Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Symbolic Computation 41 (2006) 210–221

Journal of Symbolic Computation

www.elsevier.com/locate/jsc

Cumulant varieties

Giovanni Pistone[a], Henry P. Wynn[b,*]

[a] *Politecnico, Torino, Italy*
[b] *London School of Economics, London, UK*

Received 5 April 2004; accepted 2 April 2005
Available online 2 September 2005

Abstract

For a discrete distribution in R^d on a finite support D probabilities and moments are algebraically related. If there are $n = |D|$ support points then there are n probabilities $p(x), x \in D$ and n basic moments. By suitable interpolation of the probabilities using a Gröbner basis method, high order moments can be express linearly in terms of n basic moments. A main result is that high order cumulants can also be expressed as polynomial functions of n low order moments and cumulants. This means that statistical models which can be expressed via an algebraically variety for the basic probabilities and moments, such as graphical models, induce a variety for the basic cumulants, which we shall call the "cumulant variety". It is important to stress that the cumulant variety depends on the monomial ordering defining the original Gröbner basis.

Keywords: Probability; Moments; Cumulants; Variety; Ideal; Gröbner bases

1. Introduction

In the papers Pistone and Wynn (1996), Pistone et al. (2001b) and the monograph Pistone et al. (2001a) the basic idea is that, via interpolation based on Gröbner bases, algebraic formulations can be introduced to the area of experimental design and distributions with finite, discrete support. This work can now be considered as a contribution to the wider discussion under a heading of algebraic statistics which

* Corresponding author.
E-mail address: H.Wynn@lse.ac.uk (H.P. Wynn).

0747-7171/$ - see front matter
doi:10.1016/j.jsc.2005.04.008

was launched by Diaconis and Sturmfels (1998) and was the subject of the GROSTAT series of workshops. There are important recent papers; see e.g. Geiger et al. (in press). Some material appears in Sturmfels (2002), and other recent work is collected in the current volume. In this paper the section on finitely generated cumulants draws on Pistone and Wynn (1999) and basic material on cumulants in statistics can be found in McCullagh (1988) and Barndorf-Nielsen and Cox (1989).

Notation. In all that follows we use the notation $u \le v$ for vectors u, v in R^d to denote $u_i \le v_i,\ i = 1, \dots, d$. The notation $u < v$ means $u \le v$ and, at for least one $j,\ (j = 1, \dots, d), u_j < v_j$.

Throughout we shall use the integer multi-index notation: $\alpha = (\alpha_1, \dots, \alpha_n)$ and for a point $x = (x_1, \dots, x_d)$ the corresponding monomial is written $x^\alpha = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$. We write $|\alpha| = \sum_{i=1}^d \alpha_i$. Note that, when α is a suffix, we drop the commas, e.g. $\mu_{(2,1)}$ is written μ_{21}.

1.1. Interpolation

The foundation for this paper is the interpolation method based on Gröbner bases. We summarize this now.

(1) A finite set D in $R^d\ (Q^d)$ can be expressed as the solution of a set of equations and can be thought of as a zero dimensional variety. The set of all polynomials with zeros on a D is the ideal, $I(D)$.
(2) A *monomial term ordering* τ is a total ordering of monomials $\prec_\tau$ such that for all $\gamma \ge 0$

$$x^\alpha \prec_\tau x^\beta$$

implies

$$x^{\alpha+\gamma} \prec_\tau x^{\beta+\gamma}.$$

We shall use the term monomial ordering for short.
(3) Every polynomial has a unique leading term with respect to a monomial term ordering.
(4) There is a Gröbner basis for $I(D)$ for a given monomial ordering.
(5) The quotient

$$K/I(D)$$

of the ring of polynomials $K[x_1, \dots, x_d]$ in $x_1, \dots, x_d$ form is a vector space spanned by a special set monomials: $x^\alpha, \alpha \in L_{D,\tau}$. These are all the monomials not divisible by the leading terms, with respect to $\prec_\tau$, of the G-basis and $|L| = |D|$.
(6) The set of multi-indices L has the "order ideal" property: $\alpha \in L$ implies $\beta \in L$ for any $0 \le \beta \le \alpha$.
(7) Any function $y(x)$ on D has a polynomial interpolator, which is unique given τ:

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

such that $y(x) = f(x),\ x \in D$.

Whenever we use the letter L we shall be assuming a set D and monomial term ordering τ.

1.2. Distributions, moments and cumulants

Consider a probability distribution with support D: $\{p(x),\ x \in D\}$. Technically, this should be defined relative to a base measure, but we simply take uniform for the base measure throughout. We note immediately that, by 1.1 (7), $p(x)$ can be interpolated over D relative to a particular monomial ordering:

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

Let $[p]$, $[\theta]$ be the vectors of probabilities and parameters, with entries in a suitable order. Then

$$Z = \{x^\alpha\}_{x \in D, \alpha \in L}$$

is the "design matrix", familiar to statisticians, and we can write

$$[p] = Z[\theta].$$

Since the interpolation method guarantees a non-singular basis Z is invertible and

$$[\theta] = Z^{-1}[p].$$

In the statistical terminology we have a saturated linear model for $p(x)$.

Let $X = (X_1, \ldots, X_d)$ be the multivariate random variable with distribution $\{p(x)\}$. For an integer exponent $\alpha = (\alpha_1, \ldots, \alpha_d) \geq 0$ we define the random monomial

$$X^\alpha = (X_1^{\alpha_1} \cdots X_d^{\alpha_d})$$

and the corresponding moment

$$\mu_\alpha = \mathrm{E}_X(X^\alpha),$$

where E_X denotes expectation with respect to X. Using the Z-matrix and letting $[\mu]$ be the vector of moments we have:

$$[\mu] = Z^T[p] = Z^T Z[\theta].$$

Definition 1. For a random variable X in R^d. The moment generating function is defined by

$$\begin{aligned} M_X(s) &= \mathrm{E}_X\left\{\exp\left(\sum_{i=1}^n s_i X_i\right)\right\} \\ &= \mathrm{E}_X\left\{\sum_{\alpha \geq 0} \frac{X^\alpha s^\alpha}{\alpha!}\right\} \\ &= \sum_{\alpha \geq 0} \frac{\mu_\alpha s^\alpha}{\alpha!}. \end{aligned}$$

Cumulants use the same multi-index as moments: κ_α. The simplest way to define them is via the moment generating function.

Definition 2. For a random variable X the cumulant generating function is

$$K_X(s) = \log M_X(s) = \sum_{\alpha \geq 0} \frac{\kappa_\alpha s^\alpha}{\alpha!}.$$

Note that $\mu_{0\ldots0} = 1$, $\kappa_{0\ldots0} = 0$ and the first order cumulants are equal to the (marginal) expectations (means): $\mu_{100\ldots} = \kappa_{100\ldots}$, etc.

1.3. The uses of cumulants in statistics

There are some advantages is using cumulants which arise from the following points.

(1) For n independent random variables $X = (X_1, \ldots, X_n)$ the joint cumulant generating function is the sum of the marginal cumulant generating function:

$$K_X(s_1, \ldots, s_n) = K_{X_1}(s_1) + \cdots + K_{X_n}(s_n).$$

(2) For n independent random variables the joint cumulant generating function of the sum $S_n = \sum_{i=1}^n X_i$ is the sum of the marginal cumulant generating functions:

$$K_{S_n}(s) = K_{X_1}(s) + \cdots + K_{X_n}(s).$$

(3) From (1) we have that independence of $X_1, \ldots, X_N$ is equivalent to all mixed cumulants κ_α, with at least two non-zero α_i, being zero.

There is not the space to describe all the applications of cumulants to statistical analysis. Here is a short discussion.

Based on (2) above one can easily obtain the cumulants of statistic which are sums of independent components $X_1, \ldots, X_n$ (a sample), and scaled or shifted versions thereof. For example if $\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^n X_i$ is a sample mean of independent identically distributed X_i, with cumulant generating function $K_{X_1}(s)$, then its cumulant generating function is $nK_{X_1}(\frac{s}{n})$. The cumulant generating function of the normalized statistic $S_n = \frac{S_n - n\mu}{\sqrt{n}}$ is

$$nK_{X_1-\mu}\left(\frac{s}{\sqrt{n}}\right) = \sum_{|\alpha|=2}^{\infty} \frac{1}{\alpha!}\kappa_\alpha s^\alpha n^{-\frac{|\alpha|}{2}+1}.$$

This can be extended to give cumulants, and cumulant expansions, of more complicated sample quantities, so-called sample cumulants or k-statistics. The theory is attractive when these are symmetric functions of $X_1, \ldots, X_n$.

Inversion of the moment generating function, expressed in cumulants, gives Edgeworth approximations to distributions. The Edgeworth expansions use the generalized Hermite polynomials with respect to the relevant distribution, that is, a set of polynomials of increasing degree which are orthogonal with respect to the distribution. The coefficients of these polynomials can be written as polynomials in cumulants. For a standard normal density we obtain the classical Hermite polynomials. These methods, which can be generally referred to as asymptotic methods, can be used to give approximations to the distribution of maximum likelihood and many other estimators for statistical models. See McCullagh (1988) for an introduction to Edgeworth Expansion. Cumulants are also used in special approximations such as saddle point approximations; see Jensen (1995).

2. Moment and cumulant aliasing

A main aim of this paper is to show that given a discrete distribution with finite support D and a monomial term ordering τ any moment or cumulant, μ_β, κ_β, can be expressed in terms of the moments or cumulates $\mu_\alpha, \kappa_\alpha, \alpha \in L$.

We begin with a general result which is true for any distribution on R^d.

Lemma 3. *Any μ_β, $\beta \geq 0$ can be expressed as a polynomial function of $\{\kappa_\alpha : 0 \leq \alpha \leq \beta\}$ and vice versa: any κ_β, $\beta \geq 0$ can be expressed as a polynomial function of $\{\mu_\alpha : 0 \leq \alpha \leq \beta\}$.*

Proof. For an index $\beta = \{\beta_1, \ldots, \beta_n\}$ define the mixed partial derivative:

$$D_\beta = \frac{\partial^{\beta_1+\cdots+\beta_n}}{\partial s_1^{\beta_1} \cdots \partial s_n^{\beta_n}}$$

For example, if $\beta = (2, 2)$, $D_{22} = \dfrac{\partial^4}{\partial s_1^2 \partial s_2^2}$. Then from

$$M(s) = \exp(K(s)) = \exp\left(\sum_{\alpha \geq 0} \frac{\kappa_\alpha s^\alpha}{\alpha!}\right),$$

we have

$$\mu_\beta = D_\beta \exp\left(\sum_{\alpha \geq 0} \kappa_\alpha s^\alpha\right)\Bigg|_{s=0}.$$

By studying this differential it is clear that for any $\alpha > \beta$, κ_α is eliminated when $s = 0$. Thus only κ_α with $\alpha \leq \beta$ remain. We can apply the same argument to

$$\kappa_\beta = D_\beta \log\left(\sum_{\alpha \geq 0} \mu_\alpha s^\alpha\right)\Bigg|_{s=0}$$

which completes the proof.

Moment aliasing was a term used in Pistone et al. (2001a) to show how, for a discrete random variable X on support D, high order moments are expressible in terms of low order moments. Let

$$H(x) = \exp\left(\sum_{i=1}^{n} s_i X_i\right) = \sum_{\beta \geq 0} \frac{s^\beta x^\beta}{\beta!}.$$

Now interpolate $H(x)$ on D, using (7) above:

$$H(x) = \sum_{\alpha \in L} b_\alpha(s) x^\alpha.$$

Replace x by the random variable X and take expectations to obtain:

$$M_X(s) = \sum_{\beta \geq 0} \frac{s^\beta \mu_\beta}{\beta!} = \sum_{\alpha \in L} b_\alpha(s)\mu_\alpha.$$

But then

$$\mu_\beta = \sum_{\alpha \in L} b_{\alpha,\beta}\mu_\alpha,$$

where $b_{\alpha,\beta} = D_\beta b_\alpha(s)|_{s=0}$. Note that these coefficients, coming as they do from the initial interpolation, only depend on the support and choice of monomial ordering. We summarize this as follows.

Lemma 4 (*Moment Aliasing*). *For a discrete distribution and monomial order* τ *every moment* μ_β, $\beta \geq 0$ *is expressible as a linear function of the moments* μ_α, $\alpha \in L$, *whose coefficients depend only the support and choice of monomial ordering, not the* $p(x)$.

A main result of this paper is the following.

Lemma 5 (*Cumulant Aliasing*). *For a discrete distribution and monomial order* τ *every cumulant* κ_β, $\beta \geq 0$ *is expressible as a polynomial function of the cumulants* κ_α, $\alpha \in L$, *whose form is only dependent on the support and monomial ordering.*

Proof. Use Lemma 3 to express κ_β as polynomial in moments $\mu_\beta \leq \alpha$. Then use the moment aliasing in Lemma 4 to express any such μ_β in terms of the $\mu_\gamma, \gamma \in L$. Then use Lemma 3 again to express any such μ_γ in terms of $\kappa_\delta, \delta \leq \gamma$. But, crucially, the order ideal property of L ((6), above) means that all such δ remain in L. This completes the proof.

Example 1 (*Cumulant Aliasing*). Let $d = 2$ and take D to be

$$\{(0,0),(1,0),(0,1),(2,0),(1,1),(0,2),(2,1),(1,2)\}.$$

From the special structure of D we can claim that $D = L$ for any monomial ordering. Note that $(2, 2)$ is not in L and our aim is to express κ_{22} in terms of κ_α, $\alpha \in L$. First express κ_{22} in terms of moments:

$$\begin{aligned} k_{22} \;=\; & \mu_{22} - 2\mu_{21}\mu_{01} + 2\mu_{20}\mu_{01}^2 - \mu_{20}\mu_{02} - 2\mu_{11}^2 + 8\mu_{10}\mu_{11}\mu_{01} - 2\mu_{10}\mu_{12} \\ & - 6\mu_{10}^2\mu_{01}^2 + 2\mu_{10}^2\mu_{02}. \end{aligned}$$

But we can "fold" μ_{22} into L. One method for doing this in simple problems is to use the Gröbner basis directly. The relevant G-basis element is

$$x_1^2x_2^2 - x_1^2x_2 - x_1x_2^2 + x_1x_2.$$

Since this is zero on the support D we can take expectation directly to obtain

$$\mu_{22} = \mu_{21} + \mu_{12} - \mu_{11}.$$

Substituting for μ_{22} and converting back to cumulants we have the following value for:

$$\begin{aligned}\kappa_{22} \;=\; & -\kappa_{11} + \kappa_{21} + \kappa_{12} - \kappa_{10}\kappa_{01} - \kappa_{20}\kappa_{02} + \kappa_{10}^2\kappa_{01} + \kappa_{10}\kappa_{01}^2 + \kappa_{20}\kappa_{01} \\ & + \kappa_{10}\kappa_{02} - \kappa_{10}^2\kappa_{02} - \kappa_{20}\kappa_{01}^2 + 2\kappa_{10}\kappa_{11} + 2\kappa_{01}\kappa_{11} - 2\kappa_{10}\kappa_{12} - 2\kappa_{10}\kappa_{21} \\ & - 2\kappa_{11}^2 - 4\kappa_{10}\kappa_{01}\kappa_{11} - \kappa_{10}^2\kappa_{01}^2.\end{aligned}$$

2.1. *"Binary" cumulants*

There is an algebraic 1–1 correspondence between certain products of moments and cumulants, sometimes called the exp-log relationships, when α is binary. Barndorff-Nielsen and Cox (1989) may be referred to for Lemmas 7 and 8 below. The binary case is useful because when $D \subset \{0, 1\}^d$ it is automatic that every $\alpha \in L$ is binary so that by Lemmas 4 and 5 all moments and cumulants can express in terms of binary moments and cumulants. A further reason that binary cumulants are important is that κ_{11} is the covariance of X_1 and X_2 and binary cumulants are generalizations of covariances.

Definition: A *block partition* is a separation of the index set $\{1, \ldots, d\}$ into disjoint blocks. For example, for $n = 3$ we have

$$1|2|3,\;\; 12|3,\;\; 1|23,\;\; 13|2,\;\; 123.$$

These partitions together with the refinement partial order $\prec$ define a lattice: e.g. $12|3|45 \prec 12|345$.

For any such partition we can code up the blocks: e.g. $b = 123|45|6$ can be coded alternatively as $\{111000, 000110, 000001\}$. Each such binary code has a corresponding moment and cumulant. For example 111000 has

$$\mu_{111000}, \quad \kappa_{111000}$$

in the notation of this paper.

We next introduce the product of such cumulants over all terms in a partition. With somewhat loose notation this is written $\prod_{\alpha\in b}$. For b above:

$$\prod_{\alpha\in b} \kappa_\alpha = \kappa_{11100}\kappa_{000110}\kappa_{000001}.$$

Lemma 6. *For a random variable X and binary α.*

$$\mu_\alpha = \sum_{b\prec\alpha} \prod_{\beta\in b} \kappa_\beta.$$

A formula of similar type expresses cumulants in terms of moments. The proof uses Möbius inversion and is omitted.

Lemma 7. *For a random variable X and binary α.*

$$\mu_\mu = \sum_{b\prec\alpha} (-1)^{|b|-1}(|b|-1)! \prod_{\beta\in b} \mu_\beta,$$

where $|b|$ is the number of blocks in the block partition b.

Lemma 7, gives for example,

$$\kappa_{111} = \mu_{111} - \mu_{100}\mu_{011} - \mu_{010}\mu_{101} - \mu_{001}\mu_{110} + 2\mu_{100}\mu_{010}\mu_{001}. \tag{1}$$

We can move from the binary moments and cumulants to those based on a multi-index by "copying". For example κ_{23} is obtained by taking random variables $X_1, X_1', X_2, X_2', X_2''$ and the elementary cumulant κ_{11111} and then equating $X_1 = X_1'$ and $X_2 = X_2' = X_2''$.

This provides an alternative proof of Lemma 3, as follows. Suppose we wish to express μ_α in terms of κ_β with $\beta \leq \alpha$. Split every index α_i into α_i copies of unity. For each i, associate random variables $X_{ij},\ \ j = 1, \ldots, \alpha_i$. Then take the binary moments for the $|a| = \sum \alpha_i$ random variables X_{ij}. Use Lemma 7 to express the binary moments in terms of the binary cumulants. Copy by setting all the X_{ij} equal to X_i for $i = 1, \ldots, d$. But every cumulant κ_β thus obtained has $\beta \leq \alpha$. A similar argument applies for expressing moments in terms of cumulants using the copying method with Lemma 7.

As an example, suppose $d = 2$ and we want to express κ_{21} in terms of moments we take (1) and let $X_2 = X_1'$. Then

$$\kappa_{21} = \mu_{21} - 2\mu_{10}\mu_{11} - \mu_{01}\mu_{20} + 2\mu_{10}^2\mu_{01}.$$

This is easily confirmed using Lemma 3.

2.2. *Finitely generated cumulants*

In Pistone and Wynn (1999) the authors introduced the concept of finitely generated cumulants (FGC). It concerns the relation between the first and second derivatives of the cumulant generating function $K(s)$. Thus with respect to $s = (s_1, \ldots, s_d)$ we have d first derivatives and $\frac{d(d+1)}{2}$ second (order two) derivatives of $K(s)$. We call these collections respectively K' and K''. Then the FGC property means that there is a multivariate polynomial function F such that:

$$F(K'', K') = 0.$$

Familiar in statistics is the notion of a variance function which states that (for a suitable set-up) K'' is an explicit function of K'. There is considerable interest in characterizing families of distributions in which this function is polynomial. See particularly the papers of Letac, e.g. Letac and Mora (1990). The FGC property can be considered as a generalization of this to include implicit as well as explicit polynomial relationships.

If we impose the condition that every $x \in D$ is a non-negative integer vector then any distribution on D has the FGC property. In fact D rational is enough but slightly harder to work with. The proof is as follows. Substitute $t_i = \exp(s_i)$ so that

$$\exp\left(\sum x_i s_i\right) = t_1^{s_1} \cdots t_d^{s_d} = t^s.$$

With this substitution K' and K'' become rational forms in the t_i. The polynomial F is then found by eliminating the t_i using standard elimination theory.

Our purpose here is to combine this with the cumulant aliasing of Lemma 4. Note that the function F typically depends on the raw probabilities $p(x)$ in contrast to the aliasing formulae, which do not. However, since the $p(x)$ can be expressed in terms of the cumulants we can express F in terms of K', K'' and the cumulants $\kappa_\alpha,\ \alpha \in L$. We collect this as a lemma, with the rational D condition for completeness, and omit the proof.

Lemma 8. *For rational D there is a (possibly multivariate) polynomial $F(K', K'')$ of the cumulants κ_α, $\alpha \in L$ and the first and second order derivatives K' and K'' such that $F = 0$ (identically in s).*

The equation $F = 0$ can be used to obtain all higher order derivatives of K in terms of κ_α, $\alpha \in L$, simply by repeated differentiation and setting $s = 0$. It is this idea which led to the concept of finite generation.

Example 2 (*Finitely Generated Cumulants*). Let $d = 1$ so that K' and K'' are univariate. Consider a distribution with probabilities p_0, p_1, p_2 on $\{0, 1, 2\}$ respectively. Then

$$M(s) = p_0 + \exp(s)p_1 + \exp(2s)p_2.$$

Substituting $t = \exp(s)$ we see that

$$K' = \frac{p_1 t + 2p_2 t^2}{p_0 + p_1 t + p_2 t^2}$$
$$K'' = \frac{t(p_0 p1 + p_1 p_2 t^2 + 4p_0 p_2 t)}{(p_0 + p_1 t + p_2 t^2)^2}.$$

Elimination of t gives a single equation

$$K'^4 p_1^2 - 4K'^4 p_0 p_2 + 16K'^3 p_0 p_2 - 4K'^3 p_1^2 - 8K'^2 K'' p_0 p_2 + 2K'^2 K'' p_1^2 - 16K'^2 p_0 p_2 + 5K'^2 p_1^2 + 16K'K'' p_0 p_2 - 4K'K'' p_1^2 - 2K' p_1^2 + p_1^2 K''^2 - 4K''^2 p_0 p_2 + 2K'' p_1^2 = 0.$$

Substitute $p_2 = 1 - p_0 - p_1$, $p_0 = 1 - \frac{3}{2}\kappa_1 + \frac{1}{2}\kappa_1^2 + \frac{1}{2}\kappa_2$, $p_1 = 2\kappa_1 - \kappa_1^2 - \kappa_2$ to find F.

In general, repeated differentiation of $F = 0$ and setting $s = 0$ gives higher order cumulants in terms of the cumulants $\kappa_\alpha, \alpha \in L$, recapturing the formula given by Lemma 5.

3. Models and cumulant varieties

The second main theme of this paper is that statistical models which are expressed by algebraic conditions, that is varieties, on probabilities or moments induce equivalent conditions on the cumulants. Sometimes these conditions are easy to write down in terms of cumulants; in other cases they can be obtained from conditions on probabilities or moments by elimination, using computer algebra.

Definition 9. A cumulant variety is a set of algebraic conditions on the cumulants κ_α of a distribution $\{p(x), x \in D\}$.

Example 3 (*Independence*). Suppose that we have $X_1, \ldots, X_d$ defined on a finite product space $D = D_1 \times \cdots \times D_d$, where $|D_j| = n_j$, $j = 1, \ldots, d$. Then with respect to any monomial ordering,

$$L = \{\alpha : 0 \le \alpha \le (n_1 - 1, \ldots, n_d - 1)\}$$

and we can map all cumulants into $\kappa_\alpha, \alpha \in L$. Independence is equivalent to all mixed cumulants $\kappa_\alpha, \alpha \in L$ being zero.

For example if $d = 2$ and $D = \{0, 1, 2\}^2$. Then independence is equivalent to

$$\kappa_{11} = \kappa_{21} = \kappa_{12} = \kappa_{22} = 0.$$

Example 4 (*Independent Marginals*). There is interest in statistical models expressed in terms of independence conditions on marginals. For example let $D = \{0, 1\}^4$. Then the condition that all 2×2 marginals, that is all 6 pairs $\{X_i, X_j\}_{i<j}$, are independent is

$$\kappa_{1100} = \kappa_{1010} = \kappa_{1001} = \kappa_{0110} = \kappa_{0101} = \kappa_{0011} = 0.$$

Example 5 (*Conditional Independence*). Conditional independence is the key condition in Bayesian graphical models and more complex categorical models; see Lauritzen (1996). We show how to derive the cumulant variety for conditional independence in the binary case.

For random variables (X_1, X_2, X_3) with support $\{-1, 1\}^3$ suppose that X_2 and X_3 are conditionally independent given X_1. There are several way to express this. We can start with a parametrization of the raw probabilities (with obvious notation):

$$\begin{array}{ll} p_{---} = t_1 t_3 & p_{+--} = t_5 t_6 \\ p_{-+-} = t_2 t_3 & p_{++-} = t_6 t_7 \\ p_{--+} = t_1 t_4 & p_{+-+} = t_5 t_8 \\ p_{-++} = t_2 t_4 & p_{+++} = t_6 t_8. \end{array}$$

Alternatively, we can eliminate the t_i to obtain the well-known (toric) conditions on the probabilities.

$$p_{---} p_{-++} - p_{-+-} p_{--+} = 0 \tag{2}$$

$$p_{+--} p_{+++} - p_{++-} p_{+-+} = 0. \tag{3}$$

We can express the moments in terms of probabilities using $[p] = (Z^T)^{-1}[\mu]$, from Section 1.2. The moments are expressed in terms of cumulants using Lemma 7, or equivalently by the Taylor expansion in Lemma 3:

$$\begin{aligned} \mu_{000} &= 1, \quad \mu_{100} = \kappa_{100}, \quad \mu_{010} = \kappa_{010}, \quad \mu_{001} = \kappa_{001} \\ \mu_{110} &= \kappa_{110} + \kappa_{100}\kappa_{010}, \quad \mu_{101} = \kappa_{101} + \kappa_{100}\kappa_{001}, \quad \mu_{011} = \kappa_{011} + \kappa_{010}\kappa_{001} \\ \mu_{111} &= \kappa_{111} + \kappa_{100}\kappa_{011} + \kappa_{010}\kappa_{101} + \kappa_{001}\kappa_{110}\kappa_{100}\kappa_{010}\kappa_{011}. \end{aligned}$$

By taking the sum and difference of (2) and (3), we obtain

$$\begin{aligned} \kappa_{100}\kappa_{111} + \kappa_{100}^2\kappa_{011} - \kappa_{101}\kappa_{110} + \kappa_{011} &= 0 \\ \kappa_{111} + 2\kappa_{100}\kappa_{011} &= 0. \end{aligned}$$

If the means are zero, $\kappa_{100} = \kappa_{010} = \kappa_{001} = 0$, the equations reduce to

$$\kappa_{111} = 0, \qquad \kappa_{101}\kappa_{110} - \kappa_{011} = 0. \tag{4}$$

Now in this case the covariance matrix of (X_1, X_2, X_3) is

$$C = \begin{bmatrix} 1 & \kappa_{110} & \kappa_{101} \\ \kappa_{110} & 1 & \kappa_{011} \\ \kappa_{101} & \kappa_{011} & 1 \end{bmatrix}$$

and we obtain the second equation of (4)

$$\det(C)C^{-1}_{23} = \kappa_{101}\kappa_{110} - \kappa_{k011}.$$

The matrix C^{-1} is sometimes called the "influence matrix" and when the distribution is Gaussian the condition for conditional independence is C^{-1}_{23}. We see that provided the inverse exists, the means are zero and $\kappa_{111} = 0$; this condition also applies in the binary case.

Example 6 (*4-cycle*). We again consider the binary case and $D = \{-1, 1\}^4$. Then X_1, X_2, X_3, X_4 are a 4-cycle if X_1, X_2 are conditionally independent given X_3, X_4 *and* X_3, X_4 are conditionally independent given X_1, X_2.

The conditions for the 4-cycle in terms of probabilities are the eight equations

$$p_{--++}p_{----} - p_{--+-}p_{---+} = 0 \tag{5}$$
$$p_{-+++}p_{-+--} - p_{-++-}p_{-+-+} = 0 \tag{6}$$
$$p_{+-++}p_{+---} - p_{+-+-}p_{+--+} = 0 \tag{7}$$
$$p_{++++}p_{++--} - p_{+++-}p_{++-+} = 0 \tag{8}$$
$$p_{++--}p_{----} - p_{+---}p_{-+--} = 0 \tag{9}$$
$$p_{+++-}p_{--+-} - p_{+-+-}p_{-++-} = 0 \tag{10}$$
$$p_{++-+}p_{---+} - p_{+--+}p_{-+-+} = 0 \tag{11}$$
$$p_{++++}p_{--++} - p_{+-++}p_{-+++} = 0. \tag{12}$$

These are the basis for the toric ideal of the 4-cycle. The probabilities can be expressed in terms of the moments and, again, each μ_{ijkl} can be expressed as a polynomial in the κ_{ijkl} using Lemma 3 or Lemma 7.

Some ingenuity is required. We proceed in the simpler case when all the means are zero: $\kappa_{1000} = \kappa_{0100} = \kappa_{0010} = \kappa_{0001} = 0$.

First, note that if we substitute for the κ_{ijkl} in the above equations we have

$$(5) - (6) - (7) + (8) = \frac{1}{16}\kappa_{1111} + \frac{1}{8}\kappa_{1100}\kappa_{0011}$$

which immediately gives

$$\kappa_{1111} = -2\kappa_{1100}\kappa_{0011}. \tag{13}$$

Similarly the expressions (5) + (6) − (7) + (8), (5) − (6) + (7) − (8), (9) + (10) − (11) − (12) and (9) − (10) + (11) − (12) lead to the matrix equations:

$$\begin{bmatrix} 1 & \kappa_{1100} & -\kappa_{1010} & -\kappa_{1001} \\ \kappa_{1100} & 1 & -\kappa_{0110} & -\kappa_{0101} \\ -\kappa_{1010} & -\kappa_{0110} & 1 & \kappa_{0011} \\ -\kappa_{1001} & -\kappa_{0101} & \kappa_{0101} & 1 \end{bmatrix} \begin{bmatrix} \kappa_{1110} \\ \kappa_{1101} \\ \kappa_{1001} \\ \kappa_{0111} \end{bmatrix} = 0.$$

The above matrix is, except for sign changes, the covariance matrix, $\tilde{C}$, of (X_1, X_2, X_3, X_4) and if we assume it is non-singular (equivalent to all the covariances being in the open interval $(-1, 1)$) then

$$\kappa_{1110} = \kappa_{1101} = \kappa_{1011} = \kappa_{0111}. \tag{14}$$

Substituting (13) and (14) back into the cumulant version of (5)–(12) we obtain two special equations:

$$\begin{aligned} &\kappa_{0011}^2\kappa_{1100} - \kappa_{1100} - \kappa_{1010}\kappa_{0101}\kappa_{0011} \\ &- \kappa_{1001}\kappa_{0110}\kappa_{0011} + \kappa_{1001}\kappa_{0101} + \kappa_{1010}\kappa_{0110} = 0 \\ &\kappa_{0011}\kappa_{1100}^2 - \kappa_{0011} - \kappa_{1010}\kappa_{0101}\kappa_{1100} \\ &- \kappa_{1001}\kappa_{0110}\kappa_{1100} + \kappa_{1010}\kappa_{1001} + \kappa_{0110}\kappa_{0101} = 0. \end{aligned}$$

The above expressions are respectively $\det(\tilde{C})\tilde{C}_{12}^{-1}$ and $\det(\tilde{C})\tilde{C}_{34}^{-1}$, showing that under the conditions of zero mean and non-singular covariance a zero in the $(1, 2)$ and $(3, 4)$ entries of the influence matrix is a necessary condition for the 4-cycle. But it is not sufficient as it would be in the Gaussian case because of the additional conditions (13) and (14).

4. Conclusion

The paper has demonstrated two main ideas: that for discrete distribution only a restricted finite set of cumulants need to be use, as for moments, in any algebraic statistical theory and that one can express any algebraic condition on raw probabilities, moments or cumulants in equivalent forms. Cumulants are particularly useful for expressing independence conditions and the examples show that this goes over quite well to conditional independence. We should note again that the set of moments and cumulants used depends on the support D and the monomial order τ used in the interpolation over the support using the G-basis method. The examples also show that conditions for conditional independence based on the zero entries of the inverse covariance matrix, which is standard in the Gaussian theory, can arise as necessary conditions in the binary case, and that the additional conditions needed are conveniently expressed in terms of cumulants.

References

Barndorf-Nielsen, O., Cox, D., 1989. Asymptotic Techniques for Use in Statistics. Chapman and Hall, London.

Diaconis, P., Sturmfels, B., 1998. Algorithms for sampling from a conditional distribution. Ann. Statist. 26, 363–397.

Geiger, D., Meek, C., Sturmfels, B., 2004. On the toric algebra of graphical models. Ann. Statist. (in press).

Jensen, J.L., 1995. Saddlepoint Approximations. Clarendon Press, Oxford.

Lauritzen, S.L., 1996. Graphical Models. Clarendon Press, Oxford.

Letac, G., Mora, M., 1990. Natural real exponential families with quadratic variance functions. Ann. Statist. 18, 1–37.

McCullagh, P., 1988. Tensor Methods in Statistics. Chapman and Hall, London.

Pistone, G., Riccomagno, E., Wynn, H., 2001a. Algebraic Statistics. Chapman and Hall/CRC, Boca Raton.

Pistone, G., Riccomagno, E., Wynn, H., 2001b. Computational commutative algebra in discrete statistics. Contemp. Math. 287, 267–282.

Pistone, G., Wynn, H., 1996. Generalise confounding with Gröbner bases. Biometrika 83, 653–666.

Pistone, G., Wynn, H., 1999. Finitely generated cumulants. Statist. Sinica 9, 1029–1052.

Sturmfels, B., 2002. Solving Systems of Polynomial Equations. In: CBMS Regional Conference Series, vol. 97. American Mathematical Society, Providence.