A Probabilistic Algorithm for Computing Data-Discriminants of Likelihood Equations

Jose Israel Rodriguez and Xiaoxian Tang

June 13, 2016

Abstract

An algebraic approach to the maximum likelihood estimation problem is to solve a very structured parameterized polynomial system called likelihood equations that have finitely many complex (real or non-real) solutions. The only solutions that are statistically meaningful are the real solutions with positive coordinates. In order to classify the parameters (data) according to the number of real/positive solutions, we study how to efficiently compute the discriminants, say data-discriminants (DD), of the likelihood equations. We develop a probabilistic algorithm with three different strategies for computing DDs. Our implemented probabilistic algorithm based on Maple and FGb is more efficient than our previous version [27] and is also more efficient than the standard elimination for larger benchmarks. By applying RAGlib to a DD we compute, we give the real root classification of 3 by 3 symmetric matrix model.

Key Words: Maximum likelihood estimation, Likelihood equation, Discriminant

1 Introduction

In this work, we address the statistics problem of maximum likelihood estimation (MLE). A statistical model is a subset of $\mathbb{R}^{n+1}$ whose points have nonnegative coordinates summing to one; these points represent probability distributions. We will be interested in statistical models that are defined by zero sets of polynomials restricted to the positive orthant. The study of such models is a central part of algebraic statistics.

Given a data set and statistical model, we ask, “Which point in the model best describes the data?” One way to answer this question is by maximum likelihood estimation. By determining the point in the model that maximizes the likelihood function, one finds the probability distribution that best describes the data.

For the models we consider, the data is discrete; more specifically, the data is a list of integers representing counts. The likelihood function we consider on the model is given by the monomial

$$\ell_u(p) := p_0^{u_0} p_1^{u_1} \cdots p_n^{u_n},$$

where $u = (u_0, \ldots, u_n)$ is our data.

The maximum likelihood estimate is the point in the statistical model that maximizes the likelihood function. To determine this estimate, one can solve the likelihood equations to determine critical points of the likelihood function. This set of critical points contains all local maxima. So when the maximum likelihood estimate is a local maximum, it will be one of these critical points. When the maximum likelihood estimate is not a local maximum, then it is in the boundary (in the Euclidean topology) or the singular locus of the model. In those cases, one restricts to those loci’s Zariski closure and performs an analysis in this restriction. Iterating this procedure, one solves the global optimization problem of MLE.

The number of complex solutions to the likelihood equations (for general data) is a topological invariant called the maximum likelihood degree. For discrete models, the maximum likelihood degree
was shown to be an Euler-characteristic in the smooth case \cite{22} and recently shown to be a weighted sum of Euler characteristics in the singular case \cite{1}. In this paper, we will only consider discrete data. For a recent survey for the ML degree, one can refer to \cite{23}.

The ML degree measures the algebraic complexity of maximum likelihood estimation by counting the number of complex solutions (real and nonreal solutions) to the likelihood equations. In statistics only real solutions with positive coordinates are relevant, and to analyze the real root classification problem, one studies a discriminant of the system. Discriminants are a large topic studied in algebraic geometry, and they are important because they are used in combination with Ehresmann’s Theorem (Theorem \cite{2}) for real root classification. In principle, discriminants can be computed with Gröbner bases to determine elimination ideals \cite[Chapter 3]{7} or with geometric resolutions \cite{13}. In this paper we give a probabilistic interpolation algorithm to determine discriminants of the likelihood equations. These interpolation methods can be applied to other parameterized system of equations, for example determining the Euclidean distance degree, which was defined in \cite{8}.

Our main contribution is a probabilistic algorithm (Algorithm \cite{2}) that is more efficient for larger-sized models than the standard algorithm (Algorithm \cite{1}). Tables \cite{1,2} show our contribution is nontrivial with great performance increases in particular for Model 4. Based on the discriminants we have computed by the probabilistic algorithm, we give the real root classification for 3 by 3 symmetric matrix model. The main differences between this paper and our ISSAC’15 paper \cite{27} are listed below.

- We have reimplemented \cite{27} Algorithms 1–2 based on Maple and FGb. Comparing \cite{27} Tables 1–2 in Section 4 we see the new implementation performs much better. A typical example is Model 1\cite{27}. Statements the old implementation of Algorithm 2 (Strategy 2) takes 30 days to compute the discriminant while our new implementation takes less than 30 minutes.

- In Section 3 we add Section 3.2.3 in which we discuss the motivations and effectiveness of different strategies. We add Strategy 3 which improves the efficiency of sampling over Strategy 2 in some cases. We also add Lemma 3 in Section 3.2.1 which guarantees the correctness of Strategy 3.

- In Section 4 we present the new timings in Tables 1–2 based on our new implementation. We try more examples such as Dense Models 3–6 and Models 5–9. We add Table 3 which compares Strategy 2 and Strategy 3.

Now we motivate our study with an illustrating example.

Example 1 (Illustrative Example). Suppose we have a weighted four-sided die such that the probability \( p_i \) of observing side \( i \) \((i = 0, 1, 2, 3)\) of the die satisfies the constraint \( p_0 + 2p_1 + 3p_2 - 4p_3 = 0 \). We toss the die 1000 times and record a probability vector \((u_0, u_1, u_2, u_3)\), where \( u_i \) is the number of times we observe the side \( i \). We want to determine the probability distribution \((p_0, p_1, p_2, p_3)\) \(\in\mathbb{R}^4_+\) that best explains the data subject to the constraint by maximum likelihood estimation:

Maximize the likelihood function \( p_0^{u_0} p_1^{u_1} p_2^{u_2} p_3^{u_3} \) subject to

\[
\begin{align*}
0 + 2p_1 + 3p_2 - 4p_3 &= 0, \\
p_0 + p_1 + p_2 + p_3 &= 1, \\
p_0 > 0, p_1 > 0, p_2 > 0, &\text{ and } p_3 > 0.
\end{align*}
\]

For some statistical models, the MLE problem can be solved by well known local hill climbing algorithms such as the expectation–maximization algorithm. However, the local method can fail if there is more than one local maximum. One way to remedy this, is by solving the system of likelihood equations \cite{1,2,3}:

\[
\begin{align*}
F_0 &= p_0\lambda_1 + p_0\lambda_2 - u_0 \\
F_1 &= p_1\lambda_1 + 2p_1\lambda_2 - u_1 \\
F_2 &= p_2\lambda_1 + 3p_2\lambda_2 - u_2 \\
F_3 &= p_3\lambda_1 - 4p_3\lambda_2 - u_3 \\
F_4 &= p_0 + 2p_1 + 3p_2 - 4p_3 \\
F_5 &= p_0 + p_1 + p_2 + p_3 - 1
\end{align*}
\]

where \( \lambda_1 \) and \( \lambda_2 \) are newly introduced indeterminates (Lagrange multipliers) for formulating the likelihood equations. More specifically, for given \((u_0, u_1, u_2, u_3)\), if \((p_0, p_1, p_2, p_3)\) is a critical point
of the likelihood function, then there exist numbers \( \lambda_1 \) and \( \lambda_2 \) such that \((p_0, p_1, p_2, p_3, \lambda_1, \lambda_2)\) is a solution of the polynomial system.

For general data \( u = (u_0, u_1, u_2, u_3) \), the likelihood equations have 3 complex solutions. However, only solutions with positive coordinates \( p_i \) are statistically meaningful. A solution with all positive \( p_i \) coordinates is said to be a positive solution. Real root classification (RRC) and positive root classification (PRC) are important problems:

For which \( u_i \), the polynomial system has 1, 2 and 3 real/positive solutions?

According to the theory of computational (real) algebraic geometry \( [35, 25] \), the number of (real/positive) solutions only changes when the data \( u_i \) goes across some “special” values (see Theorem \( 2 \)). The set of “special” \( u_i \) is a (projective) variety \( [25] \) Lemma 4 in \( (3 \) dimensional complex projective space) 4-dimensional complex space.

The number of real/positive solutions is uniform over each open connected component determined by the variety. In other words, the “special” \( u_i \) plays the similar role as the discriminant for univariate polynomials. The first step of RRC is calculating the “special” \( u_i \), leading to the discriminant problem:

How to effectively compute the “special” \( u_i \)?

Geometrically, the “special” \( u_i \) is a projection of a variety. So it can be computed by elimination. For instance, with the command \( \text{fgb} \_\text{basis} \_\text{elim} \) in Fgb \( [17] \), we compute that the “special” \( u_i \) in Example \( 2 \) form a hypersurface defined by a homogenous polynomial in \( u_0, u_1, u_2, u_3 \). However, for many MLE problems, the elimination computation is too expensive as shown in Tables \( 1 \) \& \( 2 \) in Section \( 3 \).

We remark that our work can be viewed as following the numerous efforts in applying computational algebraic geometry to tackle MLE and critical points problems \( [19, 3, 20, 31, 16, 12, 17, 23, 18, 26] \).

The paper is organized as follows. Section \( 2 \) proves the discriminant variety of likelihood equations is projective and gives the definition of data-discriminant. Section \( 3 \) presents the elimination algorithm and probabilistic algorithm (with three strategies). Section \( 4 \) shows the experimental results comparing the two algorithms and different strategies. Section \( 5 \) discusses the future work and gives the real root classification of 3 by 3 symmetric matrix model.

### 2 Definitions and preliminaries

In this section, we discuss how to define “data-discriminant”. We assume the reader is familiar with elimination theory \( [7] \) Chapter 3.

**Notation 1.** Let \( \mathbb{P} \) denote the projective closure of the complex numbers \( \mathbb{C} \). For homogeneous polynomials \( g_1, \ldots, g_s \) in \( \mathbb{Q}[p_0, \ldots, p_n] \), \( \mathcal{V}(g_1, \ldots, g_s) \) denotes the projective variety in \( \mathbb{P}^n \) defined by \( g_1, \ldots, g_s \). Let \( \Delta_n \) denote the \( n \)-dimensional probability simplex \( \{(p_0, \ldots, p_n) \in \mathbb{R}^{n+1} | p_0 > 0, \ldots, p_n > 0, p_0 + \cdots + p_n = 1 \} \).

**Definition 1.** (Algebraic Statistical Model and Model Invariant) The set \( \mathcal{M} \) is said to be an algebraic statistical model if \( \mathcal{M} = \mathcal{V}(g_1, \ldots, g_s) \cap \Delta_n \) where \( g_1, \ldots, g_s \) define an irreducible generically reduced projective variety. Each \( g_i \) is said to be a model invariant of \( \mathcal{M} \). If the codimension of \( \mathcal{V}(g_1, \ldots, g_s) \) is \( k \), we say \( \{h_1, \ldots, h_k\} \) is a set of general model invariants for \( \mathcal{M} \) whenever the variety \( \mathcal{V}(h_1, \ldots, h_k) \) has \( \mathcal{V}(g_1, \ldots, g_s) \) as an irreducible component.

For a given algebraic statistical model, there are several different ways to formulate the likelihood equations \( [19] \). In this section, we introduce the Lagrange likelihood equations and define the data-discriminant for this formulation. One can similarly define data-discriminants for other formulations of likelihood equations.

3
Notation 2. For any $f_1, \ldots, f_m$ in the polynomial ring $\mathbb{Q}[x_1, \ldots, x_k]$, $V_a(f_1, \ldots, f_m)$ denotes the affine variety in $\mathbb{C}^k$ defined by $f_1, \ldots, f_m$ and $(f_1, \ldots, f_m)$ denotes the ideal generated by $f_1, \ldots, f_m$. For an ideal $I$ in $\mathbb{Q}[x_1, \ldots, x_k]$, $V_a(I)$ denotes the affine variety defined by $I$.

Definition 2. (Lagrange Likelihood Equations and Correspondence) Given an algebraic statistical model $X$ of codimension $k$ and a set of general model invariants $\{h_1, \ldots, h_k\}$, the system of polynomials below is said to be the Lagrange likelihood equations of $X$ when set to zero:

\[
F_0 := p_0(\lambda_1 + \frac{\partial h_1}{\partial p_0} \lambda_2 + \cdots + \frac{\partial h_k}{\partial p_0} \lambda_{k+1}) - u_0 \\
\vdots \\
F_n := p_n(\lambda_1 + \frac{\partial h_1}{\partial p_n} \lambda_2 + \cdots + \frac{\partial h_k}{\partial p_n} \lambda_{k+1}) - u_n \\
F_{n+1} := g_1(p_0, \ldots, p_n) \\
\vdots \\
F_{n+s} := g_s(p_0, \ldots, p_n) \\
F_{n+s+1} := p_0 + \cdots + p_n - 1
\]

where $g_1, \ldots, g_s$ are the model invariants of $X$ and $u_0, \ldots, u_n, p_0, \ldots, p_n, \lambda_1, \ldots, \lambda_{k+1}$ are indeterminates (also denoted by $u, p, \Lambda$). More specifically, $p_0, \ldots, p_n, \lambda_1, \ldots, \lambda_{k+1}$ are unknowns and $u_0, \ldots, u_n$ are parameters.

$V_a(F_0, \ldots, F_{n+s+1})$, namely the set

\[
\{ (u, p, \Lambda) \in \mathbb{C}^{n+1} \times \mathbb{C}^{n+1} \times \mathbb{C}^{k+1} | F_0 = 0, \ldots, F_{n+s+1} = 0 \}
\]

is said to be the Lagrange likelihood correspondence of $X$ and denoted by $L_X$.

Notation 3. Let $\pi$ denote the canonical projection from the ambient space of the Lagrange likelihood correspondence to the $\mathbb{C}^{n+1}$ associated to the $u$ indeterminants $\pi: \mathbb{C}^{n+1} \times \mathbb{C}^{n+s+2} \to \mathbb{C}^{n+1}$. For any set $S$ in $\mathbb{C}^{n+1}$, $\mathcal{I}(S)$ denotes the ideal

\[
\{ D \in \mathbb{Q}[u] | D(a_0, \ldots, a_n) = 0, \forall (a_0, \ldots, a_n) \in S \}
\]

$\overline{S}$ denotes the affine closure of $S$ in $\mathbb{C}^{n+1}$, namely $V_a(\mathcal{I}(S))$.

Given an algebraic statistical model $M$ and a data vector $u \in \mathbb{R}_{\geq 0}^n$, the maximum likelihood estimation problem is to maximize the likelihood function $p_0^{a_0} \cdots p_n^{a_n}$ subject to $M$. The MLE problem can be solved by computing $\pi^{-1}(u) \cap L_X$. More specifically, if $p$ is a regular point of $V(g_1, \ldots, g_s)$, then $p$ is a critical point of the likelihood function if and only if there exists $\Lambda \in \mathbb{C}^{k+1}$ such that $(u, p, \Lambda) \in L_X$. Theorem 1 states that for a general data vector $u$, $\pi^{-1}(u) \cap L_X$ is a finite set and the cardinality of $\pi^{-1}(u) \cap L_X$ is constant over a dense Zariski open set, which inspires the definition of ML degree. For details, see [19].

Theorem 1. [19] For an algebraic statistical model $X$, there exists an affine variety $V \subset \mathbb{C}^{n+1}$ and a non-negative integer $N$ such that for any $u \in \mathbb{C}^{n+1} \setminus V$,

\[
\# \pi^{-1}(u) \cap L_X = N.
\]

Definition 3. [19](ML Degree) For an algebraic statistical model $X$, the non-negative integer $N$ stated in Theorem 1 is said to be the ML degree of $X$.

Definition 4. For an algebraic statistical model $X$ with a set of general model invariants $\{h_1, \ldots, h_k\}$, suppose $F_0, \ldots, F_{n+s+1}$ are defined as in Definition 2. Then, we have the following:

- $L_{X_{\infty}}$ denotes the set of non-properness of $\pi$, i.e., the set of the $u \in \pi(L_X)$ such that there does not exist a compact neighborhood $U$ of $u$ where $\pi^{-1}(U) \cap L_X$ is compact;
- $L_{X_p}$ denotes $\pi(L_X \cap V_a(\mathbb{P}_{k=0}^n p_k))$; and
• \( \mathcal{L}_{\mathcal{X}, J} \) denotes \( \pi(\mathcal{L}_{\mathcal{X} \cap V_a(J)}) \) where \( J \) denotes the determinant below

\[
\begin{vmatrix}
\frac{\partial F_1}{\partial \mathbf{p}_0} & \cdots & \frac{\partial F_1}{\partial \mathbf{p}_n} & \cdots & \frac{\partial F_1}{\partial \lambda_1} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
\frac{\partial F_m}{\partial \mathbf{p}_0} & \cdots & \frac{\partial F_m}{\partial \mathbf{p}_n} & \cdots & \frac{\partial F_m}{\partial \lambda_1} \\
1 & \cdots & 1 & 0 & 0 \\
\frac{\partial h_1}{\partial \mathbf{p}_0} & \cdots & \frac{\partial h_1}{\partial \mathbf{p}_n} & \cdots & \frac{\partial h_1}{\partial \lambda_1} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
\frac{\partial h_{k+1}}{\partial \mathbf{p}_0} & \cdots & \frac{\partial h_{k+1}}{\partial \mathbf{p}_n} & \cdots & \frac{\partial h_{k+1}}{\partial \lambda_1}
\end{vmatrix}_{(n+k+2) \times (n+k+2)}.
\]

The geometric meaning of \( \mathcal{L}_{\mathcal{X}, p} \) and \( \mathcal{L}_{\mathcal{X}, J} \) are as follows. The first, \( \mathcal{L}_{\mathcal{X}, p} \), is the projection of the intersection of the Lagrange likelihood correspondence with the coordinate hyperplanes. The second, \( \mathcal{L}_{\mathcal{X}, J} \), is the projection of the intersection of the Lagrange likelihood correspondence with the hypersurface defined by \( J \). Geometrically, \( \mathcal{L}_{\mathcal{X}, J} \) is the closure of the union of the projection of the singular locus of \( \mathcal{L}_{\mathcal{X}} \) and the set of critical values of the restriction of \( \pi \) to the regular locus of \( \mathcal{L}_{\mathcal{X}} \) [25, Definition 2].

The Lagrange likelihood equations define an affine variety. As we continuously deform the parameters \( u_i \), coordinates of a solution can tend to infinity. Geometrically, \( \mathcal{L}_{\mathcal{X}, \infty} \) is the set of data-discriminant, we study the structures of \( \mathcal{L}_{\mathcal{X}, p} \), \( \mathcal{L}_{\mathcal{X}, J} \) and \( \mathcal{L}_{\mathcal{X}, \infty} \).

**Theorem 2.** For an algebraic statistical model \( X \). If \( \mathcal{C} \) is an open connected component of

\[
\mathbb{R}^{n+1}(\mathcal{L}_{\mathcal{X}, \infty} \cup \mathcal{L}_{\mathcal{X}, J}),
\]

then over \( u \in \mathcal{C} \), the following is constant:

\[
\#\pi^{-1}(u) \cap \mathcal{L}_{\mathcal{X}} \cap \mathbb{R}^{n+s+2}.
\]

Moreover, if \( \mathcal{C} \) is an open connected component of

\[
\mathbb{R}^{n+1}(\mathcal{L}_{\mathcal{X}, \infty} \cup \mathcal{L}_{\mathcal{X}, J} \cup \mathcal{L}_{\mathcal{X}, p}),
\]

then over \( u \in \mathcal{C} \), the following is constant:

\[
\#\pi^{-1}(u) \cap \mathcal{L}_{\mathcal{X}} \cap (\mathbb{R}^{n+1} \times \mathbb{R}^{s+1}).
\]

Before we give the definition of data-discriminant, we study the structures of \( \mathcal{L}_{\mathcal{X}, p} \), \( \mathcal{L}_{\mathcal{X}, J} \) and \( \mathcal{L}_{\mathcal{X}, \infty} \). Proposition 1 shows that the structure of the likelihood equations forces \( \mathcal{L}_{\mathcal{X}, p} \) to be contained in the union of coordinate hyperplanes defined by \( \prod_{k=0}^{n} u_k \). Proposition 2 shows that the structure of the likelihood equations forces \( \mathcal{L}_{\mathcal{X}, J} \\{ 0 \} \) to be a projective variety. Similar to the proof of Proposition 2 we can also show that the structure of the likelihood equations forces \( \mathcal{L}_{\mathcal{X}, \infty} \\{ 0 \} \) to be a projective variety.

**Proposition 1.** For any algebraic statistical model \( X \),

\[
\mathcal{L}_{\mathcal{X}, p} \subset \mathcal{V}_a(\prod_{k=0}^{n} u_k).
\]

5
**Proof.** By Definition 2 for any $k$ ($0 \leq k \leq n$),
\[ u_k = p_k(\lambda_1 + \frac{\partial g_1}{\partial p_1} \lambda_2 + \cdots + \frac{\partial g_s}{\partial p_1} \lambda_s + 1) - F_k. \]
Hence,
\[ u_k \in (F_k, p_k) \cap \mathbb{Q}[u_k] \subset (F_0, \ldots, F_{n+s+1}, p_k) \cap \mathbb{C}[u] \]
So
\[ V_a((F_0, \ldots, F_{n+s+1}, p_k) \cap \mathbb{C}[u]) \subset V_a(u_k) \]
By the Closure Theorem [7],
\[ V_a((F_0, \ldots, F_{n+s+1}, p_k) \cap \mathbb{C}[u]) = \overline{\pi(L_X \cap V_a(p_k))} \]
Therefore,
\[ L_{X_p} = \overline{\pi(L_X \cap V_a(\Pi_{k=0}^n p_k))} = \overline{\pi(L_X \cap \bigcup_{k=0}^n V_a(p_k))} \]
\[ \subset \bigcup_{k=0}^n V_a(u_k) \]
\[ = V_a(\Pi_{k=0}^n u_k). \] □

**Remark 1.** Generally, $L_{X_p} \neq V_a(\Pi_{k=0}^n u_k)$. For example, suppose the algebraic statistical model $V_a(\Pi_{k=0}^n p_k)$ is a homogeneous ideal in $\mathbb{C}^{n+1}$. Then $L_{X_p} = \emptyset \neq V_a(u_0 u_1)$.

**Notation 4.** $D_{X_p}$ denotes the product $\Pi_{k=0}^n u_k$.

**Proposition 2.** For an algebraic statistical model $X$, we have $L_{X_p} \setminus \{0\}$ is a projective variety in $\mathbb{P}^n$, where $0$ is the zero vector $(0, \ldots, 0)$ in $\mathbb{C}^{n+1}$.

**Proof.** By the formulation of the Lagrange likelihood equations, we can prove that $\mathcal{I}(\pi(L_X \cap V_a(J)))$ is a homogeneous ideal by the two basic facts below, which can be proved by Definition 2 and basic algebraic geometry arguments.

**F1.** For every $u$ in $\pi(L_X \cap V_a(J))$, each scalar multiple $\alpha u$ is also in $\pi(L_X \cap V_a(J))$.

**F2.** For any $S \subset \mathbb{C}^{n+1}$, if for any $u \in S$ and for any scalar $\alpha \in \mathbb{C}$, $\alpha u \in S$, then $\mathcal{I}(S)$ is a homogeneous ideal in $\mathbb{Q}[u]$.

That means the ideal $\mathcal{I}(\pi(L_X \cap V_a(J)))$ is generated by finitely many homogeneous polynomials $D_1, \ldots, D_m$. Therefore, $L_{X_J} = V_a(\mathcal{I}(\pi(L_X \cap V_a(J))) = V_a(D_1, \ldots, D_m)$. So $L_{X_J} \setminus \{0\} = V(D_1, \ldots, D_m) \subset \mathbb{P}^n$. □

**Notation 5.** For an algebraic statistical model $X$, we define the notation $D_{X_J}$ according to the codimension of $L_{X_J} \setminus \{0\}$ in $\mathbb{P}^n$.

- If the codimension is 1, then assume $V(D_1), \ldots, V(D_K)$ are the codimension 1 irreducible components in the minimal irreducible decomposition of $L_{X_J} \setminus \{0\}$ in $\mathbb{P}^n$ and $(D_1), \ldots, (D_K)$ are radical. $D_{X_J}$ denotes the homogeneous polynomial $\Pi_{j=1}^K D_j$.

- If the codimension is greater than 1, then our convention is to take $D_{X_J} = 1$.

Similarly, we use the notation $D_{X_\infty}$ to denote the projective variety $L_{X_\infty} \setminus \{0\}$. Now we define the “data-discriminant” of Lagrange likelihood equations.

**Definition 5.** (Data-Discriminant) For a given algebraic statistics model $X$, the homogeneous polynomial $D_{X_\infty} D_{X_J} D_{X_P}$ is said to be the data-discriminant (DD) of Lagrange likelihood equations of $X$ and denoted by $D_X$. 


Remark 2. Note that DD can be viewed as a generalization of the “discriminant” for univariate polynomials. So it is interesting to compare DD with border polynomial (BP) [33] and discriminant variety (DV) [24]. DD and BP are defined for general parametric polynomial systems. DD is defined for the likelihood equations but can be generalized to generic zero-dimensional systems. Generally, for any square and generic zero-dimensional system, \( \mathcal{V}_a(DD) \subset \mathcal{V}_a(BP) \). Note that due to the special structure of likelihood equations, DD is a homogenous polynomial despite being an affine system of equations. However, generally, DD is not a projective variety and BP is not homogenous.

Example 2 (Linear Model). The algebraic statistic model for the four-sided die story in Section 1 is given by
\[
\mathcal{M} = V(p_0 + 2p_1 + 3p_2 - 4p_3) \cap \Delta_3.
\]
The Lagrange likelihood equations are \( F_0 = 0, \ldots, F_5 = 0 \) shown in Example 1. The Lagrange likelihood correspondence is \( L_X = \mathcal{V}_a(F_0, \ldots, F_5) \subset \mathbb{C}^{10} \). If we choose generic \((u_0, u_1, u_2, u_3) \in \mathbb{C}^4 \), \( \pi^{-1}(u_0, u_1, u_2, u_3) \cap L_X = 3 \), namely the ML degree is 3. The data-discriminant is the product of \( \mathcal{D}_X \), \( \mathcal{D}_p \) and \( \mathcal{D}_J \), where
\[
\mathcal{D}_X = u_0 + u_1 + u_2 + u_3, \quad \mathcal{D}_p = u_0u_1u_2u_3, \quad \text{and}
\]
\[
\mathcal{D}_J = 441u_0^2 + 4998u_0u_1^2 + 20041u_0^3 + 33320u_0u_1u_2 + 19600u_1^2 - 756u_0u_2 + 20334u_0u_1u_2 + 83370u_0u_2u_3 + 79800u_2^2 - 5346u_0u_3 + 55890u_0u_1u_3 + 119025u_1u_2 + 4860u_0u_1u_2 + 76950u_2u_3 + 18225u_2u_3 - 1596u_0u_3 - 1116u_3^2 - 1780u_0u_2u_3 + 4480u_1u_2 - 7752u_0u_1u_3 + 349680u_1u_2u_3 - 17172u_0u_2u_3 + 71460u_1u_2u_3 + 27540u_2u_3 + 216u_0u_2^2 + 6624u_0u_2u_3^2 - 4224u_1u_2^2 - 9528u_0u_1u_3 + 15264u_1u_2u_3 + 256u_2u_3^2 + 14724u_3u_2 + 1216u_0u_3 - 512u_1u_3 + 3264u_3u_1 + 256u_3.
\]
By applying the well-known partial cylindrical algebraic decomposition (PCAD) [4] method to the data-discriminant above, we get that for any \((u_0, u_1, u_2, u_3) \in \mathbb{R}_{>0}^4 \),
\begin{itemize}
  \item if \( \mathcal{D}_J(u_0, u_1, u_2, u_3) > 0 \), then the system of likelihood equations has 3 distinct real solutions and 1 of them is positive;
  \item if \( \mathcal{D}_J(u_0, u_1, u_2, u_3) < 0 \), then the system of likelihood equations has exactly 1 real solution and it is positive.
\end{itemize}
The answer above can be verified by the RealRootClassification [33] [4] command in Maple 2015. In this example, the \( \mathcal{D}_X \) does not effect the number of real/positive solutions since it is always positive when each \( u_i \) is positive. However, generally, \( \mathcal{D}_X \) plays an important role in real root classification. Also remark that the real root classification is equivalent to the positive root classification for this example but it is not true generally (see the example discussed in Section 3).

3 Algorithm

In this section, we discuss how to compute the discriminant \( \mathcal{D}_X \). We assume that \( X \) is the closure of a given statistical model, \( F_0, \ldots, F_{n+1} \) are defined as in Definition 2 and \( J \) is defined as in Definition 3. We rename \( F_0, \ldots, F_{n+1} \) as \( F_0, \ldots, F_m \). We also rename \( p_0, \ldots, p_n, \lambda_1, \ldots, \lambda_s \) as \( y_0, \ldots, y_m \). Subsection 3.1 presents the standard algorithm for reference and Subsection 3.2 presents the probabilistic algorithm.

3.1 Standard Algorithm

Considering the data-discriminant as a projection drives a natural algorithm to compute it. This is the standard elimination algorithm in symbolic computation:
\begin{itemize}
  \item we compute \( L_X \) by elimination and then get \( \mathcal{D}_X \) by the radical equidimensional decomposition [29] Definition 3]. The algorithm is formally described in the Algorithm 1.
  \item we compute \( L_X \) by the Algorithm PROPERNESSDESCRIPTS presented in [29] and then get \( \mathcal{D}_X \) by the radical equidimensional decomposition. We omit the formal description of the algorithm.
\end{itemize}
Practically, Algorithm 1 may not terminate in a reasonable time before a computer reaches its memory limit. For instance, when using FGB for our Gröbner Bases computations the memory limit is reached due to the large size of the intermediate computational results. Since the algorithms for
Gröbner Bases in FGB are based on row echelon form computations [11, Section 2], we record the sizes of matrices generated by FGB for Models [1] [9] see Table 1 in Section 4. This motivates the probabilistic algorithm found in the next subsection.

### 3.2 Probabilistic Algorithm

In Section 3.2.1, we prepare the lemmas which are used in the algorithms in Sections 3.2.2–3.2.3. We present the probabilistic algorithm with Strategy 1 (Algorithm 2) in Section 3.2.2. We discuss two different strategies (Strategy 2 and Strategy 3) for Algorithm 2 in Section 3.2.3.

#### 3.2.1 Lemmas

We prepare the lemmas which are used in the Sections 3.2.2 and 3.2.3. Lemma 1 is used to linearly transform parameter space. Corollary 1 and Lemma 2 are used to compute the totally degree of \( D_{X_J} \). Corollary 2 is used in the sampling for interpolation. Lemma 3 is used to compute the degree of \( D_{X_J} \) and to do sampling in Strategy 3. In the statements of Lemmas 1–2 and Corollaries 1–2, we say an affine variety \( V \) in \( \mathbb{C}^n \) is non-trivial if \( \mathbb{C}^n \setminus V \neq \emptyset \).

**Lemma 1.** For any \( G \in \mathbb{Q}[u] \), there exists a non-trivial affine variety \( V \) in \( \mathbb{C}^n \) such that for any \( (a_1, \ldots, a_n) \in \mathbb{C}^n \setminus V \), the total degree of \( G \) equals the degree of \( B(t_0, t_1, \ldots, t_n) \) w.r.t. to \( t_0 \), where

\[
B(t_0, t_1, \ldots, t_n) = G(t_0, a_1 t_0 + t_1, \ldots, a_n t_0 + t_n)
\]

**Proof.** Suppose the total degree of \( G \) is \( d \) and \( G_d \) is the homogeneous component of \( G \) with total degree \( d \). For any \( (1, a_1, \ldots, a_n) \in \mathbb{C}^{n+1} \setminus V_0(G_d) \), let \( B(t_0, t_1, \ldots, t_n) = G(t_0, a_1 t_0 + t_1, \ldots, a_n t_0 + t_n) \). It is easily seen that the degree of \( B \) w.r.t. \( t_0 \) equals \( d \). □

**Corollary 1.** For any \( G \in \mathbb{Q}[u] \), there exists a non-trivial affine variety \( V \) in \( \mathbb{C}^{2n+2} \) such that for any

\[
(a_0, b_0, \ldots, a_n, b_n) \in \mathbb{C}^{2n+2} \setminus V,
\]

the total degree of \( G \) equals the degree of \( B(t) \) where

\[
B(t) = G(a_0 t + b_0, \ldots, a_n t + b_n).
\]

**Lemma 2.** There exists a non-trivial affine variety \( V \) in \( \mathbb{C}^{2n+2} \) such that for any \( (a_0, b_0, \ldots, a_n, b_n) \in \mathbb{C}^{2n+2} \setminus V \), if

\[
\langle A(t) \rangle = \langle F_0(t), \ldots, F_n(t), F_{n+1}, \ldots, F_m, J \rangle \cap \mathbb{Q}[t]
\]

where \( F_i(t) \) is the polynomial by replacing \( u_i \) with \( a_i t + b_i \) in \( F_i \) (\( i = 0, \ldots, n \)) and

\[
B(t) = D_{X_J}(a_0 t + b_0, \ldots, a_n t + b_n),
\]

then \( \langle B(t) \rangle = \sqrt{\langle A(t) \rangle} \).

**Proof.** By the definition of \( D_{X_J} \) (Notation 5), there exists an affine variety \( V_1 \) such that for any \( (a_0, b_0, \ldots, a_n, b_n) \in \mathbb{C}^{2n+2} \setminus V_1 \), \( \langle B(t) \rangle \) is radical. Thus, we only need to show that there exists an affine variety \( V_2 \) in \( \mathbb{C}^{2n+2} \) such that for any \( (a_0, b_0, \ldots, a_n, b_n) \in \mathbb{C}^{2n+2} \setminus V_2 \), \( \mathcal{V}_a(\langle B(t) \rangle) = \mathcal{V}_a(\langle A(t) \rangle) \).

---

**Algorithm 1: DX-J**

**input:** \( F_0, \ldots, F_m, J \)

**output:** \( D_{X_J} \)

1. \( G_u \leftarrow \) a set of generators of the elimination ideal \( \langle F_0, \ldots, F_n, J \rangle \cap \mathbb{Q}[u] \)

2. \( D_{X_J} \leftarrow \) the codimension 1 component of the equidimensional radical decomposition of \( \langle G_u \rangle \)

3. return \( D_{X_J} \)
Suppose $\pi_t$ is the canonical projection: $\mathbb{C} \times \mathbb{C}^{n+1} \to \mathbb{C}$. For any
\[ t^* \in \pi_t(V_a(F_0(t), \ldots, F_n(t), F_{n+1}, \ldots, F_m, J)), \]
let $u_i^* = a_i t^* + b_i$ (for $i = 0, \ldots, n$), then $(u_0^*, \ldots, u_n^*) \in \pi(L_X \cap V_a(J))$. Hence $D_{X,J}(u_0^*, \ldots, u_n^*) = 0$ and so $B(t^*) = 0$. Thus
\[ B(t) \in \mathcal{I}(\pi_t(V_a(F_0(t), \ldots, F_n(t), F_{n+1}, \ldots, F_m, J))). \]
Therefore,
\[ V_a(A(t)) = V_a(\mathcal{I}(\pi_t(V_a(F_0(t), \ldots, F_n(t), F_{n+1}, \ldots, F_m, J)))) \subset V_a(B(t)). \]
For any $t^* \in V_a((B(t)))$, let $u_i^* = a_i t^* + b_i$ for $i = 0, \ldots, n$, then $(u_0^*, \ldots, u_n^*) \in V_a(D_{X,J}) \subset L_X$. By the Extension Theorem [7], there exists an affine variety $V_2 \subset \mathbb{C}^{2n+2}$ such that if $(a_0, b_0, \ldots, a_n, b_n) \notin V_2$, then $(u_0^*, \ldots, u_n^*) \in \pi(L_X \cap V_a(J))$; thus
\[ t^* \in \pi_t(V_a(F_0(t), \ldots, F_n(t), F_{n+1}, \ldots, F_m, J)) \subset V_a(A(t)). \]

**Corollary 2.** There exists a non-trivial affine variety $V$ in $\mathbb{C}^n$ such that for any $(a_1, \ldots, a_n) \in \mathbb{C}^n \setminus V$, if
\[ \langle A(u_0) \rangle = \langle F_0, F_1^*, \ldots, F_{n+1}^*, F_m, J \rangle \cap \mathbb{Q}[u_0] \]
where $F_i^*$ is the polynomial by replacing $u_i$ with $a_i$ in $F_i$ ($i = 1, \ldots, n$) and
\[ B(u_0) = D_{X,J}(u_0, a_1, \ldots, a_n), \]
then $\langle B(u_0) \rangle = \sqrt{\langle A(u_0) \rangle}$.

**Lemma 3.** Suppose $X$ is an algebraic model with ML degree $N$. Let $g$ generate the reduced codimension 1 component of $\langle F_0, \ldots, F_m \rangle \cap \mathbb{Q}[u_0, \ldots, u_n, y_0]$. If $\text{deg}(g, y_0) = N$ and $D_{y_0} = \text{resultant}(g, \frac{\partial g}{\partial y_0}, y_0)$, then $D_{y_0} \in \langle D_{X,J} \rangle$.

**Proof.** By Notation[3] [\langle D_{X,J} \rangle] is radical. So we only need to show $\mathcal{V}(D_{X,J}) \subset \mathcal{V}(D_{y_0})$. Alternatively, we show $\mathbb{C}^{n+1} \setminus \mathcal{V}(D_{y_0}) \subset \mathbb{C}^{n+1} \setminus \mathcal{V}(D_{X,J})$. For any $u \in \mathbb{C}^{n+1} \setminus \mathcal{V}(D_{y_0})$, $g(u, y_0)$ has $N$ distinct complex solutions w.r.t $y_0$. By Projective Extension Theorem [7], Page 403, Chapter 8, Corollary 10], all the $N$ distinct complex solutions can be extended in $\mathbb{P}^{n+1}$. So the likelihood equations $F_0(u) = \ldots = F_m(u) = 0$ have $N$ distinct solutions in $\mathbb{P}^{n+1}$. Therefore, $u \notin \mathcal{V}(D_{X,J})$. \[ \square \]

### 3.2.2 Probabilistic Algorithm with Strategy 1

We first give an example for the main probabilistic algorithm. We will discuss different ways the main probabilistic algorithm can be adjusted in Section 3.2.3, so we say the main algorithm presented in this subsection is the probabilistic algorithm with Strategy 1.

**Example 3** (Algorithm[2] (Strategy 1)). Suppose the radical of the elimination ideal $\langle F, J \rangle \cap \mathbb{Q}[u]$ is generated by $\mathcal{D}(u_0, u_1, u_2, u_3)$, where $F = u_0p^3 + u_1p^2 + u_2p + u_3$ and $J = \frac{\partial F}{\partial p} = 3u_0p^2 + 2u_1p + u_2$. With the standard elimination algorithm, we can compute $D$ to be the unique up to scaling homogeneous polynomial
\[ 27u_0^2u_3^2 - 18u_0u_1u_2u_3 + 4u_0^2u_2^2 + 4u_1^2u_3 - u_1^2u_2. \]

Instead, we will compute $D$ by the steps below.

First, we restrict our parameter space to a general 1-dimensional affine space by replacing each parameter $u_i$ with $\Box t + \Box$ where $\Box$'s denote a general choice. For example, we substitute $u_0 =$
\[ 7t + 11, u_1 = 3t + 2, u_2 = 5t + 6, u_3 = 4t + 13 \text{ into } F \text{ and } J. \] We compute the radical of the elimination ideal \( (F(t, p), J(t, p)) \cap \mathbb{Q}[t] \) to get an ideal generated by

\[
t^4 + \frac{173086}{17315}t^3 + \frac{632753}{17315}t^2 + \frac{972374}{17315}t + \frac{531011}{17315} .
\]

By Lemma \( \square \) and Corollary \( \blacksquare \), the total degree of \( D \) equals the degree of the univariate polynomial above, if the \( 1 \)-dimensional affine space is general (this is because we are intersecting the hypersurface defined by \( D \) with a general line). Similarly, we are able to compute the degrees of \( D \) with respect to \( u_0, u_1, u_2, u_3 \) and get 2, 3, 3, 2, respectively. Since \( D \) is only unique up to scaling, to interpolate, we must fix a consistent choice of scaling. If possible, we will fix one of the coefficients of the pure powers \( u_i^{\deg D} \) to be 1. However, this is not possible if the coefficients of each pure power is zero, which occurs when there exists no \( u_i \) such that the degree of \( D \) with respect to \( u_i \) equals the total degree. So when this is not possible, we will perform a linear coordinate change so that \( D \) has a pure power in the new coordinates with a nonzero coefficient.

Second, we perform a linear change of variables by the following substitution \( u_i = \Box v_i + \Box v_i \) in \( D \) to get the polynomial \( D^*(v_0, v_1, v_2, v_3) \). For a general substitution, e.g. \( u_0 = v_1 - 2v_0, u_1 = v_1, u_2 = v_1 - 3v_2, u_3 = v_1 - 5v_3 \), the polynomial \( D^* \) has a nonzero coefficient for the pure power \( v_1^{\deg D} \).

Now, we may assume \( D^* \) has the following structure (here we have a fixed scaling):

\[ D^* = v_1^4 + C_1 v_1^3 + C_2 v_1^2 + C_3 v_1 + C_4, \]

where \( C_i \) is a homogeneous polynomial in \( v_0, v_2, v_3 \) with total degree \( i \). This means the monomials of \( C_1 \) are \( v_0, v_2, v_3 \), and we have \( C_1 = C_{11} v_1^{3} + C_{12} v_2 + C_{13} v_3 \). Once we determine the coefficients \( C_{11}, C_{12}, C_{13} \), we are done with \( C_1 \). This will be done by interpolation; rather than determining \( D^* \) directly, we determine \( D^* \) restricted to a line which induces linear constraints on the coefficients of \( D^* \). For example, we substitute \( v_0 = 13, v_2 = 4, v_3 = 5 \) into \( F^*(v_0, v_1, v_2, v_3, p) \) and \( J^*(v_0, v_1, v_2, v_3, p) \), and then we compute the radical of the elimination ideal \( \langle F^*(13, v_1, 4, 5, p), J^*(13, v_1, 4, 5, p) \rangle \) \( \cap \mathbb{Q}[v_1] \) finding it is generated by

\[
v_1^4 + \frac{243}{2} v_1^3 + \frac{87939}{16} v_1^2 + \frac{425385}{4} v_1 + \frac{2896803}{4}.
\]

By Corollary \( \square \), the polynomial in \( \blacksquare \) is \( D^*(13, v_1, 4, 5) \). Similarly, we compute to find

\[
D^*(7, v_1, 3, 11) = v_1^4 - 168 v_1^3 + \frac{36873}{4} v_1^2 - \frac{690201}{4} v_1 + \frac{4012281}{4} .
\]

Comparing the coefficients of \( D^* \) to Equations \( \boxplus \), \( \boxplus \), we have the linear relations below

\[
13C_{11} + 4C_{12} + 5C_{13} = -\frac{243}{2}, \; 7C_{11} + 3C_{12} + 11C_{13} = -168, \; 2C_{11} + 8C_{12} + 9C_{13} = -\frac{221}{2} .
\]

Solving the linear system yields \( C_{11} = -5, C_{12} = \frac{3}{2}, C_{13} = -\frac{25}{2} \). We similarly compute \( C_2, C_3, C_4 \) to determine \( D^* \):

\[
\begin{align*}
&\frac{27}{2} v_2^2 v_1^2 - \frac{27}{2} v_2^2 v_1 v_3 + \frac{467}{16} v_2 v_1^2 v_3 - 5 v_2 v_1^3 - \frac{225}{8} v_2 v_1^2 v_3 + \frac{225}{8} v_2 v_1^2 v_3 + v_1^4 + \frac{9}{4} v_1^2 + \\
&-\frac{27}{2} v_2^2 v_1^2 + \frac{135}{16} v_2 v_1^2 v_3 - \frac{467}{16} v_2 v_1^2 v_3 + \frac{225}{8} v_2 v_1^2 v_3 + v_1^4 + \frac{9}{4} v_1^2 + \\
&-\frac{27}{2} v_2^2 v_1^2 + \frac{9 v_1^2}{16} v_2 v_3^2 - \frac{467}{8} v_2 v_1^2 v_3 + \frac{225}{8} v_2 v_1^2 v_3 - \frac{9}{4} v_1^2 v_3^2 .
\end{align*}
\]

By applying the inverse linear change of coordinates \( v_0 = -\frac{u_0}{2} + \frac{u_1}{2}, v_1 = u_1, v_2 = -\frac{u_2}{3} + \frac{u_3}{3}, v_3 = -\frac{u_3}{3} + \frac{u_4}{3} \) to \( D^* \) and removing the denominator, we recover \( D \).

We present the formal description of the probabilistic algorithm in Algorithm \( \Box \). We explain the main algorithm (Algorithm \( \blacksquare \)) and all the sub-algorithms (Algorithms \( \boxplus \), \( \boxplus \), \( \boxplus \)) below.
Algorithm 5 (Degree). The probabilistic algorithm terminates correctly by Corollary 1 and Lemma 2.

Algorithm 4 (LinearOperator). The probabilistic algorithm terminates correctly by Lemma 1.

Algorithm 3 (Intersect). The probabilistic algorithm terminates correctly by Corollary 2.

Algorithm 2 (InterpolationDX-J).

Lines 1–5. We compute the total degree of $\mathcal{D}_{XJ}$ and the degrees of $\mathcal{D}_{XJ}$ with respect to $u_0, \ldots, u_d; d, d_0, \ldots, d_n$ by Algorithm 5. Algorithm 4 guarantees that $d_0 = d$ by applying a proper linear transformation $u_1 = a_1 \cdot u_0 + u_1, \ldots, u_n = a_n \cdot u_0 + u_n$.

Lines 6–7. Suppose $\mathcal{D}_{XJ} = u_0^d + C_1 u_0^{d-1} + \ldots + C_{d-1} u_0 + C_d$ where $C_1, \ldots, C_d \in \mathbb{Q}[u_1, \ldots, u_n]$ and the total degree of $C_j$ is $j$. For $j = 1, \ldots, n$, we estimate all the possible monomials of $C_j$ by computing the set

$$\{u_1^{\alpha_1} \cdots u_n^{\alpha_n} | \alpha_1 + \cdots + \alpha_n = j, 0 \leq \alpha_i \leq d_i\}$$

Assume the cardinality of the set is $N_j$ and rename these monomials as $U_{j,1}, \ldots, U_{j,N_j}$. Then we assume

$$C_j = c_{j,1} U_{j,1} + \ldots + c_{j,N_j} U_{j,N_j}$$

where $c_{j,1}, \ldots, c_{j,N_j} \in \mathbb{Q}$. The rest of the algorithm is to compute $c_{j,1}, \ldots, c_{j,N_j}$.

Lines 8–12. For each $j$, for $k = 1, \ldots, N_j$, for a random integer vector $b_k = (b_{k,1}, \ldots, b_{k,n})$, we compute $D_{XJ} (u_0, b_k)$ by Algorithm 3. That means to compute the function value $C_j(b_k)$ without knowing $D_{XJ}$.

Lines 13–15. For each $j$, we solve a square linear equation system for the unknowns $c_{j,1}, \ldots, c_{j,N_j}$:

$$c_{j,1} U_{j,1}(b_k) + \ldots + c_{j,N_j} U_{j,N_j}(b_k) = C_j(b_k),$$

$$(k = 1, \ldots, N_j)$$

It is known that we can choose nice $b_k$ probabilistically such that the coefficient matrix of the linear equation system is non-singular.

Lines 16. We apply the inverse linear transformation in the parameter space to get the $\mathcal{D}_{XJ}$ for the original $F_0, \ldots, F_m$.

We can also apply the interpolation idea to Algorithm PROPERNESSDETECTS 25 and get a probabilistic algorithm to compute the $D_{X,\infty}$. We omit the formal description of the algorithm.

Remark 3. According to the Notation 5, when the codimension of $\mathcal{L}_{XJ} \setminus \{0\}$ ($\mathcal{L}_{X,\infty} \setminus \{0\}$) is greater than 1, we define $\mathcal{D}_{XJ}$ ($D_{X,\infty}$) is 1. In this case, the number of real/positive solutions remains constant over a dense Zariski open set of the entire parameter space.

3.2.3 Strategy 2 and Strategy 3

In this section, we discuss different strategies for the probabilistic algorithm of Section 3.2.2. Note in Algorithm 2 lines 10–13 we compute all the possible terms in $\mathcal{D}_{XJ}$ and in Algorithm 2 lines 8–15 we interpolate all these terms at once. From the timings in the “Algorithm 2 Strategy 1” columns of Tables 11–12 we see for most of examples, Strategy 1 does not perform better than Algorithm 1.

By experiments, we find the main problem of Strategy 1 is the lifting step. In fact, it is expense to compute the inverse of $M_j$ in Algorithm 2 Line 14 which can be a large size matrix with (large) rational entries. In order to overcome this problem, we interpolate one parameter by one parameter, say Strategy 2. We omit the naive formal description of the algorithm, instead to explain how Strategy 2 works, we provide Example 4.

Example 4 (Algorithm 2(Strategy 2)). We compute $D$ in Example 3 by Strategy 2. Since the first two steps of Strategy 2 are the same as that of Strategy 1 (outlined in Example 3), we only show how to compute $D^*$. Recall $D^*$ is a homogeneous polynomial in $v_0,v_1,v_2,v_3$ with total degree 4 and the degrees of $D^*$ with respect to $v_0,v_1,v_2,v_3$ are $2,4,3,2$. 

11
Algorithm 2: (Main Algorithm) InterpolationDX-J

input : $F_0, \ldots, F_m, J$
output: $D_{X,J}$
1 $a_1, \ldots, a_n \leftarrow \text{LinearOperator}(F_0, \ldots, F_m, J)$
2 for $i$ from 1 to $n$ do
3 \hspace{1em} $F_i' \leftarrow$ replace $u_i$ in $F_i$ with $a_i u_0 + u_i$
4 NewSys $\leftarrow F_0, F_1', \ldots, F_n', F_{n+1}', \ldots, F_m, J$
5 $d, d_0, \ldots, d_n \leftarrow \text{Degree}(\text{NewSys})$
6 for $j$ from 1 to $d$ do
7 \hspace{2em} Rename all the monomials of the set
8 \hspace{3em} \{ $u_1^{\alpha_1} \cdots u_n^{\alpha_n} | \alpha_1 + \ldots + \alpha_n = j, 0 \leq \alpha_i \leq d_i$ \}
9 $N \leftarrow \max(N_1, \ldots, N_d)$
10 for $k$ from 1 to $N$ do
11 \hspace{2em} $b_{k,1}, \ldots, b_{k,n} \leftarrow$ random integers
12 \hspace{2em} $A(u_0) \leftarrow \text{Intersect(} \text{NewSys}, b_{k,1}, \ldots, b_{k,n} \text{)}$
13 \hspace{2em} $C_{i,k}, C_{i,k}^\ast \leftarrow$ the coefficients of $A(u_0)$ with respect to $u_0^0, \ldots, u_0^{d-1}$
14 for $j$ from 1 to $d$ do
15 \hspace{3em} $M_j \leftarrow N_j \times N_j$ matrix whose $(k, r)$-entry is $U_j^T (b_{k,1}, \ldots, b_{k,n})$
16 \hspace{3em} $C_j \leftarrow (U_j, 1, \ldots, U_j, N_j) M_j^{-1} (C_{i,j}^\ast, \ldots, C_{i,j}^\ast)^T$
17 $D_{X,J} \leftarrow$ replace $u_1, \ldots, u_n$ in $u_0^0 + \sum_{i=1}^{d-1} C_{i,0} u_0^i$ with $u_1 - a_1 \cdot u_0, \ldots, u_n - a_n \cdot u_0$
18 return $D_{X,J}$

Algorithm 3: Intersect

input : $F_0, \ldots, F_m, J$ and integers $b_1, \ldots, b_n$
output: $D_{X,J}(u_0, b_1, \ldots, b_n)$
1 for $i$ from 1 to $n$ do
2 \hspace{1em} $F_i' \leftarrow$ replace $u_i$ in $F_i$ with $b_i$
3 $A(u_0) \leftarrow$ generator of the radical of elimination ideal $(F_0, F_1^\ast, \ldots, F_n^\ast, F_{n+1}^\ast, \ldots, F_m, J) \cap Q[u_0]$
4 return $A(u_0)$

Algorithm 4: LinearOperator

input : $F_0, \ldots, F_m, J$
output: $a_1, \ldots, a_n$ such that the total degree of $D_{X,J}$ equals the degree of $D_{X,J}(u_0, a_1 \cdot u_0 + u_1, \ldots, a_n \cdot u_0 + u_n)$
with respect to $u_0$
1 $d, d_0, \ldots, d_n \leftarrow \text{Degree}(F_0, \ldots, F_m, J)$
2 if $d = d_0$ then
3 \hspace{1em} return $0, \ldots, 0$
4 else
5 \hspace{2em} for $i$ from 1 to $n$ do
6 \hspace{3em} $a_i \leftarrow$ a random integer
7 \hspace{3em} $F_i' \leftarrow$ replace $u_i$ in $F_i$ with $a_i \cdot u_0 + u_i$
8 NewSys $\leftarrow F_0, F_1', \ldots, F_n', F_{n+1}, \ldots, F_m, J$
9 $d, d_0, \ldots, d_n \leftarrow \text{Degree}(\text{NewSys})$
10 return $a_1, \ldots, a_n$

Algorithm 5: Degree

input : $F_0, \ldots, F_m, J$
output: $d, d_0, \ldots, d_n$, where $d$ is the total degree of $D_{X,J}$ and $d_i$ is the degree of $D_{X,J}$ with respect to each $u_i$
1 for $i$ from 0 to $n$ do
2 \hspace{1em} $F_0^\ast, \ldots, F_m^\ast \leftarrow$ replace $u_0, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n$ in $F_0, \ldots, F_n$ with random integers
3 $A(u_0) \leftarrow$ generator of the radical of elimination ideal $(F_0^\ast, \ldots, F_m^\ast, F_{n+1}, \ldots, F_m, J) \cap Q[u_0]$
4 $d_i \leftarrow$ degree of $A(u_0)$
5 \hspace{2em} $a_i, b_i \leftarrow$ random integers
6 $F_0(t), \ldots, F_m(t) \leftarrow$ replace $u_0, \ldots, u_n$ with $a_0 \cdot t + b_0, \ldots, a_n \cdot t + b_n$ in $F_0, \ldots, F_n$
7 $A(t) \leftarrow$ generator of the radical of elimination ideal $(F_0(t), \ldots, F_m(t), F_{n+1}, \ldots, F_m, J) \cap Q[t]$
8 $d \leftarrow$ degree of $A(t)$
9 return $d, d_0, \ldots, d_n$
First, replace all but two parameters with general numbers, e.g. \( v_2 = 3 \) and \( v_3 = 5 \), and we will compute \( D^*(v_0, v_1, 3, 5) \). We may assume \( D^*(v_0, v_1, 3, 5) \) has the following form

\[
D^*(v_0, v_1, 3, 5) = v_1^4 + C_1 v_1^3 + C_2 v_1^2 + C_3 v_1 + C_4,
\]

where \( C_i \) (\( i = 1, 2, 3, 4 \)) is a univariate polynomial in \( v_0 \). The degree of \( C_i \) is bounded above by \( i \) and the degree of \( D^* \) with respect to \( v_0 \). For instance, the degree of \( C_4 \) is at most 2, so we may assume

\[
C_4 = C_{40} v_0^2 + C_{41} v_0 + C_{42},
\]

where the coefficients \( C_{40}, C_{41}, C_{42} \) are constants because \( v_2 \) and \( v_3 \) have been fixed. We determine these coefficients by interpolation. Now, we continue to have \( v_2 = 3, v_3 = 5 \), and also fix \( v_0 \) to be a general number, e.g. \( v_0 = 2 \), to determine the elimination ideal

\[
\langle F(2, v_1, 3, 5, p), J(2, v_1, 3, 5, p) \rangle \cap \mathbb{Q}[v_1] = \langle v_1^4 - 68v_1^3 + \frac{5733}{4} v_1^2 - \frac{26801}{4} v_1 + 17604 \rangle.
\]

By Corollary 2,

\[
D^*(2, v_1, 3, 5) = v_1^4 - 68v_1^3 + \frac{5733}{4} v_1^2 - \frac{26801}{4} v_1 + 17604.
\]

Similarly, by fixing \( v_0 \) to be different values, we also determine

\[
D^*(7, v_1, 3, 5) = v_1^4 - 93v_1^3 + \frac{6219}{4} v_1^2 - \frac{174231}{4} v_1 + \frac{837081}{4},
\]

\[
D^*(9, v_1, 3, 5) = v_1^4 - 103v_1^3 + \frac{7743}{4} v_1^2 - \frac{248103}{4} v_1 + \frac{1379997}{4}.
\]

These 3 univariate polynomials induce linear constraints on the coefficients \( C_{40}, C_{41}, C_{42} \):

\[
2^2 C_{40} + 2C_{41} + C_{42} = 17604, \quad 7^2 C_{40} + 7C_{41} + C_{42} = \frac{837081}{4}, \quad 9^2 C_{40} + 9C_{41} + C_{42} = \frac{1379997}{4}.
\]

Solving this linear system, we find

\[
C_4 = \frac{16875}{4} v_0^2 + \frac{729}{2} v_0.
\]

Similarly, we compute \( C_1, C_2, C_3 \) and get \( D^*(v_0, v_1, 3, 5) \): \( v_1^4 + \frac{27}{4} v_0^3 v_1^2 - \frac{675}{2} v_0^2 v_1 + \frac{16875}{4} v_0^3 - 5v_0 v_1^3 + \frac{549}{2} v_0 v_1^2 - 3834 v_0 v_1 + \frac{229}{2} v_0 - 58 v_1^3 + \frac{4429}{4} v_0^2 - \frac{729}{2} v_0 v_1^2 + \frac{837081}{4} \).

Second, we compute \( D^*(v_0, v_1, v_2, 5) \). According to the monomials of \( D^*(v_0, v_1, 3, 5) \), we may assume \( D^*(v_0, v_1, v_2, 5) \) has the following form:

\[
v_1^4 + C_{10} v_0^2 v_1^2 + C_{20} v_0^2 v_1 + C_{30} v_1^3 + \cdots + C_{60} v_0 v_1 + C_7 v_0 + C_8 v_1^3 + C_9 v_1^2 + C_{10} v_1,
\]

where \( C_i \) (\( i = 1, \ldots, 10 \)) is a polynomial in \( v_2 \). The degree of \( C_i \) is bounded above by the degree of \( D^* \) with respect to \( v_2 \) and the difference \( \deg(D^*) - \deg(T_i) \) where \( T_i \) corresponds to the term of \( C_i \) in \( D^*(v_0, v_1, v_2, 5) \). For instance, the degree of \( C_{10} \) is 3, so

\[
C_{10} = C_{100} v_0^2 + C_{101} v_1 v_2 + C_{102} v_2 + C_{103}.
\]

Now we solve the coefficients \( C_{100}, C_{101}, C_{102}, C_{103} \). For general specializations \( v_2 = 11, 21, 4 \), we repeat the first step and compute

\[
D^*(v_0, v_1, 11, 5) = \frac{27}{4} v_0^3 v_1^2 - \frac{675}{2} v_0^2 v_1 + \frac{16875}{4} v_0^3 - 5v_0 v_1^3 + \frac{549}{2} v_0 v_1^2 - 3834 v_0 v_1 + \frac{229}{2} v_0 - 58 v_1^3 + \frac{4429}{4} v_0^2 - \frac{729}{2} v_0 v_1^2 + \frac{837081}{4} v_1^2 - \frac{3501}{4} v_0^2 - \frac{3501}{4} v_1^2,
\]

\[
D^*(v_0, v_1, 21, 5) = \frac{27}{4} v_0^3 v_1^2 - \frac{675}{2} v_0^2 v_1 + \frac{16875}{4} v_0^3 - 5v_0 v_1^3 + 234 v_0 v_1^2 - \frac{13757}{2} v_0 v_1 - \frac{250047}{4} v_0 + v_1^4 - 31 v_1^3 + \frac{4023}{2} v_1^2 - \frac{200047}{4} v_1,
\]

\[
D^*(v_0, v_1, 4, 5) = \frac{27}{4} v_0^3 v_1^2 - \frac{675}{2} v_0^2 v_1 + \frac{16875}{4} v_0^3 - 5v_0 v_1^3 - 1089 v_0 v_1^2 - \frac{15039}{4} v_0 v_1 - 864 v_0 + v_1^4 - 113 v_1^3 + \frac{13059}{16} v_1^2 - 432 v_1.
\]

These bivariate polynomials induce linear constraints on \( C_{100}, C_{101}, C_{102}, C_{103} \):

\[
3^3 C_{100} + 3^2 C_{101} + 3C_{102} + C_{103} = -\frac{729}{4},
\]

\[
11^3 C_{100} + 11^2 C_{101} + 11 C_{102} + C_{103} = -\frac{3501}{4}.
\]
21^3 C_{100} + 21^2 C_{101} + 21 C_{102} + C_{103} = -\frac{250047}{4}

4^3 C_{100} + 4^2 C_{101} + 4 C_{102} + C_{103} = -432.

Solving this system determines $C_{100}$. Similarly, we find $C_1, \ldots, C_9$ to get $D^*(v_0, v_1, v_2, 5)$:

\begin{align*}
\frac{27}{4} v_0^2 v_1^2 - \frac{675}{2} v_0 v_1 v_3 + \frac{16875}{4} v_1 v_2^2 + 1125 v_0 v_2^2 - \frac{27}{4} v_0 v_1 v_2^2 + \frac{675}{2} v_0 v_1 v_2 - \frac{16875}{4} v_0 v_1 + \frac{27}{4} v_0 v_2^3 + v_1^3 + \frac{3}{2} v_1^2 v_2 - \frac{125}{2} v_1^2 v_2 - \frac{65}{2} v_1 v_2^3 + \frac{16875}{4} v_2 v_1^2 - \frac{225}{2} v_1 v_2^3.
\end{align*}

Third, it is easy to recover $D^*$ from $D^*(v_0, v_1, v_2, 5)$ because we know $D^*$ is homogeneous. For example, the second term in $D^*(v_0, v_1, v_2, 5)$ is $-\frac{675}{2} v_0^2 v_1$. The total degree of this term is 3. So the corresponding term in $D^*$ is $-\frac{675}{2} v_0^2 v_1 v_3$. Similarly, we recover all terms of $D^*$ and again get (4).

From the timings in the columns “Algorithm 2” and “Strategy 2” of Tables 1-2, we see Strategy 2 improves the efficiency for the larger examples, e.g. Dense Models 3-4 and Models 3-4. The advantage of Strategy 2 compared to Strategy 1 is that we only need to solve a small-sized (bounded by the degree of $D_{X,J}$ with respect to each parameter at each step) linear equation system in the lifting step. Therefore, the most time consuming part of Strategy 2 is sampling. As soon as we can compute the degree of $D_{X,J}$, we can estimate the computational timings of Strategy 2. In fact, we can first check the approximate timing of sampling once by running Algorithm 8 once, say $T_s$. Suppose the total degree of $D_{X,J}$ is $d$ and the degrees with respect to $u_0, \ldots, u_n$ are $d_0, \ldots, d_n$. Without loss of generality, assume $d = d_0$ and $d_0 \geq d_1 \geq \cdots \geq d_n$. Then the total approximate sampling timing of Strategy 2 is

$$T_s d_2 \cdots d_n.$$ (5)

So if we improve $T_s$, we might improve the efficiency. This motivates Strategy 3.

We give another two sub-algorithms Algorithm 6 and Algorithm 7 to do sampling for interpolation and to compute the degree of $D_{X,J}$, respectively. More precisely, Strategy 3 means to use Algorithms 6 and 7 instead of Algorithms 3 and 5 in Algorithm 2. The correctness of Algorithms 6 and 7 directly follows from Lemma 9. We give Example 5 to show how Algorithm 7 works (and similarly this example shows how Algorithm 6 works).

**Example 5** (Toy Example for Strategy 3). Here, we illustrate Algorithm 8 which is used to compute the degree of the discriminant. Here, we compute the total degree of $D_{X,J}$ to be 4 of the linear model in Example 3 by Algorithm 7. Define $F_0, F_1, \ldots, F_5$ as those in Example 7. Rename $p_0, \ldots, p_3, \lambda_1, \lambda_2$ as $y_0, \ldots, y_3, y_4, y_5$. The determinant of Jacobian matrix of $F_0, F_1, \ldots, F_5$ with respect to $y_0, \ldots, y_3, y_4, y_5$, say $J$, is

$$-\lambda_1^2 p_0 p_1 - 4 \lambda_1^2 p_0 p_2 - 25 \lambda_1^2 p_0 p_3 - \lambda_2^2 p_1 p_2 - 36 \lambda_1^2 p_1 p_3 - 49 \lambda_1^2 p_2 p_3 + \lambda_1 \lambda_2 p_0 p_1 + 8 \lambda_1 \lambda_2 p_0 p_2 - 125 \lambda_1 \lambda_2 p_0 p_3 + 3 \lambda_2 \lambda_2 p_1 p_2 - 144 \lambda_1 \lambda_2 p_1 p_3 - 147 \lambda_1 \lambda_2 p_2 p_3 + 12 \lambda_2^2 p_0 p_1 + 32 \lambda_2^2 p_0 p_2 - 150 \lambda_2^2 p_0 p_3 + 4 \lambda_2^2 p_2 p_1 - 108 \lambda_2^2 p_2 p_3 - 98 \lambda_2^2 p_2 p_3.$$

First, we restrict the parameter space to a general line by the substitution $u_i = \square i + \square$, e.g. $u_0 = t + 13, u_1 = 4 t + 2, u_2 = 9 t + 6, u_3 = 31 t + 5$, into $F_0, F_1, \ldots, F_5$. Next, we eliminate all but one unknown; we compute the radical of the elimination ideal $(F_0(t, y_0, \ldots, y_6), \ldots, F_5(t, y_0, \ldots, y_6)) \cap \mathbb{Q}[t, y_0]$, say $(g)$ where $g$ is

$$20250 y_0^2 t^2 + 23400 y_0^2 t - 22770 y_0^2 t^2 + 6760 y_0^3 - 45961 y_0^2 t + 1488 y_0^2 t - 18954 y_0^2 + 20582 y_0 t - 24 t^2 + 16094 y_0 - 624 t - 4056.$$

Computing the resultant $\text{resultant}(g, \frac{\partial g}{\partial y_0}, y_0)$, we get $(t + 13)(t + \frac{45}{2})G$, where

$$G = t^4 + \frac{348880664004}{24297439441} t^3 + \frac{1144507255}{1166270241} t^2 + \frac{391327027}{1166270241} t + \frac{244617385}{4065108996}.$$

By Lemma 3, we know $D_{X,J}(t + 13, 4 t + 2, 9 t + 6, 31 t + 5)$ is a factor of $\text{resultant}(g, \frac{\partial g}{\partial y_0}, y_0)$. For the three irreducible factors of $\text{resultant}(g, \frac{\partial g}{\partial y_0}, y_0)$, we check by computing elimination ideal that
\[ \langle F_0(t, y_0, \ldots, y_6), \ldots, F_5(t, y_0, \ldots, y_6), J, t + 13 \rangle \cap \mathbb{Q}[t] = (1) \]
\[ \langle F_0(t, y_0, \ldots, y_6), \ldots, F_5(t, y_0, \ldots, y_6), J, t + \frac{45}{52} \rangle \cap \mathbb{Q}[t] = (1) \]
\[ \langle F_0(t, y_0, \ldots, y_6), \ldots, F_5(t, y_0, \ldots, y_6), J, G \rangle \cap \mathbb{Q}[t] = \langle G \rangle. \]

Therefore we confirm \( D_{X,J}(t + 13, 4t + 2, 9t + 6, 31t + 5) \) is \( G \), which is consistent with the result computed by Algorithm 3. Again, we conclude the total degree of \( D_{X,J} \) is 4.

In Algorithm 6 we are eliminating all unknowns but \( t \) in \( \langle F_0, \ldots, F_m, J \rangle \). While in Algorithm 7 we are eliminating all unknowns but \( t, y_0 \) in \( \langle F_0, \ldots, F_m \rangle \) and carrying out many other eliminations to \( t \) in ideals \( \langle F_0, \ldots, F_m, J, G_i \rangle \). It is not immediately clear why Algorithm 7 would be an improvement over Algorithm 6. In fact, for the larger algebraic models, \( J \) will be a huge polynomial. In this case, eliminating from \( \langle F_0, \ldots, F_m, J \rangle \) can be very slow. Either eliminating from \( \langle F_0, \ldots, F_m \rangle \) or eliminating from \( \langle F_0, \ldots, F_m, J, G_i \rangle \) can be faster because either removing huge polynomials or including additional small polynomials might speed up elimination according to experience. Of course it is difficult to prove theoretically which strategy will be more efficient, the best way is to implement both strategies and run the benchmarks. In our case, Algorithm 7 indeed gives improvement for some larger models, see Models 4, 6, 7, 9 in Table 3, Section 4.

### Algorithm 6: IntersectStrategy3

**Input:** \( F_0, \ldots, F_m, J \) and integers \( b_1, \ldots, b_n \)

**Output:** \( D_{X,J}(u_0, b_1, \ldots, b_n) \)

1. for \( i \) from 1 to \( n \) do
2. \( F_{in} \leftarrow \text{replace } u_i \text{ in } F_0 \text{ with } b_i \)
3. \( A_i(u_0, y_0) \leftarrow \text{generator of the radical of elimination ideal } \langle F_{0}, \ldots, F_{in}, F_{n+1}, \ldots, F_m \rangle \cap \mathbb{Q}[u_0, y_0] \)
4. if degree\( (A_1, y_0) = ML Degree \) then
5. \( \bigwedge \text{resultant}(A_1, \frac{\partial A_1}{\partial y_0}, y_0) \)
6. \( \text{newA} \leftarrow 1 \)
7. for each irreducible factor \( G(u_0) \) of \( A(u_0) \) do
8. if \( \langle F_{0}, \ldots, F_{in}, F_{n+1}, \ldots, F_m, J, G(u_0) \rangle \cap \mathbb{Q}[u_0] \neq (1) \) then
9. \( \text{newA} \leftarrow \text{newA} \cdot G(u_0) \)
10. return \( \text{newA} \)

**Remark 4.** Strategy 3 is incomplete whenever the solutions to the likelihood equations along the general line do not have distinct \( y_0 \) coordinates. This means the "if" condition in Algorithm 6–Line 4 or in Algorithm 7–Lines 3–7 is not satisfied. In principle, we can modify Strategy 3 by applying a general linear coordinate change to \( y_0, \ldots, y_m \).

### 4 Implementation Details and Experimental Timings

We have implemented the probabilistic method, Algorithm 2 in Maple 2015. For comparisons, we have also implemented the standard Algorithm 1 in Maple 2015. We run our implementations of Algorithms 1 and 2 for many examples to set benchmarks by a 3.2 GHz Inter Core i5 processor (8GB total memory) under OS X 10.9.3. We record the timings for these examples in Tables 3 through 5. In this section, we give conclusions from our computational experiments and results.

The probabilistic algorithm is implemented in three different strategies described in Section 3.2.3: they are the ‘interpolate every term at once’ strategy; ‘interpolate one parameter at a time’ strategy; and ‘one variable’ strategy respectively.

There are two kinds of benchmarks, the random models and literature models. The random models, say Dense Models 1–6, are dense homogeneous polynomials generated by the Maple command

\[ \text{randpoly}([p_0, \ldots, p_n], \text{dense}, \text{degree} = m), \quad n = 2, 3; \quad m = 2, 3, 4. \]

1. See our Maple code on the website: https://sites.google.com/site/rootclassification/publications/jsc2016
2. When a computation returned no output in 2 hours we record ">2h"
Models 1–9 are examples presented in the literatures [19, 9, 17]. See the appendix for the literature models' defining equations.

The most important conclusion is that our method with new implementation based on Maple 2015 and FGb is more efficient than our implementation in [27]. For the random models, in [27, Table 1], our previous implementation of Algorithm 2 was only able to compute random dense models with matrices generated by Algorithm 1 using FGb after running for 2–3 hours. One explanation for this is seen in Table 4. We see the intermediate instance, for Model 4 [27, Example 6], the old implementation of Strategy 2 spent 30 days while the new implementation of Strategy 2 takes less than 30 minutes.

Another important take away from the tables is that the probabilistic algorithm with Strategy 1 is comparable to the standard elimination method in Table 2. We only receive great improvements when employing Strategies 2 or 3.

Example 6. As a guide to reading the tables, we summarize the first row of Table 1. The model

\[ \text{Algorithm 7: DegreeStrategy3} \]

\[
\text{input : } F_0, \ldots, F_m, J \\
\text{output: } d, d_0, \ldots, d_n, \text{ where } d \text{ is the total degree of } D_X, \text{ and } d_i \text{ is the degree of } D_{X,i} \text{ w.r.t each } u_i \ (i = 0, \ldots, n) \\
\text{for } i \text{ from } 0 \text{ to } n \text{ do} \\
\quad F_0, \ldots, F_m \leftarrow \text{replace } u_0, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n \text{ in } F_0, \ldots, F_m \text{ with random integers} \\
\quad g(u_i, y_0) \leftarrow \text{generator of the radical of elimination ideal } \langle F_0^\delta, \ldots, F_m^\delta, F_{n+1}, \ldots, F_m \rangle \cap \mathbb{Q}[u_i, y_0] \\
\quad \text{if } \text{degree}(g, y_0) = \text{ML Degree} \text{ then} \\
\quad \quad A(u_i) \leftarrow \text{resultant}(g, \frac{\partial g}{\partial y_0}, y_0) \\
\quad \text{newA} \leftarrow 1 \\
\quad \text{for each irreducible factor } G(u_i) \text{ of } A(u_i) \text{ do} \\
\quad \quad \text{if } \langle F_0^\delta, \ldots, F_m^\delta, F_{n+1}, \ldots, F_m, J, G(u_i) \rangle \cap \mathbb{Q}[u_i] \neq \{1\} \text{ then} \\
\quad \quad \quad \text{newA} \leftarrow \text{newA} \cdot G(u_i) \\
\quad \quad d_i \leftarrow \text{degree of newA} \\
\quad \quad a_i, b_i \leftarrow \text{random integers} \\
\quad F_0(t), \ldots, F_m(t) \leftarrow \text{replace } u_0, \ldots, u_n \text{ with } a_0 \cdot t + b_0, \ldots, a_n \cdot t + b_n \text{ in } F_0, \ldots, F_m \\
\quad g(t, y_0) \leftarrow \text{generator of the radical of elimination ideal } \langle F_0(t), \ldots, F_m(t), F_{n+1}, \ldots, F_m \rangle \cap \mathbb{Q}[t, y_0] \\
\quad \text{if } \text{degree}(g, y_0) = \text{ML Degree} \text{ then} \\
\quad \quad A(t) \leftarrow \text{resultant}(g, \frac{\partial g}{\partial y_0}, y_0) \\
\quad \text{newA} \leftarrow 1 \\
\quad \text{for each irreducible factor } G(t) \text{ of } A(t) \text{ do} \\
\quad \quad \text{if } \langle F_0^\delta, \ldots, F_m^\delta, F_{n+1}, \ldots, F_m, J, G(t) \rangle \cap \mathbb{Q}[t] \neq \{1\} \text{ then} \\
\quad \quad \quad \text{newA} \leftarrow \text{newA} \cdot G(t) \\
\quad d \leftarrow \text{degree of newA} \\
\text{return } d, d_0, \ldots, d_n \]

\[ \text{The dense models are available at https://sites.google.com/site/rootclassification/publications/jsc2016} \]
The lower bound of the computational timings are approximated by the times in italics. For instance, consider Model 5 in Table 3. There are 5 parameters $u_0, u_1, u_2, u_3, u_4$. By Algorithm 5 the total degree of $\mathcal{D}_{X_J}$ is 48 and the degrees with respect to the five parameters are: 31, 44, 48, 44 and 31. We check the timing for doing sample one time by Algorithm 5 is 2.0s. Then we estimate the timing for sampling of Strategy 2 is $2.0s \times 44 \times 31 \times 31 \approx 23.49h$ (1d) by formula (2) in Section 7.2.6. Similarly, we check the timing for doing sample one time by Algorithm 6 is 4.8s and estimate the total timing for sampling of Strategy 3 is $4.8s \times 44 \times 31 \times 31 \approx 56.37h$ (2d).

Remark 5. Whenever we compute an elimination ideal we use the F0b command fgb_gbasis.elim. This is done in Algorithm 1-Line 1 Algorithm 2-Line 5 Algorithm 3-Line 8 Algorithm 4-Line 5 and Algorithm 6-Line 8.

5 Future Work and Final Application

5.1 Next steps

For our future work we want to determine the discriminants for larger sized problem such as Models 5-9. The interpolation method discussed in this paper is basically linear lifting interpolation.
Section 3.2.3. Formula (5) gives the timing estimation for this linear lifting method. Although Strategy 3 improves \( T_s \) in formula (5), the interpolation is still expensive (see Table 3) since \( d_2 \cdots d_n \) is quite a large number for these models. In order to overcome this problem, one possible way is to try Newton–Hensel lifting [14, 30].

We also want to develop heuristics for checking the output of probabilistic algorithm. For all examples we have tried in Section 4, the computational results of probabilistic algorithm are the same with the results computed by the standard algorithm as soon as both of algorithms finish the computation. For some examples such as Model 4, the standard algorithm gives no output in a reasonable time. However, there are ways to check the output of probabilistic algorithm. The first way to check is to simply run the algorithm a second time. The second way is to evaluate the the discriminant on a random line. This univariate polynomial should agree by doing an elimination problem that replaces the data with the random line. The third way to check is to find a point on the discriminant. Solving the likelihood equations for this special point, we are able to check if solutions coincide by brute force. Moreover, if the discriminant has no rational points, then one could try to use numerical homotopy continuation methods instead.

The next step in this work is the complete positive root classification. The discriminant is necessary, but usually not enough for classifying positive solutions (see the next subsection regarding Model [4]). In order to complete positive root classification, we have to consider the discriminant sequence [32].

### 5.2 Final Application

We end the paper with the discussion of real root classification on the \( 3 \times 3 \) symmetric matrix model (Model [3]).

Consider the following story with dice. A gambler has a coin, and two pairs of three-sided dice. The coin and the dice are all unfair. However, the two dice in the same pair have the same weight. He plays the same game 1000 rounds. In each round, he first tosses the coin. If the coin lands on side 1, he tosses the first pair of dice. If the coin lands on side 2, he tosses the second pair of dice. After the 1000 rounds, he records a \( 3 \times 3 \) data matrix \([u_{ij}]\) (\( i, j = 1, 2, 3 \)) where \( u_{ij} \) is the number of times for him to get the sides \( i \) and \( j \) with respect to the two dice. By the matrix \([u_{ij}]\), he is trying to estimate the probability \( p_{ij} \) of getting the sides \( i \) and \( j \) with respect to the two dice.

It is easy to check that the matrix 

\[
\begin{bmatrix}
p_{11} & p_{12} & p_{13} \\
p_{21} & p_{22} & p_{23} \\
p_{31} & p_{32} & p_{33}
\end{bmatrix}
\]

is symmetric and has at most rank 2. Let

\[
p_{ij} = \begin{cases} 
p_{ij} & i = j, \\
\frac{1}{2}p_{ij} & i < j,
\end{cases}, \quad u_{ij} = \begin{cases}
p_{ij} & i = j, \\
u_{ij} + u_{ji} & i < j.
\end{cases}
\]

We have an algebraic statistical model \( \mathcal{M} = \mathcal{V}(g) \cap \Delta_5 \), whose Zariski closure is \( X \) where

\[
g = \det \begin{bmatrix}
2p_{11} & p_{12} & p_{13} \\
p_{12} & 2p_{22} & p_{23} \\
p_{13} & p_{23} & 2p_{33}
\end{bmatrix}
\]

and \( \Delta_5 = \{(p_{11}, \ldots, p_{33}) \in \mathbb{R}_{\geq 0}^3 | p_{11} + p_{12} + p_{13} + p_{22} + p_{23} + p_{33} = 1 \} \).
According to the Definition 2, we present the Langrange likelihood equations:

\[
F_0 = p_{11} \lambda_1 + (8p_{22}p_{33} - 2p_{23}^2)p_{11} \lambda_2 - u_{11}
\]
\[
F_1 = p_{12} \lambda_1 + (2p_{13}p_{23} - 4p_{12}p_{33})p_{12} \lambda_2 - u_{12}
\]
\[
F_2 = p_{13} \lambda_1 + (2p_{12}p_{23} - 4p_{13}p_{22})p_{13} \lambda_2 - u_{13}
\]
\[
F_3 = p_{22} \lambda_1 + (8p_{11}p_{33} - 2p_{13}^2)p_{22} \lambda_2 - u_{22}
\]
\[
F_4 = p_{23} \lambda_1 + (2p_{12}p_{13} - 4p_{11}p_{23})p_{23} \lambda_2 - u_{23}
\]
\[
F_5 = p_{33} \lambda_1 + (8p_{11}p_{22} - 2p_{12}^2)p_{33} \lambda_2 - u_{33}
\]
\[
F_6 = g(p_{11}, p_{12}, p_{13}, p_{22}, p_{23}, p_{33})
\]
\[
F_7 = p_{11} + p_{12} + p_{13} + p_{22} + p_{23} + p_{33} - 1
\]

where the unknowns are \(p_{11}, p_{12}, p_{13}, p_{22}, p_{23}, p_{33}, \lambda_1, \lambda_2\) and parameters are \(u_{11}, u_{12}, u_{13}, u_{22}, u_{23}, u_{33}\).

We have 8 equations in 8 unknowns with 6 parameters and the ML degree is \(6\) \([19]\). By the Algorithm 2 we have computed \(D_X\), which has 1307 terms with total degree 12. By a similar computation, we get \(D_X^\infty\) whose factors are \(g(u_{11}, \ldots, u_{33})\) or positive when each \(u_i\) is positive.

For the data-discriminant \(D_X\) we have computed above, we have also computed\(^4\) at least one rational point (sample point) from each open connected component of \(D_X \neq 0\) using \texttt{RAGlib}\(^2\) and \texttt{FGb}\(^5\). With these sample points, we can solve the real root classification problem on the open cells. By testing all 236 sample points, we see that if \(g(u_{11}, \ldots, u_{33}) \neq 0\), then

- if \(D_X\) \(u_{11}, \ldots, u_{33} > 0\), then the system has 6 distinct real solutions and there can be 6 positive solution or 2 positive solutions;
- if \(D_X\) \(u_{11}, \ldots, u_{33} < 0\), then the system has 2 distinct real (positive) solutions.

With 2 of these sample points, we see that the sign of \(D_X\) is not enough to classify the positive solutions. For example, for the sample point \((u_{11} = 1, u_{12} = 1, u_{13} = 2380264116850850, u_{22} = 1, u_{23} = 280264116850850, u_{33} = 1)\), the system has 6 distinct positive solutions. While for the sample point \((u_{11} = 1, u_{12} = 1, u_{13} = 199008, u_{22} = 30, u_{23} = 2022, u_{33} = 1)\), the system has also 6 real solutions but only 2 positive solutions.

### 6 Acknowledgments

We began this project in NIMS Thematic Program 2014 on Applied Algebraic Geometry. We had the idea for this extended version in 2015 summer before SIAM Conference: Applied Algebraic Geometry when the two authors were hosted by NIMS, Daejeon, Korea. We thank Professors Bernd Sturmfels, Hoon Hong, Jonathan Hauenstein and Frank Sottile for their valuable advice on this project. We thank Professors Mohab Safey Ei Din and Jean-Charles Faugere for their software advice on \texttt{RAGlib} and \texttt{FGb} respectively. We thank the anonymous referees of ISSAC’15 for their insightful suggestions to greatly improve our work. We thank Professor Guillaume Moroz for his nice discussion and advice during ISSAC’15. We thank Professor Agnes Szanto for the reference on Shape Lemma (multivariate version).

### References

[1] N. Budur and B. Wang. The Signed Euler Characteristic of Very Affine Varieties. *Int. Math. Res. Not. IMRN*, (14):5710–5714, 2015.

\(^4\)See \(D_X\) on the website: https://sites.google.com/site/rootclassification/publications/DD

\(^5\)The sample points were first successfully computed by one of the ISSAC’15 anonymous referees.
[2] M.-L. G. Buot, S. Hoşten, and D. Richards. Counting and locating the solutions of polynomial systems of maximum likelihood equations, ii: The behrens-fisher problem. *Statistica Sinica*, 17:1343–1354, 2007.

[3] F. Catanese, S. Hoşten, A. Khetan, and B. Sturmfels. The maximum likelihood degree. *Amer. J. Math.*, 128(3):671–697, 2006.

[4] C. Chen, J. H. Davenport, J. P. May, M. M. Maza, B. Xia, and R. Xiao. Triangular decomposition of semi-algebraic systems. In *Proceedings of ISSAC’10*, pages 187–194. ACM New York, 2010.

[5] G. E. Collins and H. Hong. *Partial Cylindrical Algebraic Decomposition for Quantifier Elimination*. Springer, 1998.

[6] M. Coste and M. Shiota. Nash triviality in families of nash manifolds. *Inventiones Mathematicae*, 108(1):349–368, 1992.

[7] D. A. Cox, J. Little, and D. Oshea. *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*. Springer, 2007.

[8] J. Draisma, E. Horobet, G. Ottaviani, B. Sturmfels, and R. R. Thomas. The Euclidean distance degree of an algebraic variety. *Found. Comp. Math.* 2014.

[9] M. Drton, B. Sturmfels, and S. Sullivant. *Lectures on algebraic statistics*. Springer, 2009.

[10] J.-C. Faugère. Fgb: A library for computing gröbner bases. In K. Fukuda, J. Hoeven, and M. Joswig, editors, *Mathematical Software - ICMS 2010*, volume 6327 of *Lecture Notes in Computer Science*, pages 84–87. Springer Berlin / Heidelberg, 2010.

[11] J.-C. Faugère and S. Lachartre. Parallel gaussian elimination for gröbner bases computations in finite fields. In *Proceedings of the 4th International Workshop on Parallel and Symbolic Computation*, pages 89–97. ACM, 2010.

[12] J.-C. Faugère, M. Safey El Din, and P.-J. Spaenlehauer. Gröbner bases and critical points: the unmixed case. In *Proceedings of ISSAC’12*, pages 162–169. ACM New York, 2012.

[13] M. Giusti, J. Heintz, J. E. Morais, J. Morgenstern, and L. M. Pardo. Straight-line programs in geometric elimination theory. *Journal of pure and applied algebra*, 124(1):101–146, 1998.

[14] M. Giusti, G. Lecerf, and B. Salvy. A Gröbner free alternative for polynomial system solving. *Journal of Complexity*, 17:154–211, 2001.

[15] A. Greuet and M. Safey El Din. Probabilistic algorithm for the global optimization of a polynomial over a real algebraic set. *SIAM Journal on Optimization*, 24(3):1313–1343, 2014.

[16] E. Gross, M. Drton, and S. Petrović. Maximum likelihood degree of variance component models. *Electronic Journal of Statistics*, 6:993–1016, 2012.

[17] E. Gross and J. I. Rodriguez. Maximum likelihood geometry in the presence of data zeros. In *Proceedings of ISSAC’14*, pages 232–239. ACM New York, 2014.

[18] J. Hauenstein, J. I. Rodriguez, and B. Sturmfels. Maximum likelihood for matrices with rank constraints. *Journal of Algebraic Statistics*, 5:18–38, 2014.

[19] S. Hoşten, A. Khetan, and B. Sturmfels. Solving the likelihood equations. *Foundations of Computational Mathematics*, 5(4):389–407, 2005.
Appendix: Literature Models

Model 1. \[ Random \ Censoring \ Model \]

\[
2p_0p_1p_2 + p_1^2p_2 + p_1p_2^2 - p_0p_{12} - p_1p_2p_{12} = 0, \quad p_0 + p_1 + p_2 + p_{12} = 1
\]

Model 2. \[ 3 \times 3 \ Zero-Diagonal \ Matrix \]

\[
\det \begin{bmatrix}
  0 & p_{12} & p_{13} \\
  p_{21} & 0 & p_{23} \\
  p_{31} & p_{32} & 0
\end{bmatrix} = 0, \quad p_{12} + p_{13} + p_{21} + p_{23} + p_{31} + p_{32} = 1
\]

Model 3. \[ Grassmannian \ of \ 2 \text{-planes} \ in \ \mathbb{C}^4 \]

\[
p_{12}p_{34} - p_{13}p_{24} + p_{14}p_{23} = 0, \quad p_{12} + p_{13} + p_{14} + p_{23} + p_{24} + p_{34} = 1
\]
Model 4. \[ \text{19} \] 3 \times 3 \text{ Symmetric Matrix}]

\[
det \begin{bmatrix}
2p_{11} & p_{12} & p_{13} \\
p_{12} & 2p_{22} & p_{23} \\
p_{13} & p_{23} & 2p_{33}
\end{bmatrix} = 0, \quad p_{11} + p_{12} + p_{13} + p_{22} + p_{23} + p_{33} = 1
\]

Model 5. \[ \text{19} \] Bernoulli 3 \times 3 \text{ Coin}]

\[
det \begin{bmatrix}
12p_0 & 3p_1 & 2p_2 \\
3p_1 & 2p_2 & 3p_3 \\
2p_2 & 3p_3 & 12p_4
\end{bmatrix} = 0, \quad p_0 + p_1 + p_2 + p_3 + p_4 = 1
\]

Model 6. \[ \text{19} \] 3 \times 3 \text{ Matrix}]

\[
det \begin{bmatrix}
p_{00} & p_{01} & p_{02} \\
p_{10} & p_{11} & p_{12} \\
p_{20} & p_{21} & p_{22}
\end{bmatrix} = 0, \quad p_{00} + p_{01} + p_{02} + p_{10} + p_{11} + p_{12} + p_{20} + p_{21} + p_{22} = 1
\]

Model 7. \[ \text{19} \] Projection of 3 \times 4 \text{ Matrix}]

\[
p_{12}p_{23}p_{34} - p_{12}p_{24}p_{33} - p_{13}p_{22}p_{34} + p_{13}p_{24}p_{32} + p_{14}p_{22}p_{33} - p_{14}p_{23}p_{32} = 0,
\]

\[
p_{12} + p_{13} + p_{14} + p_{22} + p_{23} + p_{24} + p_{32} + p_{33} + p_{34} = 1
\]

Model 8. \[ \text{19} \] Juke-Cantor Model, Example 18]

\[
9000q_{111}^2 - 9011q_{110}q_{110} = 0, \quad p_{123} + p_{dis} + p_{12} + p_{13} + p_{23} = 1
\]

where

- \( q_{111} = p_{123} + \frac{p_{dis}}{3} - \frac{p_{12}}{3} - \frac{p_{13}}{3} - \frac{p_{23}}{3} \),
- \( q_{101} = p_{123} - \frac{p_{dis}}{3} + p_{12} - \frac{p_{13}}{3} - \frac{p_{23}}{3} \),
- \( q_{011} = p_{123} - \frac{p_{dis}}{3} - \frac{p_{12}}{3} + \frac{p_{13}}{3} + \frac{p_{23}}{3} \),
- \( q_{001} = p_{123} + p_{dis} + p_{12} + p_{13} + p_{23} \).

Model 9. \[ \text{19} \] Example 16]

\[
q_{2q7} - q_{1q8} = 0, \quad q_{3q6} - q_{5q4} = 0, \quad p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7 + p_8 = 1
\]

where

- \( q_1 = p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7 + p_8 \),
- \( q_3 = p_1 + p_2 - p_3 + p_4 + p_5 + p_6 - p_7 - p_8 \),
- \( q_5 = p_1 + p_2 + p_3 + p_4 - p_5 - p_6 - p_7 - p_8 \),
- \( q_7 = p_1 + p_2 - p_3 + p_4 - p_5 - p_6 + p_7 + p_8 \),
- \( q_8 = p_1 - p_2 + p_3 + p_4 - p_5 + p_6 + p_7 - p_8 \).