MAXIMUM LIKELIHOOD DEGREE OF VARIOUS TORIC VARIETIES

A thesis presented to the faculty of
San Francisco State University
In partial fulfilment of
The Requirements for
The Degree

Master of Arts
In
Mathematics

by

Radoslav Vuchkov

San Francisco, California

August 2016

CERTIFICATION OF APPROVAL

I certify that I have read *MAXIMUM LIKELIHOOD DEGREE OF VARIOUS TORIC VARIETIES* by Radoslav Vuchkov and that in my opinion this work meets the criteria for approving a thesis submitted in partial fulfillment of the requirements for the degree: Master of Arts in Mathematics at San Francisco State University.

Serkan Hosten
Professor of Mathematics

Joseph Gubeladze
Professor of Mathematics

Matthias Beck
Professor of Mathematics

MAXIMUM LIKELIHOOD DEGREE OF VARIOUS TORIC VARIETIES

Radoslav Vuchkov
San Francisco State University
2016

Before we can speak of the maximum likelihood degree (ML degree), we should know the story of the maximum likelihood estimate (ML estimate). The ML estimate of given data (collected from sample realizations of the discrete random variable) with respect to a statistical model is a choice of parameters which maximizes the probability of observing these data. Here we define a statistical model for a discrete random variable as a subset of the probability simplex and a parametric statistical model is defined as the image of the parameter space.

In this thesis we work in the realms of algebraic geometry and algebraic statistics. In algebraic statistics, the models are images of the parameter space under a polynomial map to the probability simplex. In general, finding the ML estimate with respect to a model is accomplished by maximizing the likelihood function. For an algebraic statistical model, this function is the product of powers of polynomials, where the polynomials are the coordinates of the polynomial map. One way of finding the ML estimate is to solve the critical equations of the likelihood function, subject to the constraint that the image of the map defining the model is indeed in the probability simplex. In this thesis we define the maximum likelihood degree of an algebraic statistical model as the number of complex critical points of the

likelihood function.

We mainly work with a large class of statistical models which are known as discrete exponential models and are defined by a monomial parametrization. In algebraic statistics, these models are known as toric varieties, and they have been studied extensively. Toric varieties form an important and rich class of examples in algebraic geometry, and our focus is the computation of the ML degree for some of them. In particular, we study rational normal curves, certain Veronese varieties, and some toric surfaces known as Hirzebruch surfaces. In our work we prove that for rational normal curves, the second Veronese embedding, and for the second hyper-simplex varieties, the ML degree is equal to the degree of the respective varieties. In general, this result does not hold. For instance, in the case of Hirzebruch surfaces, we compute examples where the ML degree is smaller than the degree of the surface. Furthermore we also consider a family of rational normal curves obtained by scaling coordinates. We give a stratification of this family with respect to the ML degree by determining sets of scaling parameters for which the ML degree takes particular values. Finally, we present conjectures and computations for further study.

I certify that the Abstract is a correct representation of the content of this thesis.

Chair, Thesis Committee　　　　　　　　　　　　Date

ACKNOWLEDGMENTS

I would like to express my special appreciation and thanks to my advisor Professor Serkan Hosten, you have been a tremendous mentor for me. I would also like to thank my committee members, professor Joseph Gubeladze, professor Matthias Beck, for serving as my committee members. I also want to thank you for letting my defense be an enjoyable moment, and for your brilliant comments and suggestions.

A special thanks to my family. I would also like to express my appreciation to my fiance. She inspires me to be a better person every day.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

The goal of this thesis is to compute and, if possible, prove formulas for the maximum likelihood degree of various toric varieties. The maximum likelihood degree of a projective variety is the number of complex solutions of a system of rational equations that arise from the variety.

A parametric probability model for a discrete random variable is given by a map

$\psi : U \to \Delta_{n-1}$ where $U \subset \mathbb{R}^d$ is an open set and Δ_{n-1} is the probability simplex

$$\Delta_{n-1} = \{(p_1, p_2, \dots, p_n) \in \mathbb{R}^n \ : \ p_1 + p_2 + \cdots + p_n = 1, \ p_i \geq 0\}.$$

The model is $\psi(U)$, where $\psi = (\psi_1(\theta), \dots, \psi_n(\theta))$ and $\theta = (\theta_1, \dots, \theta_d)$. The problem of ***maximum likelihood estimation*** for a fixed data vector $u = (u_1, \dots, u_n)$ is to find a parameter θ that best explains the data vector, leading us to the problem of maximizing

$$\psi_1(\theta)^{u_1} \cdots \psi_n(\theta)^{u_n} \quad \text{subject to} \quad \psi_1 + \cdots + \psi_n = 1.$$

[8]

This optimization problem is equivalent to maximizing the ***log-likelihood function***

$$\ell = \sum_{i=1}^{n} u_i \log \psi_i \quad \text{subject to} \quad \psi_1 + \cdots + \psi_n = 1.$$

If in the above definition $\psi_1, \ldots, \psi_n$ are polynomials in $(\theta_1, \ldots, \theta_d)$, the Zariski closure of $\psi(U)$ is called an ***algebraic statistical model***. To employ tools from algebraic geometry we relax the domain of ψ to be the complex numbers

$$\psi : \mathbb{C}^d \to \mathbb{C}^n \quad \text{such that} \quad \psi_1 + \cdots + \psi_n = 1, \ \text{where} \quad \psi_i \in \mathbb{R}[\theta_1, \ldots, \theta_d].$$

Definition 1.1. ***The maximum likelihood degree*** (ML degree) of a discrete statistical model is the number of complex critical points of the log-likelihood function $\sum_{i=1}^{n} u_i \log \psi_i$ for generic data $u = (u_1, \ldots, u_n)$.

Equivalently, the ML degree is the number of complex solutions to the following system of equations which we will call ***the likelihood equations***

$$\frac{\partial(\sum_{i=1}^{n} u_i \log \psi_i)}{\partial \theta_1} = N \frac{\partial g}{\partial \theta_1},$$
$$\vdots$$
$$\frac{\partial(\sum_{i=1}^{n} u_i \log \psi_i)}{\partial \theta_n} = N \frac{\partial g}{\partial \theta_n},$$

where $g = \psi_1 + \cdots + \psi_n - 1$ and N is the sample size $\sum_{i=1}^{n} u_i$.

The maximum likelihood degree was originally defined in [3] where under suitable hypotheses, Fabrizio Catanese, Serkan Hosten, Amit Khetan, and Bernd Sturmfels were able to relate the ML degree to the number of bounded regions in the corresponding arrangement of hypersurfaces, and to the Euler characteristic of the complexified complement. The first symbolic algorithm for computing the ML degree of a projective variety was presented in [7].

Example 1.1. Let us consider the Hardy–Weinberg curve [8] as our first example for computing the ML degree:

$$\psi_0(\theta) = \theta^2$$
$$\psi_1(\theta) = 2\theta(1-\theta) \tag{1.1}$$
$$\psi_2(\theta) = (1-\theta)^2,$$

where the parameter $\theta \in \mathbb{R}$ is the probability that a biased coin lands on tails. Suppose that the coin is tossed twice. Then the equations in 1.1 represent the probability of heads appearing in the toss:

$$\psi_0(\theta) = \text{probability of 0 heads}$$
$$\psi_1(\theta) = \text{probability of 1 head}$$
$$\psi_2(\theta) = \text{ probability of 2 heads.}$$

$$u_0 = \text{number of times 0 heads were observed}$$
$$u_1 = \text{number of times 1 head was observed}$$
$$u_2 = \text{number of times 2 heads were observed.}$$

We repeat the experiment N times and construct a data vector $u = (u_0, u_1, u_2)$

where u_i is the number of times i heads appear. Note $N = u_0 + u_1 + u_2$, which is called the sample size. The likelihood function is

$$\ell_{u_0,u_1,u_2} = \psi_0(\theta)^{u_0}\psi_1(\theta)^{u_1}\psi_2(\theta)^{u_2} = (\theta^2)^{u_0}(2\theta(1-\theta))^{u_1}((1-\theta)^2)^{u_2}.$$

One can use various techniques from optimization to maximize the likelihood function over $U \subset \mathbb{R}$, where U is an open subset. The maximum likelihood degree is the number of complex critical points of the log-likelihood function $\log \ell_{u_0,u_1,u_2}$ where

$$\log \ell_{u_0,u_1,u_2} = (2u_0 + u_1)\log\theta + (u_1 + 2u_2)\log(1-\theta).$$

Remark. We used Lagrange multipliers to maximize $\log \ell_{u_0,u_1,u_2}$ with the constraint equation $\psi_0 + \psi_1 + \psi_2 = 1$, but in this example $\theta^2 + 2(\theta(1-\theta)) + (1-\theta)^2 = \theta^2 - 2\theta + 2\theta + (1 - 2\theta + \theta^2) = 1$ is automatically satisfied, so we really do not have a constraint. Taking the derivative of the log-likelihood function we obtain the equation

$$(2u_0 + 2u_1 + 2u_2)\theta - (2u_0 + u_1) = 0.$$

This is a linear polynomial with one complex root. From this we see that the ML degree of the Hardy–Weinberg curve is one.

1.1 Maximum Likelihood Degree of Toric Models

In this thesis our focus will be on maximum likelihood estimation and maximum likelihood degree of a special class of algebraic statistical models. These models are called ***toric models***. In toric models the polynomials $\psi_i(\theta)$ for $i \in \{1, \dots, n\}$ are monomials, i.e. $\psi_i(\theta) = \theta_1^{a_{i1}}\theta_2^{a_{i2}} \cdots \theta_n^{a_{in}}$; for more information on toric varieties one may refer to [1] or [5]. In algebraic statistics this type of models are also known as ***discrete exponential models***.

Toric varieties $\mathbf{V}(\psi_1, \dots, \psi_n)$ where ψ_i are monomials form an important and rich class of examples in algebraic geometry. For projective toric varieties some information about the variety is also encoded in a polytope, which creates a powerful connection of the subject with convex geometry. Also working with toric models will enable us to use Birch's theorem (Theorem 2.4) to compute the maximum likelihood degree.

Example 1.2. This example is known as the rational normal curve of degree 3, aka the twisted cubic. We will show that the ML degree of this curve is 3. The twisted cubic is the closure in $\mathbb{P}^3$ of the following map:

$$\varphi : (\mathbb{C}^*)^2 \to (\mathbb{C}^*)^4$$
$$\varphi : (s,t) \longmapsto (s, st, st^2, st^3).$$

Let $u = (u_0, u_1, u_2, u_3)$ and $\sum_{i=0}^{3} st^i = 1$ be our constraint. The log-likelihood function is $\ell_{u_0,u_1,u_2,u_3} = \sum_{i=0}^{3} u_i \log(s) + \sum_{i=1}^{3} iu_i \log(t)$. To maximize the log-likelihood function, we obtain the following critical equations:

$$\frac{u_0 + u_1 + u_2 + u_3}{s} = N(1 + t + t^2 + t^3) \tag{1.2}$$
$$\frac{u_1 + 2u_2 + 3u_3}{t} = N(s + 2st + 3st^2) \tag{1.3}$$

First we can clear the denominators and denote 1.2 by f and 1.3 by g; then to find the number of solutions we can use the resultant of f and g to eliminate s [4]. As such we obtain the Sylvester matrix and the resultant:

$$\text{Syl}(f, g, s) = \begin{bmatrix} N(t^3 + t^2 + t + 1) & N(3t^3 + 2t^2 + t) \\ -(u_0 + u_1 + 2u_2 + 3u_3) & -(u_1 + 2u_2 + 3u_3) \end{bmatrix}$$

$$\text{Res}(f, g, s) = (3a - b)Nt^3 + (2a - b)Nt^2 + (a - b)Nt - bN,$$

where $a = u_0 + u_1 + u_2 + u_3$ and $b = u_1 + 2u_2 + 3u_3$. $\text{Res}(f, g, s)$ is a polynomial in t of degree 3 . By the fundamental theorem of algebra, we have 3 solutions. One may note that we can find the special data vector u for which the maximum likelihood degree goes down. For example if we let $3a = b$, then the ML degree is 2. Later in the thesis we will return to this example.

1.2 Brief Overview of Results

In this thesis we provide explicit formulas for the maximum likelihood degree of several well known toric varieties.

Definition 1.2. The ***rational normal curve*** of degree n denoted $\mathcal{C}_n$ is the closure in $\mathbb{P}^n$ of the image of the map

$$\psi : (\mathbb{C}^*)^2 \to (\mathbb{C}^*)^{n+1}$$
$$\psi : (s, t) \longmapsto (s, st, st^2, \ldots, st^n).$$

Theorem. The ML degree of $\mathcal{C}_n$ is n.

Note we will later on denote this Theorem 3.1. Later on we will prove and expand beyond this formula. Our proof will not only calculate the ML degree, but it will also identify the Zariski closed set where the data u is not generic and subsequently the ML degree drops. We will further explore curves defined by $\psi_i = c_i s t^i$ where c_i is a complex constant:

$$\psi_c : (\mathbb{C}^*)^2 \to (\mathbb{C}^*)^{n+1}$$
$$\psi_c : (s,t) \longmapsto (c_0 s, c_1 st, c_2 st^2, \ldots, c_n st^n).$$

We will denote by $\mathcal{C}_n^c$ the curve associated to $c = (c_0, \ldots, c_n)$. In this thesis, we will introduce conditions on c that will stratify this family of varieties with respect to their ML degree. As a corollary we will see that when $c_i = \binom{n}{i}$ the ML degree of C_n^c is one.

Definition 1.3. The image of $\mathbb{P}^{d-1}$ under the following map is known as a ***Veronese variety*** and is denoted by $V_{n,d}$:

$$\psi : \mathbb{P}^{d-1} \to \mathbb{P}^{\binom{n+d-1}{n}-1}$$
$$\psi : [\theta_1 : \theta_2 : \cdots : \theta_d] \longmapsto [\theta_1^{\alpha_{1i}} \cdots \theta_d^{\alpha_{di}} \ : \ \alpha_{1i} + \cdots + \alpha_{di} = n \ \text{ for } \ i \in \{1, \ldots, n\}].$$

Theorem. The ML degree of $V_{2,d}$ is 2^{d-1}.

Note later on we will denote this Theorem 3.6. We will also look at a toric variety that is a close cousin of $V_{2,d}$. This variety is associated to a polytope known as the second hypersimplex. For simplicity we will also call the variety the second hypersimplex and denote it by $\Delta(2,d)$.

Definition 1.4. The image of the following map is the ***second hypersimplex*** and is denoted by $\Delta(2,d)$:

$$\psi : \mathbb{P}^{d-1} \to \mathbb{P}^{\binom{d}{2}-1}$$
$$\psi : [\theta_1 : \theta_2 : \cdots : \theta_d] \longmapsto [\theta_1\theta_2 : \cdots : \theta_i\theta_j : \cdots : \theta_{d-1}\theta_d]$$

where $1 \leq i < j \leq d$.

Theorem. For $d \geq 3$ the ML degree of $\Delta(2,d)$ is $2^{d-1} - d$.

Note later on we will denote this Theorem . We can think of the second hypersimplex arising from the complete graph on d vertices, where each edge ij represents $\theta_i\theta_j$: see Figure 3.1. We will get a natural corollary for the ML degree of the variety if we remove one edge from the complete graph.

Later on we will provide observations and conjectures for a particular family of toric varieties known as *Hirzebruch surfaces.*

Definition 1.5. We define the ***Hirzebruch surface*** $H_{a,b}$ as the closure in $\mathbb{P}^{a+b+1}$ of the image of the following map:

$$\psi : (\mathbb{C}^*)^3 \to (\mathbb{C}^*)^{a+b+2}$$
$$\psi : (s,t,v) \to (s, st, st^2, \ldots, st^a, sv, stv, \ldots, st^bv)$$

where $1 < a \leq b$.

1.3 Outline

In Chapter 2 we introduce definitions and theorems that we use in this thesis, mainly focusing on projective space in Section 2.1, elimination techniques in Section 2.2, and dedicating Section 2.3 to Birch's theorem. In Chapter 3 we will focus on our results and conjectures. In Section 3.1 we observe the rational normal curve and a natural generalization of it. In Section 3.2 we look at the second Veronese, proving its ML degree. In Section 3.3 we look at the second hypersimplex and a corollary where we observe the ML degree of the second hypersimplex with one missing edge. In Section 3.5 we propose conjectures for Hirzebruch surfaces with support of computational evidence.

Chapter 2

Background

This chapter is dedicated to give the necessary tools to the reader, enabling him or her to be able to follow the thesis and provide references for further exploration in the subject.

2.1 Basic Definitions

As in the Hardy–Weinberg Example 1.1, the objects of study in this thesis are known as algebraic varieties. From now on we will always consider $K = \mathbb{C}$, unless otherwise specified.

Definition 2.1 ([4])**.** Let K be a field and let $f_1, \dots, f_s$ be polynomials in $K[x_1, \dots, x_n]$. Then we define the set

$$\mathbf{V}(f_1, \ldots, f_s) = \{(a_1 : \cdots : a_n) \in K^n : f_i(a_1, \ldots, a_n) = 0 \text{ for all } 1 \leq i \leq s\}$$

to be the ***affine variety*** defined by $f_1, \ldots, f_s$.

All the varieties in this thesis are parametrized. This means that for a given variety $V = \mathbf{V}(f_1, \ldots, f_s) \subset K^n$ there exists a polynomial parametric representation of V consisting of polynomial functions $r_1, \ldots, r_n \in K[\theta_1, \ldots, \theta_d]$ such that the Zariski closure of

$$\begin{aligned} x_1 &= r_1(\theta_1, \ldots, \theta_d) \\ x_2 &= r_2(\theta_1, \ldots, \theta_d) \\ &\vdots \\ x_n &= r_n(\theta_1, \ldots, \theta_d) \end{aligned}$$

is equal to V.

We have introduced affine varieties. However, we will work with projective varieties in projective space.

Definition 2.2 ([4])**.** The n-dimensional ***projective space*** over the field K, denoted $\mathbb{P}^n(K)$, is the set of equivalence classes of $\sim$ on $K^{n+1} - \{0\}$, where we define the equivalence relation $(x_0, \ldots, x_n) \sim (y_0, \ldots, y_n)$ if and only if there is a non-zero

λ such that $(x_0, \ldots, x_n) = \lambda(y_0, \ldots, y_n)$. Thus,

$$\mathbb{P}^n(K) = (K^{n+1} - \{0\})/\sim .$$

Each nonzero $(n+1)$-tuple $[x_0 : \cdots : x_n]$ defines a point p in $\mathbb{P}^n(K)$, and we say that $[x_0 : \cdots : x_n]$ are ***homogeneous coordinates*** of p. Note that we will denote $\mathbb{P}^n(K)$ by $\mathbb{P}^n$ from this point on.

Definition 2.3 ([4]). Let K be a field and let $f_1, \ldots, f_s$ be homogeneous polynomials in $K[x_0, \ldots, x_n]$. Then we define the set

$$\mathbf{V}(f_1, \ldots, f_s) = \{[a_0 : \cdots : a_n] \in K^n : f_i(a_0, \ldots, a_n) = 0 \text{ for all } 1 \leq i \leq s\}$$

to be the ***projective variety*** defined by $f_1, \ldots, f_s$.

We will illustrate an important example of a projective variety in $\mathbb{P}^3$ known as the twisted cubic.

Example 2.1. Given the homogeneous coordinates $[x : y : z : w] \in \mathbb{P}^3$, we define the twisted cubic to be the zero locus of the polynomials $f_1 = xz - y^2$, $f_2 = yw - z^2$, $f_3 = xw - yz$,

$$\mathbf{V}(f_1, f_2, f_3) = \{[x : y : z : w] \in \mathbb{P}^3 : f_i(x, y, z, w) = 0 \text{ for all } 1 \leq i \leq 3\}.$$

Equivalently, it is given parametrically as the image of the map

$$\varphi : \mathbb{P} \to \mathbb{P}^3$$
$$\varphi : [s : t] \to [s^3 : s^2t : st^2 : t^3].$$

In algebraic geometry the twisted cubic is an example of a rational normal curve. Furthermore the twisted cubic is a Veronese variety denoted $V_{3,2}$ [5]. We will use this example to show how one may change from parametrization to implicit equations.

Let $V = \mathbf{V}(xz - y^2, yw - z^2, xw - yz)$. We aim to show that V is the image of φ. We will do that by showing

$$K[x, y, z, w]/\left\langle xz - y^2, yw - z^2, xw - yz\right\rangle = K[s^3, s^2t, st^2, t^3].$$

We can immediately see that if we define the pullback

$$\varphi^* : K[x, y, z, w] \to K[s, t]$$
$$\varphi^*(x) = s^3$$
$$\varphi^*(y) = s^2 t$$
$$\varphi^*(z) = st^2$$
$$\varphi^*(w) = t^3,$$

then $\langle xz - y^2, yw - z^2, xw - yz\rangle \subset \ker \varphi^*$. We have to show the following equality $\langle xz - y^2, yw - z^2, xw - yz\rangle = \ker \varphi^*$. Thus let $f \in \ker \varphi^*$. Just following the definition of being in the kernel we know $f(s^3, s^2t, st^2, t^3) = 0$. This implies that for $f(x, y, z, w) = h_1(xz - y^2) + h_2(yw - z^2) + h_3(xw - yz) + r(x, y, z, w)$. The terms that may appear in $r(x, y, z, w)$ may only be of the form $x^i w^j$, $x^i w^j y$, and $x^i w^j z$. If we plug in s, t we get $r(s^3, s^2t, st^2, t^3) = 0$, but the terms of r are of the form $s^{3i}t^{3j}, s^{3i+2}t^{3j+1}$, and $s^{3i+1}t^{3j+2}$. With these terms we cannot have cancelations, implying that $r(x, y, z, w) = 0$.

There is a strong connection between a projective and an affine space. Using the next proposition one can always obtain an affine variety from a projective variety and vice versa.

Proposition 2.1 ([4]). *Let* $V = \mathbf{V}(f_1, \ldots, f_s)$ *be a projective variety in* $\mathbb{P}^n$*. Then*

$W = V \cap U_0$ where $U_0 = \{[x_0, \ldots, x_n] \in \mathbb{P}^n : x_0 \neq 0\}$ can be identified with the affine variety $\mathbf{V}(g_1, \ldots, g_s) \subset K^n$, where $g_i(y_1, \ldots, y_n) = f_i(1, y_1, \ldots, y_n)$ for each $1 \leq i \leq s$.

Definition 2.4 ([4])**.** Given an affine variety $W \subset K^n$, the ***projective closure*** of W is the projective variety $\overline{W} = \mathbf{V}(\mathbf{I}_a(W)^h) \subset \mathbb{P}^n$, where $\mathbf{I}_a(W)^h \subset K[x_0, \ldots, x_n]$ is the homogenization of $\mathbf{I}_a(W) \subset K[x_1, \ldots, x_n]$, the ideal of all polynomials that vanish on W.

2.2 Elimination Theory

As we encountered the Hardy–Weinberg curve, we observed that finding the number of complex solutions to the likelihood equations was not a difficult task because of the nature of a single variable polynomial. In general the polynomials that define our varieties are in many variables, making life hard. We turn to elimination theory because not only it is used during the proofs in this thesis, but also during the computations that suggested these proofs. First, we will look into a classical tool for eliminating variables, namely ***resultants***, followed by more modern methods, namely ***Gröbner bases***.

Definition 2.5 ([4])**.** Given polynomials $f, g \in K[x]$ of positive degree, write them in the form

$$f = a_0 x^l + \cdots + a_l$$
$$g = b_0 x^m + \cdots + b_m.$$

Then the ***Sylvester matrix*** of f and g with respect to x, denoted $\mathrm{Syl}(f, g, x)$, is the matrix of size $(l+m) \times (l+m)$

$$\mathrm{Syl}(f,g,x) = \begin{bmatrix} a_0 & & & & b_0 & & & \\ a_1 & a_0 & & & b_1 & b_0 & & \\ & a_1 & \ddots & & & b_1 & \ddots & \\ \vdots & & & a_0 & \vdots & & & b_0 \\ a_l & & \ddots & \vdots & b_m & & & \vdots \\ & a_l & & & & b_m & & \\ & & \ddots & \vdots & & & \ddots & \vdots \\ & & & a_l & & & & b_m \end{bmatrix},$$

where the empty spaces are filled with zeros. The ***resultant*** of f and g with respect to x, denoted $\mathrm{Res}(f, g, x)$, is the determinant of the Sylvester matrix. Thus,

$$\mathrm{Res}(f, g, x) = \det(\mathrm{Syl}(f, g, x)).$$

The resultant is widely used in number theory and algebra. Furthermore it is a basic tool of computer algebra, and is a built-in function of most computer algebra systems [4]. One can easily see that if we have a system of polynomials in many indeterminants, we can apply the resultant to pairs of polynomials $f, g \in K[x_0, \ldots, x_n, y]$ to eliminate indeterminates. By [4, pg. 162] we know that $\text{Res}(f, g, y)$ is a polynomial in $K[x_0, \ldots, x_n]$. However this approach is not feasible for large systems of polynomials. Gröbner bases were introduced in 1965 in [2], together with an algorithm to compute them, by Bruno Buchberger. This algorithm is known as the Buchberger's algorithm. Our next tool for elimination is based on Gröbner bases.

Definition 2.6 ([4])**.** Suppose that I is an ideal in $K[x_1, \ldots, x_n]$. Then its ***initial ideal*** $in(I)$ is the ideal generated by the initial terms of all polynomials in I with respect to a term order $\prec$:

$$in_{\prec}(I) = \langle in_{\prec}(f) : f \in I \rangle.$$

Definition 2.7 ([4])**.** A finite subset $\mathcal{G}$ of I is a ***Gröbner basis*** with respect to the term order $\prec$ if the initial terms of the elements in $\mathcal{G}$ suffice to generate the initial ideal $in_{\prec}(I)$.

Definition 2.8 ([4])**.** Given $I = \langle f_1, \ldots, f_s \rangle \subset K[x_1, \ldots, x_n]$ the l-th elimination ideal I_l is the ideal of $K[x_{l+1}, \ldots, x_n]$ defined by

$$I_l = I \cap K[x_{l+1}, \ldots, x_n]$$

Thus, I_l consists of all consequences of $f_1 = \cdots = f_s = 0$ which eliminate the variables $x_1, \ldots, x_l$.

Theorem 2.2 ([4]). *Let $I \subset K[x_1, \ldots, x_n]$ be an ideal and let $\mathcal{G}$ be a Gröbner basis of I with respect to lex order where $x_1 > x_2 > \cdots > x_n$. Then, for every $0 \leq l \leq n$, the set*

$$\mathcal{G}_l = \mathcal{G} \cap K[x_{l+1}, \ldots, x_n]$$

is a Gröbner basis of the l-th elimination ideal I_l.

Definition 2.9 ([4]). Given an ideal $I \subset K[x_0, \ldots, x_n, y_0, \ldots, y_m]$ generated by homogeneous polynomials, the ***projective elimination ideal*** of I is the set

$$\hat{I} = \{f \in K[y_0, \ldots, y_m] : \text{ for each } 0 \leq i \leq n \text{ there is } e_i \geq 0 \text{ with } x_i^{e_i} f \in I\}.$$

In this thesis we will need to eliminate variables in homogeneous ideals.

Proposition 2.3 ([4]). *If $I \subset K[x_0, \ldots, x_n, y_0, \ldots, y_m]$ is an ideal, then for all* *sufficiently large integers e,*

$$\hat{I} = (I : \langle x_0^e, \ldots, x_n^e \rangle) \cap K[y_0, \ldots, y_m].$$

Remark. For more information on projective elimination and Gröbner bases one may look at [4] and [5].

2.3 Birch's Theorem

To compute the ML degree of toric varieties, the following theorem, known as Birch's theorem, will be useful. Recall that we obtain a toric variety by a monomial map $\psi : (\mathbb{C}^*)^d \mapsto (\mathbb{C}^*)^n$ where $\psi = (\theta^{a_1}, \ldots, \theta^{a_n})$. We collect the exponent vectors in a $d \times n$ matrix A.

Theorem 2.4 (Birch's Theorem)**.** *Let $A \in \mathbb{N}^{d\times n}$ and $u \in \mathbb{N}^n$ be a vector of positive counts. The maximum likelihood estimate $\hat{u}$ with respect to the toric model $\overline{\psi((\mathbb{C}^*)^d)}$ is the unique non-negative solution to the simultaneous system of equations*

$$A\hat{u} = Au \qquad \textit{and} \qquad \hat{u} \in \overline{\psi((\mathbb{C}^*)^d)}.$$

Hence the maximum likelihood degree is the number of complex solutions of these equations.

Proof. This proof is taken from [6]. Let $b_1, \ldots, b_l$ be a basis for $\ker A$. The optimization problem we wish to solve is the constrained optimization problem of maximizing $u^T \log(v)$ subject to $b_j^T \log(v) = 0$ for all $j = 1, \ldots, l$ and $\sum_{i=1}^n v_i = m$. Introducing $l+1$ Lagrange multipliers $\lambda_1, \ldots, \lambda_l, \gamma$, the critical points are the solutions to the $n + l + 1$ equations

$$\frac{u_i}{v_i} + \sum_{j=1}^{l} \lambda_j \frac{b_{ij}}{v_i} + \gamma = 0, \qquad b_j^T \log(v) = 0, \qquad \sum_{i=1}^{n} v_i = m.$$

The last two sets of conditions say that v belongs to the toric variety, and that v is a vector of frequencies with sum m. Upon clearing denominators, the first conditions say that $u + \lambda B = -\gamma v$. Now applying A to both sides of the equation implies that $A(u + \lambda B) = A(-\gamma v)$, and using the fact that $B \in \ker A$ we have $Au = A(-\gamma v)$. Now we can let $\hat{u} = -\gamma v$ and we have $Au = A\hat{u}$. The fact that there is a unique positive solution is a consequence of the strict convexity of the likelihood function for a positive u, which is a general property of the class of exponential families that encompass the discrete case discussed in this thesis. We can see this convexity considering that we are optimizing in general $f = u_1 \log(v_1) + u_2 \log(v_2) + \cdots + u_n \log(v_n)$. Thus when we take the proper partial derivative we get $\frac{\partial f}{\partial v_j} = \frac{u_j}{v_j}$. Furthermore when we compute the second partial derivatives for the second derivative test forming the Hessian matrix, we get

$$\frac{\partial^2 f}{\partial v_i \partial v_j} = \begin{cases} 0 & \text{if } i \neq j, \\ \frac{u_j}{v_j^2} & \text{if } i = j. \end{cases}$$

Thus our matrix is

$$\begin{bmatrix} -\frac{u_1}{v_1^2} & 0 & 0 & 0 & 0 \\ 0 & -\frac{u_2}{v_2^2} & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -\frac{u_n}{v_n^2} \end{bmatrix}.$$

This is a negative definite matrix. By calculus we know if the Hessian is negative definite at at a point, then f attains a local maximum at that point, implying a strict local maximizer for f.

The second statement is a tautology because $\{\hat{u} \in \mathbb{C}^n : Au = A\hat{u} \text{ and } \hat{u} \in \overline{\psi((\mathbb{C}^*)^d)}\}$ is the set of solutions to the likelihood equations for generic data u. The cardinality of this set is the ML degree by definition. □

Corollary 2.5. *The ML degree of a projective toric variety is at most the degree of the variety.*

Proof. The degree of a projective variety of dimension d is equal to the number of complex points obtained by intersecting the variety with a generic linear space of codimension d. Birch's theorem implies that the ML degree is the number of complex points we obtain when we intersect a d dimensional toric variety with a *particular* linear space of codimension d. Hence, the ML degree is less than equal to the degree of a variety.

□

Chapter 3

Results

In this chapter we introduce our contributions to the literature. We start with proving the ML degree of the rational normal curve $\mathcal{C}_n$, followed by a generalization of the rational normal curve. We introduce an algorithm for finding c for which the ML degree of $\mathcal{C}_n^c$ drops. Furthermore we prove the ML degree of some classical varieties such as the second Veronese $V_{2,d}$ and the second hypersimplex $\Delta(2,d)$. We will often refer to the ML degree of the variety defined by the parametrization ψ with $\text{MLdeg}(\psi)$.

3.1 Rational Normal Curve

In this section we consider the well known variety called the rational normal curve and a family of close cousins to it.

The rational normal curve of degree n denoted $\mathcal{C}_n$ is the closure in $\mathbb{P}^n$ of the image of the map

$$\varphi : (\mathbb{C}^*)^2 \to (\mathbb{C}^*)^{n+1}$$
$$\varphi : (s,t) \longmapsto (s, st, st^2, \ldots, st^n).$$

Theorem 3.1. *The ML degree of* $\mathcal{C}_n$ *is* n.

Proof. Let

$$\varphi : (\mathbb{C}^*)^2 \to (\mathbb{C}^*)^{n+1}$$
$$\varphi : (s,t) \longmapsto (s, st, st^2, \ldots, st^n),$$

with data vector $u = (u_0, u_1, u_2, u_3, \ldots, u_n)$ and the condition $\sum_{i=0}^{n} st^i = 1$. From this we see that our likelihood equations are

$$\frac{u_0 + u_1 + u_2 + u_3 + \cdots + u_n}{s} = N(1 + t + t^2 + t^3 + \cdots + t^n)$$
$$\frac{u_1 + 2u_2 + 3u_3 + \cdots + nu_n}{t} = N(s + 2st + 3st^2 + \cdots + nst^{n-1}).$$

We clear the denominators and form the Sylvester matrix with respect to s:

$$\text{Syl} = \begin{bmatrix} N(t^n + \cdots + t^3 + t^2 + t + 1) & N(nt^n + \cdots + 3t^3 + 2t^2 + t) \\ -a & -b \end{bmatrix},$$

where $a = u_0 + u_1 + u_2 + u_3 + \cdots + u_n$ and $b = u_1 + 2u_2 + 3u_3 + \cdots + nu_n$. Hence, $\det(\text{Syl}) = (an-b)Nt^n + (a(n-1)-b)Nt^{n-1} + \cdots + (a-b)Nt - bN$ is a polynomial of degree n in t. Thus the ML degree is n.

□

Corollary 3.2. *MLdeg($\mathcal{C}_n$) $< n$ if and only if $na = b$, where $a = u_0 + u_1 + u_2 + u_3 + \cdots + u_n$ and $b = u_1 + 2u_2 + 3u_3 + \cdots + nu_n$*

Proof. Consider the polynomial $\det(\text{Syl}) = (an-b)Nt^n + (a(n-1)-b)Nt^{n-1} + \cdots + (a-b)Nt - bN$. If $na = b$ then the degree of the polynomial will drop. Similarly showing the contrapositive, if $na \neq b$ then this is a polynomial of degree n in t and the ML degree is n. This shows our claim. □

3.1.1 Rational Normal Curves $\mathcal{C}_n^c$

In this section we will ask the natural question "How does the ML degree behave under transformations?". The answer is that the ML degree is reactive to transformations. In this section we explore only one type of transformation,

$$\psi : (\mathbb{C}^*)^2 \to (\mathbb{C}^*)^{n+1}$$
$$\psi : (s,t) \longmapsto (c_0 s, c_1 st, c_2 st^2, \ldots, c_n st^n),$$

where $(c_0, \ldots, c_n) \in (\mathbb{C}^*)^{n+1}$. This is a more generalized version of the rational normal curve. One may think of it as the standard rational normal curve transformed by multiplication with a diagonal matrix that has c_i as its ith diagonal entry. Our goal is to compute the ML degree of $\mathcal{C}_n^c$.

Example 3.1. Consider the case $n = 4$,

$$\psi : (\mathbb{C}^*)^2 \to (\mathbb{C}^*)^5$$
$$\psi : (s,t) \longmapsto (c_0 s, c_1 st, c_2 st^2, c_3 st^3, c_4 st^4).$$

The log-likelihood function is $\ell = \sum_{i=0}^4 u_i \log(c_i st^i)$ where $u = (u_0, u_1, u_2, u_3, u_4)$ is our data vector and $\sum_{i=0}^4 c_i s t^i = 1$ our constraint. The likelihood equations are

$$\frac{\sum_{i=0}^4 u_i}{s} = N(\sum_{i=0}^4 c_i t^i)$$
$$\frac{\sum_{i=1}^4 i u_i}{t} = N(\sum_{i=1}^4 i c_i s t^{i-1}),$$

where $N = \sum u_i$ is the sample size. If we denote $f(t) = \sum_{i=0}^4 c_i t^i$, then $f'(t) = \sum_{i=1}^4 i c_i t^{i-1}$. We use the first equation to obtain $s = \frac{1}{f(t)}$, and via simple substitution we get

$$\frac{\sum_{i=1}^4 i u_i}{t} = N\left(\frac{f'(t)}{f(t)}\right). \tag{3.1}$$

Assuming that our data is generic, by clearing the denominators in 3.1 we obtain a polynomial of degree four. Therefore by the fundamental theorem of algebra, we have four complex solutions. However if f and f' have a common root, then the ML degree would drop. Recall that f and f' have a common root if and only if f has a multiple root. There are different ways f can have multiple roots. We explore the following cases.

Case 1 The polynomial $f(t)$ has at most a double root:

$$f(t) = (\alpha t + \beta)^2(\gamma t + \delta)(\kappa t + \omega)$$
$$f'(t) = 2\alpha(\alpha t + \beta)(\gamma t + \delta)(\kappa t + \omega) + \gamma(\alpha t + \beta)^2(\kappa t + \omega) + \kappa(\alpha t + \beta)^2(\gamma t + \delta).$$

In 3.1 the factor $(\alpha t + \beta)$ cancels, lowering the degree of the polynomial to 3. Therefore the ML degree is 3. Observe that

$$\begin{aligned} f(t) &= (\alpha t + \beta)^2(\gamma t + \delta)(\kappa t + \omega) \\ &= (\alpha^2\gamma\kappa)t^4 + (2\alpha\beta\gamma\kappa + \alpha^2\delta\kappa + \alpha^2\gamma\omega)t^3 \\ &\quad + (\beta^2\gamma\kappa + 2\alpha\beta\delta\kappa + 2\alpha\beta\gamma\omega + \alpha^2\delta\omega)t^2 + (\beta^2\delta\kappa + \beta^2\gamma\omega + 2\alpha\beta\delta\omega)t + \beta^2\delta\omega. \end{aligned}$$

One may use this as an inspection method to determine what ML degree the coefficients yield. If the coefficients are of the form

$$\begin{aligned} c_4 &= \alpha^2\gamma\kappa \\ c_3 &= 2\alpha\beta\gamma\kappa + \alpha^2\delta\kappa + \alpha^2\gamma\omega \\ c_2 &= \beta^2\gamma\kappa + 2\alpha\beta\delta\kappa + 2\alpha\beta\gamma\omega + \alpha^2\delta\omega \\ c_1 &= \beta^2\delta\kappa + \beta^2\gamma\omega + 2\alpha\beta\delta\omega \\ c_0 &= \beta^2\delta\omega, \end{aligned}$$

then we can conclude the ML degree is 3. In practice one may form the ideal $I = \langle c_4 - h_4, c_3 - h_3, c_2 - h_2, c_1 - h_1, c_0 - h_0\rangle$ where h_i are respectively the polynomials in $K[\alpha, \beta, \gamma, \delta, \kappa, \omega]$ stated above. We form the ideal $I_c = I \cap K[c_0, c_1, c_2, c_3, c_4]$. One may use this ideal as a tool to check if a given polynomial $f(t)$ produces ML degree 3. One simply needs to take the coefficients $c = (c_0, \ldots, c_4)$ of the polynomial $f(t)$ and check if the generators of I_c vanish on c. If this is the case then the ML degree is 3. The ideal I_c can be computed via elimination and is

$$
\begin{aligned}
I_c = \langle & c_3^2c_2^2c_1^2 - 4c_4c_2^3c_1^2 - 4c_3^3c_1^3 \\
& + 18c_4c_3c_2c_1^3 - 27c_4^2c_1^4 - 4c_3^2c_2^3c_0 + 16c_4c_2^4c_0 \\
& + 18c_3^3c_2c_1c_0 - 80c_4c_3c_2^2c_1c_0 - 6c_4c_3^2c_1^2c_0 \\
& + 144c_4^2c_2c_1^2c_0 - 27c_3^4c_0^2 + 144c_4c_3^2c_2c_0^2 \\
& - 128c_4^2c_2^2c_0^2 - 192c_4^2c_3c_1c_0^2 + 256c_4^3c_0^3\rangle.
\end{aligned}
$$

Case 2 The polynomial $f(t)$ has two double roots:

$$
\begin{aligned}
f(t) &= (\alpha t + \beta)^2(\gamma t + \delta)^2 \\
f'(t) &= 2\alpha(\alpha t + \beta)(\gamma t + \delta)^2 + \gamma(\alpha t + \beta)^2(\gamma t + \delta).
\end{aligned}
$$

In (3.1) the factors $(\alpha t + \beta)$ and $(\gamma t + \delta)$ cancel, lowering the degree of the polynomial to 2. Therefore the ML degree is 2. Observe that

$$\begin{aligned} f(t) &= (\alpha t + \beta)^2(\gamma t + \delta)^2 \\ &= (\alpha^2\gamma^2)t^4 + (2\alpha^2\delta\gamma + 2\alpha\beta\gamma^2)t^3 + (\alpha^2\delta^2 + 4\alpha\beta\delta\gamma + \beta^2\gamma^2)t^2 \\ &\quad + (2\alpha\beta\delta^2 + 2\beta^2\delta\gamma)t + \beta^2\delta^2 \end{aligned}$$

If the coefficients are of the form

$$\begin{aligned} c_4 &= \alpha^2\gamma^2 \\ c_3 &= 2\alpha^2\delta\gamma + 2\alpha\beta\gamma^2 \\ c_2 &= \alpha^2\delta^2 + 4\alpha\beta\delta\gamma + \beta^2\gamma^2 \\ c_1 &= 2\alpha\beta\delta^2 + 2\beta^2\delta\gamma \\ c_0 &= \beta^2\delta^2, \end{aligned}$$

then the ML degree is 2. One may take the coefficients $c = (c_0, \ldots, c_4)$ of the polynomial $f(t)$ and check if the generators of I_c vanish on c. If this is the case then the ML degree is 2. The ideal I_c can be computed via elimination and is

$$I_c = \langle c_1^3 - 4c_2c_1c_0 + 8c_3c_0^2 c_2c_1^2 - 4c_2^2c_0 + 2c_3c_1c_0 + 16c_4c_0^2,$$

$$c_3c_1^2 - 4c_3c_2c_0 + 8c_4c_1c_0 c_4c_1^2 - c_3^2c_0 c_3^2c_1 - 4c_4c_2c_1 + 8c_4c_3c_0,$$

$$c_3^2c_2 - 4c_4c_2^2 + 2c_4c_3c_1 + 16c_4^2c_0 c_3^3 - 4c_4c_3c_2 + 8c_4^2c_1\rangle.$$

Case 3 The polynomial $f(t)$ has a triple root:

$$f(t) = (\alpha t + \beta)^3(\gamma t + \delta)$$

$$f'(t) = 3\alpha(\alpha t + \beta)^2(\gamma t + \delta) + \gamma(\alpha t + \beta)^3(\gamma t + \delta).$$

In(3.1) the factor $(\alpha t + \beta)^2$ cancel, lowering the degree of the polynomial to 2. Therefore the ML degree is 2. Observe that

$$\begin{aligned} f(t) &= (\alpha t + \beta)^3(\gamma t + \delta) \\ &= (\alpha^3\gamma)t^4 + (\alpha^3\delta + 3\alpha^2\beta\gamma)t^3 + (3\alpha^2\beta\delta + 3\alpha\beta^2\gamma)t^2 + (3\alpha\beta^2\delta + \beta^3\gamma)t + \beta^3\delta. \end{aligned}$$

If the coefficients are of the form

$$
\begin{aligned}
c_4 &= \alpha^3\gamma \\
c_3 &= \alpha^3\delta + 3\alpha^2\beta\gamma \\
c_2 &= 3\alpha^2\beta\delta + 3\alpha\beta^2\gamma \\
c_1 &= 3\alpha\beta^2\delta + \beta^3\gamma \\
c_0 &= \beta^3\delta,
\end{aligned}
$$

then the ML degree is 2. One may take the coefficients $c = (c_0, \ldots, c_4)$ of the polynomial $f(t)$ and check if the generators of I_c vanish on c. If this is the case then the ML degree is 2. The ideal I_c can be computed via elimination and is

$$
\begin{aligned}
I_c = \langle c_2^2 - 3c_3c_1 + 12c_4c_0,\, c_3c_2c_1 - 9c_4c_1^2 - 9c_3^2c_0 + 32c_4c_2c_0,& \\
c_3^2c_1^2 - 3c_4c_2c_1^2 - 3c_3^2c_2c_0 + 28c_4c_3c_1c_0 - 128c_4^2c_0^2\rangle.&
\end{aligned}
$$

Case 4 The polynomial $f(t)$ has a quartic root:

$$f(t) = (\alpha t + \beta)^4$$
$$f'(t) = 4\alpha(\alpha t + \beta)^3.$$

In 3.1 the factor $(\alpha t+\beta)^3$ cancel, lowering the degree of the polynomial making it linear. Therefore the ML is 1. Observe that

$$f(t) = (\alpha t + \beta)^4 = (\alpha^4)t^4 + (4\alpha^3\beta)t^3 + (6\alpha^2\beta^2)t^2 + (4\alpha\beta^3)t + \beta^4.$$

If the coefficients are of the form

$$c_4 = \alpha^4$$
$$c_3 = 4\alpha^3\beta$$
$$c_2 = 6\alpha^2\beta^2$$
$$c_1 = 4\alpha\beta^3$$
$$c_0 = \beta^4,$$

then the ML degree is one. One may take the coefficients $c = (c_0, \ldots, c_4)$ of the polynomial $f(t)$ and check if the generators of I_c vanish on c. In this case the ML degree is one. The ideal I_c can be computed via elimination and is

$$I_c = \langle 3c_1^2 - 8c_2c_0, c_2c_1 - 6c_3c_0, c_3c_1 - 16c_4c_0, c_2^2 - 36c_4c_0, \\ c_3c_2 - 6c_4c_1, 3c_3^2 - 8c_4c_2\rangle.$$

As one can see, this suggests that we can stratify the space of coefficients $(c_0, \dots, c_4)$ with respect to the ML degree of $\mathcal{C}_4^c$ looking at the factors of $f(t)$. The factors of $f(t)$ determine a partition of 4. For instance, 4 distinct roots correspond to the partition $(1, 1, 1, 1)$. The ML degree of $\mathcal{C}_4^c$ is summarized in the following table organized in terms of partitions of 4.

Table 3.1: Partitions of 4 and MLdeg($\mathcal{C}_4^c$).

Partitions of 4	ML Degree
(1,1,1,1)	4
(2,1,1)	3
(2,2)	2
(3,1)	2
(4)	1

The next theorem deals with the general case of $\mathcal{C}_n^c$.

Theorem 3.3. *The ML degree of* $\mathcal{C}_n^c$ *is equal to the number of distinct roots of* $f(t) = \sum_{i=0}^{n} c_i t^i$. *Furthermore for each patrition* λ *of* n, *there exists a set of coefficients* $c = (c_0, \ldots, c_n)$ *such that* $\mathcal{C}_n^c$ *has ML degree equal to the number of parts in* λ.

Proof. Let

$$\psi : (\mathbb{C}^*)^2 \to (\mathbb{C}^*)^{n+1}$$
$$\psi : (s,t) \longmapsto (c_0 s, c_1 st, \ldots, c_n st^n),$$

such that $\sum_{i=0}^{n} c_i st^i = 1$ and data vector $u = (u_0, \ldots, u_n)$. The log-likelihood function is $\ell = \sum_{i=0}^{n} u_i \log(c_i st^i)$. The likelihood equations for the model are

$$\frac{\sum_{i=0}^{n} u_i}{s} = N(\sum_{i=0}^{n} c_i t^i)$$
$$\frac{\sum_{i=1}^{n} iu_i}{t} = N(\sum_{i=1}^{n} ic_i st^{i-1}).$$

If we denote $f(t) = \sum_{i=0}^{n} c_i t^i$, then $f'(t) = \sum_{i=1}^{n} ic_i t^{i-1}$. We use the first equation to obtain $s = \frac{1}{f(t)}$, and via substitution we obtain

$$\frac{\sum_{i=1}^{n} iu_i}{t} = N\left(\frac{f'(t)}{f(t)}\right). \tag{3.2}$$

Let $S = \{(c_0, \ldots, c_n) \ : \ f(t) = \sum_{i=0}^{n} c_i t^i$ has a multiple root $\}$. This set is defined by a single polynomial called the discriminant of f and is denoted $\Delta(f)$. We propose the following symbolic algorithm for calculating the ML degree of the model.

Case 1 If $\Delta(f) \neq 0$ then the MLdeg$(\mathcal{C}_n^c) = n$. This happens because if f and f' have no common root then there is no cancelation happening in 3.2. Therefore the resulting polynomial is of degree n.

Case 2 If $\Delta(f) = 0$ then the MLdeg$(\mathcal{C}_n^c) = k$, where k is the number of distinct roots of f.

$$\begin{aligned} f(t) &= \prod_{i=1}^{k} (\alpha_i t + \beta_i)^{p_i} \\ f'(t) &= \sum_{i=1}^{k} p_i \alpha_i (\alpha_i t + \beta_i)^{p_i - 1} \prod_{j=1, j\neq i}^{k} (\alpha_j t + \beta_j)^{p_j}. \end{aligned}$$

If we substitute in 3.2 we will cancel each common root and thus the overall degree of the polynomial given by the equation 3.2 after clearing denominators is k. □

The next corollary deals with the special case when $\mathrm{MLdeg}(\mathcal{C}_n^c) = 1$.

Corollary 3.4. *The ML degree of $\mathcal{C}_n^c = 1$ if and only if $f(t) = (\alpha t + \beta)^n$, equivalently, if and only if $c_i = \binom{n}{i}\alpha^{n-i}\beta^i$ for all $i \in \{0, \ldots, n\}$ and for some α and β.*

Remark. The coefficients $c = (c_0, \ldots, c_n)$ that give rise to the rational normal curve with ML degree 1 themselves lie on a rational normal curve.

3.2 Second Veronese $V_{2,d}$

In this Section we take a look at the second Veronese $V_{2,d}$. We first state a useful formula from [7], then we provide an example where $d = 4$, and then we will generalize the proof to the general case for any d.

Theorem 3.5. *[7, Theorem 20] Let $\psi : \mathbb{C}^d \mapsto \mathbb{C}^{n+1}$ be a polynomial map. Suppose that ψ is generically finite of degree δ. For generic $u \in \mathbb{N}^{n+1}$, the number of complex solutions to the likelihood equations is equal to δ times* $\mathrm{MLdeg}(\psi)$.

Theorem 3.6. *The ML degree of $V_{2,d}$ is 2^{d-1}.*

Example 3.2. Consider the case $d = 4$

$$\varphi : \mathbb{P}^3 \to \mathbb{P}^9$$

$$\varphi : [\theta_1 : \theta_2 : \theta_3 : \theta_4] \longmapsto [\theta_1^2 : \theta_1\theta_2 : \theta_1\theta_3 : \theta_1\theta_4 : \theta_2^2 : \theta_2\theta_3 : \theta_2\theta_4 : \theta_3^2 : \theta_3\theta_4 : \theta_4^2],$$

where $\sum \theta_i\theta_j = 1$ and data vector $u = (u_{11}, u_{12}, u_{13}, u_{14}, u_{22}, u_{23}, u_{24}, u_{33}, u_{34}, u_{44})$. We have a log-likelihood function $\ell = \sum u_{ij} \log \theta_i\theta_j$ with the constraint $\sum \theta_i\theta_j = 1$. The likelihood equations are

$$\begin{aligned}
\frac{2u_{11} + u_{12} + u_{13} + u_{14}}{\theta_1} &= N(2\theta_1 + \theta_2 + \theta_3 + \theta_4) \\
\frac{u_{12} + 2u_{22} + u_{23} + u_{24}}{\theta_2} &= N(\theta_1 + 2\theta_2 + \theta_3 + \theta_4) \\
\frac{u_{13} + u_{23} + 2u_{33} + u_{34}}{\theta_3} &= N(\theta_1 + \theta_2 + 2\theta_3 + \theta_4) \\
\frac{u_{14} + u_{24} + u_{34} + 2u_{44}}{\theta_4} &= N(\theta_1 + \theta_2 + \theta_3 + 2\theta_4).
\end{aligned} \tag{3.3}$$

We introduce the substitutions $\psi_i = \frac{1}{\theta_i}$, together with

$$\begin{aligned}
a_1 &= 2u_{11} + u_{12} + u_{13} + u_{14} \\
a_2 &= u_{12} + 2u_{22} + u_{23} + u_{24} \\
a_3 &= u_{13} + u_{23} + 2u_{33} + u_{34} \\
a_4 &= u_{14} + u_{24} + u_{34} + 2u_{44}.
\end{aligned}$$

Note that now we have the property that $\psi_i\theta_i = 1$ for $i \in \{1, \ldots, 4\}$. Upon plugging in the substitutions in 3.3 we obtain the system

$$\begin{aligned}
\psi_1 a_1 &= N(2\theta_1 + \theta_2 + \theta_3 + \theta_4) \\
\psi_2 a_2 &= N(\theta_1 + 2\theta_2 + \theta_3 + \theta_4) \\
\psi_3 a_3 &= N(\theta_1 + \theta_2 + 2\theta_3 + \theta_4) \\
\psi_4 a_4 &= N(\theta_1 + \theta_2 + \theta_3 + 2\theta_4),
\end{aligned} \tag{3.4}$$

together with the relations $\psi_i\theta_i = 1$ for $i \in \{1, \ldots, 4\}$. We will homogenize the equations with respect to z, noting that the linear equations in 3.4 stay the same, while $\psi_i\theta_i = 1$ after homogenizing become $\psi_i\theta_i = z^2$. Now our job is to count the number of solutions, not at infinity, meaning $z \neq 0$ (i.e. ,$z = 1$). Our strategy is to count all the solutions including the solutions at infinity ($z = 0$), then show that there are no solutions at infinity. Consider the variety $V = \mathbf{V}(\psi_1\theta_1 = z^2, \psi_2\theta_2 = z^2, \psi_3\theta_3 = z^2, \psi_4\theta_4 = z^2)$. This is a complete intersection variety, meaning when we intersect it with a linear space of complimentary dimension such as 3.4, then the number of intersection points is equal to the product of the degrees of the polynomials defining V. In our case this is 2^4. This is, by definition, the degree of the variety V. We will argue in the following that there are no solutions at infinity. By Theorem 3.5 we know that the ML degree equals to the ML degree in the parametric space scaled by the degree of the map δ. In our case the map has degree $\delta = 2$, thus we have 8 solutions.

Define the set $\mathcal{S} = \{i \ : \ \theta_i \neq 0\}$. Suppose that $|\mathcal{S}| = k$ for some integer $1 \leq k \leq 4$.

Now without loss of generality we assume $\mathcal{S} = \{1, \ldots, k\}$. If $k = 1$ then $\theta_1 \neq 0, \theta_2 = 0, \ldots, \theta_4 = 0$ and $\psi_1 = 0, \psi_2 \neq 0, \ldots, \psi_4 \neq 0$. But this is an immediate contradiction to 3.4 forcing all $\theta_i = 0$. Similarly if $k > 1$ for $\mathcal{S} = \{1, \ldots, k\}$ we take the first k rows in $\theta_1, \ldots \theta_k$ from 3.4 and form the resulting system

$$\begin{aligned} 0 &= N(2\theta_1 + \theta_2 + \cdots + \theta_k) \\ 0 &= N(\theta_1 + 2\theta_2 + \cdots + \theta_k) \\ &\vdots \qquad\qquad \ddots \\ 0 &= N(\theta_1 + \theta_2 + \cdots + 2\theta_k). \end{aligned} \tag{3.5}$$

The matrix of this linear system is

$$A_4 = \begin{bmatrix} 2N & N & N & N \\ N & 2N & N & N \\ N & N & 2N & N \\ N & N & N & 2N \end{bmatrix}.$$

In Proposition 3.7 we will show that this type of matrix is invertible. This shows that there are no solutions at infinity, thus our claim is shown for this special case.

Proof of Theorem 3.6. Let

$$\varphi : \mathbb{P}^{d-1} \to \mathbb{P}^{\frac{d(d+1)}{2}-1}$$

$$\varphi : [\theta_1 : \cdots : \theta_d] \longmapsto [\theta_1^2 : \cdots : \theta_i\theta_j : \cdots : \theta_d^2],$$

where $1 \le i \le j \le d$ and data vector $u = (u_{11}, \ldots, u_{ij}, \ldots, u_{dd})$, subject to $\sum \theta_i\theta_j = 1$. We have a log-likelihood function $\ell = \sum u_{ij} \log \theta_i\theta_j$ with the constraint $\sum \theta_i\theta_j = 1$. The likelihood equations are

$$\begin{aligned}
\frac{2u_{11} + u_{12} + u_{13} + \cdots + u_{1d}}{\theta_1} &= N(2\theta_1 + \theta_2 + \theta_3 + \cdots + \theta_d) \\
\frac{u_{12} + 2u_{22} + u_{23} + \cdots + u_{2d}}{\theta_2} &= N(\theta_1 + 2\theta_2 + \theta_3 + \cdots + \theta_d) \\
\frac{u_{1i} + \cdots + 2u_{ii} + \cdots + u_{id}}{\theta_i} &= N(\theta_1 + \cdots + 2\theta_i + \cdots + \theta_d) \\
&\vdots \\
\frac{u_{1d} + u_{2d} + u_{3d} + \cdots + 2u_{dd}}{\theta_d} &= N(\theta_1 + \theta_2 + \theta_3 + \cdots + 2\theta_d).
\end{aligned} \tag{3.6}$$

We introduce the substitutions $\psi_i = \frac{1}{\theta_i}$ for $i \in \{1, \ldots, d\}$, together with

$$\begin{aligned}
a_1 &= 2u_{11} + u_{12} + u_{13} + u_{14} + \cdots + u_{1(d-1)} + u_{1d} \\
a_2 &= u_{12} + 2u_{22} + u_{23} + u_{24} + \cdots + u_{1(d-1)} + u_{2d} \\
a_3 &= u_{13} + u_{23} + 2u_{33} + u_{34} + \cdots + u_{1(d-1)} + u_{3d} \\
&\vdots \\
a_4 &= u_{14} + u_{24} + u_{34} + u_{44} + \cdots + u_{(d-1)d} + 2u_{dd}.
\end{aligned}$$

Note that now, we have the property that $\psi_i\theta_i = 1$ for $i \in \{1, \ldots, d\}$. Upon plugging in the substitutions in 3.6 we obtain the system

$$\begin{aligned}
\psi_1 a_1 &= N(2\theta_1 + \theta_2 + \cdots + \theta_d) \\
\psi_2 a_2 &= N(\theta_1 + 2\theta_2 + \cdots + \theta_d) \\
\vdots \quad & \qquad\qquad\qquad \ddots \\
\psi_d a_d &= N(\theta_1 + \theta_2 + \cdots + 2\theta_d),
\end{aligned} \tag{3.7}$$

together with the relations $\psi_i\theta_i = 1$ for $i \in \{1, \ldots, d\}$. We homogenize the equations with respect to z, noting that the linear equations in 3.7 stay the same, while $\psi_i\theta_i = 1$ after homogenizing become $\psi_i\theta_i = z^2$ for $i \in \{1, \ldots, d\}$. Similarly to our previous example, we count the number of solutions not at infinity, meaning $z \neq 0$ (i.e.

$z = 1$). Our strategy is simple we count all the solutions including the solutions at infinity ($z = 0$), then show that there are no solutions at infinity. Consider the variety $V = \mathbf{V}(\psi_1\theta_1 = z^2, \ldots, \psi_d\theta_d = z^2)$. V is a complete intersection variety, as such when we intersect V with a linear space of complimentary dimension such as 3.7, then the number of intersection points is equal to the product of the degrees of the polynomials defining V. In our case this is 2^d. This is by definition the degree of the variety V. We will argue in the following that there are no solutions at infinity, which means we have 2^d solutions. By Theorem 3.5 we know that ML degree equals to the ML degree in the parametric space scaled by the degree δ of the map. In our case the map has degree $\delta = 2$, thus we do have 2^{d-1} solutions.

Define the set $\mathcal{S} = \{i \ : \ \theta_i \neq 0\}$. Suppose that $|\mathcal{S}| = k$ for some integer $1 \leq k \leq d$. Now without loss of generality we assume $\mathcal{S} = \{1, \ldots, k\}$. If $k = 1$ then $\theta_1 \neq 0, \theta_2 = 0, \ldots, \theta_d = 0$ and $\psi_1 = 0, \psi_2 \neq 0, \ldots, \psi_d \neq 0$. But this is an immediate contradiction to 3.7 forcing all $\theta_i = 0$. Similarly if $k > 1$ for $\mathcal{S} = \{1, \ldots, k\}$ we take the first k rows in $\theta_1, \ldots \theta_k$ from 3.7 and form the resulting system

$$\begin{aligned}
0 &= N(2\theta_1 + \theta_2 + \cdots + \theta_k) \\
0 &= N(\theta_1 + 2\theta_2 + \cdots + \theta_k) \\
\vdots & \qquad\qquad\qquad \ddots \\
0 &= N(\theta_1 + \theta_2 + \cdots + 2\theta_k).
\end{aligned} \tag{3.8}$$

Consider the matrix that arises from 3.8

$$A_k = \begin{bmatrix} 2N & N & \dots & N \\ N & 2N & \dots & N \\ N & N & \ddots & N \\ N & N & \dots & 2N \end{bmatrix}$$

This is a system of k equations and k indeterminates, and in our next proposition we will show that A_k is invertible. This shows that there are no solutions at infinity, thus we have shown that the ML degree is 2^{d-1}. □

Proposition 3.7. *The matrix A_k is nonsingular.*

Proof. We will show that this $k \times k$ matrix is invertible by showing its determinant is nonzero. Without loss of generality we can work with the following matrix

$$A_k = \begin{bmatrix} 2 & 1 & \dots & 1 \\ 1 & 2 & \dots & 1 \\ 1 & 1 & \ddots & 1 \\ 1 & 1 & \dots & 2 \end{bmatrix}.$$

We apply the row operation $R_1 = R_1 + R_2 + \cdots + R_k$, where R_i denotes the ith row. Thus we obtain the following

$$\overline{A}_k = \begin{bmatrix} k+1 & k+1 & \dots & k+1 \\ 1 & 2 & \dots & 1 \\ 1 & 1 & \ddots & 1 \\ 1 & 1 & \dots & 2 \end{bmatrix}.$$

By using one of the properties of determinants we see

$$\det(\overline{A}_k) = (k+1)\det \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & 1 \\ 1 & 1 & \ddots & 1 \\ 1 & 1 & \dots & 2 \end{bmatrix}.$$

Now to show that the determinant of A_k is nonzero all we need to show is that the determinant of the new matrix is nonzero. We apply the following row operations $R_i = R_i - R_1$ for all $i \in \{2, 3, 4, \dots, k\}$. We obtain the matrix

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ 0 & 1 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

This matrix is nonsingular which implies that it has nonzero determinant, finishing our proof. $\square$

3.3 The Second Hypersimplex $\Delta(2, d)$

Let $\mathcal{A}_d = \{e_i + e_j : 1 \leq i < j \leq d\}$ where e_i and e_j are the standard basis vectors in $\mathbb{R}^d$. Then $\mathcal{A}_d$ is the set of column vectors of the vertex-edge incidence matrix of the complete graph K_d. The convex hull of $\mathcal{A}_d$ is called **the second hypersimplex** of order d denoted with $\Delta(2, d)$. We work with the following variety associated with the second hypersimplex

$$\psi : \mathbb{P}^{d-1} \to \mathbb{P}^{\binom{d}{2}-1}$$

$$\psi : [\theta_1 : \theta_2 : \cdots : \theta_d] \longmapsto [\theta_1\theta_2 : \cdots : \theta_i\theta_j : \cdots : \theta_{d-1}\theta_d]$$

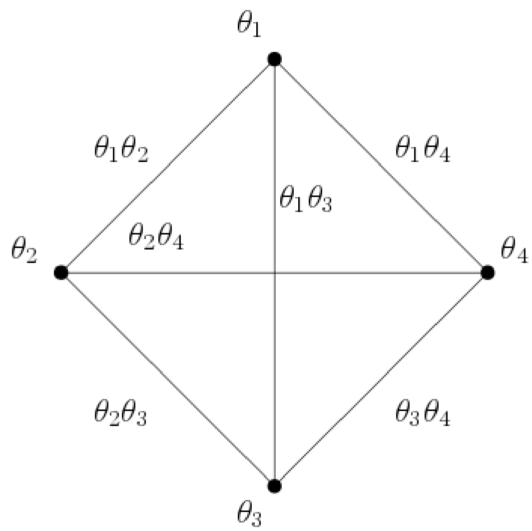
such that $1 \leq i < j \leq d$.

Theorem 3.8. *For $d \geq 3$ the ML degree of $\Delta(2, d)$ is $2^{d-1} - d$.*

First we provide the example $d = 4$ and then we will generalize the proof to the general case for any d.

Example 3.3. Let $u = (u_{12}, u_{13}, u_{14}, u_{23}, u_{24}, u_{34})$ be our data vector. One may obtain the coordinates ψ_i of the polynomial map ψ by looking at the complete graph on four vertices and identifying each edge with a coordinate.

Figure 3.1: Complete graph on four vertices



$$\psi : \mathbb{P}^3 \to \mathbb{P}^5$$

$$\psi : [\theta_1 : \theta_2 : \theta_3 : \theta_4] \longmapsto [\theta_1\theta_2 : \theta_1\theta_3 : \theta_1\theta_4 : \theta_2\theta_3 : \theta_2\theta_4 : \theta_3\theta_4]$$

such that $1 \leq i < j \leq d$. We aim to maximize the log-likelihood function

$$\ell = \sum u_{ij} \log \theta_i\theta_j \text{ for generic data } u = (u_{12}, \ldots, u_{34}).$$

We obtain the likelihood equations

$$\begin{aligned}
\frac{u_{12}+u_{13}+u_{14}}{\theta_1} &= N(\theta_2+\theta_3+\theta_4)\\
\frac{u_{12}+u_{23}+u_{24}}{\theta_2} &= N(\theta_1+\theta_3+\theta_4)\\
\frac{u_{13}+u_{23}+u_{34}}{\theta_3} &= N(\theta_1+\theta_2+\theta_4)\\
\frac{u_{14}+u_{24}+u_{34}}{\theta_4} &= N(\theta_1+\theta_2+\theta_3),
\end{aligned} \tag{3.9}$$

where the sample size $N = u_{12} + u_{13} + u_{14} + u_{23} + u_{24} + u_{34}$. We introduce the substitutions $\phi_i = \frac{1}{\theta_i}$ for $i \in \{1, \ldots, d\}$, together with

$$\begin{aligned}
u_{12}+u_{13}+u_{14} &= a_1\\
u_{12}+u_{23}+u_{24} &= a_2\\
u_{13}+u_{23}+u_{34} &= a_3\\
u_{14}+u_{24}+u_{34} &= a_4.
\end{aligned}$$

Note that now we have the property that $\phi_i\theta_i = 1$ for $i \in \{1, \ldots, 4\}$. Upon plugging in the substitutions in 3.9 we obtain the system

$$
\begin{aligned}
a_1\phi_1 &= N(\theta_2+\theta_3+\theta_4)\\
a_2\phi_2 &= N(\theta_1+\theta_3+\theta_4)\\
a_3\phi_3 &= N(\theta_1+\theta_2+\theta_4)\\
a_4\phi_4 &= N(\theta_1+\theta_2+\theta_3)
\end{aligned}
\tag{3.10}
$$

together with the relations $\phi_i\theta_i = 1$ for $i \in \{1,\dots,4\}$. We homogenize the equations with respect to z, noting that the linear equations in 3.10 stay the same, while $\phi_i\theta_i = 1$ after homogenizing become $\phi_i\theta_i = z^2$ for $i \in \{1,\dots,4\}$. Our job is to count the number of solutions not at infinity, meaning $z \neq 0$ (i.e. $z = 1$). Our strategy is to count all the solutions including the solutions at infinity ($z = 0$), then show that there are only one type of solutions at infinity and then subtract their number from the number of solutions. Consider the variety $V = \mathbf{V}(\phi_1\theta_1 = z^2,\dots,\phi_4\theta_4 = z^2)$. This is a complete intersection variety, as such when we intersect V with a linear space of complimentary dimension such as 3.10, then the number of intersection points is equal to the product of the degrees of the polynomials defining V. In our case this is 2^d. This is by definition the degree of the variety V. We argue in the following that there are d solutions at infinity of multiplicity 2, which means we have $2^d - 2d$ solutions, not at infinity. Since the degree of our map is $\delta = 2$ this will show that the ML degree is $2^{d-1} - d$.

Define the set $\mathcal{S} = \{i \ : \ \theta_i \neq 0\}$. Suppose that $|\mathcal{S}| = k$ for some integer $1 \leq k \leq d$.

Now without loss of generality we assume $\mathcal{S} = \{1, \ldots, k\}$. If $k = 1$ then $\theta_1 \neq 0, \theta_2 = 0, \ldots, \theta_d = 0$ and $\phi_1 = 0, \phi_2 \neq 0, \ldots, \phi_d \neq 0$. In particular by fixing $\theta_1 = 1$ and $\phi_2 = \frac{N}{u_{12}+u_{23}+u_{24}}$, $\phi_3 = \frac{N}{u_{13}+u_{23}+u_{34}}$, $\phi_4 = \frac{N}{u_{14}+u_{24}+u_{34}}$ yields a solution. By symmetry we can see there are three more such solutions making a count of four solutions of the sort $\theta_i = 1$ where $i \in \{1, 2, 3, 4\}$ and letting $\phi_j = \frac{N}{u_{n_0}+u_{n_1}+u_{n_2}}$ where the j ranges over $\{1, 2, 3, 4\} - \{i\}$, and the u_0, u_1, u_2 are the corresponding u's to the ϕ_j in 3.10. We have 4 such solutions, but we will show in the general proof that they are of multiplicity 2; thus there are 8 such solutions at infinity. Next we argue that there are no more solutions at infinity. Thus if $k > 1$ for $\mathcal{S} = \{1, \ldots, k\}$ we take the first k rows in $\theta_1, \ldots \theta_k$ from 3.10 and form the resulting system

$$
\begin{aligned}
0 &= N(\theta_2 + \theta_3 + \cdots + \theta_k) \\
0 &= N(\theta_1 + \theta_3 + \cdots + \theta_k) \\
&\vdots \qquad\qquad \ddots \\
0 &= N(\theta_1 + \theta_2 + \cdots + \theta_{k-1}).
\end{aligned}
\tag{3.11}
$$

The matrix of this linear system is

$$A_4 = \begin{bmatrix} 0 & N & N & N \\ N & 0 & N & N \\ N & N & 0 & N \\ N & N & N & 0 \end{bmatrix}.$$

In Proposition 3.9 we will show that this type of matrix is nonsingular. This shows that there are no other solutions at infinity. Thus we have 16 solutions and 8 of them are at infinity, i.e., we have $16 - 8 = 8$ solutions not at infinity. After dividing by the degree of the map we see that the ML degree of $\Delta(2, 4)$ is 4. Thus our claim is shown for this special case.

Proof of Theorem 3.8. Let

$$\psi : \mathbb{P}^{d-1} \to \mathbb{P}^{\binom{d}{2}-1}$$

$$\psi : [\theta_1, \theta_2, \ldots, \theta_d] \longmapsto [\theta_1\theta_2 : \cdots : \theta_i\theta_j : \cdots : \theta_{d-1}\theta_d]$$

such that $1 \leq i < j \leq d$.

We aim to maximize the log-likelihood function

$$\ell = \sum u_{ij} \log \theta_i\theta_j \text{ for generic data } u = (u_{12}, \ldots, u_{(d-1)d}).$$

The likelihood equations are

$$
\begin{aligned}
\frac{u_{12}+u_{13}+u_{14}+\cdots+u_{1d}}{\theta_1} &= N(\theta_2+\theta_3+\theta_4+\cdots+\theta_d)\\
\frac{u_{12}+u_{23}+u_{24}+\cdots+u_{2d}}{\theta_2} &= N(\theta_1+\theta_3+\theta_4+\cdots+\theta_d)\\
&\vdots\\
\frac{u_{1j}+\cdots+u_{j(j+1)}+\cdots+u_{jd}}{\theta_j} &= N(\theta_1+\theta_2+\theta_3+\cdots+\theta_{j-1}+\theta_{j+1}+\cdots+\theta_d)\\
&\vdots\\
\frac{u_{1d}+u_{2d}+u_{3d}+\cdots+u_{(d-1)d}}{\theta_d} &= N(\theta_1+\theta_2+\theta_3+\cdots+\theta_{d-1})
\end{aligned}
\tag{3.12}
$$

where the sample size $N = u_{12}+u_{13}+\cdots+u_{(d-1)d}$. We introduce the substitutions $\phi_i = \frac{1}{\theta_i}$ for $i \in \{1,\ldots,d\}$, together with

$$
\begin{aligned}
(u_{12}+u_{13}+\cdots+u_{1d}) &= a_1\\
(u_{12}+u_{23}+\cdots+u_{2d}) &= a_2\\
&\vdots\\
(u_{14}+u_{24}+\cdots+u_{(d-1)d}) &= a_d.
\end{aligned}
$$

Note that now, we have the property that $\phi_i\theta_i = 1$ for $i \in \{1,\ldots,d\}$. Upon plugging in the substitutions in 3.12 we obtain the system

$$\begin{aligned} a_1\phi_1 &= N(\theta_2+\theta_3+\cdots+\theta_d)\\ a_2\phi_2 &= N(\theta_1+\theta_3+\cdots+\theta_d)\\ &\vdots\\ a_d\phi_d &= N(\theta_1+\theta_2+\cdots+\theta_{d-1}) \end{aligned} \tag{3.13}$$

together with the relations $\phi_i\theta_i = 1$ for $i \in \{1,\ldots,d\}$. We will homogenize the equations with respect to z, noting that the linear equations in 3.13 stay the same, while $\phi_i\theta_i = 1$ after homogenizing become $\phi_i\theta_i = z^2$ for $i \in \{1,\ldots,d\}$. Similar to our previous example we want to count the number of solutions not at infinity, meaning $z \neq 0$ (i.e. $z = 1$). We do this by counting all the solutions including the solutions at infinity ($z = 0$), then show that there are only one type of solutions at infinity and then subtract their number from the number of solutions. Consider the variety $V = \mathbf{V}(\phi_1\theta_1 = z^2, \ldots, \phi_d\theta_d = z^2)$. V is a complete intersection variety, meaning when we intersect it with a linear space of complimentary dimension such as 3.13, then the number of intersection points is equal to the product of the degrees of the polynomials defining V. In our case this is 2^d. This is by definition the degree of the variety V. We will argue in the following way, there are d solutions at infinity of multiplicity 2, which means we have $2^d - 2d$ solutions not at infinity. Since the degree of our map is $\delta = 2$ this will show that the ML degree is $2^{d-1} - d$.

Define the set $\mathcal{S} = \{i \ : \ \theta_i \neq 0\}$. Suppose that $|\mathcal{S}| = k$ for some integer $1 \leq k \leq d$. Now without loss of generality we assume $\mathcal{S} = \{1, \ldots, k\}$. If $k = 1$ then $\theta_1 \neq 0, \theta_2 = 0, \ldots, \theta_d = 0$ and $\psi_1 = 0, \psi_2 \neq 0, \ldots, \psi_d \neq 0$. In particular by fixing $\theta_1 = 1$ and $\phi_j = \frac{N}{a_j}$ where $j \in \{2, \ldots, d\}$. yields a solution. By symmetry we can see there are d such solutions of the sort $\theta_i = 1$ where $i = \{1, 2, \ldots, d\}$ and letting $\phi_j = \frac{N}{a_j}$ where the j ranges over $\{1, 2, \ldots, d\} - \{i\}$. We have d such solutions, but we will show they are of multiplicity 2;

We can see this by forming the localization at p, $\mathcal{O}_p$, where p is the point $\theta_1 = 1$ and all other θ's are zero, $\phi_1 = 0$, and $\phi_j = \frac{N}{a_j}$ for all $j \neq 1$. Note $\mathcal{O}_p = (\mathcal{M}/\mathcal{I})_{\mathcal{M}}$, where $\mathcal{M} = \langle \theta_1 - 1, \phi_2 - \frac{N}{a_2}, \ldots, \phi_d - \frac{N}{a_d}, \theta_2, \theta_3, \ldots, \theta_d, \phi_1, z \rangle$ and $\mathcal{M}$ is the maximal ideal and $\mathcal{I} = \langle \theta_i \phi_i - z^2 \rangle \cap J$ were J is the ideal formed by the equations 3.13, as one might observe this is a vector space of dimension 2 showing that our point p is of multiplicity 2.

Thus they are the $2d$ solutions such solutions at infinity. Next we argue that there are no more solutions at infinity. Thus if $k > 1$ for $\mathcal{S} = \{1, \ldots, k\}$ we take the first k rows in $\theta_1, \ldots \theta_k$ from 3.13 and form the resulting system

$$
\begin{aligned}
0 &= N(\theta_2 + \theta_3 + \cdots + \theta_k) \\
0 &= N(\theta_1 + \theta_3 + \cdots + \theta_k) \\
\vdots & \qquad\qquad\quad \ddots \\
0 &= N(\theta_1 + \theta_2 + \cdots + \theta_{k-1}).
\end{aligned}
\tag{3.14}
$$

The matrix of this linear system is

$$
B_k = \begin{bmatrix} 0 & N & \dots & N \\ N & 0 & \dots & N \\ N & N & \ddots & N \\ N & N & \dots & 0 \end{bmatrix}
$$

In the next proposition we will show that this type of matrix is invertible. This shows that there are no other solutions at infinity. Thus we have $2^d - 2d$ solutions but by Theorem 3.5 we know we need to divide by the degree of the map which for us is 2 thus we have $2^{d-1} - d$ solutions. □

Proposition 3.9. *The matrix B_k is nonsingular.*

Proof. We will show that this $k \times k$ matrix is invertible by showing its determinant is nonzero. Without loss of generality we can work with the following matrix

$$B_k = \begin{bmatrix} 0 & 1 & \dots & 1 \\ 1 & 0 & \dots & 1 \\ 1 & 1 & \ddots & 1 \\ 1 & 1 & \dots & 0 \end{bmatrix}.$$

We apply the row operation $R_1 = R_1 + R_2 + \cdots + R_k$, where R_i denotes the ith row. Thus we obtain the following

$$\overline{B}_k = \begin{bmatrix} k-1 & k-1 & \dots & k-1 \\ 1 & 0 & \dots & 1 \\ 1 & 1 & \ddots & 1 \\ 1 & 1 & \dots & 0 \end{bmatrix}.$$

By using one of the properties of determinants we see

$$\det(\overline{B}_k) = (k-1)\det \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 0 & \dots & 1 \\ 1 & 1 & \ddots & 1 \\ 1 & 1 & \dots & 0 \end{bmatrix}.$$

Now to show that the determinant of B_k is nonzero all we need to show is that the determinant of the new matrix is nonzero. We apply the row operations $R_i = R_i - R_1$

for all $i \in \{2, 3, 4, \dots, k\}$. We obtain the matrix

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ 0 & -1 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & -1 \end{bmatrix}.$$

This matrix is nonsingular which implies that it has nonzero determinant finishing our proof. □

3.3.1 The Second Hypersimplex With One Missing Edge

In this Section we explore a natural corollary. We observe the second hypersimplex $\Delta(2, d)$ with one missing edge.

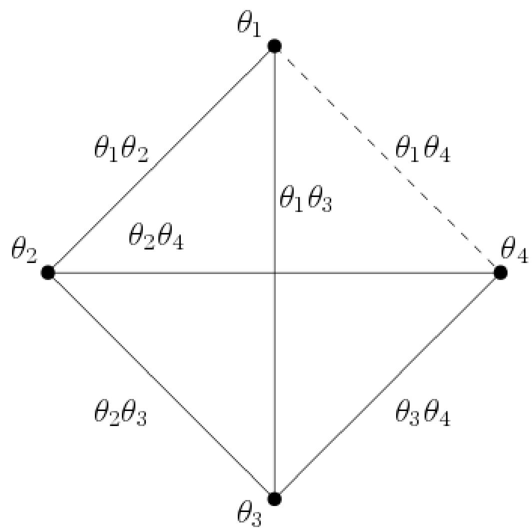
$$\psi : \mathbb{P}^{d-1} \to \mathbb{P}^{\binom{d}{2}-2}$$

$$\psi : [\theta_1, \theta_2, \dots, \theta_d] \longmapsto [\theta_1\theta_2 : \dots : \theta_1\theta_{d-1} : \theta_2\theta_3 : \dots : \theta_i\theta_j : \dots, \theta_{d-1}\theta_d]$$

such that $\sum \theta_i\theta_j = 1$ for $1 \leq i < j \leq d$.

Notice we take $\Delta(2, d)$ and forget the last edge between θ_1 and θ_d. The next corollary, with a little work, comes directly from the proof of Theorem 3.8.

Figure 3.2: Complete graph on four vertices with a missing edge



Corollary 3.10. *Let d be the number of paramaters in θ. If we consider the second hypersimplex missing the edge $\theta_1\theta_d$, then the ML degree is $2^{d-2} - (d-1)$.*

Proof. Let

$$\psi : \mathbb{P}^{d-1} \to \mathbb{P}^{\binom{d}{2}-2}$$
$$\psi : [\theta_1, \theta_2, \ldots, \theta_d] \longmapsto [\theta_1\theta_2 : \cdots : \theta_1\theta_{d-1} : \theta_2\theta_3 : \cdots : \theta_i\theta_j : \cdots : \theta_{d-1}\theta_d]$$

such that $\sum \theta_i\theta_j = 1$ for $1 \leq i < j \leq d$.

Let $u = (u_{12}, u_{13}, u_{23}, u_{24}, u_{34}, \ldots, u_{d-1,d})$, where u_{1d} is not part of the vector, be our data vector. We can obtain our likelihood equations

$$\frac{u_{12}+u_{13}+\cdots+u_{1(d-1)}}{\theta_1} = N(\theta_2+\theta_3+\cdots+\theta_{d-1})$$
$$\frac{u_{12}+u_{23}+u_{24}+\cdots+u_{2d}}{\theta_2} = N(\theta_1+\theta_3+\theta_4+\cdots+\theta_d)$$
$$\frac{u_{13}+u_{23}+u_{34}+\cdots+u_{3d}}{\theta_3} = N(\theta_1+\theta_2+\theta_4+\cdots+\theta_d)$$
$$\vdots$$
$$\frac{u_{2d}+u_{3d}+\cdots+u_{(d-1)d}}{\theta_d} = N(\theta_2+\theta_3+\cdots+\theta_{d-1}),$$

with the condition such that $\sum \theta_i\theta_j = 1$ for $1 \leq i < j \leq d$, we will denote this system L. Now looking at the first and last equations we can see that there is a nice relation

$$\frac{u_{12}+u_{13}+\cdots+u_{1(d-1)}}{\theta_1} = N(\theta_2+\theta_3+\cdots+\theta_{d-1}) = \frac{u_{2d}+u_{3d}+\cdots+u_{(d-1)d}}{\theta_d}.$$

Thus we can use the relation to obtain

$$\theta_d = \frac{u_{2d}+u_{3d}+\cdots+u_{(d-1)d}}{u_{12}+u_{13}+\cdots+u_{1(d-1)}}\theta_1. \qquad (3.15)$$

Notice we can add θ_1 to both sides of 3.15 to obtain

$$\theta_d + \theta_1 = (1 + \frac{u_{2d} + u_{3d} + \cdots + u_{(d-1)d}}{u_{12} + u_{13} + \cdots + u_{1(d-1)}})\theta_1.$$

Introducing the substitution be $\theta_1' = \theta_d + \theta_1$ we may rewrite our likelihood equations, noticing that the first and last equations are dependent. Using the mentioned substitution we have a new form for our L,

$$\begin{aligned}
\frac{u_{12} + u_{13} + \cdots + u_{1(d-1)} + u_{2d} + u_{3d} + \cdots + u_{(d-1)d}}{\theta_1'} &= N(\theta_2 + \theta_3 + \cdots + \theta_{d-1}) \\
\frac{u_{12} + u_{23} + u_{24} + \cdots + u_{2d}}{\theta_2} &= N(\theta_1 + \theta_3 + \theta_4 + \cdots + \theta_d) \\
\frac{u_{13} + u_{23} + u_{34} + \cdots + u_{3d}}{\theta_3} &= N(\theta_1 + \theta_2 + \theta_4 + \cdots + \theta_d) \\
&\vdots \\
\frac{u_{2d} + u_{3d} + \cdots + u_{(d-2)d}}{\theta_{d-1}} &= N(\theta_1 + \theta_2 + \cdots + \theta_{d-2}).
\end{aligned}$$

We can take this one step further by introducing the substitutions

$$
\begin{aligned}
u'_{12} &= u_{12} + u_{2d} \\
u'_{13} &= u_{13} + u_{3d} \\
&\vdots \\
u'_{1(d-1)} &= u_{1(d-1)} + u_{(d-1)d}.
\end{aligned}
$$

Using the substitutions from above our likelihood equations are transformed to

$$
\begin{aligned}
\frac{u'_{12} + u'_{13} + \cdots + u'_{1(d-1)'}}{\theta'_1} &= N(\theta_2 + \theta_3 + \cdots + \theta_{d-1}) \\
\frac{u'_{12} + u_{23} + u_{24} + \cdots + u_{2(d-1)}}{\theta_2} &= N(\theta'_1 + \theta_3 + \theta_4 + \cdots + \theta_d) \\
\frac{u'_{13} + u_{23} + u_{34} + \cdots + u_{3(d-1)}}{\theta_3} &= N(\theta'_1 + \theta_2 + \theta_4 + \cdots + \theta_d) \\
&\vdots \\
\frac{u'_{1d} + u_{3d} + \cdots + u_{(d-2)d}}{\theta_{d-1}} &= N(\theta'_1 + \cdots + \theta_{d-2}).
\end{aligned}
$$

This system of likelihood equations is exactly the same as in Theorem 3.8. Therefore we can apply that theorem and obtain that this system has ML degree $2^{d-2} - (d-1)$ which was indeed our claim. □

3.4 Hirzebruch Surfaces

In this section we show computations of the degree and the ML degree of families of Hirzebruch surfaces.

We defined the Hirzebruch surface $H_{a,b}$ as the closure in $\mathbb{P}^{a+b+1}$ of the image of the following map:

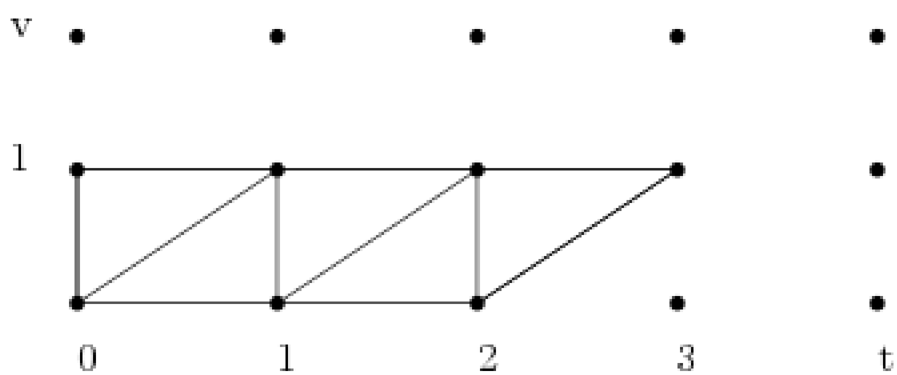
$$\psi : (\mathbb{C}^*)^3 \to (\mathbb{C}^*)^{a+b+2}$$
$$\psi : (s,t,v) \to (s, st, st^2, \ldots, st^a, sv, stv, \ldots, st^b v)$$

where $1 < a \leq b$.

Example 3.4. The surface for $a = 2$ and $b = 3$ is given by

$$\psi(s,t,v) \to (s, st, st^2, sv, stv, st^2v, st^3v).$$

Figure 3.3: $H_{2,3}$



The following tables give some of the data we obtained while computing the ML degree of the Hirzebruch surfaces. We took the matrix representation of ψ where each column is ψ_i. Using Theorem 2.4 we obtained the following computational results

Table 3.2: $H_{2,b}$

(a,b)	Degree	ML Degree
(2,2)	4	2
(2,3)	5	5
(2,4)	6	6
(2,5)	7	5
(2,6)	8	8
(2,7)	9	9
(2,8)	10	8
(2,9)	11	11
(2,10)	12	12

Table 3.3: $H_{n,n+1}$

(a,b)	Degree	ML Degree
(2,3)	5	5
(3,4)	7	7
(4,5)	9	9
(5,6)	11	11

.

In Table 3.2 we observed that the ML degree of $H_{2,b}$ was equal to the the degree of the variety in most cases. Furthermore in the cases the ML degree dropped, it was aways by two. In Table 3.2 we observed that the ML degree of $H_{n,n+1}$ was always equal to the degree of the variety.

3.5 Conclusion

In this thesis we proved formulas for computing the ML degree of classical parameterizations of toric varieties. Furthermore we generalized a few of the mappings and elaborated on some notable special cases, providing corollaries.

In Section 3.1 we observed the rational normal curve $\mathcal{C}_n$ and provided Theorem 3.1 stating the ML degree of $\mathcal{C}_n$ is n. In Section 3.1.1 we generalized the rational normal curve by by multiplication with a diagonal matrix forming $\mathcal{C}_n^c$ and provided Theorem 3.3 which stated that the ML degree of $\mathcal{C}_n^c$ is equal to the number of distinct roots of a polynomial. In Section 3.2 we observed the second Veronese $V_{2,d}$ and proved Theorem 3.6 which stated the ML degree of $V_{2,d}$ is 2^{d-1}. In Section

3.3 we observed the second hypersimplex $\Delta(2,d)$ and provided Theorem 3.8 which stated $\text{MLdeg}(\Delta(2,d)) = 2^{d-1} - d$ and a small Corollary 3.10. Finally in Section 3.5 we provided computation data in Table 3.2 and Table 3.3 for the Hirzebruch surfaces.

During our work we observed that in the cases of the rational normal curve, second Veronese, and the second hypersimplex the ML degree equals to the degree of the variety. This may suggest a more general theorem. Furthermore our computational data suggest that the ML degree of $V_{3,d}$ equals the degree for $V_{3,d}$ as well.
While working on the second hypersimplex and Corollary 3.10, computations suggested that a more general corollary may be possible. We noticed a pattern for removing edges incident to the same vertex. In particular we noticed that for removing k such edges, computations suggest the formula $2^{d-1-k} - (d-k)$.
Finally we often observed that for many of the cases of the Hirzebruch surfaces we had the ML degree equal to the degree noticeable in Table 3.3. Even in Table 3.2 we can speculate that for the majority of the cases the ML degree equals the degree of the variety.

Bibliography

[1] Winfried Bruns and Joseph Gubeladze, *Polytopes, rings, and K-theory*, Springer Monographs in Mathematics, Springer, Dordrecht, 2009.

[2] Bruno Buchberger, *An algorithm for finding the basis elements of the residue class ring of a zero dimensional polynomial ideal*, J. Symbolic Comput. **41** (2006), no. 3-4, 475–511.

[3] Fabrizio Catanese, Serkan Hoşten, Amit Khetan, and Bernd Sturmfels, *The maximum likelihood degree*, Amer. J. Math. **128** (2006), no. 3, 671–697.

[4] David A. Cox, John Little, and Donal O'Shea, *Ideals, Varieties, and Algorithms*, Undergraduate Texts in Mathematics, Springer, Cham, 2015, An introduction to computational algebraic geometry and commutative algebra.

[5] David A. Cox, John B. Little, and Henry K. Schenck, *Toric varieties*, Graduate Studies in Mathematics, vol. 124, American Mathematical Society, Providence, RI, 2011.

[6] Mathias Drton, Bernd Sturmfels, and Seth Sullivant, *Lectures on algebraic statistics*, Oberwolfach Seminars, vol. 39, Birkhäuser Verlag, Basel, 2009.

[7] Serkan Hoşten, Amit Khetan, and Bernd Sturmfels, *Solving the likelihood equations*, Found. Comput. Math. **5** (2005), no. 4, 389–407.

[8] June Huh and Bernd Sturmfels, *Likelihood geometry*, Combinatorial algebraic geometry, Lecture Notes in Math., vol. 2108, Springer, 2014, pp. 63–117.