Marginal faces of marginal polytopes

Johannes Rauh

Max Planck Institute
for Mathematics in the Sciences

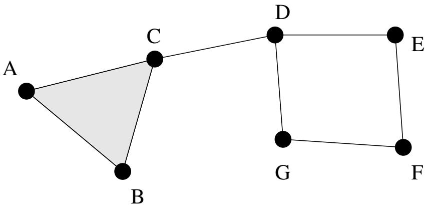
Graphical Models Oktoberfest 2019
TU München

Outline

- Graphical and hierarchical models
- Faces of marginal polytopes
- Finding faces on polytopes

Undirected graphical models

Let $G = (V, E)$ be an *undirected* graph, with V a set of *finite* r.v.s.



Definition (Parametric)

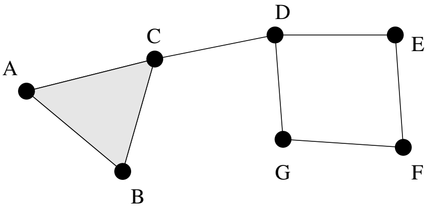
The *graphical model* $\mathcal{E}_G$ is the set of all probability distributions of the form

$$P(x_1, \ldots, x_n) = \prod_{C=\{i_1,\ldots,i_k\}\in\mathcal{C}(G)} \phi_C(x_{i_1}, \ldots, x_{i_k}),$$

where ϕ_C is a *positive* function and $\mathcal{C}(G)$ is the set of *cliques* of G (i.e. the complete subgraphs).

Undirected graphical models

Let $G = (V, E)$ be an *undirected* graph, with V a set of *finite* r.v.s.



Definition (Implicit)

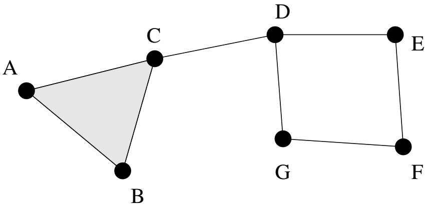
The *graphical model* $\mathcal{E}_G$ is the set of all probability distributions of *full support* such that

$$X_{V_1} \perp\!\!\!\perp X_{V_2} \,\big|\, X_{V_3} \qquad \text{whenever } V_3 \text{ separates } V_1 \text{ and } V_2.$$

(Equivalence: Hammersley-Clifford theorem)

Loglinear hierarchical models

Let $\Delta \subseteq 2^V$ be a simplicial complex, with V a set of *finite* r.v.s.



Definition (Parametric)

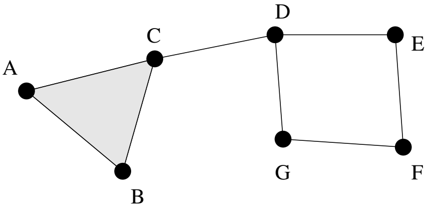
The *hierarchical model* $\mathcal{E}_\Delta$ is the set of all probability distributions of the form

$$P(x_1, \ldots, x_n) = \prod_{C=\{i_1,\ldots,i_k\}\in\Delta} \phi_C(x_{i_1}, \ldots, x_{i_k}),$$

where ϕ_C is a *positive* function.

Loglinear hierarchical models

Let $\Delta \subseteq 2^V$ be a simplicial complex, with V a set of *finite* r.v.s.



Idea

- G / Δ represents the “interaction”/“dependency” structure.
- The random variables can be understood by looking at small neighbourhoods within G / Δ.

The exponential parametrization

(Loglinear) hierarchical models are exponential families:

- Let $d_1, \ldots, d_n$ be the cardinalities of the r.v.s.
- Consider n-tensors $u \in \mathbb{R}^{d_1 \times \cdots \times d_n}$.
- For each $C \in \Delta$ let $t_C(u)$ be the *C-marginal* of u.
- Let A_Δ be the matrix that computes all C-marginals $t_C(u)$ for $C \in \Delta$ ("*sufficient statistics*"/"*moment map*").

Then $\mathcal{E}_\Delta$ consists of the distributions of the form

$$P(x_1, \ldots, x_n) = \frac{1}{Z_\theta} \exp\left(\theta^t A_{\Delta;x_1,\ldots,x_n}\right),$$

where

- θ^t is a vector of parameters;
- $A_{\Delta;x_1,\ldots,x_n}$ is the column of A_G corresponding to $x_1, \ldots, x_n$.

The marginal polytope

Definition
The convex hull of the columns of A_Δ is the *marginal polytope* $\mathbf{P}_\Delta$.

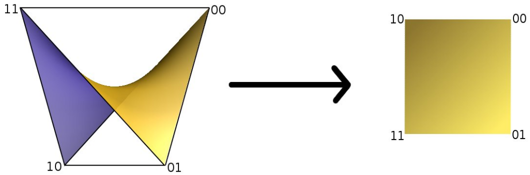
- The marginal polytope answers the question:
  Which combinations of C-marginals are compatible?
- Marginal polytopes are related to cut polytopes.
  (cut polytopes $\rightsquigarrow$ max cut problem $\rightsquigarrow$ NP completeness)

- The *moment map* $\mu : P \mapsto A_\Delta.P$ induces a bijection $\overline{\mathcal{E}_\Delta} \cong \mathbf{P}_\Delta$.
- If t are the marginals of the empirical distribution, then $\mu^{-1}(t)$ is the (generalized) MLE.

Example: Two independent binary variables

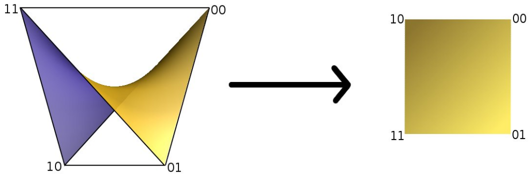
$$\Delta = \bullet \quad \bullet$$

$$A_\Delta = \begin{array}{c} \begin{array}{cccc} 00 & 01 & 10 & 11 \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \end{array}$$



The support of the GMLE

- If t are the marginals of the empirical distribution, then $\mu^{-1}(t)$ is the (generalized) MLE.
- The *support of the GMLE* corresponds to the face $\mathbf{F}$ of $\mathbf{P}_\Delta$ in which t lies:
  - Denote by $a_{\Delta,x}$ the columns of A_Δ.
  - Then $\operatorname{supp}(\mu^{-1}(t)) = \{x \in \mathcal{X} : a_{\Delta,x} \in \mathbf{F}\}$.



The support of the GMLE

- If t are the marginals of the empirical distribution, then $\mu^{-1}(t)$ is the (generalized) MLE.
- The *support of the GMLE* corresponds to the face $\mathbf{F}$ of $\mathbf{P}_\Delta$ in which t lies:
  - Denote by $a_{\Delta,x}$ the columns of A_Δ.
  - Then $\operatorname{supp}(\mu^{-1}(t)) = \{x \in \mathcal{X} : a_{\Delta,x} \in \mathbf{F}\}$.

Interpretation of the support

A support that is not full may indicate:

1. structural zeros? negligeable probabilities?
2. insufficient data?

If $\operatorname{supp}(\mu^{-1}(t))$ is not full, it highlights peculiarities of the data that are important with respect to the model.

Prominent faces of marginal polytopes

Marginal faces

For any $S \in \Delta$ and $x_S \in \bigtimes_{i \in S} \mathcal{X}_i$, the inequality $t_{S;x_S} \geq 0$ is valid.

Lemma

$t_{S;x_S} \geq 0$ *defines a facet if and only if* S *is a clique in* Δ.

Cycle faces

Every cycle in Δ contributes inequalities, the *cycle inequalities*. In the easiest case of a binary cycle x_1, x_2, x_3:

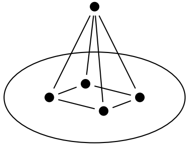
$$t_{\{1,3\};(0,0)} \leq t_{\{1,2\};(0,0)} + t_{\{2,3\};(1,0)}$$

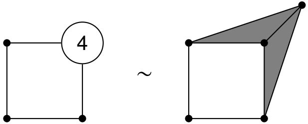
(Proof: If $t_{\{1,3\};(0,0)}(a_x) = 1$, then either $t_{\{1,2\};(0,0)}(a_x) = 1$ or $t_{\{1,3\};(1,0)}(a_x) = 1$.)

The role of marginal and cycle facets

- If Δ is a cycle, all facets are either marginal or cycle facets.
- If all variables are binary:
  - If $|S| \leq 2$ for all $S \in \Delta$[1], then all facets are either marginal or cycle facets if and only if Δ has no $K4$-minor.
  - For graphs with $|V| \leq 5$, all facets of $\mathbf{P}_G$ arise from marginal and cycle inequalities, using:

    "pyramid construction:" and "thickening:"





  - The same is true for the majority of all graphs on six nodes.

[1] I.e. $\mathbf{P}_\Delta$ is a cut polytope.

Reducible simplicial complexes

Definition

Δ is *reducible* if there exist $V_1, V_2 \subset V$ that satisfy:

1. $V \setminus V_1 \neq \emptyset$, $V \setminus V_2 \neq \emptyset$ and $V = V_1 \cup V_2$.
2. $\Delta = \Delta|_{V_1} \cup \Delta|_{V_2}$.
3. $(V_1 \cap V_2) \in \Delta$; i.e., the separator is *complete*.

If $\Delta = \Delta|_{V_1} \cup \Delta|_{V_2}$ is reducible, almost any statistical or mathematical question (about $\mathcal{E}_\Delta$ or $\mathbf{P}_\Delta$) can be answered by looking at $\Delta|_{V_1}$ and $\Delta|_{V_2}$ separately.

Concerning $\mathbf{P}_\Delta$:

Lemma (Erikson, Fienberg, Rinaldo, Sullivant 2006)

If $\Delta = \Delta|_{V_1} \cup \Delta|_{V_2}$ is reducible, then any facet-defining inequality of $\mathbf{P}_\Delta$ is a facet defining inequality of either $\mathbf{P}_{\Delta|_{V_1}}$ or $\mathbf{P}_{\Delta|_{V_2}}$.

The simplicial complex of a facet

- Sub-complexes $\Delta' \subseteq \Delta$ provide valid inequalities of $\mathbf{P}_\Delta$.
- Conversely, any facet $\mathbf{F}$ belongs to a sub-complex $\Delta(\mathbf{F})$.

Lemma
The complex $\Delta(\mathbf{F})$ of a facet $\mathbf{F}$ is irreducible.

(If $\Delta = \Delta|_{V_1} \cup \Delta|_{V_2}$ is reducible, then $\Delta(\mathbf{F}) \subseteq \Delta|_{V_1}$ or $\Delta(\mathbf{F}) \subseteq \Delta|_{V_2}$.)

Questions:
- Which sub-complexes Δ' arise in this way?
- Which facets of $\mathbf{P}_{\Delta'}$ contribute facet defining inequalities of $\mathbf{P}_\Delta$?

Finding faces on polytopes

Problem

Given a point t inside a polytope $\mathbf{P}$, determine the face $\mathbf{F}_t$ of t in $\mathbf{P}$!

Approaches:

Finding faces on polytopes

Problem

Given a point t inside a polytope $\mathbf{P}$, determine the face $\mathbf{F}_t$ of t in $\mathbf{P}$!

Approaches:

1. Compute the face lattice of $\mathbf{P}$.

Finding faces on polytopes

Problem

Given a point t inside a polytope $\mathbf{P}$, determine the face $\mathbf{F}_t$ of t in $\mathbf{P}$!

Approaches:

1. Compute the face lattice of $\mathbf{P}$.
2. Use linear programming.

Finding faces on polytopes

Problem
Given a point t inside a polytope $\mathbf{P}$, determine the face $\mathbf{F}_t$ of t in $\mathbf{P}$!

Approaches:

1. Compute the face lattice of $\mathbf{P}$.
2. Use linear programming.

Due to the relation to cut polytopes, no general easy algorithm can be expected for marginal polytopes.

3. Wang, Rauh and Massam (2019) propose *inner* and *outer approximations* of the form

$$\operatorname{conv}\{a_x : x \in F_1\} \subseteq \mathbf{F}_t \subseteq \operatorname{conv}\{a_x : x \in F_2\}.$$

Approximating faces of marginal polytopes

Observation

If $\Delta_1 \subseteq \Delta_2$, any inequality for $\mathbf{P}_{\Delta_1}$ also holds for $\mathbf{P}_{\Delta_2}$.

1. *Outer approximation* $\mathbf{F}_2 \supseteq \mathbf{F}_t$: look at sub-complex of Δ.

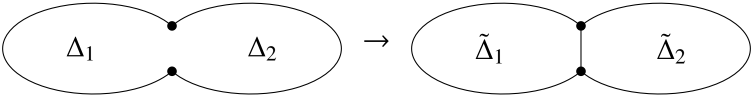
   Examples: Induced sub-complexes on few vertices, small neighbourhoods, etc.

2. *Inner approximation* $\mathbf{F}_t \subseteq \mathbf{F}_1$: look at super-complexes of Δ.

   Examples: Adding edges in order to complete separators leads to simpler marginal polytopes.

[see Wang, Rauh, Massam (Ann. Stat. 2019)]

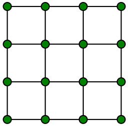
The inner approximation in detail



1. Find a small, almost-complete separator $S \subset V$.
2. Complete the separator: Let $\tilde{\Delta} = \Delta \cup \{(i, j) : i, j \in S\}$.
3. Lift t to $\tilde{t}$, by choosing a compatible S-marginal.
4. Compute the face $\mathbf{F}_{\tilde{t}}$ of $\mathbf{P}_{\tilde{\Delta}}$.
5. The approximation is:

$$\mathbf{F}_1 = \operatorname{conv}\{a_x : \tilde{a}_x \in \mathbf{F}_{\tilde{t}}\}.$$

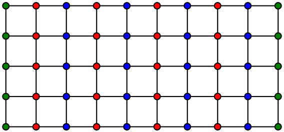
Example I: The 4×4-grid



- For each sample size, 1000 samples were generated from the model (parameters $\sim N(0, I)$).
- Outer approximation: A covering using four 3×3-grids
- Inner approx.: Use horizontal, vertical and diagonal separators.

sample size	MLE does not exist	$\mathbf{F}_1 = \mathbf{F}_t$	$\mathbf{F}_2 = \mathbf{F}_t$
10	100.0%	97.7%	100.0%
50	89.5%	100.0%	100.0%
100	71.0%	100.0%	100.0%
150	52.0%	100.0%	100.0%

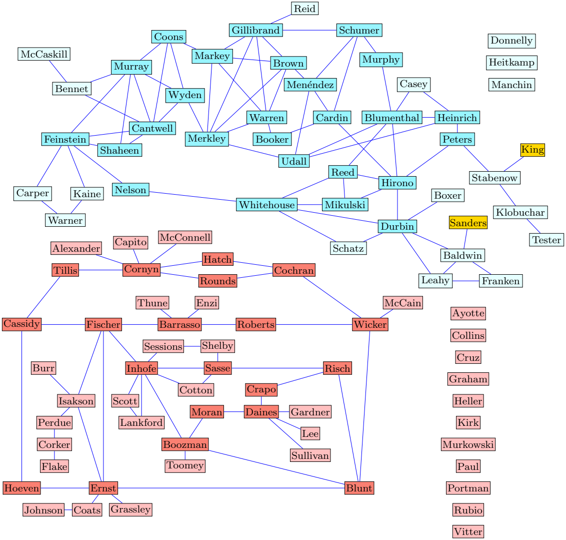
Example II: The 5×10-grid



- For each sample size, 100 samples were generated uniformly.
- Outer approximation: A covering using four 5×3-grids
- Inner approx.: Use parallel families of vertical separators

sample size	$\mathbf{F}_2 \neq \mathbf{P}$	$\mathbf{F}_1 = \mathbf{F}_2$
50	100.0%	94.3%
100	100.0%	82.5%
150	99.9%	76.5%
200	99.6%	81.2%
300	96.4%	87.7%
400	92.9%	91.5%
500	84.8%	93.9%
1000	44.7%	99.9%

Example III: US Senate voting data (2015)



Legend:

Democrat

Republican

Independent

Facets:

W. = yea ⇒ G. = yea

R. = yea ⇒ H. = yea

(W. = Warren,

G. = Gillibrand,

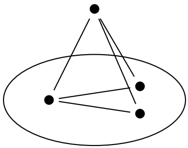
R. = Reed,

H. = Hirono)

Size of sub-complexes and facets

Question:
Do large sub-complexes give smaller facets?

- **Exception:** Pyramids
  (e.g. marginal faces)

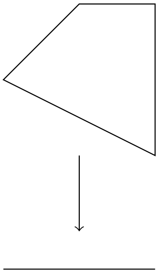


- **Positive example:** Cycle facets.

  Fraction of vertices a_x contained in:
  - an N-cycle-facet: $\mathcal{O}(c_1^{-N})$
  - an N-marginal facet: $1 - \mathcal{O}(c_2^{-N})$.

Generalization: Approximating faces of polytopes

Linear pre-images of faces are faces:



1. To find an *outer approximation*: Look at linear projections of $\mathbf{P}$.
2. To find an *inner approximation*: Look at linear liftings of $\mathbf{P}$.

For details, see Wang, Rauh, Massam (Ann. Stat. 2019)

Summary

- Marginal polytopes contain combinatorial information about discrete undirected graphical / hierarchical models.
- Faces of marginal polytopes correspond to subgraphs.
- Smaller subgraphs tend to be more important.
- This allows to efficiently approximate faces.

References:

N. Wang, J. Rauh, H. Massam
Approximating faces of marginal polytopes in discrete hierarchical models.
Annals of Statistics 47 (3), 2019.