

A Connected Positive Model with a Nonsemialgebraic Likelihood Cell

Alper Ferudun*

18 October 2026

Abstract

We give a connected compact rational polygonal model on four outcomes, all of whose probabilities are at least $1/32$, with a nonsemialgebraic logarithmic Voronoi cell at a smooth rational point. Near a specified positive data vector, the cell is the epigraph of an explicit binary entropy expression in rational coordinates. A uniform likelihood gap shows that a connecting segment does not alter this boundary. An elementary argument proves nonsemialgebraicity over the real numbers. The number of outcomes is minimal for such a cell at a positive-dimensional smooth interior model point: on at most three outcomes, log-normal containment and convexity force the cell to be an interval or a point. The construction complements existing nonalgebraicity results for likelihood cells; it does not assert that this general phenomenon is new.

1 Likelihood cells and the scope of the example

Let

$$\Delta_{m-1} = \{u \in \mathbb{R}_{\geq 0}^m : \sum_i u_i = 1\}, \quad \Delta_{m-1}^\circ = \Delta_{m-1} \cap \mathbb{R}_{> 0}^m.$$

For a model $\mathcal{M} \subset \Delta_{m-1}^\circ$, a point $p \in \mathcal{M}$, and data $u \in \Delta_{m-1}$, define

$$\ell_u(x) = \sum_i u_i \log x_i, \quad \mathcal{V}_{\mathcal{M}}(p) = \{u \in \Delta_{m-1} : \ell_u(p) \geq \ell_u(x) \text{ for every } x \in \mathcal{M}\}.$$

We use natural logarithms. The data may be arbitrary real frequencies, not only the empirical frequencies for a fixed sample size. Semialgebraic means definable by a finite Boolean combination of polynomial equalities and inequalities with *real* coefficients.

Alexandr and Heaton introduced these logarithmic Voronoi cells, proved their convexity and log-normal containment, and established polytopal descriptions for several model classes [2]. Their accompanying computation [3] already exhibits a numerically nonpolyhedral cell for two disjoint linear models in Δ_3 . Alexandr subsequently proved arithmetic transcendence of a smooth cell's endpoint and geometric nonalgebraicity of a cell at a cusp [4]. These are direct antecedents of this note. Arithmetic transcendence of an endpoint alone does not imply nonsemialgebraicity over $\mathbb{R}$: a line segment with arbitrary real endpoints is semialgebraic.

Our contribution is a closed-form connected example with uniform positivity, a smooth distinguished point, and a sharp outcome count. The source motivation is AIM Question 26 concerning the implications of semialgebraic geometry for maximum likelihood [1]. That open-ended question is not a conjecture that all likelihood cells are semialgebraic, and we do not claim to exhaust it.

*Mercury Software GmbH. Email: alper@mercurycodelab.com. GitHub: AlperTheKing.

Theorem 1.1. *There is a compact connected semialgebraic polygonal arc $\mathcal{M} \subset \Delta_3^\circ$, defined over $\mathbb{Q}$, with the following properties.*

1. *Every coordinate of every point of $\mathcal{M}$ is at least $1/32$.*
2. *The point $q = (1/8, 1/8, 3/8, 3/8)$ is a smooth interior point of the arc, but $\mathcal{V}_{\mathcal{M}}(q)$ is not semialgebraic over $\mathbb{R}$.*
3. *In rational coordinates*

$$u(r, z) = \frac{(r, 1 - r, z/2, z/2)}{1 + z},$$

the cell near $(r, z) = (1/2, \log 2 / \log(3/2))$ is exactly

$$z \geq \frac{\log 4 - H(r)}{\log(3/2)}, \quad H(r) = -r \log r - (1 - r) \log(1 - r).$$

The arc has corners away from q . We do not claim global smoothness or that the whole model is the intersection of an algebraic variety with the simplex. Its semialgebraic and connected character is part of the assertion.

2 A rational model and its competing maxima

Put $\delta = 1/16$ and define two segments

$$a(t) = (t/2, (1 - t)/2, 1/4, 1/4), \quad b(s) = (1/8, 1/8, 3s/4, 3(1 - s)/4), \quad \delta \leq t, s \leq 1 - \delta.$$

Join their left endpoints by

$$c(\lambda) = \frac{1}{64}(2 + 6\lambda, 30 - 22\lambda, 16 - 13\lambda, 16 + 29\lambda), \quad 0 \leq \lambda \leq 1. \quad (1)$$

Set

$$\mathcal{M} = a([\delta, 1 - \delta]) \cup c([0, 1]) \cup b([\delta, 1 - \delta]).$$

Direct substitution gives $c(0) = a(\delta)$ and $c(1) = b(\delta)$. The first two coordinate sums are respectively $1/2$ on a , $1/4$ on b , and $1/2 - \lambda/4$ on the connector. The segments therefore have no unintended intersections. They form a compact connected polygonal arc. All coordinates are at least $1/32$; this follows at the endpoints and hence throughout each segment. The point $q = b(1/2)$ lies on an open straight portion of the arc.

For q to be a maximum, differentiation along b gives $u_3 = u_4$. The positive data satisfying this condition are exactly $u(r, z)$, where $0 < r < 1$ and $z > 0$. The inverse coordinates are

$$r = \frac{u_1}{u_1 + u_2}, \quad z = \frac{u_3 + u_4}{u_1 + u_2}.$$

Use the rescaled score

$$S_{r,z}(x) = (1 + z)\ell_{u(r,z)}(x) = r \log x_1 + (1 - r) \log x_2 + \frac{z}{2} \log(x_3 x_4).$$

For $r \in (\delta, 1 - \delta)$, strict concavity gives the unique maximizer $a(r)$ on the first segment. Similarly, q uniquely maximizes the score on the second segment. Their difference is

$$S_{r,z}(a(r)) - S_{r,z}(q) = \log 4 - H(r) - z \log(3/2). \quad (2)$$

Thus these two segments tie on

$$z = F(r) := \frac{\log 4 - H(r)}{\log(3/2)}. \quad (3)$$

One must still exclude a higher maximum on the connecting segment.

3 A strict gap for the connector

Let $r_0 = 1/2$ and $z_0 = F(r_0) = \log 2 / \log(3/2)$. Then $0 < z_0 < 2$, since $2 < (3/2)^2$. The points $a(1/2) = (1/4, 1/4, 1/4, 1/4)$ and q have equal scores.

For $c = c(\lambda)$ define

$$A(\lambda) = 16c_1c_2 = \frac{15 + 34\lambda - 33\lambda^2}{64},$$

$$B(\lambda) = 16c_3c_4 = 1 + \lambda - \frac{377}{256}\lambda^2.$$

They are positive on $[0, 1]$. Completing the square gives the identities

$$\frac{49}{132} - A(\lambda) = \frac{33}{64} \left(\lambda - \frac{17}{33} \right)^2,$$

$$\frac{441}{377} - B(\lambda) = \frac{377}{256} \left(\lambda - \frac{128}{377} \right)^2.$$

In particular, $A < 3/8$ and $B < 5/4$. Consequently,

$$\begin{aligned} \exp(2[S_{r_0, z_0}(c(\lambda)) - S_{r_0, z_0}(a(1/2))]) &= A(\lambda)B(\lambda)^{z_0} \\ &\leq \frac{3}{8} \left(\frac{5}{4} \right)^2 = \frac{75}{128} < 1. \end{aligned} \tag{4}$$

For the middle inequality, use $B^{z_0} \leq 1$ if $B \leq 1$; otherwise use $z_0 < 2$. The gap is uniform over the connector.

By compactness of $[0, 1]$ and continuity of the scores, there is an open neighborhood U of (r_0, z_0) , inside $(\delta, 1 - \delta) \times (0, \infty)$, such that

$$S_{r,z}(c(\lambda)) < S_{r,z}(q) \quad ((r, z) \in U, 0 \leq \lambda \leq 1).$$

It follows from (2) that

$$u(r, z) \in \mathcal{V}_{\mathcal{M}}(q) \iff z \geq F(r) \quad ((r, z) \in U). \tag{5}$$

In particular, the tie curve consists of global maxima of the entire model, not merely equal values at stationary points.

4 The boundary is not semialgebraic

We include an elementary nonalgebraicity argument in order to distinguish arbitrary real coefficients from algebraic coefficients.

Lemma 4.1. *The logarithm is not semialgebraic on any nonempty open interval of positive real numbers.*

Proof. A semialgebraic graph in the plane is contained in the zero set of a nonzero polynomial. Indeed, write its defining formula as a finite disjunction of polynomial sign conditions. Each nonempty sign condition contained in the graph must contain a nontrivial polynomial equality, since a sign condition consisting only of strict inequalities is open. Multiplying these finitely many equality polynomials proves the assertion.

Thus semialgebraicity of the logarithm on an interval would give a nonzero $P(X, Y) \in \mathbb{R}[X, Y]$ with $P(x, \log x) = 0$ there. Substituting $x = e^y$ and using the identity theorem for real analytic functions gives $P(e^y, y) = 0$ for every $y \in \mathbb{R}$. Write

$$P(e^y, y) = \sum_{j=0}^m p_j(y) e^{jy}, \quad p_m \neq 0.$$

After division by e^{my} , the terms with $j < m$ tend to zero as $y \rightarrow +\infty$. Hence $p_m(y)$ tends to zero, impossible for a nonzero polynomial. The case $m = 0$ is immediate. $\square$

Lemma 4.2. *The graph of F in (3) is not semialgebraic on any nonempty open subinterval of $(0, 1)$.*

Proof. The derivative of a differentiable semialgebraic function is semialgebraic. This is a direct application of Tarski–Seidenberg to the quantified definition of the derivative. But

$$F'(r) = \frac{\log(r/(1-r))}{\log(3/2)}.$$

If F were semialgebraic on an interval, compose this derivative with the rational map $r = x/(1+x)$ and multiply by the nonzero real constant $\log(3/2)$. The result would be a semialgebraic logarithm on an open interval, contradicting Lemma 4.1. $\square$

Completion of the proof of Theorem 1.1. If $\mathcal{V}_{\mathcal{M}}(q)$ were semialgebraic, its preimage under the rational map $u(r, z)$ would be semialgebraic. Choose an open rectangle with rational endpoints around (r_0, z_0) whose closure is contained in U , and, if necessary, shrink its horizontal interval so that the graph of F stays inside its vertical interval. By (5), the relative boundary of the preimage in this rectangle is precisely the graph of F . Boundaries and intersections of semialgebraic sets are semialgebraic. Lemma 4.2 gives a contradiction. $\square$

5 A sharp outcome bound and a convex contrast

The following dimension observation uses the standard convexity and log-normal containment of likelihood cells [2]. We give the proof to specify the smooth-point hypothesis.

Proposition 5.1. *Suppose $N \subset \Delta_{m-1}^\circ$ is locally a C^1 embedded manifold without boundary of dimension $d \geq 1$ near $p \in N$. Then $\mathcal{V}_N(p)$ lies in an affine space of dimension at most $m - 1 - d$. In particular, if $m \leq 3$, the cell is semialgebraic over $\mathbb{R}$.*

Proof. For each tangent vector $v \in T_p N$, necessary stationarity gives

$$\sum_i u_i v_i / p_i = 0.$$

These d linear conditions are independent of the normalization $\sum_i u_i = 1$. To verify this, a linear combination of the tangent conditions proportional to normalization would mean $v_i / p_i = a$ for a tangent vector v , hence $v = ap$. Since $\sum_i v_i = 0$ and $\sum_i p_i = 1$, one has $a = 0$ and $v = 0$. This proves the dimension bound.

The cell is convex, being the intersection with the data simplex of the half-spaces

$$u \cdot (\log p - \log x) \geq 0, \quad x \in N.$$

When $m \leq 3$ and $d \geq 1$, it therefore is a convex subset of an affine line or a point. Such a set is semialgebraic over $\mathbb{R}$, regardless of the arithmetic nature of its endpoints. $\square$

Thus four outcomes are minimal for Theorem 1.1’s phenomenon in the stated smooth positive-dimensional class. The hypothesis excludes isolated model points, singularities, and one-sided boundary points.

Proposition 5.2. *If $N \subset \Delta_{m-1}^\circ$ is convex and semialgebraic, the graph of its global maximum likelihood estimator for positive real data is semialgebraic.*

Proof. For $p \in N$ and $u \in \Delta_{m-1}^\circ$, concavity makes global optimality equivalent to

$$\forall x \in N : \sum_i u_i(x_i - p_i)/p_i \leq 0.$$

Necessity follows by differentiation along feasible segments; sufficiency follows from the first-order concavity inequality. Multiplication by $\prod_i p_i > 0$ turns the condition into a first-order polynomial formula. Tarski–Seidenberg applies. $\square$

No closedness assumption is needed for the graph assertion; some fibers may be empty. The model of Theorem 1.1 is connected but not convex. Its nonsemialgebraic cell also shows that its joint global-MLE graph is not semialgebraic, since taking the fiber at q would preserve semialgebraicity. For a fixed integer count vector the likelihood is a polynomial in probabilities, so quantifier elimination can still describe its individual argmax. That fixed-count observation is not a joint real-data statement.

The construction supplies a concrete contribution to the broad AIM question, not a resolution of every issue it raises. It also does not address the Gaussian ellipsope conjecture of Alexandr and Høsten [5]. The prior nonalgebraicity results [4] and the earlier two-line computation [3] remain essential context.

The accompanying code checks the rational identities and performs separately labelled floating-point optimization diagnostics. Neither finite testing nor this author’s self-audit is formal verification or external peer review. AI assistance was used in deriving and checking the example, locating sources, and preparing the text and code; the author is responsible for the claims in this unrefereed note.

References

- [1] American Institute of Mathematics, *Computational algebraic statistics*, workshop problem notes (2004), Question 26, p. 7. <https://aimath.org/WWN/compalgstat/compalgstat.pdf>.
- [2] Y. Alexandr and A. Heaton, Logarithmic Voronoi cells, *Algebraic Statistics* **12** (2021), 75–95. doi:10.2140/astat.2021.12.75.
- [3] Y. Alexandr and A. Heaton, *Logarithmic Voronoi cells: nonlinear Voronoi computation*, Math-Repo companion notebook for [2]. <https://mathrepo.mis.mpg.de/logarithmicVoronoi/nonlinear-Voronoi.html>.
- [4] Y. Alexandr, *Boundaries of logarithmic Voronoi cells can be transcendental*, arXiv:2608.18434v1 (19 August 2026). <https://arxiv.org/abs/2608.18434>.
- [5] Y. Alexandr and S. Høsten, Logarithmic Voronoi cells for Gaussian models, *Journal of Symbolic Computation* **122** (2024), 102256. doi:10.1016/j.jsc.2023.102256.