Computational commutative algebra in discrete statistics

Computational Commutative Algebra in Discrete Statistics

Giovanni Pistone, Eva Riccomagno, and Henry P. Wynn

Abstract. This paper develops formally some ideas introduced in [**14**, Chapter 6]. We show that the manifold of probabilities associated to an exponential model with monomial sufficient statistics on a lattice is an algebraic variety coming from a toric ideal. This fact is relevant because of the existence of computational tools for commutative algebra that can be applied for example to log-linear models for contingency tables. The same algebraic structure is used to discuss conditional independence models on trees.

1. Introduction

In the last few years the subject of Computational Commutative Algebra has attracted much attention in a number of applied fields. This rise of interest is proved in particular by the number of textbooks and monographs recently published. In this paper we refer especially to [**6**], [**3**] and [**11**]. In the monograph [**14**, Chapter 6] a fundamental methodology is given which uses Gröbner bases (at two levels) in the construction and analysis of statistical models and sub-models. In this paper we give a condensed version of this construction together with related further rigorous results pertaining the statistics of models on lattices. In addition a number of key examples draw out the connections between some parts of the basic construction which are important in statistics.

The use we make of Computational Commutative Algebra can be described heuristically as falling in three stages. The first stage refers to the construction of polynomial functions of a variable x over $\mathbb{R}^d$ with values y in $\mathbb{R}$, giving exact interpolation of "minimal degree" at a finite set of points $D \subset \mathbb{R}^d$, called an experimental design or support. The second stage considers the special kind of functions related to probability models of the form

$$p(x) > 0, \quad x \in D, \quad \text{and} \quad \sum_{x \in D} p(x) = 1$$

We call D a support then and talk about a discrete probability model p over a support D. There are two principal cases $y = p(x)$ and $y = \log p(x)$. Suppose that

1991 *Mathematics Subject Classification.* Primary ; Secondary .

G.P. was supported by the grant MURST *Methods for planning experiments and their applications* (A. Giovagnoli) and by UFR de Mathématiques et d'Informatique, Université Louis Pasteur, Strasbourg.

in the second case the sought polynomial is $t(x)$ then

$$p(x) = \exp(t(x)), \quad x \in D$$

In a special case when D is a subset of an integer lattice a third step is possible in which by reparametrization $p(x)$ can be written as a "power product". From this, using toric ideal theory, the algebraic equations satisfied by the $p(x)$, $(x \in D)$, considered as indeterminates, are seen to belong to a toric ideal.

Statistical sub-models are set up by considering, in this construction, the class of polynomials which are candidates for $p(x)$ or $\log p(x)$ as a polynomial, or by considering the algebraic restrictions to be satisfied by parameters, such as the individual probabilities. Note here the double use of $p(x)$ which can be both a polynomial function or a vector of parameters. For example independence models which might be written

$$p(x_1, x_2) = p_1(x_1)p_2(x_2)$$

where $p_1(x_1)$ and $p_2(x_2)$ are the marginal distributions of X_1 and X_2 respectively and where $x = (x_1 : x_2)$ is some partition of the variable space. This converts naturally to

$$p(x_1, x_2) = \exp t_1(x_1) \exp t_2(x_2) = \exp(t(x_1) + t(x_2))$$

giving additivity of the exponential representation.

2. Commutative Algebra in Statistics

The set of real random variables on a probability space has the algebraic structure of a commutative ring. What we stress here is the less trivial fact that *computational* commutative algebra is relevant to problems encountered in statistics in the same sense that computational linear algebra is.

It is outside the scope of this article to give a self contained treatment of the notions in commutative algebra we found important for statistics. We restrict ourselves to a list of definitions and properties, and refer mainly to the textbooks [**6**], [**5**] and [**11**]. The actual computational feasibility of our methods is important. For the purpose of illustration in this paper we use the system CoCoA, see [**2**] and [**11**, 275–304]. Many other symbolic computation systems are available, see [**6**, Appendix C] and [**14**, Section 1.2] for short reviews. The use of Gröbner bases in statistics is just one of the possible applications of symbolic computation in statistics. Compare [**1**] for a different approach to the issue of symbolic computation in this field of application. The special application we are describing here was initiated in [**7**] and [**15**].

The key notion allowing the actual computation over polynomial rings is that of total ordering on terms. Here a *term* is a product of indeterminates (or variables), a monomial is a constant (typically a real number) multiplied by a term, a polynomial is a sum of monomials. The ring of polynomials with variables $x_1, \ldots, x_d$ and real coefficients is denoted $\mathbb{R}[x_1, \ldots, x_d]$.

DEFINITION 2.1. Let d be a positive integer. A *term-ordering* is a total ordering relation $\succ$ over the monomials in d indeterminates such that

(1) $1 \prec x^\alpha$ for all $x^\alpha = x_1^{\alpha_1} \ldots x_d^{\alpha_d}$ with $\alpha_i \in \mathbb{Z}_{\geq 0}$ for all $i = 1, \ldots, d$,
(2) if $x^\alpha \succ x^\beta$ then $x^{\alpha+\gamma} \succ x^{\beta+\gamma}$ for all vector α, β, γ with non negative entries.

The *leading term* of a polynomial $f \in \mathbb{R}[x_1, \ldots, x_d]$ with respect to the term-ordering $\succ$ is the largest of its terms with respect to $\succ$, $\mathrm{LT}_\succ(f)$.

Point (1) corresponds to the requirement that the relation $\succ$ is a well-ordering, while Point (2) expresses compatibility with simplification of monomials, that is, if x^β divides x^α then $x^\alpha \succ x^\beta$. Alternative notations for a term-ordering are $\succ_\tau$ and τ.

To a finite set $D \subset \mathbb{R}^d$ of distinct points one can associate (in a non -unique way) a set of $N = \#D$ terms, so that all the real functions over D can be expressed as real linear combinations of these terms. A technique to compute these termss is given by Gröbner bases and it assumes a term-ordering. We will refer to such a basis as a *monomial basis.*

A support $D \subset \mathbb{R}^d$ is seen as an (affine) algebraic variety that is the set of zeros of a system of polynomial equations.

Definition 2.2. Let $D \subset \mathbb{R}^d$ be a finite set of distinct points, that is a support. Let $\mathbb{R}[x_1, \ldots, x_d]$ be the set of all polynomials in the x_i's ($i = 1, \ldots, d$) and with real coefficients. The set $\mathrm{Ideal}(D) \subset \mathbb{R}[x_1, \ldots, x_d]$ is the set of all polynomials whose zeros include the points in D.

Theorem 2.3. *Let D be a support in $\mathbb{R}^d$.*

(1) $\mathrm{Ideal}(D)$ *is a polynomial ideal.*
(2) $\mathrm{Ideal}(D)$ *is generated by $f_1, \ldots, f_s \in \mathbb{R}[x_1, \ldots, x_d]$ if and only if all and only the solutions of the system of polynomial equations $f_1 = \ldots = f_s = 0$ are the points in D.*
(3) $\mathrm{Ideal}(D)$ *is a radical ideal.*

Proof. All obvious. We recall that an ideal is a subset of the ring which is closed by linear combinations with coefficients in the ring, and an ideal I is radical if $f^m \in I$ implies $f \in I$. $\square$

Definition 2.4. Given a polynomial ideal $I \subset \mathbb{R}[x_1, \ldots, x_d]$ and a term-ordering τ, a finite subset G of I is called a Gröbner basis for I with respect to τ if

$$\mathrm{Ideal}(\mathrm{LT}_\tau(g) : g \in G) = \mathrm{Ideal}(\mathrm{LT}_\tau(f) : g \in I)$$

The important *Hilbert basis theorem* states that every polynomial ideal (except $\mathrm{Ideal}(0)$) is finitely generated. That is for every polynomial ideal I there exists (non-unique) $f_1, \ldots, f_s \in \mathbb{R}[x_1, \ldots, x_d]$ such that for all $f \in I$ then $f = \sum_{i=1}^s h_i f_i$ where $h_i \in \mathbb{R}[x_1, \ldots, x_d]$. Moreover for every term-ordering there exist Gröbner bases. For a term-ordering, τ there exists a unique reduced Gröbner basis where a Gröbner basis G is reduced if for all $g \in G$ 1) the coefficient of $\mathrm{LT}_\tau(g)$ is 1 and 2) no term of g belongs to the ideal $\mathrm{Ideal}(\mathrm{LT}(f) : f \in G \setminus \{g\})$.

Given an ideal $I \subset \mathbb{R}[x_1, \ldots, x_d]$ the equivalence relation, $\sim_I$ over $\mathbb{R}[x_1, \ldots, x_d]$ modulo I is defined as: for all $f, g \in \mathbb{R}[x_1, \ldots, x_d]$ then $f \sim_I g$ if and only if $f - g \in I$. The quotient space is then defined as

$$\mathbb{R}[x_1, \ldots, x_d]/I = \{[f] : f \sim_I g \text{ for } f, g \in \mathbb{R}[x_1, \ldots, x_d]\}$$

Theorem 2.5. *Let I be a polynomial ideal in $\mathbb{R}[x_1, \ldots, x_d]$ and τ a term-ordering.*

(1) *The quotient space $\mathbb{R}[x_1, \ldots, x_d]/I$ is a vector space over the real numbers.*
(2) *Let G be a Gröbner basis of I with respect to τ.*

(a) *A vector space basis of* $\mathbb{R}[x_1, \ldots, x_d]/I$ *is given by*

$$\begin{aligned}\mathrm{Est}_\tau(D) &= \{x^\alpha : x^\alpha \text{ is not divisible by } \mathrm{LT}_\tau(g),\ g \in G\} \\ &= \{x^\alpha : \alpha \in L\}\end{aligned}$$

(b) *The full design matrix* $Z = [x^\alpha]_{x \in D; \alpha \in L}$ *is invertible.*

(3) *Let* D *be a support and* $I = \mathrm{Ideal}(D)$ *the corresponding support ideal. Then*

(a) $f \sim_I g$ *if and only if* $f(a) = g(a)$ *for all* $a \in D$

(b) *The dimension of* $\mathbb{R}[x_1, \ldots, x_d]/I$ *as* $\mathbb{R}$*-vector space is equal to the number of points in* D*.*

(c) *The design matrix for the regression model*

$$y = \sum_{\alpha \in M} \theta_\alpha x^\alpha + \varepsilon$$

with $M \subseteq L$ *is full rank.*

Note that the monomial basis obtained via Gröbner bases are just an instance of all possible monomial bases. For a discussion of monomial bases arising in binary designs, see [**9**].

3. Experimental designs and residual space

In this section we consider sub-models, first in the interpolation case. The link with the area of experimental design and with "statistical thinking" is strict. Thus let be given a set

$$\{x^\alpha : \alpha \in M\}$$

of linearly independent monomials on D. The linear independence in particular is always true if $M \subset L$, that is the model is a subset of the exponent list in Est. Let

$$Z_1 = [x^\alpha]_{x \in D; \alpha \in M}$$

be the design matrix corresponding to the statistical sub-model M. It is very common, for example in statistical analysis to fit sub-models such as linear (planar) models or quadratic models. Now extend Z_1 to a full basis for the column space (range) of Z in an orthogonal way. With abuse of notation we can replace Z by this partially orthogonal basis and write

$$Z = [Z_1 : Z_2]$$

where $Z_1^t Z_2 = 0$. In statistical modeling we consider a model

$$y = Z_1 \theta + \varepsilon$$

and the least squares estimate of θ is $\hat{\theta} = (Z_1^t Z_1)^t Z_1^t Y$. The fit is $\hat{Y} = PY$ where P is the projector $Z_1^t (Z_1^t Z_1)^{-1} Z_1$ and the orthogonal projector is

$$I - P = Z_2^t \left(Z_2^t Z_2\right)^{-1} Z_2$$

In the Gröbner case, the invertibility of $Z_1^t Z_1$ and $Z_2^t Z_2$ comes from the fact that Z_1 and Z_2 are full rank which itself derives from the Gröbner basis construction. In statistical jargon we might consider Z_2 (columns) as "spanning the space of residual".

Despite this straightforward explanation the structure of Z_2 is not so transparent. We will see that special instances are useful, for example integer valued matrices.

EXAMPLE 3.1 (One-dimensional polynomial regression). Consider the set of points $D = \{0, 1, \ldots, N-1\} \subset \mathbb{Z}$. Construct orthogonal polynomials on D of increasing order

$$1, p_1(x), , \ldots, p_{N-1}(x)$$

Let the sub-model consist of terms

$$\{1, x, x^2, \ldots, x^q\}$$

where $q < n-1$. Then it is clear that a candidate for Z_2 is

$$\left[p_j(x)\right]_{x \in D; j=q+1,\ldots,N-1}$$

Notice that since D consists of integers the $p_j(x)$ will have rational coefficients. These can be converted to integer by multiplication of a suitable integer. Then Z_2 as constructed can be replace by an integer matrix.

EXAMPLE 3.2. Consider the full factorial design 2^d with levels -1 and 1. For all term-orderings the Est set is composed of all the multi-linear terms, that is in L there are all the vectors $\alpha = \alpha_1 \ldots \alpha_d$ with $\alpha_i = 0, 1$ for all $i = 1, \ldots, d$. The choice of -1 and 1 as levels implies that the Z matrix is orthogonal. Thus for every sub-model $M \subset L$ automatically $Z_2 = Z_{|\bar{M}}$ is orthogonal to $Z_1 = Z_{|M}$.

EXAMPLE 3.3. Let D be the two-dimensional support with the particular structure

$$\begin{array}{ccc} \bullet & & \\ \bullet & & \\ \bullet & \bullet & \bullet \end{array}$$

that is $D = \{(0,2), (0,1), (0,0), (1,0), (2,0)\}$. This is called an echelon design. Echelon designs have the property that if a point $a = (a_1, \ldots, a_d)$ is in D then also $(y_1, \ldots, y_d)$ is in D if $0 \le y_i \le a_d$ for all $i = 1, \ldots, d$. Also in this case for all τ there is only one Est, namely $\{x^a : a \in D\}$.

Consider a subset of Est that is again echelon. For example $M = \{(0,0), (1,0), (0,1)\}$. Then, before conversion to the $[Z_1 : Z_2]$ form we have

$$Z = \begin{bmatrix} 1 & 0 & 2 & 0 & 4 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 2 & 0 & 4 & 0 \end{bmatrix}$$

where the first three columns refer to M and give Z_1. The last two columns are the evaluations of x_1^2 and x_2^2 at D. Now we may chose Z_2 so that its columns generate the same two-dimensional space as the last two columns of Z and moreover its columns are orthogonal to the first three columns of Z. For example we can take

$$Z_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \\ -1 & 0 \\ 0 & 2 \end{bmatrix}$$

We will see that the definition of the space of residuals of a model in the vector space sense is not enough for a fully satisfying theory of exponential models on a lattice. We will need actually a combinatorial definition, as follows.

DEFINITION 3.4. Let $D \subset \mathbb{Z}^d_{\geq 0}$ be a lattice support and let $\text{Est} = \{x^\alpha : \alpha \in L\}$ be a monomial (Gröbner) basis. If M is a monomial model and $Z_1 = \left[x^\beta\right]_{x \in D, \beta \in M}$ the model matrix, we call *integer orthogonal* every polynomial p of the quotient space $\mathbb{R}[x_1, \ldots, x_d]/\operatorname{Ideal}(D)$ such that:

(1) The polynomial p is integer valued, $p(x) \in \mathbb{Z}$, $x \in D$.
(2) The integer values $p(x), x \in D$ are relatively prime.
(3) The polynomial p is orthogonal to the model, that is

$$\sum_{x \in D} x^\beta p(x) = 0, \quad \beta \in M$$

4. Toric ideals

We recall briefly the notion of toric ideal as presented in [**11**, Tutorial 38] and [**16**], see also [**17**] and [**5**]. The relevance of this concept for the analysis of contingency tables for fixed margins and the application to simulation for exact tests has been singled out in [**7**], see also [**8**]. We present below our application to exponential models on finite lattices $D \subset \mathbb{Z}^d_{\geq 0}$. We are not going to discuss any application, but the reader will recognize, for example, one of the main theme in the classical theory of log-linear model for contingency tables, see [**10**].

Toric ideals are related to algebraic varieties defined parametrically by monomials.

DEFINITION 4.1. Let $\mathbb{R}[\zeta] = \mathbb{R}[\zeta_0, \ldots, \zeta_s]$ be a polynomial ring and let $\zeta_0^{\alpha_{i0}} \cdots \zeta_s^{\alpha_{is}}$, $\alpha_i \in \mathbb{Z}^{s+1}_{\geq 0}$, $i = 1, \ldots, n$, be a finite list of terms. Consider in the extended polynomial ring $\mathbb{R}[p, \zeta] = \mathbb{R}[p_1, \ldots, p_n, \zeta_0, \ldots, \zeta_s]$ the ideal J generated by the binomials $p_i - \zeta_0^{\alpha_{i0}} \cdots \zeta_s^{\alpha_{is}}$. If $\mathbb{R}[p] = \mathbb{R}[p_1, \ldots, p_n]$, the elimination ideal $I = J \cap R[p]$ is called the *toric ideal* of the matrix $Z_1 = \left[\alpha_{ij}\right]_{j=1,\ldots,n; i=0,\ldots,s}$.

We outline the relevant facts with a two simple non-statistical examples of application of toric ideals, both taken from the literature. The theory relevant for statistical models will be described in the next Section 5

EXAMPLE 4.2 ([**6**, Section 4.6]). In $\mathbb{R}[p_1, p_2, p_3, \zeta]$ we consider the identities $p_1 = \zeta^3$, $p_2 = \zeta^4$, $p_3 = \zeta^5$. This can be considered in several ways, all meaningful: a) a parametric description of a curve in the three dimensional space of coordinates p_1, p_2, p_3; b) a homeomorphism of the ring $\mathbb{R}[p_1, p_2, p_3]$ into the ring $\mathbb{R}[\zeta]$ derived uniquely from the rules $p_1 \mapsto \zeta^3$, $p_2 \mapsto \zeta^4$, $p_1 \mapsto \zeta^5$; c) an ideal J generated by the polynomials $p_1 - \zeta^3$, $p_2 - \zeta^4$, $p_1 - \zeta^5$. The kernel of the mapping defined in b) is the ideal I consisting of all polynomials $f(p_1, p_2, p_3)$ such that $f(\zeta_0^3, \zeta_0^4, \zeta_0^5) = 0$. This ideal is easily seen to be equal to the elimination ideal $J \cap \mathbb{R}[p_1, p_2, p_3]$. The curve a) is the variety of the ideal I. The computation using CoCoA gives the basis $-p_1^3 + p_2 p_3$, $-p_2^2 + p_1 p_3$, $p_1^2 p_2 - p_3^2$ and the corresponding equations are the implicit equations for the curve a).

Now let us consider a reduction of the previous problem to linear algebra: assume $\zeta_0 > 0$ and rewrite the initial identities as $\log p_1 = 3 \log \zeta$, $\log p_2 = 4 \log \zeta$, $p_3 = 5 \log \zeta$, that is the vector $\log[p]$, $[p] = (p_1, p_2, p_3)$, is proportional to the vector $u = (3, 4, 5)$. By introducing the vectors $v = (1, -2, 1)$ and $w = (3, -1, -2)$ which span $u^\perp$, we can write $v^t \log[p] = 0$, $w^t \log[p] = 0$, or $p_1 p_3 = p_2^2$, $p_1^3 = p_2 p_3$. Comparison with the previous results shows that one equation is missing, and, more important, that the variety now contains the axis $p_1 = p_2 = 0$ which is not part of the original curve a). Computation of the colon ideal Ideal $\left(p_1p_3 - p_2^2, p_1^3 - p_2p_3\right)$: Ideal (p_1, p_2) gives I, eliminating the extra unwanted component.

We describe now the theory relevant to fill the gap between the residual matrix and the toric ideal. Given two polynomial ideals $I, J \subset \mathbb{R}[x_1, \ldots, x_d]$ the *colon ideal* $I : J$ is defined as

$$I : J = \{g \in \mathbb{R}[x_1, \ldots, x_d] : fg \in I \text{ for all } f \in J\}$$

The saturation of the ideal $I \subset \mathbb{R}[x_1, \ldots, x_d]$ with respect to the polynomial $f \in \mathbb{R}[x_1, \ldots, x_d]$ is defined as

$$I : f^\infty = \{g \in \mathbb{R}[x_1, \ldots, x_d] : f^m g \in I \text{ for some } m > 0\}$$

THEOREM 4.3. *Let $I \subset \mathbb{R}[x_1, \ldots, x_d]$ be an ideal and let $f \in \mathbb{R}[x_1, \ldots, x_d]$ be a polynomial. Then*

(1) *$I : f^\infty$ is an ideal.*
(2) *$I : f \subset I : f^2 \subset I : f^3$.*
(3) *There exists $m > 0$ such that $I : f^m = I : f^\infty$.*
(4) *Suppose that I is generated by $f_1, \ldots, f_s$ and define the ideal $\tilde{I}$ of the extended polynomial ring $\mathbb{R}[x_1, \ldots, x_d, y]$ generated by $f_1, \ldots, f_s$ and $1 - fy$. Then $I : f^\infty = \tilde{I} \bigcap \mathbb{R}[x_1, \ldots, x_d]$.*

Saturation is a way to localize around f, that is to make f invertible.

This construction is related to the choice of the Z_2 matrix in our above development ($Z_1^t Z_2 = 0$) and in particular to the fact that multiplication or division of columns of Z_2 by a non-zero integer (and which keeps Z_2 integer) and more general column reduction of Z_2 leads to the same saturation of the ideal and lattice ideal. That is bases of the column space spanned by Z_2 can be associated to different toric ideals. But these toric ideals have the same saturation ideal with respect to the polynomial $f = x_1 \ldots x_n$, where n is the number of columns in Z_2.

EXAMPLE 4.4 ([**16**]). Consider the following system of equations

$$\begin{cases} p_1 = st \\ p_2 = s^3t^2 \\ p_3 = st^3 \\ p_4 = s^5t^2 \end{cases} \tag{4.1}$$

The elimination of the s and t indeterminates from System (4.1) is equivalent to determine the kernel of the following homomorphism of $\mathbb{R}$-algebras

$$\begin{array}{cccc} \phi: & \mathbb{R}[p_1, p_2, p_3, p_4] & \to & \mathbb{R}[s, t] \\ & (p_1, p_2, p_3, p_4) & \longmapsto & (st, s^3t^2, st^4, s^5t^2) \end{array}$$

To System (4.1) we can associate the "Z_1" matrix

$$Z_1 = \begin{bmatrix} 1 & 1 \\ 3 & 2 \\ 1 & 3 \\ 5 & 2 \end{bmatrix}$$

and the system of linear equations

$$\begin{cases} z_1 + 3z_2 + z_3 + 5z_4 = 0 \\ z_1 + 2z_2 + 3z_3 + 2z_4 = 0 \end{cases}$$

whose solutions are the integer orthogonal vectors. Here one can see that an interesting connection of toric ideals to Diophantine linear systems; this was first pointed out in [**4**]. Two solutions for this system are $(z_1, z_2, z_3, z_4) = (7, -2, -1, 0)$ and $(4, -3, 0, 1)$. Then, a basis of the "Z_2" matrix is

$$Z_2 = \begin{bmatrix} -7 & 4 \\ 2 & -3 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and the binomials $p_1^7 - p_2^2 p_3$ and $p_1^4 p_4 - p_2^3$ are in the kernel of ϕ. Also any other polynomials in $\text{Ideal}\left(p_1^7 - p_2^2 p_3, p_1^4 p_4 - p_2^3\right)$ belongs to the kernel of ϕ. Actually there are other polynomials in $\ker(\phi)$ for example $p_1^3 p_2 - p_3 p_4$. It can be proved that $\ker(\phi)$ is the saturation ideal of $\text{Ideal}\left(p_1^7 - p_2^2 p_3, p_1^4 p_4 - p_2^3\right)$ with respect to the polynomial $p_1 p_2 p_3 p_4$.

Definition 4.1 itself suggests a method to compute toric ideals as elimination ideals of s and t from the system

$$\begin{cases} p_1 - st \\ p_2 - s^3 t^2 \\ p_3 - st^3 \\ p_4 - s^5 t^2 \end{cases}$$

Another method to compute the wanted saturation ideal is to consider the following ideal in the extended space

$$\text{Ideal}\left(p_1^7 - p_2^2 p_3, p_1^4 p_4 - p_2^3, p_1 p_2 p_3 p_4 t - 1\right) \subset \mathbb{R}[p_1, p_2, p_3, p_4, t]$$

and intersect it with $\mathbb{R}[x, p_2, p_3, p_4]$.

5. Exponential models and toric ideals

In this Section we apply all the machinery of previous Sections to exponential models. Let $D \subset \mathbb{Z}^d_{\geq 0}$ be a (lattice) support and let M be a monomial model, that is the set on monomial functions on D given by x^β, $\beta \in M$ is a linearly independent set. We assume $0_d \in M$ and consider the exponential model

$$p(x; \psi) = \exp\left(\sum_{\beta \in M} \psi_\beta x^\beta\right) \tag{5.1}$$

As $x \in Z^d_{\geq 0}$, then x^β is a non-negative integer for $\beta \in M$, and we introduce the ζ-parameters defined by

$$\zeta_\beta = \exp\left(\psi_\beta\right), \quad \beta \in M \tag{5.2}$$

The p-parameters, that is the vector [p] with components $p(x; \psi)$, $x \in D$ depend on the ζ-parameters according Equations (5.1) and (5.2), and the dependence is a monomial function:

$$p(x) = \prod_{\beta \in M} \zeta_\beta^{x^b eta} \tag{5.3}$$

As before, we consider the model matrix $Z_1 = \left[x^\beta\right]_{x \in D; \beta \in M}$ and a residual integer valued matrix Z_2.

THEOREM 5.1. *Let us consider the polynomial rings* $R(p,\zeta) = \mathbb{R}[p(x), x \in D, \zeta_\beta, \beta \in M]$, $R(p) = \mathbb{R}[p(x), x \in D]$, $R(\zeta) = \mathbb{R}[\zeta_\beta, \beta \in M]$, *and let* J *be the ideal generated in* $R(p,\zeta)$ *by Equations (5.3). Let* $\mathcal{Z}_2$ *denote the set of all integer ortogonal polynomials. Let* $I = J \cap R(p)$ *be the toric ideal in* $R(p)$.

(1) *The probability parameters* $[p]$ *of the exponential model in Equation (5.1) belong to the irreductible variety* $\text{Variety}(I) \cap \text{Variety}\left(\sum_{x\in D} p(x) - 1\right)$.
(2) *The toric ideal* I *has a binomial homogenous basis. This basis can be computed by elimination of the* ζ*-variables in* J.
(3) *For each column* f *of the residual matrix* Z_2*, denote by* f_+ *and* f_- *respectively the positive and the negative part. Then the binomial*

$$\prod_{x\in D} p(x)^{f_+(x)} - \prod_{x\in D} p(x)^{f_-(x)} \tag{5.4}$$

belongs to the toric ideal I.
(4) *Let* I_0 *be the ideal generated by the binomials in Equation (5.4). Then* $I_0 \subset I$*, and* I *is the saturation of* I_0 *for the monomial* $\prod_{\beta\in M}\zeta_\beta$*. In particular, a basis of* I *can be computed by the elimination of* t *in the ideal generated by the Equations in (5.4) and* $t\prod_{\beta\in M}\zeta_\beta - 1$.
(5) *The set of binomials*

$$\prod_{x\in D} p(x)^{g_+(x)} - \prod_{x\in D} p(x)^{g_-(x)}, \quad g \in \mathcal{Z}_2$$

is a basis of the toric ideal I.

PROOF. Point 1. follows from the exponential model expressed in the ζ-parameters, plus the normalization condition. Note that the existence of a polynomial parametric representation for this variety implies that the variety is irreductible, see [**6**, Section 4.5]. Point 2. is an application of the general theory. Point 3. follows from considering positive values for the parameters, and observing that the vector of the log-probabilities $\log[p]$ is a linear combination of the columns of the model matrix Z_1, or, equivalently, orthogonal to le columns of the matrix Z_2. Point 4. follows from the general theory. Point 5. follows from the fact that any vector in $\mathcal{Z}_2$ can be a column in Z_2 and the logarithms of the binomials in a basis of the toric ideal form an integer orthogonal vector. □

There is a very close connection to the toric ideal constructed by exhibiting the $p(x)$ as power products and the residual space construction on Z_2. Suppose that D is lattice. We can always construct an orthogonal matrix Z_2 ($Z_1^t Z_2 = 0$) with integer coefficients. Any standard construction such as Gram-Schmidt or the Cholesky orthogonal polynomial construction will lead to a rational Z_2 which can be made integer by suitable multiplication by an integer. Now Z_2 gives a basis for the toric ideal as follows. For each column j of Z_2 divide the row indices in three groups according as to whether the entry ζ_{ij} is positive, zero or negative. Call the three set of indices so obtained J_j^+, J_j^0 and J_j^- respectively. Then the toric ideal has a basis

$$\prod_{i\in J_j^+} p_i^{\zeta_{ij}} - \prod_{i\in J_j^-} p_i^{\zeta_{ij}} \qquad \text{for all } j$$

In the toric ideal theory different bases for the subspace orthogonal to $\text{Span}\,(Z_1)$ (the column space of Z_1) namely different Z_2 lead to different bases for the ideal.

One may also obtain different Gröbner bases of the toric ideal by varying the monomial ordering σ, but there is only a finite number of such bases. Their union is still a Gröbner basis and is called the universal Gröbner basis of the toric ideal (see [**17**]).

Underlying the Z_2 construction and the construction obtained by varying σ is the lattice ideal. This is easily captured by considering all integer solutions to

$$Z_2^t[q] = 0$$

which defines an integer lattice. This lattice is independent of the actual Z_2 basis (or the Gröbner basis) used in the construction.

6. Modeling independence

In statistical inference many of the probability models used are based on assumptions of independence or conditional independence. In the present Section we examine the implications for indipendence and conditional independence of our assumption of finite lattice support, $D \subset \mathbb{Z}^d_{\geq 0}$. Consider first the following simple example in order to review our basic setup in this perspective.

EXAMPLE 6.1. Let X_1 and X_2 be two independent Bernoulli (0, 1 valued) random variables with marginal probabilities $p_1 = \mathrm{P}\,(X_1 = 1)$ and $p_2 = \mathrm{P}\,(X_2 = 1)$. Independence says, for example that

$$\mathrm{P}\,(\{X_1 = 1\} \cap \{X_2 = 1\}) = p_1 p_2$$

Now consider the exponential formulation. By independence, and because The interaction term $x_1 x_2$ is omitted, this is

$$\log p(x_1, x_2; \psi) = \exp(\psi_0 + \psi_1 x_1 + \psi_2 x_2)$$

Then we have

$$Z_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad Z_2 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}$$

It is easy to check that che unique column of Z_2 is also the unique integer ortogonal, then the toric ideal in obvious notation is

$$p_{00} p_{11} - p_{10} p_{01} = 0$$

This condition is very familiar to students of the 2×2 contingency table.

The attractive feature of the exponential interpolation is that essentially any conditionality structure can be captured by the structure of the terms $X^\alpha Y^\beta Z^\gamma$, $\alpha\beta\gamma \in M$ appearing in the sub-model

$$p(X, Y, Z; \psi) = \exp\left(\sum_{\alpha\beta\gamma \in M} \psi_{\alpha\beta\gamma} X^\alpha Y^\beta Z^\gamma\right) \tag{6.1}$$

Here capital letters denote block of variables.

It is enough for the purpose of the present paper to model a single conditional independence structure. Thus for three vector random variables X, Y, Z, let X and Y be conditional independent given Z, written as $X \coprod Y | Z$. Then the following proposition holds; it is a version of a well known result in graphical models, see [**12**, Chapter 4]. We partially follow the presentation given in [**13**].

LEMMA 6.2. *Let X, Y and Z be disjoint blocks of variables. Let us assume the following:*

(1) *For each $z \in D_Z$ the corresponding section of the sample space is a product $D_{X,Y,Z=z} = D_{X,Z=z} \times D_{Y,Z=z}$.*
(2) *The joint quotient ring $\mathbb{R}[X,Y,Z]/\operatorname{Ideal}(D_{X,Y,Z})$ has a monomial vector basis that contains the terms in the model $X^\alpha Y^\beta Z^\gamma$, $\alpha\beta\gamma \in M$, and a monomial vector basis both of the marginal quotient rings*

$$\mathbb{R}[X,Z]/\operatorname{Ideal}(D_{X,Z}) \text{ and } \mathbb{R}[Y,Z]/\operatorname{Ideal}(D_{Y,Z})$$

(The assumption is in particular true in the set product case, $D_{X,Y,Z} = D_X \times D_Y \times D_Z$.) Then $X \coprod Y | Z$ under the distribution in Equation (6.1) if and only if the model structure M does not contain any (α,β,γ) with α, β both nonzero, that is any X, Y interacion is excluded.

PROOF. Under the product assumpion (1) the conditional independence is equivalent to the factorization, or, taking the logarithms, the representation of $\sum_{\alpha\beta\gamma\in M} \psi_{\alpha\beta\gamma} X^\alpha Y^\beta Z^\gamma$ as a sum of a function in $\mathbb{R}[X,Z]/\operatorname{Ideal}(D_{X,Z})$ plus a function in $\mathbb{R}[Y,Z]/\operatorname{Ideal}(D_{Y,Z})$. Under the second part of the assumption this is possible only in absence of the indicated interactions. □

The omission of the terms $X^\alpha Y^\beta Z^\gamma$ with α and β, both non-zero, $(\alpha \neq 0)(\beta \neq 0) = 1$, from the model affects the form of the residual matrix Z_2 and of $\mathcal{Z}_2$ in the analysis of the previous sections and equivalently the toric ideals. The following lemma then, links the model structure implied by $X \coprod Y | Z$ to the structure of the toric ideal.

LEMMA 6.3. *In the set up of Lemma 6.2, in the product case, construct separate series of orthogonal polynomials with integer coefficients $\{p_\alpha\}$, $\{q_\beta\}$, $\{r_\gamma\}$ on D_X, D_Y and D_Z, respectively. Then one residual matrix is*

$$Z_2 = \left[p_\alpha q_\beta r_\gamma\right]_{(\alpha\neq 0)(\beta\neq 0)=1}$$

and the toric ideal can be constructed by saturation.

PROOF. The product of integer orthogonal polynomial bases is an orthogonal integer base for the product space. □

The above can be generalized to a very important group of models based on trees.

7. Conditional independence on trees

We refer to [**12**] for graph theory in graphical models. First recall that a tree is a graph $G(E,V)$ with no cycles. By identifying a source (root) vertex $v \in V$ and giving the edges directions (arrows) away from the source the tree becomes a special case of a directed graph. Directed graphs are at the foundation of Bayesian influence diagrams and hence of modern probabilistic artificial intelligence, or learning. Moreover trees are an important subclass of models which we now explain by example.

EXAMPLE 7.1. Consider six binary random variables $X_1, \ldots, X_6$ and the following tree in which each node holds one of the random variables except for the root to which we simply attach the symbol 1. The tree in Figure 1 holds the information

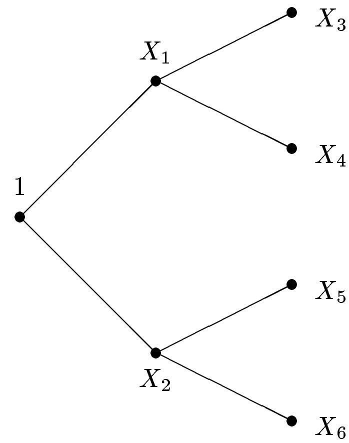

FIGURE 1. A tree of random variables.

about conditional independence as follows:

$$\begin{array}{c} X_1 \coprod X_2 \\ X_3 \coprod X_4 | X_1 \\ X_5 \coprod X_6 | X_2 \end{array}$$

The support is taken to be $D = \{0,1\}^6$ and the joint density is $p(x) > 0$, $x \in D$ can be factorized using the chain rules for condition probability. Thus with $x = (x_1, \ldots, x_6)$

$$p(x) = p_{31}(x_3|x_1)p_{41}(x_4|x_1)p_{52}(x_5|x_2)p_{62}(x_6|x_2)p_1(x_1)p_2(x_2)$$

Now switch to the exponential interpolation. The full $\mathrm{Est}_\tau(D)$ is all 2^6 multilinear terms

$$1, x_1, x_2, \ldots, x_1x_2x_3x_4x_5x_6$$

The sub-model represented by the tree is

$$\begin{array}{lllllll} M: & 1, & & & & & \\ & x_1, & x_2, & x_3, & x_4, & x_5, & x_6, \\ & x_1x_3, & x_1x_4, & x_2x_5, & x_2x_6 & & \end{array}$$

Notice that, with care, M can be read directly off the tree. Below we will give a systematic way of doing this.

DEFINITION 7.2. Let $D = \{0,1\}^d$ and let (binary) random variables $X_1, \ldots, X_d$ have a joint distribution $p(x) > 0$, $x \in D$ and $\sum_{x \in D} p(x) = 1$. The random variables are said to be a tree model with respect to a directed tree $T(E,V)$ with $\#V = 1 + d$ and some vertex v_0 if "children" are conditionally independent with respect to "parents"

$$\coprod_{e(i,j) \in E(i,j)} X_j \Bigg| X_i$$

where $E(i,j)$ is the set of directed edges out of vertex i.

The corresponding factorization of $p(x)$ takes the form

$$p(x) = \prod_{i,j:i\in V, e(i,j)\in E(i,j)} p(x_j|x_i)$$

DEFINITION 7.3. For any tree $T(E,V)$ define a maximal chain as a connected path from the root to a final branch. The maximal chains together with all their subsets form a simplicial complex. These subsets are themselves sometimes referred to as be chains, but need not be connected. Call the simplicial complex generated in this way $\mathcal{C}(T)$.

In the Example 7.1 above, the maximal chains give x_1x_3, x_1x_4, x_2x_5, x_2x_6.

THEOREM 7.4. *Let $X_1,\ldots,X_d$ be a tree model with respect to the tree $T(E,V)$ and let $\mathcal{C}(T)$ be the simplicial complex generated by the maximal chains of T. Let the distribution be $p(x) > 0$, $x \in D = \{0,1\}^d$. Then define the exponential interpolator*

$$p(x) = \exp\left(\sum_{\alpha\in M} \psi_\alpha x^\alpha\right)$$

Then for any $s \in \mathcal{C}(T)$ define

$$M_S = \{\alpha : \alpha_i = 1,\ i \in S \text{ and } \alpha_i = 0,\ i \notin S\}$$

Then

$$M = \{M_S : S \in \mathcal{C}(T)\}$$

PROOF. By repeated applications of the previous Lemma 6.2 $\square$

We are now in a position to give explicit generators for the toric ideal for the raw $p(x)$ and hence to set up algebraic formulae for the probabilistic model. We consider the setup of Theorem 7.4. The following is a direct consequence of the "automatic" orthogonality of Example 3.2.

THEOREM 7.5. *Let L be the full list for the (saturated) model consisting of all multi-linear terms. Define*

$$\overline{M} = L \setminus M$$

(where M is defined in Theorem 7.4) namely the complementary list in L. Then the model toric ideal is the saturation of the ideal generated by the binomials

$$\left\{\prod_{x:x^\alpha=1} p(x) - \prod_{x:x^\alpha=0} p(x) : \alpha \in \overline{M}\right\}$$

PROOF. By computation of the relevant Z_2 and the result on saturation. $\square$

EXAMPLE 7.6. For the Example 7.1 the list $\{x^\alpha : \alpha \in \overline{M}\}$ is

$$\begin{array}{lll}
x_1x_2, & x_1x_5, & \\
x_2x_3, & x_2x_4, & \\
x_3x_4, & x_3x_5, & x_3x_6, \\
x_4x_5, & x_6x_6, & \\
x_5x_6, & & \\
x_1x_2x_3, & \ldots, & x_4x_5x_6, \\
x_1x_2x_3x_4, & \ldots, & x_3x_4x_5x_6, \\
x_1x_2x_3x_4x_5, & \ldots, & x_2x_3x_4x_5x_6, \\
x_1x_2x_3x_4x_5x_6 & &
\end{array}$$

Each of these leads to a generator of the toric ideal. Choosing one example

$$x_3x_4 : \quad \Pi_{x:x_3x_4=1}p(x) - \Pi_{x:x_3x_4=0}p(x)$$

The algebraic conditions for the $p(x)$ are given by

$$\left\{\Pi_{x:x^\alpha=1}p(x) - \Pi_{x:x^\alpha=0}p(x) = 0 | \alpha \in \overline{M}\right\}$$

In the tree case exhibited here $\overline{M}$ can itself be seen to have its own simplicial structure.

8. Final comments

This binary case can be extended in a straightforward way to the case when each X_i is a vector random variable by using Lemmas 6.2 and 6.3.

This more general case arises as an important reduction of directed graphs which arise as the main component of Bayesian influence diagrams (for details see [**12**]). These can be explained simply as follows. We associate to each vertex v_i of a directed graph $G(E, V)$ a random variable X_i. Then the $X_1, \ldots, X_n$ (where n is the number of vertices) is said to have the Markov property with respect to $G(E, V)$ if (again) "children" are conditionally independent given "parents" and in such a way that "parents" separate "children" from "ancestors" (vertices further back but reverse reacheable from children). However when parents are grouped together so that is any two parents which have the same children are merged ("moralisation" to use the quaint term) the graph becomes a tree in which the merged parents are grouped together at a super-node (called a "clique"). Thus the general tree formula can be seen in a sense as holiding together the maximal amount of first order Markov resolution.

Returning to the first case of interpolation in Section 1 we can call the raw polynomial through the probabilities in the saturated case

$$p(x) = \sum_{\alpha \in L} \theta_\alpha x^\alpha$$

In matrix terms we have

$$\begin{array}{rcl}
[p] & = & Z[\theta] \\
[\theta] & = & Z^{-1}[p]
\end{array}$$

Adjoining to this the toric ideal conditions derived from Z_2 we can produce polynomial interpolator which satisfy the model conditions on the $p(x)$ imposed by the model $M \subset L$. We can for example compute, as possibly rational forms, all the conditional probabilities in the tree case.

References

1. D. F. Andrews and J. E. Stafford, *Symbolic computation for statistical inference*, Oxford University Pres, Oxford, 2000.
2. Antonio Capani, Gianfranco Niesi, and Lorenzo Robbiano, CoCoA*, a system for doing computations in commutative algebra*, Available via anonymous ftp from `cocoa.dima.unige.it`.
3. Arjeh M. Cohen, hans Cuypers, and Hans Sterk (eds.), *Some tapas of computer algebra*, Algorithms and Computation in Mathematics, no. 4, Springer Verlag, Berlin Heidelberg, 1999.
4. P. Conti and C. Traverso, *Buckberger algorithm and integer programming*, Proceedings of the AAECC-9, Lecture Notes in Computer Programming, no. 539, Springer, 1991, pp. 130–139.
5. D. Cox, J. Little, and D. O'Shea, *Using algebraic geometry*, Springer-Verlag, New York, 1998.
6. david Cox, John Little, and David O'Shea, *Ideal, varieties, and algorithms*, 2nd ed., Springer-Verlag, New York, 1997, 1st ed. 1992.
7. Persi Diaconis and Bernt Sturmfels, *Algebraic algorithms for sampling from conditional distributions*, The Annals of Statistics **26** (1998), no. 1, 363–397.
8. Ian H. Dinwoodie, *The Diaconis-Sturmfels algorithm and rules of succession*, Bernoulli **4** (1998), no. 3, 401–410.
9. Roberto Fontana, Giovanni Pistone, and Maria-Piera Rogantin, *Classification of two-level factorial fractions*, Journal of Statistical Planning and Inference **87** (2000), 149–172.
10. Shelby J. Haberman, *The analysis of frequency data*, The University of Chicago Press, Chicago, Ill.-London, 1974, Statistical Research Monographs, Vol. IV.
11. M. Kreuzer and Lorenzo Robbiano, *Computational commutative algebra 1*, Springer, Berlin-Heidelberg, 2000.
12. Steffen L. Lauritzen, *Graphical models*, The Clarendon Press Oxford University Press, New York, 1996, Oxford Science Publications.
13. Mauro Piccioni, *Graphical models and structural zeros*, GROSTAT IV Workshop, IUT STID Menton, 2000.
14. Giovanni Pistone, Eva Riccomagno, and Henry P. Wynn, *Algebraic statistics: Computational commutative algebra in statistics*, Chapman&Hall, 2000.
15. Giovanni Pistone and Henry P. Wynn, *Generalised confounding with Gröbner bases*, Biometrika **83** (1996), no. 1, 653–666.
16. Lorenzo Robbiano, *Toric ideals*, GROSTAT IV Workshop, IUT STID Menton, 2000.
17. B. Sturmfels, *Gröbner bases and convex polytopes*, vol. Volume 8, Ed. AMS, Providence R I, 1995, University Lecture Series.

Dipartimento di Matematica, Politecnico di Torino, Italy
E-mail address: `piston@calvino.polito.it`
Current address: UFR de Mathématiques et d'Informatique, Université Louis Pasteur, Strasbourg

EURANDOM, Eindhoven NL
E-mail address: `E.Riccomagno@tue.nl`

Department of Statistics,University of Warwick
E-mail address: `hpw@stats.warwick.ac.uk`