Maximum likelihood for dual varieties

Jose Israel Rodriguez
University of Notre Dame

SIAM AG15
August 4, 2015
A homogeneous formulation

Xiaoxian provided an affine “square” formulation, here is a homog. formulation.

- Consider the model $X$ defined by $p_1 + 2p_2 + 3p_3 - 4p_4 = 0$ and $p_1 + p_2 + p_3 + p_4 = p_s$.
  - We homogenized the equation $p_1 + p_2 + p_3 + p_4 = 1$ with respect to the unknown $p_s$.

- The model $X$ is considered as a projective variety in $\mathbb{P}^4$.
  - The codimension of $X$ is 2.

- We define the open variety $X^\circ$ as

$$X^\circ := X \setminus \{\text{coordinate hyperplanes}\}$$

$$X^\circ = \{p \in X : p \text{ has nonzero coordinates}\}.$$

- For general $u$, critical points will have nonzero coordinates
  - So it is ok to consider $X^\circ$ instead of $X$. 

A homogeneous formulation
Xiaoxian provided an affine “square” formulation, here is a homog. formulation.

- Consider the model $X$ defined by $p_1 + 2p_2 + 3p_3 - 4p_4 = 0$ and $p_1 + p_2 + p_3 + p_4 = p_s$.
  - We homogenized the equation $p_1 + p_2 + p_3 + p_4 = 1$ with respect to the unknown $p_s$.

- The model $X$ is considered as a projective variety in $\mathbb{P}^4$.
  - The codimension of $X$ is 2.

- We define the open variety $X^o$ as

\[
X^o := X \backslash \{\text{coordinate hyperplanes}\}
\]

\[
X^o = \{p \in X : p \text{ has nonzero coordinates}\}.
\]

- For general $u$, critical points will have nonzero coordinates
  - So it is ok to consider $X^o$ instead of $X$. 

Dual Varieties

Dual varieties have been studied for over a hundred years.

- Consider $X$ in $\mathbb{P}^{n+1}$.
- The **conormal variety** of $X$ is defined to be
  \[ N_X := \{(p, b) : b \perp T_p X_{\text{reg}}\} \subset \mathbb{P}^{n+1} \times \mathbb{P}^{n+1}. \]
  - Draw a picture.

- The **dual variety** of $X$, denoted $X^*$ is the projection of $N_X$ to the projective space associated to the $b$ coordinates.
  - Projecting $N_X$ to the $p$ coordinates recovers the original $X$.
- The dual of the dual gives back the original variety: $(X^*)^* = X$. 
Example 1

We compute a conormal variety.

- Consider $X$ defined by

$$p_1^3 + 5p_2^3 + 7p_3^3 = 0, \quad p_1 + p_2 + p_3 - p_s = 0.$$ 

- The conormal variety $\mathcal{N}_X$ is given by the $3 \times 3$ minors of

$$\begin{bmatrix}
  b_1 & b_2 & b_3 & b_s \\
  1 & 1 & 1 & -1 \\
  3p_1^2 & 15p_2^2 & 21p_3^2 & 0
\end{bmatrix}$$

and

$$p_1^3 + 5p_2^3 + 7p_3^3 = 0, \quad p_1 + p_2 + p_3 - p_s = 0.$$ 

- In terms of normal vectors and Lagrange multipliers we have

$$[b_1 : b_2 : b_3 : b_s] = \lambda_1 \nabla f_1 + \lambda_2 \nabla f_2.$$ 

- For singular varieties we need to saturate by the singular locus.
**Example 1**

We compute a conormal variety.

- Consider $X$ defined by

  \[ p_1^3 + 5p_2^3 + 7p_3^3 = 0, \quad p_1 + p_2 + p_3 - p_s = 0. \]

- The conormal variety $\mathcal{N}_X$ is given by the $3 \times 3$ minors of

  \[
  \begin{bmatrix}
  b_1 & b_2 & b_3 & b_s \\
  1 & 1 & 1 & -1 \\
  3p_1^2 & 15p_2^2 & 21p_3^2 & 0
  \end{bmatrix}
  \]

  and

  \[ p_1^3 + 5p_2^3 + 7p_3^3 = 0, \quad p_1 + p_2 + p_3 - p_s = 0. \]

- In terms of normal vectors and Lagrange multipliers we have

  \[ [b_1 : b_2 : b_3 : b_s] = \lambda_1 \nabla f_1 + \lambda_2 \nabla f_2. \]

- For singular varieties we need to saturate by the singular locus.
Example 2

The dual variety for a linear space is easy to compute.

- Consider $X$ defined by

\[ p_1 + 2p_2 + 3p_3 - 4p_4 = 0, \quad p_1 + p_2 + p_3 + p_4 - p_s = 0. \]

- The dual variety $X^*$ is given by the $3 \times 3$ minors of

\[
\begin{bmatrix}
  b_1 & b_2 & b_3 & b_4 & b_s \\
  1 & 1 & 1 & 1 & -1 \\
  1 & 2 & 3 & -4 & 0
\end{bmatrix}
\]

- In terms of normal vectors and Lagrange multipliers we have

\[
\]

- The conormal variety of a linear space $X$ is the product of its dual variety and self:

\[ \mathcal{N}_X \subseteq X \times X^* \quad \text{equality for a linear space } X. \]
Example 2
The dual variety for a linear space is easy to compute.

- Consider $X$ defined by

$$p_1 + 2p_2 + 3p_3 - 4p_4 = 0, \quad p_1 + p_2 + p_3 + p_4 - p_s = 0.$$ 

- The dual variety $X^*$ is given by the $3 \times 3$ minors of

$$
\begin{bmatrix}
  b_1 & b_2 & b_3 & b_4 & b_s \\
  1 & 1 & 1 & 1 & -1 \\
  1 & 2 & 3 & -4 & 0
\end{bmatrix}.
$$

- In terms of normal vectors and Lagrange multipliers we have


- The conormal variety of a linear space $X$ is the product of its dual variety and self:

$$\mathcal{N}_X \subseteq X \times X^*$$ equality for a linear space $X$. 
What is maximum likelihood estimation?

We give a formulation of the MLE problem in terms of conormal varieties.

- Maximum likelihood estimation is solving the equation below on $\mathcal{N}_X$

\[
[p_1 b_1 : p_2 b_2 : \cdots : p_s b_s] = [u_1 : \cdots : u_n : u_s], \quad u_s := -(u_1 + \cdots + u_n).
\]

- Why is this?
  - We have a critical point, if the gradient $\left[\frac{u_1}{p_1} : \frac{u_2}{p_2} : \cdots : \frac{u_n}{p_n} : \frac{u_s}{p_s}\right]$ of the likelihood function is in the row space of the Jacobian of $X$.
  - We have a point $(p, b)$ is in the conormal variety, if $b$ is in the row space of the Jacobian evaluated at $p$.
  - In other words $[b_1 : b_2 : \cdots : b_s] = \left[\frac{u_1}{p_1} : \frac{u_2}{p_2} : \cdots : \frac{u_n}{p_n} : \frac{u_s}{p_s}\right]$ on $\mathcal{N}_X$ for MLE.
Our result places maximum likelihood estimation in the context of conormal varieties.

**Theorem [-]**

Fix an algebraic statistical model $X$ and suppose $(p, b) \in N_X$. For general $u$, the following relation holds

$$[p_1 b_1 : \cdots : p_n b_n : p_s b_s] = [u_1 : \cdots : u_n : u_s].$$

iff $[p_1 : \cdots : p_n : p_s]$ is a critical point of $\ell_u(p) = p_1^{u_1} \cdots p_n^{u_n} p_s^{u_s}$ on $X$. 


A bijection between critical points
We give a bijection between critical points of two likelihood functions.

Consider $\ell_u(b) = b_1^{u_1} \cdots b_n^{u_n} b_s^{u_s}$ on $X^*$.

**Corollary [-]**

Fix an algebraic statistical model $X$ and suppose $(p, b) \in N_X$. For general $u$, the following relation holds

$$[p_1 b_1 : \cdots : p_n b_n : p_s b_s] = [u_1 : \cdots : u_n : u_s]$$

iff $[b_1 : \cdots : b_n : b_s]$ is a critical point of $\ell_u(b)$ on $X^*$.

**Corollary [-]**

For general $u$, there is a bijection between critical points of $\ell_u(p)$ on $X$ with critical point of $\ell_u(b)$ on $X^*$ given by

$$[p_1 b_1 : \cdots : p_n b_n : p_s b_s] = [u_1 : \cdots : u_n : u_s].$$
Using the bijection summary

- If we find the critical points of $\ell_u(b)$ on $X^*$ we can recover the critical points of $\ell_u(p)$ on $X$:

  $$[b_1 : \cdots : b_n : b_s] \mapsto \left[ \frac{u_1}{b_1} : \cdots : \frac{u_n}{b_n} : \frac{u_s}{b_s} \right] = [p_1 : \cdots : p_n : p_s].$$

- In the following slides we will illustrate another notion of duality called ML-duality.
Symmetric matrices

- Consider the mixture model $\mathcal{M}_m$ for pairs of $m$-sided dice. Denote its Zariski closure by $X_m$. Xiaoxian discussed the case for $m = 3$.

**Theorem**

The ML-degrees of $X_m$ include the following:

<table>
<thead>
<tr>
<th>$m$</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r = 1$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$r = 2$</td>
<td>6</td>
<td>37</td>
<td>270</td>
<td>2341</td>
</tr>
<tr>
<td>$r = 3$</td>
<td>1</td>
<td>37</td>
<td>1394</td>
<td></td>
</tr>
<tr>
<td>$r = 4$</td>
<td>1</td>
<td>270</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r = 5$</td>
<td></td>
<td>1</td>
<td>2341</td>
<td></td>
</tr>
<tr>
<td>$r = 6$</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

- Reference: “Maximum likelihood for matrices with rank constraints”
  - J. Hauenstein, [], and B. Sturmfels using Bertini.
Theorem

[Draisma and -] The symmetry above holds in general and induces a natural bijection between sets of critical points called ML-duality.

- The dual MLE problem gives a bijection between critical points in \( p \) coordinates and \( b \) coordinates.
- ML-duality gives a bijection between critical points in \( p \) coordinates and other critical points in \( p \) coordinates.
Topology

- Let $X^o$ denote the open variety $X \setminus \{\text{coordinate hyperplanes}\}$.

**Theorem [Huh]**

The ML degree of the *smooth* variety $X$ equals the signed Euler characteristic of $X^o$, i.e.

$$\chi(X^o) = (-1)^{\dim X} \text{MLdegree}(X).$$

- Generalizes of this theorem to singular varieties exist and involve Euler obstructions and Whitney stratifications.

- **Question**: Can we give a topological argument for ML-duality of matrices?
Summary

- Statistics and algebraic statistics.
- Real root classification.
- Critical points: ED degree and ML degree.
- Topological and other tools get closed form formulas.
- Symbolic and numerical computations.
- Homotopy continuation for the ML degree:
  - Draw a picture.
Thank You

Contact information

- Jose Israel Rodriguez
- jo.ro@ND.edu
- http://www.nd.edu/~jrodrig18/
Outline

- Statistics
  - Mixture model
- Applied algebraic geometry
  - Critical points
- Topology
  - ML degree
  - Euler obstructions