The Diaconis–Sturmfels algorithm and rules of succession

IAN H. DINWOODIE

Department of Mathematics, Tulane University, New Orleans, LA 70118, USA. E-mail:
ihd@math.tulane.edu

A multinomial rule of succession is derived with conditioning on partial past information. The probability computation is done with a Markov chain on tables of nonnegative integers using Groebner bases as described by Diaconis and Sturmfels.

Keywords: Laplace; Markov chain; Gröbner basis; succession probability

1. Introduction

Diaconis and Sturmfels (1993) developed a method for constructing a symmetric and irreducible Markov chain on an intersection of hyperplanes in Z_+^r. We shall show how their algorithm can be used for computing probabilities of succession in a general multi-dimensional Bayesian framework.

   The algorithm of Diaconis and Sturmfels was developed to handle practical and difficult problems in categorical data. It generalizes a useful technique for producing $r \times c$ tables with fixed row and column sums, which apparently first appeared in Besag and Clifford (1989) for testing a particular model for data in a contingency table. The manuscript of Diaconis and Sturmfels also contains methods for hypergeometric sampling, applications, and a complete theoretical treatment through toric ideals. Since their method has not yet become widely used, the present paper describes their construction in rigorous but less abstract terms. It is shown how irreducible Markov chains on hyperplanes naturally lead to telescoping sums of monomial differences. The ideals which arise from these differences are studied primarily with the division algorithm for polynomials in several variables. We use the Markov chain in Section 3 for computing probabilities of succession in a Bayesian framework conditioned on arbitrary incomplete present information. Our formula in Proposition 3.1 gives the succession probability in terms of an expectation over the set on which the event is conditioned with respect to the uniform distribution. We believe that this application is new.

2. Markov chains

An element $\mathbf{a} = (a_1, \ldots, a_r) \in Z_+^r$ will equivalently be denoted $X^{\mathbf{a}} := x_1^{a_1} x_2^{a_2} \ldots x_r^{a_r}$. Here Z_+^r denotes the set of r-tuples of non-negative integers. Let $T: \mathbf{R}^r \to \mathbf{R}^d$ be a linear

map such that $T(Z_+^r) \subset Z_+^d$. We are interested in constructing a symmetric Markov chain on the level set $S = \{\mathbf{a} \in Z_+^r: T(\mathbf{a}) = T(\mathbf{a}_0) \in Z_+^d\}$ (assumed to be finite) with moves

$$M = \{\pm f_1, \ldots, \pm f_m\}, \tag{2.1}$$

where $f_i \in Z^r$ satisfies $T(f_i) = \mathbf{0}$.

The Markov transition matrix is defined by

$$\pi(\mathbf{a}, \mathbf{a} + f_i) = \frac{1}{2}\frac{1}{m} \qquad \text{if } \mathbf{a} + f_i \in S,$$

$$\pi(\mathbf{a}, \mathbf{a} - f_i) = \frac{1}{2}\frac{1}{m} \qquad \text{if } \mathbf{a} - f_i \in S, \tag{2.2}$$

and the holding probability $\pi(\mathbf{a}, \mathbf{a})$ makes the matrix stochastic. Thus, one chooses one of the f_i uniformly; then a coin is flipped to determine the sign. If the new state $\mathbf{a} \pm f_i$ has non-negative entries, the move is made. Otherwise, one holds fast.

The main problem is finding a set of moves M which produces an irreducible Markov chain in the state space S. The set M will produce an irreducible chain if, for each $\mathbf{a}, \mathbf{b} \in S$, there is a sequence $f_{i_1}, \ldots, f_{i_k}$ of moves and a sequence of signs $z_1, \ldots, z_k$ such that

$$\mathbf{a} + z_1 f_{i_1} + \ldots + z_j f_{i_j} \in S, \qquad j = 1, \ldots, k,$$

$$\mathbf{a} + z_1 f_{i_1} + \ldots + z_k f_{i_k} = \mathbf{b}.$$

This means that there is a path in S from $\mathbf{a}$ to $\mathbf{b}$, which preserves $T(\mathbf{a})$ and which remains in Z_+^r.

The construction of the moves M for the chain (2.2) uses the algebra of polynomials of several variables. We first show how monomial differences arise naturally when studying paths using a set of moves M. This leads to Corollary 2.1, which shows the importance of a Gröbner basis when trying to connect two elements of the state space S.

Let $\mathbf{a}$ and $\mathbf{b}$ be two elements in S, and suppose that there is a path from $\mathbf{a}$ to $\mathbf{b}$ with moves in M. The path can be written $\mathbf{a}, \mathbf{a}_1 = \mathbf{a} - z_1 f_{i_1}, \mathbf{a}_2 = \mathbf{a} - z_1 f_{i_1} - z_2 f_{i_2}, \ldots, \mathbf{a}_k = \mathbf{b}$, where z_i represents the sign $+$ or $-$ and $\mathbf{a}_i \in Z_+^r$ for each $1 \leq i \leq k$. Below, $f_1^+ = \max\{f_1, 0\}$ is the positive part of f_1 and $f_1^- = \max\{-f_1, 0\}$ is the negative part, which both have non-negative entries, and $f_1^{z_1}$ will mean f_1^+ if $z_1 = +$ and f_1^- if $z_1 = -$. With this notation, we can write

$$X^{\mathbf{a}} - X^{\mathbf{b}} = (X^{\mathbf{a}} - X^{\mathbf{a}_1}) + (X^{\mathbf{a}_1} - X^{\mathbf{a}_2}) + \ldots + (X^{\mathbf{a}_{k-1}} - \mathbf{X}^b)$$

$$= z_1 X^{\mathbf{a} - f_{i_1}^{z_1}}(X^{f_{i_1}^+} - X^{f_{i_1}^-}) + \ldots + z_k X^{\mathbf{a}_{k-1} - f_{i_k}^{z_k}}(X^{f_{i_k}^+} - X^{f_{i_k}^-}). \tag{2.3}$$

We are now considering the symbol $X^{\mathbf{a}}$ as an element of the ring $Q[x_1, \ldots, x_r]$, with coefficients in the rationals Q and addition and multiplication in this structure. Then (2.3) shows that the difference $X^{\mathbf{a}} - X^{\mathbf{b}}$ is in the ideal $\langle X^{f_i^+} - X^{f_i^-}: i = 1, \ldots, m \rangle$ generated by monomial differences corresponding to the moves M.

It will be important to order elements of Z_+^r and equivalently symbols $X^{\mathbf{a}}$. We say that $\mathbf{a} \geq \mathbf{b}$ if in the difference $\mathbf{a} - \mathbf{b}$ the leftmost non-zero entry is positive. We say that $X^{\mathbf{a}} \geq X^{\mathbf{b}}$ if $\mathbf{a} \geq \mathbf{b}$ in the above sense. Thus $(1, 1, 0) \geq (1, 0, 0) \geq (0, 1, 0)$. For future convenience we shall suppose that things are arranged in (2.1) so that $f_i^+ \geq f_i^-$ and

$f_i^+ \geqslant f_{i+1}^+$ in this ordering. We shall say that the monomial difference $X^{f_i^+} - X^{f_i^-}$ corresponds to the move $f_i \in M$. Define for further use

$$I_M = \langle X^{f_i^+} - X^{f_i^-} : i = 1, \ldots, m \rangle,$$

$$I_T = \langle X^{\mathbf{a}} - X^{\mathbf{b}} : T(\mathbf{a}) = T(\mathbf{b}) \rangle.$$

Then (2.3) means that the difference $X^{\mathbf{a}} - X^{\mathbf{b}}$ is in the ideal generated by I_M, which gives the following proposition.

Proposition 2.1. *If elements $\mathbf{a}$ and $\mathbf{b}$ in S are connected by moves in M, then the monomial difference $X^{\mathbf{a}} - X^{\mathbf{b}}$ is in the ideal $I_M - \langle X^{f_i^+} - X^{f_i^-} : i = 1, \ldots, m \rangle$.*

Corollary 2.1. *If the Markov chain (2.2) is irreducible, then $I_T \subset I_M$.*

Now we are interested in conditions on M which imply that two elements of S are connected with moves in M. We shall assume that the reader is familiar with the division algorithm for polynomials in several variables. This can be found for example in Cox *et al*. (1992). It is clear that the lead term of an intermediate dividend in the division of a monomial $X^{\mathbf{b}}$ by the set of monomial differences from moves M is connected to $\mathbf{b}$ in the Markov chain defined in (2.2). It is reached from the previous lead term by substracting the move f_i corresponding to the multidegree of the divisor of the previous lead term. In order that a path with moves in M between two points $\mathbf{a}$ and $\mathbf{b}$ may always be constructed by such division, the collection of monomial differences corresponding to M will require a generous collection of lead terms, which in precise terms is the condition that these monomial differences constitute a Gröbner basis for I_T. We make this argument below to prove Theorem 2.1.

Theorem 2.1. *Suppose that M is such that $\langle LT(X^{f_i^+} - X^{f_i^-}) : i = 1, \ldots, m \rangle = \langle LT(I_M) \rangle$. Then there is a path from $\mathbf{a} \in S$ to an element $\mathbf{b} \in S$ with moves in $M = \{\pm f_1, \ldots, \pm f_m\}$ if and only if $X^{\mathbf{a}} - X^{\mathbf{b}} \in I_M$.*

Proof. Proposition 2.1 said that the difference $X^{\mathbf{a}} - X^{\mathbf{b}}$ is in I_M if there is a path from $\mathbf{a}$ to $\mathbf{b}$ using moves in M. We shall prove the converse.

Suppose that $X^{\mathbf{a}} - X^{\mathbf{b}} \in I_M$. At each step in the division of $X^{\mathbf{a}} - X^{\mathbf{b}}$ by the set of monomial differences, the intermediate dividend is the difference of two elements of I_M. Therefore it belongs to I_M and its leading term $\pm X^{\mathbf{a}_\alpha}$ by definition belongs to $\langle LT(I_M) \rangle$. Since $\langle LT(I_M) \rangle = \langle LT(X^{f_i^+} - X^{f_i^-}) : i = 1, \ldots, m \rangle$, one of the leading terms of the set of divisors goes into our leading term $\pm X^{\mathbf{a}_\alpha}$ and the division can proceed one more step. The division continues until the remainder is 0. The path can be constructed with the lead terms of the intermediate dividends and the sequence of moves by tracking the moves represented by the difference $X^{f_i^+} - X^{f_i^-}$ at each stage of the division. The path is built at both ends simultaneously (see Example 2.1). □

Corollary 2.2. *Suppose the set of moves M is such that $\langle LT(I_M) \rangle = \langle LT(X^{f_i^+} - X^{f_i^-}) : i = 1, \ldots, m \rangle$. Then the Markov chain (2.2) is irreducible if and only if $I_M = I_T$.*

Proof. Suppose that the chain is irreducible. From Proposition 2.1, $I_T \subset I_M$. Since always $I_M \subset I_T$, the two ideals are the same.

Conversely, suppose the two ideals are the same, and let $\mathbf{a}$, $\mathbf{b}$ be in S. Then the difference $X^{\mathbf{a}} - X^{\mathbf{b}} \in I_M$. With the assumption on the leading terms of I_M, it follows from Theorem 2.1 that there is a path from $\mathbf{a}$ to $\mathbf{b}$. Hence the chain is irreducible. $\square$

Corollary 2.2 means that, for the chain (2.2) to be irreducible, it is enough to find a set of monomial differences $X^{f_i^+} - X^{f_i^-}$ such that $\langle LT(X^{f_i^+} - X^{f_i^-}): i = 1, \ldots, m \rangle = \langle LT(I_M) \rangle$ and such that $I_M = I_T$. However, this is equivalent to the condition that $\langle LT(X^{f_i^+} - X^{f_i^-}): i = 1, \ldots, m \rangle = \langle LT(I_T) \rangle$, since this implies that $I_M = I_T$. Such a set is called a Gröbner basis for I_T.

We are now interested in a method to find a Gröbner basis for I_T, or in other words a method to construct a finite basis of monomial differences $\{X^{f_i^+} - X^{f_i^-}: i = 1, \ldots, m\}$ in I_T such that $\langle LT(X^{f_i^+} - X^{f_i^-}): i = 1, \ldots, m \rangle = \langle LT(I_T) \rangle$. This will give us moves $M = \{\pm f_1, \ldots, \pm f_m\}$ for an irreducible chain in S, by Corollary 2.2. Proposition 2.2 below presents a method given by Diaconis and Sturmfels (1993).

Consider the ideal I in $Q[y_1, \ldots, y_d, x_1, \ldots, x_r]$ generated by the terms $X^{\mathbf{e}_i} - Y^{T(\mathbf{e}_i)}$, where $\mathbf{e}_i$ is the basis element $(0, \ldots, 1, \ldots 0) \in Z_+^r$ with 1 in the ith position. The indeterminates are ordered $y_1 > y_2 > \ldots > y_d > x_1 > 0 \ldots x_r$. This ordering is important for Theorem 2.2.

Proposition 2.2. $I_T = I \cap Q[x_1, \ldots, x_r]$.

Proof. First we show that $I_T \subset I \cap Q[x_1, \ldots, x_r]$. This will follow if $X^{\mathbf{a}} - Y^{T(\mathbf{a})} \in I$ for every $\mathbf{a} \in Z_+^r$, since then for any $\mathbf{a}$ and $\mathbf{b}$ with $T(\mathbf{a}) = T(\mathbf{b})$ it will follow that the difference $X^{\mathbf{a}} - X^{\mathbf{b}} = X^{\mathbf{a}} - Y^{T(\mathbf{a})} - X^{\mathbf{b}} + Y^{T(\mathbf{b})}$ also belongs to the ideal I.

To see that $X^{\mathbf{a}} - Y^{T(\mathbf{a})} \in I$, observe that $x_i - Y^{T(\mathbf{e}_i)} \in I$, $x_i^2 - Y^{T(2\mathbf{e}_i)}$, etc., all belong to I, by simple factorization. Now proceed by induction on the degree $|\mathbf{a}|$ of $\mathbf{a}$, defined by $|\mathbf{a}| = a_1 + \ldots + a_r$. If the degree of $\mathbf{a}$ is 1, then we saw that $X^{\mathbf{a}} - Y^{T(\mathbf{a})} \in I$. Suppose that $|\mathbf{a}| = d$, and assume we have that $X^{\mathbf{b}} - Y^{T(\mathbf{b})} \in I$ for all $\mathbf{b}$ with $|\mathbf{b}| < d$. If the coordinate $a_\alpha > 0$ in $\mathbf{a}$, then

$$(x_\alpha - Y^{T(\mathbf{e}_\alpha)})(X^{\mathbf{a}-\mathbf{e}_\alpha} + Y^T(\mathbf{a} - \mathbf{e}_\alpha)) \in I,$$

$$(x_\alpha + Y^{T(\mathbf{e}_\alpha)})(X^{\mathbf{a}-\mathbf{e}_\alpha} - Y^{T(\mathbf{a}-\mathbf{e}_\alpha)}) \in I \qquad \text{(by the induction hypothesis),}$$

and therefore their difference $x_\alpha Y^{T(\mathbf{a}-\mathbf{e}_\alpha)} - X^{\mathbf{a}-\mathbf{e}_\alpha} Y^{T(\mathbf{e}_\alpha)} \in I$, but then the sum $X^{\mathbf{a}} - Y^{T(\mathbf{a})} = (x_\alpha + Y^{T(\mathbf{e}_\alpha)})(X^{\mathbf{a}-\mathbf{e}_\alpha} - Y^{T(\mathbf{a}-\mathbf{e}_\alpha)}) + x_\alpha Y^{T(\mathbf{a}-\mathbf{e}_\alpha)} - X^{\mathbf{a}-\mathbf{e}_\alpha} Y^{T(\mathbf{e}_\alpha)}) \in I$. This proves that $I_T \subset I \cap Q[x_1, \ldots, x_r]$.

Next we show that $I \cap Q[x_1, \ldots, x_r] \subset I_T$. Let ϕ_T be the ring homomorphism from $Q[y_1, \ldots, y_d, x_1, \ldots, x_r]$ into $Q[y_1, \ldots, y_d]$ with the defining property

$$\phi_T(X^{\mathbf{a}}) = Y^{T(\mathbf{a})},$$

$$\phi_T(Y^{\mathbf{w}}) = Y^{\mathbf{w}}.$$

This is well defined since T is linear. We shall show that $I \cap Q[x_1, \ldots, x_r] \subset \ker(\phi_T) \cap$

$Q[x_1, \ldots, x_r] \subset I_T$. It is clear that $I \cap Q[x_1, \ldots, x_r] \subset \ker(\phi_T) \cap Q[x_1, \ldots, x_r]$, since each element $X^{e_i} - Y^{T(e_i)}$ of the generating set for I belongs to $\ker(\phi_T)$.

Finally we prove that $\ker(\phi_T) \cap Q[x_1, \ldots, x_r] \subset I_T$. If this were not the case, there would be a polynomial in $\ker(\phi_T) \cap Q[x_1, \ldots, x_r]$ but not in I_T. Let $p_0(x_1, \ldots, x_r) = \sum_{i=1}^{k} z_i X^{\mathbf{a}_i}$ be such a polynomial, written with coefficients $z_i = \pm 1$, ordered but with possible repetition.

Since $\mathbf{0} = \phi_T(p_0) = \sum_{i=1}^{k} z_i Y^{T(\mathbf{a}_i)}$, it must be the case that the image $Y^{T(\mathbf{a}_1)}$ of the lead monomial $X^{\mathbf{a}_1} > 1$ is the same as the image $\phi_T(X^{\mathbf{a}_j})$ of a lesser monomial $X^{\mathbf{a}_j}$ in p_0, in order for the cancellation of monomials to take place in the image $\sum_{i=1}^{k} z_i Y^{T(\mathbf{a}_i)}$. Taking all the monomials in p_0 whose image is $Y^{T(\mathbf{a}_1)}$ and their coefficients, we get a polynomial $p_T = \sum_{j=1}^{\alpha} z_{i_j} X^{\mathbf{a}_{i_j}}$ with coefficients ± 1 which must sum to 0. It can be written then as a sum of monomial differences and $p_T \in I_T \subset \ker(\phi_T)$.

Since $\ker(\phi_T) \cap Q[x_1, \ldots, x_r]$ is an ideal and it contains p_T, it also contains $p_1 = p_0 - p_T$, which is a polynomial of strictly smaller degree than p_0. Also p_1 could not be in I_T; otherwise $p_0 = p_1 + p_T$ would also be in this ideal.

Therefore, if $\ker(\phi_T) \cap Q[x_1, \ldots, x_r]$ were not contained in I_T, there would be no polynomial of smallest degree in the difference $\ker(\phi_T) \cap Q[x_1, \ldots, x_r] - I_T$, which cannot be true. Therefore, the two ideals must be the same.　□

Recall that a Gröbner basis $G = \langle p_1, \ldots, p_g \rangle$ for an ideal J has the defining property that

$$\langle LT(p_1), \ldots, LT(p_g) \rangle = \langle LT(p) \colon p \in J \rangle,$$

which says that the set G has sufficient lead terms that the division algorithm will proceed successfully to leave eventually remainder 0 when the dividend is in the ideal generated by G.

If $G = \{X^{f_i^+} - X^{f_i^-} : i = 1, \ldots, m\}$ is a Gröbner basis of monomial differences for the ideal I_T, then the division algorithm constructs the path from one point $\mathbf{a} \in S$ to another point $\mathbf{b}$. Consider dividing $X^{\mathbf{a}} - X^{\mathbf{b}}$ by the basis until one reaches remainder 0. The lead terms of the intermediate dividends are points in the path, which decrease at first from $X^{\mathbf{a}}$. The element of the basis which is chosen as divisor of an intermediate lead term represents the move which steps from that lead term to the next. At some point, $-X^{\mathbf{b}}$ will become the lead term if it does not disappear completely. The minus sign on the chosen divisor $X^{f_i^+} - X^{f_i^-}$ flags a move backwards from $\mathbf{b}$ of $-f_i$. Thus we have put in place the final move (backwards) from $X^{\mathbf{b}}$ to $X^{\mathbf{b}-f_i}$ and we see that the division builds downward segments on both ends of the path. Which end is considered at a particular stage is determined by the sign of the lead term of the intermediate dividend. The last step in the division joins the two ends.

Not only are Gröbner bases useful for constructing paths, but also they are essential for eliminating variables. By Proposition 2.2, we can view the equations $x_i - Y^{T(e_i)} = 0$, $i = 1, \ldots, r$, as a parametrization of the variety of points on which elements of the ideal I_T vanish.

The following is the important Proposition 3.6 of Diaconis and Sturmfels (1993).

Theorem 2.2. *A Gröbner basis for I_T is found by computing a Gröbner basis for I, and then taking only those elements which involve the indeterminates $x_1, \ldots, x_r$.*

Proof. This is the elimination theorem (Cox *et al*. 1992, p. 114), together with Proposition 2.2. ☐

We give a purely illustrative example to make the symbols clear.

Example 2.1. Consider the illustrative problem of constructing the Markov chain on the set of partitions of the integer 10 into parts 1, 2 and 3. Here, $T: Z_+^3 \to Z_+$ has the form $T(a_1, a_2, a_3) = a_1 + 2a_2 + 3a_3$; so $T(e_1) = 1$, $T(e_2) = 2$, $T(e_3) = 3$. From Maple, a Gröbner basis of monomial differences for I_T is $x_1^2 - x_2$, $x_1 x_2 - x_3$, $x_1 x_3 - x_2^2$, $x_2^3 - x_3^2$. Thus an irreducible Markov chain on a level set for T is implemented by choosing $+$ or $-$ and then choosing uniformly from among the possible increments $(2, -1, 0)$, $(1, 1, -1)$, $(1, -2, 1)$, $(0, 3, -2)$.

Furthermore, the Gröbner basis enables one to construct a path between two points in S by division. The path from $(4, 0, 2)$ to $(2, 4, 0)$ is found by dividing $X^{(4,0,2)} - X^{(2,4,0)}$ by the basis, which gives first move $-(2, -1, 0)$, last move backwards $-(2, -1, 0)$, etc. This gives the sequence $-(2, -1, 0)$, $-(2, -1, 0)$, $(0, 3, -2)$, $(2, -1, 0)$.

3. Rules of succession

The rule of succession says that, after n successes in n Bernoulli trials with unknown parameter p, the probability of a further success is $(n + 1)/(n + 2)$. Zabell (1989) discusses the origin of the rule with Laplace and its connection with exchangeability. The rule can be derived in a Bayesian framework as in de Finetti (1972), which discusses the connections with inductive reasoning. If we choose a uniform prior distribution on the parameter p before the trials begin, then

$$P\{S_n + 1 = n + 1 | S_n = n\} = \left(\int_0^1 p^{n+1} \, \mathrm{d}p \right) \Big/ \left(\int_0^1 p^n \, \mathrm{d}p \right) = \frac{n+1}{n+2}.$$

Consider now independent trials with r possible outcomes on each. Place a uniform prior distribution on the standard r-simplex Δ of probabilities. Choose the parameter $\mathbf{p}$ uniformly, and then sample n times independently from the r possibilities with distribution $\mathbf{p}$. Let S_n be the r-tuple which records the number of objects in class i, $i = 1, \ldots, r$. If $A, B \subset Z_+^r$, we are interested in the succession probability $P\{S_{n+1} \in B | S_n \in A\}$.

Let $\mathbf{e}_i$ be the r-tuple given by $\mathbf{e}_i = (0, \ldots, 1, \ldots, 0)$, which is 1 in the ith place and 0 elsewhere.

Proposition 3.1. *The succession probability* $P\{S_{n+1} \in B | S_n \in A\}$ *takes the form*

$$\frac{1}{n+r} \sum_{\mathbf{a} \in A} \sum_{i=1}^{r} (a_i + 1) \frac{I_B(\mathbf{a} + \mathbf{e}_i)}{|A|}, \tag{3.1}$$

where $\mathbf{a} = (a_1, \ldots, a_r)$.

Proof. We need $P\{S_{n+1} \in B, S_n \in A\}$ and $P\{S_n \in A\}$. It is well known that the uniform prior on the probability simplex yields the uniform distribution on S_n (the so-called Bose–Einstein statistics); so

$$P\{S_n \in A\} = \frac{|A|}{\binom{n+r-1}{r-1}}.$$

Now

$$P\{S_n \in A, S_{n+1} \in B\} = \int_\Delta P(S_n \in A, S_{n+1} \in B | \mathbf{p}) \, d\sigma(\mathbf{p})$$

$$= \int_\Delta \sum_{(\mathbf{a},\mathbf{b}) \in A \times B} P(S_n = \mathbf{a}, S_{n+1} = \mathbf{b} | \mathbf{p}) \, d\sigma(\mathbf{p})$$

$$= \int_\Delta \sum_{i=1}^{r} \sum_{\mathbf{a} \in A} I_B(\mathbf{a} + \mathbf{e}_i) P(S_n = \mathbf{a}, S_{n+1} = \mathbf{a} + \mathbf{e}_i | \mathbf{p}) \, d\sigma(\mathbf{p})$$

$$= \sum_{i=1}^{r} \int_\Delta \sum_{\mathbf{a} \in A} I_B(\mathbf{a} + \mathbf{e}_i) \frac{n!}{a_1! \ldots a_r!} p_1^{a_1} \ldots p_r^{a_r} p_i \, d\sigma(\mathbf{p})$$

$$= \sum_{i=1}^{r} \sum_{\mathbf{a} \in A} I_B(\mathbf{a} + \mathbf{e}_i) \frac{n!}{a_1! \ldots a_r!} (r-1)! \frac{a_1! \ldots a_r!(a_i + 1)}{(n+1+r-1)!}.$$

Using the well-known formula for $\int p_1^{a_1} \ldots p_r^{a_r} p_i \, d\sigma(\mathbf{p})$, this equals

$$\sum_{i=1}^{r} \frac{n!(r-1)!}{(n+r)!} \sum_{\mathbf{a} \in A} (a_i + 1) I_B(\mathbf{a} + \mathbf{e}_i) = \sum_{i=1}^{r} \frac{n!(r-1)!}{(n+r)!} |A| \sum_{\mathbf{a} \in A} (a_i + 1) \frac{I_B(\mathbf{a} + \mathbf{e}_i)}{|A|}.$$

Now dividing by $P\{S_n \in A\}$, we get (3.1). $\qquad\square$

The formula above can be put in the form

$$P\{S_{n+1} \in B | S_n \in A\} = \frac{1}{n+r} E\left(\sum_{i=1}^{r} (a_i + 1) I_B(\mathbf{a} + \mathbf{e}_i) \right),$$

where E indicates expectation with respect to the uniform distribution on the set A. This means that, to compute the probability of succession, one should run a symmetric Markov chain in A and compute the expectation as a time average. We shall show how this is done in the next examples. Consider the categorical data of Snee (1974) used in Diaconis and Sturmfels (1993), which classifies 592 subjects by hair and eye colour. In our notation the table is S_{592} (Table 1).

Example 3.1. Suppose now that we are interested in the distribution of the next 4-vector of column sums conditioned on the present table S_{592}. Let B_k be the event where the kth column sum in S_{593} is one greater than its sum in the present table S_{592}. Here $A = \{S_{592} = (a_{ij})\}$; so $|A| = 1$ and the formula of Proposition 3.1. becomes

$$\frac{1}{n+r} \sum_{ij=1}^{4} (a_{ij} + 1) I_{B_k}(\mathbf{a} + \mathbf{e}_{ij}) = \frac{1}{n+r} \sum_{i=1}^{4} (a_{ik} + 1) = \frac{1}{592 + 16} \left(\sum_{i=1}^{4} a_{ik} + 1 \right).$$

Therefore, the first column sum will be 109 with conditional probability $(108 + 4)/(592 + 16)$, the second will be 287 with conditional probability $(286 + 4)/(592 + 16)$, the third will be 72 with conditional probability $(71 + 4)/(592 + 16)$, and the fourth will be 128 with conditional probability $(127 + 4/592 + 16)$.

Example 3.2. Suppose now that we are interested in the distribution of the next 4-vector of column sums conditioned on the present column sums. The present table S_{592} involves $r = 16$ classifications, and we let $(a_{ij})i = 1, \ldots, 4, j = 1, \ldots, 4$ denote a 4×4 table. We are conditioning on the event $A = \{(a_{ij}): T(\mathbf{a}) = (108, 286, 71, 127)\}$, where $T: \mathbf{R}^{16} \to \mathbf{R}^4$ sums the columns of the table.

Let B_k be the event where the kth column sum in the next table S_{593} will be 1 greater than its present value. The probability conditioned on the present column sums is

Table 1. S_{592}

Eye colour	Number of subjects with the following hair colours				
	Black	Brunette	Red	Blonde	Total
Brown	68	119	26	7	220
Blue	20	84	17	94	215
Hazel	15	54	14	10	93
Green	5	29	14	16	64
Total	108	286	71	127	592

$$\frac{1}{n+r}\sum_{ij=1}^{4}\sum_{\mathbf{a}\in A}(a_{ij}+1)\frac{I_{B_k}(\mathbf{a}+\mathbf{e}_{ij})}{|A|}=\frac{1}{n+r}\frac{1}{|A|}\sum_{i=1}^{4}\sum_{\mathbf{a}\in A}(a_{ik}+1)$$

$$=\frac{1}{n+r}\frac{1}{|A|}\sum_{\mathbf{a}\in A}\sum_{i=1}^{4}(a_{i1}+1)$$

$$=\frac{1}{n+r}\frac{1}{|A|}|A|\left(\sum_{i=1}^{4}a_{ik}+4\right)$$

$$=\frac{1}{n+r}\left(\sum_{i=1}^{4}a_{ik}+4\right).$$

The conditional probabilities work out to be the same as in Example 3.1.

Example 3.3. Suppose next that we have incomplete information from the present table S_{592}. We are interested in the distribution of the second column sum, conditioned on the present first column sum, the present fourth row sum, and the total number of observations. Then the event $A=\{(a_{ij})\in Z_{+}^{4\times4}\colon\sum_{i=1}^{4}a_{i1}=108,\ \sum_{j=1}^{4}a_{4j}=64,\ \sum_{ij=1}^{4}a_{ij}=592\}$.

Let B_k be the event where the second column sum in the next table S_{593} is no more than the integer $k\geqslant0$. Then the conditional probability of B_k given the present information A takes the form

$$\frac{1}{n+r}\sum_{ij=1}^{4}\sum_{\mathbf{a}\in A}(a_{ij}+1)\frac{I_{B_k}(\mathbf{a}+\mathbf{e}_{ij})}{|A|}.$$

Much work is needed to look at each element $\mathbf{a}\in A$ to see whether it satisfies $\mathbf{a}+\mathbf{e}_{ij}\in B_k$. The Diaconis–Sturmfels algorithm will finish the problem by using a time average from the symmetric chain on the set A to compute the expectation above.

Our 16 indeterminates are (x_{ij}), $i=1,\ldots,4$, $j=1,\ldots,4$, and our map $T\colon\mathbf{R}^{16}\to\mathbf{R}^{3}$ which defines A is given by $T(a_{ij})=(\sum_{i=1}^{4}a_{i1},\ \sum_{j=1}^{4}a_{4j},\ \sum_{ij=1}^{4}a_{ij})$. To compute a Gröbner basis in Maple for the ideal of monomial differences, we set $X:=[y1,\ y2,\ y3,\ x11,\ x12,$ $x13,\ x14,\ x21,\ x22,\ x23,\ x24,\ x31,\ x32,\ x33,\ x34,\ x41,\ x42,\ x43,\ x44]$ and $F:=[x11,$ $-y1^*y3,\ x12-y3,\ x13-y3,\ x14-y3,\ x21-y1^*y3,\ x22-y3,\ x23-y3,\ x24-y3,$ $x31-y1^*y3,\ x32-y3,\ x33-y3,\ x34-y3,\ x41-y1^*y2^*y3,\ x42-y2^*y3,\ x43-y2^*y3,$ $x44-y2^*y3]$. The command gbasis $(F,\ X,\ \text{plex})$ gives the following 13 monomial differences in the x variables: $x_{11}-x_{31},\ x_{12}-x_{34},\ x_{13}-x_{34},\ x_{14}-x_{34},\ x_{21}-x_{31},\ x_{22}-x_{34},$ $x_{23}-x_{34},\ x_{24}-x_{34},\ -x_{34}x_{41}+x_{31}x_{44},\ x_{32}-x_{34},\ x_{33}-x_{34},\ x_{42}-x_{44},\ x_{43}-x_{44}.$

The first of these corresponds to incrementing entry $(1,1)$ while decrementing entry $(3,1)$ to preserve the first column sum. It is clear that all the corresponding moves preserve the first column sum, the final row sum and the total number of observations. Since they form a Gröbner basis, they also connect all tables with the same sums as those observed in S_{592} (Corollary 2.2).

Running the chain shows that $P\{B_{286}|A\}$ is quite close to 1.

Acknowledgements

References

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

Cox, D., Little, J. and O'Shea, D. (1992) *Ideals, Varieties, and Algorithms*. New York: Springer-Verlag.

Diaconis, P. and Sturmfels, B. (1993) Algebraic algorithms for sampling from conditional distributions. Technical Report 6, Stanford University.

de Finetti, B. (1972) *Probability, Induction, and Statistics*. New York: Wiley.

Snee, R. D. (1974) Graphical display of two-way contingency tables. *Amer. Statist.* **28**, 9–12.

Zabell, S. L. (1989) The rule of succession. *Erkenntnis*, **31**, 283–321.