MLE DEGREE OF DISCRETE RANDOM CYCLES

Courtney R. Gibbons
August 4, 2017
Joint work with Carlos Améndola, Nathan Bliss, Isaac Burke, Martin Helmer, Serkan Hoşten, Evan D. Nash, Jose Israel Rodriguez, Daniel Smolkin
**Problem.** Model to use is known and data is available $\implies$ MLE.

**Parametrize.** Build a matrix $A$ to parametrize the model.

> "Statistical models are algebraic varieties." – T. Kahle

**Scale.** Scale the the model in different ways to study ML degree.

**Polytope.** Study properties of the polytope $Q$ of $A$.

**Theorems.** Prove theorems about the way $c$ changes the number of solutions to the maximum likelihood equations for the model.
Four (binary) variables

Smoker; High blood pressure;
Family history of heart disease; High lipoprotein ratio;

$X$ : Joint binary random variable ($X_1, X_2, X_3, X_4$).

\[ i, j, k, \ell \in \{0, 1\} \]

\[ p_{ijk\ell} = \text{prob}(X_1 = i, X_2 = j, X_3 = k, X_4 = \ell) \]

\[ u_{ijk\ell} = \text{data vector}; \ u_{0000} = 297, \ u_{1000} = 275, \ldots \ u_+ := \sum u_{ijk\ell} = 1841. \]
This graph encodes independence statements: $X_1$ and $X_3$ independent given $X_2$ and $X_4$ and vice versa.

Parametrize and build a matrix:

- For each vertex, record the “on states” of $X_t$ in the joint random variable $X$. For each edge $X_tX_t'$, record the combinations of on states of both $X_t$ and $X_t'$.

Parameters: $\theta_{0001}, \theta_{0010}, \theta_{0100}, \theta_{1000}, \theta_{0011}, \theta_{0110}, \theta_{1100}, \theta_{1001}$

- Make a matrix...
- Label the rows of a matrix $A$ by the $\theta$ parameters and the columns by the probabilities $p_{0000}, p_{0001}, \ldots, p_{1111}$.

- Place a 1 in an entry if the parameter label is termwise less than or equal to the probability label.

$$
\begin{align*}
A &= \begin{pmatrix}
p_{0000} & p_{0001} & p_{0010} & p_{0011} & p_{0100} & p_{0101} & p_{0110} & p_{0111} & p_{1000} & p_{1001} & p_{1010} & p_{1011} & p_{1100} & p_{1101} & p_{1110} & p_{1111} \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
\end{pmatrix}
\end{align*}
$$
Let $V$ be the Zariski closure of the image of

$$
\psi^c : (\mathbb{C}^*)^9 \longrightarrow (\mathbb{C}^*)^{16}
$$

$$
\psi^c(s, \theta_1, \ldots, \theta_8) \longmapsto (c_1 s \theta_{col_1}^{(A)}, c_2 s \theta_{col_2}^{(A)}, \ldots, c_8 s \theta_{col_8}^{(A)})
$$

for some $c \in (\mathbb{C}^*)^{16}$.

Eg, for the matrix $A$ on the previous slide,

$$
\psi^{(c_1, \ldots, c_{16})}(\theta) = (c_{0000}s, \ldots, c_{0101}s \theta_{0100} \theta_{0001}, \ldots).
$$

Let $f = \sum_{t=1}^{16} c_t \theta^{col_t(a)}$. (The image of $\theta$ is in the hyperplane $\sum_{ijk\ell} p_{ijk\ell} - 1$.)

(Statisticians: these are probabilities, so $sf = 1$)
The Zariski closure $V^{(1,\ldots,1)}$ of the image of this parametrization is a toric variety defined by the following prime ideal:

$$I = \langle p_{1011}p_{1110} - p_{1010}p_{1111}, \quad p_{0111}p_{1101} - p_{0101}p_{1111},$$

- $p_{1001}p_{1100} - p_{1000}p_{1101}, \quad p_{0110}p_{1100} - p_{0100}p_{1110}, \quad p_{0011}p_{1001} - p_{0001}p_{1011},$
- $p_{0011}p_{0110} - p_{0010}p_{0111}, \quad p_{0001}p_{0100} - p_{0000}p_{0101}, \quad p_{0010}p_{1000} - p_{0000}p_{1010},$
- $p_{0100}p_{0111}p_{1001}p_{1010} - p_{0101}p_{0110}p_{1000}p_{1011}, \quad p_{0010}p_{0101}p_{1011}p_{1100} - p_{0011}p_{0100}p_{1010}p_{1101},$
- $p_{0001}p_{0110}p_{1010}p_{1100} - p_{0010}p_{0101}p_{1001}p_{1110}, \quad p_{0000}p_{0111}p_{1010}p_{1100} - p_{0011}p_{0100}p_{1000}p_{1110},$
- $p_{0000}p_{0011}p_{1110}p_{1100} - p_{0001}p_{0010}p_{1100}p_{1111}, \quad p_{0000}p_{0111}p_{1001}p_{1110} - p_{0001}p_{0110}p_{1000}p_{1111},$
- $p_{0000}p_{0111}p_{1010}p_{1100} - p_{0011}p_{0100}p_{1000}p_{1111} \rangle.$
Given a data vector $u$, let

$$\ell_u(p) = \frac{p_{0000}^{u_{0000}} \cdot p_{0001}^{u_{0001}} \cdots p_{1111}^{u_{1111}}}{(p_{0000} + \cdots + p_{1111})^{u_{0000} + \cdots + u_{1111}}}.$$

Goal: find a probability distribution $\hat{p} = (\hat{p}_1, \ldots, \hat{p}_n)$ in $V$ which maximizes $\ell_u$. Such a probability distribution $\hat{p}$ is a **maximum likelihood estimate**, and $\hat{p}$ can be identified by computing all critical points of $\ell_u$ on $V$.

Let $u = (u_{0000}, \ldots, u_{1111})$ and recall $u_+ = \sum_{ijk\ell} u_{ijk\ell}$. Using Lagrange multipliers, we obtain the likelihood equations for the variety $V$:

$$1 = sf$$

$$\vdots$$

$$(Au)_t = u_+ s\theta_t \frac{\partial f}{\partial \theta_t} \text{ for } t = 1, \ldots, d - 1.$$

The (only real) solution to the ML equations is the only real point in the variety over these polynomials.
The binary 4-cycle parametrized with \( c = (1, \ldots, 1) \) has degree 64 (use your favorite software or theorem to prove this) and ML degree 13.\(^1\)

1. For what \( c \) does \( \text{MLdeg}(V^c) = \text{MLdeg}(V^{(1,\ldots,1)}) \)?

2. How does the choice of \( c \) affect how much \( \text{MLdeg}(V^c) \) drops from \( \deg(V^{(1,\ldots,1)}) \)?

\(^1\)This was first computed in [GMS06, p. 1484]
The Main Idea: For $A$, the associated projective variety $V$, and polynomial $f = \sum_{t=1}^{16} c_t \theta^{\text{col}_t(A)}$, define the variety

$$\nabla_A = \left\{ c \in (\mathbb{C}^*)^{16} \mid \exists \theta \in (\mathbb{C}^*)^9 \text{ where } f \text{ and its partials by } \theta_t \text{ all vanish} \right\},$$

and then look at an irreducible polynomial that vanishes on $\nabla(A)$. If this polynomial is unique (up to scalar multiple), then it is called the $A$-\textbf{discriminant}, $\Delta_A(f)$.
When the toric variety $V$ is smooth, and $Q$ is the lattice polytope whose vertices are columns of $A$, the **Principal $A$-determinant** is

$$E_A(f) = \prod_{\Gamma \text{ nonempty face of } Q} \Delta_{\Gamma \cap A}$$

where $\Gamma \cap A$ is the matrix whose columns correspond to the lattice points contained in $\Gamma$. The locus of $E_A(f)$ is denoted $\Sigma_A$.

When $V$ is a toric hypersurface, $E_A(f) = \Delta_A(f)$ (and is easy to calculate).
The binary 3-cycle can be parametrized by

\[
B = \begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1
\end{pmatrix}
\]

and has kernel \( \ker(B) = (1, -1, -1, 1, -1, 1, 1, 1)^T \).

Thus \( B \) is a hypersurface generated by \( p_{000}p_{011}p_{101}p_{110} - p_{001}p_{010}p_{100}p_{111}, \) and \( EBf = \Delta_B(f) = c_{000}c_{011}c_{101}c_{110} - c_{001}c_{010}c_{100}c_{111} \).
Theorem (MRC Likelihood Geometry Group)

Let $V_c \subset \mathbb{P}^{n-1}$ be the projective variety defined by the monomial parametrization $\psi^c : (\mathbb{C}^*)^d \rightarrow (\mathbb{C}^*)^n$ where

$$\psi^c(s, \theta_1, \theta_2, \ldots, \theta_{d-1}) = (c_1 s \theta^{a_1}, c_2 s \theta^{a_2}, \ldots, c_n s \theta^{a_n}),$$

and $c \in (\mathbb{C}^*)^n$ is fixed. Then $\text{MLdeg}(V_c) < \text{deg}(V)$ if and only if $c$ is in the principal $A$-determinant of the toric variety $V = V^{(1, \ldots, 1)}$. 

Likelihood Geometry at SIAM AAG 2017
Courtney R. Gibbons
**Proposition (MRC Likelihood Geometry Group)**

The ML degree of the binary 3-cycle is 4 unless \( c \in (\mathbb{C}^*)^{d+1} \) is in \( \Sigma_A \). If \( c \in \Sigma_A \), then \( \text{MLdeg}(V^c) = 3 \).

**Proof.**

Observe that \( V^c \) is a hypersurface with generator \( g(p) \). Then \( E_A(f) = g(c) \). Fix a useful monomial ordering.

Find a Gröbner basis for

\[
I = \langle g, \Delta_A, \text{MLE equations} \rangle.
\]

Use [GS] and a random data vector \( u \) to calculate a Gröbner basis \( \{g_1 = \Delta_A, g_2, \ldots, g_{15}\} \) for \( I \).

Analyze degrees of the leading terms under the assumption that \( c_{ijk} \in \mathbb{C}^* \) and satisfies the equation \( g_1 = \Delta_A = 0 \): \( g_2 \) is a univariate polynomial in \( p_{111} \) of degree 3, and the initial terms of \( g_3 \) through \( g_{15} \) have degree 1 in \( p_{ijk} \). \( \Box \)
For the binary random 4-cycle, here is what we know:

- The polytope $Q$ has 24 facets, of which 5 are simplices and 3 are hypersurfaces. There are 16 with nontrivial discriminants.
  - We calculated these yesterday in [GS]!
  - We have not analyzed them yet.
- There are 168 codimension two faces of $Q$, and 88 are not simplices. Of these:
  a) 24 faces have 8 vertices. They’re all simplices or arise from hypersurfaces.
  b) 32 faces have 9 vertices. There is only one whose discriminant does not lie on coordinate hyperplanes. It’s generated by

$$c_{0110}c_{1000}c_{1011}c_{1101} + c_{0100}c_{1001}c_{1011}c_{1110} - c_{0100}c_{1001}c_{1010}c_{1111}.$$  
  c) 32 faces have 10 vertices and their discriminants all lay on coordinate hyperplanes.
Thanks!
