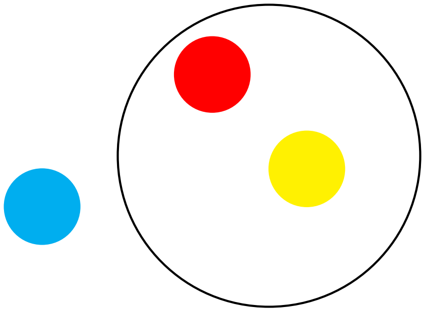
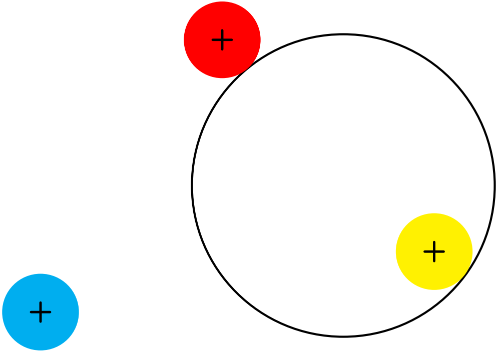
Likelihood Geometry of Determinantal Point Processes

Hannah Friedman (UC Berkeley)
joint work with Bernd Sturmfels and Maksym Zubkov

Algebra-Geometry-Combinatorics Afternoon at SFSU

November 29, 2023





Determinantal Point Processes

Definition

A *determinantal point process* is a random variable Z on the power set $2^{[n]}$ where

$$\mathbb{P}(Z = I) \sim \det(\Theta_I)$$

where Θ is an $n \times n$ symmetric positive definite matrix and Θ_I is the principal submatrix indexed by I.

Negative Correlation

Example

If $\Theta = \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{12} & \theta_{22} \end{pmatrix}$ governs a DPP Z on $\{\emptyset, \{1\}, \{2\}, \{1,2\}\}$, then we have

$$\mathbb{P}(\emptyset) \sim 1 \qquad \mathbb{P}(\{1\}) \sim \theta_{11}$$

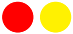
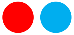
$$\mathbb{P}(\{2\}) \sim \theta_{22} \qquad \mathbb{P}(\{1,2\}) \sim \theta_{11}\theta_{22} - \theta_{12}^2$$

Since $\mathbb{P}(\{1\})\mathbb{P}(\{2\}) \geq \mathbb{P}(\{1,2\})$, the indicator variables on 1 and 2 being in the chosen subset are *negatively correlated*.

The Question

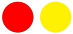
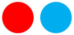
Subsets	Number of Observations
$\emptyset$	1
🔴	8
🟡	22
🔵	18
🔴🟡	151
🔴🔵	135
🟡🔵	360
🔴🟡🔵	2412

The Question

Subsets	Number of Observations
$\emptyset$	1
	8
	22
	18
	151
	135
	360
	2412

What matrix $\hat{\Theta}$ best explains this data?

The Question

Subsets	Number of Observations
$\emptyset$	1
	8
	22
	18
	151
	135
	360
	2412

What matrix $\hat{\Theta}$ best explains this data?

$$\begin{pmatrix} 8 & 5 & 3 \\ 5 & 22 & 6 \\ 3 & 6 & 18 \end{pmatrix} \begin{pmatrix} 8 & -5 & -3 \\ -5 & 22 & 6 \\ -3 & 6 & 18 \end{pmatrix}$$

$$\begin{pmatrix} 8 & -5 & 3 \\ -5 & 22 & -6 \\ 3 & -6 & 18 \end{pmatrix} \begin{pmatrix} 8 & 5 & -3 \\ 5 & 22 & -6 \\ -3 & -6 & 18 \end{pmatrix}$$

Likelihood Equation

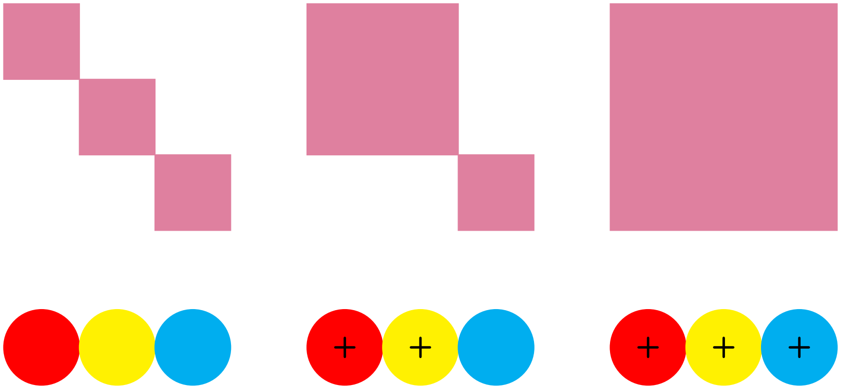
Definition

The likelihood equation for a DPP is

$$L_u = \sum_{I \subseteq [n]} u_I \log\det \Theta_I - \left(\sum_{I \subseteq [n]} u_I \right) \log\det(\Theta + Id_n)$$

where Id_n is the $n \times n$ identity matrix.

Extraneous Solutions



Previous Work

Theorem (Brunel-Moitra-Rigollet-Urschel, 2017)

The critical points with proper block structure are saddle points.

Conjecture (Brunel-Moitra-Rigollet-Urschel, 2017)

Up to sign changes, L_u has exactly one critical point for every possible block structure of an $n \times n$ matrix.

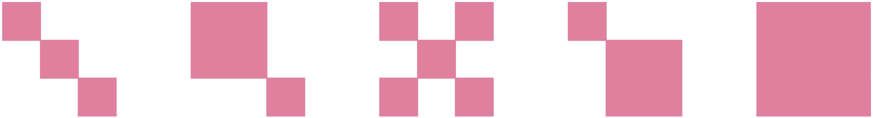
Previous Work

Theorem (Brunel-Moitra-Rigollet-Urschel, 2017)

The critical points with proper block structure are saddle points.

Conjecture (Brunel-Moitra-Rigollet-Urschel, 2017)

Up to sign changes, L_u has exactly one critical point for every possible block structure of an $n \times n$ matrix.



An Algebraic Geometer's Perspective

Implicit Likelihood Equation

The number of solutions to the implicit formulation of the likelihood equation,

$$L_u = \sum_{I \subset [n]} u_I \log p_I,$$

is called the *maximum likelihood (ML) degree* of the model.

Example

The ML degree of the model is 13 when $n = 3$.

An Algebraic Geometer's Perspective

Theorem (Holtz-Sturmfels, 2007)

The principal minors of a symmetric matrix must satisfy the hyperdeterminantal relations.

Example

When $n = 3$, the variety is cut out by the quartic

$$\begin{aligned}\mathrm{Det} = {} & p_{000}^2 p_{111}^2 + p_{001}^2 p_{110}^2 + p_{011}^2 p_{100}^2 + p_{010}^2 p_{101}^2 + 4p_{000}p_{011}p_{101}p_{110} + 4p_{001}p_{010}p_{100}p_{111} \\ & - 2p_{000}p_{001}p_{110}p_{111} - 2p_{000}p_{010}p_{101}p_{111} - 2p_{000}p_{011}p_{100}p_{111} \\ & - 2p_{001}p_{010}p_{101}p_{110} - 2p_{001}p_{011}p_{100}p_{110} - 2p_{010}p_{011}p_{100}p_{101}.\end{aligned}$$

Computing ML Degree

```
using Combinatorics
using HomotopyContinuation
using LinearAlgebra
```
✓ 4.8s

```
@var a,b,c,d,e,f
A = [a b c; b d e; c e f]
I = UniformScaling(1);

minors = cat([1], [det(A[s,s]) for s in powerset(1:size(A)[1], 1)], [det(I + A)], dims = (1,1))

@var u[1:2^3]
coefficients = cat(u, -sum(u), dims = (1,1))

phi = sum(coefficients[i] * log(minors[i]) for i in 1:2^3 + 1)
F = System(differentiate(phi,[a,b,c,d,e,f]); parameters = u)

solns = monodromy_solve(F)
```
✓ 0.2s

```
MonodromyResult
===============
• return_code → :heuristic_stop
• 52 solutions
• 416 tracked loops
• random_seed → 0xc587fb4e
```

Birational Reparametrization

$$\begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{12} & \theta_{22} & \theta_{23} \\ \theta_{13} & \theta_{23} & \theta_{33} \end{pmatrix} \longrightarrow \begin{pmatrix} x_{11} & \sqrt{x}_{12} & \sqrt{x}_{13} \\ \sqrt{x}_{12} & x_{22} & x_{23}/\sqrt{x_{12}x_{13}} \\ \sqrt{x}_{13} & x_{23}/\sqrt{x_{12}x_{13}} & x_{33} \end{pmatrix}$$

Computing ML Degree: Second Attempt

```
using Combinatorics
using HomotopyContinuation
using LinearAlgebra
```

✓ 2.7s

```
@var a,b,c,d,e,f
A = [a sqrt(b) sqrt(c); sqrt(b) d e/(sqrt(b)*sqrt(c)); sqrt(c) e/(sqrt(b)*sqrt(c)) f]
I = UniformScaling(1);

minors = cat([1], [expand(det(A[s,s])) for s in powerset(1:size(A)[1], 1)], [expand(det(I + A))], dims = (1,1))

@var u[1:2^3]
coefficients = cat(u, -sum(u), dims = (1,1))

phi = sum(coefficients[i] * log(minors[i]) for i in 1:2^3 + 1)
F = System(differentiate(phi,[a,b,c,d,e,f]); parameters = u)

solns = monodromy_solve(F)
```

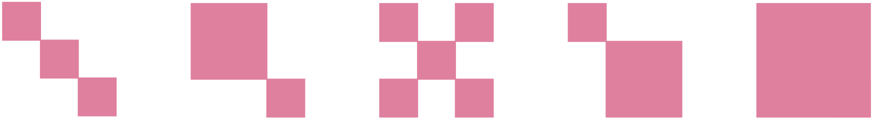
✓ 0.5s

```
MonodromyResult
===============
• return_code → :heuristic_stop
• 13 solutions
• 90 tracked loops
• random_seed → 0x23e760ad
```

Maximum Likelihood Degrees

n	ML Degree
1	1
2	1
3	13
4	3526
5	$>$30,000,000
6	???

Critical Points of the Parametric Likelihood Function: $n = 3$



μ_1^3 $\quad 2^1\mu_2 \cdot \mu_1$ $\quad 2^1\mu_2 \cdot \mu_1$ $\quad 2^1\mu_2 \cdot \mu_1$ $\quad 2^2\mu_3$

$$= 1 + 2 + 2 + 2 + 4 \cdot 13 = 59$$

Critical Points of the Parametric Likelihood Function

Theorem (F.-Sturmfels-Zubkov, 2023)

The critical points $\hat{\Theta}$ of the parametric log-likelihood function L_u are found by solving various likelihood equations on submodels $\mathcal{M}_4$ for $r \leq n$. If u is generic, then the total number of complex critical points of L_u equals

$$\sum_{\pi \in \mathcal{P}_n} \prod_{i=1}^{|\pi|} (2^{|\pi_i|-1} \mu_{|\pi_i|})$$

where $\mathcal{P}_n$ is the set of all set partitions of $[n]$.

Thank You