Strict Likelihood Multimodality and Septic Equations for Two-Component Birkhoff Models

Alper Ferudun*

28 September 2026

Abstract

We give a positive integer dataset on S_4 whose likelihood on the closed two-component Birkhoff model has at least three distinct strict local maxima in distribution space. The dataset has 2013 observations and all three probability vectors have full support. One component of each displayed mode is a boundary component. An exact nonnegative-decomposition certificate and a compactness argument rule out the possibility that the modes are artifacts of a parameter chart or component-label switching. Separately, for every $n \geq 4$, unequal positive parity-constant counts give at least three distinct global maximizers in the closed two-component model. We also determine the first nonzero homogeneous equation degree of the complex S_4 secant variety: it is seven. A 96-term septic and its 24 symmetry images supply local equations at the displayed smooth point. We do not determine the global defining ideal or assert modes with both component matrices strictly positive.

1 The model and the results

For a positive $n \times n$ matrix T, put

$$w_T(\sigma) = \prod_{i=1}^{n} T_{i,\sigma(i)}, \qquad p_T(\sigma) = \frac{w_T(\sigma)}{\operatorname{per}(T)} \quad (\sigma \in S_n).$$

The first-order permutation model is the classical Birkhoff toric model [2, 4]. Let $\mathcal{B}_n$ be its Euclidean closure in the probability simplex, and define the *closed statistical two-component model*

$$\mathcal{M}_n = \{\lambda p + (1-\lambda) q : p, q \in \mathcal{B}_n,\ 0 \leq \lambda \leq 1\}.$$

Both sets are compact. They must not be confused with the entire real or complex locus of a secant variety. For positive counts u_σ define

$$\ell_u(p) = \sum_{\sigma \in S_n} u_\sigma \log p_\sigma,$$

with value $-\infty$ at a vector with a zero coordinate.

Problem 58 of the AIM computational algebraic statistics workshop asks about equations and likelihood multimodality for a mixture of the S_4 model [1, p. 11]. The first-order model is specified on p. 3 of that report. The subsequent heading "MLE Project" starts a separate discussion. Our results concern the closed statistical model just defined. They do not settle every interpretation of the source question or its global ideal problem.

*Mercury Software GmbH. Email: alper@mercurycodelab.com. GitHub: `https://github.com/AlperTheKing`.

Theorem 1.1. *Assign counts to $\sigma \in S_4$ according to*

	σ *even*	σ *odd*
$\sigma(4) = 4$	219	95
$\sigma(4) \neq 4$	49	70

The sample size is 2013*. On $\mathcal{M}_4$, the likelihood has at least three distinct strict local maxima, all with full support. These are distinct distributions, not component-label copies.*

The strict modes will be certified by exact rational calculations and an explicit image-space argument. A second, independent construction proves global nonuniqueness for every $n \geq 4$ in Section 5. Section 6 gives the minimum equation degree and the septic certificate. Full coefficient lists and standard-library Python verifiers are included in the accompanying source package. The author used generative-AI assistance in developing and checking the arguments and preparing the text. This preprint is unrefereed; no independent certification or absolute priority claim is made.

2 An exact parameter-space maximum

Fix the last row and last column of both component matrices at 1. Row and column rescalings put every positive matrix in this gauge without changing its probability distribution. Write

$$p_\sigma = \frac{t w_A(\sigma) + w_B(\sigma)}{Z}, \qquad Z = t\,\mathrm{per}(A) + \mathrm{per}(B),$$

and consider

$$A = \begin{pmatrix} 3 & 2 & 2 & 1 \\ 2 & 3 & 2 & 1 \\ 2 & 2 & 3 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 5 & 5 & 1 \\ 5 & 0 & 5 & 1 \\ 5 & 5 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \quad t = \frac{200}{31}. \tag{1}$$

All 24 probabilities are positive, although three entries of B are zero. The gauge has 19 coordinates: t and the two upper 3×3 blocks. Its active face has dimension 16.

Lemma 2.1. *For the data of Theorem 1.1, the point* (1) *is a strict local maximum in the nonnegative gauge-fixed parameter space.*

Proof. Use logarithmic coordinates on the 16 positive parameters, and ordinary coordinates in the three zero entries B_{11}, B_{22}, B_{33}. For $f = \ell_u/2013$, all 16 active scores vanish. The negative active Hessian is positive definite. Each inward ordinary-coordinate derivative is

$$\frac{\partial f}{\partial B_{ii}} = -\frac{31}{966240} \qquad (i = 1, 2, 3). \tag{2}$$

Here these are exact rational identities, not numerical optimizer output. The certificate can be reproduced from the following formulas. If $R_s = t w_A(s) + w_B(s)$ and subscripts denote ordinary derivatives, then

$$f_i = \sum_s \frac{u_s}{2013} \frac{R_{s,i}}{R_s} - \frac{Z_i}{Z}, \tag{3}$$

$$f_{ij} = \sum_s \frac{u_s}{2013} \left(\frac{R_{s,ij}}{R_s} - \frac{R_{s,i} R_{s,j}}{R_s^2} \right) - \frac{Z_{ij}}{Z} + \frac{Z_i Z_j}{Z^2}. \tag{4}$$

The R_s are multiaffine monomial sums, so their derivatives are computed by deleting the differentiated factors, including at zero entries. The verifier `verify_boundary_mode.py` computes (3)– (4) in all 19 ordinary coordinates. It checks zero active scores, (2), all 16 positive leading principal minors of the negative face Hessian, and congruence to the independently calculated log-coordinate Hessian. No differentiation divides by a zero parameter.

Taylor expansion on the active face gives quadratic decrease. Continuity makes all three inward derivatives uniformly negative in a sufficiently small neighborhood. Integrating along a normal segment therefore gives, for active displacement h and normal displacement $v \geq 0$,

$$f(h, v) - f(0, 0) \leq -c\|h\|^2 - d \sum_i v_i$$

for some $c, d > 0$. This proves strictness in parameter space. □

A parameter-space statement alone would be insufficient: another decomposition of the same distribution could yield an improving image branch. The next section eliminates that possibility at this specific point.

3 A specific nonnegative-fiber certificate

Number the permutations lexicographically from 0 to 23. Work temporarily with unnormalized coordinates

$$U_\sigma = 200 w_A(\sigma), \qquad V_\sigma = 31 w_B(\sigma), \qquad q = U + V.$$

Lemma 3.1. *If U', V' are nonnegative weighted Birkhoff components with $U' + V' = q$, then $\{U', V'\} = \{U, V\}$.*

Proof. Let $a_\sigma \in \{0, 1\}^{16}$ be the permutation matrix written as a vector. Every binomial $p^m - p^{m'}$ with $\sum_{\sigma \in m} a_\sigma = \sum_{\sigma \in m'} a_\sigma$ vanishes on a weighted Birkhoff component, including boundary components by continuity. We use necessary quadratic and cubic binomials only; their generation of the toric ideal is not required for this argument.

Write $U' = q/2 + z$ and $V' = q/2 - z$. Subtracting their quadratic equations gives 18 independent linear constraints on z. Put these constraints in reduced row-echelon form, with lexicographic coordinate order. The free indices are

$$(14, 19, 20, 21, 22, 23).$$

Write $z = Kx$ in the resulting kernel basis, with the identity matrix on these six free rows. The known decomposition gives

$$x_0 = (1200, 600, 600, 900, 25/2, 25/2)^{\mathsf{T}}.$$

Taking the even parts of the quadratic and cubic equations yields linear constraints in the 21 products $x_i x_j$ for $i \leq j$. Twenty supplied constraints are independent. Consequently

$$x x^{\mathsf{T}} = x_0 x_0^{\mathsf{T}} + sW, \tag{5}$$

where

$$W = \begin{pmatrix} 6792/161 & 288/23 & 288/23 & 1644/161 & 357/46 & 357/46 \\ 288/23 & 72/23 & 72/23 & 36/23 & 129/46 & 129/46 \\ 288/23 & 72/23 & 72/23 & 36/23 & 129/46 & 129/46 \\ 1644/161 & 36/23 & 36/23 & -192/161 & 72/23 & 72/23 \\ 357/46 & 129/46 & 129/46 & 72/23 & 1 & 1 \\ 357/46 & 129/46 & 129/46 & 72/23 & 1 & 1 \end{pmatrix}. \tag{6}$$

Section 7 describes a direct, independent verification of these finite linear-algebra statements.

The principal 2×2 minor on the first two coordinates of the right-hand side of (5) equals

$$s\left(\frac{267840000}{161} - \frac{91584}{3703}s\right).$$

It vanishes because xx^{T} has rank at most one. Thus $s = 0$ or $s = 3565000/53$. In the second case, reconstruction by $z = Kx$ gives

$$z_{\mathrm{id}}^2 - (q_{\mathrm{id}}/2)^2 = \frac{22458880000}{54537} > 0. \tag{7}$$

For example, the known identity coordinate has $z_{\mathrm{id}} = q_{\mathrm{id}}/2 = 2700$, and $(KWK^{\mathsf{T}})_{\mathrm{id,id}} = 144896/23667$; multiplication by the second value of s gives (7). Nonnegativity of both $q_{\mathrm{id}}/2 + z_{\mathrm{id}}$ and $q_{\mathrm{id}}/2 - z_{\mathrm{id}}$ excludes this case. Therefore $s = 0$, so $xx^{\mathsf{T}} = x_0x_0^{\mathsf{T}}$ and $x = \pm x_0$. This is exactly the component swap. ☐

The lifted linear system has rank 20, not 21, and its null direction has rank two. Thus lifted rank alone does not prove uniqueness. The second rank-one possibility is excluded by statistical nonnegativity, not by an assertion of unrestricted algebraic identifiability.

4 From parameter modes to distribution modes

Lemma 4.1. *Near each of the normalized components in* (1)*, its gauge matrix can be recovered continuously from its probability vector, for every nearby point of $\mathcal{B}_4$.*

Proof. For a positive component, logarithms of the probabilities are linear in $\log \kappa$ and the nine free logarithmic entries, where $\kappa = 1/\operatorname{per}(T)$. The 24 exponent rows have rank 10, so suitable ten probabilities recover these quantities continuously near A.

For B, use cells $(7, 8, 9, 10, 12, 13, 18)$. Their probabilities are positive and their monomials avoid the three zero diagonal entries. The exponent matrix for $\log \kappa$ and the six off-diagonal logarithmic entries has rank seven. It recovers those seven positive quantities continuously. Cells $(2, 14, 6)$, respectively, contain one occurrence of T_{11}, T_{22}, T_{33} and otherwise only already recovered positive factors. Division recovers the three diagonal entries, nonnegatively and continuously.

These formulas hold on the dense positive part of $\mathcal{B}_4$. Their denominators and logarithm arguments remain positive near the displayed component, so continuity extends the formulas to every nearby closure point, not merely to points on the same boundary face. The rank and support facts are verified in the finite certificate. ☐

Proof of Theorem 1.1. Let p be the normalized vector from (1). Suppose distinct $p_k \in \mathcal{M}_4$ tend to p and satisfy $\ell_u(p_k) \geq \ell_u(p)$. Choose a decomposition of each p_k. Compactness of $\mathcal{B}_4 \times \mathcal{B}_4 \times [0, 1]$ yields a convergent subsequence. Its limiting weighted pair decomposes p; Lemma 3.1, after constant rescaling, identifies it with the displayed pair up to a swap. Both weights are nonzero.

After swapping if necessary, Lemma 4.1 gives gauge matrices converging to A, B. The relative raw weight is

$$t = \frac{\lambda \operatorname{per}(B)}{(1-\lambda)\operatorname{per}(A)},$$

so it also converges to the displayed value. The distributions $p_k \neq p$ therefore come from parameters approaching, but unequal to, the strict parameter maximum of Lemma 2.1. This contradicts the likelihood inequality. Hence p is a strict local maximum in distribution space.

Cycle the first three rows of both matrices. This preserves parity and the condition $\sigma(4) = 4$, and hence preserves the data. The resulting three probability vectors are distinct by exact fraction comparison. Row permutation is a homeomorphism of $\mathcal{M}_4$, so all three points are strict local maxima. Full support follows from the positive A component. $\square$

5 Parity and global nonuniqueness

Lemma 5.1. *For $n \geq 4$, the only point of $\mathcal{M}_n$ that is constant on the even permutations and constant on the odd permutations is the uniform distribution.*

Proof. First consider a mixture with strictly positive parameters on four rows, written $p_\sigma = c\prod_i X_{i,\sigma(i)} + d\prod_i Y_{i,\sigma(i)}$, with $c, d > 0$. Put

$$r_{ij} = X_{ij}/Y_{ij}, \qquad u_j = r_{1j}/r_{2j}, \quad v_j = r_{3j}/r_{4j}.$$

For distinct i, j, k, l, the determinant

$$D_{ijkl} = p_{(i,j,k,l)}p_{(j,i,l,k)} - p_{(j,i,k,l)}p_{(i,j,l,k)}$$

has the sign of $(u_i - u_j)(v_k - v_l)$. This follows by expanding the identity

$$\begin{aligned}(UV + ab)(U'V' + a'b') - (U'V + a'b)(UV' + ab')\\ = (Ua' - U'a)(Vb' - V'b),\end{aligned}$$

where the factors on the two row pairs are grouped separately. All omitted normalization and matrix factors are positive.

If the even and odd coordinate levels were $E \neq O$, then $\operatorname{sgn} D_{ijkl} = \epsilon \operatorname{sgn}(i, j, k, l)$, with $\epsilon = \operatorname{sgn}(E^2 - O^2) \neq 0$. All the u_j differ. Sort them by relabeling columns and absorb the relabeling parity into ϵ. For $u_1 < u_2 < u_3 < u_4$ and $\epsilon = 1$, the orders $1234, 1324, 1423$ require

$$v_3 < v_4 < v_2 < v_3,$$

which is impossible. For $\epsilon = -1$ all inequalities reverse, with the same contradiction.

The forbidden strict sign pattern is an open condition on finitely many continuous coordinate determinants. It is therefore also excluded from limits of positive mixtures, including zero parameters or weights. For $n > 4$, fix the remaining rows and symbols; their positive factors are absorbed into c, d before taking limits. This proves the lemma. $\square$

Theorem 5.2. *Let $n \geq 4$ and let a, b be distinct positive integers. Give each even permutation count a and each odd permutation count b. The likelihood on $\mathcal{M}_n$ has at least three distinct full-support global maximizers.*

Proof. The likelihood product is continuous on the compact model and is positive at the uniform vector. Every global maximizer thus has full support. The alternating group A_n acts by row permutations, preserving the data and model. Its only fixed point in $\mathcal{M}_n$ is uniform by Lemma 5.1.

Uniform is not a maximum. Mix it toward p_T with $T = I + J$, using an odd row permutation of T if $a < b$. The directional derivative of the log-likelihood at uniform is

$$\frac{n!\,|a-b|\det(I+J)}{2\operatorname{per}(I+J)} > 0,$$

since the even-minus-odd mass of p_T is $\det(T)/\operatorname{per}(T)$ and $\det(I + J) = n + 1$. Thus the A_n-orbit of any global maximizer is nontrivial. It cannot have size two: a homomorphism $A_n \to S_2$ kills every 3-cycle, and 3-cycles generate A_n. The orbit therefore has at least three points. $\square$

Theorem 5.2 concerns a different dataset from Theorem 1.1; isolation of its global maximizers is not asserted. For $n = 4, a = 2, b = 1$, mixing uniform with p_{I+J} in proportion $1/100$ already gives the exact likelihood ratio

$$\left(\frac{6819}{6500}\right)^2 \left(\frac{6531}{6500}\right)^6 \left(\frac{6483}{6500}\right)^{16} \left(\frac{6459}{6500}\right)^{12} > 1.$$

The restriction $n \geq 4$ matters. For $n = 3$, a vector with levels $2/9$ on even and $1/9$ on odd permutations is a mixture of the identity point mass of weight $7/36$ and a Birkhoff component. Its remaining unnormalized vector is $(4/9)w_T$ for

$$T = \begin{pmatrix} 1/4 & 1/2 & 1 \\ 1/2 & 1/4 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

6 First equations in degree seven

Let $X \subset \mathbb{P}^{23}_{\mathbb{C}}$ be the second secant of the S_4 Birkhoff toric variety. Its homogeneous ideal is the kernel of

$$p_\sigma \longmapsto \prod_i X_{i,\sigma(i)} + \prod_i Y_{i,\sigma(i)}. \tag{8}$$

Component scales are absorbed in a row of each matrix. Polarization and multigraded coefficient comparison are standard secant methods [3, 4]; the assertions below use complete, exact substitution matrices.

Theorem 6.1. *The smallest degree of a nonzero homogeneous polynomial in $I(X)$ is seven. There is a primitive 96-term septic with coefficients in $\{-2, -1, 1, 2\}$. Its 24 column-permutation images are linearly independent, and four independent members cut out X locally at the point of Theorem 1.1. No global ideal-generation assertion is made.*

Proof. Partition degree-d monomials in the p variables by the sum D of their permutation matrices. Under (8), a term in a fixed block is specified by its X exponent C, since its Y exponent is $D - C$. Different D blocks cannot cancel. Each block therefore gives an integer coefficient matrix whose kernel is exactly that multigraded ideal piece.

For $d = 6$ there are 475020 monomials and 132724 blocks. Every block has full column rank. Two independent implementations certify this: reduced nine-entry gauge with component degree and dense row reduction modulo 1000003; full sixteen-entry base-7 coding and sparse column reduction modulo 1000033. Digits never exceed six, so the second encoding has no carries. The block-size histograms agree. A singleton block is independent because its all-Y substitution coefficient is 1. Nonzero modular maximal minors are nonzero over $\mathbb{Q}$ and $\mathbb{C}$. Thus there is no degree-six equation; any lower-degree equation could be multiplied by a coordinate power to give one, so degrees one through six are all excluded.

In total degree seven and multidegree $2J - I$, there are 126 monomials. Exact rational reduction yields a one-dimensional kernel, with the 96-term primitive vector stated in the theorem. The coefficient list is supplied as `target_degree_7.json`. A second formal expansion, in the nine-entry gauge with both component scales retained, cancels every coefficient exactly. This chart is dense, so the nonzero polynomial vanishes on X.

The 24 column permutations have distinct multidegrees $2J - P_\sigma$ and are therefore linearly independent. At the rational mode their gradients have exact rank four. The normalized mixture map has differential rank 19 there, matching its parameter-count upper bound. In the normalized 23-dimensional affine chart, four independent equations define a smooth 19-dimensional germ. The differential of the mixture map is an isomorphism onto its tangent space, so the inverse function theorem identifies this germ locally with X. Homogeneity ensures that restricting the four gradients to the normalization chart does not reduce their rank: each is orthogonal to p, whereas the normalization normal is not. □

The dimension-19 conclusion was already present in the frozen dataset's research attempt [5]; it is not counted here as a new general nondefectivity result. The specific Jacobian calculation is used for the local equations assertion. Also, the moment variety $M_{4,(321)}$ in [4, Proposition 6 and Example 7] is the Birkhoff variety; the variety $M_{4,4}$ in that paper's Proposition 34 is different, and its displayed equations cannot be substituted for the present ones.

7 Certificates, reproducibility and limitations

The accompanying source archive fixes all indexing and arithmetic. The following construction specifies the fiber certificate without numerical tolerances. Enumerate S_4 lexicographically, enumerate unordered monomials in that order, and group them by total permutation-matrix exponent. In each group use its first monomial minus every later monomial as binomials. This gives 18 quadrics and 592 cubics (including cubic multiples of quadrics). Polarize the quadrics to form the linear equations for z, and take their unique reduced row-echelon kernel basis with the free rows used in Section 3. The even parts after $z = Kx$ are linear in the lexicographically ordered products $x_i x_j$ for $i \leq j$.

The exact certificate lists 20 selected lifted binomials, their coefficient rows, right-hand sides, the kernel basis K and the pencil W. The separate program `audit_fiber_certificate.py` checks equality of the toric exponents, expands the selected relations as full polynomials in six variables, and verifies the stated rows. It proves ranks 18 and 20 by modular elimination over $\mathbb{F}_{1000003}$, verifies the rational null vectors exactly, and checks the minor and the exclusion (7). It also checks the monomial ranks and supports required for continuous parameter recovery. None of these tests assumes generic identifiability at the boundary point.

Exact likelihood/fiber verifier	Named checks
`verify_parity_obstruction.py`	1516
`verify_boundary_mode.py`	298
`audit_fiber_certificate.py`	57

All 1871 named checks pass on Python 3.13.5 and 3.12.14, and with optimization enabled. The complete degree-six rank scan and the independent septic expansion are separate certificates. These are computational parts of the proof, not independent peer review, formalization, or a measure of novelty. The compactness and continuous-inverse arguments above remain essential even after the finite tests pass.

Our scope limitations are substantive. We exhibit modes of the closed model with boundary components, not strict modes with two strictly positive matrices. The parity construction proves global nonuniqueness without isolation. The septic orbit supplies the first equation degree and local equations, not the entire global prime ideal or its radical. A bounded primary-source search located no direct predecessor for the specific certificates, but does not certify priority. All classical modeling and polarization ingredients are credited.

References

[1] American Institute of Mathematics, *Computational algebraic statistics: workshop problem list*, prepared by Seth Sullivant, pp. 3 and 11, Problem 58. `https://aimath.org/WWN/compalgstat/compalgstat.pdf`.

[2] B. Sturmfels and V. Welker, Commutative algebra of statistical ranking, *Journal of Algebra* **361** (2012), 264–286. arXiv:1101.1597v2.

[3] J. Sidman and S. Sullivant, Prolongations and computational algebra, arXiv:math/0611696v3 (2008), especially Theorems 4.1–4.2.

[4] Y. Alexandr, J. Kileel and B. Sturmfels, Moment varieties for mixtures of products, *ISSAC 2023*, 53–60. doi:10.1145/3597066.3597084; arXiv:2301.09068v2.

[5] `ulamai/UnsolvedMath`, frozen version 1.6.0, record AIM-COMPUTATION-0073 and its associated research attempt, commit `c6e7f41b4eb6ce876feb83e7eb626be1b243fd7e`. Machine-generated research aid, not peer reviewed. `https://huggingface.co/datasets/ulamai/UnsolvedMath`.