Posets, Möbius functions and tree-cumulants

Piotr Zwiernik
(Diane Maclagan, Jim Q. Smith)

Warwick University

WOGAS2, 7th April 2010

Outline of the talk

- Part I: A motivating example: a simple naive Bayes model.
- Part II: Posets, cumulant and trees: definitions.
- Part III: Bayesian tree models: main results.

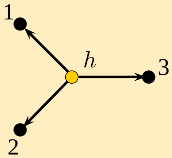
With links to some other talks*:

- Henry Wynn: there may be more than five interesting coordinate systems for discrete models (a model based approach needed?).
- Elena Stanghellini: for models on trees the algebraic statistics gives some insight into the identifiability.

* Everything is linked to everything.

The tripod tree model

- $X_1, X_2, X_3, H \in \{0,1\}$ with H hidden.



- parametric formulation of $\mathcal{M}_T$:

$$\forall_{\alpha\in\{0,1\}^3} \quad p_\alpha = \sum_{h=0}^{1} p_H(h) p_{X_1|H}(\alpha_1|h) p_{X_2|H}(\alpha_2|h) p_{X_3|H}(\alpha_3|h).$$

- 7 free parameters: $p_H(1)$ and $p_{X_i|H}(1|h)$ for $i = 1,2,3$, $h = 0,1$.
- The parameter space $\Theta = [0,1]^7$.
- The model space

$$\Delta_7 = \{p \in \mathbb{R}^8 : p_\alpha \geq 0, \sum_{\alpha\in\{0,1\}^3} p_\alpha = 1\}.$$

Change of coordinates/parameters

- "square-free" non-central moments: $\lambda_I = \mathbb{E}(\prod_{i\in I} X_i)$ for $I \subseteq \{1,2,3\}$, e.g. $\lambda_{123} = \mathbb{E}X_1X_2X_3$.
- "square-free" central moments: $\lambda_i = \mathbb{E}X_i$ and denoting $U_i = X_i - \lambda_i$

$$\mu_{ij} = \mathbb{E}U_iU_j, \quad \mu_{123} = \mathbb{E}U_1U_2U_3.$$

- $$[p_\alpha : \alpha \in \{0,1\}^3] \stackrel{1-1}{\longleftrightarrow} [\mu_I : |I| \geq 2] + [\text{means}].$$

- define $\eta_i = p_{X_i|H}(1|1) - p_{X_i|H}(1|0)$ and $\delta = 1 - 2p_H(1)$ then

$$(p_H(1), p_{X_i|H}(1,h)) \stackrel{1-1}{\longleftrightarrow} (\delta, \eta_i, \lambda_i)$$

- note that: $\eta_i = \mathrm{Cov}(X_i, H)/\mathrm{Var}(H) \Longrightarrow \mathbb{E}(U_i|H) = \eta_i(H - \mathbb{E}H)$.

The new parametrization

$$\mathcal{M}_T : \quad \begin{aligned} \mu_{12} &= \tfrac{1}{4}(1-\delta^2)\eta_1\eta_2, \\ \mu_{13} &= \tfrac{1}{4}(1-\delta^2)\eta_1\eta_3, \\ \mu_{23} &= \tfrac{1}{4}(1-\delta^2)\eta_2\eta_3, \\ \mu_{123} &= \tfrac{1}{4}(1-\delta^2)\delta\eta_1\eta_2\eta_3 \end{aligned}$$

◂ general formula

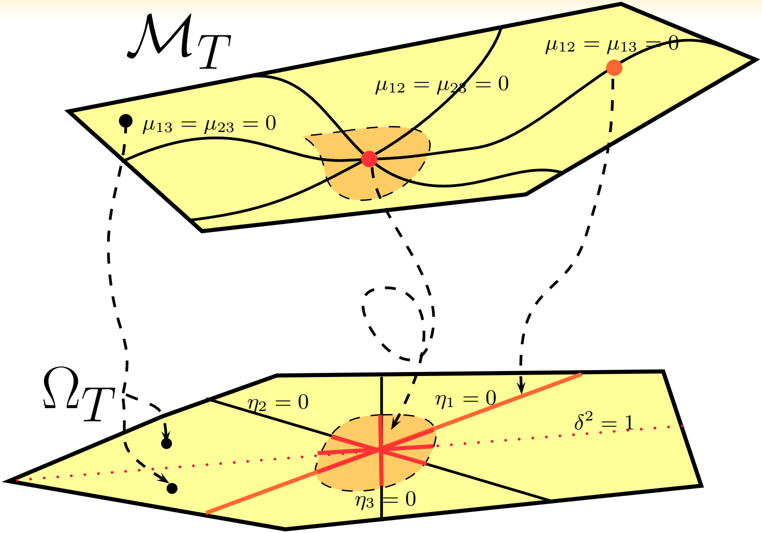
Application: Identifiability

- Case 1: If $p \in \mathcal{M}_T$ such that $\forall_{i,j}\ \mu_{ij} \neq 0$ then there are exactly two points in Θ mapping to p. For each $i = 1, 2, 3$

$$\eta_i^2 = \frac{\mu_{123}^2 + 4\mu_{12}\mu_{13}\mu_{23}}{\mu_{jk}^2}, \quad \delta^2 = \frac{\mu_{123}^2}{\mu_{123}^2 + 4\mu_{12}\mu_{13}\mu_{23}}$$

- Case 2: If $\mu_{12} = \mu_{13} = 0$ but $\mu_{23} \neq 0$ then $\eta_1 = 0$ and $\mu_{23} = \frac{1}{4}(1 - \delta^2)\eta_2\eta_3$.
- Case 3: If $\mu_{ij} = 0\,\forall_{i,j}$ then the preimage is a collection of intersecting manifolds

Application: Identifiability (cont'ed)



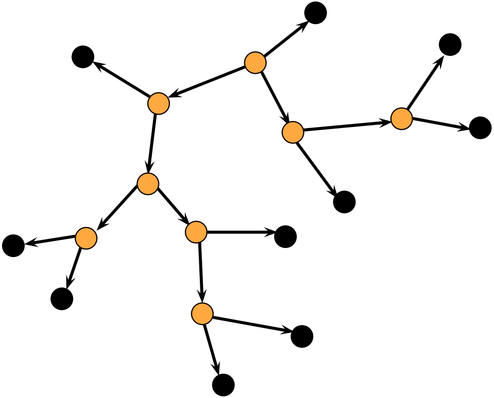
Other applications

- $\Theta = [0,1]^7$, parametrization is a polynomial map $\Longrightarrow \mathcal{M}_T$ is a semi-algebraic set
- the full description given by (Settimi,Smith 1998): e.g.

$$\mu_{12}\mu_{13}\mu_{23} = \frac{1}{64}(1-\delta^2)^3\eta_1^2\eta_2^2\eta_3^2 \geq 0.$$

- Asymptotic approximations for the marginal likelihood (Rusakov, Geiger 2005). Assume $\hat{p} \in \mathcal{M}_T$:
  - $\hat{\ell}_n - \frac{7}{2}\log n + O(1)$ if $\mu_{ij} \neq 0\ \forall_{i,j}$,
  - $\hat{\ell}_n - \frac{5}{2}\log n + O(1)$ if $\exists!_{i,j}\ \mu_{ij} \neq 0$,
  - $\hat{\ell}_n - \frac{4}{2}\log n + O(1)$ if $\mu_{ij} = 0\ \forall_{i,j}$.

Does it generalize?



Phylogenetic tree models

- $T = (V, E)$ with n leaves.
- The model given as a map $p : \Theta \to \Delta_{2^n-1}$ as

$$\mathcal{M}_T : \qquad p_x(\theta) = \sum_{\mathcal{H}} \prod_{v \in V} \theta^{(v)}_{y_v | y_{\mathrm{pa}(v)}} \text{ for } x \in \{0,1\}^n,$$

where $\theta^{(v)}_{i|j} := p(Y_v = i | Y_{\mathrm{pa}(v)} = j)$. ▸ example

- $2|E| + 1$ free parameters $\theta^{(r)}_1$ and $\forall_{(u,v)\in E}$ $\theta^{(v)}_{1|0}$ and $\theta^{(v)}_{1|1}$
- the parameter space $\Theta = [0,1]^{2|E|+1}$

Partially ordered sets

Partially ordered set (poset) Π is $(\Pi, \geq)$ such that

- For all $x \in \Pi$, $x \leq x$ (reflexivity)
- If $x \leq y$ and $y \leq x$, then $x = y$ (antisymmetry)
- If $x \leq y$ and $y \leq z$, then $x \leq z$ (trasitivity)

- All subsets of $[n] := \{1, \ldots, n\}$: $x \leq y$ iff $x \subseteq y$.
- All partitions of $[n]$: $x \leq y$ iff y is a subpartition of x. For $n = 4$
  - $x = 13|24$, $y = 1|3|24$ then $x \leq y$
  - $\hat{0} = 1234$, $\hat{1} = 1|2|3|4$
  - $|x| = 2$, $|y| = 3$, $|\hat{0}| = 1$, $|\hat{1}| = 4$

The Möbius inversion formula

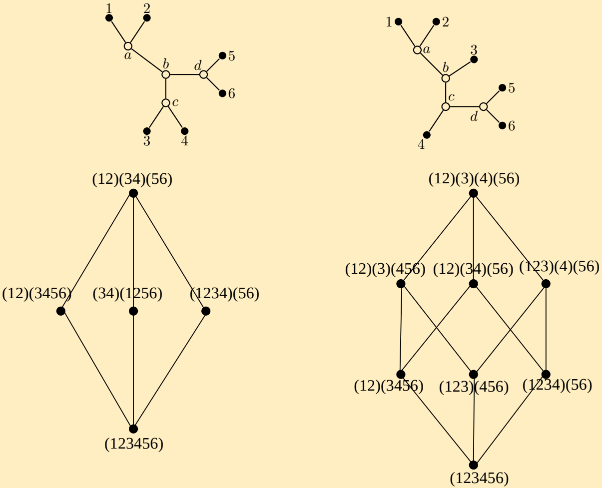
The Möbius function $m : \Pi \times \Pi \to \mathbb{R}$ such that:

- $m(x,x) = 1$ for all $x \in \Pi$.
- $m(x,y) = -\sum_{x \leq z < y} m(x,z)$ for all $x < y$.
- $m(x,y) = 0$ for all $x > y$.

Let $f, g : \Pi \to \mathbb{R}$. Then

- $g(x) = \sum_{y \leq x} f(y)$ for all $x \in \Pi$ if and only if
- $f(x) = \sum_{y \leq x} g(y) m(y,x)$

Poset of tree partitions



Tree cumulants

- T tree with n leaves, I a subset of leaves
- $\Pi_T(I)$ the poset of all the partitions of I induced by removing inner nodes together with the Möbius function m_I^T

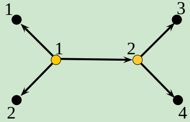
$$\kappa_I = \sum_{\pi \in \Pi_T(I)} m_I^T(\hat{0}_I, \pi) \prod_{B \in \pi} \mu_B$$

- The Möbius inversion gives the inverse map.
- Cumulants have a similar definition: $m_I(\hat{0}_I, \pi) = (-1)^{|\pi|-1}(|\pi|-1)!$
- If $|I| \leq 3$ then $\kappa_I = \mu_I$

Binary data

- $\mathbf{X} = (X_1, \ldots, X_n) \in \{0,1\}^n$ with distribution $P = [p_x]_{x \in \{0,1\}^n}$
- non-central moments: λ_I for $I \subseteq [n]$
- central moments ($+$ means): μ_I such that $|I| \geq 2$
- tree cumulants ($+$ means): κ_I such that $|I| \geq 2$

Example: The quartet tree



- $\kappa_I = \mu_I$ for all $|I| \leq 3$
- $$\begin{aligned} \kappa_{1234} &= m(\hat{0},\hat{0})\mu_{1234} + m(\hat{0},(12)(34))\mu_{12}\mu_{34} \\ &= \mu_{1234} - \mu_{12}\mu_{34} \end{aligned}$$

Reparameterization

- let $\delta_v = 1 - 2\mathbb{E}(Y_v)$ and $\eta_{uv} = \theta^{(v)}_{1|1} - \theta^{(v)}_{1|0}$ for all $(u, v) \in E$
- $\theta = (\theta^{(r)}_1, \theta^{(v)}_{1|0}, \theta^{(v)}_{1|1}) \stackrel{1-1}{\longleftrightarrow} \omega = (\delta_v, \eta_{uv})$
- note that $\eta_{uv} = \mathrm{Cov}(Y_u, Y_v)/\mathrm{Var}(Y_u)$ and $\mathbb{E}(Y_v - \mathbb{E}Y_v | Y_u) = \eta_{uv}(Y_u - \mathbb{E}Y_u)$

$T(I)$ - a subtree of T spanned on I; $r(I)$, $E(I)$, $N(I)$

$$\kappa_I = \frac{1}{4}(1 - \delta^2_{r(I)}) \prod_{v \in N(I)} \delta_v^{\deg(v)-2} \prod_{e \in E(I)} \eta_e$$

recall: tripod

Application: Identifiability

- Case 1: If $\mu_{ij} \neq 0 \, \forall_{i,j}$ then the model is identifiable up to switching labels and we easily provide the explicit formulae.
- Case 2: The preimage of p is infinite but regular.
- Case 3: The preimage is a collection of intersecting manifolds.

- The parameters identified from triples.
- The geometry of fibers determined by zeros in the covariance matrix.

Other applications

- We can list all the equations and inequalities defining the model.
- The new parameterization links to tree metrics.

- Let $\hat{p}$ be sample proportions and assume $\hat{p} \in \mathcal{M}_T$. Then if $\hat{\mu}_{ij} \neq 0$ as $N \to \infty$

$$\log Z(N) = \hat{\ell}_N - \frac{|V| + |E|}{2} \log N + O(1).$$

- The formula can be also obtained for the remaining points.

Final Comments

Quick summary

- The product like parameterization of the naive Bayes model gives a great insight into the model.
- We can obtain a similar parameterization for general tree models.

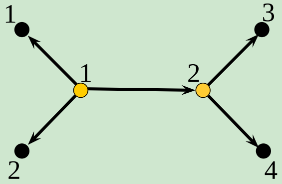
Generalizations

- Does this generalize: for general decomposable graphs, for non-binary data?
- Any other applications?

Thank you!

Example: a phylogenetic tree model

Quartet tree



- $p_x(\theta) = \sum_{h_1,h_2=0}^{1} \theta^{(5)}_{h_1} \theta^{(6)}_{h_2|h_1} \theta^{(1)}_{x_1|h_1} \theta^{(2)}_{x_2|h_1} \theta^{(3)}_{x_3|h_2} \theta^{(4)}_{x_4|h_2}$
- 11 free parameters

Go back