E_l, with $l = 1, 2, ..., S$, we define the following transformations

$$(T_{E_l}m)(i) = m(i)\frac{\sum_{\{l_v=i_v, r_v=j_v\}\in E_l} n(l_v = i_v, r_v = j_v)}{\sum_{\{l_v=i_v, r_v=j_v\}\in E_l} m(l_v = i_v, r_v = j_v)}.$$

The algorithm to solve restricted graphical log-linear models is similar to the one explained; however, we have to add on step 2 these last transformations. This is, we have to get for $r = 0, 1, 2, ...$

$$m_{r+1}(i) = (T_1T_2...T_kT_{V_1(1)}T_{V_1(2)}...T_{V_1(J)}...T_{V_T(1)}T_{V_T(2)}...T_{V_T(J)}T_{E_1}T_{E_2}....T_{E_S})m_r(i)$$

We have to consider that $T_{b_i} = T_i$ 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 $T_{V_r(j)}$.

In order to fit restricted graphical models, a computer program is needed. We have done a program and temporally called it REGRAPH. It fits the model using the modified IPF method explained above, this is because this is the method that can always be used. If it were possible to use the Newton Raphson Method, then the modified IPF method could be used to get the corresponding starting values. REGRAPH is written in Fortran 6.5, using subroutines from Haberman (1972) to fit log-linear models. For a specific model, the program calculates the fitted expected frequencies, the deviance, the non-reparametrized associated design matrix, and the degrees of freedom corresponding to the associated asymptotic chi square distribution, which are obtained using the design matrix. Notice that in some particular cases, for example quasi-symmetry models, exact distributions have been calculated (Booth *et al.*, 2003).

Numerical results obtained with REGRAPH for various models have been compared with those obtained by using other software like MIM or Splus. These models are log-linear models without restrictions, including graphical models, and symmetry and quasi-symmetry models. In REGRAPH, the user has to give the data, including how many variables and categories each variable has, the generating class, and the vertex and edges colour classes of the model he wants to fit. The generating class has to be known, if we do not know it, we could use a model search method to look for a graphical model that fits the data. This process can be done in any available software for graphical models, for example, MIM.

12 Discussion and perspectives

When one is interested in searching colour classes, one would like to have a model search method that helps to get the better colouring that fits the data. To do this, we are working in methods that join colour classes, and using the deviance, tell us if it is convenient to join the classes. These methods could be used to get a model search method, using any initial vertex and edge colour class, for example a colouring in which every vertex and every edge forms an atomic class. We want a process that joins iteratively classes whose joint is statistically significant until we get a restricted graphical model which better fits the data, in the sense that using the deviance difference we do not reject the null hypothesis that the parameters corresponding to the same colour class in the model without joining classes are equal.

We observe that in order to join vertex colour classes, it is necessary that all variables have the same number of categories as in symmetry and quasi-symmetry models. The number of models that we are eliminating by considering this are not so many, they are the models where the vertex in atomic classes have different number of categories. Because of these reasons, and also in order to simplify the programming, we are supposing that all variables have the same number of categories, *i.e.* we have square tables. We are also supposing that we do not have structural zeros, i.e $p(i) > 0$ or $m(i) > 0$ if the cells counts are realizations of Poisson or multinomial distributed random variables, respectively, although we could have sampling zeros. In this last case 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 by taking the degrees of freedom minus the number of parameters that can not be estimated minus the number of cells estimated as zeros.

It is important to notice that, depending on the model, we could determinate 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 for some data many colourings are possible.

Currently, we are working on REGRAPH in order to get a program easier to use giving more information to the user. Also, we are working with the program that joins edge colour classes. The program corresponding to vertex colour classes has already been done. At the same time, we are working with the model search algorithms, which are going to be based on some parts of the mentioned join colour classes methods. All these programs are being made in Fortran, but it will be necessary to use some graphical interface, done in Fortran or in another program, which will give the graphs corresponding to the fitted models.

Additionally, we are considering to analyze a particular simpler subset of these models, the models with restrictions of the type $\lambda_{XY}(ij) = \lambda_{ZR}(ij)$, for any i and j, and any vertex colouring. These models could have a less complicated graph representation and interpretation. For these particular models, a more interactive program could be made whose graphs are easier to understand for any user.

13 Basic References

Agresti, A. (2002). *Categorical data analysis*, 2a ed. New York: John Wiley and Sons.

Andersen, E. (1991). *The statistical analysis of categorical data.* Heidelberg: Springer-Verlag.

Bishop, Y. M. M., Fienberg, S. E., and Holand P. W. (1975). *Discrete multivariate analysis.* Cambridge, Massachusetts: MIT Press.

Bondy J. A. and Murty, U. S. R. (1976). *Graph theory with applications.* London: MacMillan Press.

Booth, J. G., Capanu, M., and Heigenhauser, L. (2003). Exact conditional p-value calculation for the quasi-symmetry model. Research Report. University of Florida.

Caussinus, H. (1965). Contribution à l′analyse statistique des tableux de corrélation. *Annales de la Faculté des Sciences de Toulouses* **29**, 77-182.

Christensen, R. (1990). *Log-linear models.* New York: Springer-Verlag.

Edwards, D. (2000). *Introduction to graphical modelling*, 2a ed. New York: Springer-Verlag.

Haberman, S. J. (1972). Algorithm AS 51: log-linear fit for contingency tables. *Applied Statistics* **21** (2), 218-225.

Haberman, S. J. (1978). *Analysis of qualitative data, Vol. 1: Introduction topics.* New York: Academic Press.

Haberman, S. J. (1979). *Analysis of qualitative data, Vol. 2: New developments.* New York: Academic Press.

Højsgaard, S. and Lauritzen, S. (2005). Restricted concentration models – graphical Gaussian models with concentration parameters restricted to being equal. *AISTATS 2005.* Published by Society for Artificial Intelligence and Statistics. Available in `http://www.gatsby.ucl.ac.uk/aistats/`.

Højsgaard S. and Lauritzen, S. (2007a). Graphical Gaussian models with edge and vertex symmetries. Work available in `http:// www.stats.ox.ac.uk/ ~steffen/ papers/ index.htm`.

Højsgaard, S. and Lauritzen, S. L. (2007b). Inference in graphical Gaussian models with edge and vertex symmetries with the gRc package for R. *Journal of Statistical Software* **23** (6). Available in `http://www.stats.ox.ac.uk/ ~steffen/ papers/index.htm`.

Lauritzen, S. L. (1989). *Lectures on contingency tables*, 3a ed. Aalborg, Denmark: University of Aalborg Press. Electronic version available in `http:// www.math.auc.dk/ ~ steffen/ cont.pdf`.

Lauritzen, S. L. (1996). *Graphical models.* Oxford, England: Clarendon Press.

Whittaker, J. (1990). *Graphical models in applied multivariate statistics.* Chichester, England: John Wiley and Sons.