

5-1-2017

# Graphical Log-Linear Models: Fundamental Concepts and Applications

Niharika Gauraha

*Indian Statistical Institute, Bangalore, India, niharika.gauraha@gmail.com*

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

---

## Recommended Citation

Gauraha, N. (2017). Graphical log-linear models: fundamental concepts and applications. *Journal of Modern Applied Statistical Methods*, 16(1), 545-577. doi: 10.22237/jmasm/1493598000

# Graphical Log-Linear Models: Fundamental Concepts and Applications

**Niharika Gauraha**

Indian Statistical Institute, Bangalore Center  
Bangalore, India

---

A comprehensive study of graphical log-linear models for contingency tables is presented. High-dimensional contingency tables arise in many areas. Analysis of contingency tables involving several factors or categorical variables is very hard. To determine interactions among various factors, graphical and decomposable log-linear models are preferred. Connections between the conditional independence in probability and graphs are explored, followed with illustrations to describe how graphical log-linear model are useful to interpret the conditional independences between factors. The problem of estimation and model selection in decomposable models is discussed.

*Keywords:* Graphical log-linear models, contingency tables, decomposable models, hierarchical log-linear models

---

## Introduction

The aim in the current study is to provide insight into graphical log-linear models (LLMs) by providing a concise explanation of the underlying mathematics and statistics, by pointing out relationships to conditional independence in probability and graphs, and providing pointers to available software and important references. LLMs are the most widely used models for analyzing cross-classified categorical data (Christensen, 1997). LLM supports various ranges of models based on non-interaction assumptions. For fairly large-dimensional tables, the analysis becomes difficult; as the number of factors increases the number of interaction terms grows exponentially. Graphical LLMs are a way of representing relationships among the factors of a contingency table using a graph. The graphical LLMs have two great advantages: from the graph structure, it is easy to read off the conditional independence relations; and graph-based algorithms usually provide efficient computational algorithms for parameter estimation and model selection.

---

*Niharika Gauraha is a PhD student. Email them at: [niharika.gauraha@gmail.com](mailto:niharika.gauraha@gmail.com).*

## GRAPHICAL LOG-LINEAR MODELS

The decomposable LLMs are a restricted class of GLLMs which are based on chordal graphs. There are several reasons for using decomposable models over an ordinary GLLM. Firstly, the maximum likelihood estimates can be found explicitly. Secondly, closed-form expressions exist for the test statistics. Another advantage is that it has triangulated graph-based efficient inference algorithms. Thus decomposable models are mostly used for analysis of high-dimensional tables.

### Graph Theory and Markov Networks

#### Graph Theory

Necessary concepts of graph theory that will be used are discussed. See West (2000) for further details on graph theory. A graph  $G$  is a pair  $G = (V, E)$ , where  $V$  is a set of vertices and  $E$  is a set of edges. A graph is said to be an undirected graph when  $E$  is a set of unordered pairs of vertices. Consider only a simple graph that has neither loops nor multiple edges.

**Definition 1 (Boundary):** Let  $G = (V, E)$  be an undirected graph. The neighbors or boundary of a subset  $A$  of vertices is a subset  $C$  of vertices such that all nodes in  $C$  are not in  $A$  but are adjacent to some vertex in  $A$ .

$$\text{bd}(A) = \{u \in V \setminus A \mid \exists v \in A : \{u, v\} \in E\}$$

**Definition 2 (Maximal Clique):** A clique of a graph  $G$  is a subset  $C$  of vertices such that all vertices in  $C$  are mutually adjacent. A clique is said to be maximal if no vertex can be added to  $C$  without violating the clique property.

**Definition 3 (Chordal (Triangulated) Graphs):** In graph theory, a chord of a cycle  $C$  is defined as an edge which is not in the edge set of  $C$  but joins two vertices from the vertex set  $C$ . A graph is said to be a chordal graph if every cycle of length four or more has a chord.

**Definition 4 (Isomorphic Graphs):** Two graphs are said to be isomorphic if they have same number of vertices, same number of edges, and they are connected in the same way.

## Conditional Independence

The concept of conditional independence in probability theory is very important and it is the basis for the graphical models. It is defined as follows:

**Definition 5 (Conditional Independence):** Let  $X$ ,  $Y$ , and  $Z$  be random variables with a joint distribution  $P$ . The random variables  $X$  and  $Y$  are said to be conditionally independent given the random variable  $Z$  if and only if the following holds:

$$\begin{aligned} P(X, Y | Z) &= P(X | Z)P(Y | Z) \\ P(X | YZ) &= P(X | Z) \end{aligned}$$

Dawid's (1979) notation,  $X \perp\!\!\!\perp Y | Z$ , is also used. Conditional independence has a vast literature in the field of probability and statistics; see also Pearl and Paz (1987).

## Markov Networks and Markov Properties

Markov network graphs, Markov networks, and different Markov properties for the Markov Networks are now defined.

**Definition 6 (Markov Network Graphs):** A Markov network graph is an undirected graph  $G = (V, E)$  where  $V = \{X_1, \dots, X_n\}$  represents random variables of a multivariate distribution.

**Definition 7 (Markov Networks):** A Markov network  $M$  is a pair  $M = (G, \Psi)$ . Where  $G$  is a Markov network graph and  $\Psi = \{\psi_1, \dots, \psi_m\}$  is a set of non-negative functions for each maximal clique  $C_i \in G \forall i = 1, \dots, m$ , and the joint probability density function (pdf) can be decomposed into factors as

$$P(x) = \frac{1}{Z} \prod_{a \in C_m} \psi_a(x)$$

where  $Z$  is a normalizing constant.

## GRAPHICAL LOG-LINEAR MODELS

**Definition 8 (Pairwise Markov Property (P)):** A probability distribution  $P$  satisfies the pairwise Markov property for a given undirected graph  $G$  if, for every pair of non-adjacent vertices  $X$  and  $Y$ ,  $X$  is independent of  $Y$  given the rest.

$$X \perp\!\!\!\perp Y \mid (V \setminus X, Y)$$

**Definition 9 (Local Markov Property (L)):** A probability distribution  $P$  satisfies the local Markov property for a given undirected graph  $G$  if every variable  $X$  is conditionally independent of its non-neighbors in the graph, given its neighbors.

$$X \perp\!\!\!\perp (V \setminus (X \cup \text{bd}(X))) \mid \text{bd}(X)$$

**Definition 10 (Global Markov Property (G)):** A probability distribution  $P$  is said to be global Markov with respect to an undirected graph  $G$  if and only if, for any disjoint subsets of nodes  $A$ ,  $B$ , and  $C$  such that  $C$  separates  $A$  and  $B$  on the graph, the distribution satisfies the following:

$$A \perp\!\!\!\perp B \mid C$$

Note the above three Markov properties are not equivalent to each other. The local Markov property is stronger than the pairwise one, while weaker than the global one. More precisely,

**Proposition 1:** For any probability measure the following holds:

$$(G) \Rightarrow (L) \Rightarrow (P)$$

See Lauritzen (1996), for proof of Proposition 1. Refer to Lauritzen (1996) and Edwards (2000) for further details on graphical models, and to Darroch, Lauritzen, and Speed (1980) for details on Markov fields for LLMs.

### Notations and Assumptions

The notations and the assumptions are now discussed. Consider three-dimensional tables for notational simplicity; this is also a true representative of  $k$ -dimensions and thus can be easily extended to any higher dimensions by increasing the

number of subscripts. See Christensen (1977) and Bishop, Fienberg, and Holland (1989).

Consider a three-dimensional table with factors  $X$ ,  $Y$ , and  $Z$ . Numeric  $\{1, 2, 3\}$  and alphabetic  $\{X, Y, Z\}$  symbols are used interchangeably to represent the factors of a contingency table. Suppose the factors  $X$ ,  $Y$ , and  $Z$  have  $I$ ,  $J$ , and  $K$  levels, respectively. Then we have an  $I \times J \times K$  contingency table.

The following notations are defined for each elementary cell  $(i, j, k)$  for  $i = 1, \dots, I, j = 1, \dots, J$ , and  $k = 1, \dots, K$ :

- $n_{ijk}$  = the observed counts in the cell  $(i, j, k)$
- $m_{ijk}$  = the expected counts in the cell  $(i, j, k)$
- $\hat{m}_{ijk}$  = the Maximum Likelihood Estimate (MLE) of  $m_{ijk}$
- $p_{ijk}$  = the probability of a count falling in cell  $(i, j, k)$
- $\hat{p}_{ijk}$  = the MLE of  $p_{ijk}$

The following notations are used for sums of elementary cell counts, where “.” represents summation over that factor. For example,

$$n_{i..} = \sum_{jk} n_{ijk}$$

$$n_{i.k} = \sum_j n_{ijk}$$

$$N = n_{...} = \text{total number of observations}$$

Similarly, the marginal totals of probabilities and the expected counts are denoted by  $p_{.jk}$ , and  $m_{.jk}$ , etc.

Denote by  $C$  the tables of sums obtained by summing over one or more factors, e.g.  $C_{12}$  represents tables of counts  $n_{ij.}$ . Subscripted  $u$ -term notation is used for main effects and interactions. For example,  $u_{ij}$  is used for two-factor interactions  $\forall i = 1, \dots, I$  and  $\forall j = 1, \dots, J$ . We may interchangeably use  $u_{12(ij)}$  and  $u_{ij}$ ; the latter is obtained by simply dropping the second set of subscript. Thus

$$u_{12} = u_{12(ij)} \quad \forall i = 1, \dots, I, j = 1, \dots, J$$

Assume that the observed cell counts are strictly positive for all models we consider throughout this article.

## Overview of Contingency Tables

A contingency table is a table of counts that summarizes the relationship between factors. In a multivariate qualitative data set where each individual is described by a set of attributes, all individual with same attributes are counted; this count is entered into a cell of a corresponding contingency table (see Bishop, Fienberg, & Holland, 1989). The term contingency was introduced by Pearson (1904). There is an extensive body of literature on contingency tables; see A. H. Andersen (1974), Bartlett (1935), and Goodman (1969).

**Example 1:** Table 1 provides an example of a three-dimensional contingency table taken from example 3.2.1 of Christensen (1997).

## Types of Contingency Tables

Based on the underlying assumption of sampling distributions, contingency tables are divided into three main categories as follows:

**The Poisson Model** In this model, it is assumed that cell counts are independent and Poisson-distributed. The total number of counts and the marginal counts are random variables. For three-dimensional tables with counts as random variables, the joint probability density function (pdf) can be written as

$$f(\{n_{ijk}\}) = \prod_i \prod_j \prod_k \frac{m_{ijk}^{n_{ijk}} e^{-m_{ijk}}}{m_{ijk}!} \quad (1)$$

**The Multinomial Model** In this model, it is assumed that the total number of subjects  $N$  is fixed. With this constraint imposed on independent Poisson distributions, the cell counts yield a multinomial distribution. For proof we refer to Fisher (1922). The pdf for this model is given as

**Table 1.** Personality type table

Personality Type	Cholesterol	Diastolic Blood Pressure	
		Normal	High
A	Normal	716	79
	High	207	25
B	Normal	819	67
	High	186	22

$$f(\{n_{ijk}\}) = \frac{N!}{\prod_i \prod_j \prod_k n_{ijk}!} \prod_i \prod_j \prod_k \left( \frac{m_{ijk}}{N} \right)^{n_{ijk}} \quad (2)$$

**The Product-Multinomial Model** In this model, it is assumed that one set of marginal counts is fixed and the corresponding table of sums follow a product-multinomial distribution. For example, consider a three-dimensional table with total counts for the first factor,  $n_{.jk}$ , fixed. The pdf is given as

$$f(\{n_{ijk}\}) = \prod_j \prod_k \left[ \frac{n_{.jk}!}{\prod_i n_{ijk}!} \prod_i \left( \frac{m_{ijk}}{n_{.jk}} \right)^{n_{ijk}} \right] \quad (3)$$

## Introduction to Log-Linear Models

As discussed previously, the distribution of cell probabilities belong to exponential family (Poisson, multinomial, and product-multinomial). Construct a linear model in the log scale of the expected cell count. A LLM for a three-factor table is defined as

$$\log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)} \quad (4)$$

with the following identifiability constraints:

$$\begin{aligned} \sum_i u_{1(i)} &= \sum_j u_{2(j)} = \sum_k u_{3(k)} = 0 \\ \sum_i u_{12(ij)} &= \sum_j u_{12(ij)} = 0 \\ \sum_j u_{12(ik)} &= \sum_k u_{12(ik)} = 0 \\ \sum_j u_{12(jk)} &= \sum_k u_{12(jk)} = 0 \\ \sum_i u_{123(ijk)} &= \sum_j u_{123(ijk)} = \sum_k u_{123(ijk)} = 0 \end{aligned}$$

The above model is called saturated or unrestricted because it contains all possible one-way, two-way, and three-way effects. In general, if no interaction terms are set to zero, it is called the saturated model.

## GRAPHICAL LOG-LINEAR MODELS

The number of terms in a LLM model depends on the dimensions or number of factors and the interdependencies between the factors; it does not depend on the number of cells (see Birch, 1963 for more details). The model given by equation (4) applies to all three kinds of contingency tables with three factors (as discussed in the previous section), but there may be differences in the interpretations of the interaction terms (see Kreiner, 1998; Lang, 1996b). There is a wide body of literature on LLMs, see for instance Agresti (2002), Christensen (1997), Zelterman (2006), and Knoke and Burke (1980).

### Log-Linear Models as Generalized Linear Models

Recall the generalized linear model (GLM). It consists of a linear predictor and a link function. The link function determines the relationship between the mean and the linear predictor. Here, we show that the LLMs are special instances of GLMs for Poisson-distributed data; see Nelder and Wedderburn (1972) for details.

Consider a  $2 \times 2$  Poisson model with two factors, say  $X$  and  $Y$ , and suppose cell counts  $n_{ij}$  are response variables such that  $n_{ij} \sim \text{Poisson}(m_{ij})$  and the factors  $X$  and  $Y$  are explanatory variables. Define a link function  $g$  as  $g(m_{ij}) = \log(m_{ij})$ . The linear predictor is defined as  $\mathbf{X}'\boldsymbol{\beta}$ , where  $\mathbf{X}$  is the design matrix and  $\boldsymbol{\beta}$  is the vector of unknown parameters. For this model,  $\mathbf{X}$  and  $\boldsymbol{\beta}$  are defined as

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ (\alpha\beta)_{11} \\ (\alpha\beta)_{12} \\ (\alpha\beta)_{21} \\ (\alpha\beta)_{22} \end{bmatrix}$$

The model can be expressed as follows:

$$\log(m_{ij}) = x_i'\boldsymbol{\beta} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

Rename the parameters as

$$\log(m_{ij}) = u + u_1 + u_2 + u_{12}$$

The above model is the same as the LLM defined for two-factor tables, where  $u$  is the overall mean,  $u_1$  and  $u_2$  are the main effects, and  $u_{12}$  is the interaction effect. LLMs can be fit as generalized linear models by using software packages available for GLMs, e.g. the `glm()` function in the stats R package.

## Classes of Log-Linear Models

### Comprehensive Log-Linear Models

The class of comprehensive LLMs is defined as follows:

**Definition 11 (Comprehensive Log-Linear Models):** A log-linear model is said to be comprehensive if it contains the main effects of all the factors.

For example, a comprehensive LLM for the three-factor contingency tables must include all the main effects  $u_1$ ,  $u_2$ , and  $u_3$ , along with other interaction effects, if any (see Zelterman, 2006).

### Hierarchical Log-Linear Models

The class of hierarchical LLMs is defined as follows:

**Definition 12 (Hierarchical Log-Linear Models):** A LLM is said to be hierarchical if it contains all the lower-order terms which can be derived from the variables contained in a higher-order term.

For example, if a model for three-dimension table includes  $u_{12}$ , then  $u_1$  and  $u_2$  must be present. Conversely, if  $u_2 = 0$ , then we must have  $u_{12} = u_{23} = u_{123} = 0$ . The hierarchical models may be represented by giving only the terms of highest order, also known as a generating class, because all the lower-order terms are implicit. The generating class is defined as follows:

**Definition 13 (Generating class):** The highest-order terms in hierarchical LLMs are called a generating class because they generate all of the lower-order terms in the model.

## GRAPHICAL LOG-LINEAR MODELS

**Example 2:** A LLM with generating classes  $C = \{[123], [34]\}$  corresponds to the following log-linear model:

$$\log(m_{hijk}) = u + u_1 + u_2 + u_3 + u_4 + u_{12} + u_{23} + u_{13} + u_{123} + u_{34}$$

Members of generating class  $[123] = \{[1], [2], [3], [12], [23], [13], [123]\}$

Members of generating class  $[34] = \{[3], [4], [34]\}$

All models considered in the remaining sections of this article are hierarchical and comprehensive LLMs unless stated otherwise.

### Graphical Log-Linear Models

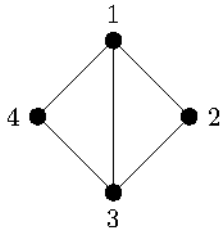
Consider a class of LLMs that can be represented by graphs, called graphical log-linear models (GLLMs).

**Definition 14 (Graphical Log-Linear Models):** A LLM is said to be graphical if it contains all the lower-order terms which can be derived from variables contained in a higher-order term, the model also contains the higher order interaction.

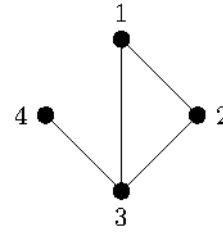
For example, if a model includes  $u_{12}$ ,  $u_{23}$ , and  $u_{31}$ , then it also contains the term  $u_{123}$ . In GLLMs, the vertices correspond to the factors and the edges correspond to the two-factor interactions. But the factors (vertices) and the two-factor interactions (edges) alone do not specify the graphical models. As mentioned previously, factorization of the probability distribution with respect to a graph must satisfy the Markov properties. For such a graph that respects the Markov properties with respect to a probability distribution, there is a one-to-one correspondence between GLLMs and graphs. It follows that every GLLM determines a graph and every graph determines a GLLM, as is illustrated by the following examples:

**Example 3:** Consider the model  $[123] [134]$ . The two-factor terms generated by  $[123]$  are  $[12]$ ,  $[13]$ , and  $[23]$ . Similarly, the two-factor terms generated by  $[134]$  are  $[13]$ ,  $[14]$ , and  $[34]$ . The corresponding graph is as given in [Figure 1](#).

Conversely, read the LLM directly from the corresponding graph. Consider a graph as given in [Figure 2](#); the edges are  $[12]$ ,  $[23]$ ,  $[13]$ , and  $[34]$ . Because the generating class for the terms  $[12]$ ,  $[23]$ , and  $[13]$  is the term  $[123]$ , we must include  $[123]$  in the model. Hence, the corresponding GLLM is  $[123] [34]$ .



**Figure 1.** Graphical model of [123] [134]



**Figure 2.** Graphical model of [123] [34]

Generating classes of GLLMs are in a one-to-one correspondence with the maximal cliques of the corresponding graph. Not all hierarchical LLMs have graphical representation. For example, the model [12] [13] [23] is hierarchical but it is not graphical because it does not contain the higher order term [123].

**Decomposable Models** Consider the class of decomposable models, which is a subclass of the GLLMs.

**Definition 15 (Decomposable Log-Linear Models):** A LLM model is decomposable if it is both graphical and chordal.

The main advantage of this model over other models is that it has closed form Maximum Likelihood Estimates (MLEs). For example, consider a decomposable model as given by Figure 1. The only conditional independence implied by the graph is that, given the factors 1 and 3, factors 2 and 4 are independent. The MLEs for the expected cell counts are factorized in a closed form in the terms of sufficient statistics as

$$\hat{m}_{ijkl} = \frac{n_{hij} \cdot n_{h.jk}}{n_{h.j}}$$

The derivation of MLE expressions, like the one above, is discussed in detail in a later section. For all the possible non-isomorphic graphical and decomposable models for the four-factor contingency tables, see Table 18 in the Appendix.

A few important articles concerned with the decomposable models are Goodman (1970, 1971b), Haberman (1974), Lauritzen, Speed, and Vijayan (1984), Meeden, Geyer, Lang, and Funo (1998) and Dahinden, Kalisch, and Bühlmann (2010).

## Statistical Properties of the Log-Linear Models

Consider statistical properties of the hierarchical LLMs, like the existence of sufficient statistics, uniqueness of the MLE, and model testing.

### The Sufficient Statistics for LLMs

The sufficient statistics exist for the hierarchical LLMs and are very easy to obtain. Consider the saturated model with simple multinomial sampling distribution for the three-factor contingency tables. The log-likelihood function of the multinomial is obtained from the pdf given by equation (1) as follows:

$$\log(f(\{n_{ijk}\})) = \log\left(\frac{N!}{\prod_i \prod_j \prod_k n_{ijk}}\right) + \sum_i \sum_j \sum_k n_{ijk} \log(m_{ijk}) - N \log N \quad (5)$$

Or, equivalently,

$$\log(f(\{n_{ijk}\})) = \sum_i \sum_j \sum_k n_{ijk} \log(m_{ijk}) + C \quad (6)$$

where  $C$  represents the constant terms. Substituting the value for  $\log(m_{ijk})$  as given by equation (4),

$$\log(f(\{n_{ijk}\})) = \sum_i \sum_j \sum_k n_{ijk} (u + u_1 + u_2 + u_3 + u_{12} + u_{13} + u_{23} + u_{123}) + C$$

The above expression can be also written as

$$\begin{aligned} f(\{n_{ijk}\}) = \exp & \left( Nu + \sum_i u_1 n_{i..} + \sum_j u_2 n_{.j.} + \sum_k u_3 n_{..k} + \sum_i \sum_j u_{12} n_{ij.} + \sum_i \sum_k u_{13} n_{i.k} \right) \\ & + \sum_j \sum_k u_{23} n_{.jk} + \sum_i \sum_j \sum_k u_{123} n_{ijk} + C \end{aligned}$$

Because the multinomial distribution belongs to exponential family sufficient statistic exists, see E. B. Andersen (1970). From the above expression it is apparent that, for the three-factor saturated model, the full table itself is the sufficient statistic since the lower-order terms are redundant and it will be

subsumed in the full table. The marginal sub-tables which correspond to the set of generating classes are the sufficient statistics for the log-linear models (see Birch, 1963).

**Example 4:** Consider a four-factor table with the following generating classes:

$$\{C_1, C_2\} = \{[123], [34]\}$$

Then  $C_1(n) = [n_{ijk.}]$  is a three-dimensional marginal sub-table and  $C_2(n) = [n_{.kl}]$  is a two-dimensional marginal sub-table. These two marginal sub-tables are the sufficient statistics for this model. For more details and proofs on the sufficient statistics for hierarchical LLMs, see Haberman (1973).

## Maximum Likelihood Estimates for the LLMs

A unique set of MLEs for every cell count can be obtained from the sufficient statistics alone; see Birch (1963) for the proof. The Birch criteria are:

1. The marginal sub-tables obtained by summing over the factors not present in the max-cliques are the sufficient statistics for the corresponding expected cell counts. e.g., for the model [123] [34],  $C_1(n) = [n_{ijk.}]$  and  $C_2(n) = [n_{.kl}]$  are sufficient statistics for  $m_{ijk.}$  and  $m_{.kl}$ , respectively.
2. All the sufficient statistics must be the same as the corresponding marginal sub-tables of their estimate means.

$$C_i(\hat{m}) = C_i(n)$$

for all  $i$  from 1 to the number of generating classes. e.g., for the model [123] [34], the estimated cell counts are

$$\begin{aligned}\hat{m}_{ijk.} &= n_{ijk.} \\ \hat{m}_{.kl} &= n_{.kl}\end{aligned}$$

Finally, the MLE of the expected cell counts for the model [123] [34] is expressed as follows:

## GRAPHICAL LOG-LINEAR MODELS

$$\hat{m}_{ijkl} = \frac{n_{ijk} \cdot n_{..kl}}{n_{..k}}$$

The closed form expressions for the MLEs will be derived below in terms of sufficient statistics for three-factor contingency tables.

The reason for choosing MLE for computing the expected cell counts is its consistency and efficiency in large samples. There is extensive research on the MLEs of LLMs; see for example Glonek, Darroch, and Speed (1988), A. H. Andersen (1974), Haberman (1974), Meeden, Geyer, Lang, and Funo (1998), Birch (1963), Fienberg and Rinaldo (2007), Lang (1996a), Lang, McDonald, and Smith (1999), and Darroch (1962).

### Testing Models

The assessment of a model's fit is very important as it describes how well it fits the data. Consider the following test statistics:

#### Pearson's $\chi^2$ Statistic

This is defined as

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

where the  $O_i$  denote the observed cell counts and the  $E_i$  the expected cell counts.

#### The Deviance Goodness-of-Fit Test Statistics

Test a model against the saturated model using the deviance goodness-of-fit test, which is defined as follows:

$$G^2 = -2 \sum_i O_i \log \frac{E_i}{O_i}$$

Under the null hypotheses, the deviance is also distributed as  $\chi^2$  with the appropriate degrees of freedom.

Significance of a test statistic is assessed by its  $p$ -value. Statistical significance is attained when the  $p$ -value is less than a predetermined minimum

level of significance, say  $\alpha$ . The significance level  $\alpha$  is often set at 0.05 or 0.01 (see Bishop, Fienberg, & Holland, 1989). Here the level  $\alpha$  is set at 0.05.

In Table 2, the degrees of freedom of all the possible models for three-factor tables are listed. For more information about the model testing refer to Davis (1968), Kreiner (1987), and Landis, Heyman, and Koch (1978).

### Analysis of Three-Factor Contingency Tables

Consider the different interaction models for three-factor tables and the mathematical formulation for the MLE of the expected counts (when it is possible) for each model.

**Table 2.** Degrees of freedom

Model	DF
[1][2][3]	$IJK - I - J - K + 2$
[12][3]	$(J - 1)(K - 1)$
[13][2]	$(IK - 1)(J - 1)$
[23][1]	$(JK - 1)(I - 1)$
[12][13]	$I(J - 1)(K - 1)$
[12][23]	$J(I - 1)(K - 1)$
[13][23]	$K(I - 1)(J - 1)$
[12][13][23]	$(I - 1)(J - 1)(K - 1)$
[123]	0

### Complete Independence Model

This is the simplest model where all the factors are mutually independent and  $u_{12} = u_{13} = u_{23} = u_{123} = 0$ . The following different equivalent notations can be used to represent this model:

$$\begin{aligned}
 & X \perp\!\!\!\perp Y \mid Z \\
 \log(m_{ijk}) &= u + u_1 + u_2 + u_3 \\
 C &= \{[1], [2], [3]\}
 \end{aligned}
 \tag{7}$$

This model can be represented graphically as given in Figure 3.

Substitute the value of  $\log(m_{ijk})$ , as given in the equation (4) to the log-likelihood kernel as given by the Equation (6) and ignoring the constant term:

## GRAPHICAL LOG-LINEAR MODELS

$$\begin{aligned}\log\left(f\left(\{n_{ijk}\}\right)\right) &= \sum_{ijk} n_{ijk} \log\left(m_{ijk}\right) \\ &= \sum_{ijk} n_{ijk} (u + u_1 + u_2 + u_3)\end{aligned}$$

After simplification, obtain

$$f\left(\{n_{ijk}\}\right) = \exp\left(Nu + \sum_j u_1 n_{i..} + \sum_j u_2 n_{.j.} + \sum_k u_3 n_{..k}\right)$$

From the above expression, obtain the sufficient statistics for this models as marginal sub-tables:  $C_1 = \{n_{i..}\}$ ,  $C_2 = \{n_{.j.}\}$ , and  $C_3 = \{n_{..k}\}$ , which are estimates of  $m_{i..}$ ,  $m_{.j.}$ , and  $m_{..k}$ , respectively.

From equation (7), by summing over  $jk$ ,  $ik$ ,  $ij$ , and  $ijk$ , we obtain  $m_{i..}$ ,  $m_{.j.}$ ,  $m_{..k}$ , and  $m_{...}$  as

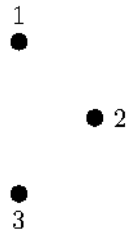
$$\begin{aligned}\{m_{i..}\} &= \exp(u + u_1) \sum_{jk} \exp(u_2 + u_3) \\ &= \exp(u + u_1) \sum_j \exp(u_2) \sum_k \exp(u_3) \\ \{m_{.j.}\} &= \exp(u + u_2) \sum_i \exp(u_1 + u_3) \\ &= \exp(u + u_2) \sum_i \exp(u_1) \sum_k \exp(u_3) \\ \{m_{..k}\} &= \exp(u + u_3) \sum_i \sum_j \exp(u_1 + u_2) \\ &= \exp(u + u_3) \sum_i \exp(u_1) \sum_j \exp(u_2) \\ \{m_{...}\} &= \exp(u) \sum_i \sum_j \sum_k \exp(u_1 + u_2 + u_3) \\ &= \exp(u) \sum_i \exp(u_1) \sum_j \exp(u_2) \sum_k \exp(u_3)\end{aligned}$$

From the above equations, get the expression for  $m_{ijk}$  as

$$m_{ijk} = \frac{m_{i..} m_{.j.} m_{..k}}{(m_{...})^2}$$

Applying Birch's result, the estimates of  $m_{ijk}$  are

$$\hat{m}_{ijk} = \frac{n_{i..} n_{.j.} n_{...k}}{(n_{...})^2}$$



**Figure 3.** The complete independence model

**Table 3.** Personality type, cholesterol, and DBP marginal sub-tables of Table 1

Personality Type		Cholesterol		Diastolic Blood Pressure	
A	1027	Normal	1681	Normal	1928
B	1094	High	440	High	193

**Table 4.** Table of estimated cell counts for Example 4

Personality Type	Cholesterol	Diastolic Blood Pressure	
		Normal	High
A	Normal	739.90	74.07
	High	193.70	19.39
B	Normal	788.20	78.90
	High	206.30	20.65

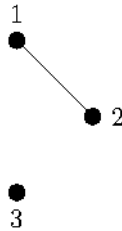
**Example 4:** Consider the contingency table as given in Table 1. Under the complete independence assumption, the sufficient statistics are the marginal sub-tables given in Table 3. The table of fitted values, under the complete independence assumption, is given in Table 4. The  $G^2$  statistic for the model is 8.723 (df: 4,  $p$ -value: 0.068), hence we conclude that the data supports the complete independence model. For details on the Chi-Squared test of independence, refer to Goodman (1971b).

**Joint Independence Model**

Under this model, two factors are jointly independent of the third factor. There are three versions of this model depending on which factor is unrelated to the other two. These three models are  $(XY) \perp\!\!\!\perp Z$ ,  $(XZ) \perp\!\!\!\perp Y$ , and  $(YZ) \perp\!\!\!\perp X$ . Consider only  $(XY) \perp\!\!\!\perp Z$  in detail as the others are comparable. Equivalent different notations are

$$\begin{aligned} \log(m_{ijk}) &= u + u_1 + u_2 + u_3 + u_{12} \\ C &= \{[12], [3]\} \end{aligned} \tag{8}$$

This model can also be represented graphically, as given in Figure 4.



**Figure 4.** The joint independence model.

---

The sufficient statistics for this model are the marginal sub-tables  $C_1 = \{n_{ij}\}$  and  $C_2 = \{n_{..k}\}$ , which are the estimates of  $m_{ij}$  and  $m_{..k}$ . From equation (8), obtain

$$\begin{aligned} m_{ij.} &= \exp(u + u_1 + u_2 + u_{12}) \sum_k \exp(u_3) \\ m_{..k} &= \exp(u + u_3) \sum_i \sum_j \exp(u_1 + u_2 + u_{12}) \\ m_{...} &= \exp(u) \sum_i \sum_j \exp(u_1 + u_2 + u_{12}) \sum_k \exp(u_3) \end{aligned}$$

From the above equations, derive the closed form expression for  $m_{ijk}$  as

$$m_{ijk} = \frac{m_{ij.} m_{..k}}{m_{...}}$$

and, applying Birch's criteria,

$$\hat{m}_{ijk} = \frac{n_{ij.} \cdot n_{.k}}{n_{...}}$$

If the previous model of the complete independence  $X \perp\!\!\!\perp Y \perp\!\!\!\perp Z$  fits a data set, then the model,  $(XY) \perp\!\!\!\perp Z$  will also fit. But the smallest model will be preferred.

**Example 5:** Consider the contingency table displayed in Table 5 to discuss this model. The sufficient statistics are given in Table 6. Under the assumptions of this model, the table of the expected cell counts is given in Table 7. The  $G^2$  statistic for this model is 5.560 (df: 5,  $p$ -value: 0.351), hence we conclude that the data supports the joint independence model.

**Table 5.** Classroom behaviour table (Everitt, 1977)

Classroom Behaviour	Adversity of School	Risk	
		Not at Risk	At Risk
Nondeviant	Low	16	7
	Medium	15	34
	High	5	3
Deviant	Low	1	1
	Medium	3	8
	High	1	3

**Table 6.** Adversity\*risk and classroom behaviour marginal sub-tables of Table 5

Adversity	Risk		Classroom Behaviour	Total
	Not at Risk	At Risk		
Low	17	8	Nondeviant	80
Medium	18	42	Deviant	17
High	6	6		

**Table 7.** Table of estimated cell counts for Example 5

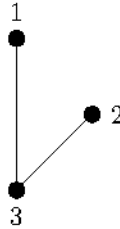
Classroom Behaviour	Adversity of School	Risk	
		Not at Risk	At Risk
Nondeviant	Low	14.020	6.597
	Medium	14.845	34.639
	High	4.948	4.948
Deviant	Low	2.979	1.402
	Medium	3.154	7.360
	High	1.051	1.051

**Conditional Independence Model**

Under this model, two factors are conditionally independent given the third factor. There are three version for this model as well, these are  $X \perp\!\!\!\perp Y \mid Z$ ,  $X \perp\!\!\!\perp Z \mid Y$ , and  $Y \perp\!\!\!\perp Z \mid X$ . Consider only  $X \perp\!\!\!\perp Y \mid Z$  in detail, as derivation for the others is similar. This model can be equivalently represented as

$$\begin{aligned} \log(m_{ijk}) &= u + u_1 + u_2 + u_3 + u_{13} + u_{23} \\ C &= \{[13],[23]\} \end{aligned} \tag{9}$$

The graph for this model is given in Figure 5.



**Figure 5.** The conditional independence model

The sufficient statistics for this model are the marginal sub-tables  $C_{13} = n_{i,k}$  and  $C_{23} = n_{,jk}$ , which are estimates of  $m_{i,k}$  and  $m_{,jk}$ . From equation (9):

$$\begin{aligned} m_{i,k} &= \exp(u + u_1 + u_3 + u_{13}) \sum_j \exp(u_2 + u_{23}) \\ m_{,jk} &= \exp(u + u_2 + u_3 + u_{23}) \sum_i \exp(u_1 + u_{13}) \\ m_{..k} &= \exp(u + u_3) \sum_i \exp(u_1 + u_{13}) \sum_j \exp(u_2 + u_{23}) \end{aligned}$$

From the above three equations, obtain the closed form expression for  $m_{ijk}$  as

$$m_{ijk} = \frac{m_{ij.} m_{.jk}}{m_{..k}}$$

As before, applying Birch's criteria derive the expected counts for each cell as

$$\hat{m}_{ijk} = \frac{n_{ij.} n_{.jk}}{n_{..k}}$$

**Example 6:** Consider Table 8, infant's survival data taken from Bishop (1969). Assuming pre-natal care and survival are independent given a clinic, the sufficient statistics are given in Table 9. The  $G^2$  statistic for this model is 0.082 (df: 2,  $p$ -value: 0.959), hence we conclude that the data supports the conditional independence model.

**Table 8.** Infant survival table

Clinic	Pre-natal care	Infant's Survival	
		Died	Survived
A	Less	3	176
	More	4	293
B	Less	17	197
	More	2	23

**Table 9.** Survival\*clinic, clinic\*pre-natal care, and clinic marginal sub-tables of Table 8

Infant's Survival			Pre-natal Care			Clinic	Total
Clinic	Died	Survived	Clinic	Less	More		
A	7	469	A	179	297	A	476
B	19	220	B	214	25	B	239

**Table 10.** Table of estimated cell counts for Example 6

Clinic	Pre-natal care	Infant's Survival	
		Died	Survived
A	Less	2.632	176.367
	More	4.367	292.632
B	Less	17.012	196.987
	More	1.987	23.012

### Uniform Association Model

This model is also known as the no three-factor interaction model, where  $u_{123} = 0$ . For this model the log-linear notation is [12] [13] [23], but there is no graphical representation for this model. Unlike the previous models, there are no closed-

## GRAPHICAL LOG-LINEAR MODELS

form estimates for the expected cell counts/probabilities under this model. The MLEs can be computed by iterative procedures such as Iterative Proportional Fitting (IPF) and the Newton-Raphson method.

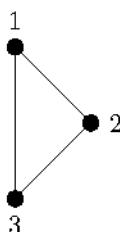
**Example 7:** Consider Table 11, auto accident data taken from Fienberg (1970). None of the models discussed in previous sections fit the data. Use the IPF algorithm to obtain the table of estimated counts as given in the Table 12. The  $G^2$  statistic for this model is 0.043 (df: 1,  $p$ -value: 0.835), hence we conclude the data supports the marginal association model. For more information on IPF, refer to Deming and Stephan (1940) and Fienberg (1970). The IPF procedure implemented in the R package `cat` was used, available at [cran.r-project.org](http://cran.r-project.org).

**Table 11.** Auto accident data table

Accident Type	Driver Ejected	Injury	
		Not Severe	Severe
Collision	No	350	150
	Yes	26	23
RollOver	No	60	112
	Yes	19	80

**Table 12.** Table of estimated cell counts for Example 7

Accident Type	Driver Ejected	Injury	
		Not Severe	Severe
Collision	No	350.48858	149.51130
	Yes	25.51142	23.48870
RollOver	No	59.51104	112.48921
	Yes	19.48896	79.51079



**Figure 6.** The saturated model

**Saturated Model**

For this model, the log-linear notation is [123]. In this case there is no independence relationship between the three factors. The expected cell counts are the same as the observed cell frequencies, e.g.  $\hat{m}_{ijk} = n_{ijk}$ . Graphical representation for the saturated model is given in Figure 6.

*Example 8:* Consider Table 13, a partial table which is based on clinical trial data from Koch, Amara, Atkinson, and Stanish (1983). None of the models fit the data; we leave this for the reader to verify.

**Table 13.** Results of a clinical trial for the effectiveness of an analgesic drug

Status	Treatment	Response		
		Poor	Moderate	Excellent
1	Active	3	20	5
	Placebo	11	4	8
2	Active	3	14	12
	Placebo	6	13	5

**Model Selection for Decomposable Models**

Model selection is now discussed for the decomposable models only, as a non-decomposable graphical model can be reduced to a decomposable one by adding a minimal number of edges to the graph. For details on minimum triangulation, refer to Rose, Tarjan, and Lueker (1970) and Heggernes (2006).

Though decomposable models are a restricted family of GLLMs, selecting an optimal model from the class of decomposable graphical models is known to be an intractable problem. Most of all existing model selection algorithms are based on forward selection, backward elimination, or a combination of the both. There is a vast literature available for model selection and inference on graphical models, e.g. see Wainwright and Jordan (2008), Dahinden, Kalisch, and Bühlmann (2010), Goodman (1971a), Ravikumar, Wainwright, and Lafferty (2010), and Allen and Liu (2012).

The Wermuth's procedure starts with the saturated model, a single clique that includes all the two-factor effects as given in Figure 7. The vertices  $a, b, c, d, e,$  and  $f$  correspond to the factors Attendance, Sex, School, Agree, Subject, and Plans, respectively.

## GRAPHICAL LOG-LINEAR MODELS

Consider the backward model selection procedure for a real data set called women and mathematics (WAM), used in Fowlkes, Freeny, and Landwehr (1988). Wermuth's (1976) backward elimination algorithm is used. The data are shown in the Table 14.

Graphical models are completely specified by their two-factor interactions. By the hierarchical principle, if a two-factor term is set to zero, then any higher-order term that contain that particular two-factor term will also be set to zero.

**Table 14.** The women and mathematics data table

		School	Suburban School			
		Sex	Female		Male	
Plan		Preference	Attend	Not	Attend	Not
College	Maths-Sciences	Agree	37	27	51	48
		Disagree	16	11	10	19
	Liberal arts	Agree	16	15	7	6
		Disagree	12	24	13	7
Job	Maths-Sciences	Agree	10	8	12	15
		Disagree	9	4	8	9
	Liberal arts	Agree	7	10	7	3
		Disagree	8	4	6	4

		School	Urban School			
		Sex	Female		Male	
Plan		Preference	Attend	Not	Attend	Not
College	Maths-Sciences	Agree	51	55	109	86
		Disagree	24	28	21	25
	Liberal arts	Agree	32	34	30	31
		Disagree	55	39	26	19
Job	Maths-Sciences	Agree	2	1	9	5
		Disagree	8	9	4	5
	Liberal arts	Agree	5	2	1	3
		Disagree	10	9	3	6

In the next step, all the  $\binom{6}{2}$  two-factor interactions are considered for elimination. Fix a backward elimination cut off level,  $\alpha = 0.05$ . Among the two-factor interactions, the terms having the largest  $p$ -value are considered for elimination, but only if the  $p$ -value exceeds  $\alpha$ . From the Table 15, choose the edge  $(bf)$  for deletion, and the resulting graphical model is  $[abcde] [acdef]$ .

In the next step, consider the cliques  $[abcde]$  and  $[acdef]$ . The edges  $ac$ ,  $ad$ ,  $ae$ ,  $cd$ ,  $ce$ , and  $de$  are common to both the cliques; they are not considered for

elimination because elimination of such edges may result in a non-decomposable model. The candidate edges for deletion are  $ab$ ,  $bc$ ,  $bd$ ,  $be$ ,  $af$ ,  $cf$ ,  $df$ , and  $ef$ . Let us examine the  $p$ -values for these edges as in the Table 16.

Delete the edge ( $af$ ); the resulting graphical model is  $[abcde][cdef]$ . Similarly, in the next step, the edge ( $ad$ ) gets deleted and the resulting graphical model becomes  $[abcde][bcde][cdef]$  as given in Figure 8.

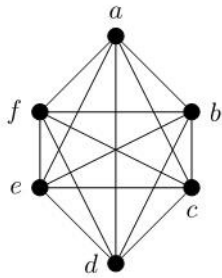


Figure 7. The saturated model for WAM

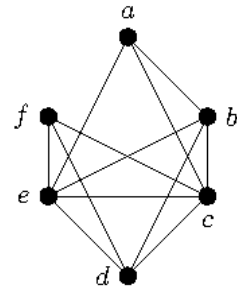


Figure 8. The fitted model for WAM

Table 15. WAM:  $[abcde]$

Edge	Clique	d.f.	$G^2$	$p$ -value
$ab$	$[acdef][bcdef]$	16	18.585	0.29078
$ac$	$[acdef][bcdef]$	16	20.689	0.19080
$ad$	$[acdef][bcdef]$	16	14.172	0.58588
$ae$	$[acdef][bcdef]$	16	18.781	0.28017
$af$	$[abcde][bcdef]$	16	11.951	0.74734
$bc$	$[acdef][abdef]$	16	26.739	0.04447
$bd$	$[acdef][abcef]$	16	34.733	0.00432
$be$	$[acdef][abcdf]$	16	56.570	0.00000
$bf$	$[acdef][abcde]$	16	11.673	0.76616
$cd$	$[abcef][abdef]$	16	29.439	0.02114
$ce$	$[abcdf][abdef]$	16	26.052	0.05329
$cf$	$[abcde][abdef]$	16	81.657	0.00000
$de$	$[abcdf][abcef]$	16	78.248	0.00000
$df$	$[abcef][abcde]$	16	46.221	0.00009
$ef$	$[abcde][abcde]$	16	17.728	0.34005

## GRAPHICAL LOG-LINEAR MODELS

**Table 16.** WAM:  $[abcde]$   $[acdef]$

Edge	Clique	d.f.	$G^2$	$p$ -value
<i>ab</i>	$[bcde]$ $[acdef]$	8	12.456	0.13198
<i>bc</i>	$[acde]$ $[acdef]$	8	18.097	0.02051
<i>bd</i>	$[acde]$ $[acdef]$	8	27.358	0.00061
<i>be</i>	$[acde]$ $[acdef]$	8	49.723	0.00000
<i>af</i>	$[abcde]$ $[cdef]$	8	5.822	0.66711
<i>cf</i>	$[abcde]$ $[acdef]$	8	73.014	0.00000
<i>df</i>	$[abcde]$ $[acef]$	8	38.845	0.00001
<i>ef</i>	$[abcde]$ $[acdf]$	8	10.881	0.20852

**Table 17.** WAM:  $[abce]$   $[bcde]$   $[cdef]$

Edge	Clique	d.f.	$G^2$	$p$ -value
<i>ab</i>	$[ace]$ $[bce]$ $[bcde]$ $[cdef]$	4	10.606	0.03137
<i>ac</i>	$[bce]$ $[ace]$ $[bcde]$ $[cdef]$	4	10.432	0.03374
<i>ae</i>	$[bce]$ $[abc]$ $[bcde]$ $[cdef]$	4	10.426	0.03383
<i>bd</i>	$[abce]$ $[cde]$ $[bce]$ $[cdef]$	4	25.507	0.00004
<i>cf</i>	$[abce]$ $[bcde]$ $[def]$ $[i]$	4	67.832	0.00000

In the next step, candidate edges for deletion are  $[ab]$ ,  $[ac]$ ,  $[ae]$ ,  $[bd]$ , and  $[cf]$ . None of the  $p$ -values are greater than  $\alpha = 0.05$  as given in Table 17. So, stop with the model  $[abce]$   $[bcde]$   $[cdef]$ .

### Computational Details

All the experimental results were carried out using R 3.1.3. For fitting LLMs, there are several function in R, for example `glm()` and `loglin()` in the stats library and `loglm()` in the MASS library. For model selection, `dmod()` and `backward()` functions implemented in the package `gRim` were used. All the packages used are available at <http://CRAN.R-project.org/>.

### Conclusion

The fundamental mathematical and statistical theory of GLLM and its applications were discussed, restricted to the complete table to make the discussion simple, because the tables having zero entries require special treatment. See Christensen (1997) for analysis of contingency tables with zero cell counts.

The limitations and open problems in the use of GLLM for recursive relationships can be further explored.

## References

- Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York, NY: Wiley-Interscience. doi: 10.1002/0471249688
- Allen, G. I., & Liu, Z. (2012, October). *A log-linear graphical model for inferring genetic networks from high-throughput sequencing data*. Paper presented at the 2012 IEEE International Conference on Bioinformatics and Biomedicine, Philadelphia, PA. doi: 10.1109/bibm.2012.6392619
- Andersen, A. H. (1974). Multidimensional contingency tables. *Scandinavian Journal of Statistics*, 1(3), 115-127. Available from <http://www.jstor.org/stable/4615563>
- Andersen, E. B. (1970). Sufficiency and exponential families for discrete sample spaces. *Journal of the American Statistical Association*, 65(331), 1248-1255. doi: 10.2307/2284291
- Bartlett, M. S. (1935). Contingency table interactions. *Supplement to the Journal of the Royal Statistical Society*, 2(2), 248-252. doi: 10.2307/2983639
- Birch, M. W. (1963). Maximum likelihood in three-way contingency tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 25(1), 220-233. Available from <http://www.jstor.org/stable/2984562>
- Bishop, Y. M. (1969). Full contingency tables, logits, and split contingency tables. *Biometrics*, 25(2), 383-400. doi: 10.2307/2528796
- Bishop, Y. M., Fienberg, S. E., & Holland, P. W. (1989). *Discrete multivariate analysis: Theory and practice*. Cambridge, MA: MIT Press.
- Christensen, R. (1997). *Log-linear models and logistic regression* (2nd ed.). New York, NY: Springer. doi: 10.1007/b97647
- Dahinden, C., Kalisch, M., & Bühlmann, P. (2010). Decomposition and model selection for large contingency tables. *Biometrical Journal*, 52(2), 233-252. doi: 10.1002/bimj.200900083
- Darroch, J. N. (1962). Interactions in multi-factor contingency tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 24(1), 251-263. Available from <http://www.jstor.org/stable/2983765>

## GRAPHICAL LOG-LINEAR MODELS

Darroch, J. N., Lauritzen, S. L., & Speed, T. P. (1980). Markov fields and log-linear interaction models for contingency tables. *The Annals of Statistics*, 8(3), 522-539. doi: 10.1214/aos/1176345006

Davis, L. J. (1968). Exact tests for  $2 \times 2$  contingency tables. *The American Statistician*, 40(2), 139-141. doi: 10.2307/2684874

Dawid, A. P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(1), 1-31. Available from <http://www.jstor.org/stable/2984718>

Deming, W. E., & Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics*, 11(4), 427-444. doi: 10.1214/aoms/1177731829

Edwards, D. (2000). *Introduction to graphical modeling* (2nd ed.). New York, NY: Springer-Verlag. doi: 10.1007/978-1-4612-0493-0

Fienberg, S. E. (1970). An iterative procedure for estimation in contingency tables. *The Annals of Mathematical Statistics*, 41(3), 901-917. doi: 10.1214/aoms/1177696968

Fienberg, S. E., & Rinaldo, A. (2007). Three centuries of categorical data analysis: Log-linear models and maximum likelihood estimation. *Journal of Statistical Planning and Inference*, 137(11), 3430-3445. doi: 10.1016/j.jspi.2007.03.022

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 222(594-604), 309-368. doi: 10.1098/rsta.1922.0009

Fowlkes, E. B., Freeny, A. E., & Landwehr, J. M. (1988). Evaluating logistic models for large contingency tables. *Journal of the American Statistical Association*, 83(403), 611-622. doi: 10.2307/2289283

Glonek, G. F., Darroch, J. N., & Speed, T. P. (1988). On the existence of maximum likelihood estimators for hierarchical loglinear models. *Scandinavian Journal of Statistics*, 15(3), 187-193. Available from <http://www.jstor.org/stable/4616100>

Goodman, L. A. (1969). How to ransack social mobility tables and other kinds of cross-classification tables. *American Journal of Sociology*, 75(1), 1-40. doi: 10.1086/224743

- Goodman, L. A. (1970). The multivariate analysis of qualitative data: Interaction among multiple classifications. *Journal of the American Statistical Association*, 65(329), 226-256. doi: 10.2307/2283589
- Goodman, L. A. (1971a). The analysis of multidimensional contingency tables: Stepwise procedures and direct estimation methods for building models for multiple classifications. *Technometrics*, 13(1), 31-66. doi: 10.2307/1267074
- Goodman, L. A. (1971b). The partitioning of chi-square, the analysis of marginal contingency tables, and the estimation of expected frequencies in multidimensional contingency tables. *Journal of the American Statistical Association*, 66(334), 339-344. doi: 10.2307/2283933
- Haberman, S. J. (1973). Log-linear models for frequency data: Sufficient statistics and likelihood equations. *The Annals of Statistics*, 1(4), 617-632. doi: 10.1214/aos/1176342458
- Haberman, S. J. (1974). *The analysis of frequency data*. Chicago, IL: University of Chicago Press.
- Heggernes, P. (2006). Minimal triangulations of graphs: A survey. *Discrete Mathematics*, 306(3), 297-317. doi: 10.1016/j.disc.2005.12.003
- Knoke, D., & Burke, P. J. (1980). *Log-linear models*. Beverly Hills, CA: Sage. doi: 10.4135/9781412984843
- Koch, G. G., Amara, I., Atkinson, S., & Stanish, W. (1983). *Overview of categorical analysis methods*. Paper presented at SAS Users Group International '83, New Orleans, LA.
- Kreiner, S. (1987). Analysis of multidimensional contingency tables by exact conditional tests: Techniques and strategies. *Scandinavian Journal of Statistics*, 14(2), 97-112. Available from <http://www.jstor.org/stable/4616054>
- Kreiner, S. (1998). Interaction model. In *Encyclopedia of Biostatistics*. Chichester, UK: Wiley.
- Landis, J. R., Heyman, E. R., & Koch, G. G. (1978). Average partial association in three-way contingency tables: A review and discussion of alternative tests. *International Statistics Review*, 46(3), 237-254. doi: 10.2307/1402373
- Lang, J. B. (1996a). Maximum likelihood methods for a generalized class of log-linear models. *The Annals of Statistics*, 24(2), 726-752. doi: 10.1214/aos/1032894462

## GRAPHICAL LOG-LINEAR MODELS

Lang, J. B. (1996b). On the comparison of multinomial and Poisson log-linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 253-266. Available from <http://www.jstor.org/stable/2346177>

Lang, J. B., McDonald, J. W., & Smith, P. W. (1999). Association-marginal modeling of multivariate categorical responses: A maximum likelihood approach. *Journal of the American Statistical Association*, 94(448), 1161-1171. doi: 10.2307/2669932

Lauritzen, S. L. (1996). *Graphical models* (2nd ed.). New York, NY: Oxford University Press, Inc.

Lauritzen, S. L., Speed, T. P., & Vijayan, K. (1984). Decomposable graphs and hypergraphs. *Journal of the Australian Mathematical Society*, 36(1), 12-29. doi: 10.1017/s1446788700027300

Meeden, G., Geyer, C., Long, J., & Funo, E. (1998). The admissibility of the maximum likelihood estimator for decomposable log-linear interaction models for contingency tables. *Communications in Statistics – Theory and Methods*, 27(2), 473-493. doi: 10.1080/03610929808832107

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3), 370-384. doi: 10.2307/2344614

Pearl, J., & Paz, A. (1987). Graphoids: A graph based logic for reasoning about relevance relations. *Advances in Artificial Intelligence*, 2, 357-363.

Pearson, K. (1904). *Mathematical contributions to the theory of evolution*. London, UK: Dulau and Co.

Ravikumar, P., Wainwright, M. J., & Lafferty, J. (2010). High-dimensional Ising model selection using  $\ell_1$ -regularized logistic regression. *The Annals of Statistics*, 38(3), 1287-1319. doi: 10.1214/09-aos691

Rose, D., Tarjan, R. E., & Lueker, G. (1976). Algorithmic aspects of vertex elimination on graphs. *SIAM Journal on Computing*, 5(2), 146-160. doi: 10.1137/0205021

Wainwright, M. J., & Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2), 1-305. doi: 10.1561/22000000001

Wermuth, N. (1976). Model search among multiplicative models. *Biometrics*, 32(2), 253-263. doi: 10.2307/2529496

West, D. B. (2000). *Introduction to graph theory* (2nd ed.). Cambridge, MA: MIT Press.

Zelterman, D. (2006). *Models for discreet data* (2nd ed.). New York, NY: Oxford University Press, Inc.

### Appendix A: Graphical Log-Linear Models for Four-Way Tables

Table 18. Graphical log-linear models for four-way tables

Model	Graph	Closed-Form Estimate
[1] [2] [3] [4]		$\hat{m}_{hijk} = \frac{n_{h...} n_{...j} n_{...k}}{n^3_{...}}$
[12] [3] [4]		$\hat{m}_{hijk} = \frac{n_{hi...} n_{...j} n_{...k}}{n^2_{...}}$
[12] [13] [4]		$\hat{m}_{hijk} = \frac{n_{hi...} n_{h..j} n_{...k}}{n_{h...} n_{...}}$
[12] [34]		$\hat{m}_{hijk} = \frac{n_{hi..} n_{..jk}}{n_{...}}$
[12] [13] [14]		$\hat{m}_{hijk} = \frac{n_{hi..} n_{h..j} n_{h..k}}{n^2_{h...}}$

Table 18, continued.

Model	Graph	Closed-Form Estimate
[12] [23] [34]		$\hat{m}_{hijk} = \frac{n_{hi..} n_{ij.} n_{..jk}}{n_{i..} n_{.j.}}$
[123] [4]		$\hat{m}_{hijk} = \frac{n_{hij.} n_{...k}}{n_{...}}$
[123] [14]		$\hat{m}_{hijk} = \frac{n_{hij.} n_{h..k}}{n_{h...}}$
[12] [23] [34] [14]		No closed-form estimate
[123] [134]		$\hat{m}_{hijk} = \frac{n_{hij.} n_{h..k}}{n_{h..j.}}$
[1234]		$\hat{m}_{hijk} = n_{hijk}$