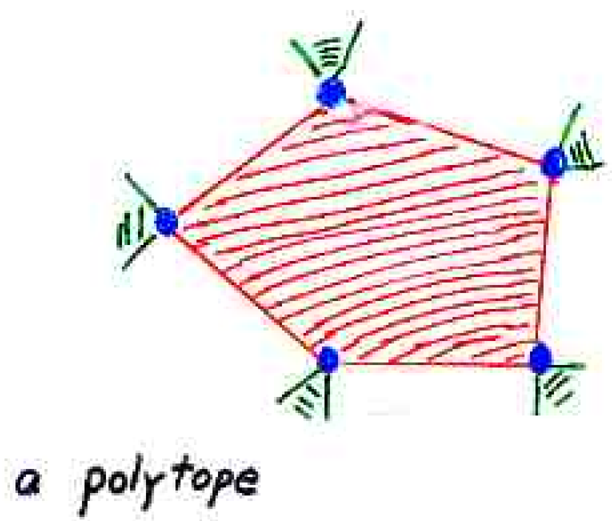
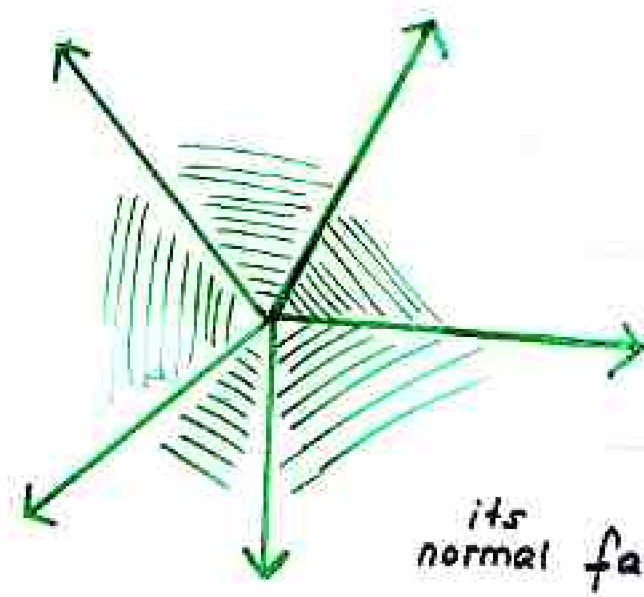
PARAMETRIC INFERENCE

BERND STURMFELS
UC Berkeley





CMI Workshop
"Algebraic Statistics and Computational Biology"

Cambridge, Mass.
Saturday, November 12, 2005

Philosophy of Algebraic Statistics

A model is given by a polynomial map

$$f : \mathbb{R}^d \longrightarrow \mathbb{R}^m$$

where d = number of parameters
m = number of observations

For $\sigma \in \{1, 2, \ldots, m\}$ and $\theta \in \mathbb{R}^d$
$f_\sigma(\theta)$ is the probability of observing σ given θ

Maximum Aposteriori Infererence means evaluating f_σ tropically, i.e. finding the largest term in f_σ

Parametric Inference means computing the Newton polytope of the polynomial f_σ or of the map f

Hidden Markov Models

Section 1.4.3
page 29-30

Fix two alphabets Σ and Σ' of size ℓ and ℓ'.
The letters in Σ are hidden states, the letters in Σ' are observed states

An observation is a word $\Upsilon \in (\Sigma')^n$

$$\boxed{m = (\ell')^n}$$

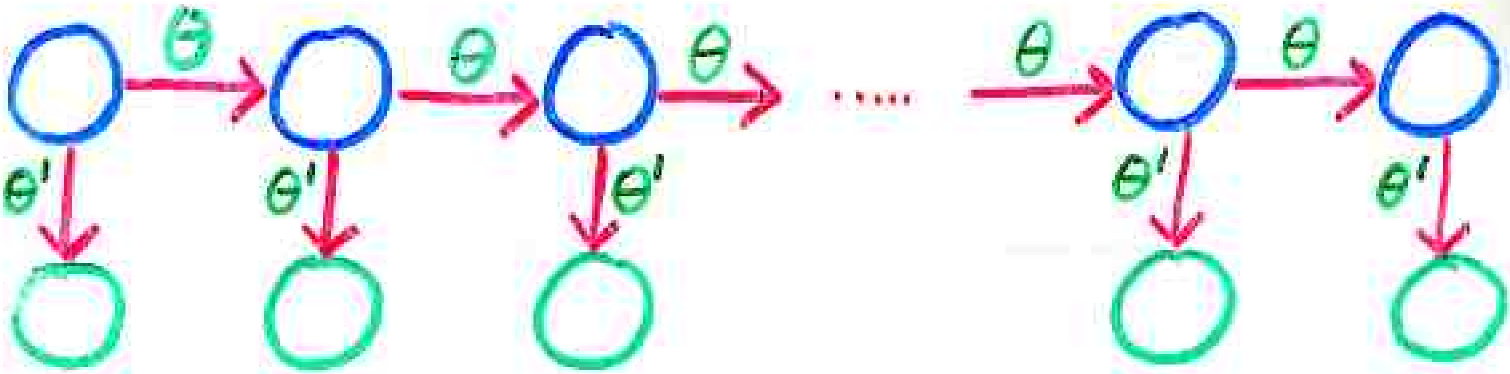
The parameters are the entries of

- the $\ell \times \ell$ transition matrix Θ
- the $\ell \times \ell'$ output matrix Θ'

$$\boxed{d = \ell^2 + \ell \cdot \ell'}$$

... or a little less

Graphical Representation of the HMM



The HMM is the polynomial map

$$f : \mathbb{R}^{\ell^2+\ell\ell'} \longrightarrow \mathbb{R}^{(\ell')^n}$$

whose coordinates are the following polynomials of degree $2n-1$

$$f_J = \sum_{\sigma\in\Sigma^n} \frac{1}{\ell}\,\theta'_{\sigma_1 J_1}\theta_{\sigma_1\sigma_2}\theta'_{\sigma_2 J_2}\theta_{\sigma_2\sigma_3}\cdots\theta_{\sigma_{n-1}\sigma_n}\theta'_{\sigma_n J_n}$$

for $J \in (\Sigma')^n$

- How to evaluate this sum of ℓ^n terms?

The Viterbi Algorithm

.... evaluates f_τ in linear time :

$$f_\tau = \frac{1}{\ell} \sum_{\sigma_1} \theta'_{\sigma_1 \tau_1} \sum_{\sigma_2} \theta_{\sigma_1 \sigma_2} \theta'_{\sigma_2 \tau_2} \sum_{\sigma_3} \theta_{\sigma_2 \sigma_3} \theta'_{\sigma_3 \tau_3} \cdots$$

Here ℓ and ℓ' are fixed but n varies

EXAMPLES

1) binary HMM $\quad \Sigma = \Sigma' = \{0, 1\}$

2) the occasionally dishonest casino

Ex. 1.21 Durbin et al

$\Sigma = \{\text{fair}, \text{loaded}\}$

$\Sigma' = \{$ ⚀, ⚁, ⚂, ⚃, ⚄, ⚅ $\}$

3) Gene Prediction (toy)

$\Sigma = \{\text{exon}, \text{intron}\} \quad \Sigma' = \{A, C, G, T\}$

4) Gene Prediction (more interesting)

Example 4.16 on page 146

Equations defining Hidden Markov Models

(Chapter 11)

Since the image of
$f : \mathbb{R}^{\ell^2 + \ell\ell'} \to \mathbb{R}^{(\ell')^n}$
is a low-dimensional variety in a high-dim. space, it is natural to ask for the polynomials that define this variety.

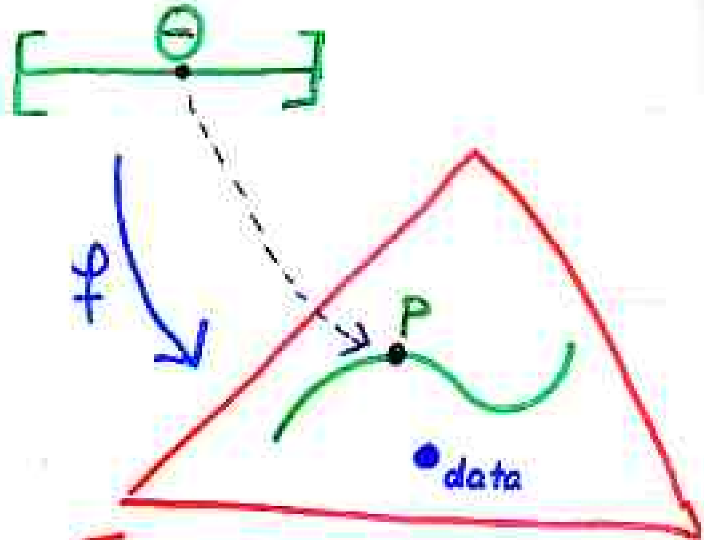
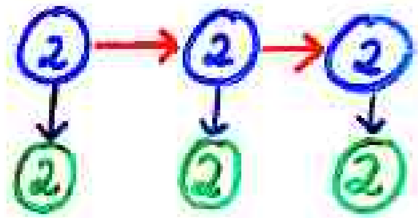


Figure 3.2, page 103

Ex. $(n=3,\ \ell=\ell'=2)$

The defining equation of



is

$$
\begin{aligned}
& p_{011}^2 p_{100}^2 - p_{001}p_{011}p_{100}p_{101} - p_{010}p_{011}p_{100}p_{101} + p_{000}p_{011}p_{101}^2 \\
& + p_{001}p_{010}p_{011}p_{110} - p_{000}p_{011}^2 p_{110} - p_{010}p_{011}p_{100}p_{110} \\
& + p_{001}p_{010}p_{101}p_{110} + p_{001}p_{100}p_{101}p_{110} - p_{000}p_{101}^2 p_{110} - p_{001}^2 p_{110}^2 \\
& + p_{000}p_{011}p_{110}^2 - p_{001}p_{010}^2 p_{111} + p_{000}p_{010}p_{011}p_{111} + p_{001}^2 p_{100}p_{111} \\
& + p_{010}^2 p_{100}p_{111} - p_{000}p_{011}p_{100}p_{111} - p_{001}p_{100}^2 p_{111} - p_{000}p_{001}p_{101}p_{111} \\
& + p_{000}p_{100}p_{101}p_{111} + p_{000}p_{001}p_{110}p_{111} - p_{000}p_{010}p_{110}p_{111}
\end{aligned}
$$

>> model invariants <<

Tropicalization

Replace $\theta_{\cdot\cdot}$ by $w_{\cdot\cdot} = -\log(\theta_{\cdot\cdot})$ and
replace f_σ by $g_\sigma = -\log(f_\sigma)$ and
replace addition by taking the minimum and
replace multiplication by classical addition.

$\boxed{3 \odot 5 = 8}$ $\boxed{3 \oplus 5 = 3}$

Section 2.1. TROPICAL ARITHMETIC & DYNAMIC PROGRAMMING

This replaces the polynomial map

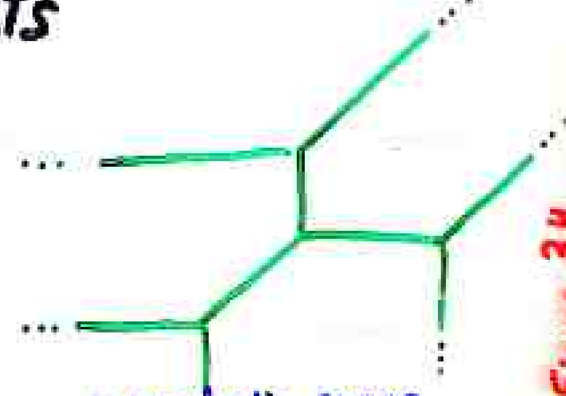
$$f : \mathbb{R}^d \longrightarrow \mathbb{R}^m$$ (i.e. our model)

by a piecewise-linear map

$$g : \mathbb{R}^d \longrightarrow \mathbb{R}^m$$ (the tropical model)

All relevant geometric objects become piecewise linear under tropicalization.

(e.g. Section 3.5. The tree of life and other tropical varieties)



a quadratic curve

Figure 3.4

MAP Inference for the HMM

Given fixed parameters $(\theta, \theta') \in \mathbb{R}^{l^2 + ll'}$ and a fixed observation $\tau \in (\Sigma')^n$ find the most likely explanation $\sigma \in \Sigma^n$

CASINO EXAMPLE

Given $\tau = 3\,5\,6\,6\,2\,1\,4\,6\,3\,5\,6 \cdots$

find $\sigma = f\,l\,l\,l\,f\,f\,f\,f\,l\,l\,l \cdots$

$$\underset{\sigma \in \Sigma^n}{\mathrm{Argmax}} \left\{ \theta_{\sigma_1 \tau_1} \theta'_{\sigma_1 \sigma_2} \theta_{\sigma_2 \tau_2} \cdots \theta_{\sigma_n \tau_n} \right\} =$$

$$\underset{\sigma \in \Sigma^n}{\mathrm{Argmin}} \left\{ w_{\sigma_1 \tau_1} + w'_{\sigma_1 \sigma_2} + w_{\sigma_2 \tau_2} + \cdots + w_{\sigma_n \tau_n} \right\}$$

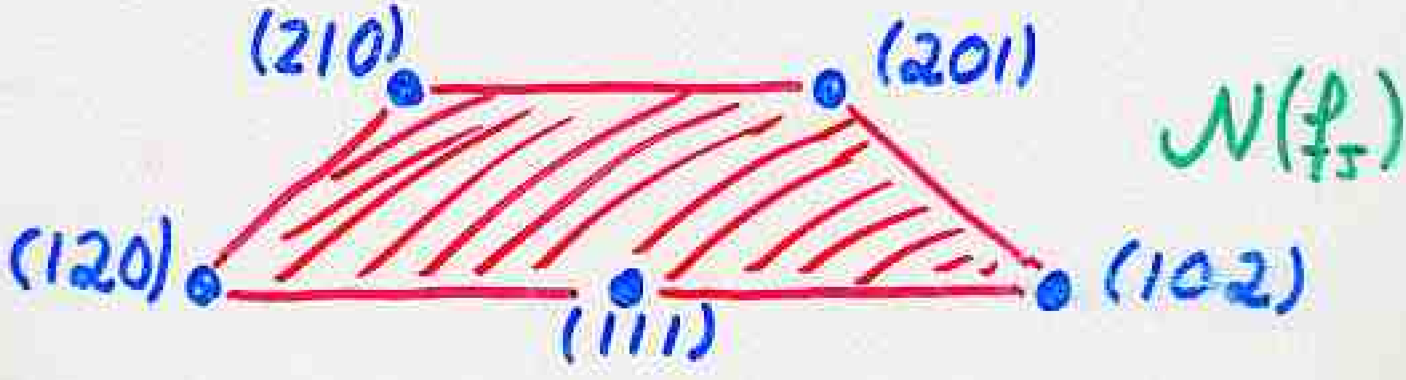
is computed by evaluating the tropical polynomial g_τ using Viterbi

Newton Polytopes

The **Newton polytope** $\mathcal{N}(f_J)$ of a polynomial f_J is the convex hull of the exponent vectors of all monomials appearing in f_J.

EXAMPLE: $(d=3)$

$$f_J = \theta_1^2\theta_2 + \theta_1^2\theta_3 + \theta_1\theta_2^2 + \theta_1\theta_2\theta_3 + \theta_1\theta_3^2$$



The **support function** of the polytope $\mathcal{N}(f_J)$ is the piecewise-linear concave function g_J

Parametric Inference - Level 1

Given a fixed observation $\tau \in (\Sigma')^n$ we precompute the function

$$\begin{array}{ccc} \mathbb{R}^d & \longrightarrow & \Sigma^n \\ \text{parameters} & \longrightarrow & \text{explanations} \\ (\theta, \theta') & \longmapsto & \sigma \end{array}$$

EQUIVALENT:

Compute the p.l. function g_τ

EQUIVALENT:

Compute the Newton polytope $\mathcal{N}(f_\tau)$

Philosophy:

"Convexity is the organizing principle that reveals the needles in the haystack"

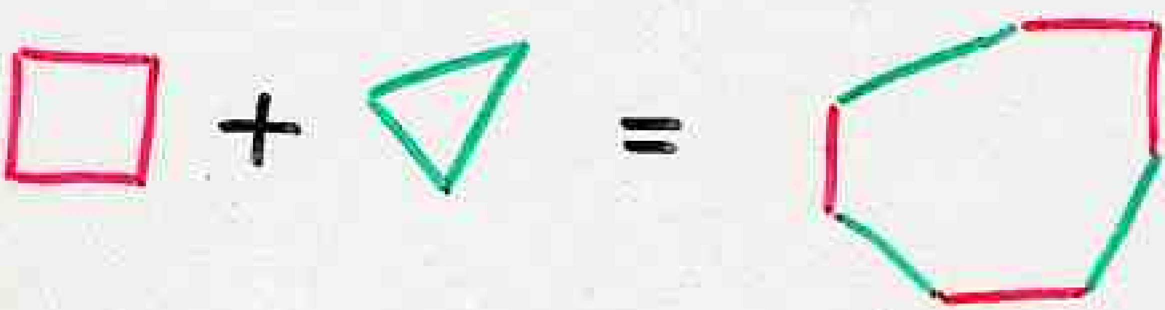
The Newton polytope of a model

$$f : \mathbb{R}^d \to \mathbb{R}^m$$

is the Minkowski sum

$$\mathcal{N}(f) := \mathcal{N}(f_1) + \mathcal{N}(f_2) + \cdots + \mathcal{N}(f_m)$$

of the Newton polytopes of all its coordinates



Aside: Statistics and Biology have already inspired new theorems in geometry & algebra, such as THEOREM 3.4.2

Parametric Inference - Level 2

The observation τ is unknown
and the parameters (θ, θ') are unknown

Q: What can we pre-compute?

A: The map $g : \mathbb{R}^{l^2 + l l'} \to \mathbb{R}^{(l')^n}$

or equivalently

the Newton polytope $\mathcal{N}(f)$
of the Hidden Markov model f

Q: What does this mean statistically?

A: The vertices of $\mathcal{N}(f)$
correspond to (equivalence classes of)

inference functions

What is an Inference Function?

Fix the binary HMM: $\Sigma = \Sigma' = \{0,1\}$

For fixed parameters (θ, θ') we have a Boolean function

$$\Phi_{\theta,\theta'} : \{0,1\}^n \longrightarrow \{0,1\}^n$$

$$\text{Observation } \tau \longmapsto \text{Best Explanation } \sigma$$

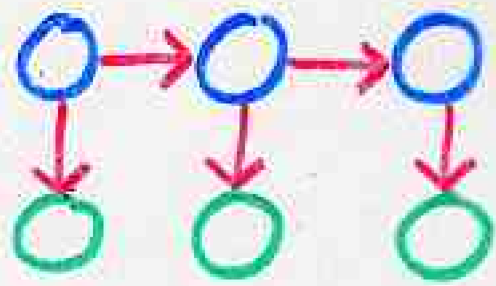
Any such function is an inference function.

There are 2^{n2^n} Boolean functions $\{0,1\}^n \to \{0,1\}^n$ but only very, very few of them are inference functions

Which ones?

All our polytopes have few vertices

Let $f : \mathbb{R}^d \to \mathbb{R}^m$ be a graphical model with E edges.



Suppose the number d of model parameters is fixed.

Theorem [§5.3] For any fixed observation σ, the number of vertices of $\mathcal{N}(f_\sigma)$ is $O\left(E^{d(d-1)/(d+1)}\right)$.

Theorem 9.3 [S. Elizalde] The number of vertices of $\mathcal{N}(f)$ is $O\left(E^{d(d-1)}\right)$

Newton polytopes of binary HMMs

Christophe Weibel computed these polytopes for $n \leq 7$

Here $d=5$, so the bound $O(n^{d(d-1)})$ is $O(n^{20})$. But, it looks like quadratic in n?

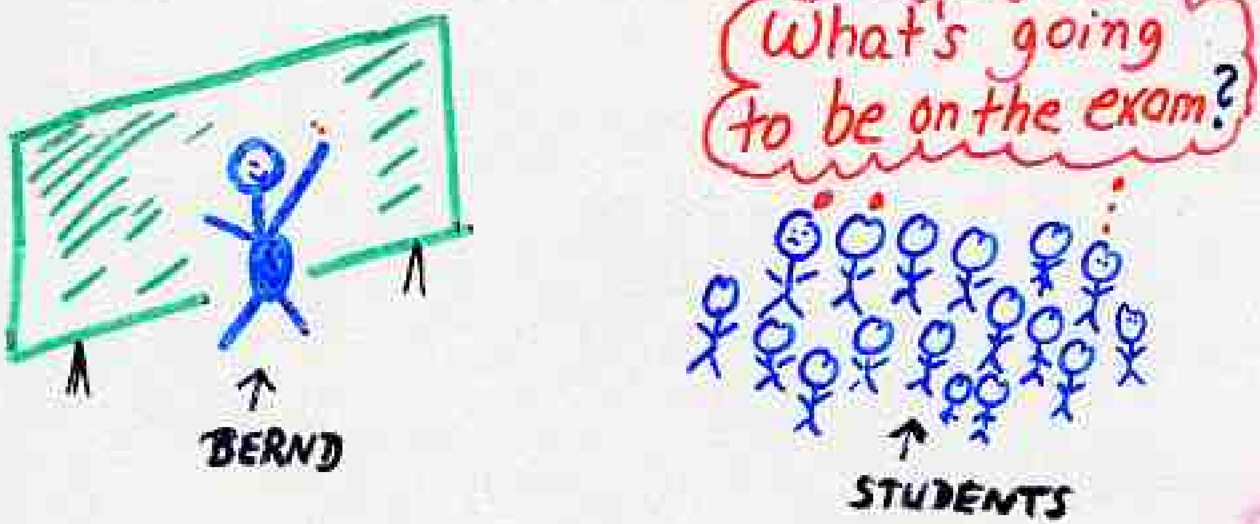
n	# vertices
2	38
3	398
4	1,570
5	5,266
6	17,354
7	55,230

Note: $2^7 = 128$

$128^{128} > 10^{269}$

What do Mathematicians do?

A: teach Calculus



TAKE-HOME EXAM

Open Book

(due today - after lunch)

What is the graphical model for sequence alignment, and what are the inference functions for that model?