# Data-Discriminants of Likelihood Equations

Jose Israel Rodriguez[1] and Xiaoxian Tang[2]

[1]University of Notre Dame, United States of America

[2]National Institute For Mathematical Sciences (NIMS), Republic of Korea

ISSAC 2015
Bath, UK

# Motivation

## First Example

# Motivation

## First Example

Assume

- $p_i$ is probability of observing side $i$ ($i = 1, 2, 3, 4$)
- the die is unfair ($\Leftrightarrow$ $\exists j$ such that $p_j$ is not 25%)

# Motivation
## First Example

Assume

- $p_i$ is probability of observing side $i$ ($i = 1, 2, 3, 4$)
- the die is unfair ($\Leftrightarrow \exists j$ such that $p_j$ is not 25%)

**Given Constraints on $p_1$, $p_2$, $p_3$ and $p_4$**

- $\{(p_1, p_2, p_3, p_4) \in \mathbb{R}^4_{>0} | \Sigma^4_{i=1} p_i = 1\}$
- We artificially assume $p_1 + 2p_2 + 3p_3 - 4p4 = 0$

# Motivation

## First Example

Assume

- $p_i$ is probability of observing side $i$ ($i = 1, 2, 3, 4$)
- the die is unfair ($\Leftrightarrow \exists j$ such that $p_j$ is not 25%)

### Given Constraints on $p_1$, $p_2$, $p_3$ and $p_4$

- $\{(p_1, p_2, p_3, p_4) \in \mathbb{R}^4_{>0} | \Sigma^4_{i=1} p_i = 1\}$
- We artificially assume $p_1 + 2p_2 + 3p_3 - 4p4 = 0$

### Data Record

We toss the die 100 times and record the times of getting each side e.g. $[u_1 = 11, u_2 = 24, u_3 = 15, u_4 = 50]$

# Motivation
## First Example

Assume

- $p_i$ is probability of observing side $i$ ($i = 1, 2, 3, 4$)
- the die is unfair ($\Leftrightarrow \exists j$ such that $p_j$ is not 25%)

### Given Constraints on $p_1$, $p_2$, $p_3$ and $p_4$

- $\{(p_1, p_2, p_3, p_4) \in \mathbb{R}^4_{>0} | \Sigma^4_{i=1} p_i = 1\}$
- We artificially assume $p_1 + 2p_2 + 3p_3 - 4p4 = 0$

### Data Record

We toss the die 100 times and record the times of getting each side e.g. $[u_1 = 11, u_2 = 24, u_3 = 15, u_4 = 50]$

### Question

For given constraints and data, how to estimate $p_1$, $p_2$, $p_3$ and $p_4$ which BEST explains the data?

### Answer

Maximize likelihood function $p_1^{11} p_2^{24} p_3^{15} p_4^{50}$ subjected to given constraints

# Motivation

## First Example

**Question**

How to maximize likelihood function $p_1^{11} p_2^{24} p_3^{15} p_4^{50}$ subjected to given constraints?

# Motivation
## First Example

**Question**

How to maximize likelihood function $p_1^{11}p_2^{24}p_3^{15}p_4^{50}$ subjected to given constraints?

**Answer.**

It is equivalent to maximize $\log(p_1^{11}p_2^{24}p_3^{15}p_4^{50})$. By the Lagrange Multiplier Method, we solve

$$\begin{aligned} p_1\lambda_1 + p_1\lambda_2 - 11 &= 0 \\ p_2\lambda_1 + 2p_2\lambda_2 - 24 &= 0 \\ p_3\lambda_1 + 3p_3\lambda_2 - 15 &= 0 \\ p_4\lambda_1 - 4p_4\lambda_2 - 50 &= 0 \\ p_1 + 2p_2 + 3p_3 - 4p_4 &= 0 \\ p_1 + p_2 + p_3 + p_4 - 1 &= 0 \end{aligned}$$

**Question**

How to maximize likelihood function $p_1^{11} p_2^{24} p_3^{15} p_4^{50}$ subjected to given constraints?

**Answer.**

It is equivalent to maximize $\log(p_1^{11} p_2^{24} p_3^{15} p_4^{50})$. By the Lagrange Multiplier Method, we solve

$$\begin{aligned} p_1\lambda_1 + \phantom{2}p_1\lambda_2 - 11 &= 0 \\ p_2\lambda_1 + 2p_2\lambda_2 - 24 &= 0 \\ p_3\lambda_1 + 3p_3\lambda_2 - 15 &= 0 \\ p_4\lambda_1 - 4p_4\lambda_2 - 50 &= 0 \\ p_1 + 2p_2 + 3p_3 - 4p_4 &= 0 \\ p_1 + p_2 + p_3 + p_4 - 1 &= 0 \end{aligned}$$

and get 3 solutions
$[p1 = 1.2691, p2 = -0.2903, p3 = -0.0862, p4 = 0.1075, \lambda_1 = 100, \lambda_2 = -91.3324]$,
$[p1 = 0.1857, p2 = 1.2980, p3 = -0.6737, p4 = 0.1901, \lambda_1 = 100, \lambda_2 = -40.7547]$,
$[p1 = 0.1232, p2 = 0.3057, p3 = 0.2214, p4 = 0.3497, \lambda_1 = 100, \lambda_2 = -10.7463]$.

# Motivation

## First Example

**Question**

How to maximize likelihood function $p_1^{u_1} p_2^{u_2} p_3^{u_3} p_4^{u_4}$ subjected to given constraints?

# Motivation

## First Example

**Question**

How to maximize likelihood function $p_1^{u_1} p_2^{u_2} p_3^{u_3} p_4^{u_4}$ subjected to given constraints?

**Answer.**

We solve

$$\begin{aligned}
p_1\lambda_1 + \phantom{2}p_1\lambda_2 - u_1 &= 0\\
p_2\lambda_1 + 2p_2\lambda_2 - u_2 &= 0\\
p_3\lambda_1 + 3p_3\lambda_2 - u_3 &= 0\\
p_4\lambda_1 - 4p_4\lambda_2 - u_4 &= 0\\
p_1 + 2p_2 + 3p_3 - 4p_4 &= 0\\
p_1 + p_2 + p_3 + p_4 - 1 &= 0
\end{aligned}$$

# Motivation

## First Example

**Question**

How to maximize likelihood function $p_1^{u_1}p_2^{u_2}p_3^{u_3}p_4^{u_4}$ subjected to given constraints?

**Answer.**

We solve

$$\begin{aligned} p_1\lambda_1 + p_1\lambda_2 - u_1 &= 0\\ p_2\lambda_1 + 2p_2\lambda_2 - u_2 &= 0\\ p_3\lambda_1 + 3p_3\lambda_2 - u_3 &= 0\\ p_4\lambda_1 - 4p_4\lambda_2 - u_4 &= 0\\ p_1 + 2p_2 + 3p_3 - 4p_4 &= 0\\ p_1 + p_2 + p_3 + p_4 - 1 &= 0 \end{aligned}$$

**Remark**

For general $[u_1, u_2, u_3, u_4]$, the system has 3 complex solutions.

# Motivation

## First Example

**Question**

For which $u_i$, the system has $0, \; 1, \; 2$ and 3 REAL/POSITIVE solutions?

# Motivation

## First Example

**Question**

For which $u_i$, the system has 0, 1, 2 and 3 REAL/POSITIVE solutions?

**Answer. Use real quantifier elimination/real root classification tools.**

For example, by `RealRootClassification` in `Maple2015` [C. Chen, J. H. Davenport, J. P. May, M. M. Maza, B. Xia and R. Xiao, 2010], for any $(u_1, u_2, u_3, u_4) \in \mathbb{R}^4_{>0}$,

- $D(u_1, u_2, u_3, u_4) > 0 \Rightarrow$ 3 distinct real solutions and 1 of them is positive;
- $D(u_1, u_2, u_3, u_4) < 0 \Rightarrow$ 1 real solution and it is positive.

where

$$D = u_1 u_2 u_3 u_4 (u_1 + u_2 + u_3 + u_4)(441 u_1^4 + 4998 u_1^3 u_2 + 20041 u_1^2 u_2^2 + 33320 u_1 u_2^3 + 19600 u_2^4 - 756 u_1^3 u_3 + 20034 u_1^2 u_2 u_3 + 83370 u_1 u_2^2 u_3 + 79800 u_2^3 u_3 - 5346 u_1^2 u_3^2 + 55890 u_1 u_2 u_3^2 + 119025 u_2^2 u_3^2 + 4860 u_1 u_3^3 + 76950 u_2 u_3^3 + 18225 u_3^4 - 1596 u_1^3 u_4 - 11116 u_1^2 u_2 u_4 - 17808 u_1 u_2^2 u_4 + 4480 u_2^3 u_4 + 7452 u_1^2 u_3 u_4 - 7752 u_1 u_2 u_3 u_4 + 49680 u_2^2 u_3 u_4 - 17172 u_1 u_3^2 u_4 + 71460 u_2 u_3^2 u_4 + 27540 u_3^3 u_4 + 2116 u_1^2 u_4^2 + 6624 u_1 u_2 u_4^2 - 4224 u_2^2 u_4^2 - 9528 u_1 u_3 u_4^2 + 15264 u_2 u_3 u_4^2 + 14724 u_3^2 u_4^2 - 1216 u_1 u_4^3 - 512 u_2 u_4^3 + 3264 u_3 u_4^3 + 256 u_4^4)$$

# Motivation

## First Example

**Question**

For which $u_i$, the system has 0, 1, 2 and 3 REAL/POSITIVE solutions?

**Answer. Use real quantifier elimination/real root classification tools.**

For example, by `RealRootClassification` in `Maple2015` [C. Chen, J. H. Davenport, J. P. May, M. M. Maza, B. Xia and R. Xiao, 2010], for any $(u_1, u_2, u_3, u_4) \in \mathbb{R}^4_{>0}$,

- $D(u_1, u_2, u_3, u_4) > 0 \Rightarrow$ 3 distinct real solutions and 1 of them is positive;
- $D(u_1, u_2, u_3, u_4) < 0 \Rightarrow$ 1 real solution and it is positive.

where

$$D = u_1 u_2 u_3 u_4 (u_1 + u_2 + u_3 + u_4)(441 u_1^4 + 4998 u_1^3 u_2 + 20041 u_1^2 u_2^2 + 33320 u_1 u_2^3 + 19600 u_2^4 - 756 u_1^3 u_3 + 20034 u_1^2 u_2 u_3 + 83370 u_1 u_2^2 u_3 + 79800 u_2^3 u_3 - 5346 u_1^2 u_3^2 + 55890 u_1 u_2 u_3^2 + 119025 u_2^2 u_3^2 + 4860 u_1 u_3^3 + 76950 u_2 u_3^3 + 18225 u_3^4 - 1596 u_1^3 u_4 - 11116 u_1^2 u_2 u_4 - 17808 u_1 u_2^2 u_4 + 4480 u_2^3 u_4 + 7452 u_1^2 u_3 u_4 - 7752 u_1 u_2 u_3 u_4 + 49680 u_2^2 u_3 u_4 - 17172 u_1 u_3^2 u_4 + 71460 u_2 u_3^2 u_4 + 27540 u_3^3 u_4 + 2116 u_1^2 u_4^2 + 6624 u_1 u_2 u_4^2 - 4224 u_2^2 u_4^2 - 9528 u_1 u_3 u_4^2 + 15264 u_2 u_3 u_4^2 + 14724 u_3^2 u_4^2 - 1216 u_1 u_4^3 - 512 u_2 u_4^3 + 3264 u_3 u_4^3 + 256 u_4^4)$$

**Question**

How to compute $D$ EFFICIENTLY?

# Maximum Likelihood Estimation Problem

**Algebraic Statistical Model**

$$X = \mathcal{V} \cap \Delta_n$$

where

$\mathcal{V}$: irreducible and generically reduced projective variety

$\{(p_0, \dots, p_n) \in \mathbb{C}^{n+1} | g_1(p_0, \dots, p_n) = 0, \dots, g_s(p_0, \dots, p_n) = 0\}$

$\Delta_n$: probability simplex

$\{(p_0, \dots, p_n) \in \mathbb{R}^{n+1}_{>0} | \Sigma_{i=0}^{n} p_i = 1\}$

# Maximum Likelihood Estimation Problem

## Algebraic Statistical Model

$$X = \mathcal{V} \cap \Delta_n$$

where

$\mathcal{V}$: irreducible and generically reduced projective variety

$\{(p_0, \ldots, p_n) \in \mathbb{C}^{n+1} | g_1(p_0, \ldots, p_n) = 0, \ldots, g_s(p_0, \ldots, p_n) = 0\}$

$\Delta_n$: probability simplex

$\{(p_0, \ldots, p_n) \in \mathbb{R}^{n+1}_{>0} | \Sigma_{i=0}^{n} p_i = 1\}$

## Data Vector

$$[u_0, u_1, \ldots, u_n]$$

# Maximum Likelihood Estimation Problem

## Algebraic Statistical Model

$$X = \mathcal{V} \cap \Delta_n$$

where

$\mathcal{V}$: irreducible and generically reduced projective variety

$\{(p_0, \ldots, p_n) \in \mathbb{C}^{n+1} | g_1(p_0, \ldots, p_n) = 0, \ldots, g_s(p_0, \ldots, p_n) = 0\}$

$\Delta_n$: probability simplex

$\{(p_0, \ldots, p_n) \in \mathbb{R}^{n+1}_{>0} | \Sigma_{i=0}^{n} p_i = 1\}$

## Data Vector

$$[u_0, u_1, \ldots, u_n]$$

## Maximum Likelihood Estimation Problem

For given model and data, how to estimate $p_0, \ldots, p_n$ which BEST explains the data?

## Method

Maximize likelihood function $\Pi_{i=0}^{n} p_i^{u_i}$ subject to the algebraic statistical model.

# Lagrange Likelihood Equations

**Question**

How to maximize likelihood function $\Pi_{i=0}^{n} p_i^{u_i}$ subject to the algebraic statistical model $\mathcal{V}(g_1, \ldots, g_s) \cap \Delta_n$?

**Answer**

For every critical point $(p_0, \ldots, p_n)$ of the likelihood function, there exists $(\lambda_1, \ldots, \lambda_{s+1}) \in \mathbb{C}^{s+1}$ such that $(p_0, \ldots, p_n, \lambda_1, \ldots, \lambda_{s+1})$ is a solution to the Lagrange likelihood equations [S. Hosten, A. Khetan and B. Sturmfels, 2005; E. Gross and J. I. Rodriguez, 2014]:

$$\begin{aligned}
F_0 &= p_0\left(\lambda_1 + \frac{\partial g_1}{\partial p_0}\lambda_2 + \cdots + \frac{\partial g_s}{\partial p_0}\lambda_{s+1}\right) - u_0 = 0\\
&\ldots\\
F_n &= p_n\left(\lambda_1 + \frac{\partial g_1}{\partial p_n}\lambda_2 + \cdots + \frac{\partial g_s}{\partial p_n}\lambda_{s+1}\right) - u_n = 0\\
F_{n+1} &= g_1(p_0, \ldots, p_n) = 0\\
&\ldots\\
F_{n+s} &= g_s(p_0, \ldots, p_n) = 0\\
F_{n+s+1} &= p_0 + \cdots + p_n - 1 = 0
\end{aligned}$$

where

- $p_0, \ldots, p_n, \lambda_1, \ldots, \lambda_{s+1}$ are unknowns,
- $u_0, \ldots, u_n$ are parameters.

# Real/Positive Root Classification Problem

**Theorem 1** (System of Lagrange likelihood equations is generically zero-dimensional)[S. Hosten, A. Khetan and B. Sturmfels, 2005]

For a given algebraic statistical model, for general data $(u_0, \ldots, u_n)$, the number of complex solutions of Lagrange likelihood equations is a non-negative constant (ML-Degree).

# Real/Positive Root Classification Problem

Theorem 1 (System of Lagrange likelihood equations is generically zero-dimensional)[S. Hosten, A. Khetan and B. Sturmfels, 2005]

For a given algebraic statistical model, for general data $(u_0, \ldots, u_n)$, the number of complex solutions of Lagrange likelihood equations is a non-negative constant (ML-Degree).

Real/Positive Root Classification Problem

Classify $(u_0, \ldots, u_n)$ according to the number of real/positive solutions of Lagrange likelihood equations.

# Real/Positive Root Classification Problem

Theorem 1 (System of Lagrange likelihood equations is generically zero-dimensional)[S. Hosten, A. Khetan and B. Sturmfels, 2005]

For a given algebraic statistical model, for general data $(u_0, \dots, u_n)$, the number of complex solutions of Lagrange likelihood equations is a non-negative constant (ML-Degree).

Real/Positive Root Classification Problem

Classify $(u_0, \dots, u_n)$ according to the number of real/positive solutions of Lagrange likelihood equations.

Standard Method for Real/Positive Root Classification [L. Yang, X. Hou and B. Xia, 2001; D. Lazard and F. Rouillier, 2005; C. Chen, J. H. Davenport, J. P. May, M. M. Maza, B. Xia and R. Xiao, 2010]

- **Step 1** Compute discriminant variety (REMARK: generally discriminant variety is not a hypersurface [D. Lazard and F. Rouillier, 2005])

# Real/Positive Root Classification Problem

**Theorem 1 (System of Lagrange likelihood equations is generically zero-dimensional)[S. Hosten, A. Khetan and B. Sturmfels, 2005]**

For a given algebraic statistical model, for general data $(u_0, \ldots, u_n)$, the number of complex solutions of Lagrange likelihood equations is a non-negative constant (ML-Degree).

**Real/Positive Root Classification Problem**

Classify $(u_0, \ldots, u_n)$ according to the number of real/positive solutions of Lagrange likelihood equations.

**Standard Method for Real/Positive Root Classification [L. Yang, X. Hou and B. Xia, 2001; D. Lazard and F. Rouillier, 2005; C. Chen, J. H. Davenport, J. P. May, M. M. Maza, B. Xia and R. Xiao, 2010]**

- **Step 1** Compute discriminant variety (REMARK: generally discriminant variety is not a hypersurface [D. Lazard and F. Rouillier, 2005])
- **Step 2** Compute cells determined by discriminant variety and number of real/positive solutions over each cell [Tarski, 1951; Collins, 1975; Arnon et al., 1988; McCallum, 1988, 1999, 2001; Grigoriev, 1988; Collins and Hong, 1991; Renegar, 1992; Basu et al., 1996, 1999, 2006; Dolzmann and Sturm 1997; Brown, 2001, 2012, 2013, 2015; McCallum and Brown 2005; Strzebonski, 2000, 2005, 2006, 2011; Hong and Safey El Din, 2012; Bradford et al., 2013; M. England et al. 2015; R. Fukasaku et al. 2015...]

# Data-Discriminant and Problem Statement

**Proposition** (See propositions 1–2 in [J. I. Rodriguez and X. Tang, 2015].)

Discriminant varieties of Lagrange likelihood equations are projective varieties.

# Data-Discriminant and Problem Statement

Proposition (See propositions 1–2 in [J. I. Rodriguez and X. Tang, 2015].)

Discriminant varieties of Lagrange likelihood equations are projective varieties.

Data-Discriminant (Remark that we do need some extra assumptions for this definition. See Definition 5 in [J. I. Rodriguez and X. Tang, 2015].)

For a given algebraic statistics model $X$, the homogeneous polynomial that generates the reduced codimension 1 component of discriminant variety of Lagrange likelihood equations is said to be data-discriminant of Lagrange likelihood equations of $X$.

# Data-Discriminant and Problem Statement

**Proposition** (See propositions 1–2 in [J. I. Rodriguez and X. Tang, 2015].)

Discriminant varieties of Lagrange likelihood equations are projective varieties.

**Data-Discriminant** (Remark that we do need some extra assumptions for this definition. See Definition 5 in [J. I. Rodriguez and X. Tang, 2015].)

For a given algebraic statistics model $X$, the homogeneous polynomial that generates the reduced codimension 1 component of discriminant variety of Lagrange likelihood equations is said to be data-discriminant of Lagrange likelihood equations of $X$.

**Problem Statement: Design Algorithm**

- **Input:** Lagrange likelihood equations
- **Output:** Data-Discriminant

# Algorithm 1 (Standard Algorithm)

# Algorithm 1 (Standard Algorithm)

**Input.** $F = u_0 p^2 + u_1 p + u_2$,

# Algorithm 1 (Standard Algorithm)

**Input.** $F = u_0 p^2 + u_1 p + u_2,$
$J = 2u_0 p + u_1$

# Algorithm 1 (Standard Algorithm)

**Input.** $F = u_0 p^2 + u_1 p + u_2$,
$J = 2u_0 p + u_1$

**Step 1.** Compute the generators of the elimination ideal $\langle F, J\rangle \cap \mathbb{Q}[u_0, u_1, u_2]$

$$\{{u_1}^2 - 4u_0 u_2\}$$

# Algorithm 1 (Standard Algorithm)

**Input.** $F = u_0 p^2 + u_1 p + u_2$,
$J = 2u_0 p + u_1$

**Step 1.** Compute the generators of the elimination ideal $\langle F, J\rangle \cap \mathbb{Q}[u_0, u_1, u_2]$

$$\{{u_1}^2 - 4u_0 u_2\}$$

**Step 2.** Compute the codimension 1 component of the equidimensional radical decomposition of $\langle {u_1}^2 - 4u_0 u_2\rangle$

$${u_1}^2 - 4u_0 u_2$$

**Output.** ${u_1}^2 - 4u_0 u_2$

# Algorithm 2 (Probabilistic Algorithm)

**Input.** $F = u_0 p^2 + u_1 p + u_2$,
$J = 2u_0 p + u_1$

# Algorithm 2 (Probabilistic Algorithm)

**Input.** $F = u_0 p^2 + u_1 p + u_2$,
$J = 2u_0 p + u_1$

**Step 1 (Compute the degree and get the possible terms).** We assume our output is $D(u_0, u_1, u_2)$.

# Algorithm 2 (Probabilistic Algorithm)

**Input.** $F = u_0 p^2 + u_1 p + u_2$,
$J = 2u_0 p + u_1$

**Step 1 (Compute the degree and get the possible terms).** We assume our output is $D(u_0, u_1, u_2)$. Substitute

$$u_0 = 1 \cdot t + 11,$$

$$u_1 = 3 \cdot t + 2,$$

$$u_2 = 5 \cdot t + 6$$

(the red coefficients are "randomly" chosen)

into $F$, $J$

# Algorithm 2 (Probabilistic Algorithm)

**Input.** $F = u_0 p^2 + u_1 p + u_2$,
$J = 2u_0 p + u_1$

**Step 1 (Compute the degree and get the possible terms).** We assume our output is $D(u_0, u_1, u_2)$. Substitute

$$u_0 = 1 \cdot t + 11,$$

$$u_1 = 3 \cdot t + 2,$$

$$u_2 = 5 \cdot t + 6$$

(the red coefficients are "randomly" chosen)

into $F$, $J$ and compute the radical of the elimination ideal $\langle F(t,p), J(t,p)\rangle \cap \mathbb{Q}[t]$

$$\langle 11t^{\mathbf{2}} + 232t + 260\rangle$$

(that means $D(t+11, 3t+2, 5t+6) = 11t^2 + 232t + 260$)

So the total degree of $D$ is **2**.

# Algorithm 2 (Probabilistic Algorithm)

**Input.** $F = u_0 p^2 + u_1 p + u_2$,
$J = 2u_0 p + u_1$

**Step 1 (Compute the degree and get the possible terms).** We assume our output is $D(u_0, u_1, u_2)$. Substitute

$$u_0 = 1 \cdot t + 11,$$

$$u_1 = 3 \cdot t + 2,$$

$$u_2 = 5 \cdot t + 6$$

(the red coefficients are "randomly" chosen)

into $F$, $J$ and compute the radical of the elimination ideal $\langle F(t,p), J(t,p)\rangle \cap \mathbb{Q}[t]$

$$\langle 11t^{2} + 232t + 260\rangle$$

(that means $D(t+11, 3t+2, 5t+6) = 11t^2 + 232t + 260$)

So the total degree of $D$ is **2**. Similarly, we compute

$$degree(D, u_0) = 1, \quad degree(D, u_1) = 2 \quad degree(D, u_2) = 1$$

(so all the possible monomials in $D$ are ${u_1}^2, u_0 u_1, u_1 u_2, u_0 u_2$)

# Algorithm 2 (Probabilistic Algorithm)

**Step 2 (Evaluation/Interpolation).** Assume

$$D(u_0, u_1, u_2) = {u_1}^2 + (C_1 u_0 + C_2 u_2)u_1 + C_3 u_0 u_2 \tag{1}$$

# Algorithm 2 (Probabilistic Algorithm)

**Step 2 (Evaluation/Interpolation).** Assume

$$D(u_0, u_1, u_2) = {u_1}^2 + (C_1 u_0 + C_2 u_2) u_1 + C_3 u_0 u_2 \tag{1}$$

**Step 2.1.** Substitute $u_0 = 13$, $u_2 = 4$ into $F$, $J$ and compute the radical of the elimination ideal $\langle F(u_1, p), J(u_1, p)\rangle \cap \mathbb{Q}[u_1]$

$$\langle {u_1}^2 - 208\rangle \tag{2}$$

(that means $D(13, u_1, 4) = {u_1}^2 - 208$)

Comparing (1) and (2), we see

$$13C_1 + 4C_2 = 0 \tag{3}$$

and $52C_3 = -208$. Therefore, $C_3 = -4$.

(We need one more evaluation to solve $C_1$ and $C_2$)

# Algorithm 2 (Probabilistic Algorithm)

**Step 2 (Evaluation/Interpolation).** Assume

$$D(u_0, u_1, u_2) = {u_1}^2 + (C_1 u_0 + C_2 u_2) u_1 + C_3 u_0 u_2 \tag{1}$$

**Step 2.1.** Substitute $u_0 = 13$, $u_2 = 4$ into $F$, $J$ and compute the radical of the elimination ideal $\langle F(u_1, p), J(u_1, p)\rangle \cap \mathbb{Q}[u_1]$

$$\langle {u_1}^2 - 208 \rangle \tag{2}$$

(that means $D(13, u_1, 4) = {u_1}^2 - 208$)

Comparing (1) and (2), we see

$$13C_1 + 4C_2 = 0 \tag{3}$$

and $52C_3 = -208$. Therefore, $C_3 = -4$.

(We need one more evaluation to solve $C_1$ and $C_2$)

**Step 2.2.** Substitute $u_0 = 7$ and $u_2 = 3$ into $F$ and $J$. Similarly, we get

$$7C_1 + 3C_2 = 0 \tag{4}$$

By (3) and (4), $C_1 = C_2 = 0$.

**Output.** ${u_1}^2 - 4u_0 u_2$

# Algorithm 2 (Probabilistic Algorithm)

**Step 2 (Evaluation/Interpolation).** Assume

$$D(u_0, u_1, u_2) = {u_1}^2 + (C_1 u_0 + C_2 u_2)u_1 + C_3 u_0 u_2 \tag{1}$$

**Step 2.1.** Substitute $u_0 = 13$, $u_2 = 4$ into $F$, $J$ and compute the radical of the elimination ideal $\langle F(u_1, p), J(u_1, p)\rangle \cap \mathbb{Q}[u_1]$

$$\langle {u_1}^2 - 208\rangle \tag{2}$$

(that means $D(13, u_1, 4) = {u_1}^2 - 208$)

Comparing (1) and (2), we see

$$13C_1 + 4C_2 = 0 \tag{3}$$

and $52C_3 = -208$. Therefore, $C_3 = -4$.

(We need one more evaluation to solve $C_1$ and $C_2$)

**Step 2.2.** Substitute $u_0 = 7$ and $u_2 = 3$ into $F$ and $J$. Similarly, we get

$$7C_1 + 3C_2 = 0 \tag{4}$$

By (3) and (4), $C_1 = C_2 = 0$.

**Output.** ${u_1}^2 - 4u_0 u_2$

*Remark: The EVALUATION/INTERPOLATION idea is NOT the first time investigated. See [M. Giusti, G. Lecerf and B. Salvy, 2001; E. Schost, 2003].*

# Experiment

## Timings for Random Models (s: seconds; h: hours)

| Random Degree 2-models | | | Random Degree 3-models | | |
|---|---|---|---|---|---|
| Algorithm 1 | Algorithm 2 | | Algorithm 1 | Algorithm 2 | |
| | Strategy 1 | Strategy 2 | | Strategy 1 | Strategy 2 |
| 4.9s | 0.8s | 0.6s | >2h | 800.4s | 901.2s |
| 3.0s | 0.7s | 0.6s | >2h | 777.3s | 871.5s |
| 5.0s | 0.8s | 0.6s | >2h | 1428.9s | 1499.5s |
| 5.4s | 0.8s | 0.7s | >2h | 1118.9s | 1192.9s |
| 6.3s | 0.8s | 0.7s | >2h | 448.9s | 489.8s |
| 3.9s | 0.7s | 0.6s | >2h | 1279.6s | 1346.1s |
| 2.0s | 0.7s | 0.5s | >2h | 1286.5s | 1409.0s |
| 1.7s | 0.7s | 0.5s | >2h | 1605.9s | 1620.9s |
| 3.8s | 0.8s | 0.6s | >2h | 1099.4s | 1242.6s |
| 5.8s | 0.8s | 0.7s | >2h | 1229.0s | 1288.7s |

Algebra System: `Macaulay 2`

Processor: 3.2 GHz Inter Core i5 (8GB total memory)

Computer System: Mac OS X 10.9.3

# Experiment

## Timings for Literature Models (s: seconds; h: hours; d: days)

| Models | Algorithm 1 | Algorithm 2 | |
|---|---|---|---|
| | | Strategy 1 | Strategy 2 |
| **Example 3** | 11.1s | 5.3s | 6.4s |
| **Example 4** | 36446.4s | 360.2s | 56.3s |
| **Example 5** | >16h | >16h | 2768.2s |
| **Example 6** | >12d | >30d | 30d |

**Example 3 (Random Censoring [M. Drton, B. Sturmfels and S. Sullivant, 2009]).**

$$2p_0p_1p_2 + p_1^2p_2 + p_1p_2^2 - p_0^2p_{12} + p_1p_2p_{12}$$

**Example 4 ($3 \times 3$ Zero-Diagonal Matrix [E. Gross and J. I. Rodriguez, 2014]).**

$$\det \begin{bmatrix} 0 & p_{12} & p_{13} \\ p_{21} & 0 & p_{23} \\ p_{31} & p_{32} & 0 \end{bmatrix}$$

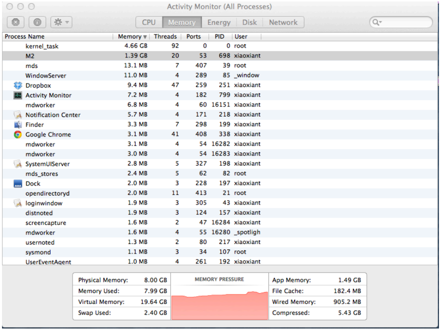
**Example 5 (Grassmannian of $2$-planes in $\mathbb{C}^4$ [S. Hosten, A, Khetan and B. Sturmfels, 2005]).**

$$p_{12}p_{34} - p_{13}p_{24} + p_{14}p_{23}$$

**Example 6 ($3 \times 3$ Symmetric Matrix Model [J. I. Rodriguez, 2014]).**

# Experiment

## Comparing Memory Pressure for Computing Example 6

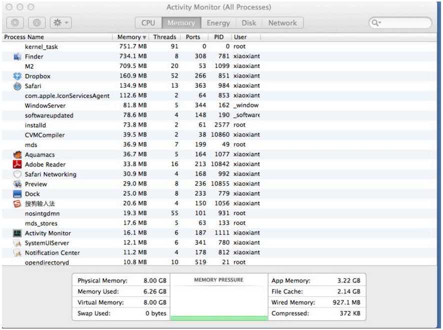


| Process Name | Memory ▾ | Threads | Ports | PID | User |
|---|---|---|---|---|---|
| kernel_task | 4.66 GB | 92 | 0 | 0 | root |
| M2 | 1.39 GB | 20 | 53 | 698 | xiaoxiant |
| mds | 13.1 MB | 7 | 407 | 39 | root |
| WindowServer | 11.0 MB | 4 | 289 | 85 | _window |
| Dropbox | 9.4 MB | 47 | 259 | 251 | xiaoxiant |
| Activity Monitor | 7.2 MB | 4 | 182 | 799 | xiaoxiant |
| mdworker | 6.8 MB | 4 | 60 | 16151 | xiaoxiant |
| Notification Center | 5.7 MB | 4 | 171 | 218 | xiaoxiant |
| Finder | 3.3 MB | 7 | 298 | 199 | xiaoxiant |
| Google Chrome | 3.1 MB | 41 | 408 | 338 | xiaoxiant |
| mdworker | 3.1 MB | 4 | 54 | 16282 | xiaoxiant |
| mdworker | 3.0 MB | 4 | 54 | 16283 | xiaoxiant |
| SystemUIServer | 2.8 MB | 5 | 327 | 198 | xiaoxiant |
| mds_stores | 2.4 MB | 5 | 62 | 82 | root |
| Dock | 2.0 MB | 3 | 228 | 197 | xiaoxiant |
| opendirectoryd | 2.0 MB | 11 | 413 | 21 | root |
| loginwindow | 1.9 MB | 3 | 305 | 43 | xiaoxiant |
| distnoted | 1.9 MB | 3 | 124 | 157 | xiaoxiant |
| screencapture | 1.6 MB | 2 | 47 | 16284 | xiaoxiant |
| mdworker | 1.6 MB | 4 | 55 | 16280 | _spotligh |
| usernoted | 1.3 MB | 2 | 80 | 217 | xiaoxiant |
| sysmond | 1.1 MB | 3 | 34 | 107 | root |
| UserEventAgent | 1.0 MB | 4 | 261 | 192 | xiaoxiant |

Physical Memory: 8.00 GB
Memory Used: 7.99 GB
Virtual Memory: 19.64 GB
Swap Used: 2.40 GB

App Memory: 1.49 GB
File Cache: 182.4 MB
Wired Memory: 905.2 MB
Compressed: 5.43 GB

Left: running standard algorithm after 3 days



| Process Name | Memory ▾ | Threads | Ports | PID | User |
|---|---|---|---|---|---|
| kernel_task | 751.7 MB | 91 | 0 | 0 | root |
| Finder | 734.1 MB | 8 | 308 | 781 | xiaoxiant |
| M2 | 709.5 MB | 20 | 53 | 1099 | xiaoxiant |
| Dropbox | 160.9 MB | 52 | 266 | 851 | xiaoxiant |
| Safari | 134.9 MB | 13 | 363 | 984 | xiaoxiant |
| com.apple.IconServicesAgent | 112.6 MB | 2 | 64 | 853 | xiaoxiant |
| WindowServer | 81.8 MB | 5 | 344 | 162 | _window |
| softwareupdated | 78.6 MB | 4 | 148 | 190 | _software |
| installd | 73.8 MB | 2 | 61 | 2577 | root |
| CVMCompiler | 39.5 MB | 2 | 38 | 10860 | xiaoxiant |
| mds | 36.9 MB | 7 | 199 | 49 | root |
| Aquamacs | 36.7 MB | 5 | 164 | 1077 | xiaoxiant |
| Adobe Reader | 33.8 MB | 16 | 213 | 10842 | xiaoxiant |
| Safari Networking | 30.9 MB | 4 | 168 | 992 | xiaoxiant |
| Preview | 29.0 MB | 8 | 236 | 10855 | xiaoxiant |
| Dock | 25.0 MB | 8 | 233 | 779 | xiaoxiant |
| 搜狗输入法 | 20.6 MB | 4 | 150 | 1056 | xiaoxiant |
| nosintgdmn | 19.3 MB | 55 | 101 | 931 | root |
| mds_stores | 17.6 MB | 5 | 63 | 133 | root |
| Activity Monitor | 16.1 MB | 6 | 187 | 1111 | xiaoxiant |
| SystemUIServer | 12.1 MB | 6 | 341 | 780 | xiaoxiant |
| Notification Center | 11.2 MB | 4 | 178 | 812 | xiaoxiant |
| opendirectoryd | 10.8 MB | 10 | 519 | 21 | root |

Physical Memory: 8.00 GB
Memory Used: 6.26 GB
Virtual Memory: 8.00 GB
Swap Used: 0 bytes

App Memory: 3.22 GB
File Cache: 2.14 GB
Wired Memory: 927.1 MB
Compressed: 372 KB

Right: running probabilistic algorithm after 3 days

# 3 × 3 Symmetric Matrix Model

A gambler has a coin and two pairs of three-sided dice. All the coin and dice are unfair. The two dice in the first pair have the same weights. The two dice in the second pair have the same weights.

# 3 × 3 Symmetric Matrix Model

A gambler has a coin and two pairs of three-sided dice. All the coin and dice are unfair. The two dice in the first pair have the same weights. The two dice in the second pair have the same weights.

He plays the same game 1000 rounds

Toss the coin.
- If the coin lands on side 1, toss the first pair of dice.
- If the coin lands on side 2, toss the second pair of dice.

# $3 \times 3$ Symmetric Matrix Model

A gambler has a coin and two pairs of three-sided dice. All the coin and dice are unfair. The two dice in the first pair have the same weights. The two dice in the second pair have the same weights.

He plays the same game 1000 rounds

Toss the coin.

- If the coin lands on side 1, toss the first pair of dice.
- If the coin lands on side 2, toss the second pair of dice.

After the 1000 rounds, he records the times of getting side i and side j with respect to the two dice every time he tosses

$$[u_{11}, u_{12}, u_{13}, u_{22}, u_{23}, u_{33}]$$

# $3 \times 3$ Symmetric Matrix Model

A gambler has a coin and two pairs of three-sided dice. All the coin and dice are unfair. The two dice in the first pair have the same weights. The two dice in the second pair have the same weights.

He plays the same game 1000 rounds

Toss the coin.
- If the coin lands on side 1, toss the first pair of dice.
- If the coin lands on side 2, toss the second pair of dice.

After the 1000 rounds, he records the times of getting side i and side j with respect to the two dice every time he tosses

$$[u_{11}, u_{12}, u_{13}, u_{22}, u_{23}, u_{33}]$$

Question

How to estimate the probability $p_{ij}$ of getting the sides $i$ and $j$ with respect to the two dice?

# $3 \times 3$ Symmetric Matrix Model

Assume that the probabilities of observing the sides 1 and 2 of the coin are $c_1$ and $c_2$, and the probabilities of observing the sides $1, 2$ and 3 of one die in the first and second pair are $[b_1, b_2, b_3]$ and $[r_1, r_2, r_3]$, respectively. We know

$$\begin{bmatrix} p_{11} & \frac{p_{12}}{2} & \frac{p_{13}}{2} \\ \frac{p_{12}}{2} & p_{22} & \frac{p_{23}}{2} \\ \frac{p_{13}}{2} & \frac{p_{23}}{2} & p_{33} \end{bmatrix} = c_1 \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} [b_1, b_2, b_3] + c_2 \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} [r_1, r_2, r_3]. \tag{5}$$

Therefore, the matrix on the left side has at most rank 2.

# 3 × 3 Symmetric Matrix Model

Assume that the probabilities of observing the sides 1 and 2 of the coin are $c_1$ and $c_2$, and the probabilities of observing the sides 1, 2 and 3 of one die in the first and second pair are $[b_1, b_2, b_3]$ and $[r_1, r_2, r_3]$, respectively. We know

$$\begin{bmatrix} p_{11} & \frac{p_{12}}{2} & \frac{p_{13}}{2} \\ \frac{p_{12}}{2} & p_{22} & \frac{p_{23}}{2} \\ \frac{p_{13}}{2} & \frac{p_{23}}{2} & p_{33} \end{bmatrix} = c_1 \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} [b_1, b_2, b_3] + c_2 \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} [r_1, r_2, r_3]. \tag{5}$$

Therefore, the matrix on the left side has at most rank 2. We have an algebraic statistical model below.

**3 × 3 Symmetric Matrix Model**

$$\mathcal{V}(g(p11, p12, p13, p22, p23, p33)) \cap \Delta_5,$$

where

$$g = det \begin{bmatrix} 2p11 & p12 & p13 \\ p12 & 2p22 & p23 \\ p13 & p23 & 2p33 \end{bmatrix},$$

$\Delta_5 = \{(p_{11}, \ldots, p_{33}) \in \mathbb{R}^6_{>0} | p11 + p12 + p13 + p22 + p23 + p33 = 1\}$

# $3 \times 3$ Symmetric Matrix Model

**Question**

How to maximize likelihood function

$$p_{11}^{u_{11}} p_{12}^{u_{12}} p_{13}^{u_{13}} p_{22}^{u_{22}} p_{23}^{u_{23}} p_{33}^{u_{33}}$$

subject to the algebraic statistical model $\mathcal{V}(g) \cap \Delta_5$?

**Answer.**

Solve the Lagrange likelihood equations

$$\begin{aligned}
F_0 &= p_{11}\lambda_1 + p_{11}\lambda_2(8p_{22}p_{33} - 2p_{23}^2) \quad - u_{11} = 0\\
F_1 &= p_{12}\lambda_1 + p_{12}\lambda_2(2p_{13}p_{23} - 4p_{12}p_{33}) - u_{12} = 0\\
F_2 &= p_{13}\lambda_1 + p_{13}\lambda_2(2p_{12}p_{23} - 4p_{13}p_{22}) - u_{13} = 0\\
F_3 &= p_{22}\lambda_1 + p_{22}\lambda_2(8p_{11}p_{33} - 2p_{13}^2) \quad - u_{22} = 0\\
F_4 &= p_{23}\lambda_1 + p_{23}\lambda_2(2p_{12}p_{13} - 4p_{11}p_{23}) - u_{23} = 0\\
F_5 &= p_{33}\lambda_1 + p_{33}\lambda_2(8p_{11}p_{22} - 2p_{12}^2) \quad - u_{33} = 0\\
F_6 &= g(p_{11}, p_{12}, p_{13}, p_{22}, p_{23}, p_{33}) = 0\\
F_7 &= p_{11} + p_{12} + p_{13} + p_{22} + p_{23} + p_{33} - 1 = 0
\end{aligned}$$

where

- $p_{11}, p_{12}, p_{13}, p_{22}, p_{23}, p_{33}$, $\lambda_1$ and $\lambda_2$ are unknowns
- $u_{11}$, $u_{12}$, $u_{13}$, $u_{22}$, $u_{23}$ and $u_{33}$ are parameters.

# $3 \times 3$ Symmetric Matrix Model

**Data-Discriminant (By Probabilistic Algorithm)**

- $\mathcal{D}_{X_p} = u_{11}u_{12}u_{13}u_{22}u_{23}u_{33}$
- $\mathcal{D}_{X_\infty} = (u_{11}+u_{22}+u_{33}+u_{12}+u_{13}+u_{23})(u_{11}+u_{22}+u_{12})(u_{11}+u_{33}+u_{13})(u_{22}+u_{33}+u_{23})(u_{12}+2u_{22}+u_{23})(u_{13}+2u_{33}+u_{23})(u_{13}+2u_{11}+u_{12})(8u_{11}u_{22}u_{33}-2u_{11}{u_{23}}^2-2{u_{12}}^2u_{33}+2u_{12}u_{13}u_{23}-2{u_{13}}^2u_{22})$.
- $\mathcal{D}_{X_J} = \underbrace{-64{u_{11}}^5{u_{22}}^3{u_{23}}^4+\ldots+{u_{13}}^4{u_{22}}^2{u_{23}}^6}_{1307\ \text{terms}}$

# $3 \times 3$ Symmetric Matrix Model

**Data-Discriminant (By Probabilistic Algorithm)**

- $\mathcal{D}_{X_p} = u_{11}u_{12}u_{13}u_{22}u_{23}u_{33}$
- $\mathcal{D}_{X_\infty} = (u_{11}+u_{22}+u_{33}+u_{12}+u_{13}+u_{23})(u_{11}+u_{22}+u_{12})(u_{11}+u_{33}+u_{13})(u_{22}+u_{33}+u_{23})(u_{12}+2u_{22}+u_{23})(u_{13}+2u_{33}+u_{23})(u_{13}+2u_{11}+u_{12})(8u_{11}u_{22}u_{33}-2u_{11}{u_{23}}^2-2{u_{12}}^2u_{33}+2u_{12}u_{13}u_{23}-2{u_{13}}^2u_{22})$.
- $\mathcal{D}_{X_J} = \underbrace{-64{u_{11}}^5{u_{22}}^3{u_{23}}^4+\ldots+{u_{13}}^4{u_{22}}^2{u_{23}}^6}_{\text{1307 terms}}$

**Real Root Classification**
**(Sample points of data-discriminant are computed by `RAGlib` [M. Safey El Din and E. Schost, 2003; H. Hong and M. Safey El Din, 2012; A. Greuet and M. Safey El Din, 2014])**

For $(u_{11},\ldots,u_{33}) \in \mathbb{R}^6_{>0}$, if $\mathcal{D}_{X_\infty}(u_{11},\ldots,u_{33}) \neq 0$, then

- $\mathcal{D}_{X_J}(u_{11},\ldots,u_{33}) > 0 \Rightarrow$ 6 distinct real solutions
- $\mathcal{D}_{X_J}(u_{11},\ldots,u_{33}) < 0 \Rightarrow$ 2 distinct real (positive) solutions.

# $3 \times 3$ Symmetric Matrix Model

### Data-Discriminant (By Probabilistic Algorithm)

– $\mathcal{D}_{X_p} = u_{11}u_{12}u_{13}u_{22}u_{23}u_{33}$
– $\mathcal{D}_{X_\infty} = (u_{11}+u_{22}+u_{33}+u_{12}+u_{13}+u_{23})(u_{11}+u_{22}+u_{12})(u_{11}+u_{33}+u_{13})(u_{22}+u_{33}+u_{23})(u_{12}+2u_{22}+u_{23})(u_{13}+2u_{33}+u_{23})(u_{13}+2u_{11}+u_{12})(8u_{11}u_{22}u_{33}-2u_{11}{u_{23}}^2-2{u_{12}}^2u_{33}+2u_{12}u_{13}u_{23}-2{u_{13}}^2u_{22})$.
– $\mathcal{D}_{X_J} = \underbrace{-64{u_{11}}^5{u_{22}}^3{u_{23}}^4+\ldots+{u_{13}}^4{u_{22}}^2{u_{23}}^6}_{1307 \text{ terms}}$

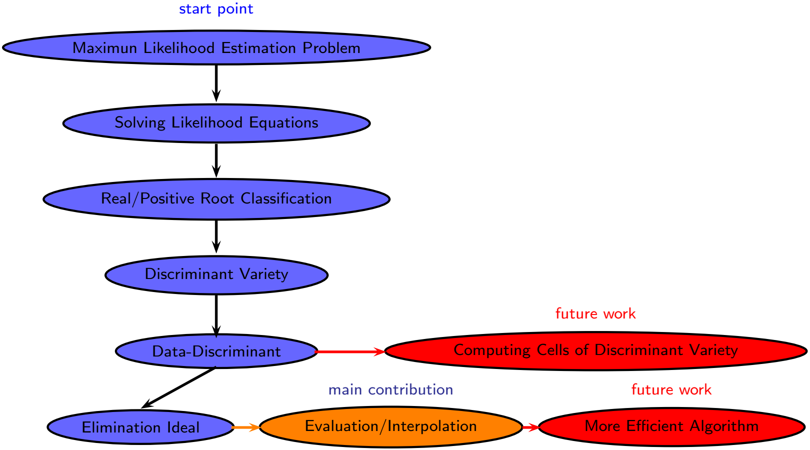
### Real Root Classification
(Sample points of data-discriminant are computed by `RAGlib` [M. Safey El Din and E. Schost, 2003; H. Hong and M. Safey El Din, 2012; A. Greuet and M. Safey El Din, 2014])

For $(u_{11},\ldots,u_{33}) \in \mathbb{R}^6_{>0}$, if $\mathcal{D}_{X_\infty}(u_{11},\ldots,u_{33}) \neq 0$, then
– $\mathcal{D}_{X_J}(u_{11},\ldots,u_{33}) > 0 \Rightarrow$ 6 distinct real solutions
– $\mathcal{D}_{X_J}(u_{11},\ldots,u_{33}) < 0 \Rightarrow$ 2 distinct real (positive) solutions.

### Remark. Sign of data-discriminant is NOT enough for classifying positive solutions.

– For data $(1, 1, \frac{280264116870825}{29514790517935282585 6}, 1, \frac{340890092055929220385 35}{14108069867573065075 9168}, \frac{32898355113670387769001}{1410806986757306507 59168})$, the system has 6 distinct positive solutions.
– For data $(1, 1, 199008, 30, 2022, 1)$, the system has also 6 real solutions but only 2 positive solutions

# Review



start point

Maximun Likelihood Estimation Problem

Solving Likelihood Equations

Real/Positive Root Classification

Discriminant Variety

Data-Discriminant

future work

Computing Cells of Discriminant Variety

Elimination Ideal

main contribution

Evaluation/Interpolation

future work

More Efficient Algorithm

Thank You for Your Attention!