Algebraic Statistics and Hidden Markov Models

Kris Reyes

Algebraic Statistics

Algebraic statistics attempts to frame statistics in an algebraic and combinatorial setting.

- Algebra, Algebraic Geometry and Discrete Geometry.
- Groebner Bases and Polytope Computations.

Statistical Models

Consider a random variable X on a discrete set $\{1, \ldots, n\}$. We can represent a distribution of X as a point $P \in \mathbb{R}^n$ where

$$P_i = \mathbb{P}\left(X = i\right).$$

where

$$\sum_i^n P_i = 1 \text{ and } P_i > 0.$$

Such points define the *probability simplex* Δ_{n-1}.

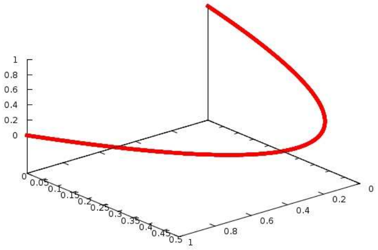
Statistical Models

If a distribution is parametric, we can view P as a function from a parameter space Ω,

$$P(\omega) : \Omega \longrightarrow \Delta_{n-1}.$$

Then the image $P(\Omega)$ cuts out a curve inside of Δ_{n-1}.

Example (*Binom*$(p, 2)$ Model)



Hidden Markov Model

A Hidden Markov Model consists of:

1. A set of *hidden states*. $\mathcal{S} = \{1, \ldots, n\}$.
2. A set of *observables*. $\Sigma = \{1, \ldots, m\}$.
3. Transition probs. $A(i,j) = \mathbb{P}\,(\text{hidden state } j | \text{hidden state } i)$.
4. Emission probs. $B(i,j) = \mathbb{P}\,(\text{observable } j | \text{hidden state } i)$.

$$\begin{array}{ccccccc} s_1 & \xrightarrow{A(s_1,s_2)} & s_2 & \xrightarrow{A(s_2,s_3)} & \ldots & \xrightarrow{A(s_{T-1},s_T)} & s_T \\ {\scriptstyle B(i_1,j_1)}\downarrow & & {\scriptstyle B(i_2,j_2)}\downarrow & & & & {\scriptstyle B(i_T,j_T)}\downarrow \\ \sigma_1 & & \sigma_2 & & \ldots & & \sigma_T \end{array}$$

Problems

A Central Problem: Inference Matching

Given an observation sequence $\sigma = (\sigma_1, \ldots, \sigma_T)$ and A, B, what is the most likely hidden sequence that produced it? That is, find the hidden sequence $s = (s_1, \ldots, s_T)$ which maximizes

$$P_\sigma(A, B, s) = \left(\prod_{i=1}^{T-1} B\left(s_i, \sigma_i\right) A\left(s_i, s_{i+1}\right) \right) B\left(s_T, \sigma_T\right).$$

Such hidden sequences are called *explanations*.

Parametric Questions

- Given a observation sequence σ and a hidden sequence s, are there parameters A, B such that s maximizes $P(A, B, s)$?
- How does the solution s obtain change with respect to A and B? Is this stable and to what degree?

Algebraic Representation

Let

$$f_\sigma(A, B) = \sum_s P_\sigma(A, B, s).$$

Definition

If Ω is the space of all pairs (A, B), define the map $f : \Omega \to \Delta_{m^T - 1}$ as

$$(A, B) \xrightarrow{f} \begin{pmatrix} f_{\sigma_1}(A, B) \\ f_{\sigma_2}(A, B) \\ \vdots \\ f_{\sigma_{m^T}}(A, B) \end{pmatrix}.$$

Then $f(\Omega)$ is called the *Algebraic Hidden Markov Model*.

Newton Polytope

Definition

For a polynomial

$$f(x_1, \ldots, x_L) = Ax_1^{\alpha_1} \cdots x_l^{\alpha_L} + Bx_1^{\beta_1} \cdots x_L^{\beta_L} + \cdots + Cx_1^{\gamma_1} \cdots x_L^{\gamma_L}.$$

then $\mathcal{N}(f)$, the *Newton Polytope* of f, is

$$\mathcal{N}(f) = conv\, \{\alpha_1, \cdots, \alpha_L), (\beta_1, \cdots, \beta_L), \cdots, (\gamma_1, \cdots, \gamma_L)\}\,.$$

For a map f with polynomial coordinate functions f_σ, $\mathcal{N}(f)$ is the *Minkowski sum* over all f_σ.

Some Key Results

Theorem

For a observation sequence σ the number of explanations s equals the number of vertices of $\mathcal{N}(f_\sigma)$. Moreover, for a fixed explanation s, the set of parameters (A, B) for which $P_\sigma(A, B, s)$ is maximized form a normal cone over a vertex of $\mathcal{N}(f\sigma)$.

Theorem

The number of inference matchings (σ, s) equals the number of vertices in $\mathcal{N}(f)$.

Theorem

The number of vertices in $\mathcal{N}(f_\sigma)$ is $\mathcal{O}(p(T))$ for polynomial p. The number of vertices in $\mathcal{N}(f)$ is $\mathcal{O}(m^{q(T)})$ for polynomial q.

Conclusions

- Not all s can serve as explanations for a fixed σ.
- The number of hidden sequences is n^T *a priori*. The number of actual explanations is polynomial in T.
- As long as we stay inside the corresponding normal cone, perturbing A and B will result in the same explanation.