Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Journal of Symbolic Computation 41 (2006) 164–172
Journal of Symbolic Computation
www.elsevier.com/locate/jsc

Markov bases and structural zeros

Fabio Rapallo

Department of Mathematics, University of Genova, Via Dodecaneso 35, 16146 Genova, Italy

Received 22 April 2004; accepted 2 April 2005
Available online 1 September 2005

Abstract

In this paper we apply the elimination technique to the computation of Markov bases, paying special attention to contingency tables with structural zeros. An algebraic relationship between the Markov basis for a table with structural zeros and the corresponding complete table is proved. In order to find the relevant Markov basis, it is enough to eliminate the indeterminates associated with the structural zeros from the toric ideal for the complete table. Moreover, we use this result for the computation of Markov bases for some classical log-linear models, such as quasi-independence and quasi-symmetry, and computations in the multi-way setting are presented.

Keywords: Contingency tables; Diaconis–Sturmfels algorithm; Elimination ideal; Log-linear models

1. Introduction

The analysis of contingency tables with structural zeros has received increasing attention in the last decades. Structural zeros are cells with true probabilities equal to zero. In such a case, the cell counts are zero regardless of the sample size. Here we consider contingency tables under the theory of log-linear models. For a general reference, see Fienberg (1980) or Agresti (2002). In particular, we are interested in exact procedures for goodness of fit tests.

In recent years, a new exact method for hypothesis testing for log-linear models has been introduced by Diaconis and Sturmfels (1998). That method is based on a Markov

E-mail address: rapallo@dima.unige.it.

doi:10.1016/j.jsc.2005.04.002

chain Monte Carlo approach, in order to navigate into the set of all contingencies with fixed value of the sufficient statistic. The Markov chain is defined through the notion of the Markov basis and the computation of Markov bases needs the computation of a system of generators of a toric ideal. As toric ideals can be computationally very intensive, especially for large tables, a number of methods have been proposed for simplifying the computations in special cases. For example see Dobra (2003) for graphical models, Aoki and Takemura (2003) for $3 \times 3 \times K$ tables with fixed two-way marginals, and Rapallo (2003) for some models for two-way tables.

In this paper we present an easy method for computing the relevant Markov basis for contingency tables with structural zeros, regardless of the dimension of the table and the form of the sufficient statistic. The method is also extended to complete tables where single cell counts are components of the sufficient statistic.

In Section 2 we recall some basic notions about log-linear models and exact goodness of fit tests, with special attention paid to the Diaconis–Sturmfels algorithm and the notion of the Markov basis, while in Section 3 we present an elimination-based method for the computation of the Markov bases. Section 4 is devoted to the presentation of a new theorem, which leads to a simple computation of the Markov basis for contingency tables with structural zeros or with single cell counts as components of the sufficient statistic. Finally, in Section 5 we show some applications to log-linear models for contingency tables with structural zeros frequently used in statistics, and examples.

2. Log-linear models and inference

We denote the sample space by $\mathcal{X}$, and its cardinality by $k = \#\mathcal{X}$, that is the number of non-zero cells of the table. We also denote the vector of the cell probabilities by p. Moreover, we denote the vector of the expected counts by μ and the vector of the observed counts by n. A generic contingency table is a function $f : \mathcal{X} \longrightarrow \mathbb{N}$.

The sufficient statistic is a function $T : \mathcal{X} \longrightarrow \mathbb{N}^s$. It can be represented by an $s \times k$ matrix A_T; see Pistone et al. (2001) for details. Let m be the rank of A_T. In the log-linear model theory, it is assumed that the vector $\log(\mu)$, the log of the expected counts, lies in the sub-vector space M of $\mathbb{R}^k$ generated by the columns of A_T. The vector space analysis of log-linear models is fully developed, e.g., in Haberman (1974). Using the linearity property, the sufficient statistic is defined for contingency tables by the formula

$$T(f) = \sum_{x \in \mathcal{X}} T(x) f(x).$$

This definition implies that $T(f) = A_T f$.

Example 1. Let us consider a 2×3 contingency table with a structural zero in the cell $(1, 2)$. The table is depicted below; 0 means a structural zero and the symbol $\bullet$ means a non-zero cell.

$$\begin{matrix} \bullet & 0 & \bullet \\ \bullet & \bullet & \bullet \end{matrix}$$

Here $\mathcal{X} = \{(1, 1), (1, 3), (2, 1), (2, 2), (2, 3)\}$ and $k = 5$. The model defined through the matrix

$$A_T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

is an independence model for this table and a sufficient statistic is $T_1 : \mathcal{X} \longrightarrow \mathbb{N}^4$ with

$$T_1(f) = (f_{11} + f_{13} + f_{21} + f_{22} + f_{23}, f_{21} + f_{22} + f_{23}, f_{22}, f_{13} + f_{23}).$$

Note that $\operatorname{rank}(A_T) = 3$. Moreover, the image of A_T is also spanned by the sufficient statistic

$$T_2(f) = (f_{11} + f_{13}, f_{21} + f_{22} + f_{23}, f_{11} + f_{21}, f_{22}, f_{13} + f_{23})$$

which is a most familiar form in the framework of contingency tables, as it is formed by the row sums and the column sums.

The goodness of fit tests for log-linear models are performed conditionally on the sufficient statistic, i.e. for a fixed value of the sufficient statistic. A number of methods for exact tests are given in the literature; see for example Agresti (2001). The most recent results on non-asymptotic goodness of fit tests are in Diaconis and Sturmfels (1998), where the theory of toric ideals is used in order to define appropriate Markov Chain Monte Carlo (MCMC) approximations of the test statistics. Previous works on the special problem of the quasi-independence model can also be found in Smith et al. (1996).

As the test statistics are computed conditionally on the sufficient statistic, we restrict the inference to the set of tables with the same value of the sufficient statistic with respect to the observed table, i.e., to the set

$$\mathcal{F}_t = \{f : \mathcal{X} \longrightarrow \mathbb{N} \mid T(f) = t\} \tag{1}$$

where t is the observed value of the sufficient statistic. This set is often called the reference set and the relevant probability distribution on $\mathcal{F}_t$ is the hypergeometric one; see Bishop et al. (1975) for details.

The Diaconis–Sturmfels algorithm defines a Markov chain on the reference set $\mathcal{F}_t$ and it is a special case of the Metropolis–Hastings algorithm which is based on a set of moves for constructing the relevant Markov chain. Once we have the right set of moves, the Markov chain is easy to implement. It is well known that a connected, reversible and aperiodic Markov chain converges to a stationary distribution. See Diaconis and Sturmfels (1998) for the convergence theorems and the detailed description of the algorithm. The following definition, from Diaconis and Sturmfels (1998), is the basis of the method.

Definition 2. A *Markov basis* of $\mathcal{F}_t$ is a set of functions $m_1, \ldots, m_L : \mathcal{X} \longrightarrow \mathbb{Z}$, called *moves*, such that for all i, $1 \le i \le L$,

$$\sum_{x \in \mathcal{X}} m_i(x) T(x) = 0, \tag{2}$$

where T is the sufficient statistic, and for any $f, f_1 \in \mathcal{F}_t$ there exist a sequence of moves $(m_{i_1}, \ldots, m_{i_A})$ and a sequence $(\epsilon_j)_{j=1}^A$ with $\epsilon_j = \pm 1$ such that

$$f_1 = f + \sum_{j=1}^{A} \epsilon_j m_{i_j} \qquad \text{and} \qquad f + \sum_{j=1}^{a} \epsilon_j m_{i_j} \geq 0 \tag{3}$$

for all $1 \leq a \leq A$.

Eqs. (2) and (3) imply that $T(m) = 0$ for all moves in the Markov basis and thus

$$T(f_1) = T\left(f + \sum_{j=1}^{A} \epsilon_j m_{i_j}\right) = T(f) + \left(\sum_{j=1}^{A} \epsilon_j T(m_{i_j})\right) = T(f),$$

i.e., the value of the sufficient statistic is constant for every table obtained with the moves in $m_1, \ldots, m_L$. A set of moves is a Markov basis if and only if the Markov chain on $\mathcal{F}_t$ is connected.

3. Computation of Markov bases

As we have mentioned in the previous section, the main problem in applying the Diaconis–Sturmfels algorithm is the computation of the relevant Markov basis. In this section we show how to address this problem using toric ideals; see Diaconis and Sturmfels (1998).

In the following, we sketch the elimination-based method for the computation of Markov bases and toric ideals. This method will be used in the proofs. More efficient algorithms can be found in Bigatti et al. (1999). As a reference in computational commutative algebra, the reader can refer to Kreuzer and Robbiano (2000).

We consider the field $\mathbb{Q}$ of the rational numbers, the indeterminates ξ_α, $\alpha \in \mathcal{X}$, and we associate with the problem the polynomial ring $\mathbb{Q}[\xi]$, i.e., we identify every cell probability with an indeterminate. Here ξ is vector notation, meaning $\xi = (\xi_\alpha)_{\alpha \in \mathcal{X}}$. The algebraic counterpart of a move m is obtained by decomposing it into a positive and a negative part $m = m^+ - m^-$ and defining the binomial

$$g_m = \xi^{m^+} - \xi^{m^-}. \tag{4}$$

For example the following move for the 2×3 tables

$$\begin{matrix} +1 & -2 & +1 \\ -1 & +2 & -1 \end{matrix}$$

is represented by the binomial

$$g_m = \xi_{11}\xi_{13}\xi_{22}^2 - \xi_{12}^2\xi_{21}\xi_{23}$$

in the polynomial ring $\mathbb{Q}[\xi_{11}, \xi_{12}, \xi_{13}, \xi_{21}, \xi_{22}, \xi_{23}]$.

Consider the matrix representation A_T of the sufficient statistic $T : \mathcal{X} \longrightarrow \mathbb{N}^s$. This matrix defines a set of power products $\{y_1, \ldots, y_k\}$ in a new polynomial ring $\mathbb{Q}[z]$, where $k = \#\mathcal{X}$ and the power products are defined by $y_i = z^{a(i)}$, where $a(i)$ is the i-th column of the matrix A_T. The power products y_i, $i = 1, \ldots, k$, induce a ring homomorphism

$$\pi_T : \mathbb{Q}[\xi] \longrightarrow \mathbb{Q}[z]$$

defined by

$$\xi_i \longmapsto y_i \tag{5}$$

for $i = 1, \ldots, k$. Now, define the binomials $\xi_i - y_i$, for $i = 1, \ldots, k$, in the polynomial ring $\mathbb{Q}[\xi, z]$ and the ideal $\mathcal{J}_T$ generated by such binomials. The toric ideal $\mathcal{I}_T$ associated with the sufficient statistic T is given by

$$\mathcal{I}_T = \mathcal{J}_T \cap \mathbb{Q}[\xi] := \mathrm{Elim}(z, \mathcal{J}_T).$$

An extended example on how this procedure works in the framework of contingency tables is presented in Rapallo (2003).

The free computer algebra system CoCoA—see CoCoA Team (2004)—has a function called `Toric` which directly computes the Gröbner basis of the toric ideal starting from the matrix representation of the sufficient statistic, and the algorithms implemented in this function are highly optimized and lead to the results in few seconds, at least in small- or medium-sized cases.

4. Markov bases and structural zeros

Suppose that we observe d random variables $Y_1, \ldots, Y_d$ on N subjects. Each random variable Y_q has a sample space $\{1, \ldots, I_q\}$, with $q = 1, \ldots, d$. If we consider the case of structural zeros, the sample space is $\mathcal{X}$, a subset of the cartesian product $\mathcal{X}' = \{1, \ldots, I_1\} \times \cdots \times \{1, \ldots, I_d\}$. The set $\mathcal{X}_0 = \mathcal{X}' \backslash \mathcal{X}$ is the set of the structural zeros of the table. The sufficient statistic for the sample of size 1 is a function $T : \mathcal{X} \longrightarrow \mathbb{N}^s$.

Definition 3. The function $T' : \mathcal{X}' \longrightarrow \mathbb{N}^s$ is an extension of T to $\mathcal{X}'$ if $T' = T$ on $\mathcal{X}$. The function $f' : \mathcal{X}' \longrightarrow \mathbb{N}$ is a contingency table on $\mathcal{X}'$ compatible with the contingency table f on $\mathcal{X}$ if $f' = f$ on $\mathcal{X}$ and $f' = 0$ on $\mathcal{X}_0$.

The relevant toric ideal for the complete table with sufficient statistic T' is $\mathcal{I}_{T'} = \mathrm{Elim}(z, \mathcal{J}_{T'})$, while the toric ideal for the incomplete table with sufficient statistic T is $\mathcal{I}_T = \mathrm{Elim}(z, \mathcal{J}_T)$. The following theorem gives a way of easily computing the toric ideal $\mathcal{I}_T$ starting from the toric ideal $\mathcal{I}_{T'}$.

Theorem 4. *Following the notation above, let $\mathcal{I}_T$ be the toric ideal associated with the sufficient statistic T on the sample space $\mathcal{X}$ and let $\mathcal{I}_{T'}$ be the toric ideal associated with the sufficient statistic T' on the sample space $\mathcal{X}'$. Finally, let $(\xi_\alpha)_{\alpha \in \mathcal{X}_0}$ be the indeterminates corresponding to the structural zeros. Then,*

$$\mathcal{I}_T = \mathrm{Elim}((\xi_\alpha)_{\alpha \in \mathcal{X}_0}, \mathcal{I}_{T'}). \tag{6}$$

Proof. Without loss of generality suppose that $\xi_1, \ldots, \xi_k$ are the indeterminates corresponding to the non-zero cells and the indeterminates $\xi_{k+1}, \ldots, \xi_{k+r}$ are the indeterminates corresponding to the structural zeros. Note that the ideal $\mathcal{I}_M$ is computed starting from the diophantine system $T(f) = t$, where t is the observed value of the sufficient statistic T. As for all structural zeros the observed counts are zero, the above system has the same solutions as

$$\begin{cases} T'(f') = t \\ f'_1 = \cdots = f'_r = 0 \end{cases} \tag{7}$$

where T' is an extension of T as in Definition 3 and f' is a contingency table on $\mathcal{X}'$ compatible with f. The system of polynomial equations is

$$\begin{cases} \xi_i = z^{a'(i)} & \text{for} \quad i = 1, \ldots, k \\ \xi_{k+j} = z_{k+j} & \text{for} \quad j = 1, \ldots, r \end{cases} \tag{8}$$

where $a'(i)$ is the ith column of the matrix $A_{T'}$, the matrix representation of T'. In particular, notice that $\xi_{k+j} = y_{k+j}$ for all $j = 1, \ldots, r$. Thus, following the theory in Chapter 2 of Kreuzer and Robbiano (2000), since $\xi_{k+j} = z_{k+j}$, the following chain of equalities holds:

$$\mathcal{I}_T = \mathrm{Elim}(z, \mathcal{J}_T) = \mathrm{Elim}((z, (\xi_{k+j})_{j=1}^r), \mathcal{J}_{T'}) = \mathrm{Elim}((\xi_{k+j})_{j=1}^r, \mathcal{I}_{T'}) .$$

The proof is now complete.

In Section 5 we will present some applications of Theorem 4 to statistical problems which occur frequently in statistics. First, we present a simple example.

Example 5. Let us consider the 2×3 table with a structural zero in the cell $(1, 2)$ as in Example 1 and the sufficient statistic

$$T_2(f) = (f_{11} + f_{13}, f_{21} + f_{22} + f_{23}, f_{11} + f_{21}, f_{22}, f_{13} + f_{23}).$$

An extension T' of T_2 is given by

$$T'(f) = (f_{11} + f_{12} + f_{13}, f_{21} + f_{22} + f_{23}, f_{11} + f_{21}, f_{12} + f_{22}, f_{13} + f_{23}).$$

The Gröbner basis for the complete table with sufficient statistic T' is

$$\mathcal{G}_{T'} = \{\xi_{11}\xi_{22} - \xi_{12}\xi_{21}, \xi_{11}\xi_{23} - \xi_{13}\xi_{21}, \xi_{12}\xi_{23} - \xi_{13}\xi_{22}\}.$$

By elimination of the indeterminate ξ_{12} corresponding to the cell $(1, 2)$, we obtain the Gröbner basis for the incomplete table with sufficient statistic T_2:

$$\mathcal{G}_{T_2} = \{\xi_{11}\xi_{23} - \xi_{13}\xi_{21}\}.$$

Thus, the Markov basis for the incomplete table is formed by making one move, namely

$$\begin{matrix} +1 & 0 & -1 \\ -1 & 0 & +1. \end{matrix}$$

The method described above for structural zeros also applies to complete tables, when single cell counts are components of the sufficient statistic.

Consistently with the above notation, let $\mathcal{X}'$ be a complete table and let $\mathcal{X}_0$ be the set of cells whose counts are components of the sufficient statistic. Define $\mathcal{X} = \mathcal{X}' \backslash \mathcal{X}_0$.

Let $\tilde{T} : \mathcal{X}' \longrightarrow \mathbb{N}^{s+r}$ be the sufficient statistic. Without loss of generality, suppose that the last r cell counts are components of the sufficient statistic. The diophantine system is

$$\tilde{T}(f') = \tilde{t} \tag{9}$$

where $\tilde{t} \in \mathbb{N}^{s+r}$ is the observed value of the sufficient statistic $\tilde{T}$. As $t_{s+1} = f_{k+1}, \ldots, t_{s+r} = f_{k+r}$, we have that $\tilde{t} = (\tilde{t}_1, f_{k+1}, \ldots, f_{k+r})^t$, with $\tilde{t}_1 \in \mathbb{N}^s$.

The system in Eq. (9) induces the polynomial system

$$\begin{cases} \xi_i = z^{\tilde{a}(i)} & \text{for } i = 1, \ldots, k \\ \xi_{k+j} = z_{k+j} & \text{for } j = 1, \ldots, r. \end{cases} \tag{10}$$

Let T be the restriction of $\tilde{T}$ to $\mathcal{X}$ and let T' be the extension of T to $\mathcal{X}'$ as in Definition 3. As $\tilde{a}_j(i) = a_j(i)$ for all $i = 1, \ldots, k$ and $j = 1 \ldots, s$, the polynomial system $\xi_i = z^{a(i)}$, $i = 1, \ldots, k$, is associated with the sufficient statistic T and the last r cells are structural zeros. Consequently, $\mathcal{I}_T = \mathcal{I}_{\tilde{T}}$. Thus, we have proved the following result.

Proposition 6. *Following the notation above, let $\mathcal{I}_{\tilde{T}}$ be the toric ideal associated with the sufficient statistic $\tilde{T}$ on the sample space $\mathcal{X}'$ and let $\mathcal{I}_{T'}$ be the toric ideal associated with the sufficient statistic T' on the sample space $\mathcal{X}$. Finally, let $(\xi_\alpha)_{\alpha \in \mathcal{X}_0}$ be the indeterminates corresponding to the cells whose counts are components of the sufficient statistic. Then,*

$$\mathcal{I}_{\tilde{T}} = \mathrm{Elim}((\xi_\alpha)_{\alpha \in \mathcal{X}_0}, \mathcal{I}_{T'}). \tag{11}$$

Note that the computation of the Markov bases does not imply any structure for the set $\mathcal{X}$. Thus, the connectedness of the Markov chain does not depend on the completeness of the table.

5. Applications and examples

There are a number of statistical models for contingency tables where structural zeros play a role. In this section we present two models for square incomplete tables, namely the quasi-independence model and the quasi-symmetry model. See Agresti (2002), Chapter 7, for references. Some specific results for the computation of Markov bases for the quasi-independence model are presented in Rapallo (2003). We show here that the computation of Markov bases for the quasi-independence model can be carried out within the above theory.

Let $\mathcal{X} = \{1, \ldots, I\} \times \{1, \ldots, I\}$ be a square two-way table. Following Agresti (2002), the sufficient statistic for the quasi-independence model on $\mathcal{X}$ is

$$\tilde{T} = \left(\sum_{j=1}^{I} f_{ij}, i = 1, \ldots I, \sum_{i=1}^{I} f_{ij}, j = 1, \ldots, I, \; f_{ii}, i = 1, \ldots, I \right).$$

Here, the counts of the main diagonal cells are components of the sufficient statistic. Hence, we apply Proposition 6. We consider the sufficient statistic

$$T' = \left(\sum_{j=1}^{I} f_{ij}, i = 1, \ldots I, \sum_{i=1}^{I} f_{ij}, j = 1, \ldots, I \right), \tag{12}$$

we compute $\mathcal{I}_{T'}$ and we obtain the relevant ideal $\mathcal{I}_{\tilde{T}}$ by elimination of the indeterminates $\xi_{11}, \ldots, \xi_{II}$.

The important issue in this case is that the toric ideal associated with T' in Eq. (12) is the toric ideal of the independence model for complete tables and does not need symbolic computations, as its Gröbner basis under the DegRevLex term order is

$$\mathcal{G} = \{\xi_{ij}\xi_{kh} - \xi_{ih}\xi_{kj}, 1 \leq i < k \leq I, 1 \leq j < h \leq I\};$$

see Diaconis and Sturmfels (1998) for details and proofs. Note that the DegRevLex term order is not the term order needed in the elimination step. It just represents an easy way to define the ideal $\mathcal{I}_{T'}$ without any symbolic computation.

The same procedure can also be applied to the quasi-symmetry model with sufficient statistic

$$T = \left(\sum_{j=1}^{I} f_{ij}, i = 1, \ldots I, \sum_{i=1}^{I} f_{ij}, j = 1, \ldots, I, f_{ij} + f_{ji}, 1 \leq i \leq j \leq I \right).$$

A number of examples of incomplete tables where our methods work are presented in Bishop et al. (1975), Chapter 5.

Finally, we show an explicit computation in the multi-way case.

Example 7. Let us consider a $3 \times 3 \times 3$ table with a structural zero in the cell $(1, 1, 1)$ under the complete independence model. The sample space is $\mathcal{X} = \{1, 2, 3\}^3 \setminus \{(1, 1, 1)\}$ and the sufficient statistic is

$$T = \left(\sum_{j,k} f_{ijk}, i = 1, 2, 3, \sum_{i,k} f_{ijk}, j = 1, 2, 3, \sum_{i,j} f_{i,j,k}, k = 1, 2, 3 \right) \tag{13}$$

where summations range over the cells in $\mathcal{X}$.

The extension T' of T to $\mathcal{X}' = \{1, 2, 3\}^3$ has the same expression as in Eq. (13), with the summations ranging over $\mathcal{X}'$. The Gröbner basis of the toric ideal $\mathcal{I}_{T'}$ associated with T' is formed by 162 binomials of degree 2. By elimination of the indeterminate ξ_{111}, we obtain the Gröbner basis for the ideal $\mathcal{I}_T$, which is formed by 160 binomials: 142 binomials of degree 2 and 18 binomials of degree 3. Note that the computation of the toric ideal $\mathcal{I}_T$ by using Theorem 4 needs about 1 s of CPU time, while the direct computation needs about 70 s.

Acknowledgements

We acknowledge Dr. Massimo Caboara (University of Pisa and CoCoA Team) for the careful revision of an earlier version of this paper. The author was partially supported by MIUR COFIN 03.

References

Agresti, A., 2001. Exact inference for categorical data: Recent advances and continuing controversies. Statist. Med. 20, 2709–2722.

Agresti, A., 2002. Categorical Data Analysis, 2nd edition. Wiley, New York.

Aoki, S., Takemura, A., 2003. Minimal basis for connected Markov chain over $3 \times 3 \times k$ contingency tables with fixed two-dimensional marginals. Aust. N. Z. J. Stat. 45 (2), 229–249.

Bigatti, A., La Scala, R., Robbiano, L., 1999. Computing toric ideals. J. Symbolic Comput. 27, 351–365.

Bishop, Y.M., Fienberg, S., Holland, P.W., 1975. Discrete Multivariate Analysis: Theory and Practice. MIT Press, Cambridge.

CoCoA Team, 2004. CoCoA: A System for Doing Computations in Commutative Algebra. Available at http://cocoa.dima.unige.it.

Diaconis, P., Sturmfels, B., 1998. Algebraic algorithms for sampling from conditional distributions. Ann. Statist. 26 (1), 363–397.

Dobra, A., 2003. Markov bases for decomposable graphical models. Bernoulli 9 (6), 1–16.

Fienberg, S., 1980. The Analysis of Cross-Classified Categorical Data. MIT Press, Cambridge.

Haberman, S.J., 1974. The Analysis of Frequency Data. The University of Chicago Press, Chicago and London.

Kreuzer, M., Robbiano, L., 2000. Computational Commutative Algebra 1. Springer, New York.

Pistone, G., Riccomagno, E., Wynn, H.P., 2001. Algebraic Statistics: Computational Commutative Algebra in Statistics. Chapman&Hall/CRC, Boca Raton.

Rapallo, F., 2003. Algebraic Markov bases and MCMC for two-way contingency tables. Scand. J. Statist. 30 (2), 385–397.

Smith, P.W.F., Forster, J.J., McDonald, J.W., 1996. Monte Carlo exact tests for square contingency tables. J. R. Stat. Soc. 159 (2), 309–321.