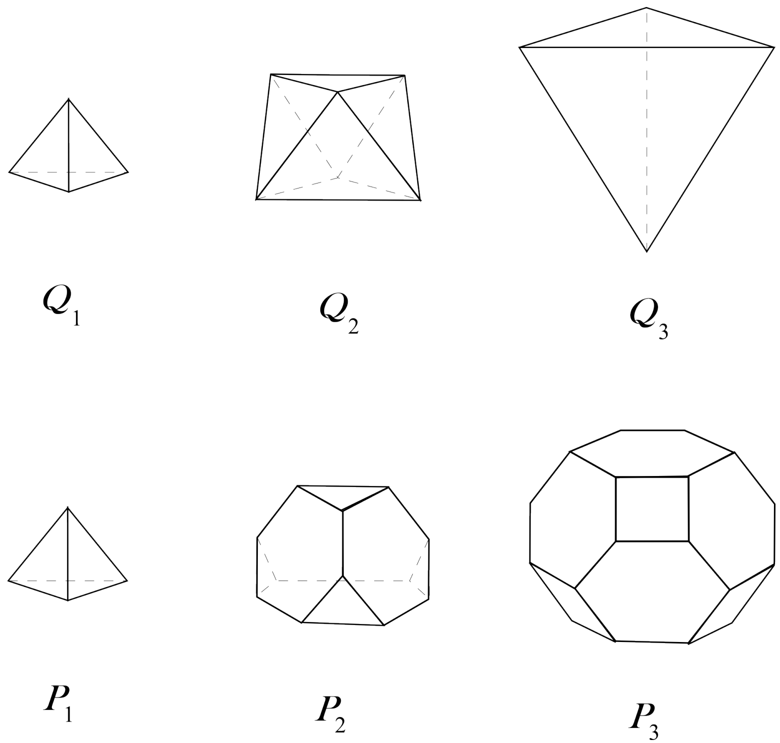
Extensions of parametric inference



Peter Huggins

Carnegie Mellon University

Statistical inference under a log-linear likelihood

- $S \subset \mathbb{Z}_+^d$ set of lattice pts (possible explanations for observed data)
- For $\mathbf{m} \in S$, log-likelihood is $\ell(\mathbf{m}) \propto c \cdot \mathbf{m}$ (linear in $\mathbf{m}$)
- c is given vector of parameters.
- Inference: Find most probable explanation, $\text{argmax}_{\mathbf{m} \in S}\, \ell(\mathbf{m})$.
- Algebraic derivation of Inference: Given polynomial $f \in \mathbb{Z}_+[x_1, \ldots, x_d]$, find leading term of f under term order c.

Recursively defined polynomials over $\mathbb{Z}_+$

- A polynomial $f \in \mathbb{Z}_+[x_1, \ldots, x_d]$ is recursively presented over $\mathbb{Z}_+$ if $f = F_N$, where
  - $F_0 := 1$
  - $F_k := G_k(F_0, \ldots, F_{k-1})$ where $G_k \in \mathbb{Z}_+[x][y_0, \ldots, y_{k-1}]$, $k = 1, \ldots, N$.
- For inference applications, definition of G_k usually depends on given observed data
- Evaluation over semirings: The entire set $F_0, F_1, \ldots, F_N$ can be efficiently evaluated in sequential order, using the recursive presentations. (Dynamic programming)
  - Example: Finding leading terms. (Inference)
  - Linear complexity: Number of semiring operations $=$ number of operations in the recursive presentations $G_1, \ldots, G_N$.

Parametric inference

- Inference: Given $f \in \mathbb{Z}_+[x_1, \ldots, x_d]$, and term order vector c, compute the leading term of f.

- Parametric inference Find all possible leading terms of f, as term order varies.

- Algebraic formulation: Compute the Newton polytope $NP(f)$.
  - Recall: $NP(f) = \text{conv}\{(m_1, \ldots, m_d) \,|\, x_1^{m_1} \cdots x_d^{m_d} \in f\}$.

- Normal cone of $(m_1, \ldots, m_d) \in NP(f)$ gives all term orders c for which $x_1^{m_1} \cdots x_d^{m_d}$ is leading term of f

Computing $NP(f)$ for recursively presented $f \in \mathbb{Z}_+[\mathbf{x}]$

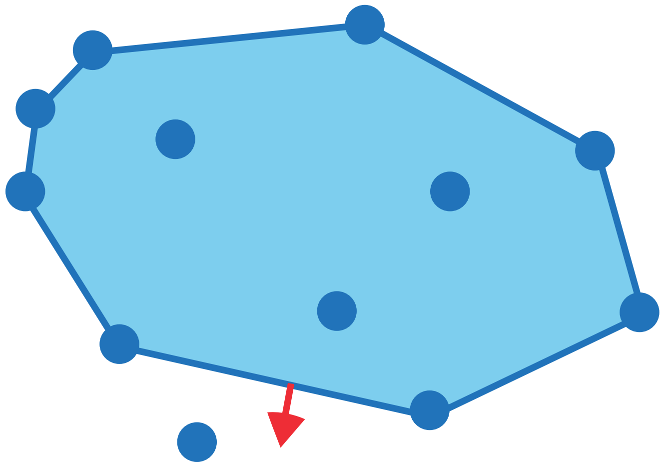
- Polytope semiring method (Pachter/Sturmfels): Evaluate recursive presentation of f, over the polytope semiring:
  - $P \odot Q :=$ Minkowski sum $P + Q$
  - $P \oplus Q :=$ conv$(P \cup Q)$
- Theorem: $f(e_1, \ldots, e_d) = NP(f)$, where $\{e_i\}$ is standard basis.

Incremental method to compute $NP(f)$

- Find new vertices by repeatedly computing leading terms of f wrt different term orders.

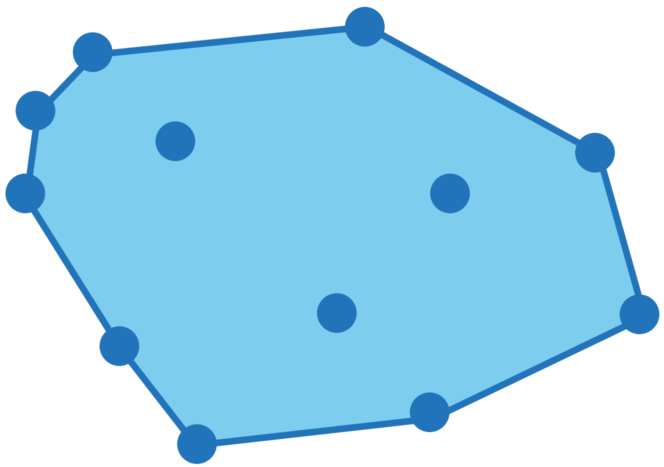
- Software `iB4e` incrementally builds polytope P, given subroutine $\texttt{Optimize}(c)$ which solves LP max $c \cdot x$ subject to $x \in P$.

How `iB4e` computes a polytope



To find new vertices, query $\texttt{Optimize}(c)$ with c orthogonal to facets of current polytope.

How $\texttt{iB4e}$ computes a polytope



To find new vertices, query $\texttt{Optimize}(c)$ with c orthogonal to facets of current polytope.

Parametric k-best inference, and k-set polytopes

Parametric k-best inference: motivation

- Correct explanation may be near-optimal likelihood, instead of optimal
- What if we want to find near-optimal explanations as parameters vary?

Parametric k-best inference

- Given $f \in \mathbb{Z}_+[x_1, \ldots, x_d]$,
- Let $S = \{(m_1, \ldots, m_d) \,|\, x_1^{m_1} \cdots x_d^{m_d} \in f\}$ be set of exponent vectors of all terms in f.
- Problem: Compute all k-tuples $(\mathbf{m}^1, \mathbf{m}^2, \ldots, \mathbf{m}^k) \subset S$ which can become the top k points in S under some linear functional.
- In other words, compute all taylor expansions of f that have k terms

k-sets and ordered k-sets

- Given $S = \{s_1, \ldots, s_n\} \subset \mathbb{R}^d$,
- Suppose $T \in \binom{S}{k}$, and $c \cdot t > c \cdot s$ for all $t \in T$ and $s \in S \backslash T$. Then T is called the k-set induced by c.
- Suppose $c \cdot s_{\sigma(1)} > c \cdot s_{\sigma(2)} > \cdots > c \cdot s_{\sigma(k)} > \cdots$. Then $\left(s_{\sigma(1)}, \ldots, s_{\sigma(k)}\right)$ is the ordered k-set induced by c.

(Unordered) k-set polytopes

- The k-set polytope $Q_k(S)$ is

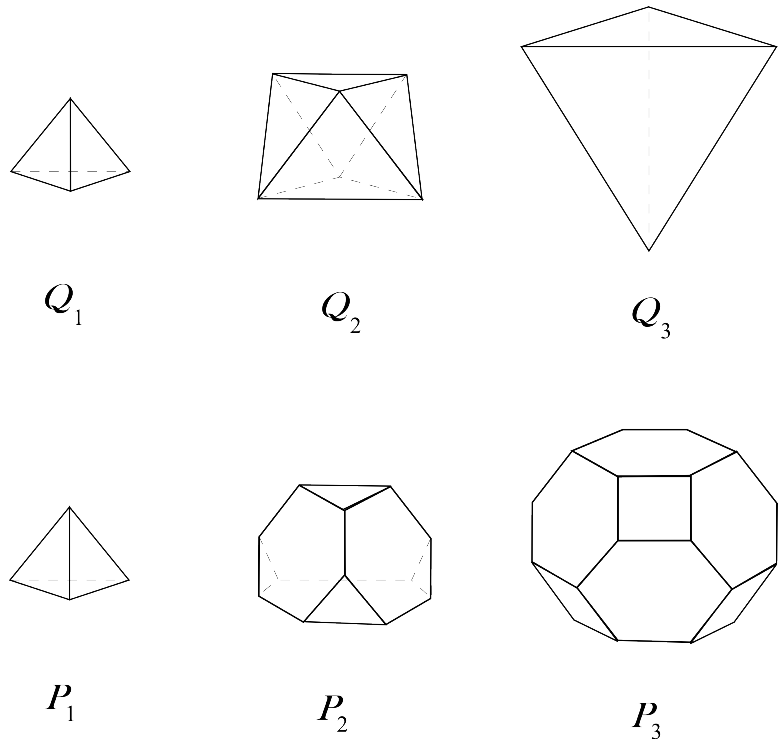
$$Q_k(S) = \text{conv}\{\sum_{s \in T} s \mid T \in \binom{S}{k}\}$$

- Studied by Edelsbrunner, Fukuda and colleagues.

- **Theorem:**
  - $\sum_{s \in T} s$ is a vertex of $Q_k(S)$ iff $T \in \binom{S}{k}$ is a k-set,
  - The normal cone $N(\sum_{s \in T} s)$ is set of c which induce k-set T.

Ordered k-set polytopes

- Note the equivalence:
  - $(s_1, \ldots, s_k)$ is the ordered k-set induced by c, iff
  - $\{s_1, \ldots, s_j\}$ is the (unordered) j-set induced by c, for each $j = 1, \ldots, k$.
- Thus vertices of $Q_1(S) + \cdots + Q_k(S)$ correspond to ordered k-sets. Normal fan characterizes how ordered k-sets depend on c.
- We call $P_k(S) := Q_1(S) + \cdots + Q_k(S)$ the ordered k-set polytope for S.



Example k-set polytopes $Q_k(S)$ and $P_k(S)$, for $S =$ vertices of a tetrahedron. Note the relation $P_3(S) = Q_1(S) + Q_2(S) + Q_3(S)$.

Parametric k-best inference

- $S =$ exponent vectors of terms in $f \in \mathbb{Z}_+[x_1, \ldots, x_d]$.
- Vertices of $P_k(S)$ give all k-tuples of top k leading terms that arise in f as term order varies.
- Normal fan of $P_k(S)$ gives the term orders which yield each choice of top k terms.

Building $P_k(S)$ for parametric k-best inference

- $S =$ exponent vectors of terms in $f \in \mathbb{Z}_+[x_1, \ldots, x_d]$.
- Vertex-finding oracle $\texttt{Optimize}(c)$ for $P_k(S)$:
  - Find the leading k terms $x^{z_1}, \ldots, x^{z_k}$ for f, under term order vector c.
  - Return vertex $\sum_{i=1}^{k}(k - i + 1)z_i$
- Thus incremental polytope construction `iB4e` can construct $P_k(S)$.

Finding top k leading terms

- Let $LT^k(f)$ denote the set of top k terms of f for given term order c.
- For $f, g \in \mathbf{Z}_+[x_1, \ldots, x_d]$ we have:
  - $LT^k(f+g) = LT^k(LT^k(f) + LT^k(g))$
  - $LT^k(fg) = LT^k(LT^k(f) \cdot LT^k(g))$
- Thus $LT^k()$ can be computed efficiently for recursively presented polynomials.

Complexity of parametric k-best inference

- For $k = 1$, best known bounds for parametric inference come from general argument of Pachter/Sturmfels.
- Let $\mathbb{V}$ be the volume of lattice polytope $P \subset \mathbb{R}^d$, d fixed.
- **Andrews theorem**: If $\mathbb{V} > 0$, P has $O\big(\mathbb{V}^{(d-1)/(d+1)}\big)$ vertices.
- **Corollary**: If $f \in \mathbb{Z}_+[x_1, \dots, x_d]$ has degree $n_i > 0$ in variable x_i, then $NP(f)$ has $O(\prod_{i=1}^{d} n_i^{(d-1)/(d+1)})$ vertices.

Complexity of parametric k-best inference

- S = exponent vectors of terms in $f \in \mathbb{Z}_+[x_1, \ldots, x_d]$. Let $\mathbb{V} = vol(NP(f))$.
- Observation: $Q_k \subset k \cdot NP(f)$, and $\dim Q_k(S) = \dim(NP(f))$.
- (Andrews theorem): If $\mathbb{V} > 0$, then $Q_k(S)$ has $O((k^d \mathbb{V})^{(d-1)/(d+1)})$ vertices, and $P_k(S)$ has $O((k^{2d} \mathbb{V})^{(d-1)/(d+1)})$ vertices.
- Polynomial in k
- For fixed k, same as best known bounds for $k = 1$ case.

Constrained parametric inference

“Reasonable” vertices of polytopes

- **General problem**: For a polytope P, we might only want vertices whose normal cones intersect a prescribed cone C.
- Often the case in parametric inference
- **Example**: Bayesian networks: logs of transition probabilities should be non-positive. Also some types of transitions might be prescribed to be more likely than others.

Envelopes of polytopes

- Given polytope $P \subset \mathbb{R}^d$, and a cone $C \subset \mathbb{R}^d$ with apex 0, the C-envelope of P is the set of faces $\{F \subset P \mid N(F) \cap C \neq \{0\}\}$.
- Graph of C-envelope is connected subgraph of graph of P.
- For parametric inference, we will assume C is full-dimensional and pointed.

Incremental construction of C-envelopes

- Can we use `iB4e` to compute C-envelope of polytope P, given oracle `Optimize()` for P?
  - Can we avoid finding vertices outside the C-envelope, to speed up the computation?
- Yes:
  - Let $u_1, \ldots, u_m$ be generators of the dual C'.
  - Initialize $P' = \text{conv}(\{Nu_i\})$ where N is a large number.
  - Then finish running `iB4e` to add vertices to P', using `Optimize()` oracle for P.
- Output P' is desired envelope. (Proof: Farkas Lemma)

Thank you

phuggins@andrew.cmu.edu