Likelihood Geometry

June Huh and Bernd Sturmfels

Introduction

Maximum likelihood estimation (MLE) is a fundamental computational problem in statistics, and it has recently been studied with some success from the perspective of algebraic geometry. In these notes we give an introduction to the geometry behind MLE for algebraic statistical models for discrete data. As is customary in algebraic statistics [LiAS], we shall identify such models with certain algebraic subvarieties of high-dimensional complex projective spaces.

The article is organized into four sections. The first three sections correspond to the three lectures given at Levico Terme. The last section will contain proofs of new results.

In Sect. 1, we start out with plane curves, and we explain how to identify the relevant punctured Riemann surfaces. We next present the definitions and basic results for likelihood geometry in $\mathbb{P}^n$. Theorems 1.6 and 1.7 are concerned with the likelihood correspondence, the sheaf of differential one-forms with logarithmic poles, and the topological Euler characteristic. The ML degree of generic complete intersections is given in Theorem 1.10. Theorem 1.15 shows that the likelihood fibration behaves well over strictly positive data. Examples of Grassmannians and Segre varieties are discussed in detail. Our treatment of linear spaces in Theorem 1.20 will appeal to readers interested in matroids and hyperplane arrangements.

Section 2 begins leisurely, with the question *Does watching soccer on TV cause hair loss?* [MSS]. This leads us to conditional independence and low rank matrices.

J. Huh (✉)
Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA
e-mail: junehuh@umich.edu

B. Sturmfels
Department of Mathematics, University of California, Berkeley, CA 94720, USA
e-mail: bernd@berkeley.edu

A. Conca et al., *Combinatorial Algebraic Geometry*, Lecture Notes in Mathematics 2108, DOI 10.1007/978-3-319-04870-3_3,

We study likelihood geometry of determinantal varieties, culminating in the duality theorem of Draisma and Rodriguez [DR]. The ML degrees in Theorems 2.2 and 2.6 were computed using the software `Bertini` [Bertini], underscoring the benefits of using numerical algebraic geometry for MLE. After a discussion of mixture models, highlighting the distinction between rank and nonnegative rank, we end Sect. 2 with a review of recent results in [ARSZ] on tensors of nonnegative rank 2.

Section 3 starts out with toric models [PS, §1.22] and geometric programming [BoydVan, §4.5]. Theorem 3.2 identifies the ML degree of a toric variety with the Euler characteristic of the complement of a hypersurface in a torus. Theorem 3.7 furnishes the ML degree of a variety parametrized by generic polynomials. Theorem 3.10 characterizes varieties of ML degree 1 and it reveals a beautiful connection to the A-discriminant of [GKZ]. We introduce the ML bidegree and the sectional ML degree of an arbitrary projective variety in $\mathbb{P}^n$, and we explain how these two are related. Section 3 ends with a study of the operations of intersection, projection, and restriction in likelihood geometry. This concerns the algebro-geometric meaning of the distinction between sampling zeros and structural zeros in statistical modeling.

In Sect. 4 we offer precise definitions and technical explanations of more advanced concepts from algebraic geometry, including logarithmic differential forms, Chern–Schwartz–MacPherson classes, and schön very affine varieties. This enables us to present complete proofs of various results, both old and new, that are stated in the earlier sections.

We close the introduction with a disclaimer regarding our overly ambitious title. There are many important topics in the statistical study of likelihood inference that should belong to "Likelihood Geometry" but are not covered in this article. Such topics include Watanabe's theory of singular Bayesian integrals [Wat], differential geometry of likelihood in information geometry [AN], and real algebraic geometry of Gaussian models [Uhl]. We regret not being able to talk about these topics and many others. Our presentation here is restricted to the setting of [LiAS, §2.2], namely statistical models for discrete data viewed as projective varieties in $\mathbb{P}^n$.

1 First Lecture

Let us begin our discussion with likelihood on algebraic curves in the complex projective plane $\mathbb{P}^2$. We fix a system of homogeneous coordinates p_0, p_1, p_2 on $\mathbb{P}^2$. The set of real points in $\mathbb{P}^2$ with $\mathrm{sign}(p_0) = \mathrm{sign}(p_1) = \mathrm{sign}(p_2)$ is identified with the open triangle

$$\Delta_2 = \left\{ (p_0, p_1, p_2) \in \mathbb{R}^3 \; : \; p_0, p_1, p_2 > 0 \;\text{ and }\; p_0 + p_1 + p_2 = 1 \right\}.$$

Given three positive integers u_0, u_1, u_2, the corresponding likelihood function is

$$\ell_{u_0,u_1,u_2}(p_0, p_1, p_2) = \frac{p_0^{u_0} p_1^{u_1} p_2^{u_2}}{(p_0 + p_1 + p_2)^{u_0+u_1+u_2}}.$$

This defines a rational function on $\mathbb{P}^2$, and it restricts to a regular function on $\mathbb{P}^2\backslash\mathcal{H}$, where

$$\mathcal{H} \;=\; \big\{ (p_0 : p_1 : p_2) \in \mathbb{P}^2 \;:\; p_0 p_1 p_2 (p_0 + p_1 + p_2) = 0 \big\}$$

is our arrangement of four distinguished lines. The likelihood function ℓ_{u_0,u_1,u_2} is positive on the triangle Δ_2, it is zero on the boundary of Δ_2, and it attains its maximum at the point

$$(\hat{p}_0, \hat{p}_1, \hat{p}_2) = \frac{1}{u_0 + u_1 + u_2}(u_0, u_1, u_2). \tag{1}$$

The corresponding point $(\hat{p}_0 : \hat{p}_1 : \hat{p}_2)$ is the only critical point of the function ℓ_{u_0,u_1,u_2} on the four-dimensional real manifold $\mathbb{P}^2\backslash\mathcal{H}$. To see this, we consider the *logarithmic derivative*

$$\mathrm{dlog}(\ell_{u_0,u_1,u_2}) = \left(\frac{u_0}{p_0} - \frac{u_0+u_1+u_2}{p_0+p_1+p_2}, \frac{u_1}{p_1} - \frac{u_0+u_1+u_2}{p_0+p_1+p_2}, \frac{u_2}{p_2} - \frac{u_0+u_1+u_2}{p_0+p_1+p_2} \right).$$

We note that this equals $(0,0,0)$ if and only if $(p_0 : p_1 : p_2)$ is the point $(\hat{p}_0 : \hat{p}_1 : \hat{p}_2)$ in (1).

Let X be a smooth curve in $\mathbb{P}^2$ defined by a homogeneous polynomial $f(p_0, p_1, p_2)$. This curve plays the role of a statistical model, and our task is to maximize the likelihood function ℓ_{u_0,u_1,u_2} over its set $X \cap \Delta_2$ of positive real points. To compute that maximum algebraically, we examine the set of all critical points of ℓ_{u_0,u_1,u_2} on the complex curve $X\backslash\mathcal{H}$. That set of critical points is the *likelihood locus*. Using Lagrange Multipliers from Calculus, we see that it consists of all points of $X\backslash\mathcal{H}$ such that $\mathrm{dlog}(\ell_{u_0,u_1,u_2})$ lies in the plane spanned by $\mathrm{d}f$ and $(1,1,1)$ in $\mathbb{C}^3$. Thus, our task is to study the solutions in $\mathbb{P}^2\backslash\mathcal{H}$ of the equations

$$f(p_0, p_1, p_2) \;=\; 0 \qquad \text{and} \qquad \det\begin{pmatrix} 1 & 1 & 1 \\ \frac{u_0}{p_0} & \frac{u_1}{p_1} & \frac{u_2}{p_2} \\ \frac{\partial f}{\partial p_0} & \frac{\partial f}{\partial p_1} & \frac{\partial f}{\partial p_2} \end{pmatrix} \;=\; 0. \tag{2}$$

Suppose that X has degree d. Then, after clearing denominators, the second equation has degree $d+1$. By Bézout's Theorem, we expect the likelihood locus to consist of $d(d+1)$ points in $\mathbb{P}^2\backslash\mathcal{H}$. This is indeed what happens when f is a generic polynomial of degree d.

We define the *maximum likelihood degree* (or *ML degree*) of our curve X to be the cardinality of the likelihood locus for generic choices of u_0, u_1, u_2. Thus a general plane curve of degree d has ML degree $d(d+1)$. However, for special curves, the ML degree can be smaller.

Theorem 1.1. *Let X be a smooth curve of degree d in $\mathbb{P}^2$, and $a = \#(X \cap \mathcal{H})$ the number of its points on the distinguished arrangement. Then the ML degree of X equals $d^2 - 3d + a$.*

This is a very special case of Theorem 1.7 which identifies the ML degree with the signed Euler characteristic of $X \backslash \mathcal{H}$. For a general curve of degree d in $\mathbb{P}^2$, we have $a = 4d$, and so $d^2 - 3d + a = d(d+1)$ as predicted. However, the number a of points in $X \cap \mathcal{H}$ can drop:

Example 1.2. Consider the case $d = 1$ of lines. A generic line has ML degree 2. The line $X = V(p_0 + cp_1)$ has ML degree 1 provided $c \notin \{0, 1\}$. The special line $X = V(p_0 + p_1)$ has ML degree 0: (2) has no solutions on $X \backslash \mathcal{H}$ unless $u_0 + u_1 = 0$. In the three cases, $X \backslash \mathcal{H}$ is the Riemann sphere $\mathbb{P}^1$ with four, three, or two points removed. $\diamondsuit$

Example 1.3. Consider the case $d = 2$ of quadrics. A general quadric has ML degree 6. The *Hardy–Weinberg curve*, which plays a fundamental role in population genetics, is given by

$$f(p_0, p_1, p_2) = \det \begin{pmatrix} 2p_0 & p_1 \\ p_1 & 2p_2 \end{pmatrix} = 4p_0p_2 - p_1^2.$$

The curve has only three points on the distinguished arrangement:

$$X \cap \mathcal{H} = \big\{ (1:0:0),\ (0:0:1),\ (1:-2:1) \big\}.$$

Hence the ML degree of the Hardy–Weinberg curve equals 1. This means that the maximum likelihood estimate (MLE) is a rational function of the data. Explicitly, the MLE equals

$$(\hat{p}_0, \hat{p}_1, \hat{p}_2) = \frac{1}{4(u_0 + u_1 + u_2)^2} \big((2u_0+u_1)^2\, ,\ 2(2u_0+u_1)(u_1+2u_2)\, ,\ (u_1+2u_2)^2 \big). \tag{3}$$

In applications, the Hardy–Weinberg curve arises via its parametric representation

$$\begin{aligned} p_0(s) &= s^2 \\ p_1(s) &= 2s(1-s) \\ p_2(s) &= (1-s)^2 \end{aligned} \tag{4}$$

Here the parameter s is the probability that a biased coin lands on tails. If we toss that same biased coin twice, then the above formulas represent the following probabilities:

$$\begin{aligned} p_0(s) &= \text{probability of 0 heads} \\ p_1(s) &= \text{probability of 1 head} \\ p_2(s) &= \text{probability of 2 heads} \end{aligned}$$

Suppose now that the experiment of tossing the coin twice is repeated N times. We record the following counts, where $N = u_0 + u_1 + u_2$ is the sample size of our repeated experiment:

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

The MLE problem is to estimate the unknown parameter s by maximizing

$$\ell_{u_0,u_1,u_2} = p_0(s)^{u_0} p_1(s)^{u_1} p_2(s)^{u_2} = 2^{u_1} s^{2u_0+u_1} (1-s)^{u_1+2u_2}.$$

The unique solution to this optimization problem is

$$\hat{s} \quad = \quad \frac{2u_0 + u_1}{2u_0 + 2u_1 + 2u_2}.$$

Substituting this expression into (4) gives the estimator $\big(p_0(\hat{s}), p_1(\hat{s}), p_2(\hat{s})\big)$ for the three probabilities in our model. The resulting rational function coincides with (3). $\diamondsuit$

The ML degree is also defined when the given curve $X \subset \mathbb{P}^2$ is not smooth, but it counts critical points of ℓ_u only in the regular locus of X. Here is an example to illustrate this.

Example 1.4. A general cubic curve X in $\mathbb{P}^2$ has ML degree 12. Suppose now that X is a cubic which meets $\mathcal{H}$ transversally but has one isolated singular point in $\mathbb{P}^2 \backslash \mathcal{H}$. If the singular point is a *node* then the ML degree of X is 10, and if the singular point is a *cusp* then the ML degree of X is 9. The ML degrees are found by saturating the equations in (2) with respect to the homogenous ideal of the singular point. $\diamondsuit$

Moving beyond likelihood geometry in the plane, we shall introduce our objects in any dimension. We fix the complex projective space $\mathbb{P}^n$ with coordinates $p_0, p_1, \ldots, p_n$, representing probabilities. We summarize the observed data in a vector $u = (u_0, u_1, \ldots, u_n) \in \mathbb{N}^{n+1}$, where u_i is the number of samples in state i. The likelihood function on $\mathbb{P}^n$ given by u equals

$$\ell_u \quad = \quad \frac{p_0^{u_0} p_1^{u_1} \cdots p_n^{u_n}}{(p_0 + p_1 + \cdots + p_n)^{u_0+u_1+\cdots+u_n}}.$$

The unique critical point of this rational function on $\mathbb{P}^n$ is the data point itself:

$$(u_0 : u_1 : \cdots : u_n).$$

Moreover, this point is the global maximum of the likelihood function ℓ_u on the probability simplex Δ_n. Throughout, we identify Δ_n with the set of all positive real points in $\mathbb{P}^n$.

The linear forms in ℓ_u define an arrangement $\mathcal{H}$ of $n+2$ distinguished hyperplanes in $\mathbb{P}^n$. The differential of the logarithm of the likelihood function is the vector of rational functions

$$\operatorname{dlog}(\ell_u) \;=\; \Big(\frac{u_0}{p_0}, \frac{u_1}{p_1}, \ldots, \frac{u_n}{p_n}\Big) - \frac{u_+}{p_+} \cdot (1, 1, \ldots, 1). \tag{5}$$

Here $p_+ = \sum_{i=0}^n p_i$ and $u_+ = \sum_{i=0}^n u_i$. The vector (5) represents a section of the sheaf of differential one-forms on $\mathbb{P}^n$ that have logarithmic singularities along $\mathcal{H}$. This sheaf is denoted

$$\Omega^1_{\mathbb{P}^n}(\log(\mathcal{H})).$$

Our aim is to study the restriction of ℓ_u to a closed subvariety $X \subseteq \mathbb{P}^n$. We will assume that X is defined over the real numbers, irreducible, and not contained in $\mathcal{H}$. Let X_{sing} denote the singular locus of X, and X_{reg} denote $X \backslash X_{\text{sing}}$. When X serves as a statistical model, the goal is to maximize the rational function ℓ_u on the semialgebraic set $X \cap \Delta_n$. To solve this problem algebraically, we determine all critical points of the log-likelihood function $\log(\ell_u)$ on the complex variety X. Here we must exclude points that are singular or lie in $\mathcal{H}$.

Definition 1.5. The *maximum likelihood degree* of X is the number of complex critical points of the function ℓ_u on $X_{\text{reg}} \backslash \mathcal{H}$, for generic data u. The *likelihood correspondence* $\mathcal{L}_X$ is the universal family of these critical points. To be precise, $\mathcal{L}_X$ is the closure in $\mathbb{P}^n \times \mathbb{P}^n$ of

$$\big\{(p, u) \,:\, p \in X_{\text{reg}} \backslash \mathcal{H} \text{ and } \operatorname{dlog}(\ell_u) \text{ vanishes at } p\big\}.$$

We sometimes write $\mathbb{P}^n_p \times \mathbb{P}^n_u$ for $\mathbb{P}^n \times \mathbb{P}^n$ to highlight that the first factor is the probability space, with coordinates p, while the second factor is the data space, with coordinates u. The first part of the following result appears in [Huh1, §2]. A precursor was [HKS, Proposition 3].

Theorem 1.6. *The likelihood correspondence $\mathcal{L}_X$ of any irreducible subvariety X in $\mathbb{P}^n_p$ is an irreducible variety of dimension n in the product $\mathbb{P}^n_p \times \mathbb{P}^n_u$. The map $\mathrm{pr}_1 : \mathcal{L}_X \to \mathbb{P}^n_p$ is a projective bundle over $X_{\text{reg}} \backslash \mathcal{H}$, and the map $\mathrm{pr}_2 : \mathcal{L}_X \to \mathbb{P}^n_u$ is generically finite-to-one.*

See Sect. 4 for a proof. The degree of the map $\mathrm{pr}_2 : \mathcal{L}_X \to \mathbb{P}^n_u$ to data space is the ML degree of X. This number has a topological interpretation as an Euler characteristic, provided suitable assumptions on X are being made. The relationship between the homology of a manifold and critical points of a suitable function on it is the topic of Morse theory.

The study of ML degrees was started in [CHKS, §2] by developing the connection to the sheaf $\Omega^1_X(\log(\mathcal{H}))$ of differential one-forms on X with logarithmic poles along $\mathcal{H}$. It was shown in [CHKS, Theorem 20] that the ML degree of X equals the signed topological Euler characteristic

$$(-1)^{\dim X} \cdot \chi(X \backslash \mathcal{H}),$$

provided X is smooth and the intersection $\mathcal{H} \cap X$ defines a normal crossing divisor in $X \subseteq \mathbb{P}^n$. A major drawback of that early result was that the hypotheses are so restrictive that they essentially never hold for varieties X that arise from statistical models used in practice. From a theoretical point view, this issue can be addressed by passing to a resolution of singularities. However, in spite of existing algorithms for resolution in characteristic zero, these algorithms do not scale to problems of the sizes of interest in algebraic statistics. Thus, whatever computations we wish to do should not be based on resolution of singularities.

The following result due to [Huh1] gives the same topological interpretation of the ML degree. The hypotheses here are much more realistic and inclusive than those in [CHKS, Theorem 20].

Theorem 1.7. *If the very affine variety $X \backslash \mathcal{H}$ is smooth of dimension d, then the ML degree of X equals the signed topological Euler characteristic of* $(-1)^d \cdot \chi(X \backslash \mathcal{H})$.

The term *very affine variety* refers to a closed subvariety of some algebraic torus $(\mathbb{C}^*)^m$. Our ambient space $\mathbb{P}^n \backslash \mathcal{H}$ is a very affine variety because it has a closed embedding

$$\mathbb{P}^n \backslash \mathcal{H} \longrightarrow (\mathbb{C}^*)^{n+1}, \qquad (p_0 : \cdots : p_n) \longmapsto \Big(\frac{p_0}{p_+}, \ldots, \frac{p_n}{p_+}\Big).$$

The study of such varieties is foundational for *tropical geometry*. The special case when $X \backslash \mathcal{H}$ is a Riemann surface with a punctures, arising from a curve in $\mathbb{P}^2$, was seen in Theorem 1.1. We remark that Theorem 1.7 can be deduced from works of Gabber–Loeser [Gabber-Loeser] and Franecki–Kapranov [Franecki-Kapranov] on perverse sheaves on algebraic tori.

The smoothness hypothesis is essential for Theorem 1.7 to hold. If X is singular then, generally, neither $X \backslash \mathcal{H}$ nor $X_{\mathrm{reg}} \backslash \mathcal{H}$ has its signed Euler characteristic equal to the ML degree of X. Varieties X that demonstrate this are the two singular cubic curves in Example 1.4.

Conjecture 1.8. For any projective variety $X \subseteq \mathbb{P}^n$ of dimension d, not contained in $\mathcal{H}$,

$$(-1)^d \cdot \chi(X \backslash \mathcal{H}) \;\geq\; \text{MLdegree}\,(X).$$

In particular, the signed topological Euler characteristic $(-1)^d \cdot \chi(X \backslash \mathcal{H})$ is nonnegative.

Analogous conjectures can be made in the slightly more general setting of [Huh1]. In particular, we conjecture that the inequality

$$(-1)^d \cdot \chi(V) \;\geq\; 0$$

holds for any closed d-dimensional subvariety $V \subseteq (\mathbb{C}^*)^m$.

Remark 1.9. We saw in Example 1.2 that the ML degree of a projective variety X can be 0. In all situations of statistical interest, the variety $X \subset \mathbb{P}^n$ intersects the open simplex Δ_n in a subset that is Zariski dense in X. If that intersection is smooth then $\text{MLdegree}(X) \geq 1$. In fact, arguing as in [CHKS, Proposition 11], it can be shown that for smooth X,

$$\text{MLdegree}\,(X) \;\geq\; \#(\text{bounded regions of } X_{\mathbb{R}} \backslash \mathcal{H}).$$

Here a *bounded region* is a connected component of the semialgebraic set $X_{\mathbb{R}} \backslash \mathcal{H}$ whose classical closure is disjoint from the distinguished hyperplane $V(p_+)$ in $\mathbb{P}^n$.

If X is singular then the number of bounded regions of $X_{\mathbb{R}} \backslash \mathcal{H}$ can exceed $\text{MLdegree}\,(X)$. For instance, let $X \subset \mathbb{P}^2$ be the cuspidal cubic curve defined by

$$(p_0+p_1+p_2)(7p_0-9p_1-2p_2)^2 = (3p_0+5p_1+4p_2)^3.$$

The real part $X_{\mathbb{R}} \backslash \mathcal{H}$ consists of 8 bounded and 2 unbounded regions, but the ML degree of X is 7. The bounded region that contains the cusp $(13 : 17 : -31)$ has no other critical points for ℓ_u. ◇

In what follows we present instances that illustrate the computation of the ML degree. We begin with the case of *generic complete intersections*. Suppose that $X \subset \mathbb{P}^n$ is a complete intersection defined by r generic homogeneous polynomials $g_1, \ldots, g_r$ of degrees $d_1, d_2, \ldots, d_r$.

Theorem 1.10. *The ML degree of X equals $Dd_1d_2\cdots d_r$, where*

$$D \quad = \sum_{i_1+i_2+\cdots+i_r \leq n-r} d_1^{i_1} d_2^{i_2} \cdots d_r^{i_r}. \tag{6}$$

Proof. By Bertini's Theorem, the generic complete intersection X is smooth in $\mathbb{P}^n$. All critical points of the likelihood function ℓ_u on X lie in the dense open subset $X \backslash \mathcal{H}$. Consider the following $(r+2)\times(n+1)$-matrix with entries in the polynomial ring $\mathbb{R}[p_0, p_1, \ldots, p_n]$:

$$\begin{bmatrix} u \\ \tilde{J}(p) \end{bmatrix} \quad = \quad \begin{bmatrix} u_0 & u_1 & \cdots & u_n \\ p_0 & p_1 & \cdots & p_n \\ p_0\frac{\partial g_1}{\partial p_0} & p_1\frac{\partial g_1}{\partial p_1} & \cdots & p_n\frac{\partial g_1}{\partial p_n} \\ p_0\frac{\partial g_2}{\partial p_0} & p_1\frac{\partial g_2}{\partial p_1} & \cdots & p_n\frac{\partial g_2}{\partial p_n} \\ \vdots & \vdots & \ddots & \vdots \\ p_0\frac{\partial g_r}{\partial p_0} & p_1\frac{\partial g_r}{\partial p_1} & \cdots & p_n\frac{\partial g_r}{\partial p_n} \end{bmatrix}. \tag{7}$$

Let Y denote the determinantal variety in $\mathbb{P}^n$ given by the vanishing of its $(r+2)\times(r+2)$ minors. The codimension of Y is at most $n-r$, which is a general upper bound for ideals of maximal minors, and hence the dimension of Y is at least r. Our genericity assumptions ensure that the matrix $\tilde{J}(p)$ has maximal row rank $r+1$ for all $p \in X$. Hence a point $p \in X$ lies in Y if and only if the vector u is in the row span of $\tilde{J}(p)$. Moreover, by Theorem 1.6,

$$(X_{\text{reg}} \backslash \mathcal{H}) \cap Y \quad = \quad X \cap Y$$

is a finite subset of $\mathbb{P}^n$, and its cardinality is the desired ML degree of X.

Since X has dimension $n-r$, we conclude that Y has the maximum possible codimension, namely $n-r$, and that the intersection of X with the determinantal variety Y is proper. We note that Y is Cohen–Macaulay, since Y has maximal codimension $n-r$, and ideals of minors of generic matrices are Cohen–Macaulay. Bézout's Theorem implies

$$\text{MLdegree}(X) \quad = \quad \text{degree}(X)\cdot\text{degree}(Y) \quad = \quad d_1\cdots d_r\cdot\text{degree}(Y).$$

The degree of the determinantal variety Y equals the degree of the determinantal variety given by generic forms of the same row degrees. By the Thom–Porteous–Giambelli formula, this degree is the complete homogeneous symmetric function of degree $\text{codim}(Y) = n-r$ evaluated at the row degrees of the matrix. Here, the row degrees are $0, 1, d_1, \ldots, d_r$, and the value of that symmetric function is precisely D. We conclude that $\text{degree}(Y) = D$. Hence the ML degree of the generic complete intersection $X = \mathcal{V}(g_1, \ldots, g_r)$ equals $D \cdot d_1 d_2 \cdots d_n$. □

Example 1.11 ($r = 1$). A generic hypersurface of degree d in $\mathbb{P}^n$ has ML degree

$$d\cdot D = d + d^2 + d^3 + \cdots + d^n.$$

Example 1.12 ($r = 2, n = 3$). A space curve that is the generic intersection of two surfaces of degree d and e in $\mathbb{P}^3$ has ML degree $de + d^2e + de^2$. ◇

Remark 1.13. It was shown in [HKS, Theorem 5] that (6) is an upper bound for the ML degree of any variety X *of codimension* r that is defined by polynomials of degree $d_1, \ldots, d_r$. In fact, the same is true under the weaker hypothesis that X is cut out by polynomials of degrees $d_1 \geq \cdots \geq d_r \geq d_{r+1} \geq \cdots \geq d_s$, so X need not be a complete intersection. However, the hypothesis $\text{codim}(X) = r$ is essential in order for $\text{MLdegree}(X) \leq (6)$ to hold. That codimension hypothesis was forgotten when this upper bound was cited in [LiAS, Theorem 2.2.6] and in [PS, Theorem 3.31]. Hence these two book references are not correct as stated.

Here is a simple counterexample. Let $n = 3$ and $d_1 = d_2 = d_3 = 2$. Then the bound (6) is the Bézout number 8, and this is also the correct ML degree for a general complete intersection of three quadrics in $\mathbb{P}^3$. Now let X be a general rational normal curve in $\mathbb{P}^3$. The curve X is defined by three quadrics, namely, the 2×2-minors of a 2×3-matrix filled with general linear forms in p_0, p_1, p_2, p_3. Since X is a Riemann sphere with 15 punctures, Theorem 1.7 tells us that $\text{MLdegree}(X) = 13$, and this exceeds the bound of 8. $\diamondsuit$

We now come to a variety that is ubiquitous in statistics, namely the model of *independence* for two binary random variables [LiAS, §1.1]. This model is represented by Segre's quadric surface X in $\mathbb{P}^3$. By this we mean the surface defined by the 2×2-determinant:

$$X \ = \ V(p_{00}p_{11} - p_{01}p_{10}) \ \subset \ \mathbb{P}^3.$$

The surface X is isomorphic to $\mathbb{P}^1 \times \mathbb{P}^1$, so it is smooth, and we can apply Theorem 1.7 to find the ML degree. In other words, we seek to determine the Euler characteristic of the open complex surface $X \backslash \mathcal{H}$ where

$$\mathcal{H} = \big\{ \, p \in \mathbb{P}^3 : p_{00}p_{01}p_{10}p_{11}(p_{00}+p_{01}+p_{10}+p_{11}) = 0 \big\}.$$

To this end, we write $X \ = \ \mathbb{P}^1 \times \mathbb{P}^1$ with coordinates $\big((x_0 : x_1), (y_0 : y_1)\big)$. Our surface is parametrized by $p_{ij} = x_i y_j$, and hence

$$\begin{aligned} X \backslash \mathcal{H} = & \ \ \big(\mathbb{P}^1 \times \mathbb{P}^1\big) \backslash \big\{ x_0x_1y_0y_1(x_0+x_1)(y_0+y_1) = 0 \big\} \\ = & \ \big(\mathbb{P}^1 \backslash \{x_0x_1(x_0+x_1) = 0\}\big) \times \big(\mathbb{P}^1 \backslash \{y_0y_1(y_0+y_1) = 0\}\big) \\ = & \ \ \big(\text{2-sphere} \backslash \{\text{three points}\}\big) \times \big(\text{2-sphere} \backslash \{\text{three points}\}\big). \end{aligned}$$

Since the Euler characteristic is additive and multiplicative,

$$\chi(X \backslash \mathcal{H}) = (-1) \cdot (-1) = 1.$$

This means that the map $u \mapsto \hat{p}$ from the data to the MLE is a rational function in each coordinate. The following "word problem for freshmen" is aimed at finding that function.

Example 1.14. Do this exercise: A biologist friend of yours wishes to test whether two binary random variables are independent. She collects data and records the matrix of counts

$$u = \begin{pmatrix} u_{00} & u_{01} \\ u_{10} & u_{11} \end{pmatrix}.$$

How to ascertain whether u lies close to the independence model

$$X \ = \ V(p_{00}p_{11} - p_{01}p_{10})\,?$$

A statistician who recently started working in her lab explains that, as the first step in the analysis of her data, the biologist should calculate the maximum likelihood estimate (MLE)

$$\hat{p} \;=\; \begin{pmatrix} \hat{p}_{00} & \hat{p}_{01} \\ \hat{p}_{10} & \hat{p}_{11} \end{pmatrix}.$$

Can you help your friend by supplying the formula for $\hat{p}$ as a rational function in u?

The solution to this word problem is as follows. The MLE is the rank 1 matrix

$$\hat{p} \;=\; \frac{1}{(u_{++})^2} \begin{pmatrix} u_{0+} \\ u_{1+} \end{pmatrix} \cdot \begin{pmatrix} u_{+0} & u_{+1} \end{pmatrix}. \tag{8}$$

We illustrate the concepts introduced above by deriving this well-known formula. The likelihood correspondence $\mathcal{L}_X$ of $X = V(p_{00}p_{11} - p_{01}p_{10})$ is the subvariety of $X \times \mathbb{P}^3$ defined by

$$U \cdot (p_{00}, p_{01}, p_{10}, p_{11})^T = 0, \tag{9}$$

where U is the matrix

$$U \;=\; \begin{pmatrix} 0 & -u_{10} - u_{11} & 0 & u_{00} + u_{01} \\ u_{11} + u_{01} & -u_{00} - u_{10} & 0 & 0 \\ u_{11} + u_{10} & 0 & -u_{01} - u_{00} & 0 \\ 0 & 0 & -u_{01} - u_{11} & u_{00} + u_{10} \end{pmatrix}.$$

We urge the reader to derive (9) from Definition 1.5 using a computer algebra system.

Note that the determinant of U vanishes identically. In fact, for generic u_{ij}, the matrix U has rank 3, so its kernel is spanned by a single vector. The coordinates of that vector are given by Cramer's rule, and we find them to be equal to the rational functions in (8).

The locus where the function $u \mapsto \hat{p}$ is undefined consists of those u where the matrix rank of U drops below 3. A computation shows that the rank of U drops to 2 on the variety

$$V(u_{00} + u_{10}, u_{01} + u_{11}) \,\cup\, V(u_{00} + u_{01}, u_{10} + u_{11}),$$

and it drops to 0 on the point $V(u_{00} + u_{01}, u_{10} + u_{11}, u_{01} + u_{11})$. In particular, the likelihood function ℓ_u given by that point u has infinitely many critical points in the quadric X. $\diamondsuit$

We note that all coefficients of the linear forms that define the exceptional loci in $\mathbb{P}^3_u$ for the independence model are positive. This means that data points u with all coordinates positive can never be exceptional. We will prove in Sect. 4 that this usually holds. Let $\mathrm{pr}_1 : \mathcal{L}_X \to \mathbb{P}^n_p$ and $\mathrm{pr}_2 : \mathcal{L}_X \to \mathbb{P}^n_u$ be the projections from the likelihood correspondence to p-space and u-space respectively. We are interested in the fibers of pr_2 over positive points u.

Theorem 1.15. *Let $u \in \mathbb{R}^{n+1}_{>0}$, and let $X \subset \mathbb{P}^n$ be an irreducible variety such that no singular points of any intersection $X \cap \{p_i = 0\}$ lies in the hyperplane at infinity $\{p_+ = 0\}$. Then*

(1) the likelihood function ℓ_u on X has only finitely many critical points in $X_{\mathrm{reg}} \backslash \mathcal{H}$;
(2) if the fiber $\mathrm{pr}_2^{-1}(u)$ is contained in X_{reg}, then its length equals the ML degree of X.

The hypothesis concerning "no singular point" will be satisfied for essentially all statistical models of interest. Here is an example which shows that this hypothesis is necessary.

Example 1.16. We consider the smooth cubic curve X in $\mathbb{P}^2$ that is defined by

$$f \;=\; (p_0 + p_1 + p_2)^3 \;+\; p_0 p_1 p_2.$$

The ML degree of the curve X is 3. Each intersection $X \cap \{p_i = 0\}$ is a triple point that lies on the line at infinity $\{p_+ = 0\}$. The fiber $\mathrm{pr}_2^{-1}(u)$ of the likelihood fibration over the positive point $u = (1:1:1)$ is the entire curve X. ◇

If u is not positive in Theorem 1.15, then the fiber of pr_2 over u may have positive dimension. We saw an instance of this at the end of Example 1.14. Such *resonance loci* have been studied extensively when X is a linear subspace of $\mathbb{P}^n$. See [CDFV] and references therein.

The following cautionary example shows that the length of the scheme-theoretic fiber of $\mathcal{L}_X \to \mathbb{P}^n_u$ over special points u in the open simplex Δ_n may exceed the ML degree of X.

Example 1.17. Let X be the curve in $\mathbb{P}^2$ defined by the ternary cubic

$$f \;=\; p_2(p_1 - p_2)^2 + (p_0 - p_2)^3.$$

This curve intersects $\mathcal{H}$ in eight points, has ML degree 5, and has a cuspidal singularity at

$$P \;:=\; (1:1:1).$$

The prime ideal in $\mathbb{R}[p_0, p_1, p_2, u_0, u_1, u_2]$ for the likelihood correspondence $\mathcal{L}_X$ is minimally generated by five polynomials, having degrees $(3,0), (2,2), (3,1), (3,1), (3,1)$. They are obtained by saturating the two equations in (2) with respect to $\langle p_0 p_2 \rangle \cap \langle p_0 - p_1, p_2 - p_1 \rangle$.

The scheme-theoretic fiber of pr_1 over a general point of X is a reduced line in the u-plane, while the fiber of pr_1 over P is the double line

$$L := \big\{ (u_0 : u_1 : u_2) \in \mathbb{P}^2 : (2u_0 - u_1 - u_2)^2 = 0 \big\}.$$

The reader is invited to verify the following assertions using a computer algebra system:

(a) If u is a general point of $\mathbb{P}^2_u$, then $\mathrm{pr}_2^{-1}(u)$ consists of 5 reduced points in $X_{\mathrm{reg}}\backslash\mathcal{H}$.
(b) If u is a general point on the line L, then the locus of critical points $\mathrm{pr}_2^{-1}(u)$ consists of four reduced points in $X_{\mathrm{reg}}\backslash\mathcal{H}$ and the reduced point P.
(c) If u is the point $(1:1:1) \in L$, then $\mathrm{pr}_2^{-1}(u)$ is a zero-dimensional scheme of length 6. This scheme consists of three reduced points in $X_{\mathrm{reg}}\backslash\mathcal{H}$ and P counted with multiplicity 3.

In particular, the fiber in (c) is not algebraically equivalent to the general fiber (a). This example illustrates one of the difficulties classical geometers had to face when formulating the "principle of conservation of numbers". See [Fulton, Chap. 10] for a modern treatment. $\diamond$

It is instructive to examine classical varieties from projective geometry from the likelihood perspective. For instance, we may study the Grassmannian in its Plücker embedding. Grassmannians are a nice test case because they are smooth, so that Theorem 1.7 applies.

Example 1.18. Let $X = G(2,4)$ denote the Grassmannian of lines in $\mathbb{P}^3$. In its Plücker embedding in $\mathbb{P}^5$, this Grassmannian is the quadric hypersurface defined by

$$p_{12}p_{34} - p_{13}p_{24} + p_{14}p_{23} = 0. \tag{10}$$

As in (7), the critical equations for the likelihood function ℓ_u are the 3×3-minors of

$$\begin{bmatrix} u_{12} & u_{13} & u_{14} & u_{23} & u_{24} & u_{34} \\ p_{12} & p_{13} & p_{14} & p_{23} & p_{24} & p_{34} \\ p_{12}p_{34} & -p_{13}p_{24} & p_{14}p_{23} & p_{14}p_{23} & -p_{13}p_{24} & p_{12}p_{34} \end{bmatrix}. \tag{11}$$

By Theorem 1.6, the likelihood correspondence $\mathcal{L}_X$ is a five-dimensional subvariety of $\mathbb{P}^5 \times \mathbb{P}^5$. The cohomology class of this subvariety can be represented by the bidegree of its ideal:

$$B_X(p,u) \quad = \quad \mathbf{4}p^5 + 6p^4u + 6p^3u^2 + 6p^2u^3 + 2pu^4. \tag{12}$$

This is the *multidegree*, in the sense of [ch3:MS, §8.5], of $\mathcal{L}_X$ with respect to the natural $\mathbb{Z}^2$-grading on the polynomial ring $\mathbb{R}[p,u]$. We can use [ch3:MS, Proposition 8.49] to compute the bidegree from the prime ideal of $\mathcal{L}_X$. Its leading coefficient 4 is the ML degree of X. Its trailing coefficient 2 is the degree of X. The polynomials $B_X(p,u)$ will be studied in Sect. 3.

The prime ideal of $\mathcal{L}_X$ is computed from the equations in (10) and (11) by saturation with respect to $\mathcal{H}$. It is minimally generated by the following eight polynomials in $\mathbb{R}[p,u]$:

(a) one polynomial of degree $(2, 0)$, namely the Plücker quadric,
(b) six polynomials of degree $(1, 1)$, given by 2×2-minors of

$$\begin{pmatrix} p_{12}-p_{34} & p_{13}-p_{24} & p_{14}-p_{23} \\ u_{12}-u_{34} & u_{13}-u_{24} & u_{14}-u_{23} \end{pmatrix} \qquad \text{and}$$

$$\begin{pmatrix} p_{12}+p_{13}+p_{23} & p_{12}+p_{14}+p_{24} & p_{13}+p_{14}+p_{34} & p_{23}+p_{24}+p_{34} \\ u_{12}+u_{13}+u_{23} & u_{12}+u_{14}+u_{24} & u_{13}+u_{14}+u_{34} & u_{23}+u_{24}+u_{34} \end{pmatrix},$$

(c) one polynomial of degree $(2, 1)$, for instance

$$2u_{24}p_{12}p_{34} + 2u_{34}p_{13}p_{24} + (u_{23} + u_{24} + u_{34})p_{14}p_{24}$$
$$- (u_{13} + u_{14} + u_{34})p_{24}^2 - (u_{12} + 2u_{13} + u_{14} - u_{24})p_{24}p_{34}.$$

For a fixed positive data vector $u > 0$, these six polynomials in (b) reduce to three linear equations, and these cut out a plane $\mathbb{P}^2$ inside $\mathbb{P}^5$. To find the four critical points of ℓ_u on $X = G(2, 4)$, we must then intersect the two conics (a) and (c) in that plane $\mathbb{P}^2$.

The ML degree of the Grassmannian $G(r, m)$ in $\mathbb{P}^{\binom{m}{r}-1}$ is the signed Euler characteristic of the manifold $G(r, m)\backslash\mathcal{H}$ obtained by removing $\binom{m}{r} + 1$ distinguished hyperplane sections. It would be very interesting to find a general formula for this ML degree. At present, we only know that the ML degree of $G(2, 5)$ is 26, and that the ML degree of $G(2, 6)$ is 156. By Theorem 1.7, these numbers give the Euler characteristic of $G(2, m)\backslash\mathcal{H}$ for $m \leq 6$. $\diamond$

We end this lecture with a discussion of the delightful case when X is a linear subspace of $\mathbb{P}^n$, and the open variety $X\backslash\mathcal{H}$ is the complement of a *hyperplane arrangement*. In this context, following Varchenko [Varchenko], the likelihood function ℓ_u is known as the *master function*, and the statement of Theorem 1.7 was first proved by Orlik and Terao in [Orlik-Terao]. We assume that X has dimension d, is defined over $\mathbb{R}$, and does not contain the vector $\mathbf{1} = (1, 1, \ldots, 1)$. We can regard X as a $(d + 1)$-dimensional linear subspace of $\mathbb{R}^{n+1}$. The orthogonal complement $X^\perp$ with respect to the standard dot product is a linear space of dimension $n - d$ in $\mathbb{R}^{n+1}$. The linear space $X^\perp + \mathbf{1}$ spanned by $X^\perp$ and the vector $\mathbf{1}$ has dimension $n - d + 1$ in $\mathbb{R}^{n+1}$, and hence can be viewed as subspace of codimension d in $\mathbb{P}^n_u$. In our next formula, the operation $\star$ is the Hadamard product or coordinatewise product.

Proposition 1.19. *The likelihood correspondence $\mathcal{L}_X$ in $\mathbb{P}^n \times \mathbb{P}^n$ is defined by*

$$p \in X \quad \textit{and} \quad u \in p \star (X^\perp + \mathbf{1}). \tag{13}$$

The prime ideal of $\mathcal{L}_X$ is obtained from these constraints by saturation with respect to $\mathcal{H}$.

Proof. If all p_i are non-zero then $u \in p \star (X^\perp + \mathbf{1})$ says that

$$u/p := \left(\frac{u_0}{p_0}, \frac{u_1}{p_1}, \ldots, \frac{u_n}{p_n}\right)$$

lies in the subspace $X^\perp + \mathbf{1}$. Equivalently, the vector obtained by adding a multiple of $(1, 1, \ldots, 1)$ to u/p is perpendicular to X. We can take that vector to be the differential (5). Hence (13) expresses the condition that p is a critical point of ℓ_u on X. $\square$

The intersection $X \cap \mathcal{H}$ is an arrangement of $n + 2$ hyperplanes in $X \simeq \mathbb{P}^d$. For special choices of the subspace X, it may happen that two or more hyperplanes coincide. Taking $\{p_+ = 0\}$ as the hyperplane at infinity, we view $X \cap \mathcal{H}$ as an arrangement of $n + 1$ hyperplanes in the affine space $\mathbb{R}^d$. A region of this arrangement is *bounded* if it is disjoint from $\{p_+ = 0\}$.

Theorem 1.20. *The ML degree of X is the number of bounded regions of the real affine hyperplane arrangement $X \cap \mathcal{H}$ in $\mathbb{R}^d$. The bidegree of the likelihood correspondence $\mathcal{L}_X$ is the h-polynomial of the broken circuit complex of the rank $d+1$ matroid associated with $X \cap \mathcal{H}$.*

We need to explain the second assertion. The hyperplane arrangement $X \cap \mathcal{H}$ consists of the intersections of the $n + 2$ hyperplanes in $\mathcal{H}$ with $X \simeq \mathbb{P}^d$. We regard these as hyperplanes through the origin in $\mathbb{R}^{d+1}$. They define a matroid M of rank $d + 1$ on $n + 2$ elements. We identify these elements with the variables $x_1, x_2, \ldots, x_{n+2}$. For each circuit C of M let $m_C = (\prod_{i\in C} x_i)/x_j$ where j is the smallest index such that $x_j \in C$. The *broken circuit complex* of M is the simplicial complex with Stanley–Reisner ring $\mathbb{R}[x_1, \ldots, x_{n+2}]/\langle m_C : C \text{ circuit of } M \rangle$. See [ch3:MS, §1.1] for Stanley–Reisner basics. The Hilbert series of this graded ring has the form

$$\frac{h_0 + h_1 z + \cdots + h_d z^d}{(1 - z)^{d+1}}.$$

What is being claimed in Theorem 1.20 is that the bidegree of $\mathcal{L}_X$ equals

$$B_X(p, u) \quad = \quad (h_0 u^d + h_1 p u^{d-1} + h_2 p^2 u^{d-2} + \cdots + h_d p^d) \cdot p^{n-d} \qquad (14)$$

Equivalently, this is the class of $\mathcal{L}_X$ in the cohomology ring

$$H^*(\mathbb{P}^n \times \mathbb{P}^n; \mathbb{Z}) = \mathbb{Z}[p, u]/\langle p^{n+1}, u^{n+1} \rangle.$$

There are several (purely combinatorial) definitions of the invariants h_i of the matroid M. For instance, they are coefficients of the following specialization of the *characteristic polynomial*:

$$\chi_M(q + 1) \;=\; q \cdot \Big(h_0 q^d - h_{d-1} q^{d-1} + \cdots + (-1)^{d-1} h_1 q + (-1)^d h_0\Big). \qquad (15)$$

Theorem 1.20 was used in [Huh0] to prove a conjecture of Dawson, stating that the sequence $h_0, h_1, \ldots, h_d$ is log-concave, when M is representable over a field of characteristic zero.

The first assertion in Theorem 1.20 was proved by Varchenko in [Varchenko]. For definitions and characterizations of the characteristic polynomial χ, and many pointers to matroid basics, we refer to [OTBook]. A proof of the second assertion was given by Denham et al. in a slightly different setting [Denham-Garrousian-Schulze, Theorem 1]. We give a proof in Sect. 4 following [Huh1, §3]. The ramification locus of the likelihood fibration $\mathrm{pr}_2 : \mathcal{L}_X \to \mathbb{P}^n_u$ is known as the *entropic discriminant* [SSV].

Example 1.21. Let $d = 2$ and $n = 4$, so X is a plane in $\mathbb{P}^4$, defined by two linear forms

$$\begin{aligned} c_{10}p_0 + c_{11}p_1 + c_{12}p_2 + c_{13}p_3 + c_{14}p_4 &= 0, \\ c_{20}p_0 + c_{21}p_1 + c_{22}p_2 + c_{23}p_3 + c_{24}p_4 &= 0. \end{aligned} \tag{16}$$

Following Theorem 1.20, we view $X \cap \mathcal{H}$ as an arrangement of five lines in the affine plane

$$\{\, p \in X \;:\; p_0 + p_1 + p_2 + p_3 + p_4 \neq 0 \,\} \quad \simeq \quad \mathbb{C}^2.$$

Hence, for generic c_{ij}, the ML degree of X is equal to 6, the number of bounded regions of this arrangement. The condition $u \in p \star (X^\perp + \mathbf{1})$ in Proposition 1.19 translates into

$$\mathrm{rank} \begin{bmatrix} u_0 & u_1 & u_2 & u_3 & u_4 \\ p_0 & p_1 & p_2 & p_3 & p_4 \\ c_{10}p_0 & c_{11}p_1 & c_{12}p_2 & c_{13}p_3 & c_{14}p_4 \\ c_{20}p_0 & c_{21}p_1 & c_{22}p_2 & c_{23}p_3 & c_{24}p_4 \end{bmatrix} \leq 3. \tag{17}$$

The 4×4-minors of this 4×5-matrix, together with the two linear forms defining X, form a system of equations that has six solutions in $\mathbb{P}^4$, for generic c_{ij}. All solutions have real coordinates. In fact, there is one solution in each bounded region of $X\backslash\mathcal{H}$. The likelihood correspondence $\mathcal{L}_X$ is the fourfold in $\mathbb{P}^4 \times \mathbb{P}^4$ given by the Eqs. (16) and (17).

We now illustrate the second statement in Theorem 1.20. Suppose that the real numbers c_{ij} are generic, so M is the uniform matroid of rank three on six elements. The Stanley–Reisner ring of the broken circuit complex of M equals

$$\mathbb{R}[x_1, x_2, x_3, x_4, x_5, x_6]/\langle x_2x_3x_4, x_2x_3x_5, x_2x_3x_6, \ldots, x_4x_5x_6\rangle.$$

The Hilbert series of this graded algebra is

$$\frac{h_0 + h_1 z + h_2 z^2}{(1-z)^3} \quad = \quad \frac{1 + 3z + 6z^2}{(1-z)^3}.$$

We conclude that the bidegree (14) of the likelihood correspondence $\mathcal{L}_X$ equals

$$B_X(p,u) \quad = \quad 6p^4 + 3p^3u + p^2u^2.$$

For special choices of the coefficients c_{ij} in (16), some triples of lines in the arrangement $X \cap \mathcal{H}$ may meet in a point. For such matroids, the ML degree drops from 6 to some integer between 0 and 5. We recommend it as an exercise to the reader to explore these cases. For instance, can you find explicit c_{ij} so that the ML degree of X equals 3? What are the prime ideal and the bidegree of $\mathcal{L}_X$ in that case? How can the ML degree of X be 0 or 1? $\diamond$

It would be interesting to know which statistical model X in $\mathbb{P}^n$ defines the likelihood correspondence $\mathcal{L}_X$ which is a complete intersection in $\mathbb{P}^n \times \mathbb{P}^n$. When X is a linear subspace of $\mathbb{P}^n$, this question is closely related to the concept of *freeness* of a hyperplane arrangement.

Proposition 1.22. *If the hyperplane arrangement $X \cap \mathcal{H}$ in X is free, then the likelihood correspondence $\mathcal{L}_X$ is an ideal-theoretic complete intersection in $\mathbb{P}^n \times \mathbb{P}^n$.*

Proof. For the definition of freeness see §1 in the paper [CDFV] by Cohen, Denman, Falk and Varchenko. The proposition is implied by their [CDFV, Theorem 2.13] and [CDFV, Corollary 3.8]. $\square$

Using Theorem 1.20, this provides a likelihood geometry proof of Terao's theorem that the characteristic polynomial of a free arrangement factors into integral linear forms [Terao].

2 Second Lecture

In our newspaper we frequently read about studies aimed at proving that a behavior or food causes a certain medical condition. We begin the second lecture with an introduction to statistical issues arising in such studies. The "medical question" we wish to address is *Does Watching Soccer on TV Cause Hair Loss?* We learned this amusing example from [MSS, §1].

In a fictional study, 296 British subjects aged between 40 and 50 were interviewed about their hair length and how many hours per week they watch soccer (a.k.a. "football") on TV. Their responses are summarized in the following *contingency table* of format 3×3:

$$U \quad = \quad \begin{array}{r} \\ \le 2\,\text{h} \\ 2\text{–}6\,\text{h} \\ \ge 6\,\text{h} \end{array} \begin{array}{c} \begin{array}{ccc} \text{lots of hair} & \text{medium hair} & \text{little hair} \end{array} \\ \begin{pmatrix} 51 & 45 & 33 \\ 28 & 30 & 29 \\ 15 & 27 & 38 \end{pmatrix} \end{array}$$

For instance, 29 respondents reported having little hair and watching between 2 and 6 h of soccer on TV per week. Based on these data, are these two random variables independent, or are we inclined to believe that watching soccer on TV and hair loss are correlated?

On first glance, the latter seems to be the case. Indeed, being independent means that the data matrix U should be close to a rank 1 matrix. However, all 2×2-minors of U are strictly positive, indeed by quite a margin, and this suggests a positive correlation.

However, this interpretation is deceptive. A much better explanation of our data can be given by identifying a certain hidden random variable. That hidden variable is *gender*. Indeed, suppose that among the respondents 126 were males and 170 were females. Our data matrix U is then the sum of the male table and the female table, maybe as follows:

$$U = \begin{pmatrix} 3 & 9 & 15 \\ 4 & 12 & 20 \\ 7 & 21 & 35 \end{pmatrix} + \begin{pmatrix} 48 & 36 & 18 \\ 24 & 18 & 9 \\ 8 & 6 & 3 \end{pmatrix}. \tag{18}$$

Both of these tables have rank 1, hence U has rank 2. Hence, the appropriate null hypothesis H_0 for analyzing our situation is not independence but it is *conditional independence*:

H_0 : *Soccer on TV and Hair Loss are Independent given Gender.*

And, based on the data U, we most definitely do not reject that null hypothesis.

The key feature of the matrix U above was that it has rank 2. We now define low rank matrix models in general. Consider two discrete random variables X and Y having m and n states respectively. Their *joint probability distribution* is written as an $m \times n$-matrix

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mn} \end{pmatrix}$$

whose entries are nonnegative and sum to 1. Here p_{ij} represents the probability that X is in state i and Y is in state j. The of all probability distributions is the standard simplex Δ_{mn-1} of dimension $mn - 1$. We write $\mathcal{M}_r$ for the manifold of rank r matrices in Δ_{mn-1}.

The matrices P in $\mathcal{M}_1$ represent independent distributions. Mixtures of r independent distributions correspond to matrices in $\mathcal{M}_r$. As always in applied algebraic geometry, we can make any problem that involves semi-algebraic sets progressively easier by three steps:

- disregard inequalities,
- replace real numbers with complex numbers,
- replace affine space by projective space.

In our situation, this leads us to replacing $\mathcal{M}_r$ with its Zariski closure in complex projective space $\mathbb{P}^{mn-1}$. This Zariski closure is the projective variety $\mathcal{V}_r$ of complex $m \times n$ matrices of rank $\leq r$. Note that $\mathcal{V}_r$ is singular along $\mathcal{V}_{r-1}$. The codimension of $\mathcal{V}_r$ is $(m-r)(n-r)$. It is a non-trivial exercise to write the degree of $\mathcal{V}_r$ in terms of m, n, r. Hint: [ch3:MS, Example 15.2].

Suppose now that i.i.d. samples are drawn from an unknown joint distribution on our two random variables X and Y. We summarize the resulting data in a contingency table

$$U \quad = \quad \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{mn} \end{pmatrix}.$$

The entries of the matrix U are nonnegative integers whose sum is u_{++}.

The *likelihood function* for the contingency table U is the following function on Δ_{mn-1}:

$$P \quad \mapsto \quad \binom{u_{++}}{u_{11}u_{12}\cdots u_{mn}} \prod_{i=1}^{m} \prod_{j=1}^{n} p_{ij}^{u_{ij}}.$$

Assuming fixed sample size, this is the likelihood of observing the data U given an unknown probability distribution P in Δ_{mn-1}. In what follows we suppress the multinomial coefficient. Furthermore, we regard the likelihood function as a rational function on $\mathbb{P}^{mn-1}$, so we write

$$\ell_U = \frac{\prod_{i=1}^{m} \prod_{j=1}^{n} p_{ij}^{u_{ij}}}{p_{++}^{u_{++}}}.$$

We wish to find a low rank probability matrix P that best explains the data U. Maximum likelihood estimation means solving the following optimization problem:

$$\text{Maximize } \ell_U(P) \text{ subject to } P \in \mathcal{M}_r. \tag{19}$$

The optimal solution $\hat{P}$ is a rank r matrix. This is the *maximum likelihood estimate* for U.

For $r = 1$, the independence model, the maximum likelihood estimate $\hat{P}$ is obtained from the data matrix U by the following formula, already seen for $m = n = 2$ in (8). Multiply the vector of row sums with the vector of column sums and divide by the sample size:

$$\hat{P} \;=\; \frac{1}{(u_{++})^2} \cdot \begin{pmatrix} u_{1+} \\ u_{2+} \\ \vdots \\ u_{m+} \end{pmatrix} \cdot \begin{pmatrix} u_{+1} \; u_{+2} \cdots u_{+n} \end{pmatrix}. \tag{20}$$

Statisticians, scientists and engineers refer to such a formula as an "analytic solution". In our view, it would be more appropriate to call this an "algebraic solution". After all, we are here using algebra not analysis. Our algebraic solution for $r = 1$ reveals the following points:

- The MLE $\hat{P}$ is a *rational function* of the data U.
- The function $U \mapsto \hat{P}$ is an algebraic function of degree 1.
- The ML degree of the independence model $\mathcal{V}_1$ equals 1.

We next discuss the smallest case when the ML degree is larger than 1.

Example 2.1. Let $m = n = 3$ and $r = 2$. Our MLE problem is to maximize

$$\ell_U \;=\; (p_{11}^{u_{11}} p_{12}^{u_{12}} p_{13}^{u_{13}} p_{21}^{u_{21}} p_{22}^{u_{22}} p_{23}^{u_{23}} p_{31}^{u_{31}} p_{32}^{u_{32}} p_{33}^{u_{33}}) / p_{++}^{u_{++}}$$

subject to the constraints $P \geq 0$ and $\mathrm{rank}(P) = 2$, where $P = (p_{ij})$ is a $3{\times}3$-matrix of unknowns. The equations that characterize the critical points of this optimization problem are

$$\det(P) = \begin{matrix} p_{11}p_{22}p_{33} - p_{11}p_{23}p_{32} - p_{12}p_{21}p_{33} \\ +p_{12}p_{23}p_{31} + p_{13}p_{21}p_{32} - p_{13}p_{22}p_{31} \end{matrix} \;=\; 0$$

and the vanishing of the 3×3-minors of the following 3×9-matrix:

$$\begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{21} & u_{22} & u_{23} & u_{31} & u_{32} & u_{33} \\ p_{11} & p_{12} & p_{13} & p_{21} & p_{22} & p_{23} & p_{31} & p_{32} & p_{33} \\ p_{11}a_{11} & p_{12}a_{12} & p_{13}a_{13} & p_{21}a_{21} & p_{22}a_{22} & p_{33}a_{33} & p_{31}a_{31} & p_{32}a_{32} & p_{33}a_{33} \end{bmatrix}$$

where $a_{ij} = \frac{\partial \det(P)}{\partial p_{ij}}$ is the cofactor of p_{ij} in P. For random positive data u_{ij}, these equations have ten solutions with $\mathrm{rank}(P) = 2$ in $\mathbb{P}^8 \backslash \mathcal{H}$. Hence the ML degree of $\mathcal{V}_2$ is 10. If we regard the u_{ij} as unknowns, then saturating the above determinantal equations with respect to $\mathcal{H} \cup \mathcal{V}_1$ yields the prime ideal of the likelihood correspondence $\mathcal{L}_{\mathcal{V}_2} \subset \mathbb{P}^8 \times \mathbb{P}^8$. See Example 4.8 for the bidegree and other enumerative invariants of the eight-dimensional variety $\mathcal{L}_{\mathcal{V}_2}$. $\diamond$

Recall from Definition 1.5 that the *ML degree* of a statistical model (or a projective variety) is the number of critical points of the likelihood function for generic data.

Theorem 2.2. *The known values for the ML degrees of the determinantal varieties $\mathcal{V}_r$ are*

	$(m,n) =$ (3,3)	(3,4)	(3,5)	(4,4)	(4,5)	(4,6)	(5,5)
$r = 1$	1	1	1	1	1	1	1
$r = 2$	10	26	**58**	**191**	**843**	**3119**	**6776**
$r = 3$	1	1	1	**191**	**843**	**3119**	**61326**
$r = 4$				1	1	1	**6776**
$r = 5$							1

The numbers 10 and 26 were computed back in 2004 using the symbolic software `Singular`, and they were reported in [HKS, §5]. The bold face numbers were found in 2012 in [HRS] using the numerical software `Bertini`. In what follows we shall describe some of the details.

Remark 2.3. Each determinantal variety $\mathcal{V}_r$ is singular along the smaller variety $\mathcal{V}_{r-1}$. Hence, the very affine variety $\mathcal{V}_r \backslash \mathcal{H}$ is singular for $r \geq 2$, so Theorem 1.7 does not apply. Here, $\mathcal{H} = \{p_{++} \prod p_{ij} = 0\}$. According to Conjecture 1.8, the ML degree above provides a lower bound for the signed topological Euler characteristic of $\mathcal{V}_r \backslash \mathcal{H}$. The difference between the two numbers reflect the nature of the singular locus $\mathcal{V}_{r-1} \backslash \mathcal{H}$ inside $\mathcal{V}_r \backslash \mathcal{H}$. For plane curves that have nodes and cusps, we encountered this issue in Examples 1.4 and 1.17.

We begin with a geometric description of the likelihood correspondence. An $m \times n$-matrix P is a regular point in $\mathcal{V}_r$ if and only if $\mathrm{rank}(P) = r$. The tangent space T_P is a subspace of dimension $rn+rm-r^2$ in $\mathbb{C}^{m\times n}$. Its orthogonal complement $T_P^{\perp}$ has dimension $(m-r)(n-r)$.

The partial derivatives of the log-likelihood function $\log(\ell_U)$ on $\mathbb{P}^{mn-1}$ are

$$\frac{\partial \log(\ell_U)}{\partial p_{ij}} = \frac{u_{ij}}{p_{ij}} - \frac{u_{++}}{p_{++}}.$$

Proposition 2.4. *An $m \times n$-matrix P of rank r is a critical point for $\log(\ell_U)$ on $\mathcal{V}_r$ if and only if the linear subspace $T_P^{\perp}$ contains the matrix*

$$\left[\frac{u_{ij}}{p_{ij}} - \frac{u_{++}}{p_{++}} \right]_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$$

In order to get to the numbers in Theorem 2.2, the geometric formulation was replaced in [HRS] with a parametric representation of the rank constraints. The following linear algebra formulation worked well for non-trivial computations. Assume $m \leq n$. Let P_1, R_1, L_1 and Λ be matrices of unknowns of formats $r \times r$, $r \times (n-r)$, $(m-r) \times r$, and $(n-r) \times (m-r)$. Set

$$L = \begin{pmatrix} L_1 & -I_{m-r} \end{pmatrix}, \quad P = \begin{pmatrix} P_1 & P_1 R_1 \\ L_1 P_1 & L_1 P_1 R_1 \end{pmatrix}, \quad \text{and} \quad R = \begin{pmatrix} R_1 \\ -I_{n-r} \end{pmatrix},$$

where I_{m-r} and I_{n-r} are identity matrices. In the next statement we use the symbol $\star$ for the Hadamard (entrywise) product of two matrices that have the same format.

Proposition 2.5. *Fix a general $m \times n$ data matrix U. The polynomial system*

$$P \star (R \cdot \Lambda \cdot L)^T \; + \; u_{++} \cdot P = U$$

consists of mn equations in mn unknowns. For generic U, it has finitely many complex solutions (P_1, L_1, R_1, Λ). The $m \times n$-matrices P resulting from these solutions are precisely the critical points of the likelihood function ℓ_U on the determinantal variety $\mathcal{V}_r$.

We next present the analogue to Theorem 2.2 for symmetric matrices

$$P \quad = \quad \begin{pmatrix} 2p_{11} & p_{12} & p_{13} & \cdots & p_{1n} \\ p_{12} & 2p_{22} & p_{23} & \cdots & p_{2n} \\ p_{13} & p_{23} & 2p_{33} & \cdots & p_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{1n} & p_{2n} & p_{3n} & \cdots & 2p_{nn} \end{pmatrix}.$$

Such matrices, with nonnegative coordinates p_{ij} that sum to 1, represent joint probability distributions for two identically distributed random variables with n states. The case $n = 2$ and $r = 1$ is the Hardy–Weinberg curve, which we discussed in detail in Example 1.3.

Theorem 2.6. *The known values for ML degrees of symmetric matrices of rank at most r (mixtures of r independent identically distributed random variables) are*

$n =$	2	3	4	5	6
$r = 1$	1	1	1	1	1
$r = 2$	1	6	**37**	**270**	**2341**
$r = 3$		1	**37**	**1394**	?
$r = 4$			1	**270**	?
$r = 5$				1	**2341**

At present we do not know the common value of the ML degree for $n = 6$ and $r = 3, 4$. In what follows we take a closer look at the model for symmetric 3×3-matrices of rank 2.

Example 2.7. Let $n = 3$ and $r = 2$, so X is a cubic hypersurface in $\mathbb{P}^5$. The likelihood correspondence $\mathcal{L}_X$ is a five-dimensional subvariety of $\mathbb{P}^5 \times \mathbb{P}^5$ having bidegree

$$B_X(p, u) = 6p^5 + 12p^4u + 15p^3u^2 + 12p^2u^3 + 3pu^4.$$

The bihomogeneous prime ideal of $\mathcal{L}_X$ is minimally generated by 23 polynomials, namely:

- One polynomial of bidegree $(3, 0)$; this is the determinant of P.
- Three polynomials of degree $(1, 1)$. These come from the underlying toric model $\{\text{rank}(P) = 1\}$. As suggested in Proposition 3.5, they are the 2×2-minors of

$$\begin{pmatrix} 2p_0 + p_1 + p_2 & p_1 + 2p_3 + p_4 & p_2 + p_4 + 2p_5 \\ 2u_0 + u_1 + u_2 & u_1 + 2u_3 + u_4 & u_2 + u_4 + 2u_5 \end{pmatrix}.$$

- One polynomial of degree $(2, 1)$,
- three polynomial of degree $(2, 2)$,
- nine polynomials of degree $(3, 1)$,
- six polynomials of degree $(3, 2)$.

It turns out that this ideal represents an expression for the MLE $\hat{P}$ in terms of radicals in U.

We shall work this out for one numerical example. Consider the data matrix U with

$$u_{11} = 10,\ u_{12} = 9,\ u_{13} = 1,\ u_{22} = 21,\ u_{23} = 3,\ u_{33} = 7.$$

For this choice, all six critical points of the likelihood function are real and positive:

p_{11}	p_{12}	p_{13}	p_{22}	p_{23}	p_{33}	$\log \ell_U(p)$
0.1037	0.3623	0.0186	0.3179	0.0607	0.1368	−82.18102
0.1084	0.2092	0.1623	0.3997	0.0503	0.0702	−84.94446
0.0945	0.2554	0.1438	0.3781	0.4712	0.0810	−84.99184
0.1794	0.2152	0.0142	0.3052	0.2333	0.0528	−85.14678
0.1565	0.2627	0.0125	0.2887	0.2186	0.0609	−85.19415
0.1636	0.1517	0.1093	0.3629	0.1811	0.0312	−87.95759

The first three points are local maxima in Δ_5 and the last three points are local minima. These six points define an algebraic field extension of degree 6 over $\mathbb{Q}$. One might expect that the Galois group of these six points over $\mathbb{Q}$ is the full symmetric group S_6. If this were the case then the above coordinates could not be written in radicals. However, that expectation is wrong. The Galois group of the likelihood fibration $\text{pr}_2 : \mathcal{L}_X \rightarrow \mathbb{P}^5_U$ given by the 3×3 symmetric problem is a subgroup of S_6 isomorphic to the solvable group S_4.

To be concrete, for the data above, the minimal polynomial for the MLE $\hat{p}_{33}$ equals

$$\begin{aligned} &9528773052286944 p_{33}^6 - 4125267629399052 p_{33}^5 + 713452955656677 p_{33}^4 \\ &\quad - 63349419858182 p_{33}^3 + 3049564842009 p_{33}^2 - 75369770028 p_{33} \\ &\qquad + 744139872 \ = \ 0. \end{aligned}$$

We solve this equation in radicals as follows:

$$
\begin{aligned}
p_{33} = \; & \tfrac{16427}{227664} + \tfrac{1}{12}\left(\zeta - \zeta^2\right)\omega_2 - \tfrac{66004846384302}{19221271018849}\omega_2^2 + \\
& \left(\tfrac{14779904193}{211433981207339}\zeta^2 - \tfrac{14779904193}{211433981207339}\zeta\right)\omega_1\omega_2^2 + \tfrac{1}{2}\omega_3,
\end{aligned}
$$

where ζ is a primitive third root of unity, $\omega_1^2 = 94834811/3$, and

$$
\begin{aligned}
\omega_2^3 = & \left(\tfrac{5992589425361}{150972770845322208}\zeta - \tfrac{5992589425361}{150972770845322208}\zeta^2\right) + \tfrac{97163}{40083040181952}\omega_1, \\
\omega_3^2 = & \tfrac{5006721709}{1248260766912} + \left(\tfrac{212309132509}{4242035935404}\zeta - \tfrac{212309132509}{4242035935404}\zeta^2\right)\omega_2 - \tfrac{2409}{20272573168}\omega_1\omega_2 \\
& - \tfrac{158808750548335}{76885084075396}\omega_2^2 + \left(\tfrac{17063004159}{422867962414678}\zeta^2 - \tfrac{17063004159}{422867962414678}\zeta\right)\omega_1\omega_2^2.
\end{aligned}
$$

The explanation for the extra symmetry stems from the duality theorem below. It furnishes an involution on the set of six critical points that allows us to express them in radicals. ◇

The tables in Theorems 2.2 and 2.6 suggest that the columns will always be symmetric. This fact was conjectured in [HRS] and subsequently proved by Draisma and Rodriguez in [DR].

Theorem 2.8. *Fix $m \leq n$ and consider the determinantal varieties $\mathcal{V}_i$ for either general or symmetric matrices. Then the ML degrees for rank r and for rank $m - r + 1$ coincide.*

In fact, the main result in [DR] establishes the following more precise statement. Given a data matrix U of format $m \times n$, we write Ω_U for the $m \times n$-matrix whose (i, j) entry equals

$$
\frac{u_{ij} \cdot u_{i+} \cdot u_{+j}}{(u_{++})^3}.
$$

Theorem 2.9. *Fix $m \leq n$ and U an $m \times n$-matrix with strictly positive integer entries. There exists a bijection between the complex critical points $P_1, P_2, \ldots, P_s$ of the likelihood function ℓ_U on $\mathcal{V}_r$ and the complex critical points $Q_1, Q_2, \ldots, Q_s$ of ℓ_U on $\mathcal{V}_{m-r+1}$ such that*

$$
P_1 \star Q_1 \; = \; P_2 \star Q_2 \; = \; \cdots \; = \; P_s \star Q_s = \Omega_U .
$$

Thus, this bijection preserves reality, positivity, and rationality.

The key to computing the ML degree tables and to formulating the duality conjectures in [HRS], was the use of numerical algebraic geometry. The software `Bertini` allowed for the computation of thousands of instances in which the formula of Theorem 2.9 was confirmed.

`Bertini` is numerical software, based on homotopy continuation, for finding all complex solutions to a system of polynomial equations (and much more). The software is available at [Bertini]. The developers, Daniel Bates, Jonathan

Hauenstein, Andrew Sommese, Charles Wampler, have just completed a new textbook [BHSW] on the mathematics behind `Bertini`.

For the past two decades, algebraic geometers have increasingly employed computational methods as a tool for their research. However, these computations have almost always been symbolic (and hence exact). They relied on Gröbner-based software such as `Singular` or `Macaulay2`. Algebraists often feel a certain discomfort when asked to trust a numerical computation. We encourage discussion about this issue, by raising the following question.

Example 2.10. In the rightmost column of Theorem 2.6, it is asserted that the solution to a certain enumerative geometry problem is **2341**. Which of these would **you** trust most:

- the output of a symbolic computation?
- the output of a numerical computation?
- a proof written by an algebraic geometer?

In the authors' view, it always pays off to be critical and double-check all computations, regardless of how they were carried out. And, this applies to all three of the above. ◇

One of the big advantages of numerical algebraic geometry over Gröbner bases when it comes to MLE is the separation between *Preprocessing* and *Solving*. For any particular variety $X \subset \mathbb{P}^n$, such as $X = \mathcal{V}_r$, we preprocess by solving the likelihood equations once, for a generic data set U_0 chosen by us. The coordinates of U_0 may be complex (rather than real) numbers. We can chose them with stable numerics in mind, so as to compute all critical points up to high accuracy. This step can take a long time, but the output is highly reliable.

After solving the equations once, for that generic U_0, all subsequent computations for any other data set U are very fast. In particular, the computation is fully parallelizable. If we have m processors at our disposal, where $m = \text{MLdegree}\,(X)$, then each processor can track one of the paths. To be precise, homotopy continuation starts from the critical points of ℓ_{U_0} and transform them into the critical points of ℓ_U. Geometrically speaking, for fixed X, the homotopy amounts to walking on the sheets of the likelihood fibration $\text{pr}_2 : \mathcal{L}_X \to \mathbb{P}^n_u$.

To illustrate this point, here are the timings (in seconds) that were reported in [HRS] for the determinantal varieties $X = \mathcal{V}_r$. Those computations were carried out in `Bertini` on a 64-bit Linux cluster with 160 processors. The first row is the preprocessing time for solving the equations once. The second row is the time needed to solve any subsequent instance:

(m, n, r)	$(4,4,2)$	$(4,4,3)$	$(4,5,2)$	$(4,5,3)$	$(5,5,2)$	$(5,5,4)$
Preprocessing	257	427	1938	2902	348555	146952
Solving	4	4	20	20	83	83

This table suggests that combining numerical algebraic geometry with existing tools from computational statistics might lead to a viable tool for certifiably solving MLE problems.

We are now at the point where it is essential to offer a disclaimer. The low rank model $\mathcal{M}_r$ does not correctly represent the notion of conditional independence. The model we should have used instead is the *mixture model* Mix_r. By definition, Mix_r is the set of probability distributions P in Δ_{mn-1} that are convex combinations of r independent distributions, each taken from $\mathcal{M}_1$. Equivalently, the mixture model Mix_r consists of all matrices

$$P = A \cdot \Lambda \cdot B, \tag{21}$$

where A is a nonnegative $m \times r$-matrix whose rows sum to 1, Λ is a nonnegative $r \times r$ diagonal matrix whose entries sum to 1, and B is a nonnegative $r \times n$-matrix whose columns sum to 1. The formula (21) expresses Mix_r as the image of a trilinear map between polytopes:

$$\phi \,:\, (\Delta_{m-1})^r \times \Delta_{r-1} \times (\Delta_{n-1})^r \;\to\; \Delta_{mn-1}\,, \quad (A, \Lambda, B) \;\mapsto\; P.$$

The following result is well-known; see e.g. [LiAS, Example 4.1.2].

Proposition 2.11. *Our low rank model $\mathcal{M}_r$ is the Zariski closure of the mixture model Mix_r in the probability simplex Δ_{mn-1}. If $r \leq 2$ then $\mathrm{Mix}_r = \mathcal{M}_r$. If $r \geq 3$ then $\mathrm{Mix}_r \subsetneq \mathcal{M}_r$.*

The point here is the distinction between the rank and the nonnegative rank of a nonnegative matrix. Matrices in $\mathcal{M}_r$ have rank $\leq r$ and matrices in Mix_r have nonnegative rank $\leq r$. Thus elements of $\mathcal{M}_r \backslash \mathrm{Mix}_r$ are matrices whose nonnegative rank exceeds its rank.

Example 2.12. The following 4×4-matrix has rank 3 but nonnegative rank 4:

$$P \quad = \quad \frac{1}{8} \cdot \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

This is the slack matrix of a regular square. It is an element of $\mathcal{M}_3 \backslash \mathrm{Mix}_3$. $\diamond$

Engineers and scientists care more about Mix_r than $\mathcal{M}_r$. In many applications, nonnegative rank is more relevant than rank. The reason can be seen in (18). In such a low-rank decomposition, we do not want the female table or the male table to have a negative entry.

This raises the following important questions: How to maximize the likelihood function ℓ_U over Mix_r? What are the algebraic degrees associated with that optimization problem?

Statisticians seek to maximize the likelihood function ℓ_U on Mix_r by using the *expectation-maximization* (EM) algorithm in the space $(\Delta_{m-1})^r \times \Delta_{r-1} \times (\Delta_{n-1})^r$ of parameters (A, Λ, B). In each iteration, the EM algorithm strictly decreases the *Kullback–Leibler divergence* from the current model point $P = \phi(A, \Lambda, B)$ to the empirical distribution $\frac{1}{u_{++}} \cdot U$. The hope in running the EM algorithm for given data U is that it converges to the global maximum $\hat{P}$ on Mix_r. For a presentation of the EM algorithm for discrete algebraic models see [PS, §1.3]. A study of the geometry of this algorithm for the mixture model Mix_r is undertaken in [KRS].

If the EM algorithm converges to a point that lies in the interior of the parameter polytope, and is non-singular with respect to ϕ, then that point will be among the critical points on $\mathcal{M}_r$. These are characterized by Proposition 2.4. However, since Mix_r is properly contained in $\mathcal{M}_r$, it frequently happens that the true MLE $\hat{P}$ lies on the boundary of Mix_r. In that case, $\hat{P}$ is not a critical point of ℓ_U on $\mathcal{M}_r$, meaning that $(\hat{P}, U)$ is not in the likelihood correspondence on $\mathcal{V}_r$. Such points will never be found by the method described above.

In order to address this issue, we need to identify the divisors in the variety $\mathcal{V}_r \subset \mathbb{P}^{mn-1}$ that appear in the algebraic boundary of Mix_r. By this we mean the irreducible components $W_1, W_2, \ldots, W_s$ of the Zariski closure of $\partial\mathrm{Mix}_r$. Each of these W_i has codimension 1 in $\mathcal{V}_r$. Once the W_i are identified, one would need to examine their ML degree, and also the ML degree of the various strata $W_{i_1} \cap \cdots \cap W_{i_s}$ in which ℓ_U might attain its maximum. At present we do not have this information even in the smallest non-trivial case $m = n = 4$ and $r = 3$.

Example 2.13. We illustrate this issue by describing one of the components W of the algebraic boundary for the mixture model Mix_3 when $m = n = 4$. Consider the equation

$$\begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{pmatrix} = \begin{pmatrix} 0 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ a_{31} & 0 & a_{33} \\ a_{41} & a_{42} & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & b_{12} & b_{13} & b_{14} \\ b_{21} & 0 & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & 0 \end{pmatrix}$$

This parametrizes a 13-dimensional subvariety W of the hypersurface $\mathcal{V}_3 = \{\det(P) = 0\}$ in $\mathbb{P}^{15}$. The variety W is a component in the algebraic boundary of Mix_3. To see this, we choose the a_{ij} and b_{ij} to be positive, and we note that P lies outside Mix_3 when precisely one of the 0 entries gets replaced by $-\epsilon$. The prime ideal of W in $\mathbb{Q}[p_{11}, \ldots, p_{44}]$ is obtained by eliminating the 17 unknowns a_{ij} and b_{ij} from the 16 scalar equations. A direct computation with `Macaulay 2` shows that the variety W is Cohen–Macaulay of codimension-2. By the Hilbert–Burch Theorem, it is defined by the 4×4-minors of the 4×5-matrix. This following specific matrix representation was suggested to us by Aldo Conca and Matteo Varbaro:

$$\begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} & 0 \\ p_{21} & p_{22} & p_{23} & p_{24} & 0 \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{34}(p_{11}p_{22} - p_{12}p_{21}) \\ p_{41} & p_{42} & p_{43} & p_{44} & p_{41}(p_{12}p_{24} - p_{14}p_{22}) + p_{44}(p_{11}p_{22} - p_{12}p_{21}) \end{pmatrix}.$$

Tte algebraic boundary of Mix_3 consists of precisely 304 irreducible components, namely the 16 coordinate hyperplanes and 288 hypersurfaces that are all isomorphic to W. This is proved in [KRS]. In that paper, it is also shown that the ML degree of W equals 633. $\diamond$

The definition of rank varieties and mixture models extends to m-dimensional tensors P of arbitrary format $d_1 \times d_2 \times \cdots \times d_m$. We refer to Landsberg's book [Land] for an introduction to tensors and their rank. Now, $\mathcal{V}_r$ is the variety of tensors of borderrank $\leq r$, the model $\mathcal{M}_r$ is the set of all probability distributions in $\mathcal{V}_r$, and the model Mix_r is the subset of tensors of nonnegative rank $\leq r$. Unlike in the matrix case $m = 2$, the mixture model for borderrank $r = 2$ is already quite interesting when $m \geq 3$. We state two theorems that characterize our objects. The set-theoretic version of Theorem 2.14 is due to Landsberg and Manivel [LM]. The ideal-theoretic statement was proved more recently by Raicu [Rai].

Theorem 2.14. *The variety $\mathcal{V}_2$ is defined by the 3×3-minors of all flattenings of P.*

Here, *flattening* means picking any subset A of $[n] = \{1, 2, \ldots, n\}$ with $1 \leq |A| \leq n - 1$ and writing the tensor P as an ordinary matrix with $\prod_{i\in A} d_i$ rows and $\prod_{j\notin A} d_j$ columns.

Theorem 2.15. *The mixture model* Mix_2 *is the subset of supermodular distributions in $\mathcal{M}_2$.*

This theorem was proved in [ARSZ]. Being *supermodular* means that P satisfies a natural family of quadratic binomial inequalities. We explain these for $m = 3, d_1 = d_2 = d_3 = 2$.

Example 2.16. We consider $2\times 2\times 2$ tensors. Since secant lines of the Segre variety $\mathbb{P}^1\times\mathbb{P}^1\times\mathbb{P}^1$ fill all of $\mathbb{P}^7$, we have that $\mathcal{V}_2 = \mathbb{P}^7$ and $\mathcal{M}_2 = \Delta_7$. The mixture model Mix_2 is an interesting, full-dimensional, closed, semi-algebraic subset of Δ_7. By definition, Mix_2 is the image of a 2-to-1 map $\phi : (\Delta_1)^7 \to \Delta_7$ analogous to (21). The branch locus is the $2\times2\times2$-hyperdeterminant, which is a hypersurface in $\mathbb{P}^7$ of degree 4 and ML degree 13.

The analysis in [ARSZ, §2] represents the model Mix_2 as the union of four *toric cells*. One of these toric cells is the set of tensors satisfying

$$\begin{array}{lll} p_{111}p_{222} \geq p_{112}p_{221} & p_{111}p_{222} \geq p_{121}p_{212} & p_{111}p_{222} \geq p_{211}p_{122} \\ p_{112}p_{222} \geq p_{122}p_{212} & p_{121}p_{222} \geq p_{122}p_{221} & p_{211}p_{222} \geq p_{212}p_{221} \\ p_{111}p_{122} \geq p_{112}p_{121} & p_{111}p_{212} \geq p_{112}p_{211} & p_{111}p_{221} \geq p_{121}p_{211} \end{array} \tag{22}$$

A nonnegative $2\times2\times2$-tensor P in Δ_7 is *supermodular* if it satisfies these inequalities, possibly after label swapping $1 \leftrightarrow 2$. We visualize Mix_2 by restricting to the three-dimensional subspace H given by $p_{111} = p_{222}$, $p_{112} = p_{221}$, $p_{121} = p_{212}$ and $p_{211} = p_{122}$. The intersection $H \cap \Delta_7$ is a tetrahedron, and we consider $H \cap \mathrm{Mix}_2$ inside that tetrahedron. The restricted model $H \cap \mathrm{Mix}_2$ is shown on the left in Fig. 1. It consists of four toric cells as shown on the right side. The boundary

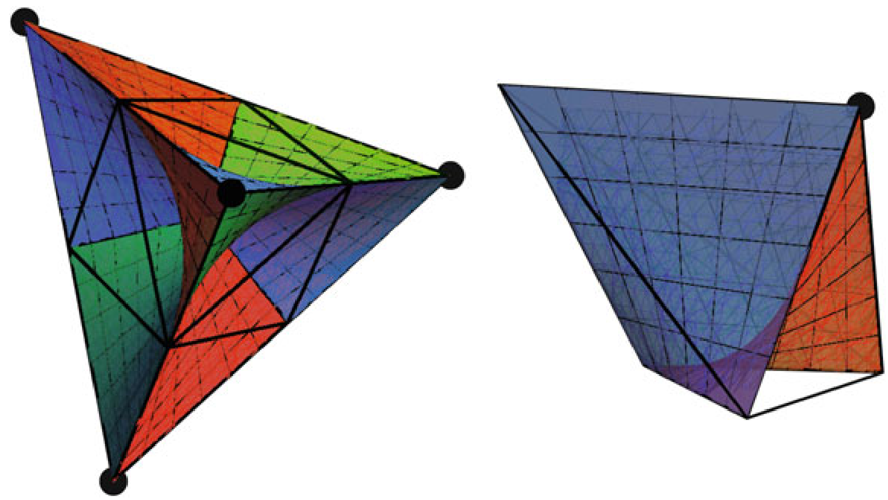

Fig. 1 A three-dimensional slice of the seven-dimensional model of 2×2×2 tensors of nonnegative rank ≤ 2. Each toric cell is bounded by 3 quadrics and contains a vertex of the tetrahedron

is given by three quadratic surfaces, shown in red, green and blue, and which are obtained from either the first or the second row in (22) by restriction to H.

The boundary analysis suggested in Example 2.13 turns out to be quite simple in the present example. All boundary strata of the model Mix_2 are varieties of ML degree 1.

One such boundary stratum for Mix_2 is the five-dimensional toric variety

$$X \;=\; V(p_{112}p_{222} - p_{122}p_{212},\, p_{111}p_{122} - p_{112}p_{121},\, p_{111}p_{222} - p_{121}p_{212}) \;\subset\; \mathbb{P}^7.$$

As a preview for what is to come, we report its ML bidegree and its sectional ML degree:

$$\begin{aligned} B_X(p,u) &= \; p^7 \,+\, 2p^6u \,+\, 3p^5u^2 \,+\, 3p^4u^3 \,+\, 3p^3u^4 \,+\, 3p^2u^5, \\ S_X(p,u) &= \; p^7 + 14p^6u + 30p^5u^2 + 30p^4u^3 + 15p^3u^4 + 3p^2u^5. \end{aligned} \tag{23}$$

In the next section, we shall study the class of toric varieties and the class of varieties having ML degree 1. Our variety X lies in the intersection of these two important classes. ◇

3 Third Lecture

In our third lecture we start out with the likelihood geometry of embedded toric varieties. Fix a $(d+1)\times(n+1)$ integer matrix $A = (a_0, a_1, \ldots, a_n)$ of rank $d+1$ that has $(1,1,\ldots,1)$ as its last row. This matrix defines an effective action of the torus $(\mathbb{C}^*)^d$ on projective space $\mathbb{P}^n$:

$$(\mathbb{C}^*)^d \times \mathbb{P}^n \longrightarrow \mathbb{P}^n, \qquad t\times(p_0 : p_1 : \cdots : p_n) \longmapsto (t^{\tilde{a}_0}\cdot p_0 : t^{\tilde{a}_1}\cdot p_1 : \cdots : t^{\tilde{a}_n}\cdot p_n).$$

Here $\tilde{a}_i$ is the column vector a_i with the last entry 1 removed. We also fix

$$c = (c_0, c_1, \ldots, c_n) \in (\mathbb{C}^*)^{n+1},$$

viewed as a point in $\mathbb{P}^n$. Let X_c be the closure in $\mathbb{P}^n$ of the orbit $(\mathbb{C}^*)^d \cdot c$. This is a projective toric variety of dimension d, defined by the pair (A, c). The ideal that defines X_c is the familiar *toric ideal* I_A as in [LiAS, §1.3], but with $p = (p_0, \ldots, p_n)$ replaced by

$$p/c = \left(\frac{p_0}{c_0}, \frac{p_1}{c_1}, \ldots, \frac{p_n}{c_n}\right). \tag{24}$$

Example 3.1. Fix $d = 2$ and $n = 3$. The matrix

$$A = \begin{pmatrix} 0 & 3 & 0 & 1 \\ 0 & 0 & 3 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

specifies the following family of toric surfaces of degree three in $\mathbb{P}^3$:

$$X_c = \overline{\{(c_0 : c_1 x_1^3 : c_2 x_2^3 : c_3 x_1 x_2) \; : \; (x_1, x_2) \in (\mathbb{C}^*)^2\}} = V(c_3^3 \cdot p_0 p_1 p_2 - c_0 c_1 c_2 \cdot p_3^3).$$

Of course, the prime ideal of any particular surface X_c is the principal ideal generated by

$$\frac{p_0}{c_0} \frac{p_1}{c_1} \frac{p_2}{c_2} - \left(\frac{p_3}{c_3}\right)^3.$$

How does the ML degree of X_c depend on the parameter $c = (c_0, c_1, c_2, c_3) \in (\mathbb{C}^*)^4$? ◇

We shall express the ML degree of the toric variety X_c in terms of the complement of a hypersurface in the torus $(\mathbb{C}^*)^d$. The pair (A, c) define the sparse Laurent polynomial

$$f(x) = c_0 \cdot x^{\tilde{a}_0} + c_1 \cdot x^{\tilde{a}_1} + \cdots + c_n \cdot x^{\tilde{a}_n}.$$

Theorem 3.2. *The ML degree of the d-dimensional toric variety $X_c \subset \mathbb{P}^n$ is equal to $(-1)^d$ times the Euler characteristic of the very affine variety*

$$X_c \backslash \mathcal{H} \;\; \simeq \;\; \left\{x \in (\mathbb{C}^*)^d \; : \; f(x) \neq 0\right\}. \tag{25}$$

For generic c, the ML degree agrees with the degree of X_c, which is the normalized volume of the d-dimensional lattice polytope conv(A) *obtained as the convex hull of the columns of A.*

Proof. We first argue that the identification (25) holds. The map

$$x \longmapsto p = (c_0 \cdot x^{\tilde{a}_0} : c_1 \cdot x^{\tilde{a}_1} : \cdots : c_n \cdot x^{\tilde{a}_n})$$

defines an injective group homomorphism from $(\mathbb{C}^*)^d$ into the dense torus of $\mathbb{P}^n$. Its image is equal to the dense torus of X_c, so we have an isomorphism between $(\mathbb{C}^*)^d$ and the dense torus of X_c. Under this isomorphism, the affine open set $\{f \neq 0\}$ in $(\mathbb{C}^*)^d$ is identified with the affine open set $\{p_0 + \cdots + p_n \neq 0\}$ in the dense torus of X_c. The latter is precisely $X_c \backslash \mathcal{H}$. Since $(\mathbb{C}^*)^d$ is smooth, we see that $X_c \backslash \mathcal{H}$ is smooth, so our first assertion follows from Theorem 1.7. The second assertion is a consequence of the description of the likelihood correspondence $\mathcal{L}_{X_c}$ via linear sections of X_c that is given in Proposition 3.5 below. □

Example 3.3. We return to the cubic surface X_c in Example 3.1. For a general parameter vector c, the ML degree of X_c is 3. For instance, the surface $V(p_0p_1p_2 - p_3^3) \subset \mathbb{P}^3$ has ML degree 3. However, the ML degree of X_c drops to 2 whenever the plane curve defined by

$$f(x_1, x_2) = c_0 + c_1x_1^3 + c_2x_2^3 + c_3x_1x_2$$

has a singularity in $(\mathbb{C}^*)^2$. For instance, this happens for $c = (1 : 1 : 1 : -3)$. The corresponding surface $V(27p_0p_1p_2 + p_3^3) \subset \mathbb{P}^3$ has ML degree 2. ◇

The isomorphism (25) has a nice interpretation in terms of Convex Optimization. Namely, it implies that maximum likelihood estimation for toric varieties is equivalent to global minimization of *posynomials*, and hence to the most fundamental case of *Geometric Programming*. We refer to [BoydVan, §4.5] for an introduction to posynomials and geometric programming.

We write $| \cdot |$ for the one-norm on $\mathbb{R}^{n+1}$, we set $b = Au$, and we assume that $c = (c_0, c_1, \ldots, c_n)$ is in $\mathbb{R}^{n+1}_{>0}$. Maximum likelihood estimation for toric models is the problem

$$\text{Maximize } \frac{p^u}{|p|^{|u|}} \text{ subject to } p \in X_c \cap \Delta_n. \tag{26}$$

Setting $p_i = c_i \cdot x^{\tilde{a}_i}$ as above, this problem becomes equivalent to the geometric program

$$\text{Minimize } \frac{f(x)^{|u|}}{x^b} \text{ subject to } x \in \mathbb{R}^d_{>0}. \tag{27}$$

By construction, $f(x)^{|u|}/x^b$ is a posynomial whose Newton polytope contains the origin. Such a posynomial attains a unique global minimum on the open orthant $\mathbb{R}^d_{>0}$. This can be seen by convexifying as in [BoydVan, §4.5.3]. This global minimum of (27) corresponds to the solution of (26), which exists and is unique by Birch's Theorem [PS, Theorem 1.10].

Example 3.4. Consider the geometric program for the surfaces in Example 3.1, with

$$A \;=\; \begin{pmatrix} 0 & 3 & 0 & 1 \\ 0 & 0 & 3 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \quad \text{and} \quad u \;=\; (0,0,0,1).$$

The problem (27) is to find the global minimum, over all positive $x = (x_1, x_2)$, of the function

$$\frac{f(x_1,x_2)}{x_1x_2} = c_0 x_1^{-1}x_2^{-1} + c_1 x_1^2 x_2^{-1} + c_2 x_1^{-1} x_2^2 + c_3.$$

This is equivalent to maximizing p_3/p_+ subject to $p \in V(c_3^3 \cdot p_0p_1p_2 - c_0c_1c_2 \cdot p_3^3) \cap \Delta_3$. ◇

We now describe the toric likelihood correspondence $\mathcal{L}_{X_c}$ in $\mathbb{P}^n \times \mathbb{P}^n$ associated with the pair (A, c). This is the likelihood correspondence of the toric variety $X_c \subset \mathbb{P}^n$ defined above.

Proposition 3.5. *On the open subset* $(X_c \backslash \mathcal{H}) \times \mathbb{P}^n$, *the toric likelihood correspondence* $\mathcal{L}_{X_c}$ *is defined by the* 2×2*-minors of the* $2 \times (d+1)$*-matrix*

$$\begin{pmatrix} p/c \cdot A^T \\ u/c \cdot A^T \end{pmatrix}. \tag{28}$$

Here the notation p/c *is as in (24). In particular, for any fixed data vector* u, *the critical points of* ℓ_u *are characterized by a linear system of equations in* p *restricted to* X_c.

Proof. This is an immediate consequence of Birch's Theorem [PS, Theorem 1.10]. □

Example 3.6. The *Hardy–Weinberg curve* of Example 1.3 is the subvariety $X_c = V(p_1^2 - 4p_0p_2)$ in the projective plane $\mathbb{P}^2$. As a toric variety, this plane curve is given by

$$A = \begin{pmatrix} 0 & 1 & 2 \\ 2 & 1 & 0 \end{pmatrix} \quad \text{and} \quad c = (1, 2, 1).$$

The likelihood correspondence of X_c is the surface in $\mathbb{P}^2 \times \mathbb{P}^2$ given by

$$\det \begin{pmatrix} 2p_0 & p_1 \\ p_1 & 2p_2 \end{pmatrix} = \det \begin{pmatrix} p_1 + 2p_2 & 2p_0 + p_1 \\ u_1 + 2u_2 & 2u_0 + u_1 \end{pmatrix} = 0. \tag{29}$$

Note that the second determinant equals the determinant of the 2×2-matrix (28) times 4. Saturating (29) with respect to $p_0 + p_1 + p_2$ reveals two further equations of degree $(1, 1)$:

$$2(u_1 + 2u_2)p_0 = (2u_0 + u_1)p_1 \quad \text{and} \quad (u_1 + 2u_2)p_1 = 2(2u_0 + u_1)p_2.$$

For fixed u, these equations have a unique solution in $\mathbb{P}^2$, given by the formula in (3). ◇

Toric varieties are rational varieties that are parametrized by monomials. We now examine those varieties that are parametrized by generic polynomials. Understanding these is useful for statistics since many widely used models for discrete data are given in the form

$$f : \Theta \to \Delta_n,$$

where Θ is a d-dimensional polytope and f is a polynomial map. The coordinates $f_0, f_1, \ldots, f_n$ are polynomial functions in the parameters $\theta = (\theta_1, \ldots, \theta_d)$ satisfying $f_0 + f_1 + \cdots + f_n = 1$. Such models include the mixture models in Proposition 2.11, phylogenetic models, Bayesian networks, hidden Markov models, and many others arising in computational biology [PS].

The model specified by the polynomials $f_0, \ldots, f_n$ is the semialgebraic set $f(\Theta) \subset \Delta_n$. We study its Zariski closure $X = \overline{f(\Theta)}$ in $\mathbb{P}^n$. Finding its equations is hard and interesting.

Theorem 3.7. *Let $f_0, f_1, \ldots, f_n$ be polynomials of degrees $b_0, b_1, \ldots, b_n$ satisfying $\sum f_i = 1$. The ML degree of the variety X is at most the coefficient of z^d in the generating function*

$$\frac{(1-z)^d}{(1-zb_0)(1-zb_1)\cdots(1-zb_n)}.$$

Equality holds when the coefficients of $f_0, f_1, \ldots, f_n$ are generic relative to $\sum f_i = 1$.

Proof. This is the content of [CHKS, Theorem 1]. □

Example 3.8. We examine the case of quartic surfaces in $\mathbb{P}^3$. Let $d = 2, n = 3$, pick random affine quadrics f_1, f_2, f_3 in two unknowns and set $f_0 = 1 - f_1 - f_2 - f_3$. This defines a map

$$f : \mathbb{C}^2 \to \mathbb{C}^3 \subset \mathbb{P}^3.$$

The ML degree of the image surface $X = \overline{f(\mathbb{C}^2)}$ in $\mathbb{P}^3$ is equal to 25 since

$$\frac{(1-z)^2}{(1-2z)^4} = 1 + 6z + \mathbf{25}z^2 + 88z^3 + \cdots$$

The rational surface X is a Steiner surface (or Roman surface). Its singular locus consists of three lines that meet in a point P. To understand the graph of f, we observe that the linear span of $\{f_0, f_1, f_2, f_3\}$ in $\mathbb{C}[x, y]$ has a basis $\{1, L^2, M^2, N^2\}$ where L, M, N represent lines in $\mathbb{C}^2$. Let l denote the line through $M \cap N$ parallel to L, m the line through $L \cap N$ parallel to M, and n the line through $L \cap M$ parallel to N. The map $\mathbb{C}^2 \to X$ is a bijection outside these three lines, and it maps each line 2-to-1 onto one of the lines in X_{sing}. The fiber over the special point P on X consists of three points, namely, $l \cap m$, $l \cap n$ and $m \cap n$. If the quadric f_0 were also picked at random, rather than as $1 - f_1 - f_2 - f_3$, then we would still get a Steiner surface $X \subset \mathbb{P}^3$. However, now the ML degree of X increases to 33.

On the other hand, if we take X to be a general quartic surface in $\mathbb{P}^3$, so X is a smooth K3 surface of Picard rank 1, then X has ML degree 84. This is the formula in Example 1.11 evaluated at $n = 3$ and $d = 4$. Here $X \backslash \mathcal{H}$ is the generic quartic surface in $\mathbb{P}^3$ with five plane sections removed. The number 84 is the Euler characteristic of that open K3 surface.

In the first case, $X \backslash \mathcal{H}$ is singular, so we cannot apply Theorem 1.7 directly to our Steiner surface X in $\mathbb{P}^3$. However, we can work in the parameter space and consider the smooth very affine surface $\mathbb{C}^2 \backslash V(f_0 f_1 f_2 f_3)$. The number 25 is the Euler characteristic of that surface.

It is instructive to verify Conjecture 1.8 for our three quartic surfaces in $\mathbb{P}^3$. We found

$$\begin{aligned}
\chi(X \backslash \mathcal{H}) = 38 \;&>\; 25 = \text{MLdegree}\,(X),\\
\chi(X \backslash \mathcal{H}) = 49 \;&>\; 33 = \text{MLdegree}\,(X),\\
\chi(X \backslash \mathcal{H}) = 84 \;&=\; 84 = \text{MLdegree}\,(X).
\end{aligned}$$

The Euler characteristics of the three surfaces were computed using Aluffi's method [AluJSC]. ◇

We now turn to the following question: *which projective varieties X have ML degree one?* This question is important for likelihood inference because a model having ML degree one means that the MLE $\hat{p}$ is a rational function in the data u. It is known that Bayesian networks and decomposable graphical models enjoy this property, and it is natural to wonder which other statistical models are in this class. The answer to this question was given by the first author in [Huh2]. We shall here present the result of [Huh2] from a slightly different angle.

Our point of departure is the notion of the A-discriminant, as introduced and studied by Gel'fand, Kapranov and Zelevinsky in [GKZ]. We fix an $r \times m$ integer matrix $A = (a_1, a_2, \ldots, a_m)$ of rank r which has $(1, 1, \ldots, 1)$ in its row space. The Zariski closure of

$$\left\{(t^{a_1} : t^{a_2} : \cdots : t^{a_m}) \in \mathbb{P}^{m-1} \;:\; t \in (\mathbb{C}^*)^r\right\}$$

is an $(r-1)$-dimensional toric variety Y_A in $\mathbb{P}^{m-1}$. We here intentionally changed the notation relative to that used for toric varieties at the beginning of this section. The reason is that d and n are always reserved for the dimension and embedding dimension of a statistical model.

The *dual variety* Y_A^* is an irreducible variety in the dual projective space $(\mathbb{P}^{m-1})^\vee$ whose coordinates are $x = (x_1 : x_2 : \cdots : x_m)$. We identify points x in $(\mathbb{P}^{m-1})^\vee$ with hypersurfaces

$$\left\{ t \in (\mathbb{C}^*)^r \;:\; x_1 \cdot t^{a_1} + x_2 \cdot t^{a_2} + \cdots + x_m \cdot t^{a_m} = 0 \right\}. \tag{30}$$

The dual variety Y_A^* is the Zariski closure in $(\mathbb{P}^{m-1})^\vee$ of the locus of all hypersurfaces (30) that are singular. Typically, Y_A^* is a hypersurface. In that case, Y_A^* is defined by a unique (up to sign) irreducible polynomial $\Delta_A \in \mathbb{Z}[x_1, x_2, \ldots, x_m]$. The homogeneous polynomial Δ_A is called the *A-discriminant*. Many classical discriminants and resultants are instances of Δ_A. So are determinants and hyperdeterminants. This is the punch line of the book [GKZ].

Example 3.9. Let $m = 4, r = 2$, and $A = \begin{pmatrix} 3 & 2 & 1 & 0 \\ 0 & 1 & 2 & 3 \end{pmatrix}$. The associated toric variety is the twisted cubic curve

$$Y_A = \overline{\left\{ (1 : t : t^2 : t^3) \,|\, t \in \mathbb{C} \right\}} \subset \mathbb{P}^3.$$

The variety Y_A^* that is dual to the curve Y_A is a surface in $(\mathbb{P}^3)^\vee$. The surface Y_A^* parametrizes all planes that are tangent to the curve Y_A. These represent univariate cubics

$$x_1 + x_2 t + x_3 t^2 + x_4 t^3$$

that have a double root. Here the A-discriminant is the classical discriminant

$$\Delta_A \;=\; 27x_1^2x_4^2 - 18x_1x_2x_3x_4 + 4x_1x_3^3 + 4x_2^3x_4 - x_2^2x_3^2.$$

The surface Y_A^* in $\mathbb{P}^3$ defined by this equation is the discriminant of the univariate cubic. $\diamondsuit$

Theorem 3.10. *Let $X \subseteq \mathbb{P}^n$ be a projective variety of ML degree* 1. *Each coordinate $\hat{p}_i$ of the rational function $u \mapsto \hat{p}$ is an alternating product of linear forms in $u_0, u_1, \ldots, u_n$.*

The paper [Huh2] gives an explicit construction of the map $u \mapsto \hat{p}$ as a *Horn uniformization*. A precursor was [Kapranov]. We explain this construction. The point of departure is a matrix A as above. We now take Δ_A to be any non-zero homogenous polynomial that vanishes on the dual variety Y_A^* of the toric variety Y_A. If Y_A^* is a hypersurface then Δ_A is the A-discriminant.

First, we write Δ_A as a Laurent polynomial by dividing it by one of its monomials:

$$\frac{1}{\text{monomial}} \cdot \Delta_A \;=\; 1 - c_0 \cdot x^{b_0} - c_1 \cdot x^{b_1} - \cdots - c_n \cdot x^{b_n}. \tag{31}$$

This expression defines an $m \times (n+1)$ integer matrix $B = (b_0, \ldots, b_n)$ satisfying $AB = 0$. Second, we define X to be the rational subvariety of $\mathbb{P}^n$ that is given parametrically by

$$\frac{p_i}{p_0 + p_1 + \cdots + p_n} = c_i \cdot x^{b_i} \qquad \text{for } i = 0, 1, \ldots, n. \tag{32}$$

The defining ideal of X is obtained by eliminating $x_1, \ldots, x_m$ from the equations above. Then X has ML degree 1, and, by Huh [Huh2], every variety of ML degree 1 arises in this manner.

Example 3.11. The following curve in $\mathbb{P}^3$ happens to be a variety of ML degree 1:

$$X = V\big(9p_1p_2 - 8p_0p_3\,,\; p_0^2 - 12(p_0+p_1+p_2+p_3)p_3\big).$$

This curve comes from the discriminant of the univariate cubic in Example 3.9:

$$\frac{1}{\text{monomial}} \cdot \Delta_A = 1 - \Big(\frac{2}{3}\frac{x_2x_3}{x_1x_4}\Big) - \Big(-\frac{4}{27}\frac{x_2^3}{x_1^2x_4}\Big) - \Big(-\frac{4}{27}\frac{x_3^3}{x_1x_4^2}\Big) - \Big(\frac{1}{27}\frac{x_2^2x_3^2}{x_1^2x_4^2}\Big).$$

We derived the curve X from the four parenthesized monomials via the formula (32). The maximum likelihood estimate for this model is given by the products of linear forms

$$\hat{p}_0 = \frac{2}{3}\frac{x_2x_3}{x_1x_4} \qquad \hat{p}_1 = -\frac{4}{27}\frac{x_2^3}{x_1^2x_4} \qquad \hat{p}_2 = -\frac{4}{27}\frac{x_3^3}{x_1x_4^2} \qquad \hat{p}_3 = \frac{1}{27}\frac{x_2^2x_3^2}{x_1^2x_4^2}$$

where

$$\begin{array}{llll} x_1 = -u_0 - u_1 - 2u_2 - 2u_3 & x_2 = u_0 + 3u_2 + 2u_3 \\ x_3 = u_0 + 3u_1 + 2u_3 & x_4 = -u_0 - 2u_1 - u_2 - 2u_3 \end{array}$$

These expressions are the alternating products of linear forms promised in Theorem 3.10. ◇

We now give the formula for $\hat{p}_i$ in general. This is the *Horn uniformization* of [GKZ, §9.3].

Corollary 3.12. *Let $X \subset \mathbb{P}^n$ be the variety of ML degree 1 with parametrization (32) derived from a scaled A-discriminant (31). The coordinates of the MLE function $u \mapsto \hat{p}$ are*

$$\hat{p}_k \quad = \quad c_k \cdot \prod_{j=1}^{m} \Big(\sum_{i=0}^{n} b_{ij} u_i\Big)^{b_{kj}}.$$

It is not obvious (but true) that $\hat{p}_0 + \hat{p}_1 + \cdots + \hat{p}_n = 1$ holds in the formula above. In light of its monomial parametrization, our variety X is toric in $\mathbb{P}^n \backslash \mathcal{H}$. In general, it is not toric in $\mathbb{P}^n$, due to appearances of the factor $(p_0 + p_1 + \cdots + p_n)$ in equations for X. Interestingly, there are numerous instances when this factor does not appear and X is toric also in $\mathbb{P}^n$.

One toric instance is the independence model $X = V(p_{00}p_{11} - p_{01}p_{10})$, whose MLE was derived in Example 1.14. What is the matrix A in this case? We shall answer this question for a slightly larger example, which serves as an illustration for *decomposable graphical models*.

Example 3.13. Consider the conditional independence model for three binary variables given by the graph $\bullet\!-\!\!-\!\bullet\!-\!\!-\!\bullet$. We claim that this graphical model is derived from

$$
A \;=\; \begin{array}{c} \\ x \\ y \\ z \\ w \end{array}
\!\!\begin{array}{c} \begin{array}{ccccccccccc} a_{00} & a_{10} & a_{01} & a_{11} & b_{00} & b_{01} & b_{10} & b_{11} & c_0 & c_1 & d \end{array} \\
\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0
\end{pmatrix} \end{array}.
$$

The discriminant of the corresponding family of hypersurfaces

$$
\big\{(x,y,z,w) \in (\mathbb{C}^*)^4 \mid (a_{00}+a_{10})x + (a_{01}+a_{11})y + (b_{00}+b_{01})z + (b_{10}+b_{11})w + c_0xz + c_1yw + d = 0\big\}
$$

equals

$$
\begin{aligned}
\Delta_A = {} & c_0c_1d - a_{01}b_{10}c_0 - a_{11}b_{10}c_0 - a_{01}b_{11}c_0 - a_{11}b_{11}c_0 \\
& - a_{00}b_{00}c_1 - a_{10}b_{00}c_1 - a_{00}b_{01}c_1 - a_{10}b_{01}c_1.
\end{aligned}
$$

We divide this A-discriminant by its first term c_0c_1d to rewrite it in the form (31) with $n = 7$. The parametrization of $X \subset \mathbb{P}^7$ given by (32) can be expressed as

$$
p_{ijk} \;=\; \frac{a_{ij} \cdot b_{jk}}{c_j \cdot d} \qquad \text{for } i,j,k \in \{0,1\}. \tag{33}
$$

This is indeed the desired graphical model $\bullet\!-\!\!-\!\bullet\!-\!\!-\!\bullet$ with implicit representation

$$
X = V\big(p_{000}p_{101} - p_{001}p_{100}\,,\; p_{010}p_{111} - p_{011}p_{110}\big) \;\subset\; \mathbb{P}^7.
$$

The linear forms used in the Horn uniformization of Corollary 3.12 are

$$
a_{ij} = u_{ij+} \qquad b_{jk} = u_{+jk} \qquad c_j = u_{+j+} \qquad d = u_{+++}
$$

Substituting these expressions into (33), we obtain

$$\hat{p}_{ijk} = \frac{u_{ij+} \cdot u_{+jk}}{u_{+j+} \cdot u_{+++}} \qquad \text{for } i, j, k \in \{0, 1\}.$$

This is the formula in Lauritzen's book [Lau] for MLE of decomposable graphical models. ◇

We now return to the likelihood geometry of an arbitrary d-dimensional projective variety X in $\mathbb{P}^n$, as always defined over $\mathbb{R}$ and not contained in $\mathcal{H}$. We define the *ML bidegree* of X to be the bidegree of its likelihood correspondence $\mathcal{L}_X \subset \mathbb{P}^n \times \mathbb{P}^n$. This is a binary form

$$B_X(p, u) \quad = \quad (b_0 \cdot p^d + b_1 \cdot p^{d-1}u + \cdots + b_d \cdot u^d) \cdot p^{n-d},$$

where $b_0, b_1, \ldots, b_d$ are certain positive integers. By definition, $B_X(p, u)$ is the multidegree [ch3:MS, §8.5] of the prime ideal of $\mathcal{L}_X$, with respect to the natural $\mathbb{Z}^2$-grading on the polynomial ring $\mathbb{R}[p, u] = \mathbb{R}[p_0, \ldots, p_n, u_0, \ldots, u_n]$. Equivalently, the ML bidegree $B_X(p, u)$ is the class defined by $\mathcal{L}_X$ in the cohomology ring

$$H^*(\mathbb{P}^n \times \mathbb{P}^n; \mathbb{Z}) = \mathbb{Z}[p, u]/\langle p^{n+1}, u^{n+1}\rangle.$$

We already saw some examples, for the Grassmannian $G(2, 4)$ in (12), for arbitrary linear spaces in (14), and for a toric model of ML degree 1 in (23). We note that the bidegree $B_X(p, u)$ can be computed conveniently using the command `multidegree` in `Macaulay2`.

To understand the geometric meaning of the ML bidegree, we introduce a second polynomial. Let L_{n-i} be a sufficiently general linear subspace of $\mathbb{P}^n$ of codimension i, and define

$$s_i \;=\; \text{MLdegree}\,(X \cap L_{n-i}).$$

We define the *sectional ML degree* of X to be the polynomial

$$S_X(p, u) \quad = \quad (s_0 \cdot p^d + s_1 \cdot p^{d-1}u + \cdots + s_d \cdot u^d) \cdot p^{n-d},$$

Example 3.14. The sectional ML degree of the Grassmannian $G(2, 4)$ in (10) equals

$$S_X(p, u) = 4p^5 + 20p^4u + 24p^3u^2 + 12p^2u^3 + 2pu^4.$$

Thus, if H_1, H_2, H_3 denote generic hyperplanes in $\mathbb{P}^5$, then the threefold $G(2, 4) \cap H_1$ has ML degree 20, the surface $G(2, 4) \cap H_1 \cap H_2$ has ML degree 24, and the curve $G(2, 4) \cap H_1 \cap H_2 \cap H_3$ has ML degree 12. Lastly, the coefficient 2 of pu^4 is simply the degree of $G(2, 4)$ in $\mathbb{P}^5$. ◇

Conjecture 3.15. The ML bidegree and the sectional ML degree of any projective variety $X \subset \mathbb{P}^n$, not lying in $\mathcal{H}$, are related by the following involution on binary forms of degree n:

$$B_X(p,u) = \frac{u \cdot S_X(p, u-p) - p \cdot S_X(p,0)}{u-p},$$
$$S_X(p,u) = \frac{u \cdot B_X(p, u+p) + p \cdot B_X(p,0)}{u+p}.$$

This conjecture is a theorem when $X \backslash \mathcal{H}$ is smooth and its boundary is *schön*. See Theorem 4.6 below. In that case, the ML bidegree is identified, by Huh [Huh1, Theorem 2], with the Chern–Schwartz–MacPherson (CSM) class of the constructible function on $\mathbb{P}^n$ that is 1 on $X \backslash \mathcal{H}$ and 0 elsewhere. Aluffi proved in [Alu, Theorem 1.1] that the CSM class of an locally closed subset of $\mathbb{P}^n$ satisfies such a *log-adjunction formula*. Our formula in Conjecture 3.15 is precisely the homogenization of Aluffi's involution. The combination of [Alu, Theorem 1.1] and [Huh1, Theorem 2] proves Conjecture 3.15 in cases such as generic complete intersections (Theorem 1.10) and arbitrary linear spaces (Theorem 1.20). In the latter case, it can also be verified using matroid theory. Conjecture 3.15 says that this holds for any X, indicating a deeper connection between likelihood correspondences and CSM classes.

We note that $B_X(p,u)$ and $S_X(p,u)$ always share the same leading term and the same trailing term, and this is compatible with our formulas. Both polynomials start and end like

$$\text{MLdegree}\,(X) \cdot p^n \; + \; \cdots \; + \text{degree}\,(X) \cdot p^{\text{codim}(X)} u^{\dim(X)}.$$

We now illustrate Conjecture 3.15 by verifying it computationally for a few more examples.

Example 3.16. Let us examine some cubic fourfolds in $\mathbb{P}^5$. If X is a generic hypersurface of degree 4 in $\mathbb{P}^5$ then its sectional ML degree and ML bidegree satisfy the conjectured formula:

$$\begin{aligned} S_X(p,u) &= 1364p^5 + 448p^4u + 136p^3u^2 + 32p^2u^3 + 3pu^4, \\ B_X(p,u) &= \;1364p^5 + 341p^4u + 81p^3u^2 + 23p^2u^3 + 3pu^4. \end{aligned}$$

Of course, in algebraic statistics, we are more interested in special hypersurfaces that are statistically meaningful. One such instance was seen in Example 2.7. The mixture model for two identically distributed ternary random variables is the fourfold $X \subset \mathbb{P}^5$ defined by

$$\det \begin{pmatrix} 2p_{11} & p_{12} & p_{13} \\ p_{12} & 2p_{22} & p_{23} \\ p_{13} & p_{23} & 2p_{33} \end{pmatrix} \quad = \quad 0. \tag{34}$$

The sectional ML degree and the ML bidegree of this determinantal fourfold are

$$\begin{aligned} S_X(p,u) &= 6p^5 + 42p^4u + 48p^3u^2 + 21p^2u^3 + 3pu^4 \\ B_X(p,u) &= 6p^5 + 12p^4u + 15p^3u^2 + 12p^2u^3 + 3pu^4. \end{aligned}$$

For the toric fourfold $X = V(p_{11}p_{22}p_{33} - p_{12}p_{13}p_{23})$, ML bidegree and sectional ML degree are

$$\begin{aligned} B_X(p,u) &= \mathbf{3}p^5 + 3p^4u + 3p^3u^2 + 3p^2u^3 + 3pu^4, \\ S_X(p,u) &= \mathbf{3}p^5 + 12p^4u + 18p^3u^2 + 12p^2u^3 + 3pu^4. \end{aligned}$$

Now, taking $X = V(p_{11}p_{22}p_{33} + p_{12}p_{13}p_{23})$ instead, the leading coefficient 3 changes to 2. $\diamond$

Remark 3.17. Conjecture 3.15 is true when X_c is a toric variety with c generic, as in Theorem 3.2. Here we can use Proposition 3.5 to infer that all coefficients of B_X are equal to the normalized volume of the lattice polytope $\mathrm{conv}(A)$. In symbols, for generic c, we have

$$B_{X_c}(p,u) \quad = \quad \text{degree}\,(X_c) \cdot \sum_{i=0}^{d} p^{n-i}u^i.$$

It is now an exercise to transform this into a formula for the sectional ML degree $S_{X_c}(p,u)$.

In general, it is hard to compute generators for the ideal of the likelihood correspondence.

Example 3.18. The following submodel of (34) was featured prominently in [HKS, §1]:

$$\det \begin{pmatrix} 12p_0 & 3p_1 & 2p_2 \\ 3p_1 & 2p_2 & 3p_3 \\ 2p_2 & 3p_3 & 12p_4 \end{pmatrix} \quad = \quad 0. \tag{35}$$

This cubic threefold X is the secant variety of a rational normal curve in $\mathbb{P}^4$, and it represents the mixture model for a binomial random variable (tossing a biased coin four times). It takes several hours in `Macaulay2` to compute the prime ideal of the likelihood correspondence $\mathcal{L}_X \subset \mathbb{P}^4 \times \mathbb{P}^4$. That ideal has 20 minimal generators one in degree $(1,1)$, one in degree $(3,0)$, five in degree $(3,1)$, ten in degree $(4,1)$ and three in degree $(3,2)$. After passing to a Gröbner basis, we use the formula in [ch3:MS, Definition 8.45] to compute the bidegree of $\mathcal{L}_X$:

$$B_X(p,u) \quad = \quad 12p^4 + 15p^3u + 12p^2u^2 + 3pu^3.$$

We now intersect X with random hyperplanes in $\mathbb{P}^4$, and we compute the ML degrees of the intersections. Repeating this experiment many times reveals the sectional ML degree of X:

$$S_X(p,u) \quad = \quad 12p^4 + 30p^3u + 18p^2u^2 + 3pu^3.$$

The two polynomials satisfy our transformation rule, thus confirming Conjecture 3.15. We note that Conjecture 1.8 also holds for this example: using Aluffi's method [AluJSC], we find $\chi(X\backslash\mathcal{H}) = -13$. $\diamond$

Our last topic is the operation of restriction and deletion. This is a standard tool for complements of hyperplane arrangements, as in Theorem 1.20. It was developed in [Huh1] for arbitrary very affine varieties, such as $X\backslash\mathcal{H}$. We motivate this by explaining the distinction between *structural zeros* and *sampling zeros* for contingency tables in statistics [BFH, §5.1.1].

Returning to the "hair loss due to TV soccer" example from the beginning of Sect. 2, let us consider the following questions. What is the difference between the data set

$$U \quad = \quad \begin{array}{r c} & \begin{array}{ccc} \text{lots of hair} & \text{medium hair} & \text{little hair} \end{array} \\ \begin{array}{r} \leq 2\,\text{h} \\ 2\text{–}6\,\text{h} \\ \geq 6\,\text{h} \end{array} & \begin{pmatrix} \quad 15 \quad & \quad 0 \quad & \quad 9 \quad \\ 20 & 24 & 12 \\ 10 & 12 & 6 \end{pmatrix} \end{array}$$

and the data set

$$\tilde{U} \quad = \quad \begin{array}{r c} & \begin{array}{ccc} \text{lots of hair} & \text{medium hair} & \text{little hair} \end{array} \\ \begin{array}{r} \leq 2\,\text{h} \\ 2\text{–}6\,\text{h} \\ \geq 6\,\text{h} \end{array} & \begin{pmatrix} \quad 10 \quad & \quad 0 \quad & \quad 5 \quad \\ 9 & 3 & 6 \\ 7 & 9 & 8 \end{pmatrix} \end{array} ?$$

How should we think about the zero entries in row 1 and column 2 of these two contingency tables? Would the rank 1 model $\mathcal{M}_1$ or the rank 2 model $\mathcal{M}_2$ be more appropriate?

The first matrix U has rank 2 and it can be completed to a rank 1 matrix by replacing the zero entry with 18. Thus, the model $\mathcal{M}_1$ fits perfectly except for the *structural zero* in row 1 and column 2. It seems that this zero is inherent in the structure of the problem: planet Earth simply has no people with medium hair length who rarely watch soccer on TV.

The second matrix $\tilde{U}$ also has rank two, but it cannot be completed to rank 1. The model $\mathcal{M}_2$ is a perfect fit. The zero entry in $\tilde{U}$ appeared to be an artifact of the particular group that was interviewed in this study. This is a *sampling zero*. It arose because, by chance, in this cohort nobody happened to have medium hair length and watch soccer on TV rarely. We refer to the book of Bishop et al. [BFH, Chap. 5] for an introduction.

We now consider an arbitrary projective variety $X \subseteq \mathbb{P}^n$, serving as our statistical model. Suppose that structural zeros or sampling zeros occur in the last coordinate u_n. Following [Rapallo, Theorem 4], we model structural zeros by the projection $\pi_n(X)$. This model is the variety in $\mathbb{P}^{n-1}$ that is the closure of the image of X under the rational map

$$\pi_n : \mathbb{P}^n \dashrightarrow \mathbb{P}^{n-1}, \qquad (p_0 : p_1 : \cdots : p_{n-1} : p_n) \longmapsto (p_0 : p_1 : \cdots : p_{n-1}).$$

Which projective variety is a good representation for sampling zeros? We propose that sampling zeros be modeled by the intersection $X \cap \{p_n{=}0\}$. This is now to be regarded as a subvariety in $\mathbb{P}^{n-1}$. In this manner, both structural zeros and sampling zeros are modeled by closed subvarieties of $\mathbb{P}^{n-1}$. Inside that ambient $\mathbb{P}^{n-1}$, our standard arrangement $\mathcal{H}$ consists of $n + 1$ hyperplanes. Usually, none of these hyperplanes contains $X \cap \{p_n{=}0\}$ or $\pi_n(X)$.

It would be desirable to express the (sectional) ML degree of X in terms of those of the intersection $X \cap \{p_n = 0\}$ and the projection $\pi_n(X)$. As an alternative to the ML degree of the projection $\pi_n(X)$ into $\mathbb{P}^{n-1}$, here is a quantity in $\mathbb{P}^n$ that reflects the presence of structural zeros even more accurately. We denote by

$$\text{MLdegree}\,(X|_{u_n=0})$$

the number of critical points $\hat{p} = (\hat{p}_0 : \hat{p}_1 : \cdots : \hat{p}_{n-1} : \hat{p}_n)$ of ℓ_u in $X_{\text{reg}} \backslash \mathcal{H}$ for those data vectors $u = (u_0, u_1, \ldots, u_{n-1}, 0)$ whose first n coordinates u_i are positive and generic.

Conjecture 3.19. The maximum likelihood degree satisfies the inductive formula

$$\text{MLdegree}\,(X) = \text{MLdegree}\,(X \cap \{p_n{=}0\}) \,+\, \text{MLdegree}\,(X|_{u_n=0}), \tag{36}$$

provided X and $X \cap \{p_n{=}0\}$ are reduced, irreducible, and not contained in their respective $\mathcal{H}$.

We expect that an analogous formula will hold for the sectional ML degree $S_X(p,u)$. The intuition behind equation (36) is as follows. As the data vector u moves from a general point in $\mathbb{P}^n_u$ to a general point on the hyperplane $\{u_n = 0\}$, the corresponding fiber $\text{pr}_2^{-1}(u)$ of the likelihood fibration splits into two clusters. One cluster has size MLdegree $(X|_{u_n=0})$ and stays away from $\mathcal{H}$. The other cluster moves onto the hyperplane $\{p_n = 0\}$ in $\mathbb{P}^n_p$, where it approaches the various critical points of ℓ_u in that intersection. This degeneration is the perfect scenario for a numerical homotopy, e.g. in `Bertini`, as discussed in Sect. 2. These homotopies are currently being studied for determinantal varieties by Elizabeth Gross and Jose Rodriguez [GR]. The formula (36) has been verified computationally for many examples. Also, Conjecture 3.19 is known to be true in the slightly different setting of [Huh1], under a certain smoothness assumption. This is the content of [Huh1, Corollary 3.2].

Example 3.20. Fix the space $\mathbb{P}^8$ of 3×3-matrices as in Sect. 2. For the rank 2 variety $X = \mathcal{V}_2$, the formula (36) reads $10 = 5+5$. For the rank 1 variety $X = \mathcal{V}_1$, it reads $1 = 0 + 1$. ◇

Example 3.21. If X is a generic (d, e)-curve in $\mathbb{P}^3$, then

$$\text{MLdegree}\,(X) = d^2e + de^2 + de \quad \text{and} \quad X \cap \{p_3 = 0\} = (d \cdot e \text{ distinct points}).$$

Computations suggest that

$$\text{MLdegree}\,(X|_{u_3=0}) = d^2e + de^2 \quad \text{and} \quad \text{MLdegree}\,(\pi_3(X)) = d^2e + de^2.$$

To derive the second equality geometrically, one may argue as follows. Both curves $X \subset \mathbb{P}^3$ and $\pi_3(X) \subset \mathbb{P}^2$ have degree de and genus $\frac{1}{2}(d^2e + de^2) - 2de + 1$. Subtracting this from the expected genus $\frac{1}{2}(de-1)(de-2)$ of a plane curve of degree de, we find that $\pi_3(X)$ has $\frac{1}{2}d(d-1)e(e-1)$ nodes. Example 1.4 suggests that each node decreases the ML degree of a plane curve by 2. Assuming this to bet the case, we conclude

$$\text{MLdegree}\,(\pi_3(X)) = de(de+1) - d(d-1)e(e-1) = d^2e + de^2.$$

Here we are using that a general plane curve of degree de has ML degree $de(de+1)$. ◇

This example suggests that, in favorable circumstances, the following identity would hold:

$$\text{MLdegree}\,(X|_{u_n=0}) = \text{MLdegree}\,(\pi_n(X)). \tag{37}$$

However, this is certainly not true in general. Here is a particularly telling example:

Example 3.22. Suppose that X is a generic surface of degree d in $\mathbb{P}^3$. Then

$$\begin{aligned}
\text{MLdegree}\,(X) &= d + d^2 + d^3,\\
\text{MLdegree}\,(X \cap \{p_3 = 0\}) &= d + d^2,\\
\text{MLdegree}\,(X|_{u_3=0}) &= d^3,\\
\text{MLdegree}\,(\pi_3(X)) &= 1.
\end{aligned}$$

Indeed, for most hypersurfaces $X \subset \mathbb{P}^n$, the same will happen, since $\pi_n(X) = \mathbb{P}^{n-1}$. ◇

As a next step, one might conjecture that (37) holds when the map is birational and the center $(0 : \cdots : 0 : 1)$ of the projection does not lie on the variety X. But this also fails:

Example 3.23. Let X be the twisted cubic curve in $\mathbb{P}^3$ defined by the 2×2-minors of

$$\begin{pmatrix} p_0 + p_1 - p_2 & 2p_0 - p_2 + 9p_3 & p_0 - 6p_1 + 8p_2 \\ 2p_0 - p_2 + 9p_3 & p_0 - 6p_1 + 8p_2 & 7p_0 + p_1 + 2p_2 \end{pmatrix}.$$

The ML degree of X is $13 = 3 + 10$, and X intersects $\{p_3 = 0\}$ in three distinct points. The projection of the curve X into $\mathbb{P}^2$ is a cuspidal cubic, as in Example 1.4. We have

$$\text{MLdegree}\,(X|_{u_3=0}) = 10 \quad \text{and} \quad \text{MLdegree}\,(\pi_3(X)) = 9.$$

It is also instructive to compare the number $13 = -\chi(X\backslash\mathcal{H})$ with the number 11 one gets in Theorem 3.7 for the special twisted cubic curve with $d = 1, n = 3$ and $b_0 = b_1 = b_2 = b_3 = 3$. There are many mysteries still to be explored in likelihood geometry, even within $\mathbb{P}^3$. ◇

4 Characteristic Classes

We start by giving an alternative description of the likelihood correspondence which reveals its intimate connection with the theory of Chern classes on possibly noncompact varieties. An important role will be played by the Lie algebra and cotangent bundle of the algebraic torus $(\mathbb{C}^*)^{n+1}$. This section ties our discussion to the work of Aluffi [AluJSC, AluLectures, Alu] and Huh [Huh1, Huh0, Huh2]. In particular, we introduce and explain Chern–Schwartz–MacPherson (CSM) classes. And, most importantly, we present proofs for Theorems 1.6, 1.7, 1.15, and 1.20.

Let $X \subseteq \mathbb{P}^n$ be a closed and irreducible subvariety of dimension d, not contained in our distinguished arrangement of $n + 2$ hyperplanes,

$$\mathcal{H} = \{(p_0 : p_1 : \cdots : p_n) \in \mathbb{P}^n \mid p_0 \cdot p_1 \cdots p_n \cdot p_+ = 0\}, \qquad p_+ = \sum_{i=0}^{n} p_i.$$

Let φ_i denote the restriction of the rational function p_i/p_+ to $X\backslash\mathcal{H}$. The closed embedding

$$\varphi : X\backslash\mathcal{H} \longrightarrow (\mathbb{C}^*)^{n+1}, \qquad \varphi = (\varphi_0, \ldots, \varphi_n),$$

shows that the variety $X\backslash\mathcal{H}$ is *very affine*. Let x be a smooth point of $X\backslash\mathcal{H}$. We define

$$\gamma_x \;:\; T_xX \;\longrightarrow\; T_{\varphi(x)}(\mathbb{C}^*)^{n+1} \;\longrightarrow\; \mathfrak{g} := T_1(\mathbb{C}^*)^{n+1} \tag{38}$$

to be the derivative of φ at x followed by that of left-translation by $\varphi(x)^{-1}$. Here $\mathfrak{g}$ is the Lie algebra of the algebraic torus $(\mathbb{C}^*)^{n+1}$. In local coordinates $(x_1, \ldots, x_d)$ around the smooth point x, the linear map γ_x is represented by the logarithmic Jacobian matrix

$$\left(\frac{\partial \log \varphi_i}{\partial x_j}\right), \quad 0 \le i \le n, \quad 1 \le j \le d.$$

The linear map γ_x in (38) is injective because φ is injective. We write $q_0, \ldots, q_n$ for the coordinate functions on the torus $(\mathbb{C}^*)^{n+1}$. These functions define a $\mathbb{C}$-linear basis of the dual Lie algebra $\mathfrak{g}^\vee$ corresponding to differential forms

$$\mathrm{dlog}(q_0), \ldots, \mathrm{dlog}(q_n) \ \in \ H^0\Big((\mathbb{C}^*)^{n+1}, \Omega^1_{(\mathbb{C}^*)^{n+1}}\Big) \ \simeq \ \mathfrak{g}^\vee \ \simeq \ \mathbb{C}^{n+1}.$$

We fix this choice of basis of $\mathfrak{g}^\vee$, and we identify $\mathbb{P}(\mathfrak{g}^\vee)$ with the space of data vectors $\mathbb{P}^n_u$:

$$\mathfrak{g}^\vee \ \simeq \ \Big\{ \sum_{i=0}^{n} u_i \cdot \mathrm{dlog}(q_i) \mid u = (u_0, \ldots, u_n) \in \mathbb{C}^{n+1} \Big\}.$$

Consider the vector bundle homomorphism defined by the pullback of differential forms

$$\gamma^\vee : \mathfrak{g}^\vee_{X_{\mathrm{reg}} \backslash \mathcal{H}} \longrightarrow \Omega^1_{X_{\mathrm{reg}} \backslash \mathcal{H}}, \qquad (x, u) \longmapsto \sum_{i=0}^{n} u_i \cdot \mathrm{dlog}(\varphi_i)(x). \tag{39}$$

Here $\mathfrak{g}^\vee_{X_{\mathrm{reg}} \backslash \mathcal{H}}$ is the trivial vector bundle over $X_{\mathrm{reg}} \backslash \mathcal{H}$ modeled on the vector space $\mathfrak{g}^\vee$. The induced linear map $\gamma^\vee_x$ between the fibers over a smooth point x is dual to the injective linear map $\gamma_x : T_x X \longrightarrow \mathfrak{g}$. Therefore $\gamma^\vee$ is surjective and $\ker(\gamma^\vee)$ is a vector bundle over $X_{\mathrm{reg}} \backslash \mathcal{H}$. This vector bundle has positive rank $n - d + 1$, and hence its projectivization is nonempty.

Proof of Theorem 1.6. Under the identification $\mathbb{P}(\mathfrak{g}^\vee) \simeq \mathbb{P}^n_u$, the projective bundle $\mathbb{P}(\ker \gamma^\vee)$ corresponds to the following constructible subset of dimension n:

$$\mathcal{L}_X \cap \Big((X_{\mathrm{reg}} \backslash \mathcal{H}) \times \mathbb{P}^n_u\Big) \ \subseteq \ \mathbb{P}^n_p \times \mathbb{P}^n_u.$$

Therefore its Zariski closure $\mathcal{L}_X$ is irreducible of dimension n, and $\mathrm{pr}_1 : \mathcal{L}_X \to \mathbb{P}^n_p$ is a projective bundle over $X_{\mathrm{reg}} \backslash \mathcal{H}$. The likelihood vibration $\mathrm{pr}_2 : \mathcal{L}_X \to \mathbb{P}^n_u$ is generically finite-to-one because the domain and the range are algebraic varieties of the same dimension. $\square$

Our next aim is to prove Theorem 1.15. For this we fix a resolution of singularities

$$\begin{tikzcd} \pi^{-1}(X_{\mathrm{reg}}\backslash\mathcal{H}) \arrow[r] \arrow[d] & \tilde{X} \arrow[d, "\pi"] & \\ X_{\mathrm{reg}}\backslash\mathcal{H} \arrow[r] & X \arrow[r] & \mathbb{P}^n, \end{tikzcd}$$

where π is an isomorphism over $X_{\mathrm{reg}}\backslash\mathcal{H}$, the variety $\tilde{X}$ is smooth and projective, and the complement of $\pi^{-1}(X_{\mathrm{reg}}\backslash\mathcal{H})$ is a simple normal crossing divisor in $\tilde{X}$ with irreducible components $D_1, \ldots, D_k$. Each φ_i lifts to a rational function on $\tilde{X}$ which is regular on $\pi^{-1}(X\backslash\mathcal{H})$. If $u = (u_0, \ldots, u_n)$ is an integer vector in $\mathbb{Z}^{n+1}$, then these functions satisfy

$$\mathrm{ord}_{D_j}(\ell_u) \quad = \quad \sum_{i=0}^{n} u_i \cdot \mathrm{ord}_{D_j}(\varphi_i). \tag{40}$$

If $u \in \mathbb{C}^{n+1}\backslash\mathbb{Z}^{n+1}$ then $\mathrm{ord}_{D_j}(\ell_u)$ is the complex number defined by the Eq. (40) for $j = 1, \ldots, k$. We write $H_i := \{p_i = 0\}$ and $H_+ := \{p_+ = 0\}$ for the $n+2$ hyperplanes in $\mathcal{H}$.

Lemma 4.1. *Suppose that $X \cap H_i$ is smooth along H_+, and let D_j be a divisor in the boundary of $\tilde{X}$ such that $\pi(D_j) \subseteq \mathcal{H}$. Then the following three statements hold:*

(1) If $\pi(D_j) \not\subseteq H_+$ then $\mathrm{ord}_{D_j}(\varphi_i)$ is $\begin{cases} \textit{positive} & \textit{if } \pi(D_j) \subseteq H_i, \\ \textit{zero} & \textit{if } \pi(D_j) \not\subseteq H_i. \end{cases}$

(2) If $\pi(D_j) \subseteq H_+$ then $-\mathrm{ord}_{D_j}(\varphi_i)$ is $\begin{cases} \textit{positive} & \textit{if } \pi(D_j) \not\subseteq H_i, \\ \textit{nonnegative} & \textit{if } \pi(D_j) \subseteq H_i. \end{cases}$

(3) In each of the above two cases, $\mathrm{ord}_{D_j}(\varphi_i)$ is non-zero for at least one index i.

Proof. Write H_i' and H_+' for the pullbacks of H_i and H_+ to X respectively. Note that $\mathrm{ord}_{D_j}(\pi^*(H_i'))$ is positive if D_j is contained in $\pi^{-1}(H_i')$ and otherwise zero. Since

$$\mathrm{ord}_{D_j}(\varphi_i) \;=\; \mathrm{ord}_{D_j}(\pi^*(H_i')) \;-\; \mathrm{ord}_{D_j}(\pi^*(H_+')),$$

this proves the first and second assertion, except for the case when $\pi(D_j) \subseteq H_i \cap H_+$. In this case, our assumption that H_i' is smooth along H_+' shows that $\pi(D_j) \subseteq X_{\mathrm{reg}}$ and the order of vanishing of H_i' along $\pi(D_j)$ is 1. Therefore

$$-\mathrm{ord}_{D_j}(\varphi_i) \;=\; \mathrm{ord}_{D_j}(\pi^*(H_+')) - 1 \geq 0.$$

The third assertion of Lemma 4.1 is derived by the following set-theoretic reasoning:

- If $\pi(D_j) \nsubseteq H_+$, then $\pi(D_j) \subseteq H_i$ for some i because $\pi(D_j) \subseteq \mathcal{H}$ is irreducible.
- If $\pi(D_j) \subseteq H_+$, then $\pi(D_j) \nsubseteq H_i$ for some i because $\bigcap_{i=0}^{n} H_i = \emptyset$.

$\square$

From Lemma 4.1 and Eq. (40) we deduce the following result. In Lemmas 4.2 and 4.3 we retain the hypothesis from Lemma 4.1 which coincides with that in Theorem 1.15.

Lemma 4.2. *If $\pi(D_j) \subseteq \mathcal{H}$ and $u \in \mathbb{R}^{n+1}_{>0}$ is strictly positive, then $\mathrm{ord}_{D_j}(\ell_u)$ is nonzero.*

Consider the sheaf of logarithmic differential one-forms $\Omega^1_{\tilde{X}}(\log D)$, where D is the sum of the irreducible components of $\pi^{-1}(\mathcal{H})$. If u is an integer vector, then the corresponding likelihood function ℓ_u on $\tilde{X}$ defines a global section of this sheaf:

$$\mathrm{dlog}(\ell_u) \;=\; \sum_{i=0}^{n} u_i \cdot \mathrm{dlog}(\varphi_i) \;\in\; H^0\big(\tilde{X}, \Omega^1_{\tilde{X}}(\log D)\big). \tag{41}$$

If $u \in \mathbb{C}^{n+1} \backslash \mathbb{Z}^{n+1}$ then we define the global section $\mathrm{dlog}(\ell_u)$ by the above expression (41).

Lemma 4.3. *If $u \in \mathbb{R}^{n+1}_{>0}$ is strictly positive, then $\mathrm{dlog}(\ell_u)$ does not vanish on $\pi^{-1}(\mathcal{H})$.*

Proof. Let $x \in \pi^{-1}(\mathcal{H})$ and $D_1, \ldots, D_l$ the irreducible components of D containing x, with local equations $g_1, \ldots, g_l$ on a small neighborhood G of x. Clearly, $l \geq 1$. By passing to a smaller neighborhood if necessary, we may assume that $\Omega^1_{\tilde{X}}(\log D)$ trivializes over G, and

$$\mathrm{dlog}(\ell_u) \quad = \quad \sum_{j=1}^{l} \mathrm{ord}_{D_j}(\ell_u) \cdot \mathrm{dlog}(g_j) \;+\; \psi,$$

where ψ is a regular 1-form. Since the $\mathrm{dlog}(g_j)$ form part of a free basis of a trivialization of $\Omega^1_{\tilde{X}}(\log D)$ over G, Lemma 4.2 implies that $\mathrm{dlog}(\ell_u)$ is nonzero on $\pi^{-1}(\mathcal{H})$ if $u \in \mathbb{R}^{n+1}_{>0}$. $\square$

Proof of Theorem 1.7. In the notation above, the logarithmic Poincaré–Hopf theorem states

$$\int_{\tilde{X}} c_d\big(\Omega^1_{\tilde{X}}(\log D)\big) = (-1)^d \cdot \chi\big(\tilde{X} \backslash \pi^{-1}(\mathcal{H})\big).$$

See [AluLectures, Sect. 3.4] for example. If $X \backslash \mathcal{H}$ is smooth, then Lemma 4.3 shows that, for generic u, the zero-scheme of the Eq. (41) is equal to the likelihood locus

$$\{x \in X\backslash\mathcal{H} \mid \mathrm{dlog}(\ell_u)(x) = 0\}.$$

Since the likelihood locus is a zero-dimensional scheme of length equal to the ML degree of X, the logarithmic Poincaré–Hopf theorem implies Theorem 1.7. □

Proof of Theorem 1.15. Suppose that the likelihood locus $\{x \in X_{\mathrm{reg}}\backslash\mathcal{H} \mid \mathrm{dlog}(\ell_u)(x) = 0\}$ contains a curve. Let C and $\tilde{C}$ denote the closures of that curve in X and $\tilde{X}$ respectively. Let $\pi^*(\mathcal{H})$ be the pullback of the divisor $\mathcal{H} \cap X$ of X. If $u \in \mathbb{R}^{n+1}_{>0}$ then Lemma 4.3 implies that $\pi^*(\mathcal{H}) \cdot \tilde{C}$ is rationally equivalent to zero in $\tilde{X}$. It then follows from the Projection Formula that $\mathcal{H} \cdot C$ is also rationally equivalent to zero in $\mathbb{P}^n$. But this is impossible. Therefore the likelihood locus does not contain a curve. This proves the first part of Theorem 1.15.

For the second part, we first show that $\mathrm{pr}_2^{-1}(u)$ is contained in $X\backslash\mathcal{H}$ for a strictly positive vector u. This means there is no pair $(x, u) \in \mathcal{L}_X$ with $x \in \mathcal{H}$ which is a limit of the form

$$(x, u) = \lim_{t\to 0}(x_t, u_t), \qquad x_t \in X_{\mathrm{reg}}\backslash\mathcal{H}, \qquad \mathrm{dlog}(\ell_{u_t})(x_t) = 0.$$

If there is such a sequence (x_t, u_t), then we can take its limit over $\tilde{X}$ to find a point $\tilde{x} \in \tilde{X}$ such that $\mathrm{dlog}(\ell_u)(\tilde{x}) = 0$, but this would contradict Lemma 4.3.

Now suppose that the fiber $\mathrm{pr}_2^{-1}(u)$ is contained in X_{reg}, and hence in $X_{\mathrm{reg}}\backslash\mathcal{H}$. By Theorem 1.6, this fiber $\mathrm{pr}_2^{-1}(u)$ is contained the smooth variety $(\mathcal{L}_X)_{\mathrm{reg}}$. Furthermore, by the first part of Theorem 1.15, $\mathrm{pr}_2^{-1}(u)$ is a zero-dimensional subscheme of $(\mathcal{L}_X)_{\mathrm{reg}}$. The assertion on the length of the fiber now follows from a standard result on intersection theory on Cohen–Macaulay varieties. More precisely, we have

$$\mathrm{MLdegree}(X) \;=\; (U_1 \cdot \ldots \cdot U_n)_{\mathcal{L}_X} \;=\; (U_1 \cdot \ldots \cdot U_n)_{(\mathcal{L}_X)_{\mathrm{reg}}} \;=\; \deg(\mathrm{pr}_2^{-1}(u)),$$

where the U_i are pullbacks of sufficiently general hyperplanes in $\mathbb{P}^n_u$ containing u, and the two terms in the middle are the intersection numbers defined in [Fulton, Definition 2.4.2]. The fact that $(\mathcal{L}_X)_{\mathrm{reg}}$ is Cohen–Macaulay is used in the last equality [Fulton, Example 2.4.8]. □

Remark 4.4. If X is a curve, then the zero-scheme of the Eq. (41) is zero-dimensional for generic u, even if $X\backslash\mathcal{H}$ is singular. Furthermore, the length of this zero-scheme is at least as large as ML degree of X. Therefore

$$-\chi(X\backslash\mathcal{H}) \;\geq\; -\chi\big(\tilde{X}\backslash\pi^{-1}(\mathcal{H})\big) \;\geq\; \mathrm{MLdegree}\,(X).$$

This proves that Conjecture 1.8 holds for $d = 1$.

Next we give a brief description of the Chern–Schwartz–MacPherson (CSM) class. For a gentle introduction we refer to [AluLectures]. The group $C(X)$ of *constructible functions* on a complex algebraic variety X is a subgroup of the group of integer valued functions on X. It is generated by the characteristic functions $\mathbf{1}_Z$ of all closed subvarieties Z of X. If $f : X \to Y$ is a morphism between complex algebraic varieties, then the pushforward of constructible functions is the homomorphism

$$f_* : C(X) \longrightarrow C(Y), \qquad \mathbf{1}_Z \longmapsto \Big(y \longmapsto \chi\big(f^{-1}(y) \cap Z\big), \quad y \in Y\Big).$$

If X is a compact complex manifold, then the characteristic class of X is the Chern class of the tangent bundle $c(TX) \cap [X] \in H_*(X;\mathbb{Z})$. A generalization to possibly singular or noncompact varieties is provided by the Chern–Schwartz–MacPherson class, whose existence was once a conjecture of Deligne and Grothendieck.

In the next definition, we write C for the functor of constructible functions from the category of complete complex algebraic varieties to the category of abelian groups.

Definition 4.5. The *CSM class* is the unique natural transformation

$$c_{SM} : C \longrightarrow H_*$$

such that $c_{SM}(\mathbf{1}_X) = c(TX) \cap [X] \in H_*(X;\mathbb{Z})$ when X is smooth and complete.

The uniqueness follows from the naturality, the resolution of singularities over $\mathbb{C}$, and the requirement for smooth and complete varieties. We highlight two properties of the CSM class which follow directly from Definition 4.5:

1. The CSM class satisfies the inclusion–exclusion relation

$$c_{SM}(\mathbf{1}_{U\cup U'}) = c_{SM}(\mathbf{1}_U) + c_{SM}(\mathbf{1}_{U'}) - c_{SM}(\mathbf{1}_{U\cap U'}) \in H_*(X;\mathbb{Z}). \qquad (42)$$

2. The CSM class captures the topological Euler characteristic as its degree:

$$\chi(U) = \int_X c_{SM}(\mathbf{1}_U) \in \mathbb{Z}. \qquad (43)$$

Here U and U' are arbitrary constructible subsets of a complete variety X.

What kind of information on a constructible subset is encoded in its CSM class? In likelihood geometry, U is a constructible subset in the complex projective space $\mathbb{P}^n$, and we identify $c_{SM}(\mathbf{1}_U)$ with its image in $H_*(\mathbb{P}^n,\mathbb{Z}) = \mathbb{Z}[p]/\langle p^{n+1}\rangle$. Thus $c_{SM}(\mathbf{1}_U)$ is a polynomial of degree $\leq n$ is one variable p. To be consistent with the earlier sections, we introduce a homogenizing variable u, and we write $c_{SM}(\mathbf{1}_U)$ as a binary form of degree n in (p,u).

The CSM class of U carries the same information as the *sectional Euler characteristic*

$$\chi_{\mathrm{sec}}(\mathbf{1}_U) \quad = \quad \sum_{i=0}^{n} \chi(U \cap L_{n-i}) \cdot p^{n-i} u^i .$$

Here L_{n-i} is a generic linear subspace of codimension i in $\mathbb{P}^n$. Indeed, it was proved by Aluffi in [Alu, Theorem 1.1] that $c_{SM}(\mathbf{1}_U)$ is the transform of $\chi_{\mathrm{sec}}(\mathbf{1}_U)$ under a linear involution on binary forms of degree n in (p, u). In fact, our involution in Conjecture 3.15 is nothing but the signed version of the Aluffi's involution. This is explained by the following result.

Theorem 4.6. *Let $X \subset \mathbb{P}^n$ be closed subvariety of dimension d that is not contained in $\mathcal{H}$. If the very affine variety $X\backslash\mathcal{H}$ is schön then, up to signs, the ML bidegree equals the CSM class and the sectional ML degree equals the sectional Euler characteristic. In symbols,*

$$c_{SM}(\mathbf{1}_{X\backslash\mathcal{H}}) = (-1)^{n-d} \cdot B_X(-p, u) \quad \text{and} \quad \chi_{\mathrm{sec}}(\mathbf{1}_{X\backslash\mathcal{H}}) = (-1)^{n-d} \cdot S_X(-p, u).$$

Proof. The first identity is a special case of [Huh1, Theorem 2], here adapted to $\mathbb{P}^n$ minus $n+2$ hyperplanes, and the second identity follows from the first by way of [Alu, Theorem 1.1]. □

To make sense of the statement in Theorem 4.6, we need to recall the definition of *schön*. This term was coined by Tevelev in his study of tropical compactifications [Tevelev]. Let U be an arbitrary closed subvariety of the algebraic torus $(\mathbb{C}^*)^{n+1}$. In our application, $U = X\backslash\mathcal{H}$. We consider the closures $\overline{U}$ of U in various (not necessarily complete) normal toric varieties Y with dense torus $(\mathbb{C}^*)^{n+1}$. The closure $\overline{U}$ is complete if and only if the support of the fan of Y contains the tropicalization of U [Tevelev, Proposition 2.3]. We say that $\overline{U}$ is a *tropical compactification* of U if it is complete and the multiplication map

$$m : (\mathbb{C}^*)^{n+1} \times \overline{U} \longrightarrow Y, \quad (t, x) \longmapsto t \cdot x$$

is flat and surjective. Tropical compactifications exist, and they are obtained from toric varieties Y defined by sufficiently fine fan structures on the tropicalization of U [Tevelev, §2]. The very affine variety U is called *schön* if the multiplication is smooth for some tropical compactification of U. Equivalently, U is schön if the multiplication is smooth for every tropical compactification of U, by Tevelev [Tevelev, Theorem 1.4].

Two classes of schön very affine varieties are of particular interest. The first is the class of complements of essential hyperplane arrangements. The second is the class of nondegenerate hypersurfaces. What we need from the schön hypothesis is the existence of a simple normal crossings compactification which admits sufficiently many differential one-forms which have logarithmic singularities along the boundary. For complements of hyperplane arrangements, such a compactification is provided by the wonderful compactification of De Concini and Procesi [DP]. For nondegenerate hypersurfaces, and more generally for nondegenerate complete intersections, the needed compactification has been constructed by Khovanskii [Hovanskii].

We illustrate this in the setting of likelihood geometry by a d-dimensional linear subspace of $X \subset \mathbb{P}^n$. The intersection of X with distinguished hyperplanes $\mathcal{H}$ of $\mathbb{P}^n$ is an arrangement of $n+2$ hyperplanes in $X \simeq \mathbb{P}^d$, defining a matroid M of rank $d+1$ on $n+2$ elements.

Proposition 4.7. *If X is a linear space of dimension d then the CSM class of $X \backslash \mathcal{H}$ in $\mathbb{P}^n$ is*

$$c_{SM}(\mathbf{1}_{X\backslash\mathcal{H}}) = \sum_{i=0}^{d} (-1)^i h_i u^{d-i} p^{n-d+i}.$$

where the h_i are the signed coefficients of the shifted characteristic polynomial in (15).

Proof. This holds because the recursive formula for a triple of arrangement complements

$$c_{SM}(\mathbf{1}_{U_1}) \ = \ c_{SM}(\mathbf{1}_U - \mathbf{1}_{U_0}) \ = \ c_{SM}(\mathbf{1}_U) - c_{SM}(\mathbf{1}_{U_0}),$$

agrees with the usual deletion-restriction formula [OTBook, Theorem 2.56]:

$$\chi_{M_1}(q+1) = \chi_M(q+1) - \chi_{M_0}(q+1).$$

Here our notation is as in [Huh1, §3]. We now use induction on the number of hyperplanes. □

Proof of Theorem 1.20. The very affine variety $X\backslash\mathcal{H}$ is schön when X is linear. Hence the asserted formula for the ML bidegree of X follows from Theorem 4.6 and Proposition 4.7. □

Rank constraints on matrices are important both in statistics and in algebraic geometry, and they provide a rich source of test cases for the theory developed here. We close our discussion with the enumerative invariants of three hypersurfaces defined by 3×3-determinants. It would be very interesting to compute these formulas for larger determinantal varieties.

Example 4.8. We record the *ML bidegree*, the *CSM class*, the *sectional ML degree*, and the *sectional Euler characteristic* for three singular hypersurfaces seen earlier in this paper. These examples were studied already in [HKS]. The classes we present are elements of $H^*(\mathbb{P}^n_p \times \mathbb{P}^n_u)$ and of $H^*(\mathbb{P}^n_p; \mathbb{Z})$ respectively, and they are written as binary forms in (p, u) as before.

- The 3×3 determinantal hypersurface in $\mathbb{P}^8$ (Example 2.1) has

$$\begin{aligned} B_X(p,u) = {} & 10p^8 + 24p^7u + 33p^6u^2 + 38p^5u^3 + 39p^4u^4 \\ & +33p^3u^5 + 12p^2u^6 + 3pu^7, \end{aligned}$$

$$\begin{aligned} c_{SM}(\mathbf{1}_{X\backslash\mathcal{H}}) &= -11p^8 + 26p^7u - 37p^6u^2 + 44p^5u^3 - 45p^4u^4 \\ &\quad +33p^3u^5 - 12p^2u^6 + 3pu^7, \\ S_X(p,u) &= 11p^8 + 182p^7u + 436p^6u^2 + 518p^5u^3 + 351p^4u^4 \\ &\quad +138p^3u^5 + 30p^2u^6 + 3pu^7, \\ \chi_{\sec}(\mathbf{1}_{X\backslash\mathcal{H}}) &= -11p^8 + 200p^7u - 470p^6u^2 + 542p^5u^3 - 357p^4u^4 \\ &\quad +138p^3u^5 - 30p^2u^6 + 3pu^7. \end{aligned}$$

- The 3×3 symmetric determinantal hypersurface in $\mathbb{P}^5$ (Example 2.7) has

$$\begin{aligned} B_X(p,u) &= 6p^5 + 12p^4u + 15p^3u^2 + 12p^2u^3 + 3pu^4, \\ c_{SM}(\mathbf{1}_{X\backslash\mathcal{H}}) &= 7p^5 - 14p^4u + 19p^3u^2 - 12p^2u^3 + 3pu^4, \\ S_X(p,u) &= 6p^5 + 42p^4u + 48p^3u^2 + 21p^2u^3 + 3pu^4, \\ \chi_{\sec}(\mathbf{1}_{X\backslash\mathcal{H}}) &= 7p^5 - 48p^4u + 52p^3u^2 - 21p^2u^3 + 3pu^4. \end{aligned}$$

- The secant variety of the rational normal curve in $\mathbb{P}^4$ (Example 3.18) has

$$\begin{aligned} B_X(p,u) &= 12p^4 + 15p^3u + 12p^2u^2 + 3pu^3, \\ c_{SM}(\mathbf{1}_{X\backslash\mathcal{H}}) &= -13p^4 + 19p^3u - 12p^2u^2 + 3pu^3, \\ S_X(p,u) &= 12p^4 + 30p^3u + 18p^2u^2 + 3pu^3, \\ \chi_{\sec}(\mathbf{1}_{X\backslash\mathcal{H}}) &= -13p^4 + 34p^3u - 18p^2u^2 + 3pu^3. \end{aligned}$$

In all known examples, the coefficients of $B_X(p,u)$ are less than or equal to the absolute value of the corresponding coefficients of $c_{SM}(\mathbf{1}_{X\backslash\mathcal{H}})$, and similarly for $S_X(p,u)$ and $\chi_{\sec}(\mathbf{1}_{X\backslash\mathcal{H}})$. That this inequality holds for the first coefficient is Conjecture 1.8 which relates the ML degree of a singular X to the signed Euler characteristic of the very affine variety $X\backslash\mathcal{H}$. ◇

Acknowledgements We thank Paolo Aluffi and Sam Payne for helpful communications, and the Mathematics Department at KAIST, Daejeon, for hosting both authors in May 2013. Bernd Sturmfels was supported by NSF (DMS-0968882) and DARPA (HR0011-12-1-0011).

References

[ARSZ] E. Allmann, J. Rhodes, B. Sturmfels, P. Zwiernik, Tensors of nonnegative rank two, in *Linear Algebra and its Applications*. Special Issue on Statistics. http://www.sciencedirect.com/science/article/pii/S0024379513006812

[AluJSC] P. Aluffi, Computing characteristic classes of projective schemes. J. Symb. Comput. **35**, 3–19 (2003)

[AluLectures] P. Aluffi, Characteristic classes of singular varieties, in *Topics in Cohomological Studies of Algebraic Varieties*. Trends in Mathematics (Birkhäuser, Basel, 2005), pp. 1–32

[Alu] P. Aluffi, Euler characteristics of general linear sections and polynomial Chern classes. Rend. Circ. Mat. Palermo. **62**, 3–26 (2013)

[AN] S. Amari, H. Nagaoka, in *Methods of Information Geometry*. Translations of Mathematical Monographs, vol. 191 (American Mathematical Society, Providence, 2000)

[Bertini] D.J. Bates, J.D. Hauenstein, A.J. Sommese, C.W. Wampler, *Bertini: Software for Numerical Algebraic Geometry* (2006). www.nd.edu/~sommese/bertini

[BHSW] D.J. Bates, J.D. Hauenstein, A.J. Sommese, C.W. Wampler, *Numerically Solving Polynomial Systems with Bertini* (Society for Industrial and Applied Mathematics, Philelphia, 2013). http://www.ec-securehost.com/SIAM/SE25.html

[BFH] Y. Bishop, S. Fienberg, P. Holland, *Discrete Multivariate Analysis: Theory and Practice* (Springer, New York, 1975)

[BoydVan] S. Boyd, L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004)

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

[CDFV] D. Cohen, G. Denham, M. Falk, A. Varchenko, Critical points and resonance of hyperplane arrangements. Can. J. Math. **63**, 1038–1057 (2011)

[DP] C. De Concini, C. Procesi, Wonderful models of subspace arrangements. Selecta Math. New Ser. **1**, 459–494 (1995)

[Denham-Garrousian-Schulze] G. Denham, M. Garrousian, M. Schulze, A geometric deletion-restriction formula. Adv. Math. **230**, 1979–1994 (2012)

[DR] J. Draisma, J. Rodriguez, Maximum likelihood duality for determinantal varieties. Int. Math. Res. Not. [arXiv:1211.3196]. http://imrn.oxfordjournals.org/content/early/2013/07/03/imrn.rnt128.full.pdf

[LiAS] M. Drton, B. Sturmfels, S. Sullivant, in *Lectures on Algebraic Statistics*. Oberwolfach Seminars, vol. 39 (Birkhäuser, Basel, 2009)

[Franecki-Kapranov] J. Franecki, M. Kapranov, The Gauss map and a noncompact Riemann-Roch formula for constructible sheaves on semiabelian varieties. Duke Math. J. **104**, 171–180 (2000)

[Fulton] W. Fulton, in *Intersection Theory*, 2nd edn. Ergebnisse der Mathematik und ihrer Grenzgebiete. A Series of Modern Surveys in Mathematics, vol. 2 (Springer, Berlin, 1998)

[Gabber-Loeser] O. Gabber, F. Loeser, Faisceaux pervers l-adiques sur un tore. Duke Math. J. **83**, 501–606 (1996)

[GKZ] I.M. Gel'fand, M. Kapranov, A. Zelevinsky, *Discriminants, Resultants, and Multidimensional Determinants* (Birkhäuser, Boston, 1994)

[GR] E. Gross, J. Rodriguez, *Maximum likelihood geometry in the presence of data zeros*, http://front.math.ucdavis.edu/1310.4197

[HRS] J. Hauenstein, J. Rodriguez, B. Sturmfels, Maximum likelihood for matrices with rank constraints. J. Algebr. Stat. (to appear) [arXiv:1210.0198]

[HKS] S. Hoşten, A. Khetan, B. Sturmfels, Solving the likelihood equations. Found. Comput. Math. **5**, 389–407 (2005)

[Hovanskii] A. Hovanskiĭ, Newton polyhedra and toroidal varieties. Akademija Nauk SSSR. Funkcional'nyi Analiz i ego Priloženija **11**, 56–64 (1977)

[Huh1] J. Huh, The maximum likelihood degree of a very affine variety. Compositio Math. **149**, 1245–1266 (2013)

[Huh0] J. Huh, *h-vectors of matroids and logarithmic concavity*. Preprint. http://arxiv.org/abs/1201.2915. [arXiv:1201.2915]

[Huh2] J. Huh, *Varieties with maximum likelihood degree one*. J. Algeb. Stat. (to appear). [arXiv:1301.2732]

[Kapranov] M. Kapranov, A characterization of A-discriminantal hypersurfaces in terms of the logarithmic Gauss map. Math. Ann. **290**, 277–285 (1991)

[KRS] K. Kubjas, E. Robeva, B. Sturmfels, Fixed points of the EM algorithm and nonnegative rank boundaries, http://arxiv.org/abs/1312.5634
(in preparation)

[Land] J.M. Landsberg, *Tensors: Geometry and Applications*. Graduate Studies in Mathematics, vol. 128 (American Mathematical Society, Providence, 2012)

[LM] J.M. Landsberg, L. Manivel, On ideals of secant varieties of Segre varieties. Found. Comput. Math. **4**, 397–422 (2004)

[Lau] S. Lauritzen, *Graphical Models* (Oxford University Press, Oxford, 1996)

[ch3:MS] E. Miller, B. Sturmfels, in *Combinatorial Commutative Algebra*. Graduate Texts in Mathematics, vol. 227 (Springer, New York, 2004)

[MSS] D. Mond, J. Smith, D. van Straten, Stochastic factorizations, sandwiched simplices and the topology of the space of explanations. R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci. **459**, 2821–2845 (2003)

[OTBook] P. Orlik, H. Terao, in *Arrangements of Hyperplanes*. Grundlehren der Mathematischen Wissenschaften, vol. 300 (Springer, Berlin, 1992)

[Orlik-Terao] P. Orlik, H. Terao, The number of critical points of a product of powers of linear functions. Inventiones Math. **120**, 1–14 (1995)

[PS] L. Pachter, B. Sturmfels, *Algebraic Statistics for Computational Biology* (Cambridge University Press, Cambridge, 2005)

[Rai] C. Raicu, Secant varieties of Segre–Veronese varieties. Algebra Number Theory **6**, 1817–1868 (2012)

[Rapallo] F. Rapallo, Markov bases and structural zeros. J. Symb. Comput. **41**, 164–172 (2006)

[SSV] R. Sanyal, B. Sturmfels, C. Vinzant, The entropic discriminant. Adv. Math. **244**, 678–707 (2013)

[Terao] H. Terao, Generalized exponents of a free arrangement of hyperplanes and the Shepherd-Todd-Brieskorn formula. Invent. Math. **63**, 159–179 (1981)

[Tevelev] J. Tevelev, Compactifications of subvarieties of tori. Am. J. Math. **129**, 1087–1104 (2007)

[Uhl] C. Uhler, Geometry of maximum likelihood estimation in Gaussian graphical models. Ann. Stat. **40**, 238–261 (2012)

[Varchenko] A. Varchenko, Critical points of the product of powers of linear functions and families of bases of singular vectors. Compositio Math. **97**, 385–401 (1995)

[Wat] S. Watanabe, in *Algebraic Geometry and Statistical Learning Theory*. Monographs on Applied and Computational Mathematics, vol. 25 (Cambridge University Press, Cambridge, 2009)