

Exact Tests to Compare Contingency Tables Under Quasi-independence and Quasi-symmetry

Cristiano Bocci¹, Fabio Rapallo^{2,*}

¹ *Department of Information Engineering and Mathematics, University of Siena, Siena, Italy*

² *Department of Science and Technological Innovation, University of Piemonte Orientale, Alessandria, Italy*

Abstract. In this work we define log-linear models to compare several square contingency tables under the quasi-independence or the quasi-symmetry model, and the relevant Markov bases are theoretically characterized. Through Markov bases, an exact test to evaluate if two or more tables fit a common model is introduced. Two real-data examples illustrate the use of these models in different fields of applications.

2000 Mathematics Subject Classifications: 62H17

Key Words and Phrases: Algebraic Statistics, Markov bases, MCMC algorithms, Rater agreement, Social mobility tables

1. Introduction

Complex models for contingency tables have received an increasing interest in the last decades from researchers and practitioners in different fields, from Biology to Medicine, from Economics to Social Science. For a general introduction to the statistical models for contingency tables see for instance [1], [4] and [11]. Quasi-symmetry and quasi-independence models are well known log-linear models for square contingency tables. Starting from Caussinus in [6], several authors have considered such models from the point of view of both theory and applications, and it is impossible to give a complete account on all the papers where quasi-independence and quasi-symmetry are studied or used in data analysis. In the next section we will recall the basic facts on the quasi-independence and quasi-symmetry models, while for a full presentation and an historical overview the reader can refer to [4] and [9]. Quasi-symmetry is also the topic of a special issue of the *Annales de la Faculté des Sciences de Toulouse*, edited in 2002 by S. Fienberg and P. G. M. van der Heijden [8].

*Corresponding author.

Email addresses: cristiano.bocci@unisi.it (C. Bocci), fabio.rapallo@uniupo.it (F. Rapallo)

Within Algebraic Statistics, quasi-independence and quasi-symmetry are very important models for contingency tables, for several reasons. We briefly review why the synergy between Algebraic Statistics and quasi-independence has been fruitful. Firstly, Algebraic Statistics provides an exact goodness-of-fit test based on the Diaconis-Sturmfels algorithm. Such a test is very flexible when applied to complex models, and it allows us to make exact inference also outside the basic independence model, where the classical Fisher's exact test is available. When the sample size is small, the use of the asymptotic tests based on the chi-square approximation of the test statistics may lead to wrong conclusions, and this fact is even more relevant in this kind of models, where the asymptotics fail also with moderately large sample sizes, see an example in [13]. Secondly, under quasi-independence it is possible to fix the diagonal counts, or even to analyze incomplete tables where the diagonal counts (or an arbitrary subset of cells) are undefined or unavailable. To include structural zeros in the analysis, the notion of toric statistical model is a generalization of log-linear model that allows us to study also the boundary. Toric models are described by non-linear polynomials, but in several cases it is possible to describe the geometry of such models, or at least it is possible to write their invariants through Computer Algebra systems. Quasi-independence and quasi-symmetry from the point of view of Algebraic Statistics can be found in [13], [3], [7], and [2]. Applications of quasi-symmetry to the problem of rater agreement in biomedical experiments are presented in [14].

In this paper, we use classical techniques from Algebraic Statistics in order to compare several contingency tables under the quasi-independence and quasi-symmetry models. This is accomplished by the construction of a three-way table and by the definition of suitable log-linear models in order to determine if two or more tables fit a common quasi-independence (resp., quasi-symmetry) model, versus the alternative hypothesis that each table follows a specific quasi-independence (resp., quasi-symmetry) model with its own parameters. A third model is also introduced, as its matrix representation is a well known object in Combinatorics, namely the Lawrence lifting of a matrix. For the first two models, the relevant Markov bases are computed theoretically using a distance-reducing argument, while for the third model the Markov bases are characterized only in the case of two tables, and some advices are presented to efficiently run the exact test in the general case. We show two applications of these models on datasets coming from different areas: the first example comes from a rater agreement problem in a biomedical experiments, while the second one concerns the analysis of social mobility tables.

This research suggests several future directions. Since one of the models described here can be written in the form of a toric fiber product, it would be interesting to study other forms of toric fiber products with interesting statistical meaning. From the point of view of Applied Statistics, it would be interesting to study the use of this technique to make inference on other measures of mobility based on log-linear models, also including one-sided tests and their semi-algebraic characterization. For an introductory overview of these measures, with several examples from surveys in European countries, refer to [22] and [5].

The paper is organized as follows. In Sect. 2 we recall some definitions and basic results about log-linear models and toric models, with special attention to quasi-independence

and quasi-symmetry. In particular, we collect here several scattered results on the Markov bases for these models. In Sect. 3 we show how to define suitable log-linear models to compare two or more square tables. Given a base log-linear model, we define new log-linear models through the specification of their model matrices. For quasi-independence and quasi-symmetry on several tables, the Markov bases are theoretically computed. Sect. 4 is devoted to the illustration of two real-data examples.

2. Markov bases for quasi-independence and quasi-symmetry

In this section we recall some basic definitions and properties of log-linear models, with special attention to quasi-independence and quasi-symmetry for square two-way tables. A probability distribution on a finite sample space \mathcal{X} with K elements is a normalized vector of K non-negative real numbers. Thus, the most general probability model is the simplex

$$\Delta = \left\{ (p_1, \dots, p_K) : p_k \geq 0, \sum_{k=1}^K p_k = 1 \right\}.$$

A statistical model is therefore a subset of Δ .

A classical example of finite sample space is the case of a multi-way contingency table where the cells are the joint counts of two or more random variables with a finite number of levels each. In the case of square two-way contingency tables, where the sample space is usually written as a cartesian product of the form $\mathcal{X} = \{1, \dots, I\} \times \{1, \dots, I\}$, we will use the notation $p_{i,j}$ to ease the readability. In such a case, the two categorical variables are denoted with X and Y .

In the classical theory of log-linear models, under the Poisson sampling scheme the cell counts are independent and identically distributed Poisson random variables with expected values Np_1, \dots, Np_K , where N is the sample size, and the statistical model is defined by constraints on the raw parameters p_1, \dots, p_K . A model is log-linear if the log-probabilities lie in an affine subspace of the vector space \mathbf{R}^K . Given d real parameters $\alpha_1, \dots, \alpha_d$, a log-linear model is described, apart from normalization, through the equations:

$$\log(p_k) = \sum_{r=1}^d A_{k,r} \alpha_r \quad (1)$$

for $k = 1, \dots, K$, where A is the model matrix (or design matrix, [12]). Exponentiating Eq. (1), we obtain the expression of the corresponding toric model

$$p_k = \prod_{r=1}^d \zeta_r^{A_{k,r}} \quad (2)$$

for $k = 1, \dots, K$, where $\zeta_r = \exp(\alpha_r)$, $r = 1, \dots, d$, are new non-negative parameters. Allowing the ζ_r 's to be non-negative (instead of strictly positive) leads us to consider also the boundary of the models, with points having some entries equal to zero. It follows that

the model matrix A is also the matrix representation of the minimal sufficient statistic of the model. The matrix representation of the toric models as in Eq. (2) is widely discussed in, e.g., [15] and [7]. It is easy to see from Eq. (1) that different model matrices with the same image as vector space generate the same log-linear model.

In the two-way case, the simplest (and widely studied) log-linear model is the independence model, which models the stochastic independence between the two categorical variables X and Y . Its log-linear form is

$$\log(p_{i,j}) = \mu + \alpha_i^{(X)} + \beta_j^{(Y)} \quad (3)$$

with the constraints

$$\sum_{i=1}^I \alpha_i^{(X)} = 0, \quad \sum_{j=1}^I \beta_j^{(Y)} = 0. \quad (4)$$

Quasi-independence and quasi-symmetry are both derived from the independence model adding constraints on given subsets of cells (typically, the cells on the main diagonal) and constraints on the symmetry of the table. Although quasi-independence can be defined for general rectangular tables with fixed counts on an arbitrary subset of \mathcal{X} , see [2], here we restrict our attention to the case of square tables with fixed counts on the main diagonal. The log-linear form of the quasi-independence model is

$$\log(p_{i,j}) = \mu + \alpha_i^{(X)} + \beta_j^{(Y)} + \gamma_i \delta_{i,j} \quad (5)$$

where $\delta_{i,j}$ is the Kronecker delta. Also in the quasi-independence model the constraints in Eq. (4) hold.

The log-linear form of the quasi-symmetry model is

$$\log(p_{i,j}) = \mu + \alpha_i^{(X)} + \beta_j^{(Y)} + \gamma_{i,j} \quad (6)$$

with the constraints

$$\sum_{i=1}^I \alpha_i^{(X)} = 0, \quad \sum_{j=1}^I \beta_j^{(Y)} = 0, \quad \gamma_{i,j} = \gamma_{j,i}, \quad i, j = 1, \dots, I.$$

In Eq. (6), the $\alpha_i^{(X)}$ are the parameters of the row effect, the $\beta_j^{(Y)}$ are the parameters of the column effect, while the parameters $\gamma_{i,j}$ force the quasi-symmetry. Comparing Eqs. (1) and (6) it is easy to explicitly write the model matrix A_{qs} for the quasi-symmetry model. The first non-trivial example of quasi-independence and quasi-symmetry models is the 3×3 case, and in this first case the two models coincide as log-linear models. A possible choice is reported in Fig. 1.

Notice that in A_{qs}^t each column represents a cell of the table (the cells are ordered lexicographically for convenience), and each row represents a parameter. Analyzing the structure of A_{qs}^t , the first 7 rows of A_{qs}^t form the model matrix of the independence model A_{ind}^t , while the last three rows of A_{qs}^t define one real parameter for each diagonal cell, and

$$A_{\text{qs}}^t = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Figure 1: The model matrix of the quasi-symmetry model for $I = 3$.

hence force the diagonal cells to be fitted exactly. This parametrization is redundant, since the model has 1 degree of freedom, and therefore 8 parameters are sufficient to describe the model. The ideal of the independence model with model matrix A_{ind} is the set of all 2×2 minors of the table of probabilities, i.e.,

$$\begin{aligned} \mathcal{I}_{A_{\text{ind}}} = \text{Ideal} & (p_{1,1}p_{2,2} - p_{1,2}p_{2,1}, p_{1,1}p_{2,3} - p_{1,3}p_{2,1}, p_{1,1}p_{3,2} - p_{1,2}p_{3,1}, \\ & p_{1,1}p_{3,3} - p_{1,3}p_{3,1}, p_{1,2}p_{2,3} - p_{1,3}p_{2,2}, p_{1,2}p_{3,3} - p_{3,2}p_{2,3}, \\ & p_{2,1}p_{3,2} - p_{3,1}p_{2,2}, p_{2,1}p_{3,3} - p_{3,1}p_{2,3}, p_{2,2}p_{3,3} - p_{3,2}p_{2,3}), \end{aligned} \quad (7)$$

while for the quasi-independence model and for the quasi-symmetry model from the matrix A_{qs} we have only one binomial:

$$\mathcal{I}_{A_{\text{qind}}} = \mathcal{I}_{A_{\text{qs}}} = \text{Ideal}(p_{1,2}p_{2,3}p_{3,1} - p_{1,3}p_{3,2}p_{2,1}).$$

From the ideals above one can easily derive the corresponding Markov bases. Given a model matrix A , recall that a move is a table m with integer entries such that $Am = 0$, and that a set of moves \mathcal{M}_A is a Markov basis if all fibers

$$\mathcal{F}_{A,b} = \{f \in \mathbb{N}^k : A^t f = b\}$$

are connected. Following [7], from the point of view of computations, the easiest way to build a Markov basis is to compute the binomials in a system of generators of the toric ideal \mathcal{I}_A of A and to transform such binomials through the logs: $x^{m^+} - x^{m^-} \mapsto \pm m = \pm(m^+ - m^-)$. For instance, the ideal $\mathcal{I}_{A_{\text{qs}}}$ yields a Markov basis with only two moves:

$$m = \pm \begin{pmatrix} 0 & +1 & -1 \\ -1 & 0 & +1 \\ +1 & -1 & 0 \end{pmatrix}.$$

To conclude this section, we collect some results on Markov bases for quasi-independence and quasi-symmetry models to be found in [3], [2], [7], and [17].

A loop of degree r on \mathcal{X} is an $I \times I$ a move $m = \pm m_r(i_1, \dots, i_r; j_1, \dots, j_r)$ for $1 \leq i_1, \dots, i_r \leq I$, $1 \leq j_1, \dots, j_r \leq I$, where $m_r(i_1, \dots, i_r; j_1, \dots, j_r)$ has entries

$$\begin{aligned} m_{i_1, j_1} &= m_{i_2, j_2} = \dots = m_{i_{r-1}, j_{r-1}} = m_{i_r, j_r} = 1, \\ m_{i_1, j_2} &= m_{i_2, j_3} = \dots = m_{i_{r-1}, j_r} = m_{i_r, j_1} = -1, \end{aligned}$$

and all other elements are zero. The indices i_1, i_2, \dots , are all distinct, as well as the indices j_1, j_2, \dots , i.e.

$$i_m \neq i_n \text{ and } j_m \neq j_n \text{ for all } m \neq n,$$

A loop of degree 2, $m_2(i_1, i_2; j_1, j_2)$, is called a basic move, and a loop m_r is called df 1 if its support does not contain the support of any other loop. A loop m_r is called a symmetric loop if $\{i_1, \dots, i_r\} = \{j_1, \dots, j_r\}$.

The first result concerns quasi-independence with possible structural zeros. Let \mathcal{S} be the set of structural zeros.

Proposition 2.1. *The set of df 1 loops of degree 2, \dots, I with support on $\mathcal{X} \setminus \mathcal{S}$ forms a unique minimal Markov basis for $I \times I$ contingency tables under the quasi-independence model with possible structural zeros. When the fixed cells of the table are located only on the main diagonal, the minimal Markov basis is formed by the df 1 loops of degree 2 and 3.*

For the quasi-symmetry model we have the following

Proposition 2.2. *The set of symmetric loops of degree 3, \dots, I with support outside the main diagonal form a unique minimal Markov basis for $I \times I$ contingency tables with structural zeros under the quasi-symmetry model. This set of moves is also a Graver basis.*

3. Comparison of several tables under quasi-independence and quasi-symmetry

As outlined in the Introduction, in this section we define three log-linear models to compare two or more square tables under quasi-independence and quasi-symmetry, and we study the corresponding Markov bases. Let us consider H tables ($H \geq 2$) and define a three-way contingency table T by stacking the H tables. Conversely, each original table is a layer of the table T . Let $K' = HI^2$ be the number of cells of T . Since the definition of the new models can be done starting from a generic log-linear model on the two-way table, we present the models in a general context, and then we write the explicit log-linear representation in the case of quasi-independence and quasi-symmetry in order to highlight the meaning of such new models in our cases.

Definition 3.1. Let A be the model matrix of a log-linear model. We define three log-linear models for T :

- Under the model M_0 we assume that all the layers follow a common model with model matrix A ;
- Under the model M_1 we assume that each layer of the table T follows a model with model matrix A with its own parameters, without further constraints;
- Under the model M_2 we assume that each layer of the table T follows a model with model matrix A with its own parameters, and with the additional constraint of fixed marginal sums over the layers.

The model matrices of M_0 , M_1 and M_2 have a simple block structure. In fact:

$$A_{M_0}^t = \left(\begin{array}{c|c|c} A^t & \cdots & A^t \\ \hline \mathbf{1}_K & & \\ \hline & \ddots & \\ \hline & & \mathbf{1}_K \end{array} \right) \quad A_{M_1}^t = \left(\begin{array}{c|c|c} A^t & & \\ \hline & \ddots & \\ \hline & & A^t \end{array} \right) \quad A_{M_2}^t = \left(\begin{array}{c|c|c} A^t & & \\ \hline & \ddots & \\ \hline & & A^t \\ \hline \mathbf{I}_K & \cdots & \mathbf{I}_K \end{array} \right)$$

where $\mathbf{1}_K$ is a row vector of 1's with length K , \mathbf{I}_K is a the identity matrix with dimensions $K \times K$ and each empty block means a block filled with 0's. The matrix $A_{M_2}^t$ is the H -th order Lawrence lifting of A^t and its properties in terms of Markov and Graver bases have been studied in [19] and [18].

As noticed in the Introduction, it is easy to see from the matrix representations above that the model M_0 is built up from a base model (with model matrix A) taking the same model with the same parameters on all the layers. On the other hand, the model M_1 takes the same base model on all the layers, but it allows different model parameters on each layer. Writing explicitly the log-linear form of the three models in the case of quasi-independence we have the equations below. For M_0 :

$$(M_0) \quad \log(p_{i,j,h}) = \mu + \mu_h + \alpha_i^{(X)} + \beta_j^{(Y)} + \gamma_i \delta_{i,j} \tag{8}$$

with the constraint $\sum_{h=1}^H \mu_h = 0$ in addition to the constraints on $\alpha_i^{(X)}$ and $\beta_j^{(Y)}$ naturally derived from the basic quasi-independence model in Eq. (5). The second model M_1 is defined by

$$(M_1) \quad \log(p_{i,j,h}) = \mu + \mu_h + \alpha_{i,h}^{(X)} + \beta_{j,h}^{(Y)} + \gamma_{i,h} \delta_{i,j} \tag{9}$$

with the constraint $\sum_{h=1}^H \mu_h = 0$ in addition to the constraints on $\alpha_{i,h}^{(X)}$ and $\beta_{j,h}^{(Y)}$ derived from the basic quasi-independence model in Eq. (5) and valid on each layer of the table T . The third model M_2 is defined by

$$(M_2) \quad \log(p_{i,j,h}) = \mu + \mu_h + \mu_{i,j} + \alpha_{i,h}^{(X)} + \beta_{j,h}^{(Y)} + \gamma_{i,h} \delta_{i,j} \tag{10}$$

with the same constraints as in M_1 plus the additional constraints $\sum_{i=1}^I \mu_{i,j} = 0, j = 1, \dots, I$ and $\sum_{j=1}^I \mu_{i,j} = 0, i = 1, \dots, I$.

In the case of quasi-symmetry we obtain the expressions below. For M_0 :

$$(M_0) \quad \log(p_{i,j,h}) = \mu + \mu_h + \alpha_i^{(X)} + \beta_j^{(Y)} + \gamma_{i,j} \tag{11}$$

with the constraint $\sum_{h=1}^H \mu_h = 0$ in addition to the constraints on $\alpha_i^{(X)}$, $\beta_j^{(Y)}$ and $\gamma_{i,j}$ naturally derived from the basic quasi-symmetry model in Eq. (6). The second model M_1 is defined by

$$(M_1) \quad \log(p_{i,j,h}) = \mu + \mu_h + \alpha_{i,h}^{(X)} + \beta_{j,h}^{(Y)} + \gamma_{i,j,h} \quad (12)$$

with the constraint $\sum_{h=1}^H \mu_h = 0$ in addition to the constraints on $\alpha_{i,h}^{(X)}$, $\beta_{j,h}^{(Y)}$ and $\gamma_{i,j,h}$ derived from the basic quasi-symmetry model in Eq. (6) and valid on each layer of the table T . The third model M_2 is defined by

$$(M_2) \quad \log(p_{i,j,h}) = \mu + \mu_h + \mu_{i,j} + \alpha_{i,h}^{(X)} + \beta_{j,h}^{(Y)} + \gamma_{i,j,h} \quad (13)$$

with the same constraints as in M_1 plus the additional constraints $\sum_{i=1}^I \mu_{i,j} = 0, j = 1, \dots, I$ and $\sum_{j=1}^I \mu_{i,j} = 0, i = 1, \dots, I$.

With a simple linear algebra argument, it easy to see that $M_0 \subset M_1 \subset M_2$. As a consequence, these models can be used in two ways. They can be applied separately, to define goodness-of-fit test to contrast the observed table with a given model, or they can be used to define a test for nested models, see e.g. [1] where several examples are introduced and discussed. In the next section we will focus on tests for nested models.

Now, we compute the Markov bases for the models M_0, M_1 and M_2 defined above. Let \mathcal{M} be a Markov basis for quasi-independence (or quasi-symmetry). In the proofs below we will make use of a distance-reducing argument, introduced in [21].

For the model M_0 , define the following two types of moves b :

- type 1: fix a move $m \in \mathcal{M}$ and split it on the different layers with the condition that each row of m belongs to one layer.
- type 2: choose integers $1 \leq i_1 < i_2 \leq I$ and $1 \leq j_1 < j_2 \leq I$ and define the moves $\pm b$ where b has zero coordinate except for

$$\begin{aligned} b_{i_1, j_1, h_1} &= 1 \\ b_{i_2, j_2, h_1} &= -1 \\ b_{i_1, j_1, h_2} &= -1 \\ b_{i_2, j_2, h_2} &= 1 \end{aligned}$$

for $1 \leq h_1 < h_2 \leq H$.

Consider as a set of moves

$$\mathcal{M}_0 = \mathcal{B}_1 \cup \mathcal{B}_2$$

where \mathcal{B}_i is the set of moves of type i .

Proposition 3.2. *The set \mathcal{M}_0 above is a Markov basis for the model M_0 .*

Proof. Let $v \in \mathbb{Z}^{I \times I \times H}$. Then $v \in \text{Ker}_{\mathbb{Z}}(A_{M_0}^t)$ if and only if

- i) $A^t \left(\sum_{h=1}^H v_{\bullet, \bullet, h} \right) = 0$ that is $\sum_{h=1}^H v_{\bullet, \bullet, h} \in \text{Ker}_{\mathbb{Z}}(A^t)$;

$$\text{ii) } \sum_{i,j=1}^I v_{i,j,h} = 0 \text{ for all } 1 \leq h \leq H$$

where ii) follows directly looking at the vectors of ones $\mathbf{1}_K$ in the definition of $A_{M_0}^t$.

Let u, v be vectors with same value of the sufficient statistic, i.e. $A_{M_0}^t u = A_{M_0}^t v$. We want to prove that there exists $b \in \mathcal{B}$ such that

$$u + b \geq 0$$

and

$$\|u + b - v\|_1 < \|u - v\|_1.$$

Since u and v are distinct with $A_{M_0}^t u = A_{M_0}^t v$ then there exists a positive entry in $u - v$, say $u_{i_1, j_1, h_1} - v_{i_1, j_1, h_1} > 0$. Suppose that such entry belongs to the main diagonal, i.e., $i_1 = j_1$. Then there exists another layer h_2 such that $u_{i_1, j_1, h_2} - v_{i_1, j_1, h_2} < 0$. Moreover, by the condition ii), there exists a positive entry of $u - v$ in the layer h_2 and a negative entry in the layer h_1 : $u_{i_2, j_2, h_1} - v_{i_2, j_2, h_1} < 0$ and $u_{i_3, j_3, h_2} - v_{i_3, j_3, h_2} > 0$ for some indices i_2, i_3, j_2, j_3 . Now consider the move $b \in \mathcal{B}_2$ defined by

$$\begin{aligned} b_{i_1, j_1, h_1} &= -1 \\ b_{i_1, j_1, h_2} &= +1 \\ b_{i_3, j_3, h_2} &= -1 \\ b_{i_3, j_3, h_1} &= +1 \end{aligned}$$

that satisfies $\|u - v\|_1 > \|u + b - v\|_1$.

Thus, we can consider only the case where $u - v$ is zero on the main diagonal of all layers. Let U and V be the sum over the H layers of u and v , respectively. By condition i), $U - V \in \text{Ker}_{\mathbb{Z}}(A^t)$. By Props. 2.1 and 2.2, there exists a distance reducing move m which is a df 1 loop (in the case of quasi-independence) or a symmetric loop (in the case of quasi-symmetry):

$$\|U + m - V\|_1 < \|U - V\|_1$$

Let $\mathcal{I} = \{(i_1, j_1), (i_1, j_2), \dots, (i_t, j_t), (i_t, j_1)\}$ be the set of the indices where m has nonzero entries. Without loss of generality we can suppose that $m_{i_1, j_1} = -1$. The move m is defined such that

$$\begin{aligned} m_{\alpha, \beta} &= +1 && \text{if and only if} && U_{\alpha, \beta} - V_{\alpha, \beta} < 0 \\ m_{\alpha, \beta} &= -1 && \text{if and only if} && U_{\alpha, \beta} - V_{\alpha, \beta} > 0 \end{aligned}$$

for all $(\alpha, \beta) \in \mathcal{I}$, except at most for the last index (i_t, j_1) , where $m_{i_t, j_1} = +1$.

Now, we split m in the H layers. For this aim, notice that $U_{\alpha, \beta} - V_{\alpha, \beta} < 0$ implies that there exists a layer h such that $u_{\alpha, \beta, h} - v_{\alpha, \beta, h} < 0$, and similarly $U_{\alpha, \beta} - V_{\alpha, \beta} > 0$ implies that there exists a layer h such that $u_{\alpha, \beta, h} - v_{\alpha, \beta, h} > 0$. We then split m by putting the +1 in the layer with a negative entry of $u - v$ and -1 in the layer with a positive entry of $u - v$. This can be done for all $(\alpha, \beta) \in \mathcal{I} \setminus (i_t, j_1)$. The last +1 can be assigned to an arbitrary layer. Thus we have a matrix b defined by a sequence of indices

$$\mathcal{I}' = \{(i_1, j_1, h_{11}), (i_1, j_2, h_{12}), \dots, (i_t, j_t, h_{tt}), (i_t, j_1, h_{t1})\}$$

Now, we arrange the entries in such a way that the entries in each row belong to the same layer. This ensures that the split move b satisfies the condition ii) above.

Let us consider the first row and the corresponding indices $(i_1, j_1, h_{11}), (i_1, j_2, h_{12})$ where b is nonzero. By construction, $u_{i_1, j_1, h_{11}} - v_{i_1, j_1, h_{11}} > 0$ and $u_{i_1, j_2, h_{12}} - v_{i_1, j_2, h_{12}} < 0$. If $h_{11} = h_{22}$ we have concluded, otherwise there is an entry (α, β, h_{12}) in the layer h_{12} such that $u_{\alpha, \beta, h_{12}} - v_{\alpha, \beta, h_{12}} > 0$. Now, consider the preliminary move $b^{(p)} \in \mathcal{B}_2$ defined by

$$\begin{aligned} b_{\alpha, \beta, h_{12}}^{(p)} &= -1 \\ b_{i_1, j_1, h_{12}}^{(p)} &= +1 \\ b_{i_1, j_1, h_{11}}^{(p)} &= -1 \\ b_{\alpha, \beta, h_{11}}^{(p)} &= +1. \end{aligned}$$

If $u_{i_1, j_1, h_{12}} - v_{i_1, j_1, h_{12}} < 0$, then the move $b^{(p)}$ is distance-reducing and we have concluded. Otherwise, $\|u + b^{(p)} - v\|_1 = \|u - v\|_1$ and $u_{i_1, j_1, h_{12}} + b_{i_1, j_1, h_{12}}^{(p)} - v_{i_1, j_1, h_{12}} > 0$, and therefore we can move the -1 into the layer h_{12} . Now, the remaining $+1$'s in b can be simply moved, row by row, in the same layer of the corresponding -1 and this is enough to conclude: there is a move $b \in \mathcal{B}_1$ such that $\|u + b^{(p)} + b - v\|_1 < \|u - v\|_1$.

Remark 3.3. The Markov basis for the model M_0 can be computed also within the framework of toric fiber products, see [20], by passing to consider the ideal $I \subset K[x_{1,1}, \dots, x_{I,I}]$ of the common model for all the layers and the empty ideal $J = (0) \subset K[y_1, \dots, y_H]$. The ideal of the model M_0 is given by the kernel $I \times J$ of the map

$$\begin{array}{ccc} K[p_{1,1,1}, \dots, p_{I,I,H}] & \rightarrow & K[x_{1,1}, \dots, x_{I,I}]/I \otimes_K K[y_1, \dots, y_H]/J \\ p_{i,j,h} & \mapsto & x_{i,j} \otimes y_h \end{array}$$

The procedure defined in [20] to write the generators of such an ideal through the lift of the binomials of the base model plus suitable quadratic moves yields more generators than required. Thus we prefer to prove Prop. 3.2 by a more constructive proof, based on the distance reduction method.

However, our proof strongly depends on the form of the base model and on the corresponding Markov basis. Thus, Prop. 3.2 is valid only when the base model is the quasi-independence or the quasi-symmetry model. Of course a similar proof can be carried out for some other special base models, but a proof for a generic model matrix A would require the toric fiber product, and this is beyond the scopes of this paper.

Example 3.4. If $I = 3$ and $H = 2$ an example of move of type 1, for both the quasi-independence and the quasi-symmetry models, is drawn in Fig. 2.

We remark that, while in the quasi-independence model the diagonal cells are fixed, this is no longer true when two or more tables are compared under the model M_0 . In fact, the moves in \mathcal{B}_2 act also on the diagonal cells.

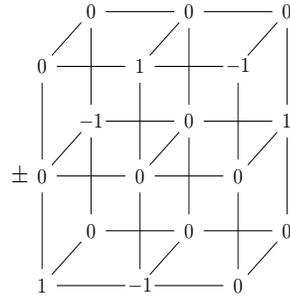


Figure 2: An example of split move of type 1 for the model M_0 .

Remark 3.5. Let $H = 2$. In this case the minimal Markov basis for M_0 has $I^2(I^2 - 1)/2$ moves of type 2 and $I + 1$ moves of type 1 for any move of degree d in the base model. These last ones are obtained by putting in the first layer the first t rows, with $0 \leq t \leq I$, of the move in the base model, and the remaining rows in the second layer.

In general, it is possible to move a row of a move of type 1 in a different layer using two moves of type 2. For example, if we consider the move of Fig. 2, which is not in the minimal Markov basis for the model M_0 , this can be written as in Fig. 3. Notice that in such decomposition the -1 's appear, for the first time, where they are in the move to decompose, or where previously there was a $+1$. Thus, if the original move can be applied then also the three successive moves in Fig. 3 can be applied; these are, in order, a move of type 2, a move of type 1 in the minimal Markov basis and another move of type 2. In

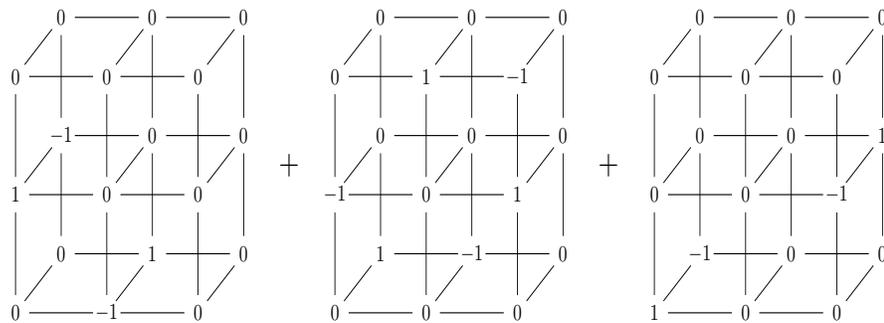


Figure 3: The decomposition of the move for the model M_0 in Fig. 2 using moves in the minimal Markov basis.

the proof of Prop. 3.2, we did not restrict to use only minimal moves to make the proof more readable.

By the previous remark we get that the degrees of moves in the model M_0 are not greater than the degrees of moves in the base model.

The model M_1 is easy to analyze, and the moves are given by the following proposition. Although other proofs of this result are possible, we provide a proof using again a distance-reducing argument.

Proposition 3.6. *The set of moves*

$$\mathcal{M}_1 = \{(m, 0, \dots, 0), (0, m, 0, \dots, 0), \dots, (0, \dots, 0, m) : m \in \mathcal{M}\}$$

is a Markov basis for M_1 .

Proof. Let $v \in \mathbb{Z}^{I \times I \times H}$ and denote by $v|_h$ the h -th layer of v , for $1 \leq h \leq H$. Then $v \in \ker_{\mathbb{Z}}(A_{M_1}^t)$ if and only if $v|_h \in \ker_{\mathbb{Z}}(A^t)$ for all $h = 1, \dots, H$.

Let u, v be two vectors with same value of the sufficient statistic, i.e. $A_{M_1}^t u = A_{M_1}^t v$. We want to prove that there exists $b \in \mathcal{M}_1$ such that

$$u + b \geq 0 \quad \text{and} \quad \|u + b - v\|_1 < \|u - v\|_1. \tag{14}$$

Since u and v are distinct with $A_{M_1}^t u = A_{M_1}^t v$ then there exists a positive entry in $u - v$, say $u_{i,j,h} - v_{i,j,h} > 0$. Let us fix our attention on layer h . By hypothesis $(u - v)|_h \in \text{Ker}_{\mathbb{Z}}(A^t)$, hence one has $A^t u|_h = A^t v|_h$. For the base model we know there exists a move m such that

$$u|_h + m \geq 0 \quad \text{and} \quad \|u|_h + m - v|_h\|_1 < \|u|_h - v|_h\|_1.$$

Define the move b by putting m on the h -th layer and 0 on all other layers. It is immediate to see that such a move satisfies the conditions in Eq. (14), and this is enough to conclude.

Example 3.7. If $I = 3$ and $H = 2$ the moves for the model M_1 for quasi-independence and quasi-symmetry are drawn in Fig. 4.

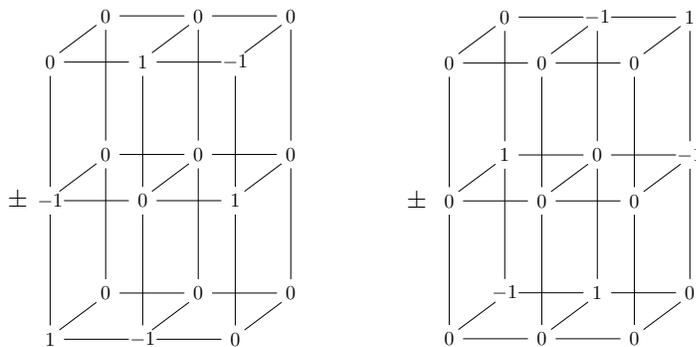


Figure 4: The moves for the model M_1 .

For the model M_2 , we restrict to the case of two layers. Let us consider a Graver basis \mathcal{G} for the model matrix A^t , and consider the set of moves

$$\mathcal{L} = \{(m, -m) : m \in \mathcal{G}\}.$$

Remark 3.8. For quasi-symmetry, the Markov basis from Prop. 2.2 is also a Graver basis, while in the case of quasi-independence we need to consider the set of all df 1 loops of degree $2, \dots, I$.

Proposition 3.9. *The set \mathcal{L} is a Graver basis (and thus also a Markov basis) for the model M_2 .*

Proof. This follows from Theorem 7.1 in [19].

Example 3.10. For example, if $I = 3$ there are only two moves for the model M_2 for both the quasi-independence and the quasi-symmetry models. They are drawn in Fig. 5.

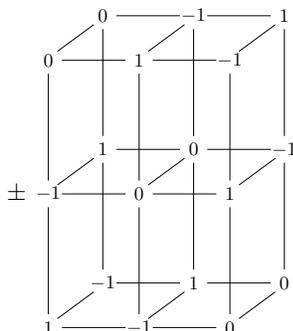


Figure 5: The moves for the model M_2 .

For more than 2 layers (i.e., for higher Lawrence configurations), the Markov basis for M_2 may be computed through 4ti2 [10], but the number of moves increases rapidly with I and H . A valid alternative in this case is to run the Markov chain of the Diaconis-Sturmfels algorithm without a Markov basis, as described in Chapter 16 of [2].

We point out that the model M_2 , based on the Lawrence lifting of the model matrix A , has mainly a theoretical interest. Nevertheless, in several applications of Algebraic Statistics the Lawrence lifting has been used, see e.g. [17], and therefore we have reported here some relevant results for the sake of completeness.

4. Examples

In this section we present two numerical examples where the comparison of two quasi-symmetry tables may be used. Before introducing the two numerical examples, we briefly recall the Diaconis-Sturmfels algorithm, by adapting the notation to the case of a test for nested models to contrast model M_0 inside the model M_1 . For details on the Diaconis-Sturmfels algorithm, see [7].

Let f be the observed table of counts, and write f as a vector of length K' according to the row labels of A_{M_0} . Moreover, let p be the vector of probabilities associated to f . The test for nested models with null hypothesis $H_0 : p \in M_0 \subset M_1$ versus $H_1 : p \in M_1$ can be done using the log-likelihood ratio statistic

$$G^2 = 2 \sum_{k=1}^{K'} f_k \log \left(\frac{\hat{f}_{0k}}{\hat{f}_{1k}} \right),$$

where \hat{f}_{0k} and \hat{f}_{1k} are the maximum likelihood estimates of the expected cell counts under the models M_0 and M_1 respectively. In the asymptotic theory, the value of G^2 must be compared with the quantiles of the chi-square distribution with the appropriate number of degrees of freedom, depending on the dimensions of the table.

We use here the Diaconis-Sturmfels algorithm based on a Markov basis \mathcal{M}_0 of the model M_0 . Given the observed table f , its reference set under M_0 is

$$\mathcal{F}(f) = \left\{ f' \in \mathbf{N}^{K'} : A_{M_0}^t f' = A_{M_0}^t f \right\}$$

and it is connected using the algorithm below. At each step:

1. let f be the current table;
2. choose with uniform probability a move $m \in \mathcal{M}_0$;
3. define the candidate table as $f_+ = f + m$;
4. generate a random number u with uniform distribution over $[0, 1]$. If $f_+ \geq 0$ and

$$\min \left\{ 1, \frac{\mathcal{H}(f_+)}{\mathcal{H}(f)} \right\} > u$$

then move the chain in f_+ ; otherwise stay at f . Here \mathcal{H} denotes the hypergeometric distribution on $\mathcal{F}(t)$.

After an appropriate burn-in-period and taking only tables at fixed times to reduce correlation between the sampled tables, the proportion of sampled tables with test statistics greater than or equal to the test statistic of the observed one is the Monte Carlo approximation of p -value of the log-likelihood ratio test. The results in this section are based on Monte Carlo samples of size 10,000, and the corresponding asymptotic p -values are displayed for comparison.

4.1. Rater agreement data

The data in Tab. 1 summarize the results of a medical experiment involving the evaluation of agreement among raters. Two independent raters are asked to assign a set of medical images to 5 different stages of a disease (levels 1 to 5 in increasing order of severity of the disease). To check the relevance of a thorough training of the raters, a first set of images has been classified before the training session, while a second set has been classified after the training session.

We use the quasi-symmetry model, and the Markov basis \mathcal{M}_0 for the model M_0 consists of $2 \cdot 1004$ moves (i.e., 1004 moves, each of them with the two signs). First, we run two exact tests for the goodness-of-fit of the two tables separately under quasi-symmetry, and we obtain exact p -values equal to 0.912 and 0.791 respectively ($G^2 = 1.578$ and $G^2 = 4.303$ respectively, with 7 df). Running the test for a unique quasi-symmetry model described in the previous section, the exact test gives a p -value equal to 0.029 ($G^2 = 30.589$ with 18 df,

Table 1: Rater agreement data. Columns represent the grading assigned by the first rater, rows represent the grading assigned by the second rater. The data in the left panel have been collected before the training session, the data in the right panel have been collected after the training session.

	1	2	3	4	5		1	2	3	4	5
1	10	2	1	4	0	1	16	5	1	0	0
2	4	8	4	1	1	2	0	15	2	1	0
3	0	0	10	3	1	3	1	3	14	1	1
4	1	1	4	11	0	4	0	2	0	14	3
5	0	1	3	3	10	5	0	2	0	3	14

corresponding to an asymptotic p -value equal to 0.014). While both the layers fit a quasi-symmetry model very well, they do not fit a common quasi-symmetry model. This means that there are significant differences in the classification of the medical images before and after the training. In other words, there are counts in the same position but in different layers which make not plausible the use of a common quasi-symmetry model. In order to check where such counts are actually located, and therefore to detect which pairs of ratings are to be considered the cause of the departure from a common quasi-symmetry, one could consider the squared (standardized) residuals, assuming that the cells with the greatest squared residuals correspond to the cells with strong departure from the null hypothesis. Applying this analysis, one sees that several cells with high squared residual (namely, 5 out of 6 cells with squared residual higher than 2) pertain to the stage 3, meaning that the training session has changed (hopefully, improved) the agreement on the classification of this stage of the disease.

Further analyses can be performed within the Algebraic Statistics framework using the tests for outliers and patterns of outliers, and therefore the theory of Markov bases for incomplete tables, as introduced in [16], where several numerical examples are discussed.

4.2. Social mobility data

As a second numerical example we consider the data reported in Tab. 2 (adapted from [5] and originally collected during the “Italian Household Longitudinal Survey”) where the inter-generational social mobility has been recorded on a sample of 4,343 Italian workers in 1997. The data take into account the gender, and thus we have separate tables for men and women. There are 4 categories of workers. A : “High level professionals”; B : “Employees and commerce”; C : “Skilled working class and artisans”; D : “Unskilled working class”. In [5] these data are analyzed extensively with a thorough presentation of a lot of models to describe special patterns of mobility. Here we merely use the simplified version displayed in Tab. 2 to show the practical applicability of the methodology introduced in Sect. 3 also in this context.

We use again the quasi-symmetry model, and the Markov basis \mathcal{M}_0 for the model M_0 is formed by $2 \cdot 200$ moves. If we consider the two layers separately, we obtain exact p -values are equal to 0.051 and 0.088 respectively ($G^2 = 6.703$ and $G^2 = 8.279$ respectively, with 3 df). Although this fit may appear weak, nevertheless the situation dramatically

Table 2: Table of social mobility in Italy (1997). Columns represent the father's occupation, rows represent the son's (or daughter's) occupation. Male respondents in the left panel, female respondents in the right panel.

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>		<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
<i>A</i>	172	31	31	28	<i>A</i>	137	52	29	15
<i>B</i>	108	49	24	46	<i>B</i>	78	46	14	23
<i>C</i>	174	84	301	272	<i>C</i>	142	100	124	145
<i>D</i>	225	148	236	664	<i>D</i>	164	181	141	35

changes when considering the test for nested models. Running the test for a unique quasi-symmetry model, the exact test produces a p -value equal to 0 ($G^2 = 112.687$ with 12 df, corresponding to an asymptotic p -value less than 10^{-15}), meaning that there is a strong departure from the null hypothesis. Combining these results, one can conclude that the two genders present strong differences in terms of patterns of mobility. To see what cells contribute to that difference, one can use the techniques illustrated in the previous example.

References

- [1] Alan Agresti. *Categorical Data Analysis*. Wiley, New York, 3 edition, 2013.
- [2] Satoshi Aoki, Hisayuki Hara, and Akimichi Takemura. *Markov Bases in Algebraic Statistics*. Springer, New York, 2012.
- [3] Satoshi Aoki and Akimichi Takemura. Markov chain Monte Carlo exact tests for incomplete two-way contingency tables. *J. Stat. Comput. Simul.*, 75(10):787–812, 2005.
- [4] Yvonne M. Bishop, Stephen E. Fienberg, and Paul W. Holland. *Discrete multivariate analysis: Theory and practice*. MIT Press, Cambridge, 1975.
- [5] Richard Breen. *Social Mobility in Europe*. Oxford University Press, Oxford, 2007.
- [6] Henri Caussinus. Contribution à l'analyse statistique des tableaux de corrélation. *Ann. Fac. Sci. Toulouse Math. (4)*, 29:77–183, 1965.
- [7] Mathias Drton, Bernd Sturmfels, and Seth Sullivant. *Lectures on Algebraic Statistics*. Birkhauser, Basel, 2009.
- [8] Stephen E. Fienberg and Peter G. M. Van der Heijden. Introduction to special issue on quasi-symmetry and categorical data analysis. *Ann. Fac. Sci. Toulouse Math. (6)*, 11:439–441, 2002.
- [9] Leo A. Goodman. Contributions to the statistical analysis of contingency tables: Notes on quasi-symmetry, quasi-independence, log-linear models, log-bilinear models, and correspondence analysis models. *Ann. Fac. Sci. Toulouse*, 11(4):525–540, 2002.

- [10] Ralf Hemmecke, Raymond Hemmecke, and Peter Malkin. 4ti2 version 1.2—computation of Hilbert bases, Graver bases, toric Gröbner bases, and more. Available at www.4ti2.de, 2005.
- [11] Maria Kateri. *Contingency Table Analysis. Methods and Implementation Using R*. Springer, New York, 2014.
- [12] Giovanni Pistone, Eva Riccomagno, and Henry P. Wynn. *Algebraic Statistics: Computational Commutative Algebra in Statistics*. Chapman&Hall/CRC, Boca Raton, 2001.
- [13] Fabio Rapallo. Algebraic Markov bases and MCMC for two-way contingency tables. *Scand. J. Statist.*, 30(2):385–397, 2003.
- [14] Fabio Rapallo. Algebraic exact inference for rater agreement models. *Stat. Methods Appl.*, 14(1):45–66, 2005.
- [15] Fabio Rapallo. Toric statistical models: Parametric and binomial representations. *Ann. Inst. Statist. Math.*, 59(4):727–740, 2007.
- [16] Fabio Rapallo. Outliers and patterns of outliers in contingency tables with algebraic statistics. *Scand. J. Statist.*, 39(4):784–797, 2012.
- [17] Fabio Rapallo and Ruriko Yoshida. Markov bases and subbases for bounded contingency tables. *Ann. Inst. Stat. Math.*, 62(4):785–805, 2010.
- [18] Francisco Santos and Bernd Sturmfels. Higher Lawrence configurations. *J. Combin. Theory, Ser. A*, 103(1):151–164, 2003.
- [19] Bernd Sturmfels. *Gröbner Bases and Convex Polytopes*. Memoirs of the American Mathematical Society. American Mathematical Society, Providence, RI, 1996.
- [20] Seth Sullivant. Toric fiber products. *J. Algebra*, 316(2):560–577, 2007.
- [21] Akimichi Takemura and Satoshi Aoki. Distance-reducing markov bases for sampling from a discrete sample space. *Bernoulli*, 11(5):793–813, 2005.
- [22] Yu Xie. The log-multiplicative layer effect model for comparing mobility tables. *American Sociological Review*, 57:380–395, 1992.