Relational models for contingency tables

Anna Klimova[a], Tamás Rudas[b], and Adrian Dobra[c]

[a]Corresponding Author,
Department of Statistics, University of Washington,
Box 355845, Seattle WA 98195-4322, USA;
e-mail: klimova@u.washington.edu
[b]Department of Statistics, Eötvös Loránd University,
Pazmany Peter setany 1/A, H-1117, Budapest, Hungary;
e-mail: rudas@tarki.hu
[c]Department of Statistics, University of Washington,
Box 354322, Seattle WA 98195-4322, USA;
e-mail: adobra@u.washington.edu

Abstract

The paper considers general multiplicative models for complete and incomplete contingency tables that generalize log-linear and several other models and are entirely coordinate free. Sufficient conditions for the existence of maximum likelihood estimates under these models are given, and it is shown that the usual equivalence between multinomial and Poisson likelihoods holds if and only if an overall effect is present in the model. If such an effect is not assumed, the model becomes a curved exponential family and a related mixed parameterization is given that relies on non-homogeneous odds ratios. Several examples are presented to illustrate the properties and use of such models.

KEY WORDS: Contingency tables, curved exponential family, exponential family,
generalized odds ratios, maximum likelihood estimate,
multiplicative model

Introduction

The main objective of the paper is to develop a new class of models for the set of all strictly positive distributions on contingency tables and on some sets of cells that have a more general structure. The proposed relational models are motivated by traditional log-linear models, quasi models, and some other multiplicative models for discrete distributions that have been discussed in the literature.

Under log-linear models (Bishop et al., 1975), cell probabilities are determined by multiplicative effects associated with various subsets of the variables in the contingency table. However, some cells may have other characteristics in common, and there always has been interest in models that also allow for multiplicative effects that are associated with those characteristics. Examples, among others, include quasi models (Goodman, 1968, 1972), topological models (Hauser, 1978; Hout, 1983), indicator models (Zelterman & Youn, 1992), rater agreement-disagreement models (Tanner & Young, 1985a,b), two-way subtable sum models (Hara et al., 2009). All these models, applied in different contexts, have one common idea behind them. A model is generated by a class of subsets of cells, some of which may not be induced by marginals of the table, and, under the model, every cell probability is the product of effects associated with subsets the cell belongs to. This idea is generalized in the relational model framework.

The outline of the paper is as follows. The definition of a table and the definition of a relational model generated by a class of subsets of cells in the table are given in Section 1. The cells are characterized by strictly positive parameters (probabilities or intensities); a table is a structured set of cells. Under the model, the parameter of each cell is the product of effects associated with the subsets in the generating class, to which the cell belongs. Two examples are given to illustrate this definition. Example 1.1 shows how traditional log-linear models fit into the framework, and Example 1.2 describes how multiplicative models for incomplete contingency tables are handled.

The degrees of freedom and the dual representation of relational models are discussed in Section 2. Every relational model can be stated in terms of generalized odds ratios. The minimal number of generalized odds ratios required to specify the model is equal to the number of degrees of freedom of this model.

The models for probabilities that include the overall effect and all relational models for intensities are regular exponential families. Under known conditions (cf. Barndorff-Nielsen, 1978), the maximum likelihood estimates for cell frequencies exist and are unique; the mean-value parameters of the MLE, associated with the subsets of the model, are equal to the corresponding mean-value parameters of the observed distribution. The maximum likelihood estimates for cell frequencies under a model for intensities and under a model for probabilities, when the model matrix is the same, are equal if and only if the model for probabilities is a regular family. These facts are proven in Section 3.

The main results of the paper, given in Section 4, discuss the properties of the MLE under relational models without the overall effect. If the overall effect is not present, a relational model for probabilities forms a curved exponential family. The maximum likelihood estimates in the curved case exist and are unique under the same condition as for regular families. For any relational model, the mean-value parameters of the MLE, associated with the subsets of the model, are proportional to the corresponding mean-value parameters of the observed distribution. The parameter space and the relational models are also described in terms of algebraic geometry.

A mixed parameterization of finite discrete exponential families is discussed in Section 5. Any relational model is naturally defined under this parameterization: the corresponding generalized odds ratios are fixed and the model is parameterized by remaining mean-value parameters. The distributions of observed values of subset sums and generalized odds ratios are variation independent and, in the regular case, specify the table uniquely.

Two applications of the framework are presented in Section 6. These are analyses of social mobility data and of a valued network with given attributes. These two examples suggest that the flexibility of the framework and substantive interpretations of parameters make relational models appealing in many settings.

1 Definition and Log-linear Representation of Relational Models

Let $Y_1, \dots, Y_K$ be the discrete random variables modeling certain characteristics of the population of interest. Denote the domains of the variables by $\mathcal{Y}_1, \dots, \mathcal{Y}_K$ respectively. A point $(y_1, y_2, \dots, y_K) \in \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_K$ generates a cell if and only if the outcome $(y_1, y_2, \dots, y_K)$ appears in the population. A cell $(y_1, y_2, \dots, y_K)$ is called empty if the combination is not included in the design.

Let $\mathcal{I}$ denote the lexicographically ordered set of non-empty cells in $\mathcal{Y}_1 \times \cdots \times \mathcal{Y}_K$, and $|\mathcal{I}|$ denote the cardinality of $\mathcal{I}$. Since the case, when $\mathcal{I} = \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_K$, corresponds to a classical complete contingency table, then the set $\mathcal{I}$ is also called a table.

Depending on the procedure that generates data on $\mathcal{I}$, the population may be characterized by cell probabilities or cell intensities. The parameters of the true distribution will be denoted by $\boldsymbol{\delta} = \{\delta(i), \text{ for } i \in \mathcal{I}\}$. In the case of probabilities, $\delta(i) = p(i) \in (0, 1)$, where $\sum_{i \in \mathcal{I}} p(i) = 1$; in the case of intensities, $\delta(i) = \lambda(i) > 0$.

Write $\mathcal{P} = \{P_{\boldsymbol{\delta}} : \ \boldsymbol{\delta} \in \Omega\}$ for the set of all positive distributions on the table $\mathcal{I}$. Here the parameter space Ω is an open subset of $\mathbb{R}^{|\mathcal{I}|}$. Suppose $\Theta \subset \Omega$. Then the set $\mathcal{P}_\Theta = \{P_{\boldsymbol{\delta}} \in \mathcal{P} : \boldsymbol{\delta} \in \Theta\}$ is a model in $\mathcal{P}$.

Definition 1.1. Let $\mathbf{S} = \{S_1, \dots, S_J\}$ be a class of non-empty subsets of the table $\mathcal{I}$ and $\mathbf{A}$ be a $J \times |\mathcal{I}|$ matrix with entries

$$a_{ji} = \mathbf{I}_j(i) = \begin{cases} 1, & \text{if the } i\text{-th cell is in } S_j, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } i = 1, \dots, |\mathcal{I}| \text{ and } \ j = 1, \dots, J. \tag{1}$$

A relational model $RM(\mathbf{S})$ with the model matrix $\mathbf{A}$ is the following subset of $\mathcal{P}$:

$$RM(\mathbf{S}) = \{P_{\boldsymbol{\delta}} \in \mathcal{P} : \ \log \boldsymbol{\delta} = \mathbf{A}'\boldsymbol{\beta}, \text{ for some } \ \boldsymbol{\beta} \in \mathbb{R}^J\}. \tag{2}$$

Under the model (2) the parameters of the distribution can also be written as

$$\delta(i) = \exp \{\sum_{j=1}^{J} \mathbf{I}_j(i)\beta_j\} = \prod_{j=1}^{J} (\theta_j)^{\mathbf{I}_j(i)}, \tag{3}$$

where $\theta_j = \exp(\beta_j)$, for $j = 1, \dots, J$.

The parameters $\boldsymbol{\beta}$ in (2) are called the log-linear parameters. The parameters $\boldsymbol{\theta}$ in (3) are called the multiplicative parameters. If the subsets in $\mathbf{S}$ are cylinder sets, the parameters $\boldsymbol{\beta}$ coincide with the parameters of the corresponding log-linear model.

In the case $\boldsymbol{\delta} = \boldsymbol{p}$ it must be assumed that $\cup_{j=1}^{J} S_j = \mathcal{I}$, i.e. there are no zero columns in the matrix $\mathbf{A}$. A zero column implies that one of the probabilities is 1 under the model and the model is thus trivial.

The example below describes a model of conditional independence as a relational model.

Example 1.1. Consider the model of conditional independence $[Y_1Y_3][Y_2Y_3]$ of three binary variables Y_1, Y_2, Y_3, each taking values in $\{0, 1\}$. The model is expressed as

$$p_{ijk} = \frac{p_{i+k}p_{+jk}}{p_{++k}},$$

where $p_{i+k}, p_{+jk}, p_{++k}$ are marginal probabilities in the standard notation (Bishop et al., 1975). Let $\mathbf{S}$ be the class consisting of the cylinder sets associated with the empty marginal and with the marginals Y_1, Y_2, Y_3, Y_1Y_3, Y_2Y_3. The model matrix computed from (1) is not full row rank and thus the model parameters are not identifiable (cf. Section 2). A full row rank model matrix can be obtained by setting, for instance, the level 0 of each variable as the reference level. After that, the model matrix is equal to

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{4}$$

Table 1: Poisson intensities by bait type

	Fish	
Sugarcane	Yes	No
Yes	λ_{00}	λ_{01}
No	λ_{10}	-

The first row corresponds to the cylinder set associated with the empty marginal. The next three rows correspond to the cylinder sets generated by the level 1 of Y_1, Y_2, Y_3 respectively. The fifth row corresponds to the cylinder set generated by the level 1 for both Y_1 and Y_3, and the last row - to the cylinder set corresponding to the level 1 for both Y_2 and Y_3. $\square$

In the next example, one of the cells in the Cartesian product of the domains of the variables is empty and the sample space $\mathcal{I}$ is a proper subset of this product.

Example 1.2. The study described by Kawamura et al. (1995) compared three bait types for trapping swimming crabs: fish alone, sugarcane alone, and sugarcane-fish combination. The observed frequencies are given in Tables 2 and 3. During the experiment, catching crabs without bait was not considered. Three Poisson random variables are used to model the amount of crabs caught in the three traps. The notation for the intensities is shown in Table 1. The model assuming that there is a multiplicative effect of using both bait types at the same time will be tested in this paper. The hypothesis of interest is

$$\lambda_{00} = \lambda_{01}\lambda_{10}. \tag{5}$$

The effect can be tested using the relational model for intensities on the class $\mathbf{S}$ consisting of two subsets - $\mathbf{S} = \{S_1, S_2\}$, where $S_1 = \{(0,0),(0,1)\}$ and $S_2 = \{(0,0),(1,0)\}$:

$$\log \boldsymbol{\lambda} = \mathbf{A}'\boldsymbol{\beta}.$$

Here, the model matrix is

$$\mathbf{A} = \left(\begin{array}{ccc} 1 & 1 & 0 \\ 1 & 0 & 1 \end{array} \right),$$

and $\boldsymbol{\beta} = (\beta_1, \beta_2)'$. The relationship between the two forms of the model will be explored in the next section. $\square$

2 Parameterizations and Degrees of Freedom

A choice of subsets in $\mathbf{S} = \{S_1, \ldots, S_J\}$ is implied by the statistical problem, and the relational model $RM(\mathbf{S})$ can be parameterized with different model matrices, which may be

Table 2: Number of trapped *Charybdis japonica* by bait type

	Fish	
Sugarcane	Yes	No
Yes	36	2
No	11	-

Table 3: Number of trapped *Portunuspelagicus* by bait type

	Fish	
Sugarcane	Yes	No
Yes	71	3
No	44	-

useful depending on substantive meaning of the model. Sometimes a particular choice of subsets leads to a model matrix $\mathbf{A}$ with linearly dependent rows and thus non-identifiable model parameters. To ensure identifiability, a reparameterization, that is sometimes referred to as model matrix coding, is needed. Examples of frequently used codings are reference coding, effects coding, orthogonal coding, polynomial coding (cf. Christensen, 1997).

Write $R(\mathbf{A})$ for the row space of $\mathbf{A}$ and call it the design space of the model. The elements of $R(\mathbf{A})$ are $|\mathcal{I}|$-dimensional row-vectors and $\mathbf{1}$ denotes the row-vector with all components equal to 1. Reparameterizations of the model have form $\boldsymbol{\beta} = \mathbf{C}\boldsymbol{\beta}_1$, where $\boldsymbol{\beta}_1$ are the new parameters of the model and $\mathbf{C}$ is a $J \times [rank(\mathbf{A})]$ matrix such that the modified model matrix $\mathbf{C}'\mathbf{A}$ has full row rank and $R(\mathbf{A}) = R(\mathbf{C}'\mathbf{A})$. Then $R(\mathbf{A})^{\perp} = R(\mathbf{C}'\mathbf{A})^{\perp}$, that is $Ker\,(\mathbf{A}) = Ker\,(\mathbf{C}'\mathbf{A})$.

Of course, the reparameterization does not affect the number of the degrees of freedom. The number of degrees of freedom of a model $\mathcal{P}_{\Theta} \subset \mathcal{P}$ is the difference between dimensionalities of Ω and Θ.

Theorem 2.1. *The number of degrees of freedom in a relational model $RM(\mathbf{S})$ is $|\mathcal{I}| - dimR(\mathbf{A})$.*

Proof. Let $\boldsymbol{\delta} = \boldsymbol{p} = (p(1), \ldots, p(|\mathcal{I}|)'$. Since $\sum_{i \in \mathcal{I}} p(i) = 1$, then the parameter space Ω is $|\mathcal{I}| - 1$-dimensional. If $RM(\mathbf{S})$ is a relational model for probabilities (3), its multiplicative parameters $\boldsymbol{\theta}$ must satisfy the normalizing equation

$$\sum_{i \in \mathcal{I}} \prod_{j=1}^{J} (\theta_j)^{\mathbf{I}_j(i)} = 1. \tag{6}$$

Since the model matrix is full row rank, then the set $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}_+^J : \ \sum_{i \in \mathcal{I}} \prod_{j=1}^{J} (\theta_j)^{\mathbf{I}_j(i)} = 1\}$ is a $J-1$-dimensional manifold in $\mathbb{R}^J$. Therefore, the number of degrees of freedom of $RM(\mathbf{S})$ is $dim\Omega - dim\Theta = |\mathcal{I}| - 1 - (J-1) = |\mathcal{I}| - dimR(\mathbf{A})$.

Let $\boldsymbol{\delta} = \boldsymbol{\lambda}$ and $RM(\mathbf{S})$ be a model for intensities. In this case, $\Omega = \{\boldsymbol{\lambda} \in \mathbb{R}_+^{|\mathcal{I}|}\}$ and $\Theta \subset \Omega$ consists of all $\boldsymbol{\lambda}$ satisfying (3). Since no normalization is needed, $dim\Omega = |\mathcal{I}|$ and $dim\Theta = dimR(\mathbf{A})$ and thence the number of degrees of freedom of $RM(\mathbf{S})$ is equal to $|\mathcal{I}| - dimR(\mathbf{A})$. □

The theorem implies that the number of degrees of freedom of the relational model coincides with $dim\, Ker(\mathbf{A})$. This is in coherence with the fact that the kernel of the model matrix is invariant of reparameterizations of the model (2). To restrict further analysis to models with a positive number of degrees of freedom suppose in the sequel that $Ker(\mathbf{A})$ is non-trivial. Without loss of generality, suppose further that the model matrix is full row rank.

Definition 2.1. A matrix $\mathbf{D}$ with rows that form a basis of $Ker(\mathbf{A})$ is called *a kernel basis matrix* of the relational model $RM(\mathbf{S})$.

The representation (2) is a primal (intuitive) representation of relational models; a dual representation is described in the following theorem.

Theorem 2.2. *(i) The distribution, parameterized by* $\boldsymbol{\delta}$*, belongs to the relational model* $RM(\mathbf{S})$ *if and only if*

$$\mathbf{D} log\, \boldsymbol{\delta} = \mathbf{0}. \tag{7}$$

(ii) The matrix $\mathbf{D}$ *may be chosen to have integer entries.*

Proof. (i) By the definition of a relational model,

$$P_{\boldsymbol{\delta}} \in RM(\mathbf{S}) \ \Leftrightarrow\ \log \boldsymbol{\delta} = \mathbf{A}'\boldsymbol{\beta}.$$

The orthogonality of the design space and the null space implies that $\mathbf{AD}' = \mathbf{0}$ for any kernel basis matrix $\mathbf{D}$. The rows of $\mathbf{D}$ are linearly independent. Therefore,

$$P_{\boldsymbol{\delta}} \in RM(\mathbf{S}) \ \Leftrightarrow\ \mathbf{D}\log \boldsymbol{\delta} = \mathbf{DA}'\boldsymbol{\beta} = \mathbf{0}.$$

(ii) Since $\mathbf{A}$ has full row rank, the dimension of $Ker\,(\mathbf{A})$ is equal to $K_0 = |\mathcal{I}| - J$.

By Corollary 4.3b (Schrijver, 1986, pg. 49), there exists a unimodular matrix $\mathbf{U}$, i.e. $\mathbf{U}$ is integer and $det\, \mathbf{U} = \pm 1$, such that $\mathbf{AU}$ is the Hermite normal form of $\mathbf{A}$, that is

(a) $\mathbf{AU}$ has form $[\mathbf{B}, \mathbf{0}]$;

(b) $\mathbf{B}$ is a non-negative, non-singular, lower triangular matrix;

(c) $\mathbf{AU}$ is an $n \times m$ matrix with entries c_{ij} such that $c_{ij} < c_{ii}$ for all $i = 1, \dots, n,\ j = 1, \dots, m,\ i \neq j$.

Let $\mathbf{I}_{K_0}$ stand for the $K_0 \times K_0$ identity matrix, $\mathbf{0}$ denote the $J \times K_0$ zero matrix, and $\mathbf{Z}$ be the following $|\mathcal{I}| \times K_0$ matrix:

$$\mathbf{Z} = \begin{pmatrix} \mathbf{0} \\ \mathbf{I}_{K_0} \end{pmatrix}.$$

Since the matrix $\mathbf{AU}$ has form $[\mathbf{B}, \mathbf{0}]$, where $\mathbf{B}$ is the nonsingular, lower triangular, $J \times J$ matrix, then $(\mathbf{AU})\mathbf{Z} = \mathbf{0}$.

Set $\mathbf{D}' = \mathbf{UZ}$. Then

$$\mathbf{AD}' = \mathbf{AUZ} = \mathbf{0}. \tag{8}$$

The matrix $\mathbf{U}$ is integer and nonsingular, the columns of $\mathbf{Z}$ are linearly independent. Therefore, the matrix $\mathbf{D}'$ is integer and has linearly independent columns. Hence the matrix $\mathbf{D}$ is an integer kernel basis matrix of the model. □

Example 1.1 (Revisited) For the model of conditional independence, $dim\, Ker(\mathbf{A}) = 2$. If the kernel basis matrix is chosen as

$$\mathbf{D} = \begin{pmatrix} 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 \end{pmatrix},$$

the equation $\mathbf{D}\log \boldsymbol{p} = \mathbf{0}$ is equivalent to the following constraints:

$$\frac{p_{000}p_{110}}{p_{010}p_{100}} = 1, \;\; \frac{p_{001}p_{111}}{p_{011}p_{101}} = 1.$$

The latter is a well-known representation of the model $[Y_1Y_3][Y_2Y_3]$ in terms of the conditional odds ratios (Bishop et al., 1975). □

The dual representation (7) of a relational model is, in fact, a model representation in terms of some monomials in $\boldsymbol{\delta}$. All types of polynomial expressions that may arise in the dual representation of a relational model are captured by the following definition.

Definition 2.2. Let $u(i), v(i) \in \mathbb{Z}_{\geq 0}$ for all $i \in \mathcal{I}$, $\boldsymbol{\delta}^{\boldsymbol{u}} = \prod_{i\in\mathcal{I}} \delta(i)^{u(i)}$ and $\boldsymbol{\delta}^{\boldsymbol{v}} = \prod_{i\in\mathcal{I}} \delta(i)^{v(i)}$. *A generalized odds ratio* for a positive distribution, parameterized by $\boldsymbol{\delta}$, is a ratio of two monomials:

$$\mathcal{OR} = \boldsymbol{\delta}^{\boldsymbol{u}}/\boldsymbol{\delta}^{\boldsymbol{v}}. \tag{9}$$

The odds ratio $\mathcal{OR} = \boldsymbol{\delta}^{\boldsymbol{u}}/\boldsymbol{\delta}^{\boldsymbol{v}}$ is called homogeneous if $\sum_{i\in\mathcal{I}} u(i) = \sum_{i\in\mathcal{I}} v(i)$.

To express a relational model $RM(\mathbf{S})$ in terms of generalized odds ratios, write the rows $\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_{K_0} \in \mathbb{Z}^{|\mathcal{I}|}$ of a kernel basis matrix $\mathbf{D}$ in terms of their positive and negative parts:

$$\boldsymbol{d}_l = \boldsymbol{d}_l^+ - \boldsymbol{d}_l^-,$$

where $\boldsymbol{d}_l^+$, $\boldsymbol{d}_l^- \geq \mathbf{0}$ for all $l = 1, 2, \ldots, K_0$. Then the model (7) takes the form

$$\boldsymbol{d}_l^+ \log \boldsymbol{\delta} = \boldsymbol{d}_l^- \log \boldsymbol{\delta}, \;\; \text{for } l = 1, 2, \ldots, K_0,$$

which is equivalent to the model representation in terms of generalized odds ratios:

$$\boldsymbol{\delta}^{\boldsymbol{d}_l^+}/\boldsymbol{\delta}^{\boldsymbol{d}_l^-} = 1, \;\; \text{for } l = 1, 2, \ldots, K_0. \tag{10}$$

The number of degrees of freedom is equal to the minimal number of generalized odds ratios required to uniquely specify a relational model.

Example 1.2 (Revisited) The model $\lambda_{00} = \lambda_{01}\lambda_{10}$ can be expressed in the matrix form as

$$\mathbf{D}\log \boldsymbol{\lambda} = 0, \tag{11}$$

where $\mathbf{D} = (1, -1, -1)$. The matrix $\mathbf{D}$ is a kernel basis matrix of the relational model, as one would expect. Finally, the model representation in terms of generalized odds ratios is

$$\frac{\lambda_{00}}{\lambda_{01}\lambda_{10}} = 1.$$

□

The role of generalized odds ratios in parameterizing distributions in $\mathcal{P}$ will be explored in Section 5.

3 Relational Models as Exponential Families: Poisson vs Multinomial Sampling

The representation (3) implies that a relational model is an exponential family of distributions. The canonical parameters of a relational model are β_j's and the canonical statistics are indicators of subsets $\mathbf{I}_j$. Relational models for intensities and relational models for probabilities are considered in this section in more detail.

Let $RM_{\boldsymbol{\lambda}}(\mathbf{S})$ denote a relational model for intensities and $RM_{\boldsymbol{p}}(\mathbf{S})$ denote a relational model for probabilities with the same model matrix $\mathbf{A}$, that has a full row rank J.

Theorem 3.1. *A model $RM_{\boldsymbol{\lambda}}(\mathbf{S})$ is a regular exponential family of order J.*

Proof. The model matrix $\mathbf{A}$ in (2) has full row rank; no normalization is needed for intensities. Therefore, the representation (3) is minimal and the exponential family is regular, of order J. □

Relational models for probabilities may have a more complex structure than relational models for intensities and, in some cases, become curved exponential families (Efron, 1975; Brown, 1988; Kass & Vos, 1997).

Theorem 3.2. *If $\mathbf{1} \in R(\mathbf{A})$, a model $RM_{\boldsymbol{p}}(\mathbf{S})$ is a regular exponential family of order $J-1$; otherwise, it is a curved exponential family of order $J - 1$.*

Proof. Suppose that $\mathbf{1} \in R(\mathbf{A})$. Without loss of generality, $\mathcal{I} = S_1 \in \mathbf{S}$ and thus

$$p(i) = \exp\{\beta_1\}\exp\{\sum_{j=2}^{J} \mathbf{I}_j(i)\beta_j\}. \tag{12}$$

The exponential family representation given by (12) is minimal; the model $RM_{\boldsymbol{p}}(\mathbf{S})$ is a regular exponential family of order $J - 1$.

If $\mathbf{1} \notin R(\mathbf{A})$ then, independently of what parameterization is used, the model matrix does not include the row of all 1s. Then normalization is required and thus the parameter space is a manifold of dimension $J - 1$ in $\mathbb{R}^J$ (see e.g. Rudin, 1976, p.229). In this case, $RM_{\boldsymbol{p}}(\mathbf{S})$ is a curved exponential family of order $J - 1$ (Kass & Vos, 1997). □

If a relational model is a regular exponential family, the maximum likelihood estimate of the canonical parameter exists if and only if the observed value of the canonical statistic is contained in the interior of the convex hull of the support of its distribution (Barndorff-Nielsen, 1978). In this case, the MLE is also unique.

If the distribution of a random vector $\boldsymbol{Y}$ is parameterized by intensities $\boldsymbol{\lambda}$, then, under the model $RM_{\boldsymbol{\lambda}}(\mathbf{S})$,

$$P(\boldsymbol{Y} = \boldsymbol{y}) = \frac{1}{\prod_{i \in \mathcal{I}} y(i)!} \exp\ \{\boldsymbol{\beta}'\mathbf{A}\boldsymbol{y} - \mathbf{1}\exp\mathbf{A}'\boldsymbol{\beta}\}. \tag{13}$$

If the distribution of $\boldsymbol{Y}$ is multinomial, with parameters N and $\boldsymbol{p}$, then, under the model $RM_{\boldsymbol{p}}(\mathbf{S})$,

$$P(\boldsymbol{Y} = \boldsymbol{y}) = \frac{N!}{\prod_{i \in \mathcal{I}} y(i)!} \exp\ \{\boldsymbol{\beta}'\mathbf{A}\boldsymbol{y}\}. \tag{14}$$

Set

$$\boldsymbol{T}(\boldsymbol{Y}) = \mathbf{A}\boldsymbol{Y} = (T_1(\boldsymbol{Y}), T_2(\boldsymbol{Y}), \ldots, T_J(\boldsymbol{Y}))'. \tag{15}$$

For each $j \in 1, \ldots, J$, the statistic $T_j(\boldsymbol{Y}) = \sum_{i \in \mathcal{I}} \mathbf{I}_j(i)Y(i)$ is the subset sum corresponding to the subset S_j.

It is well known for log-linear models that the kernel of the likelihood is the same for the multinomial and Poisson sampling schemes, if the sample sizes are equal, and thus the maximum likelihood estimates of the cell frequencies, obtained under either sampling scheme, are equal (see e.g. Birch (1963) and Bishop et al. (1975), p.448). The following theorem is an extension of this result.

Theorem 3.3. *Assume that, for a given set of observations, the maximum likelihood estimates $\hat{\boldsymbol{\lambda}}$, under the model $RM_{\boldsymbol{\lambda}}(\mathbf{S})$, and $\hat{\mathbf{p}}$, under the model $RM_{\boldsymbol{p}}(\mathbf{S})$, exist. The following four conditions are equivalent:*

(A) The MLEs for the cell frequencies obtained under either model are the same.

(B) The vector $\mathbf{1}$ *is in the design space* $R(\mathbf{A})$*.*

(C) Both models may be defined by homogeneous odds ratios.

(D) The model for intensities is scale invariant.

Proof. (A) $\Longleftarrow$ (B)

Under the model $RM_{\boldsymbol{p}}(\mathbf{S})$, the probabilities can be written in the form (3):

$$p(i) = \prod_{j=1}^{J} (\theta_j)^{\mathbf{I}_j(i)}, \;\; i \in \mathcal{I},$$

where $\beta_j = \log \theta_j$, for $j = 1, \dots, J$. The problem of maximization, with respect to $\boldsymbol{\theta}$, of the likelihood (14) under the normalization condition (6) is equivalent to maximizing the Lagrangian

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i \in \mathcal{I}} y(i) \sum_{j=1}^{J} \mathbf{I}_j(i) \log\, \theta_j - \alpha (\sum_{i \in \mathcal{I}} \prod_{j=1}^{J} (\theta_j)^{\mathbf{I}_j(i)} - 1).$$

Setting the derivatives of $\mathcal{L}$ with respect to θ_j, $j = 1, \dots, J$, equal to zero and rearranging terms lead to the likelihood equations

$$\begin{aligned} \mathbf{A}\boldsymbol{y} &= \alpha \mathbf{A}\hat{\boldsymbol{p}}, \\ \mathbf{1}\hat{\boldsymbol{p}} &= 1. \end{aligned} \tag{16}$$

Here $\hat{\boldsymbol{p}}$ are the maximum likelihood estimates for probabilities under the model $RM_{\boldsymbol{p}}(\mathbf{S})$.

If $\mathbf{1} \in R(\mathbf{A})$, then there exists a $\boldsymbol{k} \in \mathbb{R}^J$ such that $\boldsymbol{k}'\mathbf{A} = \mathbf{1}$. Multiplying both sides of the first equation in (16) by $\boldsymbol{k}'$ yields $\alpha = N$ and hence

$$\mathbf{A}\boldsymbol{y} = N\mathbf{A}\hat{\boldsymbol{p}}. \tag{17}$$

Under the model $RM_{\boldsymbol{\lambda}}(\mathbf{S})$, the problem of maximization of the likelihood (13) leads to the likelihood equations

$$\mathbf{A}\boldsymbol{y} = \mathbf{A}\hat{\boldsymbol{\lambda}}, \tag{18}$$

where $\hat{\boldsymbol{\lambda}}$ are the maximum likelihood estimates for intensities.

From the equations (17) and (18):

$$\hat{\boldsymbol{\lambda}} - N\hat{\boldsymbol{p}} \in Ker(\mathbf{A}).$$

The latter implies that $\mathbf{1}(\hat{\boldsymbol{\lambda}} - N\hat{\boldsymbol{p}}) = 0$ and $N = \mathbf{1}\hat{\boldsymbol{\lambda}}$. Therefore,

$$\hat{\boldsymbol{p}} = \frac{\hat{\boldsymbol{\lambda}}}{\mathbf{1}\hat{\boldsymbol{\lambda}}}$$

and the maximum likelihood estimates for the cell frequencies obtained under either model are the same:

$$\hat{\boldsymbol{y}} = N\hat{\boldsymbol{p}} = \hat{\boldsymbol{\lambda}}.$$

(A) $\Longrightarrow$ (B)

Suppose that $\hat{\boldsymbol{y}} = N\hat{\boldsymbol{p}} = \hat{\boldsymbol{\lambda}}$. Under the model $RM_{\boldsymbol{\lambda}}(\mathbf{S})$

$$\log\,(\hat{\boldsymbol{\lambda}}) = \mathbf{A}'\hat{\boldsymbol{\beta}}_1$$

for some $\hat{\boldsymbol{\beta}}_1$. On the other hand, under the model $RM_{\boldsymbol{p}}(\mathbf{S})$,

$$\log\,(\hat{\boldsymbol{\lambda}}) = \log\,(N\hat{\boldsymbol{p}}) = \mathbf{A}'\hat{\boldsymbol{\beta}}_2 + \log\,N\mathbf{1}'$$

for some $\hat{\boldsymbol{\beta}}_2$. The condition $\mathbf{A}'\hat{\boldsymbol{\beta}}_1 = \mathbf{A}'\hat{\boldsymbol{\beta}}_2 + \log\,N\mathbf{1}'$ can only hold if $\mathbf{1} \in R(\mathbf{A})$.

(B) $\Longleftrightarrow$ (C)

The vector $\mathbf{1} \in R(\mathbf{A})$ if and only if all rows of a kernel basis matrix $\mathbf{D}$ are orthogonal to $\mathbf{1}$, or the sum of entries in every row of $\mathbf{D}$ is zero. The latter is equivalent to the generalized odds ratios obtained from rows of $\mathbf{D}$ being homogeneous.

(D) $\Longleftrightarrow$ (B)

Let $t > 0$, $t \neq 1$.

$$\mathbf{D}\log\,(t\boldsymbol{\lambda}) = \mathbf{0} \Longleftrightarrow \log\,t \cdot (\mathbf{D}\mathbf{1}') = \mathbf{0} \Longleftrightarrow \mathbf{D}\mathbf{1}' = \mathbf{0}, \text{ or } \mathbf{1} \in R(\mathbf{A}).$$

□

4 Existence and Properties of the Maximum Likelihood Estimates

The condition $\mathbf{1} \notin R(\mathbf{A})$ affects the properties of the MLE and the model structure.

Theorem 4.1. *Under a model $RM_{\boldsymbol{p}}(\mathbf{S})$, the sums of the MLEs of the cell frequencies in the subsets $S_1, \ldots, S_J$ are equal to their observed values for any observed distribution if and only if $\mathbf{1} \in R(\mathbf{A})$.*

Proof. If $\mathbf{1} \in R(\mathbf{A})$, the model $RM_{\boldsymbol{p}}(\mathbf{S})$ is a regular exponential family and the statement holds.

Suppose that the subset sums of the MLEs are equal to their observed values for any observed distribution. To prove that $\mathbf{1} \in R(\mathbf{A})$ it suffices to show that every element of $Ker(\mathbf{A})$ is orthogonal to $\mathbf{1}$. Let $\boldsymbol{u}$ be an arbitrary vector in $Ker(\mathbf{A})$. There exists a frequency distribution $\boldsymbol{y}$, such that $\boldsymbol{y}+\boldsymbol{u}$ is also a frequency distribution. The kernels of the log-likelihoods of $\boldsymbol{y}$ and $\boldsymbol{y}+\boldsymbol{u}$ are $\boldsymbol{y}'\mathbf{A}'\boldsymbol{\beta}$ and $(\boldsymbol{y}+\boldsymbol{u})'\mathbf{A}'\boldsymbol{\beta}$ respectively. The vector $\boldsymbol{u} \in Ker(\mathbf{A})$ and thus $u'\mathbf{A}' = \mathbf{0}$, so the two log-likelihoods coincide. Therefore, the MLEs for cell probabilities are equal:

$$\hat{\boldsymbol{p}}_{\boldsymbol{y}} = \hat{\boldsymbol{p}}_{\boldsymbol{y}+\boldsymbol{u}},$$

where $\hat{\boldsymbol{p}}_{\boldsymbol{y}}$ denotes the MLE for $\boldsymbol{p}_{\boldsymbol{y}} = \boldsymbol{y}/\mathbf{1}\boldsymbol{y}$ and $\hat{\boldsymbol{p}}_{\boldsymbol{y}+\boldsymbol{u}}$ denotes the MLE for $\boldsymbol{p}_{\boldsymbol{y}+\boldsymbol{u}} = (\boldsymbol{y}+\boldsymbol{u})/\mathbf{1}(\boldsymbol{y}+\boldsymbol{u})$. Under the initial assumption about the subset sums of the MLEs,

$$\mathbf{A}\hat{\boldsymbol{p}}_{\boldsymbol{y}} = \mathbf{A}\boldsymbol{p}_{\boldsymbol{y}} \text{ and } \mathbf{A}\hat{\boldsymbol{p}}_{\boldsymbol{y}+\boldsymbol{u}} = \mathbf{A}\boldsymbol{p}_{\boldsymbol{y}+\boldsymbol{u}}.$$

Therefore, using that $\mathbf{A}\boldsymbol{u} = \mathbf{0}$,

$$\mathbf{A}\frac{\boldsymbol{y}}{\mathbf{1}\boldsymbol{y}} = \mathbf{A}\hat{\boldsymbol{p}}_{\boldsymbol{y}} = \mathbf{A}\hat{\boldsymbol{p}}_{\boldsymbol{y}+\boldsymbol{u}} = \mathbf{A}\frac{\boldsymbol{y}+\boldsymbol{u}}{\mathbf{1}(\boldsymbol{y}+\boldsymbol{u})} = \mathbf{A}\frac{\boldsymbol{y}}{\mathbf{1}(\boldsymbol{y}+\boldsymbol{u})},$$

implying the equality $\mathbf{1}\boldsymbol{y} = \mathbf{1}(\boldsymbol{y}+\boldsymbol{u})$, which is possible if and only if $\mathbf{1}\boldsymbol{u} = 0$.

□

Corollary 4.2. *Suppose $\mathbf{1} \notin R(\mathbf{A})$. For a given set of observations, the MLEs, if exist, of the subset sums under a model $RM_{\boldsymbol{p}}(\mathbf{S})$ are proportional to their observed values.*

Proof. In this case the value of α cannot be found from (16) and one can only assert that

$$\mathbf{A}\boldsymbol{y} = \frac{\alpha}{N}\mathbf{A}\hat{\boldsymbol{y}}.$$

□

Example 1.2 illustrates a situation when a relational model for intensities is not scale invariant. This model is a curved exponential family. The existence and uniqueness of the maximum likelihood estimates in such relational models are proven next.

Theorem 4.3. *Let $\boldsymbol{Y} \sim M(N, \boldsymbol{p})$, $\boldsymbol{y}$ be a realization of $\boldsymbol{Y}$, and $RM_{\boldsymbol{p}}(\mathbf{S})$ be a relational model, such that $\mathbf{1} \notin R(\mathbf{A})$. The maximum likelihood estimate for $\boldsymbol{p}$, under the model $RM_{\boldsymbol{p}}(\mathbf{S})$, exists and is unique if and only if $\boldsymbol{T}(\boldsymbol{y}) > 0$.*

Proof. A point in the canonical parameter space of the model $RM_{\boldsymbol{p}}(\mathbf{S})$ that maximizes the log-likelihood subject to the normalization constraint is a solution to the optimization problem:

$$\max_{\text{s.t. } \boldsymbol{\beta} \in \mathcal{D}} l(\boldsymbol{\beta}; \boldsymbol{y}),$$

where

$$l(\boldsymbol{\beta};\boldsymbol{y}) = T_1(\boldsymbol{y})\beta_1 + \cdots + T_J(\boldsymbol{y})\beta_J$$

and

$$\mathcal{D} = \{\boldsymbol{\beta} \in \mathbb{R}^J_- : \sum_{i \in \mathcal{I}} \exp\{\sum_{j=1}^{J} \mathbf{I}_j(i)\beta_j\} - 1 = 0\}.$$

The set $\mathcal{D}$ is non-empty and is a level set of a convex function. The level sets of convex functions are not convex in general. However, the sub-level sets of convex functions and hence the set

$$\mathcal{D}_{\leq} = \{\boldsymbol{\beta} \in \mathbb{R}^J_- : \sum_{i \in \mathcal{I}} \exp\{\sum_{j=1}^{J} \mathbf{I}_j(i)\beta_j\} - 1 \leq 0\}$$

are convex.

The set of maxima of $l(\boldsymbol{\beta};\boldsymbol{y})$ over the set $\mathcal{D}_{\leq}$ is nonempty and consists of a single point if and only if (Bertsekas, 2009, Section 3)

$$R_{\mathcal{D}_{\leq}} \cap R_{-l} = L_{\mathcal{D}_{\leq}} \cap L_{-l}.$$

Here $R_{\mathcal{D}_{\leq}}$ is the recession cone of the set $\mathcal{D}_{\leq}$, R_{-l} is the recession cone of the function $-l$, $L_{\mathcal{D}_{\leq}}$ is the lineality space of $\mathcal{D}_{\leq}$, and L_{-l} is the lineality space of $-l$.

The recession cone of $\mathcal{D}_{\leq}$ is the orthant $\mathbb{R}^J_-$, including the origin; the lineality space is $L_{\mathcal{D}_{\leq}} = \{0\}$. The lineality space of the function $-l$ is the plane passing through the origin, with the normal $\mathbf{T}(\boldsymbol{y})$; the recession cone of $-l$ is the half-space above this plane. The condition $R_{\mathcal{D}_{\leq}} \cap R_{-l} = L_{\mathcal{D}_{\leq}} \cap L_{-l} = \{0\}$ holds if and only if all components of $\mathbf{T}(\boldsymbol{y}) = (T_1(\boldsymbol{y}), \ldots, T_J(\boldsymbol{y}))'$ are positive.

The function $l(\boldsymbol{\beta};\boldsymbol{y})$ is linear; its maximum is achieved on $\mathcal{D}$. Therefore, there exists one and only one $\boldsymbol{\beta}$ which maximizes the likelihood over the canonical parameter space, and the maximum likelihood estimate for $\boldsymbol{p}$, under the model $RM_{\boldsymbol{p}}(\mathbf{S})$, exists and is unique. □

Table 4: The MLEs for the Number of trapped *Charybdis japonica* by bait type

	Fish	
Sugarcane	Yes	No
Yes	35.06	2.94
No	11.94	-

Table 5: The MLEs for the Number of trapped *Portunuspelagicus* by bait type

	Fish	
Sugarcane	Yes	No
Yes	72.31	1.69
No	42.69	-

Example 1.2 (Revisited) In this example, the relational model for intensities is not scale invariant. The maximum likelihood estimates for the cell frequencies exist and are shown in

Tables 4 and 5. The observed Pearson's statistics are $X^2 = 0.40$ and $X^2 = 1.07$ respectively, on one degree of freedom. □

The relational model framework deals with models generated by subsets of cells, and the model matrix for a relational model is an indicator matrix that has only 0-1 entries. Theorems 2.2 and 3.3 hold if the model matrix has non-negative integer entries. The next example illustrates how the techniques and theorems apply to those more general exponential families.

Example 4.1. This example, given in (Agresti, 2002), describes a study carried out to determine if a pneumonia infection has an immunizing effect on dairy calves. Within 60 days after birth, the calves were exposed to a pneumonia infection. The calves that got the infection were then classified according to whether or not they got the secondary infection within two weeks after the first infection cleared up. The number of the infected calves is thus a random variable with the multinomial distribution $M(N, (p_{11}, p_{12}, p_{22})')$, where N denotes the total number of calves in the sample. Suppose further that p_{11} is the probability to get both the primary and the secondary infection, p_{12} is the probability to get only the primary infection and not the secondary one, and p_{22} is the probability not to catch either the primary or the secondary infection. Let $0 < \pi < 1$ denote the probability to get the primary infection. The hypothesis of no immunizing effect of the primary infection is expressed as (cf. Agresti, 2002)

$$p_{11} = \pi^2, \; p_{12} = \pi(1-\pi), \; p_{22} = 1 - \pi. \tag{19}$$

The model given in (19) does not contain the overall effect and can be expressed in terms of a non-homogeneous odds ratio:

$$\frac{p_{11}p_{22}^2}{p_{12}^2} = 1.$$

Write N_{11}, N_{12}, N_{22} for the number of calves, as a random variable in each category, and n_{11}, n_{12}, n_{22} for their realizations. The log-likelihood is proportional to

$$(2n_{11} + n_{12})\text{log } \pi + (n_{12} + n_{22})\text{log } (1 - \pi).$$

The sufficient statistic $\boldsymbol{T} = (2N_{11} + N_{12}, N_{12} + N_{22})$ is two-dimensional. The canonical parameter space $\{(\log \pi, \log (1-\pi)) : \pi \in (0, 1)\}$ is the curve in $\mathbb{R}^2$ shown in Figure 1. The model (19) is thus a curved exponential family of order 1.

The likelihood is maximized by

$$\hat{\pi} = \frac{2n_{11} + n_{12}}{2n_{11} + 2n_{12} + n_{22}} = \frac{T_1}{T_1 + T_2},$$

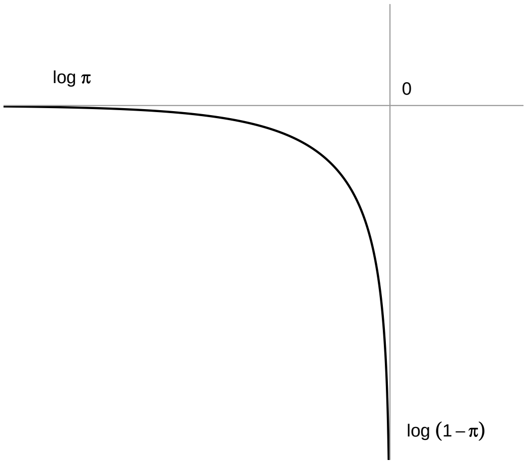

Figure 1: The canonical parameter space in Example 4.1

where $T_1 = 2n_{11} + n_{12}$ and $T_2 = n_{12} + n_{22}$ are the observed components of the sufficient statistic, or the subset sums. The MLEs of the subset sums can be expressed in terms of their observed values as

$$N(2\hat{\pi}^2 + \hat{\pi}(1-\hat{\pi})) = N(\frac{2T_1^2}{(T_1+T_2)^2} + \frac{T_1T_2}{(T_1+T_2)^2}) = T_1\frac{N(2T_1+T_2)}{(T_1+T_2)^2},$$
$$N(\hat{\pi}(1-\hat{\pi}) + (1-\hat{\pi})) = N(\frac{T_1T_2}{(T_1+T_2)^2} + \frac{T_2}{T_1+T_2}) = T_2\frac{N(2T_1+T_2)}{(T_1+T_2)^2}.$$

Thus, under the model (19), the MLEs of the subset sums differ from their observed values by the factor $\frac{N(2T_1+T_2)}{(T_1+T_2)^2}$. For the data and the MLEs in Table 6, this factor is approximately 0.936. □

Table 6: Observed (Expected) Counts for Primary and Secondary Pneumonia Infection of Calves (Agresti, 2002)

	Secondary Infection	
Primary Infection	Yes	No
Yes	30 (38.1)	63 (39.0)
No	-	63 (78.9)

Let $RM_{\boldsymbol{p}}(\mathbf{S})$ be a relational model for probabilities with the model matrix $\mathbf{A}$ of full row rank. Then for any two distributions $P, Q \in \mathcal{P}$, with parameters $\boldsymbol{p}$ and $\boldsymbol{q}$ respectively, the relation

$$P \underset{\mathbf{A}}{\sim} Q \qquad \text{iff} \qquad \mathbf{A}\boldsymbol{p} = \alpha \mathbf{A}\boldsymbol{q} \qquad \text{for some } \alpha > 0. \tag{20}$$

is an equivalence relation and, thus, defines a partition of $\mathcal{P}$. The following statement summarizes Theorem 4.1, Corollary 4.2, and Theorem 4.3; the proof is thus omitted.

Theorem 4.4. *Suppose $H \subset \mathcal{P}$ is a class of the partition defined by $\underset{\mathbf{A}}{\sim}$. Then the following holds:*

(a) If $\mathbf{1} \in R(\mathbf{A})$, then $\alpha = 1$ for every pair of distributions $P, Q \in H$.

(b) $|RM_{\boldsymbol{p}}(\mathbf{S}) \cap H| = 1$. Say, $RM_{\boldsymbol{p}}(\mathbf{S}) \cap H = \{T\}$.

(c) For every $P \in H$, its MLE under the model $RM_{\boldsymbol{p}}(\mathbf{S})$ is T.

Theorem 4.4 is an extension of the results of Birch (1963) and Csiszár (1975), which apply to the regular case, and has a clear geometric interpretation. A generalization of Birch's theorem for toric models in terms of algebraic geometry is given by Pachter & Sturmfels (2005), p.14. This generalization can be applied to the relational models that are regular exponential families. In this case, for the observed frequency distribution $\boldsymbol{y_0}$, the MLE, if exists, is the unique point of the intersection of the projective toric variety $\mathcal{V}$ and the polytope $\mathcal{P}_{\boldsymbol{y}_0}$ defined by the equations $\mathbf{A}\boldsymbol{y} = \mathbf{A}\boldsymbol{y_0}$. The variety $\mathcal{V}$ is the vanishing set of the homogeneous toric ideal $I_{\mathbf{A}}$ spanned by the binomials $\boldsymbol{p}^{\boldsymbol{u}} - \boldsymbol{p}^{\boldsymbol{v}}$, where $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{Z}_{\geq 0}^{|\mathcal{I}|} \cap Ker(\mathbf{A})$ (cf. Sturmfels, 1996, p.31). The set of frequency distributions which have the same subset sums as the observed table

$$\mathcal{F}_{\boldsymbol{y}_0} = \{\boldsymbol{y} \in \mathcal{Y} : \ \mathbf{A}\boldsymbol{y} = \mathbf{A}\boldsymbol{y_0}\}$$

is called the fiber of $\boldsymbol{y_0}$. If the equivalence relation is extended to frequency distributions, the fiber $\mathcal{F}_{\boldsymbol{y}_0}$ becomes an equivalence class under $\underset{\mathbf{A}}{\sim}$ and all distributions in it have the same MLE. A fiber is a finite set and any two frequency distributions in it are connected by a "walk" along the elements of this fiber. The set of moves that is sufficient to connect any two distributions in fibers $\mathcal{F}_{\boldsymbol{y}}$ for all $\boldsymbol{y} \in \mathcal{Y}$ is called a Markov basis. The moves in a Markov basis belong to the kernel of the model matrix $\mathbf{A}$ and can be derived from a lattice basis of the relational model by, for example, the Saturation algorithm (cf. Sturmfels, 1996, p.114).

However, a relational model for probabilities that is a curved exponential family is not a toric model. The ideal $I_{\mathbf{A}}$ spanned by the binomials $\boldsymbol{p}^{\boldsymbol{u}} - \boldsymbol{p}^{\boldsymbol{v}}$, where $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{Z}_{\geq 0}^{|\mathcal{I}|} \cap Ker(\mathbf{A})$, is not homogeneous in this case. Theorem 4.3 implies that the MLE under such a model is the unique point of the intersection of the affine toric variety $\mathcal{V}$ (the vanishing set of $I_{\mathbf{A}}$), the polytope $\mathbf{A}\boldsymbol{p} = \alpha\mathbf{A}\boldsymbol{y_0}$ (for some constant $\alpha > 0$) and the normalizing equation $\mathbf{1}\boldsymbol{p} = 1$, which defines a hyper-surface in $\mathbb{R}^{|\mathcal{I}|}$. As it follows from Theorem 4.4, the equivalence classes induced by $\underset{\mathbf{A}}{\sim}$ on the sample space have more complex structure than a fiber in the regular case. Every equivalence class includes distributions with the same maximum likelihood estimates and the frequency distributions in it are connected by a walk, but the coefficient of proportionality varies over the distributions in this class. This fact is illustrated in the next example.

Example 4.2. Let $\mathcal{I}$ be a table with only three cells and $\boldsymbol{p} = (p_1, p_2, p_3)$, where $p_i \in (0, 1)$ for $i = 1, 2, 3$, and $p_1 + p_2 + p_3 = 1$. The relational model $p_3 = p_1 p_2$ is a curved exponential family. Its model matrix and the kernel basis matrix are, respectively,

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \ \mathbf{D} = (1, 1, -1).$$

Let $T_1 = p_1 + p_3$ and $T_2 = p_2 + p_3$ denote the subset sums. If $p_1 \neq p_2$, the MLEs are

$$\hat{p}_1 = (-T_2 + \sqrt{T_1^2 + T_2^2})/T_1, \ \ \hat{p}_2 = (T_1 - T_2)/(-T_2) + T_1/T_2 * \hat{p}_1, \ \ \hat{p}_3 = 1 - \hat{p}_1 - \hat{p}_2$$

and the ratio of the subset sums of the given distribution to their MLEs is $\alpha = (p_1+p_3)/(\hat{p}_1+\hat{p}_3)$. If $p_1 = p_2$, the MLE is $\hat{p}_1 = \hat{p}_2 = -1 + \sqrt{2}$ and $\hat{p}_3 = 3 - 2\sqrt{2}$ and the ratio of the subset sums equals $\alpha = (p_1 + p_3)/(\hat{p}_1 + \hat{p}_3) = (p_1 + p_3)/(2 - \sqrt{2})$.

For the distribution $\boldsymbol{p} = (0.5, 0.2, 0.3)$, the MLE is $\hat{\boldsymbol{p}} = (0.554, 0.287, 0.159)$ and the ratio $\alpha_1 = (p_1 + p_3)/(\hat{p}_1 + \hat{p}_3) = 1.12$. Another distribution from the same equivalence class is $\boldsymbol{q} = (54/99, 27/99, 18/99)$. One can check that $\hat{\boldsymbol{q}} = \hat{\boldsymbol{p}}$ and the ratio $\alpha_2 = (q_1+q_3)/(\hat{q}_1+\hat{q}_3) = 1.02$. □

5 Mixed Parameterization of Exponential Families

Let $\mathcal{P}_{\boldsymbol{\delta}}$ be an exponential family formed by all positive distributions on $\mathcal{I}$ and $\log \boldsymbol{\delta}$ be the canonical parameters of this family. Denote by $\mathcal{P}_{\boldsymbol{\gamma}}$ the reparameterization of $\mathcal{P}_{\boldsymbol{\delta}}$ defined by the following one-to-one mapping:

$$\log \boldsymbol{\delta} = \mathbf{M}'\boldsymbol{\gamma}, \tag{21}$$

where $\mathbf{M}$ is a full rank, $|\mathcal{I}| \times |\mathcal{I}|$, integer matrix, and $\boldsymbol{\gamma} \in \mathbb{R}^{|\mathcal{I}|}$. It was shown by Brown (1988) that $\mathcal{P}_{\boldsymbol{\gamma}}$ is an exponential family with the canonical parameters $\boldsymbol{\gamma}$.

Theorem 5.1. *The canonical parameters of $\mathcal{P}_{\boldsymbol{\gamma}}$ are the generalized log odds ratios in terms of $\boldsymbol{\delta}$.*

Proof. Since the matrix $\mathbf{M}$ is full rank, then

$$\boldsymbol{\gamma} = (\mathbf{M}')^{-1}\log \boldsymbol{\delta}. \tag{22}$$

Let $\mathbf{B}$ denote the adjoint matrix to $\mathbf{M}'$ and write $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{|\mathcal{I}|}$ for the rows of $\mathbf{B}$. The components of $\boldsymbol{\gamma}$ can be expressed as:

$$\gamma_i = \frac{1}{\det(\mathbf{M})} \log \boldsymbol{\delta}^{\boldsymbol{b}_i}, \text{ for } i = 1, \ldots, |\mathcal{I}|. \tag{23}$$

All rows of $\mathbf{B}$ are integer vectors and thus the components of $\boldsymbol{\gamma}$ are multiples of the generalized log odds ratios. The common factor $1/\det(\mathbf{M}) \neq 0$ can be included in the canonical statistics, and the canonical parameters become equal to the generalized log odds ratios. □

Let $\mathbf{A}$ be a full row rank $J \times |\mathcal{I}|$ matrix with non-negative integer entries, and $\mathbf{D}$ denote a kernel basis matrix of $\mathbf{A}$. Set

$$\mathbf{M} = \left[\begin{array}{c} \mathbf{A} \\ \mathbf{D} \end{array} \right], \tag{24}$$

find the inverse of $\mathbf{M}$ and partition it as

$$\mathbf{M}^{-1} = \left[\mathbf{A}^{-}, \mathbf{D}^{-} \right].$$

Since $\mathbf{D}\mathbf{A}' = \mathbf{0}$, then $(\mathbf{D}^{-})'\mathbf{A}^{-} = \mathbf{0}$. This matrix $\mathbf{M}$ can be used to derive a mixed parameterization of $\mathcal{P}$ with variation independent parameters (cf. Brown, 1988; Hoffmann-Jørgensen, 1994). Under this parameterization,

$$\boldsymbol{\delta} \longmapsto \left(\begin{array}{c} \boldsymbol{\zeta}_1 \\ \boldsymbol{\zeta}_2 \end{array} \right), \tag{25}$$

where $\boldsymbol{\zeta}_1 = \mathbf{A}\boldsymbol{\delta}$ (mean-value parameters) and $\boldsymbol{\zeta}_2 = \mathbf{D}^{-}\log \boldsymbol{\delta}$ (canonical parameters), and the range of the vector $(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2)'$ is the Cartesian product of the separate ranges of $\boldsymbol{\zeta}_1$ and $\boldsymbol{\zeta}_2$.

Another mixed parameterization, which does not require calculating the inverse of $\mathbf{M}$, may be obtained as follows. Notice first that for any $\boldsymbol{\delta} \in \mathbb{R}_{+}^{|\mathcal{I}|}$ there exist unique vectors $\boldsymbol{\beta} \in \mathbb{R}^{J}$ and $\boldsymbol{\theta} \in \mathbb{R}^{|\mathcal{I}|-J}$ such that

$$\log \boldsymbol{\delta} = \mathbf{A}'\boldsymbol{\beta} + \mathbf{D}'\boldsymbol{\theta}. \tag{26}$$

By orthogonality,

$$\begin{aligned} \mathbf{D}\log \boldsymbol{\delta} &= \mathbf{0} + \mathbf{D}\mathbf{D}'\boldsymbol{\theta}, \\ \boldsymbol{\theta} &= (\mathbf{D}\mathbf{D}')^{-1}\mathbf{D}\log \boldsymbol{\delta}. \end{aligned} \tag{27}$$

Because of the uniqueness, $\mathbf{D}^{-} = (\mathbf{D}\mathbf{D}')^{-1}\mathbf{D}$. Moreover, since there is one-to-one correspondence between $\boldsymbol{\zeta}_2$ and $\tilde{\boldsymbol{\zeta}}_2 = \mathbf{D}\log \boldsymbol{\delta}$, then, in the mixed parameterization, the parameter $\boldsymbol{\zeta}_2$ can be replaced with $\tilde{\boldsymbol{\zeta}}_2$. The components of $\tilde{\boldsymbol{\zeta}}_2 = \mathbf{D}\log \boldsymbol{\delta}$ are some generalized log odds ratios as well.

A relational model is clearly defined and parameterized in the mixed parameterization derived from the model matrix of this model. In this parameterization the model requires logs of the generalized odds ratios to be zero and distributions in this model are parameterized by the remaining mean-value parameters.

The following two examples illustrate the proposed mixed parameterization.

Example 1.1 (Revisited) Consider a $2 \times 2 \times 2$ contingency table and matrices $\mathbf{A}$ and $\mathbf{D}$ as in Example 1.1. From (26):

$$\log \boldsymbol{p} \;=\; \mathbf{A}'\boldsymbol{\beta} + \theta_1 \cdot (1,0,-1,0,-1,0,1,0)' + \theta_2 \cdot (0,1,0,-1,0,-1,0,1)', \tag{28}$$

for some $\boldsymbol{\beta} \in \mathbb{R}^6$ and $\theta_1, \theta_2 \in \mathbb{R}$.

Since the rows of $\mathbf{D}$ are mutually orthogonal, then

$$\begin{aligned}(1,0,-1,0,-1,0,1,0)\log \boldsymbol{p} &= 4\theta_1,\\ (0,1,0,-1,0,-1,0,1)\log \boldsymbol{p} &= 4\theta_2.\end{aligned}$$

Thus, $\theta_1 = \frac{1}{4}\log\,(p_{111}p_{221})/(p_{121}p_{211})$ and $\theta_2 = \frac{1}{4}\log\,(p_{112}p_{222})/(p_{122}p_{212})$, as it is well known (see e.g. Bishop et al., 1975).

The parameters $\boldsymbol{\beta}$ can be expressed as generalized log odds ratios by applying (23):

$$\begin{aligned}\beta_1 &= \log\,\frac{p_{111}^3 p_{121} p_{211}}{p_{221}}, & \beta_2 &= \log\,\frac{p_{211}^2 p_{221}^2}{p_{111}^2 p_{121}^2},\\ \beta_3 &= \log\,\frac{p_{121}^2 p_{221}^2}{p_{111}^2 p_{211}^2}, & \beta_4 &= \log\,\frac{p_{112}^3 p_{122} p_{212} p_{221}}{p_{111}^3 p_{121} p_{211} p_{222}},\\ \beta_5 &= \log\,\frac{p_{111}^2 p_{121}^2 p_{212}^2 p_{222}^2}{p_{112}^2 p_{122}^2 p_{211}^2 p_{221}^2}, & \beta_6 &= \log\,\frac{p_{111}^2 p_{122}^2 p_{211}^2 p_{222}^2}{p_{112}^2 p_{121}^2 p_{212}^2 p_{221}^2}.\end{aligned}$$

The mean-value parameters for this family are $\boldsymbol{\zeta}_1 = N\mathbf{A}\boldsymbol{p}$ (the expected values of the subset sums). The mixed parameterization consists of the mean-value parameters and the canonical parameters $\boldsymbol{\zeta}_2 = (\theta_1, \theta_2)'$ or $\tilde{\boldsymbol{\zeta}}_2 = \mathbf{D}\log \boldsymbol{p}$. □

Some models, more general than relational models, can be specified by setting generalized odds ratios equal to positive constants. An example of such a model is given next.

Example 5.1. The Hardy-Weinberg distribution arising in genetics was discussed as an exponential family by Barndorff-Nielsen (1978) and Brown (1988), among others. Assume that a parent population contains alleles G and g with probabilities π and $1-\pi$ respectively. The number of genotypes GG, Gg, and gg, that appear in a generation of N descendants, is a random variable with $M(N, \boldsymbol{p})$ distribution. Under the model of random mating and no selection, the vector of probabilities $\boldsymbol{p}$ has components

$$p_1 = \pi^2,\; p_2 = 2\pi(1-\pi),\; p_3 = (1-\pi)^2. \tag{29}$$

The model (29) is a one-parameter regular exponential family with the canonical parameter $\log\frac{\pi}{1-\pi}$. This model is slightly more general than relational models, but the techniques used for relational models apply. The model representation in terms of homogeneous odds ratios is

$$\frac{p_2^2}{p_1p_3} = 4. \tag{30}$$

If the kernel basis matrix is chosen as $\mathbf{D} = (-1, 2, -1)$ and the model matrix is

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \end{pmatrix},$$

the model (30) can be expressed as

$$\mathbf{D}\text{log } \boldsymbol{p} = 2\text{log } 2.$$

There exists a mixed parameterization of the family of multinomial distributions of the form

$$\log \boldsymbol{p} = \mathbf{A}'\boldsymbol{\beta} + \mathbf{D}'\theta. \tag{31}$$

Here $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ and $\theta \in (-\infty, \infty)$. From the equation (27):

$$\theta = \frac{1}{6}\text{log } \frac{p_2^2}{p_1p_3}.$$

The parameter θ may be interpreted as a measure of the strength of selection in favor of the heterozygote character Gg (cf. Brown, 1988).

The condition $\mathbf{D}\text{log } \boldsymbol{p} = \log 4$ is equivalent to setting the parameter θ equal to $\frac{1}{6}\text{log } \frac{1}{4}$. □

It is well known for a multidimensional contingency table that marginal distributions are variation independent of conditional odds ratios. Properly selected conditional odds ratios and sets of marginal distributions determine the distribution of the table uniquely (Barndorff-Nielsen, 1976; Rudas, 1998; Bergsma & Rudas, 2003). A generalization of this fact to the set $\mathcal{I}$ is given in the following theorem.

Theorem 5.2. *Let $\mathcal{P}$ be the set of all positive distributions on the table $\mathcal{I}$. Suppose $\mathbf{A}$ is a non-negative integer matrix of full row rank and $\mathbf{D}$ is a kernel basis matrix of $\mathbf{A}$. Then the following statements hold:*

(i) For any $P_{\boldsymbol{\delta}_1}$, $P_{\boldsymbol{\delta}_2} \in \mathcal{P}$ there exist a distribution $P_{\boldsymbol{\delta}} \in \mathcal{P}$ and a scalar α such that

$$\mathbf{A}\boldsymbol{\delta} = \alpha\mathbf{A}\boldsymbol{\delta}_1 \textit{ and } \mathbf{D}\textit{log } \boldsymbol{\delta} = \mathbf{D}\textit{log } \boldsymbol{\delta}_2.$$

(ii) The coefficient of proportionality $\alpha = 1$ for any $P_{\boldsymbol{\delta}_1}$, $P_{\boldsymbol{\delta}_2} \in \mathcal{P}$ if and only if $\mathbf{1} \in R(\mathbf{A})$.

The proof is straightforward, by Theorem 4.1 and Corollary 4.2, and is omitted here.

6 Applications

The first example features relational models as a potential tool for modeling social mobility tables. A model of independence is considered on a space that is not the Cartesian product of the domains of the variables in the table.

Example 6.1. Social mobility tables often express a relation between statuses of two generations, for example, the relation between occupational statuses of respondents and their fathers, as in Table 7 (Blau & Duncan, 1967). To test the hypothesis of independence between respondent's mobility and father's status, consider the respondent's mobility variable with three categories: Upward mobile (moving up compared to father's status), Immobile (staying at the same status), and Downward mobile (moving down compared to father's status). The initial table is thence transformed into Table 8.

Table 7: Occupational Changes in a Generation, 1962

Father's occupation	Respondent's occupation		
	White-collar	Manual	Farm
White-collar	6313	2644	132
Manual	6321	10883	294
Farm	2495	6124	2471

Table 8: Father's occupation vs Respondent's mobility. The MLEs are shown in parentheses.

Father's occupation	Respondent's mobility		
	Upward	Immobile	Downward
White-collar	-	6313 (7518.17)	2776 (1570.83)
Manual	6321 (8823.66)	10883 (7175.18)	294 (1499.17)
Farm	8619 (6116.34)	2471 (4973.66)	-

Since respondents cannot move up from the highest status or down from the lowest status, then the cells $(1,1)$ and $(3,3)$ in Table 8 do not exist. The set of cells $\mathcal{I}$ is a proper subset of the Cartesian product of the domains of the variables in the table. Let $\mathbf{S}$ be the class consisting of the cylinder sets associated with the marginals, including the empty one. The relational model generated by $\mathbf{S}$ has the model matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

and is expressed in terms of local odds ratios as follows:

$$\frac{p_{12}p_{23}}{p_{13}p_{22}} = 1, \;\; \frac{p_{21}p_{32}}{p_{22}p_{31}} = 1.$$

This model is a regular exponential family of order 4; the maximum likelihood estimates of cell frequencies exist and are unique. (The estimates are shown in Table 8 next to the observed values.) The observed $X^2 = 6995.83$ on two degrees of freedom provides an evidence of strong association between father's occupation and respondent's mobility. □

The next example illustrates the usefulness of relational models for network analysis.

Example 6.2. Table 9 shows the total trade data between seven European countries that were collected from *United Nations Commodity Trade Statistics Database* (2007). Every cell contains the value of trade volume for a pair of countries; cell counts are assumed to have Poisson distribution. The two hypotheses of interest are: countries with larger economies generate more trade, and trade volume between two countries is higher if they use the same currency. In this example, GDP (gross domestic product) is chosen as the characteristic of economy and Eurozone membership is chosen as the common currency indicator. The class **S** includes five subsets of cells reflecting the GDP size:

$$\begin{array}{l}
\{GDP < 0.1 \cdot 10^6 \;\text{ vs }\; GDP < 0.1 \cdot 10^6\}, \\
\{GDP < 0.1 \cdot 10^6 \;\text{ vs }\; 0.1 \cdot 10^6 \leq GDP < 0.6 \cdot 10^6\}, \\
\{GDP < 0.1 \cdot 10^6 \;\text{ vs }\; GDP \geq 0.6 \cdot 10^6\}, \\
\{0.1 \cdot 10^6 \leq GDP < 0.6 \cdot 10^6 \;\text{ vs }\; 0.1 \cdot 10^6 \leq GDP < 0.6 \cdot 10^6\}, \\
\{0.1 \cdot 10^6 \leq GDP < 0.6 \cdot 10^6 \;\text{ vs }\; GDP \geq 0.6 \cdot 10^6\},
\end{array}$$

and three subsets reflecting Eurozone membership:

$$\begin{array}{l}
\{\text{cells showing trade between two Eurozone members }\}, \\
\{\text{cells showing trade between a Eurozone member and a non-member }\}, \\
\{\text{cells showing trade between two Eurozone non-members}\}.
\end{array}$$

Under the model generated by **S**, trade volume is the product of the GDP effect and the Eurozone membership effect.

This model is a regular exponential family of order 6. The maximum likelihood estimates for cell frequencies exist and are unique. The observed $X^2 = 20.16$ on 14 degrees of freedom yields the asymptotic p-value of 0.12; so the model fits the trade data well. Alternatively, sensitivity of the model fit to other choices regarding GDP could also be studied. □

Acknowledgments

The authors wish to thank Thomas Richardson for his comments.

Table 9: Total trade between seven countries (in billion US dollars). The MLEs are shown in parentheses.

	LV	NLD	FIN	EST	SWE	BEL	LUX
LV	[0]	0.7 (3.29)	1 (1.17)	2 (2.0)	1.3 (1.17)	0.4 (1.17)	0.01 (0.01)
NLD	-	[0]	10 (17)	1 (1.17)	17 (15)	102 (102)	2.1 (2.29)
FIN	-	-	[0]	4 (1.17)	18 (15)	4 (2.29)	0.1 (2.29)
EST	-	-	-	[0]	2.6 (1.17)	0.5 (1.17)	0.01 (0.01)
SWE	-	-	-	-	[0]	15 (15)	0.35 (2.29)
BEL	-	-	-	-	-	[0]	9 (6.41)
LUX	-	-	-	-	-	-	[0]

The work of Anna Klimova and Adrian Dobra was supported in part by NIH Grant R01 HL092071.

Tamás Rudas was supported in part by Grant No TAMOP 4.2.1./B-09/1/KMR-2010-0003 from the European Union and the European Social Fund. He is also a Recurrent Visiting Professor at the Central European University and the moral support received is acknowledged.

References

Agresti, A. (2002). *Categorical data analysis.* New York: Wiley.

Barndorff-Nielsen, O. E. (1976). Factorization of likelihood functions for full exponential families. *J. Roy. Statist. Soc. Ser.B*, *38*, 37–44.

Barndorff-Nielsen, O. E. (1978). *Information and exponential families.* New York: Wiley.

Bergsma, W., & Rudas, T. (2003). On conditional and marginal association. *Ann. Fac. Sci. Toulouse*, *11*, 455–468.

Bertsekas, D. P. (2009). *Convex optimization theory.* Nashua, NH: Athena Scientific.

Birch, M. W. (1963). Maximum likelihood in three-way contingency tables. *J. Roy. Statist. Soc. Ser.B*, *25*, 220–233.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice.* MIT.

Blau, P. M., & Duncan, O. D. (1967). *The American occupational structure.* New York: John Wiley and Sons, Inc.

Brown, L. D. (1988). *Fundamentals of statistical exponential families.* Hayward, Calif.: Institute of Mathematical Statistics.

Christensen, R. (1997). *Log-linear models and logistic regression.* New York: Springer.

Csiszár, I. (1975). I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, *3*, 146–158.

Efron, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). *Ann. Statist.*, *3*, 1189–1242.

Goodman, L. A. (1968). The analysis of cross-classified data: independence, quasi-independence, and interaction in contingency tables with or without missing cells. *J. Amer. Statist. Assoc.*, *63*, 1091–1131.

Goodman, L. A. (1972). Some multiplicative models for the analysis of cross-classified data. *Proceedings of the Sixth Berkley Symposium on Mathematical Statistics and Probability.*

Hara, H., Takemura, A., & Yoshida, R. (2009). Markov bases for two-way subtable sum problems. *J. Pure Appl. Algebra*, *213*, 1507–1521.

Hauser, R. M. (1978). A structural model of the mobility table. *Social Forces*, *56*, 919–953.

Hoffmann-Jørgensen, J. (1994). *Probability with a view toward statistics* (Vol. 2). New York: Chapman & Hall.

Hout, M. (1983). *Mobility tables* (Vol. 31). Sage Publications, Inc.

Kass, R. E., & Vos, P. W. (1997). *Geometrical foundations of asymptotic inference.* New York: Wiley.

Kawamura, G., Matsuoka, T., Tajiri, T., Nishida, M., & Hayashi, M. (1995). Effectiveness of a sugarcane-fish combination as bait in trapping swimming crabs. *Fisheries Research*, *22*, 155–160.

Pachter, L., & Sturmfels, B. (Eds.). (2005). *Algebraic statistic for computational biology.* Cambridge university press.

Rudas, T. (1998). *Odds ratios in the analysis of contingency tables.* Sage Publications, Inc.

Rudin, W. (1976). *Principles of mathematical analysis.* McGraw-Hill.

Schrijver, A. (1986). *Theory of linear and integer programming.* New York: Wiley.

Sturmfels, B. (1996). *Gröbner bases and convex polytopes.* Providence RI: AMS.

Tanner, M. A., & Young, M. A. (1985a). Modeling agreement among raters. *J. Amer. Statist. Assoc.*, *80*, 175–180.

Tanner, M. A., & Young, M. A. (1985b). Modeling agreement among raters. *Psychological Bulletin*, *98*, 408–415.

United nations commodity trade statistics database. (2007). Available from `http://comtrade.un.org/`

Zelterman, D., & Youn, T. I. (1992). Indicator models for social mobility tables. *Comput. Statist. Data Anal.*, *14*.