

Likelihood Correspondence of Toric Statistical Models

International Symposium on Symbolic and Algebraic Computation 2026

David Barnhill¹, John Cobb², Matt Faust³

¹ U.S. Naval Academy

² Auburn University

³ Michigan State University

Outline

- 1 Toric Models
- 2 Likelihood Geometry and Correspondence
- 3 Computational Results

Toric Models

Consider a process resulting in outcomes, each with probability p_i of occurring.

Definition

A *discrete statistical model* is a variety, $\mathcal{M}$, restricted to the probability simplex

$$\Delta^n = \left\{ p \in \mathbb{R}_{\geq 0}^{n+1} \mid \sum_{i=0}^n p_i = 1 \right\}.$$

The estimated probability that we observe some data $u = (u_0, \dots, u_n) \in \mathbb{N}^{n+1}$ is given by the likelihood function

$$\ell_u(p) = \prod_{i=0}^n p_i^{u_i}.$$

Goal: Find the parameters most likely to produce the observed data after restricting to a model of interest $\mathcal{M}$.

Toric Models

Toric models represent a large class statistical models which include

Toric $\supset$ (Hierarchical) Log-linear $\supset$ Undirected Graphical $\supset$ Independence.

Definition

A model $\mathcal{M}$ is *toric* if $I_{\mathcal{M}}$ is binomial and prime. Fixing a matrix $A = (a_{ij}) \in \mathbb{Z}^{k \times r}$ such the column sums of A are equal, the corresponding *toric ideal*, I_A , is generated by binomials $p^v - p^w$ with $v, w \in \mathbb{N}^{n+1}$ such that $Av = Aw$ and $p^v = p_0^{v_0} p_1^{v_1} \cdots p_n^{v_n}$.

Letting A_i represent the column vectors of A and reparameterizing p in terms of $\theta = (\theta_1, \dots, \theta_k)$ with u being our data vector, $\ell(p)$ becomes

$$\ell(p) = \ell(\theta) = \prod_{i=1}^k (\theta^{A_i})^{u_i} = \theta^{Au}$$

Example.

I went out and asked 100 people if

- 1 they liked this talk,
- 2 they think cilantro tastes like soap.

I summarize the results in the *contingency table*:

$$\begin{array}{cc} & \begin{array}{cc} \text{liked talk} & \text{disliked talk} \end{array} \\ \begin{array}{c} \text{not soapy} \\ \text{soapy} \end{array} & \begin{pmatrix} 73 & 11 \\ 7 & 9 \end{pmatrix} \end{array} \in \mathbb{R}^4$$

.

Example.

I went out and asked 100 people if

- 1 they liked this talk,
- 2 they think cilantro tastes like soap.

I summarize the results in the *contingency table*: $\begin{pmatrix} 73 & 11 \\ 7 & 9 \end{pmatrix} \in \mathbb{R}^4$

Given this sample, what are the empirical probabilities $\begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$ of the occurrence of each event if one person is selected at random?

$$\begin{pmatrix} .73 & .11 \\ .07 & .09 \end{pmatrix}$$

"Cilantro is not soapy and this talk is awesome!"

Example.

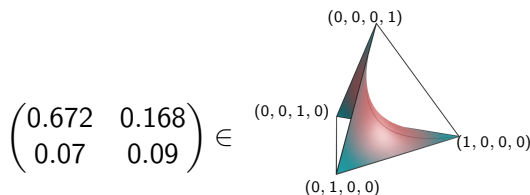
I went out and asked 100 people if

- 1 they liked this talk,
- 2 they think cilantro tastes like soap.

Suppose I know that the outcomes from each of these questions are statistically independent.
This is equivalent to:

$$p_{11}p_{22} - p_{21}p_{12} = 0.$$

Goal: Maximize $\ell(p) = p_{11}^3 p_{12}^{11} p_{21}^7 p_{22}^9$ over $V(p_{11}p_{22} - p_{21}p_{12})$.



Example: From our first 2×2 contingency table we have

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \quad u = \begin{pmatrix} 73 \\ 11 \\ 7 \\ 9 \end{pmatrix} \quad Au = (\underbrace{84 \quad 16 \quad 80 \quad 20}_{\text{sufficient statistics}})$$

$$p_{11} = \theta_1 \theta_3 \quad p_{12} = \theta_1 \theta_4 \quad p_{21} = \theta_2 \theta_3 \quad p_{22} = \theta_2 \theta_4$$

Likelihood Geometry

We make all questions easier by making the replacements

semi-algebraic sets	→	algebraic sets
real numbers	→	complex numbers
affine space	→	projective space

Probabilities, p , now live in $\mathbb{P}_p^n$ and data, u , live in the *data* space $\mathbb{P}_u^n$. The likelihood function then becomes

$$\ell_u(p) = \frac{p_0^{u_0} p_1^{u_1} \cdots p_n^{u_n}}{\underbrace{(p_0 + p_1 + \cdots + p_n)}_{p_+} \underbrace{^{u_0 + u_1 + \cdots + u_n}}_{u_+}}$$

Goal: Maximize the restriction of $\ell_u(p) = (\prod p_i^{u_i} / p_+^{u_+})$ on a variety $\mathcal{M} \subseteq \mathbb{P}_p^n$.

- We replace ℓ_u with the log-likelihood function $\log \ell_u$.
- $\log \ell_u$ has singularities on the $n + 2$ hyperplanes $V(p_0), \dots, V(p_n)$ and $V(p_+)$. We denote this hyperplane arrangement $\mathcal{H}$.

Example: Let $\mathcal{M} = \mathbb{P}_p^n$ and fix u . The maxima of ℓ_u are in the critical set $V(d \log \ell_u)$.

$$\frac{\partial}{\partial p} \log \ell_u = \left(\frac{u_0}{p_0}, \frac{u_1}{p_1}, \dots, \frac{u_n}{p_n} \right) - \left(\frac{u_+}{p_+}, \dots, \frac{u_+}{p_+} \right)$$

is zero on the single point $p = u \in \mathbb{P}_p^n \setminus \mathcal{H}$.

Likelihood Correspondence

Collecting these critical sets together into a family over $\mathbb{P}_u^n$ of possible counts gives us the likelihood correspondence.

Definition

The *likelihood correspondence* is the universal family of all complex critical points of the likelihood function.

$$L_{\mathcal{M}} = \left\{ (p, u) \in (\mathcal{M}_{\text{reg}} \setminus \mathcal{H}) \times \mathbb{P}_u^n : d \log \ell_u|_{\mathcal{M}}(p) = 0 \right\}.$$

The vanishing ideal of $L_{\mathcal{M}}$ is called the *likelihood ideal*.

Example: Let $\mathcal{M} = \mathbb{P}_p^n$. The likelihood correspondence is given by the diagonal $\Delta = \{(u, u) \in \mathbb{P}_p^n \times \mathbb{P}_u^n\}$, which has the likelihood ideal $I_2 \begin{pmatrix} u_0 & \cdots & u_n \\ p_0 & \cdots & p_n \end{pmatrix}$ inside the coordinate ring $\mathbb{C}[p_0, \dots, p_n, u_0, \dots, u_n]$ of $\mathbb{P}_p^n \times \mathbb{P}_u^n$.

Likelihood Correspondence: What's Known?

- In 2012, June Huh proved that $L_{\mathcal{M}}$ is an irreducible variety of dimension n and the map onto the second factor is k -to-1, where k is the maximum likelihood degree (ML degree).
- Much focus within algebraic statistics has been given to computing properties of the likelihood ideal such as the ML degree. For example, models with ML degree 1 have maximum likelihood estimators that are a rational function of the input data.
- It is generally unknown which properties of $\mathcal{M}$ induce nice properties for $L_{\mathcal{M}}$, which stem from a lack of tools to explicitly construct the likelihood ideal.
- Kahle, et al. (2026) study the likelihood correspondence involving strictly normal crossings (SNC), however, statistical models do not satisfy the SNC hypothesis.

Question: What are generators for the likelihood ideal of $\mathcal{M}$? What properties does it possess?

Likelihood Correspondence: Lagrange Multipliers

The only general algorithm for computing the likelihood ideal requires computing minors of size $\text{codim}(\mathcal{M}) + 1$ of the matrix:

$$\begin{pmatrix} u_0 & u_1 & \cdots & u_n \\ p_0 & p_1 & \cdots & p_n \\ p_0 \frac{\partial f_0}{\partial p_0} & p_1 \frac{\partial f_0}{\partial p_1} & \cdots & p_n \frac{\partial f_0}{\partial p_n} \\ \vdots & \vdots & \ddots & \vdots \\ p_0 \frac{\partial f_r}{\partial p_0} & p_1 \frac{\partial f_r}{\partial p_1} & \cdots & p_n \frac{\partial f_r}{\partial p_n} \end{pmatrix} \text{ where } I_{\mathcal{M}} = \langle f_0, \dots, f_r \rangle.$$

Problems

- The number of minors can be enormous requiring saturation by $\mathcal{H}$ and singular locus of $\mathcal{M}$.
- Gives little theoretical insight to $L_{\mathcal{M}}$.

Likelihood Correspondence: Lagrange Multipliers

Example: Recall that $\mathcal{M} = V(p_{11}p_{22} - p_{12}p_{21})$ corresponds to the complete independence model on two binary variables. The likelihood correspondence is a 3-dimensional subvariety $L_{\mathcal{M}} \subset \mathbb{P}_p^3 \times \mathbb{P}_u^3$ and can be computed using Lagrangian multipliers with the following Jacobian matrix:

$$J = \begin{pmatrix} u_{11} & u_{12} & u_{21} & u_{22} \\ p_{11} & p_{12} & p_{21} & p_{22} \\ p_{11}p_{22} & -p_{12}p_{21} & -p_{21}p_{12} & p_{11}p_{22} \end{pmatrix}.$$

The likelihood ideal of $\mathcal{M}$ inside the coordinate ring $\mathbb{C}[p_{11}, p_{12}, p_{21}, p_{22}, u_{11}, u_{12}, u_{21}, u_{22}]$ is

$$(\langle p_{11}p_{22} - p_{12}p_{21} \rangle + I_2(J)) : \langle p_{11}p_{12}p_{21}p_{22}p_{++} \rangle^{\infty}.$$

This computation takes roughly two minutes in Macaulay2.

Computing the Likelihood Correspondence

Theorem 1 (B, Cobb, Faust, 2023)

The likelihood ideal for complete and joint independence models on $c > 1$ discrete random variables is given explicitly by the 2×2 minors of at most c matrices. These form a Gröbner basis.

Example. The likelihood ideal of $\mathcal{M} = V(p_{11}p_{22} - p_{12}p_{21})$ is:

$$I_2 \begin{pmatrix} p_{11} & p_{12} & u_{1+} \\ p_{21} & p_{22} & u_{2+} \end{pmatrix} + I_2 \begin{pmatrix} p_{11} & p_{21} & u_{+1} \\ p_{12} & p_{22} & u_{+2} \end{pmatrix}.$$

Likelihood Correspondence

Example: Consider discrete random variables X_1 , X_2 , and X_3 corresponding to the space of $2 \times 3 \times 4$ contingency tables. The likelihood ideal of the joint independence model $X_1 \perp \{X_2, X_3\}$ is given by 2-minors of each matrix $M_{\{X_i\}}$.

$$M_{\{X_1\}} = \begin{pmatrix} p_{111} & p_{112} & p_{113} & p_{114} & p_{121} & p_{122} & p_{123} & p_{124} & p_{131} & p_{132} & p_{133} & p_{134} & u_{1++} \\ p_{211} & p_{212} & p_{213} & p_{214} & p_{221} & p_{222} & p_{223} & p_{224} & p_{231} & p_{232} & p_{233} & p_{234} & u_{2++} \end{pmatrix}$$

$$M_{\{X_2, X_3\}} = \begin{pmatrix} p_{111} & p_{211} & u_{+11} \\ p_{112} & p_{212} & u_{+12} \\ p_{113} & p_{213} & u_{+13} \\ p_{114} & p_{214} & u_{+14} \\ p_{121} & p_{221} & u_{+21} \\ p_{122} & p_{222} & u_{+22} \\ p_{123} & p_{223} & u_{+23} \\ p_{124} & p_{224} & u_{+24} \\ p_{131} & p_{231} & u_{+31} \\ p_{132} & p_{232} & u_{+32} \\ p_{133} & p_{233} & u_{+33} \\ p_{134} & p_{234} & u_{+34} \end{pmatrix} \quad (\textit{sufficient statistics of } \mathcal{M})$$

Computing the Likelihood Correspondence

Theorem 2 (B., Cobb, Faust, 2023)

Let $\mathcal{M} \subseteq \mathbb{P}_p^n$ be a toric model defined by a matrix A . Let

$$M = \begin{pmatrix} u_0 & p_0 \\ \vdots & \vdots \\ u_n & p_n \end{pmatrix} \text{ then,}$$

$$I_{\mathcal{M}} + I_2(AM) : \langle p_0 \cdots p_n p_+ \rangle^\infty$$

defines the likelihood ideal of $\mathcal{M}$.

* Do we always need to saturate?

Likelihood Correspondence

The following table organizes the computation times for 2-way complete independence models using minors (Theorem 1, I) and the toric method (Theorem 2, T).

$i \times j$ computation times (s)						
	2	3	4	5	6	Method
2	.0042050	.0026273	.0030742	.0037202	.0042469	I
	.0126488	.0121614	.019614	.0155284	.0170114	T
3		.0030728	.0038790	.0052260	.0058382	I
		.0142195	.0172784	.0271773	.0276863	T
4			.0049143	.0064623	.0142544	I
			.0198774	.0350445	.0484354	T
5				.0134698	.0167110	I
				.0443950	.0581353	T
6					.0233014	I
					.0864135	T

For comparison, the Lagrange method took 2 minutes for the 2×2 case, with higher not terminating even after several days.

Thanks!