The maximum likelihood degree of rank 2 matrices via Euler characteristics

Jose Israel Rodriguez
University of Chicago
Joint work with Botong Wang

AMS
University of Loyola
October 3, 2015

A mixture of independence models

1. Consider a pair of four sided dice: one red die and one blue die R_1, B_1.
2. Consider a second pair of four sided dice: one red die and one blue die R_2, B_2.
3. Consider a biased coin $C = [c_1, c_2]$

- The following map induces a set of probability distributions denoted $\mathscr{M}_{44} \subset \triangle_{15} \subset \mathbb{R}^{16}$ and is called the model.

$$\Delta_1 \times (\Delta_3 \times \Delta_3) \times (\Delta_3 \times \Delta_3) \to \mathscr{M}_{44} \subset \Delta_{15} \subset \mathbb{R}^{16}$$

$$c_1 R_1 B_1^T + c_2 R_2 B_2^T = [p_{ij}]$$

- $\mathscr{M}_{44}$ is the set of 4×4 nonnegative rank at most 2 matrices.
- $\mathscr{M}_{44}$ is a mixture of two independence models.

A mixture of independence models

1. Consider a pair of four sided dice: one red die and one blue die R_1, B_1.
2. Consider a second pair of four sided dice: one red die and one blue die R_2, B_2.
3. Consider a biased coin $C = [c_1, c_2]$

- The following map induces a set of probability distributions denoted $\mathscr{M}_{44} \subset \triangle_{15} \subset \mathbb{R}^{16}$ and is called the model.

$$\Delta_1 \times (\Delta_3 \times \Delta_3) \times (\Delta_3 \times \Delta_3) \to \mathscr{M}_{44} \subset \Delta_{15} \subset \mathbb{R}^{16}$$

$$c_1 R_1 B_1^T + c_2 R_2 B_2^T = [p_{ij}]$$

- $\mathscr{M}_{44}$ is the set of 4×4 nonnegative rank at most 2 matrices.
- $\mathscr{M}_{44}$ is a mixture of two independence models.

Collecting data and the likelihood function

Roll the dice

- Rolling the dice we may observe the following data:

$$u = [u_{ij}] = \begin{bmatrix} 160 & 8 & 16 & 24 \\ 32 & 200 & 16 & 8 \\ 8 & 24 & 176 & 32 \\ 16 & 40 & 8 & 232 \end{bmatrix}$$

- To each p in the set of probability distributions $\mathscr{M}_{44}$ we assign the likelihood of p with respect to u by the likelihood function:

$$\ell_u(p) = \binom{\sum u_{ij}}{u_{11}, \ldots, u_{44}}^{-1} \prod_{ij} p_{ij}^{u_{ij}}.$$

- The probability distribution maximizing $\ell_u(p)$ on the set of distributions $\mathscr{M}_{44}$ is called the maximum likelihood estimate (mle).
- The mle is the best point of $\mathscr{M}_{44}$ to describe the observed data.
- The statistics problem is to determine mle's.

Collecting data and the likelihood function

Roll the dice

- Rolling the dice we may observe the following data:

$$u = [u_{ij}] = \begin{bmatrix} 160 & 8 & 16 & 24 \\ 32 & 200 & 16 & 8 \\ 8 & 24 & 176 & 32 \\ 16 & 40 & 8 & 232 \end{bmatrix}$$

- To each p in the set of probability distributions $\mathscr{M}_{44}$ we assign the likelihood of p with respect to u by the likelihood function:

$$\ell_u(p) = \binom{\sum u_{ij}}{u_{11}, \ldots, u_{44}}^{-1} \prod_{ij} p_{ij}^{u_{ij}}.$$

- The probability distribution maximizing $\ell_u(p)$ on the set of distributions $\mathscr{M}_{44}$ is called the maximum likelihood estimate (mle).
- The mle is the best point of $\mathscr{M}_{44}$ to describe the observed data.
- The statistics problem is to determine mle's.

Collecting data and the likelihood function

Roll the dice

- Rolling the dice we may observe the following data:

$$u = [u_{ij}] = \begin{bmatrix} 160 & 8 & 16 & 24 \\ 32 & 200 & 16 & 8 \\ 8 & 24 & 176 & 32 \\ 16 & 40 & 8 & 232 \end{bmatrix}$$

- To each p in the set of probability distributions $\mathscr{M}_{44}$ we assign the likelihood of p with respect to u by the likelihood function:

$$\ell_u(p) = \binom{\sum u_{ij}}{u_{11}, \ldots, u_{44}}^{-1} \prod_{ij} p_{ij}^{u_{ij}}.$$

- The probability distribution maximizing $\ell_u(p)$ on the set of distributions $\mathscr{M}_{44}$ is called the maximum likelihood estimate (mle).
- The mle is the best point of $\mathscr{M}_{44}$ to describe the observed data.
- The statistics problem is to determine mle's.

Applied Algebraic Geometry

The mle can be determined by solving the likelihood equations.

- Instead of $\mathscr{M}_{44}$, we consider its Zariski closure X_{44}.
- The Zariski closure is described by zero sets of homogeneous polynomials.
- The defining polynomials of X_{44} are the 3×3 minors of

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix}$$

  and the linear constraint $p_{11} + p_{12} + \cdots + p_{44} - p_s = 0$.
- The equations define a projective variety of $\mathbb{P}^{16}$: rank at ≤ 2 matrices
- We consider the homogenized likelihood function $\ell_u(p) = \prod_{ij}(p_{ij}/p_s)^{u_{ij}}$ on X_{44}.

Applied Algebraic Geometry

The mle can be determined by solving the likelihood equations.

- Instead of $\mathscr{M}_{44}$, we consider its Zariski closure X_{44}.
- The Zariski closure is described by zero sets of homogeneous polynomials.
- The defining polynomials of X_{44} are the 3×3 minors of

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix}$$

  and the linear constraint $p_{11} + p_{12} + \cdots + p_{44} - p_s = 0$.
- The equations define a projective variety of $\mathbb{P}^{16}$: rank at ≤ 2 matrices
- We consider the homogenized likelihood function $\ell_u(p) = \prod_{ij} (p_{ij}/p_s)^{u_{ij}}$ on X_{44}.

Applied Algebraic Geometry

The mle can be determined by solving the likelihood equations.

- Instead of $\mathscr{M}_{44}$, we consider its Zariski closure X_{44}.
- The Zariski closure is described by zero sets of homogeneous polynomials.
- The defining polynomials of X_{44} are the 3×3 minors of

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix}$$

  and the linear constraint $p_{11} + p_{12} + \cdots + p_{44} - p_s = 0$.
- The equations define a projective variety of $\mathbb{P}^{16}$: rank at ≤ 2 matrices
- We consider the homogenized likelihood function $\ell_u(p) = \prod_{ij} (p_{ij}/p_s)^{u_{ij}}$ on X_{44}.

Geometric definition of critical points

Critical points can be determined by solving a system of polynomial equations.

- For the models in this talk, the mle is a critical point of the homogenized likelihood function.
- The solutions to the likelihood equations are critical points.
- One way to formulate the likelihood equations is to use Lagrange multipliers.
  - We omit a formal description of the likelihood equations, but instead give a geometric description of critical points.

Geometric definition of critical points (cont.)

Critical points can be determined by solving a system of polynomial equations.

- Let X^o denote the open variety $X \setminus \{\text{coordinate hyperplanes}\}$.
  - X^o is the set of points in X which have nonzero coordinates.
- The gradient of the likelihood function up to scaling equals

$$\nabla \ell_u(p) = \begin{bmatrix} \frac{u_{11}}{p_{11}} & \frac{u_{12}}{p_{12}} & \cdots & \frac{u_{44}}{p_{44}} & \frac{u_s}{p_s} \end{bmatrix}, \quad u_s := -\sum_{ij} u_{ij}.$$

  - The gradient is defined on X^o.
- We say $p \in X^o$ is a complex critical point, whenever $\nabla \ell_u(p)$ is orthogonal to the tangent space of X at p and $p \in X^o_{reg}$.
- The mle is a critical point (in the cases we consider).

Two experiments and ML degree

Two experiments

- Consider *vectorized* datasets u for likelihood function $\ell_u(p)$ on X_{44}.
    - $u = \{160, 8, 16, 24, 32, 200, 16, 8, 8, 24, 176, 32, 16, 40, 8, 232\}$
        - 191 complex: 25 real and 166 nonreal
    - $u = \{292, 45, 62, 41, 142, 51, 44, 42, 213, 75, 67, 63, 119, 85, 58, 70\}$
        - 191 complex : 3 real and 188 nonreal
- The # of complex solutions was always 191 (this is the ML degree).
- For general choices of u we get the same number of complex critical points.
    - This number is called the *ML degree* of a variety.

Two experiments and ML degree

- Consider *vectorized* datasets u for likelihood function $\ell_u(p)$ on X_{44}.
  - $u = \{160, 8, 16, 24, 32, 200, 16, 8, 8, 24, 176, 32, 16, 40, 8, 232\}$
    - ★ 191 complex: 25 real and 166 nonreal
  - $u = \{292, 45, 62, 41, 142, 51, 44, 42, 213, 75, 67, 63, 119, 85, 58, 70\}$
    - ★ 191 complex : 3 real and 188 nonreal
- The $\#$ of complex solutions was always 191 (this is the ML degree).
- For general choices of u we get the same number of complex critical points.
  - This number is called the *ML degree* of a variety.

Previous Computational Results

- Consider the mixture model $\mathscr{M}_{mn}$ for m-sided red dice and n-sided blue dice. Denote its Zariski closure by X_{mn}.

Theorem

The ML-degrees of X_{mn} include the following:

(m,n)	3	4	5	6	7	8	9	10	11	12
3	10	26	**58**	**122**	**250**	**506**	**1018**	**2042**	*4090*	*8186*
4	26	**191**	**843**	**3119**	**6776**	?	?	?	?	?

- Reference: "Maximum likelihood for matrices with rank constraints"
  - J. Hauenstein, [], and B. Sturmfels using Bertini.
- Any conjectures for the first row? (Hint add 6.)
- "Maximum likelihood geometry in the presence of sampling and model zeros" gave supporting evidence for up to $n = 15$.
  - E. Gross and [] using Macaulay2.

Previous Computational Results

- Consider the mixture model $\mathcal{M}_{mn}$ for m-sided red dice and n-sided blue dice. Denote its Zariski closure by X_{mn}.

Theorem

The ML-degrees of X_{mn} include the following:

(m,n)	3	4	5	6	7	8	9	10	11	12
3	10	26	**58**	**122**	**250**	**506**	**1018**	**2042**	*4090*	*8186*
4	26	**191**	**843**	**3119**	**6776**	?	?	?	?	?

- Reference: "Maximum likelihood for matrices with rank constraints"
  - J. Hauenstein, [], and B. Sturmfels using Bertini.
- Any conjectures for the first row? (Hint add 6.)
- "Maximum likelihood geometry in the presence of sampling and model zeros" gave supporting evidence for up to $n = 15$.
  - E. Gross and [] using Macaulay2.

Euler characteristics and ML degrees

Huh proves that the ML degrees are an Euler characteristic in the smooth case.

- Let X be a smooth variety of $\mathbb{P}^{n+1}$ defined by homogeneous polynomials and the linear constraint

$$p_0 + p_1 + \cdots + p_n - p_s = 0.$$

- 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.

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

- The independence model (one sided coin) is smooth, but the mixture model is not.

Euler characteristics and ML degrees

Huh proves that the ML degrees are an Euler characteristic in the smooth case.

- Let X be a smooth variety of $\mathbb{P}^{n+1}$ defined by homogeneous polynomials and the linear constraint

$$p_0 + p_1 + \cdots + p_n - p_s = 0.$$

- 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.

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

- The independence model (one sided coin) is smooth, but the mixture model is not.

Independence model ML degree

Use Huh's result to give a topological proof.

- Let Z denote the Zariski closure of the independence model, a variety of $\mathbb{P}^{16}$.
- The following map gives an algebraic geometry parameterization of Z.

$$\mathbb{P}^3 \times \mathbb{P}^3 \to Z$$

$$([r_1, r_2, r_3, r_4], [b_1, b_2, b_3, b_4]) \to \left[r_i b_j, \sum_{ij} r_i b_j \right] \text{ where } i, j \in \{1,2,3,4\}.$$

- Let $\mathscr{O}$ denote $\mathbb{P}^3 \setminus \mathbf{V}(x_0 x_1 x_2 x_3 (x_0 + x_1 + x_2 + x_3))$. Then we have a parameterization of X^o given by

$$\mathscr{O} \times \mathscr{O} \to X^o$$

  because $\sum_{ij} r_i b_j = \left(\sum_i r_i\right)\left(\sum_j b_j\right)$.

- Using inclusion-exclusion and the additive properties of Euler characteristics we see that $\chi(\mathscr{O}) = -1$.
- By the product property $\chi(\mathscr{O} \times \mathscr{O}) = 1$.
- This parameterization is a homeomorphism thus $\chi(\mathscr{O} \times \mathscr{O}) = \chi(X^o)$.

Independence model ML degree

Use Huh's result to give a topological proof.

- Let Z denote the Zariski closure of the independence model, a variety of $\mathbb{P}^{16}$.
- The following map gives an algebraic geometry parameterization of Z.

$$\mathbb{P}^3 \times \mathbb{P}^3 \to Z$$

$$([r_1, r_2, r_3, r_4], [b_1, b_2, b_3, b_4]) \to \left[r_i b_j, \sum_{ij} r_i b_j \right] \text{ where } i, j \in \{1, 2, 3, 4\}.$$

- Let $\mathscr{O}$ denote $\mathbb{P}^3 \setminus \mathbf{V}(x_0 x_1 x_2 x_3 (x_0 + x_1 + x_2 + x_3))$. Then we have a parameterization of X^o given by

$$\mathscr{O} \times \mathscr{O} \to X^o$$

  because $\sum_{ij} r_i b_j = (\sum_i r_i)(\sum_j b_j)$.
- Using inclusion-exclusion and the additive properties of Euler characteristics we see that $\chi(\mathscr{O}) = -1$.
- By the product property $\chi(\mathscr{O} \times \mathscr{O}) = 1$.
- This parameterization is a homeomorphism thus $\chi(\mathscr{O} \times \mathscr{O}) = \chi(X^o)$.

Independence model ML degree

Use Huh's result to give a topological proof.

- Let Z denote the Zariski closure of the independence model, a variety of $\mathbb{P}^{16}$.
- The following map gives an algebraic geometry parameterization of Z.

$$\mathbb{P}^3 \times \mathbb{P}^3 \to Z$$

$$([r_1, r_2, r_3, r_4], [b_1, b_2, b_3, b_4]) \to \left[r_i b_j, \sum_{ij} r_i b_j \right] \text{ where } i, j \in \{1, 2, 3, 4\}.$$

- Let $\mathscr{O}$ denote $\mathbb{P}^3 \setminus \mathbf{V}(x_0 x_1 x_2 x_3 (x_0 + x_1 + x_2 + x_3))$. Then we have a parameterization of X^o given by

$$\mathscr{O} \times \mathscr{O} \to X^o$$

because $\sum_{ij} r_i b_j = (\sum_i r_i)(\sum_j b_j)$.
- Using inclusion-exclusion and the additive properties of Euler characteristics we see that $\chi(\mathscr{O}) = -1$.
- By the product property $\chi(\mathscr{O} \times \mathscr{O}) = 1$.
- This parameterization is a homeomorphism thus $\chi(\mathscr{O} \times \mathscr{O}) = \chi(X^o)$.

ML degrees of singular models

The ML degree is a stratified topological invariant.

- Let $(S_1, S_2, \ldots, S_k)$ denote a Whitney stratification of X^o.
  - When X^o is smooth the Whitney stratification is (X^o).
  - When $k = 2$, $S_1 = X^o_{reg}$ and $S_2 = X^o_{sing}$.

Theorem

Given reduced irreducible X^o with Whitney stratification $(S_1, \ldots, S_k)$, we have

$$\chi\left(X^o_{reg}\right) = e_{11}\mathrm{MLdegree}\left(\bar{S}_1\right) + e_{21}\mathrm{MLdegree}\left(\bar{S}_2\right) + \cdots + e_{k1}\mathrm{MLdegree}\left(\bar{S}_k\right).$$

- The e_{ij} are topological invariants called Euler obstructions, which can be considered as the topological multiplicity of the singularities.
- This theorem is a corollary of Botong Wang and Nero Budur's result that relates ML degrees to Gaussian degrees.
- The Euler obstruction e_{11} always equals $(-1)^{\dim X^o}$.

ML degrees of singular models

The ML degree is a stratified topological invariant.

- Let $(S_1, S_2, \ldots, S_k)$ denote a Whitney stratification of X^o.
  - When X^o is smooth the Whitney stratification is (X^o).
  - When $k = 2$, $S_1 = X^o_{reg}$ and $S_2 = X^o_{sing}$.

Theorem

Given reduced irreducible X^o with Whitney stratification $(S_1, \ldots, S_k)$, we have

$$\chi\left(X^o_{reg}\right) = e_{11}\text{MLdegree}\left(\bar{S}_1\right) + e_{21}\text{MLdegree}\left(\bar{S}_2\right) + \cdots + e_{k1}\text{MLdegree}\left(\bar{S}_k\right).$$

- The e_{ij} are topological invariants called Euler obstructions, which can be considered as the topological multiplicity of the singularities.
- This theorem is a corollary of Botong Wang and Nero Budur's result that relates ML degrees to Gaussian degrees.
- The Euler obstruction e_{11} always equals $(-1)^{\dim X^o}$.

ML degrees of singular models

The ML degree is a stratified topological invariant.

- Let $(S_1, S_2, \ldots, S_k)$ denote a Whitney stratification of X^o.
  - When X^o is smooth the Whitney stratification is (X^o).
  - When $k = 2$, $S_1 = X^o_{reg}$ and $S_2 = X^o_{sing}$.

Theorem

Given reduced irreducible X^o with Whitney stratification $(S_1, \ldots, S_k)$, we have

$$\chi\left(X^o_{reg}\right) = e_{11}\text{MLdegree}\left(\bar{S}_1\right) + e_{21}\text{MLdegree}\left(\bar{S}_2\right) + \cdots + e_{k1}\text{MLdegree}\left(\bar{S}_k\right).$$

- The e_{ij} are topological invariants called Euler obstructions, which can be considered as the topological multiplicity of the singularities.
- This theorem is a corollary of Botong Wang and Nero Budur's result that relates ML degrees to Gaussian degrees.
- The Euler obstruction e_{11} always equals $(-1)^{\dim X^o}$.

Ternary Cubic Example for Singular Case

We determine the ML degree of a singular X using the previous theorem.

- Let X be defined by

$$p_2 (p_1 - p_2)^2 - (p_0 - p_2)^3 = p_0 + p_1 + p_2 - p_s = 0.$$

- The Whitney stratification of X^o consists of S_1 the regular points (so $\bar{S}_1 = X$) and S_2 the singular point which is $[1:1:1:3]$,

$$\chi(S_1) = e_{11}\text{MLdegree}(X) + e_{21}\text{MLdegree}(\bar{S}_2).$$

- S_2 is a point so $S_2 = \bar{S}_2$ and $\text{MLdegree}(\bar{S}_2) = 1$.
- The Euler obstruction e_{21} is the signed multiplicity of the singular point, i.e. $e_{21} = -2$.
  - In general, the sign depends on the dimension of S_2 and the multiplicity is actually the Euler characteristic of a link [Kashiwara].
- The Euler obstruction e_{11} always equals $(-1)^{\dim X}$.

Ternary Cubic Example for Singular Case

We determine the ML degree of a singular X using the previous theorem.

- Let X be defined by

$$p_2 (p_1 - p_2)^2 - (p_0 - p_2)^3 = p_0 + p_1 + p_2 - p_s = 0.$$

- The Whitney stratification of X^o consists of S_1 the regular points (so $\bar{S}_1 = X$) and S_2 the singular point which is $[1 : 1 : 1 : 3]$,

$$\chi (S_1) = e_{11} \text{MLdegree} (X) + e_{21} \text{MLdegree} (\bar{S}_2).$$

- S_2 is a point so $S_2 = \bar{S}_2$ and $\text{MLdegree} (\bar{S}_2) = 1$.
- The Euler obstruction e_{21} is the signed multiplicity of the singular point, i.e. $e_{21} = -2$.
  - In general, the sign depends on the dimension of S_2 and the multiplicity is actually the Euler characteristic of a link [Kashiwara].
- The Euler obstruction e_{11} always equals $(-1)^{\dim X}$.

Ternary Cubic Example for Singular Case

We determine the ML degree of a singular X using the previous theorem.

- Let X be defined by

$$p_2 (p_1 - p_2)^2 - (p_0 - p_2)^3 = p_0 + p_1 + p_2 - p_s = 0.$$

- The Whitney stratification of X^o consists of S_1 the regular points (so $\bar{S}_1 = X$) and S_2 the singular point which is $[1:1:1:3]$,

$$\chi (S_1) = e_{11} \text{MLdegree} (X) + e_{21} \text{MLdegree} (\bar{S}_2).$$

- S_2 is a point so $S_2 = \bar{S}_2$ and $\text{MLdegree} (\bar{S}_2) = 1$.
- The Euler obstruction e_{21} is the signed multiplicity of the singular point, i.e. $e_{21} = -2$.
  - In general, the sign depends on the dimension of S_2 and the multiplicity is actually the Euler characteristic of a link [Kashiwara].
- The Euler obstruction e_{11} always equals $(-1)^{\dim X}$.

Returning to the mixture model

We apply the Whitney stratification-ML degree theorem to X_{mn}^o.

- The Whitney stratification of $X^o = X_{mn}^o$ is given by (S_1, S_2) where S_1 are the regular points $X_{mn}^o \backslash Z_{mn}^o$ and S_2 are the singular points Z_{mn}^o.
  - Denote the singular points of X_{mn}^o by Z_{mn}^o.
  - Z_{mn}^o should be thought of as the set of rank 1 matrices (Z_{mn} is the Zariski closure of the independence model)
- By the theorem we have

$$\chi(X_{mn}^o \backslash Z_{mn}^o) = e_{11}\text{MLdegree}(X_{mn}) + e_{21}\text{MLdegree}(Z_{mn}).$$

- It is already well known $e_{11} = -1$ and $\text{MLdegree}(Z_{mn}) = 1$.
- The first lemma we would prove determines e_{21}:

$$e_{21} = (-1)^{m+n-1}(\min\{m,n\} - 1).$$

- If we knew $\chi(X_{mn}^o \backslash Z_{mn}^o)$, then we would know $\text{MLdegree}(X_{mn})$.

Returning to the mixture model

We apply the Whitney stratification-ML degree theorem to X^o_{mn}.

- The Whitney stratification of $X^o = X^o_{mn}$ is given by (S_1, S_2) where S_1 are the regular points $X^o_{mn} \backslash Z^o_{mn}$ and S_2 are the singular points Z^o_{mn}.
    - Denote the singular points of X^o_{mn} by Z^o_{mn}.
    - Z^o_{mn} should be thought of as the set of rank 1 matrices (Z_{mn} is the Zariski closure of the independence model)
- By the theorem we have

$$\chi(X^o_{mn} \backslash Z^o_{mn}) = e_{11}\text{MLdegree}(X_{mn}) + e_{21}\text{MLdegree}(Z_{mn}).$$

- It is already well known $e_{11} = -1$ and $\text{MLdegree}(Z_{mn}) = 1$.
- The first lemma we would prove determines e_{21}:

$$e_{21} = (-1)^{m+n-1}(\min\{m,n\} - 1).$$

- If we knew $\chi(X^o_{mn} \backslash Z^o_{mn})$, then we would know $\text{MLdegree}(X_{mn})$.

Returning to the mixture model

We apply the Whitney stratification-ML degree theorem to X^o_{mn}.

- The Whitney stratification of $X^o = X^o_{mn}$ is given by (S_1, S_2) where S_1 are the regular points $X^o_{mn} \backslash Z^o_{mn}$ and S_2 are the singular points Z^o_{mn}.
  - Denote the singular points of X^o_{mn} by Z^o_{mn}.
  - Z^o_{mn} should be thought of as the set of rank 1 matrices (Z_{mn} is the Zariski closure of the independence model)
- By the theorem we have

$$\chi\left(X^o_{mn} \backslash Z^o_{mn}\right) = e_{11}\text{MLdegree}\left(X_{mn}\right) + e_{21}\text{MLdegree}\left(Z_{mn}\right).$$

- It is already well known $e_{11} = -1$ and $\text{MLdegree}\left(Z_{mn}\right) = 1$.
- The first lemma we would prove determines e_{21}:

$$e_{21} = (-1)^{m+n-1}\left(\min\{m, n\} - 1\right).$$

- If we knew $\chi\left(X^o_{mn} \backslash Z^o_{mn}\right)$, then we would know $\text{MLdegree}\left(X_{mn}\right)$.

Determining the Euler characteristic $\chi(X^o_{mn} \backslash Z^o_{mn})$

This is our main theorem.

- If we knew $\chi(X^o_{mn} \backslash Z^o_{mn})$, then we would know $\mathrm{MLdegree}(X_{mn})$.
- Let Λ_m be a sequence of $m-1$ integers $(\lambda_1, \lambda_2, \ldots, \lambda_{m-1})$.

Theorem [- and B. Wang]

Fix m greater than or equal to 2. Then, there exists Λ_m such that

$$\chi(X^o_{mn} \backslash Z^o_{mn}) = (-1)^{n-1} \sum_{1 \leq i \leq m-1} \frac{\lambda_i}{i+1} - \sum_{1 \leq i \leq m-1} \frac{\lambda_i}{i+1} \cdot i^{n-1}.$$

- Now we prove the conjecture of Hauenstein, [], Sturmfels.

Using the main theorem

Fix $m = 3$.

$$\chi(X_{3n}^{o} \backslash Z_{3n}^{o}) = (-1)^{n-1}\left(\frac{\lambda_1}{2} + \frac{\lambda_2}{3}\right) - \left(\frac{\lambda_1}{2}\cdot 1^{n-1} + \frac{\lambda_2}{3}\cdot 2^{n-1}\right).$$

$$\chi(X_{mn}^{o} \backslash Z_{mn}^{o}) = -\text{MLdegree}(X_{3n}) + (-1)^{3+n-1}(\min\{3, n\} - 1).$$

- $\text{MLdegree}(X_{32}) = 1$ yields the relation $-\lambda_1 - \lambda_2 = 0$.
- $\text{MLdegree}(X_{33}) = 10$ yields the relation $-\lambda_2 = -12$.

$$\text{MLdegree}(X_{3n}) = (2^{n+1} - 6) + (-1)^n((\min\{3, n\} - 3))$$

- Main idea: For fixed m, if we knew

$$\text{MLdegree}(X_{m2}), \text{MLdegree}(X_{m3}), \ldots, \text{MLdegree}(X_{mm})$$

then we can solve for $\Lambda_m = (\lambda_1, \ldots, \lambda_{m-1})$ thereby giving a closed form expression for $\text{MLdegree}(X_{mn})$ for all n.

Using the main theorem

Fix $m = 3$.

$$\chi(X_{3n}^o \backslash Z_{3n}^o) = (-1)^{n-1}\left(\frac{\lambda_1}{2} + \frac{\lambda_2}{3}\right) - \left(\frac{\lambda_1}{2} \cdot 1^{n-1} + \frac{\lambda_2}{3} \cdot 2^{n-1}\right).$$

$$\chi(X_{mn}^o \backslash Z_{mn}^o) = -\mathrm{MLdegree}(X_{3n}) + (-1)^{3+n-1}(\min\{3, n\} - 1).$$

- $\mathrm{MLdegree}(X_{32}) = 1$ yields the relation $-\lambda_1 - \lambda_2 = 0$.
- $\mathrm{MLdegree}(X_{33}) = 10$ yields the relation $-\lambda_2 = -12$.

$$\mathrm{MLdegree}(X_{3n}) = (2^{n+1} - 6) + (-1)^n((\min\{3, n\} - 3))$$

- Main idea: For fixed m, if we knew

  $$\mathrm{MLdegree}(X_{m2}), \mathrm{MLdegree}(X_{m3}), \ldots, \mathrm{MLdegree}(X_{mm})$$

  then we can solve for $\Lambda_m = (\lambda_1, \ldots, \lambda_{m-1})$ thereby giving a closed form expression for $\mathrm{MLdegree}(X_{mn})$ for all n.

Do Better

- Main idea (from previous slide): For fixed m, if we knew

$$\mathrm{MLdegree}(X_{m2}), \mathrm{MLdegree}(X_{m3}), \ldots, \mathrm{MLdegree}(X_{mm})$$

  then we can solve for $\Lambda_m = (\lambda_1, \ldots, \lambda_{m-1})$ thereby giving a closed form expression for $\mathrm{MLdegree}(X_{mn})$ for all n.

- Main idea (this slide): Recursively determine Λ_m thereby giving a closed form formula for $\mathrm{MLdegree}(X_{mn})$ for fixed m but *any* n.
  - Note $\mathrm{MLdegree}(X_{mn}) = \mathrm{MLdegree}(X_{nm})$.
  - Prove λ_{m-1} of Λ_m is $(m-1)m!$.

- Closed form expressions for fixed m and $n \geq m$:

$$\mathrm{MLdeg} X_{4n} = 25 \cdot 1^{n-1} - 40 \cdot 2^{n-1} + 23 \cdot 3^{n-1}$$

$$\mathrm{MLdeg} X_{5n} = -90 \cdot 1^{n-1} + 260 \cdot 2^{n-1} - 270 \cdot 3^{n-1} + 96 \cdot 4^{n-1}$$

$$\mathrm{MLdeg} X_{6n} = 301 \cdot 1^{n-1} - 1400 \cdot 2^{n-1} + 2520 \cdot 3^{n-1} - 2016 \cdot 4^{n-1} + 600 \cdot 5^{n-1}$$

Do Better

- Main idea (from previous slide): For fixed m, if we knew

$$\text{MLdegree}(X_{m2}), \text{MLdegree}(X_{m3}), \ldots, \text{MLdegree}(X_{mm})$$

  then we can solve for $\Lambda_m = (\lambda_1, \ldots, \lambda_{m-1})$ thereby giving a closed form expression for $\text{MLdegree}(X_{mn})$ for all n.
- Main idea (this slide): Recursively determine Λ_m thereby giving a closed form formula for $\text{MLdegree}(X_{mn})$ for fixed m but *any* n.
  - Note $\text{MLdegree}(X_{mn}) = \text{MLdegree}(X_{nm})$.
  - Prove λ_{m-1} of Λ_m is $(m-1)m!$.
- Closed form expressions for fixed m and $n \geq m$:

$$\text{MLdeg} X_{4n} = 25 \cdot 1^{n-1} - 40 \cdot 2^{n-1} + 23 \cdot 3^{n-1}$$

$$\text{MLdeg} X_{5n} = -90 \cdot 1^{n-1} + 260 \cdot 2^{n-1} - 270 \cdot 3^{n-1} + 96 \cdot 4^{n-1}$$

$$\text{MLdeg} X_{6n} = 301 \cdot 1^{n-1} - 1400 \cdot 2^{n-1} + 2520 \cdot 3^{n-1} - 2016 \cdot 4^{n-1} + 600 \cdot 5^{n-1}$$

Do Better

- Main idea (from previous slide): For fixed m, if we knew

$$\text{MLdegree}(X_{m2}), \text{MLdegree}(X_{m3}), \ldots, \text{MLdegree}(X_{mm})$$

  then we can solve for $\Lambda_m = (\lambda_1, \ldots, \lambda_{m-1})$ thereby giving a closed form expression for $\text{MLdegree}(X_{mn})$ for all n.
- Main idea (this slide): Recursively determine Λ_m thereby giving a closed form formula for $\text{MLdegree}(X_{mn})$ for fixed m but *any* n.
  - Note $\text{MLdegree}(X_{mn}) = \text{MLdegree}(X_{nm})$.
  - Prove λ_{m-1} of Λ_m is $(m-1)m!$.
- Closed form expressions for fixed m and $n \geq m$:

$$\text{MLdeg}X_{4n} = 25 \cdot 1^{n-1} - 40 \cdot 2^{n-1} + 23 \cdot 3^{n-1}$$

$$\text{MLdeg}X_{5n} = -90 \cdot 1^{n-1} + 260 \cdot 2^{n-1} - 270 \cdot 3^{n-1} + 96 \cdot 4^{n-1}$$

$$\text{MLdeg}X_{6n} = 301 \cdot 1^{n-1} - 1400 \cdot 2^{n-1} + 2520 \cdot 3^{n-1} - 2016 \cdot 4^{n-1} + 600 \cdot 5^{n-1}$$

Using Numerical Algebraic Geometry

Witness sets allow us to use parallelizable algorithms.

- Treat the u_{ij} as parameter values that we can adjust,
- If we have a set of critical points for generic data, then we can solve any specific instance of data quickly using a parameter homotopy.
- Critical points of ℓ_u for $u_{general}$ are taken to
  - critical points of ℓ_u for $u_{specific}$
  - by a parameter homotopy

$$\begin{bmatrix} \square & \square & \square & \square \\ \square & \square & \square & \square \\ \square & \square & \square & \square \\ \square & \square & \square & \square \end{bmatrix} ---> \begin{bmatrix} 160 & 8 & 16 & 24 \\ 32 & 200 & 16 & 8 \\ 8 & 24 & 176 & 32 \\ 16 & 40 & 8 & 232 \end{bmatrix}$$

$$191 \text{ points } ---> 191 \text{ points}$$

  - $\square$ denotes a random complex number.

Thank You

- Contact information
  - Jose Israel Rodriguez
  - JoIsRo@UChicago.edu
  - http://home.uchicago.edu/~joisro

Outline

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