

Maximum information divergence from linear and toric models

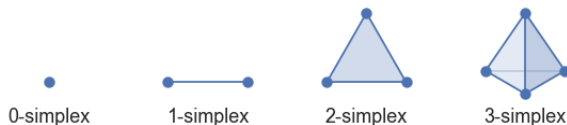
Yulia Alexandr (UCLA)
joint work with Serkan Hoşten

Caltech Information, Geometry & Physics Seminar
May 6, 2026

Notation

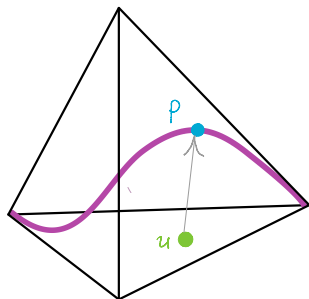
- A *probability simplex* is defined as

$$\Delta_{n-1} = \{(p_1, \dots, p_n) : p_1 + \dots + p_n = 1, p_i \geq 0 \text{ for } i \in [n]\}.$$



- A *statistical model* is a subset of Δ_{n-1} .
- A *variety* is the set of solutions to a system of polynomial equations.
- An *algebraic statistical model* is a subset $\mathcal{M} = \mathcal{V} \cap \Delta_{n-1}$ for some variety $\mathcal{V} \subseteq \mathbb{C}^n$.

The log-likelihood function



Let $\mathcal{M} \subseteq \Delta_{n-1}$ be a statistical model.

For an empirical data point $u = (u_1, \dots, u_n) \in \Delta_{n-1}$, the *log-likelihood function* with respect to u assuming distribution $p = (p_1, \dots, p_n) \in \mathcal{M}$ is

$$\ell_u(p) = u_1 \log p_1 + u_2 \log p_2 + \dots + u_n \log p_n.$$

Maximum likelihood estimation

Fix an algebraic statistical model $\mathcal{M} \subseteq \Delta_{n-1}$

- 1 The maximum likelihood estimation problem (MLE):

Given a sampled empirical distribution $u \in \Delta_{n-1}$, which point $p \in \mathcal{M}$ did it most likely come from? In other words, we wish to maximize $\ell_u(p)$ over all points $p \in \mathcal{M}$.

Maximum likelihood estimation

Fix an algebraic statistical model $\mathcal{M} \subseteq \Delta_{n-1}$

- 1 The maximum likelihood estimation problem (MLE):

Given a sampled empirical distribution $u \in \Delta_{n-1}$, which point $p \in \mathcal{M}$ did it most likely come from? In other words, we wish to maximize $\ell_u(p)$ over all points $p \in \mathcal{M}$.

- 2 Computing logarithmic Voronoi cells:

Given a point $q \in \mathcal{M}$, what is the set of all points $u \in \Delta_{n-1}$ that have q as a global maximum when optimizing the function $\ell_u(p)$ over \mathcal{M} ?

The set of all such elements $u \in \Delta_{n-1}$ is the *logarithmic Voronoi cell* at q .

Linear and toric models

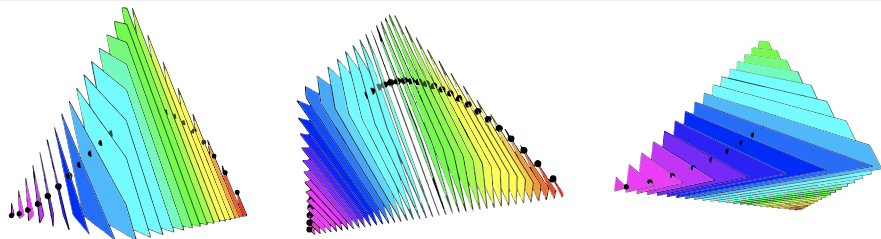
Theorem (A.-Heaton)

If \mathcal{M} is a linear model or a toric model, the logarithmic Voronoi cell at any point $p \in \mathcal{M}$ is a polytope.

We will denote the *logarithmic Voronoi polytope* at $p \in \mathcal{M}$ by Q_p .

Example (The twisted cubic.)

The curve is given by $\theta \mapsto (\theta^3, 3\theta^2(1 - \theta), 3\theta(1 - \theta)^2, (1 - \theta)^3)$.



Maximum KL-divergence

For two distributions $p, q \in \Delta_{n-1}$, the *Kullback-Leibler (KL) divergence* is

$$D(p||q) = \sum_{i=1}^n p_i \log \left(\frac{p_i}{q_i} \right).$$

For fixed $u \in \Delta_{n-1}$ maximizing $\ell_u(p) =$ minimizing $D(u||p)$ over $p \in \mathcal{M}$.

What is the maximum and the maximizers of $\max_{u \in \Delta_{n-1}} \min_{p \in \mathcal{M}} D(u||p)$?

In other words, which point in the simplex is the farthest to its MLE?

- problem formulated by Ay '02 when \mathcal{M} is a discrete exponential family
- many information-theoretic results by Ay, Matúš, Montúfar, Rauh, etc.
- bio-neural networks develop in such a way to maximize the mutual information between the input and output of each layer.

Maximum KL divergence

In this talk, \mathcal{M} will be a linear or a toric model.

- Let $D_{\mathcal{M}}(u) := \min_{p \in \mathcal{M}} D(u||p)$ be the **divergence** from u to \mathcal{M} .
- We study the **maximum divergence** $D(\mathcal{M}) := \max_{u \in \Delta_{n-1}} D_{\mathcal{M}}(u)$ and the points which achieve $D(\mathcal{M})$.
- For fixed $q \in \mathcal{M}$, the function $D(u||q)$ is strictly convex in u over Δ_{n-1} .
- Hence, the maximum of $D_{\mathcal{M}}(u)$ restricted to the logarithmic Voronoi polytope Q_q is achieved at a vertex of Q_q .

Main idea:

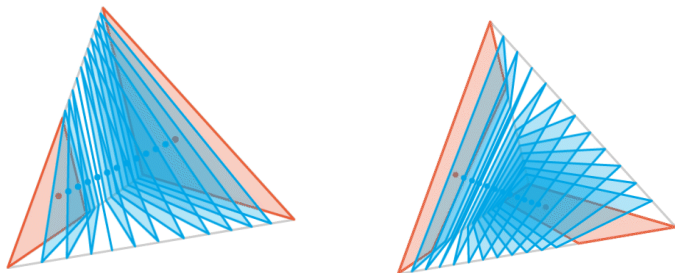
In order to compute $D(\mathcal{M})$ and its maximizers, we will systematically keep track of the vertices of Q_p as we vary p over the model \mathcal{M} .

Linear models

A *discrete linear model* is defined parametrically by linear polynomials in the parameters $\theta_1, \dots, \theta_d$.

Theorem (A.)

Logarithmic Voronoi cells of all interior points in a linear model have the same combinatorial type.



Maximum divergence to linear models

- Let $\mathcal{M} = \{c - Bx : x \in \Theta\}$ where B is $n \times d$ whose rows sum to 0 and the entries of c sum to 1.
- By a *co-circuit* of B we mean a nonzero $z \in \mathbb{R}^n$ of minimal support so that $z^T B = 0$.
- The vertices of the logarithmic Voronoi polytope at $q \in \mathcal{M}$ are in bijection with the positive co-circuits z of B such that $\sum_{i=1}^n z_i q_i = 1$:
 $V_z(q) = (z_1 q_1, \dots, z_n q_n)$.
- For a fixed co-circuit z of B , the information divergence $D(V_z(q), q) = \sum_{i=1}^n z_i \log(z_i) q_i$ is linear in $q \in \mathcal{M}$.

Theorem (A.-Hoşten)

The maximum divergence of a linear model \mathcal{M} is achieved at a vertex of the logarithmic Voronoi polytope Q_q where q itself is a vertex of \mathcal{M} .

Toric models (aka exponential families)

- Consider a *discrete* exponential family $\mathcal{E}_{\omega,A}$ in Δ_{n-1} .
- The matrix A has integer entries.
- Let

$$A = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_n \end{pmatrix}$$

with $a_i \in \mathbb{N}^d$ and $\text{rank}(A) = d + 1$.

- $X_{\omega,A}$ is the Zariski closure of $(\mathbb{C}^*)^d$ under the monomial map.

$$z \mapsto (\omega_1 z^{a_1}, \omega_2 z^{a_2}, \dots, \omega_n z^{a_n}).$$

- The associated *toric model* is $\mathcal{M}_A = X_{\omega,A} \cap \Delta_{n-1}$. It is equal to $\overline{\mathcal{E}_{\omega,A}}$.
- Let $q \in \mathcal{M}_A$. The logarithmic Voronoi polytope at q is of the form

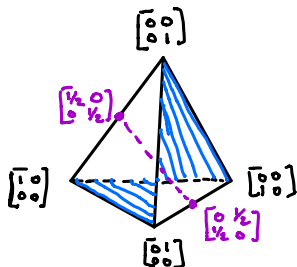
$$Q_b = \{p \in \Delta_{n-1} : Ap = Aq = b\}.$$

Critical points

A vertex v of Q_b is *complementary* if there exists a face F of Q_b such that $\text{supp}(F) = [n] \setminus \text{supp}(v)$.

Theorem (Ay '02, Matúš '07, A.-Hoşten '25.)

Every critical point p of $D_{\mathcal{M}_A}$ is a complementary vertex of Q_q where q is the MLE of p . A complementary vertex v of Q_q with the complementary face F is a critical point if and only if the line passing through v and q intersects the relative interior of F .



The chamber complex

Let's study these vertices systematically!

Let $\text{conv}(A) = \text{conv}(a_1, \dots, a_n)$. The *chamber complex* \mathcal{C}_A of $\text{conv}(A)$ is the common refinement of all triangulations of $\text{conv}(A)$.

Theorem (A.-Hoşten)

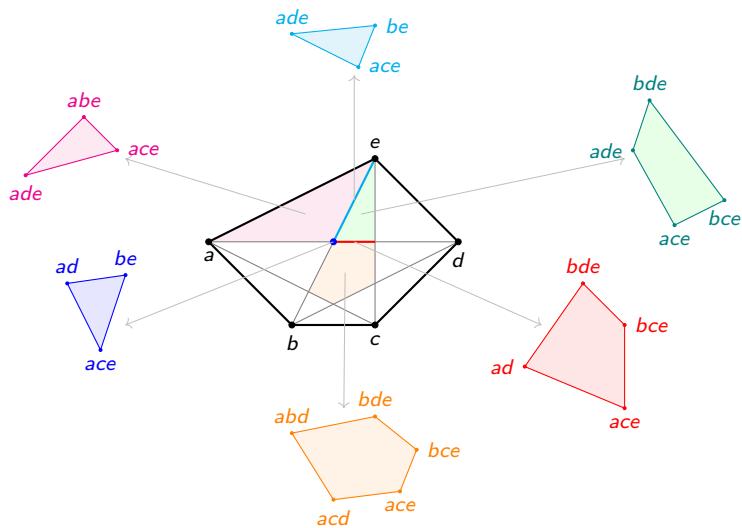
Fix a chamber $C \in \mathcal{C}_A$. As b varies in the relative interior of C , the support of each face of Q_b as well as the combinatorial type of Q_b does not change.

Let

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 2 \\ 1 & 0 & 0 & 1 & 2 \end{pmatrix},$$

and denote the columns of A by a, b, c, d , and e .

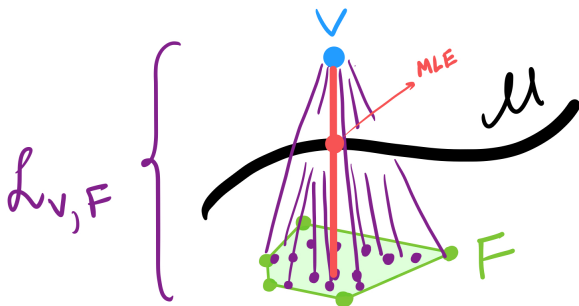
The chamber complex



Given a logarithmic Voronoi polytope Q_b where $b \in \text{conv}(A)$, we need to identify complementary vertices of Q_b and decide whether any of these vertices are critical points. These are potential maximizers of $D_{\mathcal{M}_A}$.

Proposition (A.-Hoşten)

Let v be a complementary vertex of Q_b with the complementary face F . Let $\mathcal{L}_{v,F}$ be the collection of the lines passing through v and each point on F . Then v is a critical point if and only if $\mathcal{L}_{v,F}$ intersects \mathcal{M}_A .



For each $Q_b \dots$

To check whether a complementary vertex v of a *fixed* logarithmic Voronoi polytope Q_b is a critical point:

- 1 Let F be the complementary face of dimension k and assume v_1, \dots, v_{k+1} are vertices of F that are affinely independent.
- 2 Then $\overline{\mathcal{L}_{v,F}}$ is the image of the map

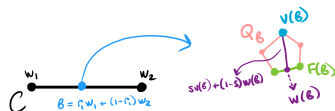
$$(s, t_1, \dots, t_{k+1}) \mapsto sv + (1-s)(t_1v_1 + \dots + t_{k+1}v_{k+1})$$

where $t_1 + \dots + t_{k+1} = 1$

- 3 To intersect $\overline{\mathcal{L}_{v,F}}$ with \mathcal{M}_A plug in the image into the binomial equations defining the toric variety X_A .
- 4 Note $\overline{\mathcal{L}_{v,F}}$ and X_A intersect in finitely many points.
- 5 Compute them using numerical algebraic geometry.
- 6 Checks if this finite set contains a point with positive coordinates.

General algorithm

- ★ Compute the equations of X_A .
- ★ Compute the chamber complex \mathcal{C}_A .
- ★ For each chamber C in \mathcal{C}_A do:



- 1 Let w_1, \dots, w_m be the vertices of C , so $b = \sum_i r_i w_i$.
 - 2 Let $(v(b), F(b))$ be a complementary vertex-face pair in Q_b .
The coordinates of $v(b)$ and vertices of $F(b)$ are linear functions of r_i .
 - 3 Parametrize a general point $w(b)$ on $F(b)$ via $w(b) = \sum t_i v_i(b)$.
 - 4 The line segment between $v(b)$ and $w(b)$ is parametrized by $sv(b) + (1-s)w(b)$ where $0 \leq s \leq 1$.
 - 5 Substitute the coordinates of $sv(b) + (1-s)w(b)$ into the equations of X_A , check whether this system of equations has positive solutions.
- ★ Locate the global maximizer(s) among these local maximizers contributed by each chamber C .

Reducing chambers

Proposition (A.-Hoşten)

Fix a chamber C and $b \in C^\circ$.

- If C has dimension k where $k + 1 > n/2$, then Q_b does not contain complementary vertices.
- If (v, F) is a complementary vertex-face pair, where both v and F are contained in the same facet F' of Q_b , then v is not a critical point.
- If no two vertices of Q_b have disjoint supports, then the same is true for any chamber C' containing C .
- Suppose $\text{conv}(A)$ is a simplicial polytope where each column of A is a vertex. Suppose C intersects the boundary as well as the interior of $\text{conv}(A)$. Then Q_b does not contain complementary vertices.

Also, we can (and should) employ *symmetries* to deal with less chambers!

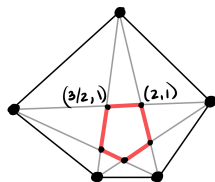
Example (pentagon)

The toric variety X_A is defined by the equations

$$p_2^2 p_4^2 - p_3^3 p_5 = p_1 p_3^3 - p_2^3 p_4 = p_1 p_4 - p_2 p_5 = 0.$$

Only need to consider the boundary edges of the pentagonal chamber!

- The edge between $(3/2, 1)$ and $(2, 1)$ has one complementary vertex/face pair (v, F) .
- (v, F) have supports $\{a, d\}$ and $\{b, c, e\}$.
- The parametrization of the line segment between v and F is given by
 $(r, s) \mapsto \left(s\left(\frac{1}{6}r + \frac{1}{3}\right), (1-s)\frac{r}{2}, (1-s)\frac{1-r}{2}, s\left(-\frac{1}{6}r + \frac{2}{3}\right), \frac{1-s}{2} \right)$,
where we parametrized b on the edge by $r(3/2, 1) + (1-r)(2, 1)$.
- Plugging it into the equations, we get:



Pentagon continued

$$s^2 r^2 + 7s^2 r - 8s^2 - 18sr + 9r = 0$$

$$197s^4 r - 194s^4 - 1401s^3 r - 3sr^3 + 1014s^3 + 4398s^2 r + 246sr^2 - 2094s^2 \\ - 5837sr - 81r^2 + 2s + 2349r = 0$$

$$885s^4 - 31312s^3 r - 294sr^3 + 32392s^3 + 179435s^2 r + 17016sr^2 - 117350s^2 \\ - 295438sr - 6165r^2 + 2560s + 129141r - 591 = 0.$$

- This is a zero-dimensional system that has 11 solutions (Bertini). Four are complex and seven are real.
- There is a unique real solution where $0 < r, s < 1$, namely

$$r = 0.4702953126494577 \text{ and } s = 0.4106301713351522.$$

- The corresponding KL-divergence at the vertices v and F are 0.890062259952966 and 0.528701425022976.

Pentagon continued

For each of the remaining four edges we also get a pair of critical vertices with corresponding KL-divergences

0.729916767214609 and 0.657681783609608

0.736523721240758 and 0.651574202843057

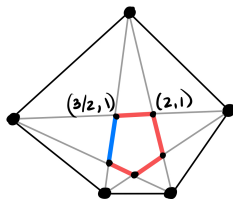
0.927851227501820 and 0.503192212618303

0.856820834934792 and 0.552532602066626.

The global maximizer is the vertex

$$v = (0, 0.6722451790633609, 0, 0, 0.3277548209366391)$$

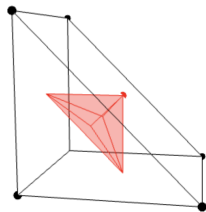
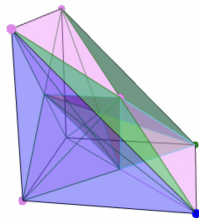
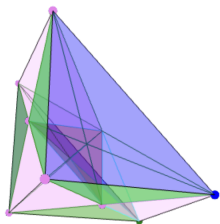
It is a vertex of the polytope Q_b where $b = (1.3277548209, 0.655509642)$ lies on the blue edge:



Example (2×3 independence)

Consider the independence model of a binary and ternary random variables X and Y .

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



2×3 independence continued

- 109 chambers in the chamber complex!
- But $\text{conv}(A)$ is highly symmetrical due to the action of $S_2 \times S_3$ on the states of X and $Y \implies$ only need to study one chamber in each orbit.
- After eliminating chambers, we are left with three edges e_1 , e_2 , and e_3 .
- Parameterize b on e_1 as $r(1/2, 1/2, 0) + (1 - r)(1/2, 0, 1/2)$.
- The only vertex-face pair we need to consider is the pair $v = 1/2(1, 0, 0, 0, r, 1 - r)$ and $w = 1/2(0, r, 1 - r, 1, 0, 0)$.
- The parametrization of the line between them gives rise to the single equation $(s - 1)^2 - s^2 = 0 \implies s = 1/2, 0 \leq r \leq 1$.
- Upon substituting $s = 1/2$ into the divergence function $D(v \| mle(v))$, we get the constant value $\log 2$.
- Therefore, the divergence at every point b of the edge e_1 is $\log 2$.
- By symmetry, the same is true of e_2 and e_3 .
- The maximum divergence from this model is $\log 2$ and there are infinitely many maximizers!

More in the paper...

- Maximum divergence from reducible hierarchical log-linear models.
 - ▶ decomposition theory of logarithmic Voronoi polytopes
 - ▶ study how to use this decomposition to obtain and bound information divergence to reducible models
- Maximum divergence from toric models of ML degree one:
 - ▶ Multinomial distributions revisited
 - ▶ Box model
 - ▶ Trapezoid model
 - ▶ Some three-dimensional models

Thanks!