Differential cumulants, hierarchical models and monomial ideals

Daniel Bruynooghe,

Department of Statistics, London School of Economics, London, UK

d.hawellek@lse.ac.uk

Henry P. Wynn

Department of Statistics, London School of Economics, London, UK

h.wynn@lse.ac.uk

February 11, 2011

Abstract

For a joint probability density function $f_X(x)$ of a random vector X the mixed partial derivatives of $\log f_X(x)$ can be interpreted as limiting cumulants in an infinitesimally small open neighborhood around x. Moreover, setting them to zero everywhere gives independence and conditional independence conditions. The latter conditions can be mapped, using an algebraic differential duality, into monomial ideal conditions. This provides an isomorphism between hierarchical models and monomial ideals. It is thus shown that certain monomial ideals are associated with particular classes of hierarchical models.

Keywords: Differential cumulants, conditional independence, hierarchical models, monomial ideals.

1 Introduction

This paper draws together three areas: a new concept of differential cumulants, hierarchical models and the theory of monomial ideals in algebra. The central idea is that for a strictly positive density $f_X(x)$ of a p-dimensional random vector

1

X, the mixed partial derivative of the log density $g_X(x) = \log f_X(x)$ can be used to express independence and conditional independence statements. Thus, for random variables X_1, X_2, X_3 in $\mathbb{R}$, the condition

$$\frac{\partial^2}{\partial x_1 \partial x_2} g_{X_1, X_2, X_3}(x_1, x_2, x_3) = 0 \text{ for all } (x_1, x_2, x_3) \text{ in } \mathbb{R}^3 \qquad (1)$$

is equivalent to the conditional independence statement

$$X_1 \perp\!\!\!\perp X_2 | X_3.$$

In the next section we show how such mixed partial derivatives can be interpreted as differential cumulants. Then, in section 3, we show how collections of differential equations like (1) can be used to express independence and conditional independence models. Section 4 shows that, more generally, these collections can be used to define hierarchical statistical models of exponential form.

Section 5 maps the hierarchical model conditions to monomial ideals, which are increasingly being used within algebraic statistics. This isomorphism maps, for example, the mixed partial derivative condition (1) to the monomial ideal $< x_1 x_2 >$ within the polynomial ring $k[x_1, x_2, x_3]$. The equivalence allows ideal properties to be interpreted as hierarchical model properties, opening up an algebraic-statistical interface with some potential.

2   Local and differential cumulants

This section can be considered as a development from a body of work on local correlation. Good examples are the papers of Holland & Wang (1987), Jones (1996) and Bairamov et al. (2003). We draw particularly on Mueller & Yan (2001).

Let $X \in \mathbb{R}^p$ be a random vector. We assume X has a $p + 1$ times continuously differentiable density f_X. Once we introduce the concept of differential cumulants, we further require f_X be strictly positive.

For x, k in $\mathbb{R}^p$ we set $x^k := \prod_{i=1}^p x_i^{k_i}$, $x! := \prod_{i=1}^p x_i!$ and $m_k = \mathbb{E}(X^k)$. Let $M_X : \mathbb{R}^p \longrightarrow \mathbb{R}$ and $K_X : \mathbb{R}^p \longrightarrow \mathbb{R}$ denote the moment and cumulant generating functions of X respectively. For a vector k in $\mathbb{N}^p$ we set

$$D^k f(x) := \frac{\partial^{\|k\|_1}}{\prod_{i=1}^p \partial x_i^{k_i}} f(x),$$

where $\|k\|_1 := \sum_{i=1}^p |k_i|$ is the Manhattan norm. By convention $D^0 f(x) := f(x)$.

The cumulant κ_k can be found by evaluating $D^k(\log(M_X(t))$ at zero. We use the multivariate chain rule (given e.g. in Hardy, 2006) stated in Theorem 1. At the heart of the chain rule is an identification of differential operators with multisets:

Definition 1 (Multiset, multiplicity, size). A multiset M is a set which may hold multiple copies of its elements. The number of occurrences of an element is its multiplicity. The multiplicity of a multiset is the vector of multiplicities of its elements, denoted by ν_M. The total number of elements $|M|$ in M is the size. A multiset which is a set is called degenerate.

Example 1 (Partial derivative and multiset). The partial derivative $\frac{\partial^3}{\partial x_1 \partial x_3^3} f(x)$ has associated multiset $\{1, 3, 3, 3\}$ with multiplicity $(1, 0, 3)$ and size four.

Definition 2 (Partition of a multiset). Let I be some index set and $(M_i)_{i \text{ in } I}$ be a family of multisets with associated family of multiplicities $(\nu_{M_i})_{i \text{ in } I}$. A partition π of a multiset M is a multiset of multisets $\{(M_i)_{i \in I}\}$ such that $\nu_M = \sum_{i \in I} \nu_{M_i}$. Being a multiset itself, a partition can hold multiple copies of one or more multisets.

Example 2 (Partition of a multiset). The multiset $\{\{x_1, x_3\}, \{x_1, x_3\}, \{x_3\}\}$ is a partition of $\{x_1, x_1, x_3, x_3, x_3\}$, since $(1, 0, 1) + (1, 0, 1) + (0, 0, 1) = (2, 0, 3)$. In the following, we will use the shorthand $\{x_1 x_3 | x_1 x_3 | x_3\}$.

Associated with a partition π of a multiset M is a combinatorial quantity to which we refer as the collapse number $c(\pi)$. It is defined as

$$c(\pi) := \frac{\nu_M!}{\prod_{i \in I} \nu_{M_i}! \nu_\pi!}.$$

See Hardy (2006) for a combinatorial interpretation of $c(\pi)$.

Theorem 1 (Higher order derivative of chain functions).

$$D^k g(h(x)) = \sum_{\pi \in \Pi(k)} c(\pi) D^{|\pi|} g(h) \prod_{j=1}^{|\pi|} D^{\nu_{M_j}} h(x),$$

where $\Pi(k)$ is the set of all partitions of a multiset with multiplicity k and M_j is the j-th multiset in the partition π.

Proof. See Hardy (2006). $\square$

3

Corollary 1 (Cumulants as functions of moments). *Let κ_k be the k-th cumulant. Then*

$$\kappa_k = \sum_{\pi \in \Pi(k)} c(\pi)(-1)^{(|\pi|-1)}(|\pi|-1)! \prod_{j=1}^{p} m_{\nu_{M_j}}. \tag{2}$$

Proof. Set $g(h) = \log(h)$, $\quad h(t) = M_X(t)$ and evaluate at $t = 0$. $\qquad\square$

Example 3 (Partial derivative). Consider the partial derivative $\frac{\partial^3}{\partial x \partial z^2} g(h(x, y, z))$. The associated multiset is $\{1, 3, 3\}$ with partitions $\{133\}, \{13|3\}, \{1|33\}, \{1|3|3\}$. The multivariate chain rule tells us that

$$\begin{aligned}
D^{102} g(h(x, y, z)) = {} & Dg D^{102} h \\
& + 2D^2 g D^{101} h D^{001} h \\
& + D^2 g D^{100} h D^{002} h \\
& + D^3 g D^{100} h (D^{001} h)^2,
\end{aligned}$$

where function arguments have been suppressed on the right hand side for better readability. In particular we may conclude that

$$\kappa_{102} = m_{102} - 2m_{101} m_{001} - m_{100} m_{002} + m_{100} m_{001}^2.$$

The expression for cumulants in terms of moments is particularly simple in what we shall call the square-free case, that is for cumulants κ_k, whose index vector k is binary. In that case, the multiset associated with k is degenerate and $c(\pi) = 1$. Equation (2) simplifies to

$$\kappa_k = \sum_{\pi \in \Pi(k)} (-1)^{(|\pi|-1)}(|\pi|-1)! \prod_{j=1}^{p} m_{\nu_{M_j}}.$$

In this form it is often stated and derived via the classical Faa Di Bruno formula applied to an exponential function followed by a Moebius inversion (see e.g. Barndorff-Nielsen & Cox, 1989).

Local analogues to moments and cumulants can be derived as one considers their limiting counterparts in the neighborhood of a fixed point ξ in $\mathbb{R}^p$, an idea proposed by Mueller & Yan (2001). This section derives formulae for local moments and cumulants and local moment generating functions provided its global counterpart exists.

For a strictly positive edge length ϵ in $\mathbb{R}_+$, let $A(a, \epsilon) := [\xi - \frac{\epsilon}{2}, \xi + \frac{\epsilon}{2}]^p$ denote the hyper cube centralized at ξ. Let $|A| = \epsilon^p$ denote its volume. The density of the random variable X in $\mathbb{R}^p$ conditional on being in A is given by

$$f_X^A(x) = \frac{f_X(x)\mathbb{1}_A(x)}{pr(X \in A)},$$

where $\mathbb{1}_A(x)$ is the indicator function which returns unity if x is in A and zero otherwise. The conditional moments about ξ are denoted by

$$m_k^A = \mathbb{E}\left(\prod_{i=1}^p (X_i - \xi_i)^{k_i} \,\middle|\, X \in A \right).$$

Let $2\mathbb{N}$ and $2\mathbb{N}+1$ denote the set of positive even and odd integers respectively. For symmetry reasons, even and odd orders of individual components have different effects on local moments, which motivates the following definition:

$$\|\alpha\|_1^+ := \|\alpha\|_1 + \sum_{i=1}^p \mathbb{1}\left(\alpha_i \in 2\mathbb{N}+1\right).$$

$\|\cdot\|_1^+$ increments the total sum of the components of a vector by one additional unit for each odd component (it is not to be interpreted as a norm).

Theorem 2 (Local moments). *Let X in $\mathbb{R}^p$ be an absolutely continuous random vector with density f_X which is p times differentiable in ξ in $\mathbb{R}^p$. Let k in $\mathbb{N}^p$ determine the order of moment. Then, for $|A|$ sufficiently small, X has local moment*

$$m_k^A = r(\epsilon, k)\left(\frac{D^\alpha f_X(\xi)}{f_X(\xi)} + O(\epsilon^2) \right), \tag{3}$$

where $r(\epsilon, k) := \epsilon^{\|k\|_1^+} \prod_{\substack{i=1, \\ k_i \in 2\mathbb{N}}}^p \frac{1}{k_i+1} \prod_{\substack{i=1, \\ k_i \in 2\mathbb{N}+1}}^p \frac{1}{k_i+2}$ *and* $\alpha := \sum_{\substack{i=1, \\ k_i \in 2\mathbb{N}+1}}^p e_i$.

Proof. Consider

$$m_k^A = \frac{\int_A \prod_{i=1}^p (x_i - \xi_i)^{k_i} f_X(x)dx}{\int_A f_X(x)dx} \tag{4}$$

Approximate f_X through its p-th order Taylor expansion, integrate (4) term by term and exploit the point symmetry of odd order terms about the origin. □

Example 4 (Local moment m_{120}). Consider a tri-variate random variable X with local moment $m_{120}^A = E((X_1 - \xi_1)(X_2 - \xi_2)^2 | X \in A)$. Then $r(\epsilon, k) = \frac{\epsilon^4}{9}, \alpha := (1, 0, 0)'$ and we obtain

$$m_{120}^A = \frac{\epsilon^4}{9} \frac{\partial f(x_1, x_2, x_3)}{\partial x_1} + O(\epsilon^6).$$

A natural way to extend the concept of a local moment is to consider the limiting case when $\epsilon \to 0$. This leads to our definition of differential moments.

Definition 3 (Differential moment). The differential moment of an absolutely continuous random vector X in $\mathbb{R}^p$ in ξ in $\mathbb{R}^p$ is defined as:

$$m_k^\xi := \lim_{\epsilon \to 0} \frac{m_k^A}{r(\epsilon, k)}.$$

Corollary 2 (Differential moment). *For a differential moment of order k in $\mathbb{N}^p$ in ξ in $\mathbb{R}^p$ it holds that*

$$m_k^\xi = \frac{D^\alpha f_X(\xi)}{f_X(\xi)}.$$

Proof. This follows from Theorem 2 upon taking the limit as $\epsilon \to 0$. $\qquad \square$

From (3) it is clear that the choice of α in the derivative $D^\alpha f_X$ depends only on the pattern of odd and even components of the moment. To be precise, α holds a unity corresponding to odd components and a zero corresponding to even component entries. Consequently, the differential moment m_k^ξ depends on k only via the pattern of odd and even values.

This suggests defining an equivalence relation on $\mathbb{N}^p \times \mathbb{N}^p$: For $u, k \in \mathbb{N}^p$ set $u \sim_m k \iff m_{u_1 \cdots u_p} = m_{k_1 \cdots k_p}$. The relation $\sim_m$ partitions the product space $\mathbb{N}^p \times \mathbb{N}^p$ into 2^p equivalence classes of same differential moments. The graph corresponding to $\sim_m$ is depicted in Figure 1 for the bivariate case. The axes give the order of the moment for the two components. Different symbols represent different equivalence classes. For instance, $(2, 2) \sim_m (4, 2)$, since $m_{22}^\xi = m_{42}^\xi$. Note that $u \sim_m k \iff \|u - k\|_1 \in 2\mathbb{N}$.

Similarly to local moments, for any measurable set A we can define a local moment generating function:
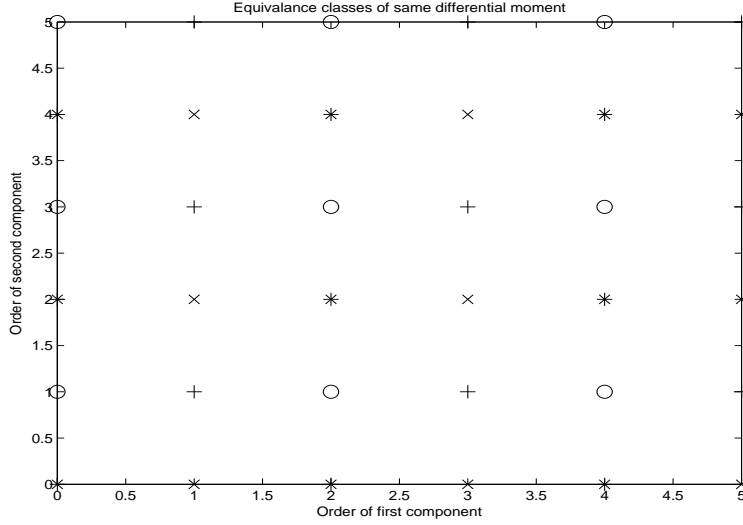
$$M_X^A(t) := \mathbb{E}(e^{t'X} | X \in A).$$

Figure 1: Graph of the equivalence classes induced by $\sim_m$ (bivariate case). Each equivalence class is depicted with a different symbol.

Being a conditional expectation, it exists if M_X exists. We have the following expansion:

$$
\begin{aligned}
M_X^A(t) &= \frac{1}{pr(X \in A)} \int_A e^{\sum_{i=1}^p t_i x_i} f_X(x) dx \\
&= 1 + \sum_{i=1}^p t_i \left. \frac{\partial f_X(x)}{\partial x_i} \right|_{x=\xi} \left(\frac{\epsilon^2}{3 f_X(\xi)} + O(\epsilon^4) \right) \\
&\quad + \sum_{i=1}^p t_i^2 \left. \frac{\partial^2 f_X(x)}{\partial x_i^2} \right|_{x=\xi} \left(\frac{\epsilon^2}{6 f_X(\xi)} + O(\epsilon^4) \right) \\
&\quad + \sum_{i=1}^p \sum_{j>i}^p t_i t_j \left. \frac{\partial^2 f_X(x)}{\partial x_i \partial x_j} \right|_{x=\xi} \left(\frac{\epsilon^4}{9 f_X(\xi)} + O(\epsilon^6) \right) + O(\epsilon^4 \|t^3\|).
\end{aligned}
$$

The local moments can be computed from the local moment generating function via differentiation to appropriate order and evaluation at $t = 0$. The natural logarithm of the local moment generating function defines the local cumulant gener-

ating function $K_X^A(t) : \mathbb{R}^p \longrightarrow \mathbb{R}$:

$$K_X^A(t) := \log(M_X^A(t)).$$

Corollary 3 (Local cumulants). *Under the conditions of Theorem 2 it holds for the local cumulants that*

$$\kappa_k^A = \sum_{\pi \in \Pi(k)} c(\pi)(-1)^{(|\pi|-1)}(|\pi|-1)! \prod_{j=1}^{|\pi|} r(\epsilon, \nu_{M_j}) \left(\frac{D^{\alpha_j} f_X(\xi)}{f_X(\xi)} + O(\epsilon^2) \right),$$

where α_j is a function of the partition π and defined as

$$\alpha_j := \sum_{i=1}^{p} e_i \mathbb{1}\left(\nu_{M_j}(i) \in 2\mathbb{N}+1 \right),$$

that is, α_j is binary and holds ones corresponding to odd elements of ν_{M_j}. Furthermore,

$$r(\epsilon, \nu_{M_j}) := \epsilon^{\left\| \nu_{M_j} \right\|_1^+} \prod_{\substack{i=1, \\ \nu_{M_j}(i) \in 2\mathbb{N}}}^{p} \frac{1}{\nu_{M_j}(i)+1} \prod_{\substack{i=1, \\ \nu_{M_j}(i) \in 2\mathbb{N}+1}}^{p} \frac{1}{\nu_{M_j}(i)+2}.$$

Proof. Combine the chain rule and Theorem 2. $\qquad\square$

Similarly to differential moments we can define differential cumulants at ξ. Two different ways of doing so are natural. First, taking the limiting quantity of the local cumulants as $\epsilon \to 0$ or, second, taking the series of differential moments and requiring that the mapping between moments and cumulants is preserved which is induced through the ex-log relation of the associated generating functions, see also the discussion in (McCullagh, 1987, page 62).

As demonstrated below, the two quantities just described differ in general and coincide only in the square-free case. In order to retain the intuitive and familiar relation between cumulants and moments, we define differential cumulants in terms of differential moments.

Definition 4 (Differential cumulant). For an index vector k in $\mathbb{N}^p$, the differential cumulant in a in $\mathbb{R}^p$ is defined as

$$\kappa_k^a := \sum_{\pi \in \Pi(k)} c(\pi)(-1)^{(|\pi|-1)}(|\pi|-1)! \prod_{i=1}^{|\pi|} m_{\nu_{M_i}}^a.$$

We are now in a position to state the main result of this section, namely that mixed partial derivatives of the log density can be interpreted as differential cumulants.

Lemma 1 (Differential cumulant). *For a differential cumulant in ξ in $\mathbb{R}^p$ of order k in $\mathbb{N}^p$ it holds that*

$$\kappa_k^\xi = D^\alpha \log(f_X(\xi)),$$

where $\alpha := \sum_{\substack{i=1, \\ k_i \in 2\mathbb{N}+1}}^p e_i$ projects odd elements of k onto one and even elements of k onto zero.

Proof. Apply the chain rule to $D^\alpha \log(f_X(\xi))$. ☐

This is a multivariate generalization of the local dependence function introduced by Holland & Wang (1987). The next theorem relates differential cumulants to the limit of local cumulants.

Theorem 3 (Differential and limiting local cumulant). *A differential cumulant κ_k^ξ equals the limit of the local cumulant $lim_{\epsilon \to 0} \frac{1}{r(\epsilon,k)} \kappa_k^A$ if and only if k is binary, i.e. κ_k is a square-free cumulant.*

Proof. First, let $k \in \{0,1\}^p$ be binary and $\pi = \{(M_j)_{1 \leq j \leq |\pi|}\}$ be a partition of the lattice corresponding to k. One can show that $r(\epsilon, k) = \prod_{j=1}^{|\pi|} r(\epsilon, \nu_{M_j})$. With that

$$\frac{1}{r(\epsilon, k)} \kappa_k^A = \sum_{\pi \in \Pi(k)} (-1)^{(|\pi|-1)} (|\pi|-1)! \prod_{\substack{j=1, \\ M_j \in \pi}}^{|\pi|} \frac{D^{\nu_{M_j}} f_X(\xi)}{f_X(\xi)} + O(\epsilon^2). \quad (5)$$

Now take limits as $\epsilon \to 0$ to obtain $lim_{\epsilon \to 0} \frac{1}{r(\epsilon,k)} \kappa_k^A = \kappa_k^\xi$.

Conversely, suppose k is not binary. Express κ_k^A as a linear combination of local moments. Consider the degenerate partition π, which holds only one multiset M with multiplicity $\nu_M = k$. The quantity associated with π converges to $c \frac{D^k f_X(\xi)}{f_X(\xi)}$ for some constant c in $\mathbb{R}$. k not being binary, this cannot be a differential moment, which are proportional to $D^\alpha f_X(\xi)$ for some binary α. Differential cumulants are linear combinations of differential moment products only. Hence κ_k^A does not converge to a differential cumulant. ☐

Of particular interest to us are differential cumulants which vanish everywhere. We refer to them as zero-cumulants. Writing $g = \log f_X$, we shall usually write $D^\alpha g = 0$ to denote the zero-cumulant associated with α in the understanding that this holds for all x.

The next section shows that sets of zero cumulants are isomorphic to conditional independence statements. As a consequence of lemma 1 zero-cumulants are invariant under diagonal transformations of the random vector X. In particular, they are not affected by the probability integral transformation and hence any result below holds also true for the copula density of X.

3 Independence and conditional independence

From now on, we shall assume that f_X is strictly positive everywhere. Sets of zero-cumulants are equivalent to conditional and unconditional dependency structures.

Proposition 1 (Independence in the bivariate case)**.** *Let X in $\mathbb{R}^2$. Then $X_1 \perp\!\!\!\perp X_2 \iff \kappa_{11}^x = 0$ for all x in $\mathbb{R}^2$.*

Proof.

$$0 = \kappa_{11}^x = \frac{\partial^2}{\partial x_1 \partial x_2} \log(f_{X_1,X_2}(x_1, x_2) \iff f_{X_1,X_2}(x_1, x_2) = e^{h_1(x_1)+h_2(x_2)}$$

for some functions $h_1, h_2 : \mathbb{R} \to \mathbb{R}$. □

In the multivariate case, we can express conditional independence of any pair given the remaining variables through square free differential cumulants.

Proposition 2 (Conditional independence of two random variables)**.** *Let X in $\mathbb{R}^p$. Then*

$$X_i \perp\!\!\!\perp X_j | X_{-ij} \iff \kappa_k^x = 0 \quad \text{for all } x \text{ in } \mathbb{R}^p,$$

where

$$X_{-ij} := (X_1, ..., X_{i-1}, X_{i+1}, ..., X_{j-1}, X_{j+1}, ..., X_p)$$

and $k = e_i + e_j$, $(i, j) \in \{1, ..., p\}^2$, $i \neq j$.

Proof. By analogy with the bivariate case. □

Setting several square-free differential cumulants to zero simultaneously allows us to express conditional independence statements.

Proposition 3 (Multivariate conditional independence). *Given three index sets* I, J, K *which partition* $\{1, ..., p\}$, *let* $S = \{e_i + e_j, i \in I, j \in J\}$. *Then*

$$X_I \perp\!\!\!\perp X_J | X_K \iff \kappa_k^x = 0 \text{ for all } k \in S \text{ and for all } x \text{ in } \mathbb{R}^p.$$

Proof. From proposition 2 it is clear, that this is equivalent to the conditional independence statement

$$X_I \perp\!\!\!\perp X_J | X_K \iff X_i \perp\!\!\!\perp X_j | X_{-ij} \quad \text{for all } (i, j) \in I \times J.$$

Sufficiency ($\Rightarrow$) and necessity ($\Leftarrow$) are semi-graphoid and graphoid axioms referred to as decomposition and intersection respectively. Both hold true for strictly positive conditional densities (see for instance Cozman & Walley, 2005). $\square$

Pairwise conditional independence of all pairs is equivalent to independence.

Theorem 4 (Pairwise conditional independence if and only if independence). *The random variables* $X_1, ..., X_n$ *are independent if and only if* $\kappa_{e_i + e_j} = 0 \quad$ *for all* $(i, j) \in \{1, ..., n\}^2, \ i \neq j$.

Proof. Sufficiency follows from differentiation of the log-density. Necessity can be proved by induction on the number of variables n. The statement is true for $n = 2$ by proposition 1. Let the statement be true for n and let the $\binom{n+1}{2}$ differential cumulants $\kappa_{e_i + e_j}$ vanish, where e_i and e_j are unit vectors in $\mathbb{R}^{n+1}$. Consider $\kappa_{e_1 + e_2} = 0$. Integration with respect to x_1 and x_2 yields

$$f_{X_1, ..., X_{n+1}}(x_1, ..., x_{n+1}) = e^{h_1(x_{-1}) + h_2(x_{-2})} \tag{6}$$

for some functions $h_1 : \mathbb{R}^n \longrightarrow \mathbb{R}$ and $h_2 : \mathbb{R}^n \longrightarrow \mathbb{R}$. Now integrate again with respect to x_1 to obtain

$$f_{X_{-1}}(x_{-1}) = e^{h_1(x_{-1})} \int_{\mathbb{R}} e^{h_2(x_{-2})} dx_1.$$

The left hand side is an n-dimensional marginal density which factorises into n marginals by induction assumption: $f_{X_{-1}}(x_{-1}) = \prod_{i=2}^{n+1} f_{X_i}(x_i)$. This allows us to conclude that $h_1(x_{-1})$ can be split into a sum of two functions, $g_1 : \mathbb{R}^{n-1} \longrightarrow \mathbb{R}$ and $g_2 : \mathbb{R} \longrightarrow \mathbb{R}$, where the latter is a function of x_2 only, i.e. $h_1(x_{-1}) = g_1(x_{-12}) + g_2(x_2)$. Considering (6) again we see that the density $f_{X_1, ..., X_{n+1}}$ factorises

$$f_{X_1, ..., X_{n+1}}(x_1, ..., x_{n+1}) = e^{g_2(x_2) + g_1(x_{-12}) + h_2(x_{-2})}.$$

Hence $X_2 \perp\!\!\!\perp X_{-2}$ and the density of X_{-2} factorises by induction assumption. $\square$

4 Hierarchical models

The analysis of the last section makes clear that setting certain mixed two-way partial derivatives of $g(x) = \log f_X(x)$ equal to zero, is equivalent to independence or conditional independence statements. We can go further and define a generalized hierarchical model using the same process.

The basic structure of a hierarchical model can be define via a simplicial complex. Thus let $\mathcal{N} = \{1, \ldots, p\}$ be the vertex set representing the random variables $X_1, ..., X_p$. A collection $\mathcal{S}$ of index sets $J \subseteq \mathcal{N}$ is a simplicial complex if it is closed under taking subsets, i.e. if J in $\mathcal{S}$ and $K \subseteq J$ then K in $\mathcal{S}$.

Definition 5. Given a simplicial complex $\mathcal{S}$ over an index set $\mathcal{N} = \{1, \ldots, p\}$ and an absolutely continuous random vector X a hierarchical model for the joint distribution function $f_X(x)$ takes the form:

$$f_X(x) = \exp \left\{ \sum_{J \text{ in } \mathcal{S}} h_J(x_J) \right\},$$

where $h_J : \mathbb{R}^J \longrightarrow \mathbb{R}$ and x_J in $\mathbb{R}^J$ is the canonical projection of x in $\mathbb{R}^p$ onto the subspace associated with the index set J.

This is equivalent to a quasi-additive model for $g(x) = \sum_{J \in \mathcal{S}} h_J(x_J)$, and we also refer to this model for $g(x)$ as being hierarchical. It is clear that we may write the model over the maximal cliques only, namely simplexes which are not contained in a larger simplex. In the terminology of Lauritzen (1996) we require f_X be positive and factorise according to $\mathcal{S}$ for it to be a hierarchical model with respect to $\mathcal{S}$.

Associated to an index set $K \subseteq \mathcal{N}$ is a differential operator D^k, where $k = \sum_{i \in K} e_i \in \{0, 1\}^p$ holds ones for every member of K and zeros otherwise. In the following, we overload the differential operator by allowing it to be superscripted by a set or by a vector. Thus, for an index set K we set $D^K := D^k$ and similarly $\kappa_K^x := \kappa_k^x$. D^K returns the differential cumulant κ_K^x, when applied to $g(x)$.

Example 5. Let $K = \{2, 4, 6\}$. We obtain $k = (0, 1, 0, 1, 0, 1)$ and

$$D^K g(x) = \kappa_K^x = \kappa_k^x = \frac{\partial^3}{\partial x_2 \partial x_4 \partial x_6} g(x).$$

We collect the results of the last section into a comprehensive statement. First, we define the complementary complex to a simplicial complex $\mathcal{S}$ on $\mathcal{N}$.

Definition 6. Given a simplicial complex $\mathcal{S}$ on an index set $\mathcal{N}$ we define the complementary complex as the collection $\bar{\mathcal{S}}$ of every index set K which is not a member of $\mathcal{S}$.

Note immediately that $\bar{\mathcal{S}}$ is closed under unions, i.e. K, K' in $\bar{\mathcal{S}} \Rightarrow K \cup K' \in \bar{\mathcal{S}}$. It is a main point of this paper that there is a duality between setting collections of mixed differential cumulants equal to zero and a general hierarchical model:

Theorem 5. *Given a simplicial complex $\mathcal{S}$ on an index set $\mathcal{N}$, a model g is hierarchical, based on $\mathcal{S}$ if and only if all differential cumulants on the complementary complex vanish everywhere, that is*

$$\kappa_K^x = 0, \ \text{for all } x \text{ in } \mathbb{R}^p \text{ and for all } K \text{ in } \bar{\mathcal{S}}.$$

Proof. First, let g be hierarchical with respect to $\mathcal{S}$, that is g is a log-density with representation $g(x) = \sum_{J \text{ in } \mathcal{S}} h_J(x_J)$. Then, for K in $\bar{\mathcal{S}}$, the associated differential operator D^K annihilates any term h_J in g, since $K \nsubseteq J$ for any J in S.

Conversely, suppose $\kappa_K^x = 0$ for all x in $\mathbb{R}^p$ and for all K in $\bar{\mathcal{S}}$. Then, by proposition 2, f_X is pairwise Markov with respect to $\mathcal{S}$ and hence factorises over maximal cliques of $\mathcal{S}$ by the Hammersley-Clifford theorem. The reader is referred to Lauritzen (1996) for a detailed discussion of factorization and Markov properties. $\square$

5 The duality with monomial ideals

The growing area of algebraic statistics makes use of computational commutative algebra particularly for discrete probability model, notably Poisson and multinomial log-linear models. Work connecting the algebraic methods to continuous probability models is sparser although considerable process has been made in the Gaussian case. For an overview see Drton et al. (2009). Our link to the algebra is via monomial ideals.

A monomial in $x, ..., x_p$ is a product of the form $x^\alpha = \prod_{j=1}^p x_j^{\alpha_j}$, where α in $\mathbb{N}^p$. A monomial ideal I is a subset of a polynomial ring $k[x_1, ..., x_p]$ such that any $m \in I$ can be written as a finite polynomial combination $m = \sum_{k \in K} h_k x^{\alpha_k}$, where $h_k \in k[x_1, ..., x_p]$ and $\alpha_k \in \mathbb{N}^p$ for all $k \in K$. We write $I = < x^{\alpha_1}, ..., x^{\alpha_K} >$ to express that I is generated by the family of monomials $(x^{\alpha_k})_{k \in K}$.

The full set M of monomials contained in monomial ideal I has the hierarchical structure:

$$x^\alpha \in M \Rightarrow x^{\alpha+\gamma} \in M, \tag{7}$$

13

for any index set $\gamma \in \mathbb{N}^p$. A monomial ideal is square-free if its generators $(x^{\alpha_k})_{1 \le k \le K}$ are square free, i.e. $\alpha_k \in \{0,1\}^p$ for all $1 \le k \le K$.

The following discussion shows that there is complete duality between the structure of square-free monomial ideals and hierarchical models. Associated with a simplicial complex $\mathcal{S}$ is its *Stanley-Reisner ideal* $I_\mathcal{S}$. This is the ideal generated by all square-free monomial in the complementary complex $\bar{\mathcal{S}}$. For a face $K \in \bar{\mathcal{S}}$ let $m_K(x) := \prod_{k \in K} x_k$ denote the associated square-free monomial. Then

$$I_\mathcal{S} = < (m_K)_{K \in \bar{\mathcal{S}}} > .$$

The second step, which is a main point of the paper, is to associate the differential operator D^K with the monomial $m_K(x)$. We need only confirm that the hierarchical structure implied by (7) is consistent with differential conditions of Theorem 5.

Without loss of generality include all differential operators which are obtained by continued differentiation. Then, (7) is mapped exactly to

$$D^\alpha g(x) = 0, \text{ for all } x \in \mathbb{R}^p \Rightarrow D^{\alpha+\gamma} g(x) = 0, \text{ for all } x \in \mathbb{R}^p$$

simply by continued differentiation. This bijective mapping from monomial ideals into differential operators, is sometimes referred to as a "polarity" and within differential ideal theory has its origins in "Seidenberg's differential nullstellensatz" (Seidenberg, 1956). It allows us to map properties of hierarchical models in statistics to monomial ideal properties and vice versa.

One of the main conditions discussed in the theory of hierarchical models in statistics is the decomposability of a joint density function into a product of certain marginal probabilities. Simple conditional probability is a canonical case. Thus with $p = 3$ the conditional independence $X_1 \perp\!\!\!\perp X_2 | X_3$ is represented by the graph $1 - 3 - 2$. In this case the graph has the model simplicial complex: $\mathcal{S} = \{13, 23\}$, where, again, we write $\mathcal{S}$ in terms of its maximal cliques. The Stanley-Reisner ideal is $I_\mathcal{S} = < x_1 x_2 >$.

There is a factorization:

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3) = \frac{f_{X_1,X_3}(x_1, x_3) f_{X_2,X_3}(x_2, x_3)}{f_{X_3}(x_3)}.$$

Decomposable graphical models, discussed below, are a generalization of this simple case. There are other cases, however, where one or more factorizations are associated with the same simplicial complex. An example is the 4-cycle: $\mathcal{S} = \{12, 23, 34, 41\}$ with The Stanley-Reisner ideal $I_\mathcal{S} = < x_1 x_3, x_2 x_4 >$. Although

14

this ideal is rather simple from an algebraic point of view the 4-cycle from a statistical point of view is rather complex. By considering special ideals we obtain general classes of models, in a subsection 5.2.

Another issue is that the structure of $\mathcal{S}$ may suggest factorizations even when they are problematical. Perhaps the first such case is the 3-cycle: $\mathcal{S} = \{12, 13, 23\}$. The Stanley-Reisner ideal is $I_\mathcal{S} = \langle x_1 x_2 x_3 \rangle$. The maximal clique log-density representation has no three-way interaction:

$$g(x_1, x_2, x_3) = h_{12}(x_1, x_2) + h_{13}(x_1, x_3) + h_{14}(x_1, x_4).$$

This might suggest the factorization

$$f_{X_1, X_2, X_3}(x_1, x_2, x_3) = \frac{f_{X_1, X_2}(x_1, x_2) f_{X_1, X_3}(x_1, x_3) f_{X_2, X_3}(x_2, x_3)}{f_{X_1}(x_1) f_{X_2}(x_2) f_{X_3}(x_3)} \qquad (8)$$

A factorization of this kind is the continuous analogue to a perfect three-dimensional table in the discrete case (Darroch, 1962). However, except when X_1, X_2, X_3 are independent we have not been able to provide a standard density for which (8) holds.

5.1 Decomposability and marginality

Our use of the index set notation makes its straightforward to define decomposability.

Definition 7. Let $\mathcal{N} = \{1, \ldots, p\}$ be the vertex set of a graph $\mathcal{G}$ and I, J vertex sets such that $I \cup J = \mathcal{N}$. Then $\mathcal{G}$ is decomposable if and only if $I \cap J$ is complete and I forms a maximal clique or the subgraph based on I is decomposable and similarly for J.

Under this condition the corresponding hierarchical model has a factorization

$$f_V(x_V) = \frac{\prod_{J \in C} f_J(x_J)}{\prod_{K \in S} f_K(X_K)},$$

where the numerator on the right hand side corresponds to cliques and the denominator to separators which arise in the continued factorization under the definition.

It is important to realize that in order to proceed with the factorization at each stage a marginalisation step is required. Consider the simple case based on the
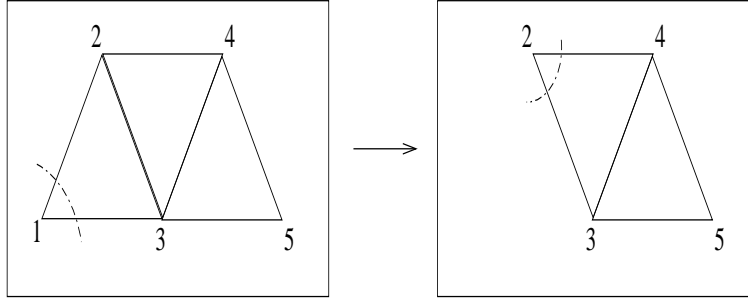
Figure 2: Factorization and marginalisation of a hierarchical model.

simplicial complex $\mathcal{S} = \{123, 234, 345\}$. One choice of factorization at first stage is (with simplified notation):

$$f_{12345} = \frac{f_{123} f_{2345}}{f_{23}}$$

and we continue the factorization to give

$$f_{12345} = \frac{f_{123} f_{234} f_{345}}{f_{23} f_{34}}.$$

The process of marginalisation is shown in Figure 2. At any stage, we may choose to marginalise with respect to any variable that is member of just a single clique. In the first step these are X_1 and X_5 and suppose we chose to single out X_1. Once f_X has been integrated with respect to x_1, the marginal model for $X_2, ..., X_5$ is obtained. The removal of a the clique 123 leads to X_2 being exposed and we may continue with X_2 or X_5 etc.

The Stanley-Reisner ideal $I_\mathcal{S} = < x_1 x_4, x_1 x_5, x_2 x_5 >$ is an ideal in $k[x_1, x_2, x_3, x_4, x_5]$. The factorization of f_{2345} is, however, mapped into the monomial ideal $< x_2 x_5 >$ which is an ideal in $k[x_2, x_3, x_4, x_5]$. A marginalisation has allowed us to drop from five dimensions to four. This is clear from the exponential expression of the model:

$$f_{12345} = \exp \left\{ h_{123}(x_1, x_2, x_3) + h_{234}(x_2, x_3, x_4) + h_{345}(x_3, x_4, x_5) \right\}.$$

Integrating with respect to x_1 we obtain a hierarchical model for the marginal joint distribution of (X_2, X_3, X_4, X_5). This marginalisation is possible because x_1 appears only in the single clique $\{1, 2, 3\}$.

16

We have exposed an interesting relationship between the statistical and algebra formulation: in order to reduce the dimensionality and obtain the Stanley-Reisner ideal for a reduced set of variables, we must first perform a marginalisation, which is a non-algebraic operation, at least, not in general a finite dimensional operation. We capture this in the following Lemma:

Lemma 2. *Whenever a simplicial complex of hierarchical model has a subset of vertexes which form a facet of a unique maximal clique (simplex) then the marginal model obtained by deleting this facet (and its connections) is valid. Moreover the monomial ideal representation is obtained by deleting any generators containing the corresponding variables and is in the ring without these variables.*

Proof. This follows the lines of the example. If J is the subset of vertexes and K, with $J \subset K$, is the unique maximal clique, then in the exponential expression for the density there will be a unique term $\exp(h_K(x_K))$ in which x_J appears. Integrating with respect to x_J to obtain the marginal distribution for $X_{V \setminus J}$ gives the reduced model. The monomial ideal representation follows accordingly. $\square$

5.2 Artinian closure and polynomial exponential models

The terms $h_J(x_J)$ which appear in the hierarchical models have not been given any special form. In fact it is a main point of this paper that this is not required to give the monomial ideal equivalence. We note, again, that we always use square-free monomial ideals.

Certain classes of hierarchical models can, however, be obtained by imposing further differential conditions. The following lemma shows that the log-density is polynomial if we impose univariate derivative restriction.

Lemma 3. *If in addition to the differential conditions in Theorem 5 we impose conditions of the form*

$$\frac{\partial^{n_i}}{\partial x_i^{n_i}} g(x) = 0, \ \ for \ all \ 1 \leq i \leq p \ and \ n \in \mathbb{N}^p \tag{9}$$

the h-functions in the corresponding hierarchical model are polynomials, in which the degree of x_i does not exceed $n_i - 1$, for all $1 \leq i \leq p$.

Proof. Repeated integration with respect to x_i shows that g is indeed a polynomial in x_i of degree less than n_i, when the other variable are fixed. Since this holds for all $1 \leq i \leq p$ the result follows. $\square$

The simultaneous inclusions of derivative operators with respect to one indeterminate in (9) constitutes an *Artinian closure* of the differential version of the Stanley-Reisner ideal $I_{\mathcal{S}}$.

Example 6 (BEC density). Suppose X is bivariate and we impose the symmetric Artinian closure conditions

$$\frac{\partial^2}{\partial x_i^2} g(x_1, x_2) = 0, \quad \text{for } i = 1, 2.$$

Then integration yields

$$g(x_1, x_2) = x_1 h_1(x_2) + h_2(x_2)$$

and

$$g(x_1, x_2) = x_2 h_3(x_1) + h_4(x_1).$$

A comparison of these functionals identifies $h_1(x_2) = a_3 x_2 + a_1$, $h_2(x_2) = a_0 + a_2 x_2$, $h_3(x_1) = a_3 x_1 + a_2$, $h_4(x_1) = a_1 x + a_0$, for some $a_i \in \mathbb{R}$ for all $1 \leq i \leq 4$, so that $g(x_1, x_2)$ can be written as

$$g(x_1, x_2) = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1 x_2. \tag{10}$$

It can be shown that X_1 is distributed exponentially conditional on $X_2 = x_2$ for all $x_2 > 0$ and vice versa. A distributions with that property is called bivariate exponential conditionals (BEC) distribution. BEC distributions are completely described by g in the sense that any BEC density is of the form (10) (Arnold & Strauss, 1988). In particular, the independence case is included, if we force $a_3 = 0$ by imposing the additional restriction

$$\frac{\partial^2}{\partial x_1 x_2} g(x_1, x_2) = 0.$$

This confirms Proposition 1 for this particular example.

The previous example extends readily into higher dimension. We call a distribution multivariate exponential conditionals (MEC) distribution, if X_j is distributed exponentially conditional on $X_i = x_i$ for all $1 \leq i, j \leq p, i \neq j$. We capture the extension to the p-dimensional case in the following lemma:

18

Lemma 4 (MEC distributions and Artinian closure). *The following statements are equivalent:*

1. *A distribution belongs to the class of MEC distributions*

2. *g is multi-linear, i.e there exist 2^p indices $a_s \in \mathbb{R}$ such that $g = \sum_{s \in \zeta} a_s x^s$, where $\zeta = \{0,1\}^p$ denotes the set of p dimensional binary vectors*

3. *$\frac{\partial^2}{\partial x_i^2} g(x) = 0$, for all $1 \leq i \leq p$.*

Proof. For a proof of $(1) \iff (2)$ see Arnold & Strauss (1988). The proof of $(2) \iff (3)$ follows the lines of the example. $\square$

Another case of considerable importance is the Gaussian distribution. Here

$$g(x) = \sum_{K \in \mathcal{S}} h_K(X_K),$$

and the maximal cliques are of degree two. The latter condition is partly obtained with an Artinian closure with $n_i = 3, \ i = 1, \ldots, p$. However, more is required. We can guess, from the fact that for a normal distribution all (ordinary) cumulants of degree three and above are zero, that if we impose all degree-three differential cumulant to be zero we obtain polynomial terms of maximum degree 2. This is, in fact the correct set of conditions to make the models terms of degree at most two. In the α-notation the conditions are

$$D^\alpha g = 0, \text{ for all } \alpha \in \mathbb{N}^p \text{ with } \|\alpha\|_1 = 3,$$

which includes the Artinian closure conditions. The corresponding ideal is generated by all polynomials of degree three. For a non-singular multivariate Gaussian, we, of course, require non-negative definiteness of the degree-two part of the model, considered as a quadratic form.

The hierarchical model is given by additional restrictions which are equivalent to removing certain terms of the form $x_i x_j, \ i \neq j$. This is the same as setting the corresponding $\{ij\}$-th entry in the inverse covariance matrix (influence matrix) equal to zero. The removed $x_i x_j$ generate the Stanley-Reisner ideal so that the zero structure of the influence matrix completely determines the ideal.

5.3 Ideal-generated models

The duality between monomial ideals and hierarchical models encourages the investigation of the properties of hierarchical models for different types of ideals. There are some important properties and features of monomial ideals which may be linked to the corresponding hierarchical models and we mention just a few here in an attempt to introduce a larger research programme.

We begin with the sub-class of decomposable models. It is well know from the statistical literature (see Lauritzen, 1996) that the decomposability property of the model based on a simplicial complex $\mathcal{S}$ is equivalent to the chordal property: there is no chord-less 4-cycle. Remarkably, the latter is equivalent to a property of the Stanley-Reisner ideal $I_{\mathcal{S}}$, namely: that the minimal free resolution of $I_{\mathcal{S}}$ be linear(see below for a brief explanation). This is a result of Fröberg (1988), see also Dochtermann & Engström (2009). Petrovic & Stokes (2010) adapt a result of Geiger et al. (2006) to show that $I_{\mathcal{S}}$, in this case, is generated in degree 2, that is all its generators have degree 2.

Theorem 6. *A decomposable graphical model $I_{\mathcal{S}}$ has a "2-linear" resolution.*

The term linear refers to the structure of the minimal free resolution of $I_{\mathcal{S}}$. In this resolution there are monomial maps between the stages of the resolution sequence. Linear means that these maps are linear. As a simple example consider again the simplicial complex $\mathcal{S} = \{123, 234, 345\}$ with Stanley-Reisner ideal $I_{\mathcal{S}} = <x_1 x_4, x_1 x_5, x_2 x_5 >$ The minimal free resolution of $I_{\mathcal{S}}$ is given by:

$$[x_1 x_4, x_1 x_5, x_2 x_5] \xrightarrow{\begin{bmatrix} -x_5 & 0 \\ x_4 & -x_2 \\ 0 & x_1 \end{bmatrix}} 0,$$

and one sees that the map is linear. By contrast, the 4-cycle is generated in degree 2, but is not linear:

$$[x_1 x_3, x_2 x_4] \xrightarrow{\begin{bmatrix} x_2 x_4 \\ -x_1 x_3 \end{bmatrix}} 0,$$

giving a non-linear map.

A special case of 2-linear resolutions are Ferrer ideals. A Ferrer ideal $I_{\mathcal{S}}$ is one in which the degree-two linear generators can be placed in a table with an inverse stair-case. Such staircases arose historically in the study of integer partitions. As an example take the Stanley-Reisner ideal

$$I_{\mathcal{S}} = < x_1 x_6, x_1 x_7, x_1 x_8, x_2 x_6, x_2 x_7, x_3 x_6, x_3 x_7, x_4 x_6, x_5 x_6 > \subseteq k[x_1, ..., x_9].$$

The Ferrer table is:

	6	7	8	9
1	x_1x_6	x_1x_7	x_1x_8	
2	x_2x_6	x_2x_7		
3	x_3x_6	x_3x_7		
4	x_4x_6			
5	x_5x_6			

Considering the non-empty cells as given by edges this corresponds to a special type of bi-partite graph between nodes $\{1, 2, 3, 4, 5\}$ and $\{6, 7, 8, 9\}$. Corso & Nagel (2009) show that, among the class of bi-partite graphs, Ferrer ideals are indeed uniquely characterized as having a 2-linear minimal free resolution.

It is straightforward to show that the corresponding hierarchical model is decomposable by exhibiting the decomposition given by Lemma 2. First take two simplices based on the variables defining, respectively, the rows and columns. In the example these are $J_1 = 12345$, $J_2 = 6789$. Then join all nodes corresponding to the complement of the Ferrer diagram to give:

	6	7	8	9
1				x_1x_9
2			x_2x_8	x_2x_9
3			x_3x_8	x_3x_9
4		x_4x_7	x_4x_8	x_4x_9
5		x_5x_7	x_5x_8	x_5x_9

The maximal cliques are easily seen to be given by a simple rule on this complementary table. For each non-empty row take the variable which defines that row together with every other variables for nonempty columns in that row *and* all the variables for the rows below that row. In this example we find, working down the rows, that the maximal cliques are:

$$123459, 234589, 34589, 45789, 5789.$$

Note how to this example we can apply Lemma 2, by successively stripping off variables in the order: x_1, x_2, x_3, x_4, x_5. The separators are $23459, 34589, 4589, 5789, 789$. The rule provides a proof of the following.

Theorem 7. *Hierarchical models generated by Ferrer ideals are decomposable.*

As another illustration of the duality between monomial ideals and conditional independence structures, we next consider two terminal networks. In Sáenz de Cabezón & Wynn
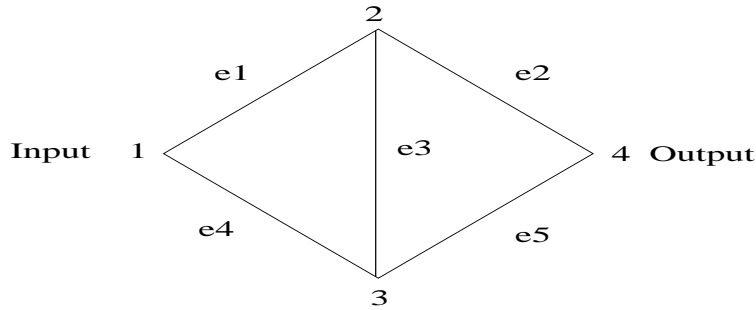
Figure 3: A two-terminal network.

(2010) the authors apply the theory and construction of minimal free resolutions to the theory of reliability. One sub-class of these is to networks, in the classical sense of network reliability. Consider a connected graph $G = (E, V)$, with two identified nodes called *input* and *output*, respectively. A *cut* is a set of edges, which if removed from the graph disconnects input and output. A *path* is connected set of edges from input to output. A minimal cut is a cut for which no proper subset is a cut and minimal path is a path for which no proper subset is a path.

As a simple example consider the network depicted in Figure 3 with input $= 1$ and output $= 4$ and edges:

$$e_1 = 1 - 2, \ e_2 = 2 - 4, \ e_3 = 2 - 3, \ e_4 = 1 - 3, \ e_5 = 3 - 4.$$

The minimal cuts are $\{e_1, e_4\}$, $\{e_2, e_5\}$, $\{e_1, e_3, e_5\}$, $\{e_2, e_3, e_4\}$. If we associate a variable x_i with each edge e_i then the minimal cuts generate an ideal. In this example we write

$$I_{\mathcal{S}} = < x_1 x_4, x_2 x_5, x_1 x_3 x_5, x_2 x_3 x_4 > .$$

The maximal cliques of $\mathcal{S}$ for the corresponding model simplicial complex $\mathcal{S}$ are

$$\{15, 24, 123, 345\}$$

We could, on the other hand define $I_{\mathcal{S}^*}$ as being the collection of all paths on the network. In this case the $I_{\mathcal{S}^*}$ is generated by the minimal paths giving:

$$< x_1 x_2, x_4 x_5, x_1 x_3 x_5, x_2 x_3 x_4 >,$$

and S^* consists of the complements of the cuts and has maximal cliques $\{15, 24, 134, 235\}$.

There is a duality between cuts and path models for two-terminal networks:

22

Lemma 5. *The model simplicial complex $\mathcal{S}$ based on the cut ideal $I_{\mathcal{S}}$ of a two terminal network is formed from the complement of all paths on the network. Conversely, the model simplicial complex $\mathcal{S}^*$ based on the path ideal $I_{\mathcal{S}^*}$, is formed from the complement of all cuts. Moreover: $(\mathcal{S}^*)^* = \mathcal{S}$.*

For example, the term 15 of $\mathcal{S}$ is the complement of the (non-minimal) path 234 and the term 14 in $\mathcal{S}^*$ is the complement the (non-minimal) cut 235.

This duality is a special example of Alexander duality and we omit the proof, see Miller & Sturmfels (2005), Proposition 1.37. The general result says that for a square-free $\mathcal{S}$, if we define $\mathcal{S}^*$ as the complement of all non-faces of $\mathcal{S}$, then $(\mathcal{S}^*)^* = \mathcal{S}$.

It will have been noticed that for this network S and S^* are self-dual in the sense that the two simplicial complexes have the same structure and only differ in the labelling of the vertexes. Both models have two separate conditional independence properties. Thus for $\mathcal{S}$ we have $X_1 \perp\!\!\!\perp X_4 | (X_2, X_3, X_5)$ and $X_2 \perp\!\!\!\perp X_5 | (X_1, X_3, X_4)$.

5.4   Geometric constructions

Simplicial complexes are at the heart of algebraic topology and it is natural to look in that field for classes of simplicial complexes whose abstract version may be used to support hierarchical models. We mention briefly one class here arising from the fast-growing area of persistent homology, see Edelsbrunner & Harer (2010). This class has already been used by Lunagómez et al. (2009) to construct graphical models using so-called Alpha complexes. We give the construction now. It is to based on the cover provided by a union of balls in R^d, a construction used by Edelsbrunner (1995) in the context of computational geometry and in Naiman & Wynn (1992) and Naiman & Wynn (1997) to study Bonferroni bounds in statistics.

Thus, let $z_1, \ldots, z_p$ be p points in R^d and define the solid balls with radius r centered at the points:

$$B_i(r) = \{z : ||x - z_i|| \leq r\}, \ i = 1, \ldots, p$$

The *nerve* of the cover represented by the union of balls is the simplicial complex $\mathcal{S}$ derived form the intersections of the balls, and is called the Alpha complex. It consists of exactly all index sets J for which $\cap_{i \in J} B_i(r) \neq \emptyset$.

When the radius, r, is small $\mathcal{S}$ consists of unconnected vertexes and the hierarchical model gives independence of the $X_1, ..., X_p$. As $r \to \infty$ there is a value

of r at and beyond which $\cap_{i=1}^{p} B_i(r) \neq \emptyset$ and $\mathcal{S}$ consists of a single complete clique. In that case we have a full hierarchical model. Between these two cases, and depending on the position of the z_i and the value of r we obtain a rich classes of simplicial complexes and hence hierarchical models. Some of these will be decomposable and we refer to the discussion in Lunagómez et al. (2009).

It is the study of the topology of the nerve as r changes, and in particular the behavior of its Betti numbers, which drives the area of persistent homology. A important theoretical and computational result is that this topology is also that of the reduced simplicial complex based on the Delauney complex associated with their Voronoi diagram. That is to say, for fixed r it is enough, from a topological (homotopy) viewpoint, to use the sub-complex of the Delauney complex $\mathcal{S}^-$ contained in $\mathcal{S}$. The theory derives from classical results of Borsuk (1948) and Leray (1945). One beautiful fact is that the Delauney dual complex based on the furthest point Voronoi diagram (Okabe et al., 2000), is obtained by the Alexander duality mentioned in the last subsection.

In this paper we have concentrated on the correspondence between $\mathcal{S}$ and its Stanley-Reisner ideal $I_\mathcal{S}$. The use of $I_\mathcal{S}$ is not always explicit in persistent homology but is implicit in the underlying homology theory: see Sáenz de Cabezón (2008) for a thorough investigation, including algorithms. Also, although the topology of $\mathcal{S}$ and its reduced Delaunay version $\mathcal{S}^-$ is the same, if their actual structure is different they lead to different hierarchical models. One can also use non-Euclidean metrics to define the cover and, indeed, work in different spaces and with other kinds of cover. Notwithstanding these many interesting technical issues the use of geometric constructions to define interesting classes of hierarchical model promises to be very fruitful.

6   Conclusion

There are many features and properties of monomial ideals which remain to be exploited in statistics via the isomorphism discussed in the last section. We should mention minimal free resolutions, the closely related Hilbert series, Betti numbers (including graded and multi-graded versions) and Alexander duality. It is pleasing that in the general case the development of the last section only requires consideration of square-free ideals, whose theory is a little easier than the full polynomial case. Fast algorithms are available for symbolic operations covering all these areas so that as further links are made they can be implemented. We have not covered statistical analysis in this paper. Further work is in progress to fit and

test the zero-cumulant conditions $D^\alpha g = 0$ using, for example, kernel methods.

References

ARNOLD, B. & STRAUSS, D. (1988). Bivariate distributions with exponential conditionals. *Journal of the American Statistical Association* **83**, 522–527.

BAIRAMOV, I., KOTZ, S. & KOZUBOWSKI, T. (2003). A new measure of linear local dependence. *Statistics* **37**, 243–258.

BARNDORFF-NIELSEN, O. & COX, D. (1989). Asymptotic techniques for use in statistics .

BORSUK, K. (1948). On the imbedding of systems of compacta in simplicial complexes. *Fund. Math* **35**, 217–234.

CORSO, A. & NAGEL, U. (2009). Monomial and toric ideals associated to Ferrers graphs. *Transactions of The American Mathematical Society* **361**, 1371–1395.

COZMAN, F. & WALLEY, P. (2005). Graphoid properties of epistemic irrelevance and independence. *Annals of Mathematics and Artificial Intelligence* **45**, 173–195.

DARROCH, J. (1962). Interactions in multi-factor contingency tables. *Journal of the Royal Statistical Society. Series B (Methodological)* **24**, 251–263.

DOCHTERMANN, A. & ENGSTRÖM, A. (2009). Algebraic properties of edge ideals via combinatorial topology. *the electronic journal of combinatorics* **16**, R2.

DRTON, M., STURMFELS, B. & SULLIVANT, S. (2009). *Lectures on algebraic statistics*. Birkhauser.

EDELSBRUNNER, H. (1995). The union of balls and its dual shape. *Discrete and Computational Geometry* **13**, 415–440.

EDELSBRUNNER, H. & HARER, J. (2010). *Computational topology: an introduction*. American Mathematical Society.

FRÖBERG, R. (1988). On Stanley-Reisner rings, in "Topics in Algebra". *Banach Center Public* **26**, 57–69.

GEIGER, D., MEEK, C. & STURMFELS, B. (2006). On the toric algebra of graphical models. *The Annals of Statistics* **34**, 1463–1492.

HARDY, M. (2006). Combinatorics of partial derivatives. *The Electronic Journal of Combinatorics* **13**.

HOLLAND, P. & WANG, Y. (1987). Dependence function for continuous bivariate densities. *Communications in Statistics-Theory and Methods* **16**, 863–876.

JONES, M. (1996). The local dependence function. *Biometrika* **83**, 899.

LAURITZEN, S. (1996). *Graphical models*. Oxford University Press, USA.

LERAY, J. (1945). Sur la forme des espaces topologiques et sur les points fixes des représentations. *J. Math. Pures Appl., IX. Sér.* **24**, 95–167.

LUNAGÓMEZ, S., MUKHERJEE, S. & WOLPERT, R. (2009). Geometric representations of hypergraphs for prior specification and posterior sampling. *Duke University Department of Statistical Science Discussion Paper* .

MCCULLAGH, P. (1987). *Tensor methods in statistics*. Chapman and Hall London.

MILLER, E. & STURMFELS, B. (2005). *Combinatorial Commutative Algebra*. Springer Verlag.

MUELLER, H. & YAN, X. (2001). On local moments. *Journal of Multivariate Analysis* **76**, 90–109.

NAIMAN, D. & WYNN, H. (1992). Inclusion-exclusion-Bonferroni identities and inequalities for discrete tube-like problems via Euler characteristics. *The Annals of Statistics* **20**, 43–76.

NAIMAN, D. & WYNN, H. (1997). Abstract tubes, improved inclusion-exclusion identities and inequalities and importance sampling. *The Annals of Statistics* **25**, 1954–1983.

OKABE, A., BOOTS, B., SUGIHARA, K. & CHIU, S. (2000). *Spatial tessellations: Concepts and applications of Voronoi diagrams (POD)*. New York: John Wiley & Sons.

PETROVIC, S. & STOKES, E. (2010). Markov degrees of hierarchical models determined by betti numbers of stanley-reisner ideals ArXiv:0910.1610v2.

SÁENZ DE CABEZÓN, E. (2008). *Combinatorial Koszul Homology: Computations and Applications*. Ph.D. thesis, Universidad de La Rioja.

SÁENZ DE CABEZÓN, E. & WYNN, H. (2010). Mincut ideals of two-terminal networks. *Applicable Algebra in Engineering, Communication and Computing* **21**, 443–457.

SEIDENBERG, A. (1956). Some remarks on Hilbert's Nullstellensatz. *Archiv der Mathematik* **7**, 235–240.