Hierarchical models and monomial ideals

Daniel Bruynooghe and Henry Wynn

London School of Economics
H.Wynn@lse.ac.uk

Fields Institute, April 2012

Summary diagram

Conditional indep statements

$\nearrow\swarrow$ $\searrow\nwarrow$

Graphical models $\Longleftrightarrow$ zero local cumulants

$\nwarrow\searrow$ $\nearrow\swarrow$

flag complex $\Leftrightarrow$ monomial ideal

$\Updownarrow$

"new" classes of model

Local cumulants and monomial ideals

- Notation
- $\log f$ and local cumulants
- Independence and conditional independence
- Hierarchical models
- The duality with monomial ideals
- Some “new” examples

Notation

- Random variables: $X_1, \ldots, X_p$
- pdf: $f(X) = f(X_1, \ldots, X_n)$
- $g = \log f$
- Index set: $J \in N = \{1, \ldots, p\}$
- Margin: X_J
- Multi-index: $\alpha = (\alpha_1, \ldots, \alpha_p),\ |\alpha| = \sum \alpha_i$
- Monomial: $x^\alpha = x_1^{\alpha_1} \cdots x_p^{\alpha_p}$
- Differential:
$$D^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots x_p^{\alpha_p}}$$

Ordinary moments and cumulants

$$
\begin{aligned}
\kappa_{100} &= \mu_{100} \\
\kappa_{010} &= \mu_{010} \\
\kappa_{001} &= \mu_{010} \\
\kappa_{110} &= \mu_{110} - \mu_{100}\mu_{010} \\
\kappa_{101} &= \mu_{101} - \mu_{010}\mu_{101} \\
\kappa_{011} &= \mu_{011} - \mu_{010}\mu_{001} \\
\kappa_{111} &= \mu_{111} - \mu_{100}\mu_{011} - \mu_{010}\mu_{101} - \mu_{110}\mu_{001} + 2\mu_{100}\mu_{010}\mu_{001}
\end{aligned}
$$

$$
\begin{aligned}
\mu_{100} &= \kappa_{100} \\
\mu_{010} &= \kappa_{010} \\
\mu_{001} &= \kappa_{010} \\
\mu_{110} &= \kappa_{110} + \kappa_{100}\kappa_{010} \\
\mu_{101} &= \kappa_{101} + \kappa_{010}\kappa_{101} \\
\mu_{011} &= \kappa_{011} + \kappa_{010}\kappa_{001} \\
\mu_{111} &= \mu_{111} + \kappa_{100}\kappa_{011} + \kappa_{010}\kappa_{101} + \kappa_{110}\kappa_{001} + \mu_{100}\mu_{010}\mu_{001}
\end{aligned}
$$

Local moments and cumulants

1. Take a box $B(\epsilon) = \prod_{i=1}^{P}[\xi_i - \frac{\epsilon}{2}]$,
2. Multivariate Taylor expansion of the pdf, f, at $(\xi_1, \xi_2, \ldots)$
3. Truncate and normalise the expansion over $B(\epsilon)$ (care needed)
4. Compute moments and cumulants
5. Let $\epsilon \to 0$
6. Investigate the dominant terms
7. Result: square free cumulants dominate!

Theorems

Theorem (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

$$m_k^A = r(\epsilon, k)\left(\frac{D^\alpha f_X(\xi)}{f_X(\xi)} + O(\epsilon^2)\right), \tag{1}$$

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* $\alpha := \sum_{\substack{i=1,\\ k_i \in 2\mathbb{N}+1}}^{p} e_i$.

Corollary (Local cumulants)

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

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,

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

Differential cumulants

Definition (Differential cumulant)

For an index vector k in $\mathbb{N}^p$, the differential cumulant in a in $\mathbb{R}^p$ is defined as

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

Lemma (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 even elements of k onto zero.*

Independence and conditional independence via differential cumulants

Proposition (Independence in the bivariate case)

Let X in $\mathbb{R}^2$. Then $X_1 \perp\!\!\!\perp X_2 \iff \kappa^x_{11} = 0 \quad$ *for all x in $\mathbb{R}^2$.*

Proposition (Conditional independence of two random variables)

Let X in $\mathbb{R}^p$. Then

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

where

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

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

Multivariate conditional independence

Proposition (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

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

Hierarchical models

Definition

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.

Exponential family hierarchical models: parameters which appear in a linear way, (log-linear models)

Hierarchical models via differential cumulants

Theorem

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

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

Monomial ideals

- An ideal: $\langle g(x), \ldots, g_m(x) \rangle$ is the set of all polynomials:

$$s_1(x)g(x) + \cdots s_m(x)g_m(x)$$

- A monomial ideal: all the $g_j(x)$ are monomials. Diagram

$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
x_2^5	$x_1x_2^5$	$x_1^2x_2^5$	$x_1^3x_2^5$	$x_1^4x_2^5$	$x_1^5x_2^5$	$\ldots$
x_2^4	$x_1x_2^4$	$x_1^2x_2^4$	$x_1^3x_2^4$	$x_1^4x_2^4$	$x_1^5x_2^4$	$\ldots$
x_2^3	$x_1x_2^3$	$x_1^2x_2^3$	$x_1^3x_2^3$	$x_1^4x_2^3$	$x_1^5x_2^3$	$\ldots$
x_2^2	$x_1x_2^2$	$x_1^2x_2^2$	$x_1^3x_2^2$	$x_1^4x_2^2$	$x_1^5x_2^2$	$\ldots$
x_2	x_1x_2	$x_1^2x_2$	$x_1^3x_2$	$x_1^4x_2$	$x_1^5x_2$	$\ldots$
1	x_1	x_1^2	x_1^3	x_1^4	x_1^5	$\ldots$

- : $< x_1x_2^4, x_1^3x_2^2, x_1^5 >$

Stanley-Reisner ideal and duality

- Let $\mathcal{S}$ be a hierarchical model simplicial complex, eg cliques: $\{13, 23\}$ (conditional independence).
- Take all facets NOT in $\mathcal{S}$: $\{12, 123\}$
- Construct the corresponding monomial ideal (Stanley-Reisner)

$$I_{\mathcal{S}} = < x_1 x_2 >$$

- Duality (Seidenberg nullstellensatz)

$$D^{110} g(x_1, x_2, x_3) \longleftrightarrow < x_1 x_2 >$$

- One-one correspondence

$$\text{hierarchical models} \quad \longleftrightarrow \quad \text{square free monomial ideals}$$

Examples

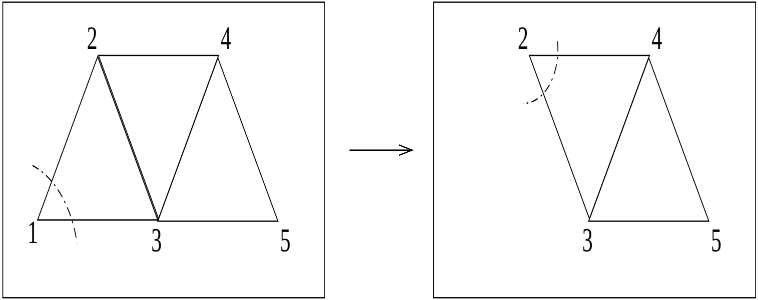
CI	$\{13,23\} \longleftrightarrow < x_1x_2 >$
3-cycle	$\{12,23,13\} \longleftrightarrow < x_1x_2x_3 >$
4-cycle	$\{12,23,34,14\} \longleftrightarrow < x_1x_3, x_2x_4 >$
decomp	$\{123,234,345\} \longleftrightarrow < x_1x_4, x_1x_5, x_2x_5 >$

Decomposability

Definition

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.

Example of decomposability



Decomposability and marginalisation

Important point: the decomposition of f_{2345} at stage 2 requires a marginalisation.

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

Two stages

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

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

Minimal free resolution theorem

- LCM for monomial ideals:

$$x_1x_2x_3 \wedge x_xx_3x_5 = x_1x_2x_3x_5$$

- Resolution: monomial maps between successive “levels”, forming an exact sequence (see alg top)
- Minimal free resolution: maps have minimal rank
- Length of MFR: projective dimension
- Linear resolution: all matrices in the resolution have linear terms
- 2-linear: linear and every generator of I_S is of degree 2 (simple interactions)

Alexander duality

- Take the model simplicial complex S.
- Construct the S-R ideal I_S
- The Alexander dual (in monomial form) is all complements of terms in I_S. eg if $n = 5$ the complement of $x_1x_3x_5$ is x_2x_4

Example:

$$S = \{123, 234\}$$

Non-faces:

$$\{14, 124, 134, 1234\}$$

Complements:

$$\{\emptyset, 2, 3, 23\}$$

Dirac's Theorem

For a model based on a graph G the following are equivalent

1. G is chordal
2. I_S has a 2-linear resolution
3. The projective dimension of I_{S^*} is 1.

Counter example

Model, S: $\{123, 124, 134, 234, 235, 15\}$

Stanley-Reisner ideal, I_S: $\langle 45.125, 135, 1234 \rangle$

$$0 \longrightarrow S \longrightarrow_C S^4 \longrightarrow_B S^4 \longrightarrow_A S \longrightarrow 0$$

$$A = [45, 125, 135, 1234]$$

$$B = \begin{pmatrix} 0 & -12 & -13 & -123 \\ 3 & 4 & 0 & 0 \\ -2 & 0 & 4 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix}$$

$$C = -[-4, 4, -2, 0]^T$$

$$AB = 0, \quad BC = 0$$

Not 2-linear $\Rightarrow$ not decomposable

Ferrer ideals

	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			

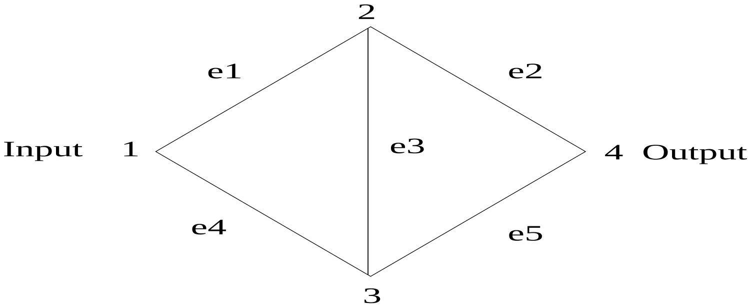
	6	7	8	9
1				x_1x_9
2			x_2x_8	x_2x_9
3			x_3x_8	x_3x_9
4		x_4x_7	x_4x_8	x_4x_9
5		x_5x_7	x_5x_8	x_5x_9

- Ferrer is 2-linear
- Take Ferrer as the Stanley-Reisner ideal $I_{\mathcal{S}}$
- The model $\mathcal{S}$ is decomposable
- Construction: work down the rows of the complementary table.
- Max cliques for the example

$$\{123459, 234589, 34589, 45789, 5789\}$$

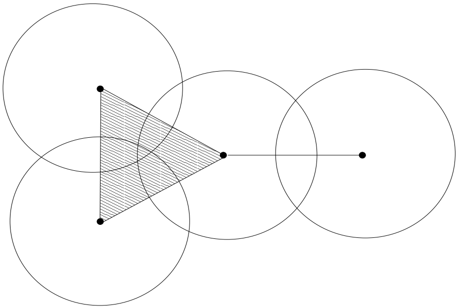
Interpretation in statistics?

Network ideals



Cut ideal: $I_{\mathcal{S}} =< x_1x_4, x_2x_5, x_1x_3x_5, x_2x_3s_4 >$
Path ideal: $I_{\mathcal{S}^*} < x_1x_2, x_4x_5, x_1x_3x_5, x_2, x_3x_5 >$
Alexander duality: $I_{\mathcal{S}^*} = \bar{I}_{\mathcal{S}}$
Extend to "all terminal" reliability: generators (for "paths") are all minimal spanning trees. Cuts are multipartite graphs.

Persistent homology constructions



- Simplicial complex depends on the centres of the sphere and the radius
- Nerve of the cover has the same topology as cover (Borsuk)
- Delauney complex $\cap$ nerve, has same topology (Naiman and W)
- Different metrics, different types of cover
- Building models using persistent homology ideas

Polynomials and Artinian closure

Lemma

Impose additional differential conditions of the form

$$\frac{\partial^{n_i}}{\partial x_i^{n_i}} g(x) = 0, \quad \text{for all } 1 \leq i \leq p \text{ and } n \in \mathbb{N}^p \tag{2}$$

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 $1 \leq i \leq p$.

- Set all $D^\alpha g = 0$ with $|\alpha| = 3$, gives quadratic. Add a NND conditions gives Gaussian
- Set all

$$\frac{\partial^2}{\partial x_i^2} g = 0$$

gives MEC: multivariate exponential class.

More general differential closure?

- How do we deal with other non-polynomials models?
- Are there any interesting continuous exponential family graphical models outside: exponential and Gaussian?
- Yes: eg multivariate von Mises:

$$f = \exp(h\text{-functions})$$

  h-functions are terms like

$$\cos(x), \sin(2x), \cos(2x+y), \sin(x+y), \ldots.$$

  But cos and sin satisfy differential equations
  Conclusion: new classes of distributions

$$f = \exp(\text{Weyl, D-modules} \ldots.)$$

  Parameters appear naturally as constants of integration
- Big problem: closed form for normalising constant (partition/potential function). Takayama et al.

Shellability

Many ideal properties for I_S need to be investigated for their implications for the model S. Shellability is one. Hibbi and Herzog (2011).

- Basic idea: we can build up a complex by adding cliques, of the same max dimension in a special order
- The “join” has one less dimension, but need not be a simplex
- Weaker than decomposability
- But has some decomposability hidden inside
- Can be generalised
- Related to other properties: Cohen-Macauley, projective dimension, Krull dimension ....

Further work

- Testing the cumulant condition: DB PhD theseis.
- Connect decomposability to ideal properties
- The key construction is the MFR: make more use of it
- Betti numbers: ordinary, graded, multigraded
- Building models by “growing” simplicial complexes
- Interpretation of “interactions”