Markov Chains, Quotient Ideals, and Connectivity with Positive Margins

Edited by
Yuguo Chen, University of Illinois at Urbana-Champaign

Ian H. Dinwoodie, Duke University

Ruriko Yoshida, University of Kentucky

Abstract

We present algebraic methods for studying connectivity of Markov moves with margin positivity. The purpose is to develop Markov sampling methods for exact conditional inference in statistical models where a Markov basis is hard to compute. In some cases positive margins are shown to allow a set of Markov connecting moves that are much simpler than the full Markov basis.

0.1 Introduction

Advances in algebra have impacted in a fundamental way the study of exponential families of probability distributions. In the 1990s, computational methods of commutative algebra were brought into statistics to solve both classical and new problems in the framework of exponential family models. In some cases, the computations are of an algebraic nature or could be made algebraic with some work, as in the cumulant methods of Pistone and Wynn (1999). In other cases, the computations are ultimately Monte Carlo averages, and the algebra plays a secondary role in designing algorithms. This is the nature of the work of Diaconis and Sturmfels (1998). Commutative algebra is also used in statistics for experimental design (Pistone, Riccomagno, and Wynn, 2001) where exponential families are not the focus.

Diaconis and Sturmfels (1998) showed how computing a generating set for a toric ideal is fundamental to irreducibility of a Markov chain on a set of constrained tables. This theory gives a method for obtaining Markov chain moves, such as the genotype sampling method of Guo and Thompson (1992), extensions to graphical models (Geiger, Meek, and Sturmfels, 2006) and beyond (Hosten and Sullivant, 2004).

It has been argued that irreducibility is not essential (Besag and Clifford, 1989), but that view is not conventional. Sparse tables in high dimensions can be very difficult to study.

Algorithms and software have been developed for toric calculations that are much faster than early methods. The monograph of Sturmfels (1996) and Kreuzer and Robbiano (2000) are good introductions to toric ideals and some algorithms for computation. In addition, the software 4ti2 (4ti2 Team, 2006) is essential to research on statistics and algebra. It is easy to use and very fast (Hemmecke and Malkin, 2005).

Despite these significant computational advances, there are applied problems where one may never be able to compute a Markov basis. Recall that a Markov basis is a collection of vector increments that preserve the table constraints, and connect all tables with the same constraints (see Section

0.2). Models of no-3-way interaction and constraint matrices of Lawrence type seem to be arbitrarily difficult, in that the degree and support of elements of a minimal Markov basis can be arbitrarily large (De Loera and Onn, 2005). Thus, it is useful to compute a smaller number of moves which connect tables with given constraints rather than all constraints. The purpose of this paper is to develop algebraic tools for understanding sets of Markov moves that connect tables with positive margins, because sets of Markov moves that work with certain margins may be much simpler than a full Markov basis. Such connecting sets were formalized in Chen, Dinwoodie, and Sullivant (2006) with the terminology Markov subbasis.

Connectivity of a set of Markov moves is traditionally studied through primary decomposition (Diaconis, Eisenbud, and Sturmfels, 1998). As a practical tool, this is problematic because the primary decomposition is very difficult to compute, and also it can be hard to interpret in a useful way. In our experience, the computation is very slow or impossible with 20 or more cells in the table (giving 20 or more indeterminates). Theoretical results on primary decomposition of lattice ideals are relevant (for example Hosten and Shapiro, 2000) but are generally not sufficient to determine connecting properties of sets of Markov moves. Therefore we believe that developing algebraic tools based on quotient operations and radical ideals may be more practical in large problems.

A motivating example is the following, which is treated on a small scale in Example 5.2. In logistic regression at 10 levels of an integer covariate, one has a table of counts that gives the number of “yes” responses and the number of “no” responses at each covariate level $i = 1, 2, \dots, 10$. The sufficient statistics for logistic regression are 1) the total number of “yes” responses over all levels, 2) the quantity which is the sum over i of the “yes” count at level i multiplied by the covariate level i, and 3) the total counts of “yes” and ”no” responses at each level i. Conditional inference requires that one works with all tables that fix these 12 values, and which have nonnegative entries. A Markov chain with 2465 moves from “primitive partition identities” (p. 47 of Sturmfels (1996)) is irreducible in this collection of constrained tables, no matter what the 12 constraint values are. However, when each of the 10 sums over “yes” and ”no” counts at the 10 levels of i is positive, a Markov chain with only 36 moves is irreducible (Chen *et al.*, 2005). Therefore the property of positive margins can greatly simplify computations.

0.2 Arbitrary Margins and Toric Ideals

A *contingency table* records counts of events at combinations of factors, and it is used to study the relationship between the factors. All possible combinations of factor labels or "levels" make "cells" in an array, and the count in each cell may be viewed as the outcome of a multinomial probability distribution.

In this section a contingency table is written as a vector of length c, and this representation comes from numbering the cells in a multiway table. Let A be an $r \times c$ matrix of nonnegative integers with columns $\mathbf{a}_1, \ldots, \mathbf{a}_c$ in Z_+^r. The matrix A is the *design matrix* or *constraint matrix*, and the r rows are the vectors for computing sufficient statistics. The total number of constraints when sufficient statistics are fixed is r, which is also the number of parameters in a loglinear representation of the cell probabilities p_i:

$$p_i = \frac{e^{\theta' \mathbf{a}_i}}{z_\theta}$$

where z_θ is the normalizing constant, and θ is a column vector of parameters in $\mathbb{R}^r$. Then the points $(p_1, \ldots, p_c)$ are in the toric variety defined by the matrix A, while also being nonnegative and summing to 1.

For example, for 2×3 tables under the independence model, A is the 5×6 matrix given by

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

and the rows of A compute row and column sums of the contingency table.

Assume that a strictly positive vector is in the row space of A. The toric ideal I_A in the ring $Q[\mathbf{x}] = Q[x_1, x_2, \ldots, x_c]$ is defined by

$$I_A = \langle \mathbf{x}^{\mathbf{a}} - \mathbf{x}^{\mathbf{b}} : A\mathbf{a} = A\mathbf{b} \rangle$$

where $\mathbf{x}^{\mathbf{a}} = x_1^{a_1} x_2^{a_2} \cdots x_c^{a_c}$ is the usual monomial notation. Define the fiber $\Omega_{\mathbf{t}} := \{\mathbf{n} \in Z_+^c : A\mathbf{n} = \mathbf{t}\}$ (nonnegative integer lattice points) for $\mathbf{t} = (t_1, \ldots, t_r) \in Z_+^r$. That is, the fiber is the set of all contingency tables satisfying the given constraints. It is known that a generating set of binomials $\{\mathbf{x}^{\mathbf{a}_i^+} - \mathbf{x}^{\mathbf{a}_i^-}\}$ for I_A provide increments $\{\pm(\mathbf{a}_i^+ - \mathbf{a}_i^-)\}$ that make an irreducible Markov chain in $\Omega_{\mathbf{t}}$, whatever the value of $\mathbf{t}$ (Diaconis and Sturmfels, 1998). Here $\mathbf{a}_i^+ = \max\{\mathbf{a}_i, 0\}$ and $\mathbf{a}_i^- = \max\{-\mathbf{a}_i, 0\}$. Such a generating set is called a Markov basis. The Markov chain is run by randomly choosing one of the increments $\mathbf{a}_i^+ - \mathbf{a}_i^-$ and randomly choosing a sign, then adding the increment to the current state if the result is nonnegative. Irreducible means that for any two nonnegative integer vectors $\mathbf{m}, \mathbf{n}$ that satisfy $A\mathbf{m} = A\mathbf{n} = \mathbf{t}$, there is a sequence of signed vectors $\sigma_j(\mathbf{a}_{i_j}^+ - \mathbf{a}_{i_j}^-)$, $j = 1, 2, \ldots, J$ $(\sigma_j = \pm 1)$, that connects $\mathbf{m}$ and $\mathbf{n}$. That is, $\mathbf{n} = \mathbf{m} + \sum_{j=1}^{J} \sigma_j(\mathbf{a}_{i_j}^+ - \mathbf{a}_{i_j}^-)$ and furthermore every intermediate point in the path remains in the domain:

$$\mathbf{m} + \sum_{j=1}^{I} \sigma_j(\mathbf{a}_{i_j}^+ - \mathbf{a}_{i_j}^-) \in \Omega_{\mathbf{t}},\ 1 \leq I \leq J.$$

In particular, intermediate points on the path are nonnegative.

When one allows entries in the table to go negative, connecting Markov chains are easier to find. The following proposition uses some standard terminology. Let $M := \{\pm \mathbf{a}_i \in Z^c : i = 1, \ldots, g\} \subset \ker(A)$ be signed Markov moves (that is, integer vectors in $\ker(A)$ that are added or subtracted randomly from the current state), not necessarily a Markov basis. Let $I_M := \langle \mathbf{x}^{\mathbf{a}_i^+} - \mathbf{x}^{\mathbf{a}_i^-},\ i = 1, \ldots, g \rangle$ be the corresponding ideal, which satisfies $I_M \subset I_A$. The *radical* of an ideal I is $\sqrt{I} = \{f \in Q[\mathbf{x}] : f^i \in I \text{ for some } i \in \mathbb{Z}_+\}$. If $I = \sqrt{I}$, then we say that I is a *radical ideal* (p. 35 of Cox, Little, and O'Shea (1997)).

A set of integer vectors $M \subset Z^c$ is called a *lattice basis* for A if every integer vector in $\ker(A)$ can be written as an integral linear combination of the vectors (or moves) in M. Computing a lattice basis is very simple and does not require symbolic computation.

Proposition 0.2.1 *Suppose I_M is a radical ideal, and suppose the moves in M form a lattice basis. Then the Markov chain using the moves in M that allows entries to drop down to -1 connects a set that includes the set $\Omega_{\mathbf{t}}$.*

Proof Let $\mathbf{m}, \mathbf{n}$ be two elements in $\Omega_{\mathbf{t}}$. By allowing entries to drop down to -1 in the Markov chain, it is enough to show that $\mathbf{m} + \mathbf{1}$ and $\mathbf{n} + \mathbf{1}$ are connected with a nonnegative path using moves in M. By Theorem 8.14 of Sturmfels (2002), $\mathbf{m}+\mathbf{1}$ and $\mathbf{n}+\mathbf{1}$ are connected in this way if $\mathbf{x}^{\mathbf{m}+\mathbf{1}} - \mathbf{x}^{\mathbf{n}+\mathbf{1}}$ are in the ideal $I_M \subset Q[\mathbf{x}]$. Let $p = x_1 \cdot x_2 \cdot \ldots \cdot x_c$. Since the moves are a lattice basis, it follows that $I_M : p^n = I_A$ for some integer $n > 0$ (Lemma 12.2 of Sturmfels (1996)). Thus $p^n(\mathbf{x}^{\mathbf{m}} - \mathbf{x}^{\mathbf{n}}) \in I_M$ by the definition of the quotient ideal. Hence $p^n(\mathbf{x}^{\mathbf{m}} - \mathbf{x}^{\mathbf{n}})^n \in I_M$, and since I_M is radical it follows that $\mathbf{x}^{\mathbf{m}+\mathbf{1}} - \mathbf{x}^{\mathbf{n}+\mathbf{1}} = p\,(\mathbf{x}^{\mathbf{m}} - \mathbf{x}^{\mathbf{n}}) \in I_M$. $\square$

The idea of allowing some entries to drop down to -1 appears in Bunea and Besag (2000) and Chen *et al.* (2005). In high dimensional tables (c large), the enlarged state space that allows entries to drop down to -1 may be much larger than the set of interest $\Omega_{\mathbf{t}}$, even though each dimension is only slightly extended. Nevertheless, Proposition 0.2.1 makes it possible to use the following approach on large tables: compute a lattice basis, compute the radical of the ideal of binomials from the lattice basis, run the Markov chain in the larger state space, and do computations on $\Omega_{\mathbf{t}}$ by conditioning. To be precise, suppose $\Omega_{\mathbf{t}} \subset \Omega_0$ where the set Ω_0 is the connected component of the Markov chain that is allowed to drop down to -1 as above. Suppose the desired sampling distribution μ on $\Omega_{\mathbf{t}}$ is uniform. If one runs a symmetric Markov chain $X_1, X_2, X_3, \dots, X_n$ in Ω_0, then a Monte Carlo estimate of $\mu(A)$ for any subset $A \subset \Omega_{\mathbf{t}}$ is

$$\mu(A) = \frac{\sum_{i=1}^n I_A(X_i)}{\sum_{i=1}^n I_{\Omega_{\mathbf{t}}}(X_i)}$$

where I_A is the indicator function of the set A.

0.3 Survey of Computational Methods

A loglinear model for a multiway table of counts can be fit and evaluated many ways. Maximum likelihood fitting and asymptotic measures of goodness-of-fit are available from Poisson regression on a data frame, part of any generalized linear model package such as the one in R (R Development Core Team, 2004). The R command `loglin` also does table fitting, using iterative proportional fitting, and this is more convenient than Poisson regression when the data is in a multidimensional array. Both methods rely on χ^2 asymptotics on either the Pearson χ^2 statistic or likelihood ratio statistic for goodness-of-fit. For sparse tables, one often wants exact conditional methods to avoid asymptotic doubts. The basic command `chisq.test` in R has an option for the exact method on two-way tables, usually called Fisher's exact test.

For higher-way tables, the package `exactLoglinTest` is maintained by Brian Caffo (Caffo, 2006). This implements an importance sampling method of Booth and Butler (1999). There are certain examples where it has difficulty generating valid tables, but user expertise can help.

Markov chains can be run with a set of Markov moves that come from generators of a toric ideal. Computing these generators can be done in many algebra software packages, including COCOA (CoCoATeam, 2007), Macaulay 2 (Grayson and Stillman, 2006), and Singular (Greuel, Pfister, and Schoenemann, 2006) which implement several algorithms. Finally, 4ti2 (4ti2 Team, 2006) was used for computing Markov bases in this paper. It is very fast, it has a natural coding language for statistical problems, and it has utilities for filtering output.

A Monte Carlo method that is extremely flexible and does not require algebraic computations in advance is sequential importance sampling (Chen, Dinwoodie, and Sullivant, 2006). This method uses linear programming to generate tables that in practice satisfy constraints with very high probability. Efficient implementation requires a good proposal distribution.

0.4 Margin Positivity

The Markov basis described in Section 0.2 is a very powerful construction – it can be used to construct an irreducible Markov chain for any margin values $\mathbf{t}$. It is possible that a smaller set of moves may connect tables when $\mathbf{t}$ is strictly positive. The notion of Markov subbasis was introduced in Chen, Dinwoodie, and Sullivant (2006) to study connecting sets of moves in $\Omega_{\mathbf{t}}$ for certain values of $\mathbf{t}$.

Now a lattice basis for $\ker(A)$ has the property that any two tables can be connected by its vector increments if one is allowed to swing negative in the connecting path (see p. 47 of Schrijver (1986) and Chapter 12 of Sturmfels (1996) for definitions and properties of a lattice basis). One may expect that if the margin values $\mathbf{t}$ are sufficiently large positive numbers, then the paths can be drawn out of negative territory and one may get nonnegative connecting paths and so remain in $\Omega_{\mathbf{t}}$. However in general large positive margin values do not make every lattice basis a connecting set, as illustrated below.

Example 4.1 The following example is from p. 112 of Sturmfels (2002). With moves of adjacent minors (meaning the nine adjacent $\begin{matrix} + & - \\ - & + \end{matrix}$ sign pattern vector increments in the matrix), it is clear that one cannot connect the following tables, no matter how large the margins $3n$ may be:

$$\begin{array}{|cccc|} \hline n & n & 0 & n \\ 0 & 0 & 0 & n \\ n & 0 & 0 & n \\ n & 0 & n & n \\ \hline \end{array}\,, \quad \begin{array}{|cccc|} \hline n & n & 0 & n \\ n & 0 & 0 & n \\ 0 & 0 & 0 & n \\ n & n & 0 & n \\ \hline \end{array}\,.$$

Adjacent minors have been studied in depth, see for example Hosten and Sullivant (2002).

Proposition 0.4.1 *Let A be a 0-1 matrix. Suppose there is an integer lower bound $b > 0$ on all the constraint values:*

$$t_m \geq b,\ \ m = 1, 2, \ldots, r.$$

Let $I_m = \langle x_k \rangle_{A_{m,k}>0}$ be the monomial ideal generated by all the indeterminates for the cells that contribute to margin m. If

$$I_A \cap \bigcap_{m=1}^{r} I_m^b \subset I_M$$

where $I_m^b = \langle x_{i_1} x_{i_2} \cdots x_{i_b} \rangle_{A_{m,i_k}>0}$, then the moves in M connect all tables in $\Omega_{\mathbf{t}}$.

Proof Let $\mathbf{m}$ and $\mathbf{n}$ be two tables in $\Omega_{\mathbf{t}}$. It is sufficient to show that $\mathbf{x}^{\mathbf{m}} - \mathbf{x}^{\mathbf{n}} \in I_M$, by Theorem 8.14 of Sturmfels (2002). Now clearly $\mathbf{x}^{\mathbf{m}} - \mathbf{x}^{\mathbf{n}} \in I_A$. Since all the constraint values t_m are positive and A has 0-1 entries, it follows that each monomial $\mathbf{x}^{\mathbf{m}}$ and $\mathbf{x}^{\mathbf{n}}$ belongs to $I_m^b = \langle x_{i_1} x_{i_2} \cdots x_{i_b} \rangle_{A_{m,i_k}>0}$. Thus the binomial $\mathbf{x}^{\mathbf{m}} - \mathbf{x}^{\mathbf{n}} \in I_A \cap \bigcap_{m=1}^{r} I_m^b$.

Thus it is sufficient to show that

$$I_A \cap \bigcap_{m} I_m^b \subset I_M$$

which is the condition of the proposition. □

This result can establish connectivity in examples where the primary decomposition is hard to compute. It does not require I_M to be radical.

Let $p = x_1 x_2 \cdots x_c$ and let $I_M : p^\infty$ be the *saturation* of I_M by p, namely,

$$I_M : p^\infty := \{g \in Q[\mathbf{x}] : p^k \cdot g \in I_M \text{ for some } k \geq 0\}.$$

Then $I_A = I_M : p^\infty$ when the moves in M form a lattice basis (Lemma 12.2 of Sturmfels (1996)). One can show easily that

$$I_A \cap \bigcap_{m=1}^{r} I_m \subset (I_M \cap \bigcap_{m=1}^{r} I_m) : p^\infty$$

but the right hand side seems hard to compute directly, so this way of computing moves for tables with positive margins does not seem efficient. The ideal $\bigcap_m I_m$ is a monomial ideal for the Stanley-Reisner complex given by subsets of sets of cell indices *not* in the margins. For example, for 2×3 tables with fixed row and column sums as in Example 5.1, and cells labeled left to right, the ideals are $\langle x_1, x_2, x_3 \rangle \cap \langle x_4, x_5, x_6 \rangle \cap \langle x_1, x_4 \rangle \cap \langle x_2, x_5 \rangle \cap \langle x_3, x_6 \rangle$ and the simplicial complex is all subsets of the sets $\{\{4,5,6\},\{1,2,3\},\{2,3,5,6\},$ $\{1,3,4,6\},\{1,2,4,5\}\}$.

Example 4.2. Consider the collection of 3×3 tables with fixed row and column sums. If the margin values are all positive, then the collection of four moves of adjacent minors is not necessarily a connecting set – consider the two tables below:

$$\begin{array}{|c|c|c|}\hline 1&0&0\\\hline 0&0&1\\\hline 0&1&0\\\hline\end{array}\,, \quad \begin{array}{|c|c|c|}\hline 0&1&0\\\hline 0&0&1\\\hline 1&0&0\\\hline\end{array}\,.$$

However, if all the six margin values are at least $b=2$, then one can apply Proposition 0.4.1 to the moves in M of adjacent minors, which do not form a radical ideal. The toric ideal I_A can be computed and the containment required can be shown with $I_M : (I_A \cap \bigcap_{m=1}^{6} I_m^2) = \langle 1\rangle$.

Theorem 0.4.1 *Suppose I_M is a radical ideal, and suppose M is a lattice basis. Let $p = x_1 \cdot x_2 \cdot \ldots \cdot x_c$. For each row index m with $t_m > 0$, let $I_m = \langle x_k\rangle_{A_{m,k}>0}$ be the monomial ideal generated by indeterminates for cells that contribute to margin m. Let $\mathcal{M}$ be the collection of indices m with $t_m > 0$. Define*

$$I_{\mathcal{M}} = I_M : \prod_{m\in\mathcal{M}} I_m.$$

If

$$I_{\mathcal{M}} : (I_{\mathcal{M}} : p) = \langle 1\rangle$$

then the moves in M connect all tables in $\Omega_{\mathbf{t}}$.

Proof Let $\mathbf{m}$ and $\mathbf{n}$ be two tables in $\Omega_{\mathbf{t}}$ with margins $\mathcal{M}$ positive. It is sufficient to show that $\mathbf{x}^{\mathbf{m}} - \mathbf{x}^{\mathbf{n}} \in I_M$, by Theorem 8.14 of Sturmfels (2002). Now clearly $\mathbf{x}^{\mathbf{m}} - \mathbf{x}^{\mathbf{n}} \in I_A$, and since the margins $\mathcal{M}$ are positive it follows that $\mathbf{x}^{\mathbf{m}} - \mathbf{x}^{\mathbf{n}} \in \cap_{m\in\mathcal{M}} I_m$. Thus it is sufficient to show that

$$I_A \cap \bigcap_{m\in\mathcal{M}} I_m \subset I_M.$$

Since I_M is radical, this will follow if

$$I_A \cdot \prod_{m\in\mathcal{M}} I_m \subset I_M,$$

which holds if $I_M : (\prod_{m\in\mathcal{M}} I_m \cdot I_A) = (I_M : \prod_{m\in\mathcal{M}} I_m) : I_A = \langle 1\rangle$. This condition follows if $I_A \subset I_M : \prod_{m\in\mathcal{M}} I_m = I_{\mathcal{M}}$.

If $I_{\mathcal{M}} : (I_{\mathcal{M}} : p) = \langle 1 \rangle$, it follows that $I_{\mathcal{M}} = I_{\mathcal{M}} : p$. Then furthermore, $I_{\mathcal{M}} = I_{\mathcal{M}} : p^{\infty}$.

Since M is a lattice basis, it follows (Lemma 12.2 of Sturmfels (1996)) that $I_A = I_M : p^{\infty} \subset I_{\mathcal{M}} : p^{\infty} = I_{\mathcal{M}} : p$. This shows that $I_A \subset I_{\mathcal{M}} : p = I_{\mathcal{M}}$ and the result is proven. □

0.5 Additional Examples

In this section we apply the results on further examples, starting with the simplest for illustration and clarification of notation. We also do an example of logistic regression where the results are useful, and an example of no-3-way interaction where it is seen that the results are not useful.

Example 5.1. Consider the simplest example, the 2×3 table with fixed row and column sums, which are the constraints from fixing sufficient statistics in an independence model. If the second column sum is positive, then tables can be connected with adjacent minors. This is well-known based on primary decomposition. Indeed, the two moves corresponding to increments

$+1$	-1	0
-1	$+1$	0

,

0	$+1$	-1
0	-1	$+1$

make the radical ideal $I_M = \langle x_{11}x_{22} - x_{12}x_{21}, x_{12}x_{23} - x_{13}x_{22} \rangle$ in $Q[x_{11}, x_{12}, x_{13}, x_{21}, x_{22}, x_{23}]$. Then I_M has primary decomposition equal to

$$I_A \cap \langle x_{12}, x_{22} \rangle$$

which shows that the binomial $\mathbf{x}^{\mathbf{m}} - \mathbf{x}^{\mathbf{n}}$ for two tables $\mathbf{m}, \mathbf{n}$ with the same row and column sums can be connected by the two moves of adjacent minors if either x_{12} or x_{22} is present in $\mathbf{x}^{\mathbf{m}}$ and either is present in $\mathbf{x}^{\mathbf{n}}$ – in other words if the second column sum is positive.

Also, Theorem 0.4.1 applies. The set $\mathcal{M}$ has one index for the second column margin, and $I_{\mathcal{M}} = I_M : \langle x_{12}, x_{22} \rangle = I_A$. Hence $I_{\mathcal{M}} : (I_{\mathcal{M}} : x_{11}x_{12}x_{13}x_{21}x_{22}x_{23}) = I_A : (I_A : x_{11}x_{12}x_{13}x_{21}x_{22}x_{23}) = \langle 1 \rangle$.

Example 5.2. Consider the logistic regression problem with a 2×7 table and constraints of fixed row and column sums (9 constraints) in addition to fixed regression weighted sum $\sum_{i=1}^{7} i\, n_{1,i}$. The setup and connection with exponential families is described on p. 387 of Diaconis and Sturmfels (1998). Consider the fifteen moves like

0	$+1$	-1	0	-1	$+1$	0
0	-1	$+1$	0	$+1$	-1	0

.

The ideal I_M is radical, even though initial terms in a Groebner basis are not square-free. It is known that such moves connect tables with positive column sums (Chen *et al.*, 2005). This was not deduced from the primary decomposition, which we have not yet computed. Theorem 0.4.1 does apply, and computing the radical ideal in order to verify the conditions of the theorem is not difficult. We have seven monomial ideals for the column sums given by $I_i = \langle x_{1,i}, x_{2,i} \rangle$ and the quotient ideal $I_{\mathcal{M}} = I_M : (I_1 \cdot I_2 \cdots I_7)$ is the toric ideal I_A with 127 elements in the reduced Groebner basis.

A widely used class of models in applications is the no-3-way interaction class. For example, if one has four factors A, B, C, D for categorical data, each with several levels, the no-3-way interaction model is the loglinear model described with the common notation [A, B], [A, C], [A, D], [B, C], [B, D], [C, D] (see Christensen (1990) for notation and definitions). That is, the sufficient statistics are given by sums of counts that fix all pairs of factors at specified levels. The Markov basis calculations for these models are typically hard, even for the $4 \times 4 \times 4$ case. Whittaker (1990) presents an 8-way binary table of this type, for which we have not yet computed the Markov basis but which can be approached with sequential importance sampling.

Given the difficulty of these models, it would be interesting and useful if positive margins lead to simpler Markov bases. The answer seems to be no. Consider the natural class of moves $M = \{(e_{i,j,\mathbf{k}} + e_{i',j',\mathbf{k}} - e_{i',j,\mathbf{k}} - e_{i,j',\mathbf{k}}) - (e_{i,j,\mathbf{k}'} + e_{i',j',\mathbf{k}'} - e_{i',j,\mathbf{k}'} - e_{i,j',\mathbf{k}'}), \ldots\}$. Also, permute the location of $i, j, \mathbf{k}$. That is, choose two different coordinates from the d coordinates (above it is the first two), and choose two different levels i, i' and j, j' from each. Choose two different vectors $\mathbf{k}, \mathbf{k}'$ for all the remaining coordinates. This collection is in $\ker(A)$. The example below shows that these moves do not connect tables with positive margins.

Example 5.3. Consider 4-way binary data, and order the 2^4 cells 0000, 1000, 0100, 1100, …, 1111. There are 20 moves M of degree 8 as described above which preserve sufficient statistics for the no-3-way interaction model. More precisely, the representation of moves M above $(e_{i,j,\mathbf{k}} + e_{i',j',\mathbf{k}} - e_{i',j,\mathbf{k}} - e_{i,j',\mathbf{k}}) - (e_{i,j,\mathbf{k}'} + e_{i',j',\mathbf{k}'} - e_{i',j,\mathbf{k}'} - e_{i,j',\mathbf{k}'})$ gives square-free degree-8 moves, including for example $(e_{1100} + e_{0000} - e_{0100} - e_{1000}) - (e_{1101} + e_{0001} - e_{0101} - e_{1001})$. The representation is redundant,and only 20 of them are needed to connect the same set of tables. To see this, first compute a Groebner basis using 4ti2 for the model, which gives 61 moves and 20 square-free moves of lowest total degree 8, under a graded term order. Each of the degree-8 moves in M reduces to 0 under long division by the Groebner basis, and this division process can only use the degree-8 moves of the Groebner basis, since the dividend has degree 8. Now the degree-8 moves in the Groebner basis are the 20 degree-8 moves from M. Therefore these 20 moves connect everything that M connects.

Consider two tables given by

$$(0,0,1,0,1,0,0,2,0,1,0,0,0,0,1,0),(0,0,0,1,0,1,2,0,1,0,0,0,0,0,0,1).$$

These tables have the same positive margin vectors, but the 20 moves do not connect the two tables. This can be verified in Singular (2006) by division – long division of the binomial $x_3x_5x_8^2x_{10}x_{15} - x_4x_6x_7^2x_9x_{16}$ by a Groebner basis for the ideal of 20 moves does not leave remainder 0.

Example 5.4. Consider $4 \times 4 \times 2$ tables with constraints [A, C], [B, C], [A, B] for factors A, B, C, which would arise for example in case-control data with two factors A and B at four levels each.

The constraint matrix that fixes row and column sums in a 4×4 table gives a toric ideal with a $\binom{4}{2} \times \binom{4}{2}$ element Groebner basis. Each of these moves can be paired with its signed opposite to get 36 moves of $4 \times 4 \times 2$ tables that preserve sufficient statistics:

$$\begin{array}{|cccc|}\hline 0 & 0 & 0 & 0 \\ +1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & +1 & 0 \\ \hline \end{array}\,, \quad \begin{array}{|cccc|}\hline 0 & 0 & 0 & 0 \\ -1 & 0 & +1 & 0 \\ 0 & 0 & 0 & 0 \\ +1 & 0 & -1 & 0 \\ \hline \end{array}\,.$$

These elements make an ideal with a Groebner basis that is square-free in the initial terms, and hence the ideal is radical (Proposition 5.3 of Sturmfels (2002)). Then applying Theorem 0.4.1 with sixteen margins of case-control counts shows that these 36 moves do connect tables with positive case-control sums. The full Markov basis has 204 moves. This example should generalize to a useful proposition on extending Markov moves for simple models to an extra binary variable. The results of Bayer, Popescu, and Sturmfels (2001) on Lawrence liftings may be useful for a more general result.

Fallin *et al.* (2001) present case-control data with four binary factors, which are nucleotides at four loci related to Alzheimer's disease. The statistical question is whether the model of independence of nucleotides at these loci fits the data. One has five factors: L1, L2, L3, L4, for the four loci, and C for the binary case-control variable. The constraint matrix for exact conditional analysis is the Lawrence lifting of the independence model on L1, L2, L3, L4, which is described in loglinear notation as [L1, C], [L2, C], [L3, C], [L4, C], [L1, L2, L3, L4]. The next example is an algebraic treatment of the situation with three loci L1, L2, L3. A general result for any number of binary factors would be interesting. Further examples of case-control data where such results could be applied are in Chen, Dinwoodie, and MacGibbon (2007).

Example 5.5. Consider the 4-way binary model [L1, C], [L2, C], [L3, C], [L1, L2, L3]. There is a natural set of twelve degree 8 moves that comes from putting the degree 4 moves from the independence model [L1], [L2], [L3] at level C=1, and matching them with the opposite signs at level C=0. This construction is very general for case-control data. The resulting ideal I_M is radical. Suppose the case-control sums are positive, or in other words suppose that the 2^3 constraints described by [L1, L2, L3] are positive. Then one can show that these twelve moves connect all tables.

0.6 Conclusions

We have presented algebraic methods for studying connectivity of moves with margin positivity. The motivation is that two kinds of constraint matrices lead to very difficult Markov basis calculations, and they arise often in applied categorical data analysis. The first kind are the matrices of Lawrence type, which come up in case-control data. The second kind are the models of no-3-way interaction, which come up when three or more factors are present and one terminates the model interaction terms at 2-way interaction.

The examples that we have studied suggest that further research on connecting moves for tables with constraints of Lawrence type and with positive margins would have theoretical and applied interest. In this setting it does appear that there can be Markov connecting sets simpler than the full Markov basis. On the other hand, margin positivity does not seem to give much simplification of a Markov connecting set in problems of no-3-way interaction. Finally, radical ideals of Markov moves have valuable connectivity properties, and efficient methods for computing radicals and verifying radicalness would be useful. When the full toric ideal is too complicated, working with a radical ideal may be possible.

0.7 Acknowledgments

Yuguo Chen was partly supported under NSF grant DMS-0503981.

References

4ti2 Team (2006). *4ti2 – A software package for algebraic, geometric and combinatorial problems on linear spaces*. Available at www.4ti2.de.

Bayer, D., Popescu, S., and Sturmfels, B. (2001). Syzygies of unimodular Lawrence ideals. *Journal für die reine und angewandte Mathematik*, **534**, 169-186.

Besag, J., and Clifford, P. (1989). Generalized Monte Carlo significance tests. *Biometrika*, **76**, 633-642.

Booth, J. G., and Butler, J. W. (1999). An importance sampling algorithm for exact conditional tests in loglinear models. *Biometrika*, **86**, 321-332.

Bunea, F., and Besag, J. (2000). MCMC in $I \times J \times K$ contingency tables. *Monte Carlo Methods*. N. Madras ed. American Mathematical Society, Providence, Rhode Island. 25-36.

Caffo, B. (2006). *exactLoglinTest: A Program for Monte Carlo Conditional Analysis of Log-linear Models*. Available at http://www.cran.r-project.org/src/contrib/Descriptions/exactLoglinTest.html.

Chen, Y., Dinwoodie, I. H., Dobra, A., and Huber, M. (2005). Lattice points, contingency tables and sampling. *Integer Points in Polyhedra–Geometry, Number Theory, Algebra, Optimization*. A. Barvinok, M. Beck, C. Haase, B. Reznick and V. Welker eds. American Mathematical Society. 65-78.

Chen, Y., Dinwoodie, I. H., and MacGibbon, B. (2007). Sequential importance sampling for case-control data. *Biometrics*. **63**, 845-855.

Chen, Y., Dinwoodie, I. H., and Sullivant, S. (2006). Sequential importance sampling for multiway tables. *The Annals of Statistics*. **34**, 523-545

Christensen, R. (1990). *Log-Linear Models*. New York: Springer-Verlag.

CoCoATeam (2007). *CoCoA: a system for doing Computations in Commutative Algebra*. Available at http://cocoa.dima.unige.it.

Cox, D., Little, J., and O'Shea, D. (1997). *Ideals, Varieties, and Algorithms*, 2nd Edition. Springer, New York.

De Loera, J., and Onn, S. (2005). Markov bases of three-way tables are arbitrarily complicated. *Journal of Symbolic Computation*, **41**, 173–181.

Diaconis, P., Eisenbud, D., and Sturmfels, B. (1998). Lattice walks and primary decomposition. In *Mathematical Essays in Honor of Gian-Carlo Rota,* eds. B. E. Sagan, R. P. Stanley, Birkhauser, Boston, 173-193.

Diaconis, P., and Sturmfels, B. (1998). Algebraic methods for sampling from conditional distributions. *The Annals of Statistics*, **26**, 363-397.

Fallin, D., Cohen, A., Essioux, L., Chumakov, I., Blumenfeld, M., Cohen, D., and Schork, N. J. (2001). Genetic analysis of case/control data using estimated haplotype frequencies: application to APOE locus variation and Alzheimer's disease. *Genome Research*, **11**, 143-151.

Geiger, D., Meek, C., and Sturmfels, B. (2006). On the toric algebra of graphical models. *The Annals of Statistics*, **34**, 1463-1492.

Grayson, D. and Stillman, M. (2006). *Macaulay 2, a software system for research in algebraic geometry*, http://www.math.uiuc.edu/Macaulay2/.

Greuel, G.-M., Pfister, G., and Schoenemann, H. (2006). *SINGULAR: A computer algebra system for polynomial computations*. http://www.singular.uni-kl.de.

Guo, S. W., and Thompson, E. A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, **48**, 361-372.

Hemmecke, R., and Malkin, P. (2005). Computing generating sets of lattice ideals. arXiv:math.CO/0508359 v1.

Hosten, S., and Shapiro, J. (2000). Primary decomposition of lattice basis ideals. *Journal of Symbolic Computation*, **29**, 625-639.

Hosten, S., and Sullivant, S. (2002). Gröbner basis and polyhedral geometry of reducible and cyclic models. *Journal of Combinatorial Theory A*, **100**, 277-301.

Hosten, S., and Sullivant, S. (2004). Ideals of adjacent minors. *Journal of Algebra*, **277**, 615-642.

Kreuzer, M., and Robbiano, L. (2000). *Computational Commutative Algebra.* Springer, New York.

Pistone, G., Riccomagno, E., and Wynn, H. (2001). *Algebraic Statistics.* Chapman and Hall, Boca Raton.

Pistone, G., and Wynn, H. (1999). Finitely generated cumulants. *Statistica Sinica*, **9**, 1029-1052.

R Development Core Team (2004). *R: A Language and Environment for Statistical Computing.* http://www.R-project.org.

Schrijver, A. (1989). *Theory of linear and integer programming.* Wiley-Interscience Series in Discrete Mathematics, A Wiley-Interscience Publication, John Wiley & Sons Ltd., Chichester.

Sturmfels, B. (1996). *Gröbner Bases and Convex Polytopes.* AMS, Providence, Rhode Island.

Sturmfels, B. (2002). *Solving Systems of Polynomial Equations.* AMS, Providence, Rhode Island.

Whittaker, J. (1990). *Graphical Models in Applied Mathematical Multivariate Statistics.* Wiley, New York.