Scandinavian Journal of Statistics

Theory and Applications

Scandinavian Journal of Statistics, Vol. 38: 447–465, 2011
doi: 10.1111/j.1467-9469.2010.00713.x

Quasi-Symmetric Graphical Log-Linear Models

ANNA GOTTARD and GIOVANNI MARIA MARCHETTI

Department of Statistics, University of Florence

ALAN AGRESTI

Department of Statistics, University of Florida

ABSTRACT. We propose an extension of graphical log-linear models to allow for symmetry constraints on some interaction parameters that represent homologous factors. The conditional independence structure of such quasi-symmetric (QS) graphical models is described by an undirected graph with coloured edges, in which a particular colour corresponds to a set of equality constraints on a set of parameters. Unlike standard QS models, the proposed models apply with contingency tables for which only some variables or sets of the variables have the same categories. We study the graphical properties of such models, including conditions for decomposition of model parameters and of maximum likelihood estimates.

Key words: conditional independence, decomposition, exchangeability, graphical models, homologous variables

1. Introduction

Log-linear models are useful in many fields, such as for describing multivariate association structure among response variables measured in sample surveys for social science research. Graphical log-linear models, described by Darroch *et al.* (1980), are a subclass of hierarchical log-linear models. Each such model corresponds to an undirected graph in which each variable is represented by a node and the absence of an edge connecting nodes represents conditional independence. In the log-linear model formula, interaction terms are set to zero according to the edges that are missing in the graph.

In some applications, many or all the response variables are measured with the same categorical scale; that is, they are homologous. This is common in medical longitudinal studies that repeatedly observe whether some condition is present at various times, or attitudinal research that observes subjects' opinions about some issue under a variety of conditions. In such cases, the contingency table that cross-classifies the response variables has specialized structure, and certain models can be useful when we expect the joint distribution to exhibit a structure that is exchangeable in certain aspects. An example is the standard quasi-symmetric (QS) structure by which the single-factor marginal distributions may differ but the two-way and higher-order terms in the log-linear model have a symmetric form (Caussinus, 1966; Bishop *et al.*, 1975).

To illustrate, we present and later analyse two data sets from the US *General Social Survey*. The first presents attitudes about legalized abortion and about the death penalty. The data, displayed in Table 1, consists of four binary variables observed in 2002, 2004 and 2006. Three variables provide responses to whether abortions should be legal: when there is a strong chance of a serious defect in the baby (D), when the woman's health is seriously endangered (H) and when the pregnancy is the result of a rape (R). The fourth variable is whether a subject favours the death penalty for people convicted of murder (P). All variables use the same scale, (yes, no). It is plausible that exchangeability could occur among D, H and R,

Table 1. *Data on attitudes about legalized abortion in three cases (D: defect in the baby, H: woman's health problems, R: rape) and whether favour the death penalty (P)*

		P	Yes		No	
D	H	R	Yes	No	Yes	No
Yes	Yes		1590	105	634	60
	No		12	9	10	10
No	Yes		159	128	66	93
	No		30	172	9	131

but there is no reason to expect exchangeability of these three items with P except for the fact that this item also deals with potential taking of a human life.

The second data set relates to the degree of satisfaction with some aspects of US government policy in 2006. The data, shown in Table 2, contain homologous variables on how successful US government policy has been on protecting the environment (E), fighting unemployment (U) and providing a decent standard of living for the old (O), and a non-homologous variable, gender (G). The three homologous variables have levels s, successful; n, neither successful nor unsuccessful; and u, unsuccessful.

For such data sets, questions of interest include the following: Do subsets of the variables have exchangeable structure for interactions? When certain sets of variables form cliques that have interaction structure simpler than the usual one for graphical models, do standard results from graphical models still apply about properties such as decomposability and collapsibility?

In this article, we introduce a new class of graphical log-linear models, called *QS graphical models*, in which certain interaction parameters are restricted to be identical for sets of variables. Different sets of constraints have different graphical configurations, with the standard QS model being a particular case. To depict the QS constraints in the graph, we use coloured edges, adapting the idea of Højsgaard & Lauritzen (2008) of using colours to represent symmetries in the association form. QS graphical models extend ordinary QS models and other types of generalized symmetry models, such as a subclass for hypercubic contingency tables proposed by Lovison (2000).

After introducing graphical QS models, we study their graphical properties. In particular, we investigate the consequences on decomposition and collapsibility of imposing symmetry constraints and equality constraints on certain interaction terms. Such properties are useful for reducing the complexity of a model and for facilitating interpretation. We will see that, as in ordinary models, separation implies parametric collapsibility, in terms of certain parameters remaining constant when tables are collapsed in certain ways. However, a new,

Table 2. *Data on degree of satisfaction with government policy*

		O	s			n			u		
E	G	U	s	n	u	s	n	u	s	n	u
s	m		58	22	13	26	30	7	39	13	19
	f		52	21	17	21	31	7	28	29	32
n	m		18	12	8	27	31	14	10	22	34
	f		9	13	11	26	55	23	13	39	39
u	m		13	12	12	14	25	33	17	29	112
	f		14	12	10	12	32	24	21	38	112

coloured, version of decomposition implies collapsibility of the estimates from maximum likelihood (ML) inference.

The article is organized as follows. Section 2 defines QS graphical models. Section 3 discusses model fitting and special issues that arise in the comparison of coloured graphical models in model selection. In section 4, we describe characteristics and properties of coloured graphical models. Section 5 uses coloured graphical models to analyse the data introduced before. Section 6 discusses possible extensions and makes some concluding remarks.

2. Graphical QS models

2.1. Graphical log-linear models

Let $X_1, X_2, \dots, X_d$ be discrete random variables with X_v taking values in $[r_v]=\{1,\dots,r_v\}$, for $v \in V=\{1,\dots,d\}$. Let $\mathcal{I}=\prod_{v\in V}[r_v]$, be the d-dimensional table of cells from cross-classifying the variables, with $\imath=(i_1,\dots,i_d)$ denoting a generic cell in the table. The probabilities

$$p(\imath)=P(X_1=i_1,\dots,X_d=i_d), \quad \imath=(i_1,\dots,i_d)\in\mathcal{I}$$

specify the joint distribution of $(X_1, X_2, \dots, X_d)$. We shall assume that each $p(\imath)$ is strictly positive. Under multinomial sampling with n observations, let $n(\imath)$ be the observed count in cell $\imath$, with $\mu(\imath)=np(\imath)$ its expected value. For each subset a of V, let $\mathcal{I}_a=\prod_{v\in A}[r_v]$ be a marginal table with $\imath_a$ denoting a corresponding marginal cell. The joint probability table admits a log-linear expansion

$$\log\mu(\imath)=\sum_{a\subseteq V}\lambda^a(\imath_a), \tag{1}$$

where $\lambda^a(\imath_a)$ is a function, defining the log-linear parameters, indexed by the subset a of V. There is a one-to-one and smooth transformation between the joint probabilities p and the log-linear parameters λ, and model (1) is the saturated log-linear model. The non-zero parameters $\lambda^a(\imath_a)$ are commonly called log-linear interactions of order $|a|$. To ensure identifiability we shall adopt sum-to-zero constraints, whereby the sum over values of any index for a λ term equals 0. For many models, one could equivalently use reference-level coding, such as $\lambda^a(\imath_a)=0$ whenever at least one index in $\imath_a$ is equal to 1. However, for certain models in the class discussed in this article, the use of such coding results in a non-equivalent model having unnatural constraints in patterns of association. We discuss this further in section 2.4.

Submodels of the saturated log-linear model are obtained by deleting some $\lambda^a(\imath_a)$ or imposing equality constraints on them. An especially useful class is that of *graphical log-linear models*, in which the terms deleted are dictated by the structure of an undirected graph $G=(V,E)$, defined over the set V of nodes and by a set of edges E. The appendix presents a short summary of graph theory.

In the theory of graphical models, the nodes index the random vector $X=(X_v\mid v\in V)$ while absence of an edge $vw=(v,w)$ in E implies conditional independence between X_v and X_w, given the other variables. We summarize graphical log-linear models briefly here, referring to Lauritzen (1996) for further details. Given an undirected graph $G=(V,E)$, a graphical log-linear model associated with G is defined by the set of strictly positive discrete probability distributions with log-linear expansion (1) having constraints

$$\lambda^a(\imath_a)=0 \text{ whenever } a \text{ is a subset of nodes of } G \text{ that is not complete,} \tag{2}$$

in the sense that not all pairs of nodes in a are joined by an edge. It can be proved (Darroch *et al.*, 1980; Lauritzen, 1996) that if p has the form given by (1) and (2), then p satisfies

$$X_v \perp\!\!\!\perp X_w \mid X_{V\setminus\{v,w\}}, \qquad \text{for all } vw \notin E. \tag{3}$$

This is called the pairwise Markov property. In words, the variables corresponding to each missing edge in the graph are conditionally independent given the remaining variables. Identifying the generating class of a graphical log-linear model $\mathcal{M}(G)$ defined by constraints (2) by the list of the higher-order interaction parameters, it can be verified that this class is formed by the cliques of the graph G. We next present a simple example of a graphical log-linear model.

Example 1. For a three-way contingency table with a generic cell having indices (i,j,k), the saturated log-linear model (using superscripts to identify the variables) is:

$$\log \mu_{ijk} = \lambda + \lambda_i^1 + \lambda_j^2 + \lambda_k^3 + \lambda_{ij}^{12} + \lambda_{ik}^{13} + \lambda_{jk}^{23} + \lambda_{ijk}^{123}.$$

The saturated model corresponds to a complete graph with three nodes. For the graph obtained by deleting the edge 13 from the complete graph (see the graph later in this article in Fig. 3A), the associated graphical log-linear model satisfies $\lambda_{ijk}^{123}=0$ and $\lambda_{ik}^{13}=0$, for all i,j,k. The resulting model is equivalent to the conditional independence model, $X_1 \perp\!\!\!\perp X_3 \mid X_2$.

2.2. QS interaction structure

When some variables in vector X are homologous, log-linear models can add appropriate constraints for parameters, such as exchangeability for associations and interactions. This results in models that combine conditional independence between some pairs of variables with symmetry restrictions for the interaction terms within other sets of variables.

QS models are a subclass of log-linear models, specified for contingency tables with homologous variables, defined by equality constraints on higher-order interaction parameters. This class of models has been defined by Caussinus (1966), also by Bishop *et al.* (1975) for up to three-dimensional tables, and by Bhapkar & Darroch (1990) for tables of general order. With two homologous variables, the model has a symmetric interaction term. With three homologous variables (X_1, X_2, X_3) forming an $I \times I \times I$ contingency table, the ordinary QS model is:

$$\log \mu_{ijk} = \lambda + \lambda_i^1 + \lambda_j^2 + \lambda_k^3 + \lambda_{ij}^{12} + \lambda_{ik}^{13} + \lambda_{jk}^{23} + \lambda_{ijk}^{123}, \tag{4a}$$

$$\lambda_{ij}^{12} = \lambda_{ji}^{12} = \lambda_{ij}^{13} = \lambda_{ji}^{13} = \lambda_{ij}^{23} = \lambda_{ji}^{23}, \qquad \text{for all } i,j=1,\ldots,I, \tag{4b}$$

$$\lambda_{ijk}^{123} = \lambda_{\text{perm}(ijk)}^{123} \qquad \text{for all } i,j,k=1,\ldots,I, \tag{4c}$$

where perm(ijk) denotes any permutation of the set of indices in the argument. Unlike the complete symmetry model, this model does not impose restrictions on the single-factor terms, the consequence being that the observed and fitted values are identical in the one-dimensional margins of the table.

Restrictions such as (4b) and (4c) treat all variables alike, and do not allow only a subset of variables to be conditionally independent. In this article, we generalize this by taking a graphical log-linear model and then imposing QS restrictions on selected non-zero parameters involving homologous variables. The following examples present possible generalizations of the QS model. Each proposed model is represented by a graph, whose characteristics will be described in the next section.

Example 2. Suppose X_1, X_2 and X_3 are homologous variables and that $X_1 \perp\!\!\!\perp X_3 \mid X_2$. Then, we could consider an extended QS model

$$\log \mu_{ijk} = \lambda + \lambda_i^1 + \lambda_j^2 + \lambda_k^3 + \lambda_{ij}^{12} + \lambda_{jk}^{23}, \tag{5a}$$

$$\lambda_{ij}^{12} = \lambda_{ji}^{12} = \lambda_{ij}^{23} = \lambda_{ji}^{23} \quad \text{for all } i, j = 1, \dots, I. \tag{5b}$$

This model is represented by Fig. 1A, where the QS constraints are represented by colours in the edges.

Example 3. Consider four variables, X_i, $i = 1, \dots, 4$, such that the first three variables are homologous with I levels and X_4 is non-homologous with J levels. Consider the conditional independence model, $X_4 \perp\!\!\!\perp (X_1, X_2) \mid X_3$, specified by the graphical log-linear model

$$\log \mu_{ijkl} = \lambda + \lambda_i^1 + \lambda_j^2 + \lambda_k^3 + \lambda_l^4 + \lambda_{ij}^{12} + \lambda_{ik}^{13} + \lambda_{jk}^{23} + \lambda_{kl}^{34} + \lambda_{ijk}^{123}, \tag{6}$$

which has two cliques, [123][34]. A simplification of this model uses QS structure for the three homologous variables, with the constraints defined by (4b) and (4c). This model is represented by Fig. 1B.

Sometimes it is convenient to relax some of the restrictions of the standard QS model by removing the assumption of homogeneity of the interactions of the same order. The following example presents such a model.

Example 4. When the data are not compatible with the standard QS restrictions (4b) and (4c), less severe restrictions impose symmetry on the array of three-factor interactions and on the separate two-factor interactions, but without imposing homogeneity for the two-factor terms,

$$\lambda_{ij}^{12} = \lambda_{ji}^{12}, \quad \lambda_{ij}^{13} = \lambda_{ji}^{13}, \quad \lambda_{ij}^{23} = \lambda_{ji}^{23}, \quad \text{for all } i, j = 1, \dots, I, \tag{7a}$$

$$\lambda_{ijk}^{123} = \lambda_{\text{perm}(ijk)}^{123}, \quad \text{for all } i, j, k = 1, \dots, I. \tag{7b}$$

Then, each pair of variables is conditionally QS of different forms. This model is represented by Fig. 1C, in the context of another special case of model (6), but one that is more general than the model represented by Fig. 1B.

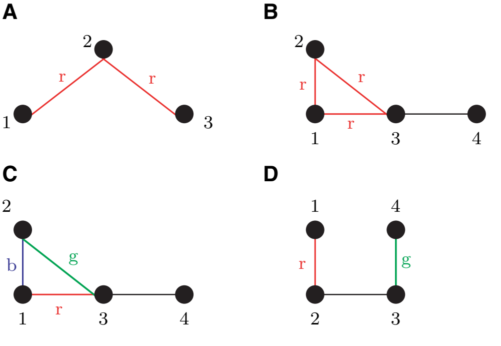


Fig. 1. Edge-coloured graphs for: (A) example 2, (B) example 3, (C) example 4 and (D) example 5.

Some applications have more than one subgroup of homologous variables. For such cases, there are other possible extensions of QS models.

Example 5. Consider a $I^2 \times J^2$ table for two pairs of homologous variables, X_1 and X_2 with I levels, and X_3 and X_4 with J levels. One possible log-linear graphical model is:

$$\log \mu_{ijkl} = \lambda + \lambda_i^1 + \lambda_j^2 + \lambda_k^3 + \lambda_l^4 + \lambda_{ij}^{12} + \lambda_{jk}^{23} + \lambda_{kl}^{34} \tag{8a}$$

with generators [12][23][34], specifying the conditional independencies

$$(X_1, X_2) \perp\!\!\!\perp X_4 \mid X_3, \quad (X_3, X_4) \perp\!\!\!\perp X_1 \mid X_2.$$

If each pair of homologous variables is QS, we can impose the restrictions

$$\lambda_{ij}^{12} = \lambda_{ji}^{12}, \ \ \lambda_{kl}^{34} = \lambda_{lk}^{34}, \qquad \text{for all } i,j = 1, \ldots, I, \ \ k,l = 1, \ldots, J. \tag{8b}$$

This model is represented by Fig. 1D.

The models in these examples belong to a special kind of graphical log-linear model defined by imposing both conditional independence constraints and equality and quasi-symmetry constraints on interaction parameters. Each model has a related undirected graph encoding the conditional independencies, but without some modification the usual graph does not portray all the details of the model. In fact, as we shall see, the presence of the equality constraints partly modifies the statistical properties of the graphical model. It is therefore convenient to enrich the interpretation by giving a special graphical code to edges affected by the QS constraints. We propose here to use the class of coloured graphs recently utilized by Højsgaard & Lauritzen (2008) for Gaussian undirected graph models with constraints on the parameters. The following section will give the general definition of this new class of graphical models and of their graphical representation.

2.3. QS graphical models

We define next a class of discrete graphical models with equality constraints on particular subsets of interaction parameters. The models apply to situations in which the variables can be partitioned into subsets of homologous variables and possibly a subset of non-homologous variables. This class of models is defined by a graph with coloured edges, where edges with the same colour correspond to interaction parameters constrained to be equal.

In this article, by a *coloured graph*, we mean a triplet $G = (V, E, \mathscr{E})$ where V is a set of nodes, E is a set of unordered pairs of nodes, $vw = (v, w), v, w \in V$ and $\mathscr{E}$ is a partition of E into disjoint *colour classes* $E_0, \ldots, E_s$, each collecting edges with the same colour. Conventionally, we assign black edges to E_0. The coloured edges could also be marked by specific symbols (such as the first letter of a colour name) to facilitate the reading in black-and-white prints. We define the *skeleton graph* G^* of a coloured graph G to be the undirected graph obtained by replacing all colours by black. Moreover, we call the *induced coloured subgraph* G_a^{col} of a coloured graph $G = (V, E, \mathscr{E})$, $a \subseteq V$, the subgraph induced by a after deleting all the black edges. We denote by col(a) the *set of colours* of the edges in G_a^{col}. Note that black is not included in the set of colours. Therefore, col(V) is the set of colours which are present in the graph G, whose cardinality is the number of colour classes minus 1.

Definition 1. *A QS graphical model with associated coloured graph* $G = (V, E, \mathscr{E})$ *is a log-linear model for the discrete joint probability distribution* $p(\imath)$, *with parameters* $\lambda^a(\imath_a)$ *for* $a \subseteq V$, *such that*:

(i) *the model satisfies the conditional independence constraints of the graphical model associated with the skeleton graph G^*, that is, $\lambda^a(\imath_a)=0$ for all subsets not complete in G^*;*

(ii) *for each edge vw in colour class E_k with $k \neq 0$,*

$$\lambda_{ij}^{vw}=\lambda_{ji}^{vw}, \quad \text{for all } i,j=1,\ldots,I$$

and all non-zero higher-factor parameters $\lambda^a(\imath_a)$ involving both v and w, with $b=a\setminus\{v,w\}$ satisfy the QS equality constraints

$$\lambda^a(i_v,i_w,i_b)=\lambda^a(i_w,i_v,i_b);$$

(iii) *For all complete subsets $C_1,\ldots,C_{k(s)}$ in $s\geq 2$ nodes with the same edge colour, the interaction parameters of order s satisfy the equality constraints*

$$\lambda^{C_1}(\imath_{C_1})=\cdots=\lambda^{C_{k(s)}}(\imath_{C_{k(s)}}).$$

Therefore, equality constraints are imposed for all interactions associated with complete subsets with the same colour in $s=2,3,\ldots$ nodes.

To achieve identifiability, model parameters $\lambda^a(\imath_a)$ are constrained to sum to 0 *when summed over any index.*

Variables associated with a coloured edge should be homologous, to have sensible QS constraints. Thus, if the set of nodes V is partitioned into subsets $V_0,V_1,\ldots,V_m$, where $V_1,\ldots,V_m$ refer each to a set of homologous variables, and V_0 refers to non-homologous variables, an edge vw may be coloured (i.e. not black) only if v,w are in the same class V_r for some $r\neq 0$. In addition, note that a black edge in a coloured graph denotes a set of interaction parameters with no constraints. For this reason, we separate the class E_0 of the black edges from the other colour classes and do not consider black as a colour.

Notice that in QS graphical models the colour of edges has a double meaning, representing both internal and external constraints. The internal constraints are symmetry constraints within an interaction parameter [see definition 1(ii)]. The external constraints are equality constraints, among sets of interaction parameters [see definition 1(iii)]. In Højsgaard & Lauritzen (2008), instead, colours of edges imply exclusively external constraints, that is, equality between parameters, and each colour has to be shared by at least two edges. This difference of semantics reflects the fact that association parameters in Gaussian graphical models cannot be asymmetric.

We next illustrate definition 1 by examples, some of which continue those introduced in the previous section.

Example 6 (example 2 continued). Consider the graphical QS model associated with the graph of Fig. 1A, with set of edges $E=\{12,23\}$. By definition 1(i), it satisfies the conditional independence $X_1 \perp\!\!\!\perp X_3 \mid X_2$, because of the missing edge 23. The two coloured edges 12 and 23 imply, by definition 1(ii), separate QS constraints on the two-factor interaction terms: $\lambda_{ij}^{12}=\lambda_{ji}^{12}$ and $\lambda_{ij}^{23}=\lambda_{ji}^{23}$. Moreover, the fact that the two edges have the same colour implies, by definition 1(iii), that such interactions are equal, thus giving the constraints in (5b).

Example 7. Consider an edge-coloured graph corresponding to a QS graphical model, having a complete subgraph in three nodes with edges 12 and 23 coloured (i.e. not black). Then, the associated QS graphical model has the constraints $\lambda_{ij}^{12}=\lambda_{ji}^{12}$ and $\lambda_{ij}^{23}=\lambda_{ji}^{23}$. By the second part of definition 1(ii), the three-factor terms must then satisfy

$$\lambda^{123}_{ijk} = \lambda^{123}_{jik} \quad \text{and} \quad \lambda^{123}_{ijk} = \lambda^{123}_{ikj}.$$

However, these two sets of constraints imply also $\lambda_{ijk} = \lambda_{jki}$ and $\lambda_{ijk} = \lambda_{kji}$. Thus, the array λ_{ijk} is symmetric with respect to all possible permutations: $\lambda_{ijk} = \lambda_{\mathrm{perm}(ijk)}$. This result suggests proposition 1.

Proposition 1. *Let $\mathcal{M}^{\mathrm{QS}}(G)$ be a QS graphical model associated with a coloured graph G. Let G_c be a complete subgraph of G, with $c \subseteq V$ and $|c| > 1$. Let G^{col}_c be the induced coloured subgraph of G_c obtained by deleting all the black edges. Then, if G^{col}_c is connected, the QS graphical model $\mathcal{M}^{\mathrm{QS}}(G)$ contains the parameters $\lambda^c(\imath_c)$ and this term is fully symmetric with respect to all the indices, that is, $\lambda^c(\imath_c) = \lambda^c(\mathrm{perm}(\imath_c))$.*

Proof. Because of definition 1(ii), every QS constraint corresponds to a transposition to the indices of the interaction term $\lambda^c(\imath_c)$. As G^{col}_c is connected, every node in c is connected to at least another node in c by a coloured edge, so that there is at least one QS constraint imposed on $\lambda^c(\imath_c)$ for each index in $\imath_c$. Thus, each possible permutation of the set $\imath_c$ can be generated by applying at least one sequence of transpositions to the indices. Therefore, the entire set of QS constraints is $\lambda^c(\imath_c) = \lambda^c(\mathrm{perm}(\imath_c))$.

Notice that the equality $\lambda^c(\imath_c) = \lambda^c(\mathrm{perm}(\imath_c))$ is trivially true whenever G_c is not complete, as all the interaction parameters vanish.

Example 8 (example 3 continued). The QS graphical model for this example is associated with the graph portrayed in Fig. 1B. Here the set of nodes is $V = \{1, 2, 3, 4\}$, and, in the corresponding skeleton graph, node 3 separates node 4 from nodes 1 and 2. As a consequence, by definition 1(i), $X_4 \perp\!\!\!\perp (X_1, X_2) \mid X_3$. The set of edges $\mathscr{E}$ is partitioned into two colour classes, $\{E_0, E_1\}$, with $E_0 = \{34\}$ and $E_1 = \{12, 13, 23\}$. According to definition 1(ii), QS constraints are imposed on the two-factor interaction terms, while by definition 1(iii), $\lambda^{12}_{ij} = \lambda^{12}_{ji} = \lambda^{13}_{ij} = \lambda^{13}_{ji} = \lambda^{23}_{ij} = \lambda^{23}_{ji}$. Moreover, proposition 1 implies that $\lambda^{123}_{ijk} = \lambda^{123}_{\mathrm{perm}(ijk)}$, as in (7b). The black edge 34 indicates that the corresponding interaction terms are not involved in any QS constraint and are set free.

Example 9 (example 4 continued). The QS graphical model illustrated in this example, in Fig. 1C, has the same skeleton graph as the previous model. Consequently by definition 1(i) the two models have the same conditional independence structure. The set of edges $\mathscr{E}$ is partitioned into four colour classes, with $E_0 = \{34\}$, $E_1 = \{12\}$, $E_2 = \{13\}$ and $E_3 = \{23\}$, so that each edge has a different colour. By definition 1(ii), QS constraints are separately imposed on the two-factor interaction terms $\lambda^a_{ij} = \lambda^a_{ji}$, with $a = 12, 13, 23$. The colours on these three edges act also on the three-factor terms, which satisfy the constraints in (4c) because of proposition 1.

Note that this situation, with colour classes containing only single edges, is effective only when $I > 2$. This agrees with ordinary QS models being of interest for $I \times I$ tables only when $I > 2$ (Bishop *et al.*, 1975, p. 281).

Example 10 (example 5 continued). The model presented for this case is represented by the graph in Fig. 1D. This graph has $\mathscr{E} = \{E_0, E_1, E_2\}$, with $E_0 = \{23\}$, $E_1 = \{12\}$ and $E_2 = \{34\}$, corresponding to the constraints in (8b).

2.4. Parameter coding for identifiability and sensible models

In fitting standard log-linear models (i.e. without equality constraints), to ensure identifiability it is common in the model matrix to use sum-to-zero constraints on the parameters or else reference-level coding by which a particular level of each factor (usually, either the first level or the last level) is the reference category and has value 0 for parameters at that index level. The choice is unimportant, because the model-fitted values and the estimates of relevant parameters (such as odds ratios) are the same for each choice. For QS graphical models, however, in definition 1 we specified that the parameters satisfy sum-to-zero constraints. This is because for certain models, different forms of reference coding do not satisfy a certain invariance property and do not yield equivalent models to those with sum-to-zero constraints but rather can result in non-sensible models.

Specifically, by the nature of the equality constraints on the parameters in QS graphical models and the nature of the sum-to-zero constraints for identifiability, any QS graphical model is invariant to permutations of categories for all variables that are homologous. This is a basic property of any sensible model for nominal-scale variables (regardless of whether variables are homologous), and it applies for any of the standard coding methods for ordinary log-linear models and for standard QS models for a set of variables in one clique for which all edges have the same colour. However, for certain models having the same colour in different cliques or a common colour for only some of the variables in a clique, this property does not hold with reference-level coding. For this reason, definition 1 for QS graphical models does not permit such coding, as in some cases the entire meaning of the model would then depend on the parameter coding.

To illustrate, consider the model for a four-way cross-classification of binary variables,

$$\log \mu_{hijk} = \lambda + \lambda_h^1 + \lambda_i^2 + \lambda_j^3 + \lambda_k^4 + \lambda_{hi}^{12} + \lambda_{hj}^{13} + \lambda_{ij}^{23} + \lambda_{hij}^{123} + \lambda_{jk}^{34},$$

in which all edges have the same colour. (See model 3 in Table 3 for the graph of this case, considered in an example discussed later.) In particular, $\lambda^{12}_{\mathrm{perm}(ab)} = \lambda^{13}_{\mathrm{perm}(ab)} = \lambda^{23}_{\mathrm{perm}(ab)} = \lambda^{34}_{\mathrm{perm}(ab)}$, and we denote its common value by $\lambda_{\min(a,\,b),\,\max(a,\,b)}$. The conditional log odds ratio between variables 3 and 4 at levels h and i of variables 1 and 2 equals $(\lambda_{11} + \lambda_{22} - 2\lambda_{12})$. This equals $4\lambda_{11}$ with zero-sum coding and λ_{11} with (1, 0) reference-level coding that sets parameters equal to 0 at the second level of a variable. We now compare this with the conditional log odds ratio between variables 1 and 2 at levels j and k of variables 3 and 4, which equals

$$\log \left[\frac{\mu_{11jk}\mu_{22jk}}{\mu_{12jk}\mu_{21jk}} \right] = (\lambda_{11} + \lambda_{22} - 2\lambda_{12}) + (\lambda_{11j} - 2\lambda_{12j} + \lambda_{22j}).$$

With zero-sum coding, this equals $4(\lambda_{11} + \lambda_{11j})$, which is $4(\lambda_{11} + \lambda_{111})$ when $j = 1$ and $4(\lambda_{11} - \lambda_{111})$ when $j = 2$. In contrast, with (1, 0) reference-level coding, this equals $(\lambda_{11} + \lambda_{111})$ when $j = 1$ and λ_{11} when $j = 2$. That is, this coding constrains the conditional log odds ratio between variables 1 and 2 to equal the log odds ratio between variables 3 and 4 when $j = 2$ but not when $j = 1$. Similarly, the (0, 1) reference-level coding constrains the conditional log odds ratio between variables 1 and 2 to equal the log odds ratio between variables 3 and 4 when $j = 1$ but not when $j = 2$. Because of this, the fit and statistical significance of various terms then depends on the coding used and is not invariant to a like permutation of categories for all the variables.

In practice, the use of a common colour with different cliques would normally be considered mainly when cliques have the same case, such as the chain graph in Fig. 1A. However, this example shows that in more complex models, using reference-level coding can impose constraints under which the model would not be considered scientifically sensible. Another such example is the model with generators [123][234] and with all edges having the same colour, in which the 23 relationship is involved in two higher-order terms. Using reference-level coding constrains some of the 23 conditional odds ratios to equal 12 and 34 odds ratios and some to differ, and the ones that are equal differ according to which category is the reference category. In contrast, in each of these models the sum-to-zero constraints impose a sensible relationship among the odds ratios and the fit is invariant when categories of all variables are permuted in the same way.

It is an interesting question for future research to determine the class of QS graphical models that, when instead used with reference-level coding, would have identical fit to the QS graphical model as we have defined it with sum-to-zero constraints. One case in which it does seem to happen is for a restricted class of such models in which all the a-factor interactions with $|a|>2$ equal zero.

3. Fitting and selecting QS graphical models

3.1. Model fitting

In general no closed-form expression of the ML estimates can be found for QS graphical models. Nevertheless, ML estimates can be easily computed by any package for generalized linear models, by using a log link function assuming independent Poisson counts and modifying the model matrix to represent the assumed parameter constraints. Both analyses in section 5 were performed using the software R (R Development Core Team, 2009).

In practice, the model matrix for a certain QS graphical model $\mathcal{M}^{\mathrm{QS}}(G)$ can be constructed by suitably simplifying the model matrix for the saturated unconstrained model, to satisfy the elements of definition 1. The saturated model is an ordinary log-linear associated to a skeleton (black) complete graph and can be written as $\log\mu(\iota)=Z\lambda$, where Z is the model matrix and λ is the set of log-linear parameters. Columns corresponding to zero parameters because of missing edges in the skeleton graph G^* must be deleted [definition 1(i)]. Columns corresponding to parameters constrained to be equal, which correspond to coloured edges in G, must be summed to obtain new columns [definition 1(ii)]. Columns corresponding to parameters that refer to edges sharing the same colour must be summed to obtain new columns [definition 1(iii)]. The resulting model can be then written as $\log\mu(\iota)=X\lambda=ZQ\lambda$, in which Q is a matrix operator summing up columns of Z according to parameter constraints.

The rank of the resulting model matrix X corresponds to the number of distinct parameters in the model. Similarly as for ordinary log-linear models, the likelihood equations equate the observed counts to the expected counts, through $\mu'X=n'X$, or, equivalently, $\mu'ZQ=n'ZQ$. The deviance has a large sample-approximate chi-squared distribution with degrees of freedom equal to the difference between the rank of the saturated model matrix Z and the rank of the reduced model matrix X.

3.2. Comparing models and model selection

As in other contexts, any particular data set may be described well by several graphical log-linear models. In some cases, existing theory often suggests a particular model, with the research study perhaps focusing on whether the model needs a certain interaction or conditional dependence term. In exploratory studies, it is often useful to conduct a model selection procedure to find a simple model that adequately describes the data and to suggest questions for future research. As in ordinary graphical log-linear modelling, in either case such analyses may compare candidate models that differ by the presence/absence of some edges.

In either type of study, once a tentative working model of standard graphical form is selected, one can analyse whether terms involving homologous variables can be simplified to a relevant QS graphical model. As with ordinary graphical log-linear models, there may be more than one acceptable model. Interpretation is often simplified, as shown in the first example in section 5, if we can remove at least one of the higher-order interactions, even if the resulting model is no longer graphical.

In a model selection process, typically the likelihood-ratio test statistic is used to compare a model $\mathcal{M}(G_1)$ with a reduced model $\mathcal{M}(G_2)$, where $G_1=(V,E_1)$ and $G_2=(V,E_2)$ are such that $E_2\subset E_1$. In comparing QS graphical models, however, careful consideration must be given to whether two models are appropriately nested (one a special case of the other), because one model may result from two kinds of linear constraints on the other model: (i) a zero constraint, when an edge is removed; and (ii) an equality constraint, when an edge is taken to be coloured or certain edges are constrained to have the same colour. A reduced model results whenever one of the following actions is performed to a model:

(a) a black edge is removed
(b) a black edge is coloured
(c) all the edges of a same colour are removed
(d) two or more colour classes are joined.

Case (a) corresponds to imposing a zero constraint on a set of parameters; case (b) corresponds to imposing equality constraints on a set of parameters, while case (c) corresponds to imposing a zero constraint on a set of parameters that were already constrained to be equal. Finally, in case (d) two or more sets of parameters, already equal inside the sets, are fixed to be equal. For instance, in Fig. 2, graph B corresponds to a reduced model of graph A but not of graph C. Indeed, the models associated with graphs B and C have the same number of parameters.

4. Properties of QS graphical models

The interpretation of QS graphical models is aided by the fact that they satisfy the Markov properties of the skeleton graph, according to definition 1(i). However, their behaviour regarding decomposability and collapsibility requires consideration. In ordinary graphical log-linear models, it has been shown (see Asmussen & Edwards, 1983) that, given a decomposition a and b of a graph G (as defined in the Appendix),

$$\mu(\imath)=\frac{\mu(\imath_a)\mu(\imath_b)}{\mu(\imath_{a\cap b})}, \tag{9a}$$

where $\mu(\imath_a)$ is the expected count in cell $\imath_a$ of the marginal table $\mathcal{I}_a$. Equivalently, $\log\mu(\imath)=\log\mu(\imath_a)+\log\mu(\imath_b)-\log\mu(\imath_{a\cap b})$. Moreover, also the ML estimates, the deviance and the degrees of freedom admit a similar decomposition, so that inference for a graphical log-linear model $\mathcal{M}(G)$ can be based on the three lower-dimensional models on the subgraphs induced by the decompositions: $\mathcal{M}_a(G_a)$, $\mathcal{M}_b(G_b)$ and $\mathcal{M}_{a\cap b}(G_{a\cap b})$. In terms of estimated expected counts,

$$\hat{\mu}(\imath)=\frac{\hat{\mu}_a(\imath_a)\hat{\mu}_b(\imath_b)}{\hat{\mu}_{a\cap b}(\imath_{a\cap b})}, \tag{9b}$$

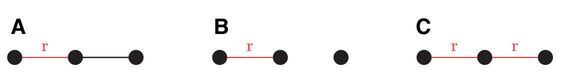


Fig. 2. Examples of undirected graphs.

where $\hat{\mu}_a(\imath_a)$ is the ML estimate of $\mu(\imath_a)$ obtained by fitting the lower-dimensional model $\mathcal{M}_a(G_a)$ to the marginal table $\mathcal{I}_a$. Such a decomposition in the ML estimates happens with QS graphical models only in some special cases. Because of the equality constraints, the factorization of the joint distribution with respect to the variables in a model does not always correspond to a consequent parameter-based factorization of the likelihood function (Cox & Wermuth, 1999), and the ML estimates cannot be obtained by fitting separate lower-dimensional models.

4.1. Coloured decomposition for QS graphical models

We introduce next a definition of decomposition for QS graphical models.

Definition 2. *Two subsets a and b of an edge-coloured graph $G=(V,E,\mathscr{E})$ form a coloured decomposition of V relative to a QS graphical model $\mathscr{M}(G)$ if*

(i) *a and b form a decomposition for the skeleton graph G^*, and*
(ii) $\mathrm{col}(a)\cap\mathrm{col}(b)=\emptyset$.

This definition adds a further condition to the common definition of decomposition to avoid the case of QS constraints coupling parameters from different parts of the decomposition. Notice that, because of this added condition, the subgraph induced by $a\cap b$ can have only black edges. For QS graphical models satisfying this new condition for decomposition, proposition 2 holds.

Proposition 2. *Suppose two subsets a and b of an edge-coloured graph $G=(V,E,\mathscr{E})$ form a coloured decomposition of V relative to a QS graphical model $\mathscr{M}^{\mathrm{QS}}(G)$. Then*

$$\log\hat{\mu}(\imath)=\log\hat{\mu}_a(\imath_a)+\log\hat{\mu}_b(\imath_b)-\log\hat{\mu}_{a\cap b}(\imath_{a\cap b}) \text{ and } \log\hat{\mu}(\imath_a)=\log\hat{\mu}_a(\imath_a),$$

where $\hat{\mu}_a(\imath_a)$ denotes the ML estimates of $\mu(\imath_a)$ obtained for the marginal model $\mathscr{M}_a^{\mathrm{QS}}(G_a)$.

Proof. As a and b form a coloured decomposition of V in the coloured graph G, they also form a decomposition in the skeleton graph G^* that corresponds to the ordinary graphical log-linear model $\mathscr{M}(G^*)$, which is the unconstrained version of $\mathscr{M}^{\mathrm{QS}}(G)$. Denoting $c=a\cap b$, according to Asmussen & Edwards (1983), the ML estimates for $\mathscr{M}(G^*)$ can be achieved by maximizing separately the three terms of the decomposition, $\mathscr{M}_a(G_a^*)$, $\mathscr{M}_b(G_b^*)$ and $\mathscr{M}_c(G_c^*)$, each term depending on the observed counts of separate marginal tables. The QS model $\mathscr{M}^{\mathrm{QS}}(G)$ is obtained by adding constraints according to the colours of the edge in the coloured graph G. As a and b form a coloured decomposition, colours are added separately to the subgraphs $G_{a\setminus c}$ and $G_{b\setminus c}$. No coloured edge in a (excluding the black edges) has the same colour as an edge in b. Consequently, no equality constraints couple parameters regarding a with those in b. Therefore, the sufficient statistics for the constrained parameters concerning a part of the decomposition, say G_a, depend only on the sufficient statistics of the corresponding unconstrained model, say $\mathscr{M}_a(G_a^*)$. The subgraph G_c is complete and not coloured, so the model equates the fitted marginal table to the corresponding observed marginal table. Then, the ML estimates can be obtained by separately maximizing the three terms of the decompositions, $\mathscr{M}_a(G_a)$, $\mathscr{M}_b(G_b)$ and $\mathscr{M}_c(G_c)$, as each term depends only on the observed counts of the separate marginal tables, as in the unconstrained model.

Coloured decompositions are useful for QS graphical models, as they imply that the model and its fit can be obtained from separate lower-dimensional models. Example 11 illustrates the use of decomposition in fitting.

Example 11. Consider Fig. 3C; the graph has $V=\{1,2,3\}$ and $E=\{12,23\}$. The two edges are coloured by different colours, so that the corresponding model is $\log\mu_{ijk}=\lambda+\lambda_i^1+\lambda_j^2+\lambda_k^3+\lambda_{ij}^{12}+\lambda_{jk}^{23}$, with QS constraints $\lambda_{ij}^{12}=\lambda_{ji}^{12}$ and $\lambda_{ij}^{23}=\lambda_{ji}^{23}$. Its skeleton graph, in Fig. 3A, is decomposable into its cliques [12][23]. Thus, in the ordinary model represented by the skeleton graph, $\log\mu_{ijk}=\log\mu_{ij.}+\log\mu_{.jk}-\log\mu_{.j.}$, the dot denoting summation with respect to the index. This is true also for the QS graphical model, in the sense that it can be specified by the decomposition of the model formula associated with the skeleton graph, and by two different QS models for the margins [12] and [23], that is,

$$\begin{aligned}
&\log\mu_{ij.}=\alpha+\alpha_i^1+\alpha_j^2+\alpha_{ij}^{12}, \quad \alpha_{ij}^{12}=\alpha_{ji}^{12}\\
&\log\mu_{.jk}=\gamma+\gamma_j^2+\gamma_k^3+\gamma_{jk}^{23}, \quad \gamma_{jk}^{23}=\gamma_{kj}^{23}\\
&\log\mu_{.j.}=\xi+\xi_j^2.
\end{aligned}$$

Here, $\lambda=\alpha+\gamma-\xi$ and $\lambda_j^2=\alpha_j^2+\gamma_j^2-\xi_j^2$, while the other parameters coincide (e.g. $\alpha_i^1=\lambda_i^1$). The same decomposition holds for the ML estimates, so that, using the notation λ^a for the vector of the components $\lambda^a(\iota_a)$:

$$\begin{aligned}
&\hat{\lambda}=\hat{\alpha}+\hat{\gamma}-\hat{\xi}, \quad \hat{\lambda}^1=\hat{\alpha}^1, \quad \hat{\lambda}^3=\hat{\gamma}^3,\\
&\hat{\lambda}^2=\hat{\alpha}^2+\hat{\gamma}^2-\hat{\xi}^2, \quad \hat{\lambda}^{12}=\hat{\alpha}^{12}, \quad \hat{\lambda}^{23}=\hat{\gamma}^{23}.
\end{aligned}$$

The possibility of obtaining the ML estimates by combining separate lower-dimensional models on the cliques and separators depends however on the different parts of the decomposition having different colours, as we can see in example 12.

Example 12. Consider the QS graphical model corresponding to the graph in Fig. 3B. The model formula is $\log\mu_{ijk}=\lambda+\lambda_i^1+\lambda_j^2+\lambda_k^3+\lambda_{ij}^{12}+\lambda_{jk}^{23}$, with $\lambda_{ij}^{12}=\lambda_{ji}^{12}=\lambda_{ij}^{23}=\lambda_{ji}^{23}$. Here the ML estimates cannot be obtained from the marginal tables as the QS constraints couple some parameters in the two different QS models for the margins [12] and [23]. In fact, the sufficient statistics for this model are $\{n_{i..}\}$, $\{n_{.i.}\}$, $\{n_{..i}\}$ and $\{n_{ij.}+n_{ji.}+n_{.ij}+n_{.ji}\}$ (see Bishop *et al.*, 1975, sections 8.2–8.3), where n_{ijk} denotes the observed count for the cell (i,j,k). The ML estimates can be obtained by equating the sufficient statistics to their expectation under the imposed constraints. Therefore, the likelihood equations for $\lambda_{ij}^{12}=\lambda_{ij}^{23}$ are:

$$\hat{\mu}_{ij.}+\hat{\mu}_{ji.}+\hat{\mu}_{.ij}+\hat{\mu}_{.ji}=n_{ij.}+n_{ji.}+n_{.ij}+n_{.ji} \quad \forall i<j=1,\dots I$$

depending both on the marginal tables [12] and [23].

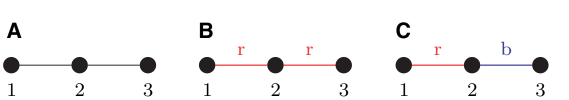


Fig. 3. (A) An undirected graph; (B) a coloured graph non-decomposable in cliques and colours; (C) a coloured graph decomposable in cliques and colours.

4.2. Collapsibility for QS graphical models

Sometimes it is of interest to reduce a large contingency table to a smaller one, focusing on a subset of variables of primary interest if the remaining variables have no effect on the interaction structure among those primary variables. However, only in certain cases is such interaction structure not affected by collapsing over other variables. Different definitions of collapsibility have been given in the literature to describe the property of having the marginal distribution of a subset of variables the same as in a larger model. See, for example, Whittaker (1990).

Consider first the case of parametric collapsibility with respect to log-linear parameters. According to Whittemore (1978), a discrete probability distribution $p(\imath)$ is *collapsible* onto a marginal distribution $p_M(\imath_M)$ for $M \subseteq V$, with respect to the log-linear interaction parameter λ^A, with $A \subseteq M$, if λ^A coincides with the marginal log-linear parameter λ^A_M obtained from $p_M(\imath_M)$.

Let now a, b form a decomposition of the skeleton graph G^* of a QS graphical model with an edge-coloured graph $G=(V,E,\mathscr{E})$. For $A=a\setminus b$, $B=b\setminus a$ and $C=a\cap b$, we have the conditional independence $X_A \perp\!\!\!\perp X_B \mid X_C$. This conditional independence by Whittemore (1978, corollary 2) is a sufficient condition for log-linear parametric collapsibility onto a, $\lambda^{A'}=\lambda^{A'}_a$ for all non-empty $A' \subseteq A$. These equalities are preserved also by the QS graphical model, because the added constraints do not affect the conditional independence structure, but involve only further equality restrictions among non-zero interaction parameters.

The previous discussion ensures that after marginalization, in the presence of a decomposition, there is parametric collapsibility with respect to selected parameters. Therefore, QS graphical models behave like standard log-linear graphical models as far as parametric collapsibility is concerned.

Two other concepts used in graphical models are model collapsibility and estimate collapsibility. These concepts have been discussed by Asmussen & Edwards (1983) who also showed their equivalence in the case of hierarchical (and thus also graphical) log-linear models. A hierarchical log-linear model $\mathcal{M}(G)$ defined on V is *model collapsible* onto a subset a of V whenever for all $\mu(\imath)\in\mathcal{M}(G)$, $\mu(\imath_a)\in\mathcal{M}_a(G_a)$, that is, the model for the marginal distribution of a does not differ from the one represented by the induced subgraph G_a. On the other hand, we say that there is *estimate collapsibility* if the ML estimates for the marginal table $\mathcal{I}_a$ in $\mathcal{M}(G)$ coincide with the ML estimates for the marginal table $\mathcal{I}_a$ in the marginal model $\mathcal{M}_a(G_a)$, that is $\hat{\mu}(\imath_a)=\hat{\mu}_a(\imath_a)$, for all $\imath_a$.

In the case of QS graphical models, model collapsibility is implied by parametric collapsibility, but, by similar arguments as given for decompositions in the previous section, the equivalence between model collapsibility and estimate collapsibility does not occur in finite samples. To have estimate collapsibility we need the further condition that the QS constraints for the set we are collapsing onto must be disjoint from the QS constraints for the complementary set.

Proposition 3. *Let a be a subset of V in a QS graphical model $\mathcal{M}^{\mathrm{QS}}(G)$, $G=(V,E,\mathscr{E})$. For $a^c=V\setminus a$, let* $\mathrm{cl}(a^c)$ *denote the closure of a^c. A necessary and sufficient condition for estimating collapsibility onto a is that*

(i) *the boundary of every connected component of a^c is complete, and*
(ii) $\mathrm{col}(a)\cap\mathrm{col}(\mathrm{cl}(a^c))=\emptyset$.

Proof. Sufficiency: Let $b=\mathrm{cl}(a^c)$. Then, a and b form a decomposition, with $a\cap b=\mathrm{bd}(a^c)$ complete and separating $a\setminus b$ from $b\setminus a$. Moreover, condition (ii) implies that this is a coloured decomposition. Thus, the result follows because, by proposition 2, we have $\hat{\mu}(\imath_a) = \hat{\mu}_a(\imath_a)$.

Necessity: Suppose that $\mathcal{M}^{\mathrm{QS}}(G)$ is collapsible onto a in its estimates. Then, it is also model collapsible and condition (i) necessarily holds (Asmussen & Edwards, 1983, theorem 2.3). Suppose now that condition (i) holds, while condition (ii) is false. As $\mathrm{col}(a) \cap \mathrm{col}(\mathrm{cl}(a^c)) \neq \emptyset$, some constraints involve both elements of a and $\mathrm{cl}(a^c)$. Then, some of the likelihood equations for model $\mathcal{M}^{\mathrm{QS}}(G)$ involve both sufficient statistics and observed margins in a and in $\mathrm{cl}(a^c)$. This results, in general, in estimates that differ from those derived from the likelihood equations for model $\mathcal{M}_a^{\mathrm{QS}}(G_a)$, which are based only on the sufficient statistics and observed margins in a, unless the set of constraints is empty. This completes the proof.

To illustrate the concept of collapsibility, we return to example 11.

Example 13 (example 11 continued). Let us consider again the graph in Fig. 3C. Marginal analysis on the collapsed tables can be conducted both for the margins [12] and [23]. For instance, consider the margin [12]. The set of nodes can be partitioned into $a = \{1, 2\}$ and $a^c = \{3\}$. The closure $\mathrm{cl}(a^c)$ consists in the clique [23], which is complete. Moreover, the colour of the edge 12 (the only edge in G_a) is red, while the colour of 23 is blue. As a consequence, both conditions of proposition 3 are fulfilled and the model can be collapsed onto the marginal table [12]. Similar arguments are valid for margin [23]. These conditions would not be satisfied if the graph had both edges of the same colour, as, for example, in Fig. 3B.

Notice that, as a consequence of condition (ii) in proposition 3, nodes in the boundary of a^c can be connected only by black edges. In particular, the boundary of a^c belongs to a, so that a coloured edge between two nodes in the boundary implies that $\mathrm{col}(a) \cap \mathrm{col}(\mathrm{cl}(a^c)) \neq \emptyset$.

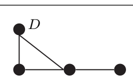
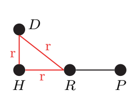
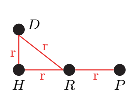
5. Examples

In the following two sections, the two data sets introduced in section 1 are analysed with QS graphical models.

5.1. Opinions about abortion and the death penalty

Table 1 concerns attitudes of 3218 US citizens who gave valid responses to four variables considered: attitude towards death penalty (P), and attitude towards abortion if there is a strong

Table 3. *Models comparison for data on attitudes towards abortion and the death penalty*

Model	Graph	No. of parameters	*df*	Deviance
Model 1		10	6	6.65
Model 2		8	8	8.80
Model 3		7	9	304.2

chance of a serious defect in the baby (D), if the woman's health is seriously endangered (H) and if the pregnancy is a result of a rape (R). A preliminary analysis using ordinary graphical log-linear models leads to a graphical model (model 1 in Table 3) by which $(D, H) \perp\!\!\!\perp P \mid R$:

$$\log \mu_{ijkl} = \lambda + \lambda_i^H + \lambda_j^D + \lambda_k^R + \lambda_l^P + \lambda_{kl}^{RP} + \lambda_{ij}^{HD} + \lambda_{ik}^{HR} + \lambda_{jk}^{DR} + \lambda_{ijk}^{HDR}. \tag{10}$$

This model corresponds to the undirected graph $G_1 = (V, E)$ in which E contains HR, HD, DR, RP. To summarize the fit of this and other models, we will use the deviance only as an overly stringent indication of lack of fit that would provide lower bounds on actual P-values, because the sampling design for the General Social Survey is more complex than simple random sampling and the code book (Davis *et al.*, 2009) for the survey suggests that sampling variances tend to be about 50 per cent larger than for simple random sampling. For this model, the deviance is 6.65 with $df = 6$, suggesting an adequate fit.

All four variables are homologous, but the death penalty opinion is of a different nature from the abortion items, so it seems sensible to consider simplifying by imposing QS structure among the abortion items, leaving the association between P and R to be different. This gives the QS graphical model, model 2 in Table 3, corresponding to the coloured graph $G_2 = (V, E, \mathcal{E})$, with $\mathcal{E} = \{E_0, E_1\}$, $E_0 = \{RP\}$ and $E_1 = \{HR, HD, DR\}$, for which $\lambda_{ij}^{HD} = \lambda_{ij}^{HR} = \lambda_{ij}^{DR}$. This model also fits adequately (deviance $= 8.80, df = 8$).

For model 2, achieving identifiability by using zero-sum constraints, we obtain ML estimate $\hat{\lambda}_{111} = -0.046\,(SE = 0.031)$ for the three-factor term. After removing this non-significant interaction, we obtain a model which is no longer graphical but simpler to interpret. This reduced model has a deviance of 11.0 on nine degrees of freedom. The ML estimate of the common two-factor term is then $\hat{\lambda}_{11} = 0.616\,(SE = 0.016)$. Thus, for those with a given response on one of the abortion items, the estimated odds ratio between the other two items is $\exp[4(0.616)] = 11.7$. For this model, the estimated odds ratio between R and P is only 1.77.

As an aside, we mention model 3 in Table 3, which also has exchangeability including the association between R and P. On subject-matter grounds, one would not expect this association to be similar to the other three in that model, as P deals with a different subject. That model indeed fits poorly, with a deviance of 304.2 ($df = 9$). The poor fit reflects the quite different estimated odds ratio for R and P in the model just discussed.

5.2. *Degree of satisfaction with US government policy*

In this second example, we analyse the data displayed in Table 2 on the degree of satisfaction on some aspects of US government policy about the environment (E), the unemployment (U) and a decent standard of living for the old (O), all cross-classified with gender (G). A preliminary analysis with ordinary graphical log-linear models leads to model A, showing an adequate fit (deviance $= 20.85, df = 24$). The model assumes $(E, O) \perp\!\!\!\perp G \mid U$, with

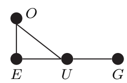
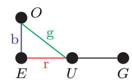
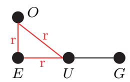
$$\log \mu_{ijkl} = \lambda + \lambda_i^E + \lambda_j^O + \lambda_k^U + \lambda_l^G + \lambda_{kl}^{UG} + \lambda_{ij}^{EO} + \lambda_{ik}^{EU} + \lambda_{jk}^{OU} + \lambda_{ijk}^{EOU}. \tag{11}$$

This model ignores that E, O and U are homologous. We can so adapt a QS graphical model adding the following equality constraints to model (11)

$$\lambda_{ij}^{EO} = \lambda_{ji}^{EO} = \lambda_{ij}^{EU} = \lambda_{ji}^{EU} = \lambda_{ij}^{OU} = \lambda_{ji}^{OU} \quad \text{and} \quad \lambda_{ijk}^{EOU} = \lambda_{\text{perm}(ijk)}^{EOU} \quad \forall i, j, k = 1, \ldots, I.$$

This includes in the same colour class all the edges in the subgraph G_{EOU}. The so-defined QS graphical model is model C in Table 4. This model implies a symmetric and common conditional association among the homologous variables, given all the other variables, including the non-homologous gender. Model C fits adequately, with a deviance of 43.06 ($df = 37$). In

Table 4. *Models comparison for data on the degree of satisfaction with US government policy*

Model	Graph	No. of parameters	*df*	Deviance
Model A		30	24	20.85
Model B		23	31	30.03
Model C		17	37	43.06

model B we relax the model C assumption of a common conditional association among E, O and U by specifying separate QS constraints as follows:

$$\lambda_{ij}^{EO}=\lambda_{ji}^{EO},\quad \lambda_{ij}^{EU}=\lambda_{ji}^{EU},\quad \lambda_{ij}^{OU}=\lambda_{ji}^{OU}\quad \forall i,j=1,\ldots,I$$
$$\lambda_{ijk}^{EOU}=\lambda_{\mathrm{perm}(ijk)}^{EOU},\quad \forall i,j,k=1,\ldots,I.$$

The corresponding graph has three edges with different colours and a black edge, UG. This model has a deviance of 30.03 with 31 degrees of freedom. All three models seem compatible with the data.

All the models in this section and models 1 and 2 in section 5.1 have the two cliques forming a coloured decomposition. The estimates reported before can be obtained by fitting separate models, using proposition 2.

6. Discussion

We have proposed a class of graphical log-linear models, called QS graphical models, in which some interaction parameters are constrained to be equal. These models have as a special case the QS model for two-way tables proposed by Caussinus (1966) and its generalizations to higher-way tables by Bishop *et al.* (1975) and Bhapkar & Darroch (1990). Our model extends to other structures, such as sets of variables for which only some are homologous or different sets are homologous. The symmetry in the conditional association structure is represented using coloured edges in a conditional independence graph, in the spirit of Højsgaard & Lauritzen (2008) for concentration graphical models. The coloured graph represents both the conditional independence structure and the QS constraints, so that also the form of the association is represented in the graph. Moreover, the graph is useful for determining the conditions for decomposability and collapsibility. Specifically, decomposable QS models are models with a triangulated structure for the skeleton graphs that have also cliques not sharing colours. In ordinary log-linear models, decomposability ensures that the ML estimates can be calculated from marginal tables using closed-form expressions. In QS graphical models, closed-form expressions are not available in the presence of symmetry constraints, but decomposability can still be very useful to reduce model complexity, for example, in large graphs. In general, QS graphical models can be estimated as generalized linear models in which the model matrix has a certain structure. The standard Fisher scoring algorithm is usually adequate for fitting the model, but the iterative proportional fitting algorithm can also be used to obtain fitted values.

As an extension, it would be of interest to consider such equality constraints with more highly structured association and interaction terms. This would be useful when some variables are ordinal and associations are more parsimoniously described with models such as the RC association model and its special cases (Goodman, 1979, possibly with identical scores for each variable in a set) or an ordinal QS model. For example, such models would exploit the ordinal nature of the homologous variables in Table 2. Other possible extensions could be made in the framework of Rasch models, whose connections with QS models have been pointed out by Agresti (2002) and Erosheva *et al.* (2002). Moreover, different kinds of graphs, such as a coloured version of directed acyclic graphs and chain graphs could be considered to enlarge the proposed class of models to the case of one or more response variables. Further useful conditions for collapsibility in directed acyclic graph models which may have some relevance here have been given by Kim & Kim (2006).

References

Agresti, A. (2002). Links between binary and multi-category logit item response models and quasi-symmetric loglinear models. *Annales de la Faculté des Sciences de Toulouse Sér. 6* **11**, 485–505.

Asmussen, S. & Edwards, D. (1983). Collapsibility and response variables in contingency tables. *Biometrika* **70**, 567–578.

Bhapkar, V. & Darroch, J. (1990). Marginal symmetry and quasi symmetry of general order. *J. Multivariate Anal.* **34**, 173–184.

Bishop, Y., Fienberg, S. & Holland, P. (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press.

Caussinus, H. (1966). Contribution à l'analyse statistique des tableaux de corrélation. *Annales de la Faculté des Sciences de l'Université de Toulouse* **29**, 77–182.

Cox, D. & Wermuth, N. (1999). Likelihood factorizations for mixed discrete and continuous variables. *Scand. J. Statist.* **26**, 209–220.

Darroch, J. N., Lauritzen, S. L. & Speed, T. P. (1980). Markov fields and log-linear interaction models for contingency tables. *Ann. Statist.* **8**, 522–539.

Davis, J., Smith, T. & Marsden, P. (2009). *General Social Surveys 1972–2008 cumulative codebook*. University of Chicago, National Opinion Research Center (NORC).

Erosheva, E., Fienberg, S. & Junker, B. (2002). Alternative statistical models and representations for large sparse multi-dimensional contingency tables. *Annales de la Faculté des Sciences de Toulouse Sér. 6* **11**, 485–505.

Goodman, L. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *J. Amer. Statist. Assoc.* **74**, 537–552.

Højsgaard, S. & Lauritzen, S. L. (2008). Graphical Gaussian models with edge and vertex symmetries. *J. Roy. Statist. Soc. Ser. B* **70**, 1005–1027.

Kim, S. H. & Kim, S. H. (2006). A note on collapsibility in DAG models of contingency tables. *Scand. J. Statist.* **14**, 97–112.

Lauritzen, S. L. (1996). *Graphical models*. Oxford Science Publications, Oxford.

Lovison, G. (2000). Generalized symmetry models for hypercubic concordance tables. *Int. Stat. Rev.* **68**, 323–338.

R Development Core Team (2009). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria; ISBN 3-900051-07-0.

Whittaker, J. L. (1990). *Graphical models in applied multivariate statistics*. Wiley, New York.

Whittemore, A. (1978). Collapsibility of multidimensional contingency tables. *J. Roy. Statist. Soc. Ser. B (Methodol.)* **40**, 328–340.

Received May 2009, in final form August 2010

Anne Gottard, Department of Statistics, University of Florence, Florence, 50134, Italy.
E-mail: gottard@ds.unifi.it

Appendix: graph theory

This appendix summarizes the graph theory required for this article. See Lauritzen (1996) for a full presentation of this topic.

An undirected graph $G=(V,E)$ is defined by a set of nodes V and a set of edges $E\subseteq V\times V$. The graph is assumed to have no loops, that is $(v,v)\notin E$ for all node v. An edge $(v,w)\in E$ is undirected if $(w,v)\in E$ and it is represented by a line joining v and w, which are said to be *neighbours*. If a is a subset of nodes, the subgraph induced by a is the graph obtained from $G=(V,E)$, by taking the nodes in a and by keeping the edges in E with both endpoints in a. A graph is *complete* if all nodes are joined by an edge, and a subset is complete if it induces a complete subgraph. A *clique* is a complete subset that is maximal, in the sense that if any other node was added to the subset it would no longer be complete. The *boundary* bd(a) of a subset a of nodes is the set of nodes in $V\setminus a$ that are neighbours to nodes in a, while the *closure* cl(a) is the union of a and its boundary.

In a graph, a *path* is a sequence of distinct nodes $v_1, v_2, \dots, v_s$ such that each $\{v_i, v_{i+1}\}\in E$, $i=1,\dots,s-1$, where s is the *length* of the path. A *cycle* is a path with $v_1=v_s$ and $s>2$. A subset a of V is a *connected component* if every couple of nodes is connected by a path. A set of nodes c is a *separator* of two distinct set of nodes a and b if every path from a to b has at least one node in c.

Two non-empty subsets a and b of V form a decomposition of an undirected graph G relative to a graphical log-linear model $\mathscr{M}(G)$, if $a\cup b=V$, and $a\cap b$ is complete and separates $a\setminus c$ from $b\setminus c$. Furthermore, a graph is said to be *decomposable* if it can be recursively decomposed in subgraphs, until all subgraphs are complete. A necessary and sufficient condition for decomposability is the graph to be triangulated.