hexagon:2611.00734v1

THE MAXIMUM LIKELIHOOD DEGREE OF NO-THREE-WAY INTERACTION MODELS

SHANKAR SIVARAJAN

Department of Computational Medicine and Bioinformatics
University of Michigan, Ann Arbor, MI, USA
sivarajn@umich.edu

Abstract. We prove that for all $I, J, K \geq 2$ the maximum likelihood degree of the no-three-way interaction model of $I \times J \times K$ contingency tables is

$$(-1)^{(I-1)(J-1)} \int \beta(M_A)^K \, d\chi(A),$$

an integral with respect to the Euler characteristic over the $I \times J$ matrices A with nonzero entries modulo row and column scaling, where $\beta(M_A)$ is the beta invariant of the matroid of the columns of $[\mathrm{Id}_I \mid A]$. By a theorem of Clarke, Hoşten, Kushnerchuk, and Oldekop, $\beta(M_A)$ is the maximum likelihood degree of the independence model scaled by A. Closed forms were previously known only when one variable is binary, and the formula recovers them. For fixed I and J the maximum likelihood degree is an exponential polynomial in K; the coefficient of its largest base $\binom{I+J-2}{I-1}$ is the maximum likelihood degree of the moduli space $X(I, I+J)$ of points in linearly general position, which counts solutions to the generalized scattering equations of Cachazo, Early, Guevara, and Mizera. The same argument for a generically scaled model expresses the degree of $\mathcal{M}_{I,J,K}$ as an Euler-characteristic integral of products of beta invariants. For 3×3 matrices the integral reduces to counting strata of line arrangements, and the maximum likelihood degree of the $3 \times 3 \times K$ model is $26 \cdot 6^K - 60 \cdot 5^K + 36 \cdot 4^K + 6 \cdot 3^K - 9 \cdot 2^K + 1$; in particular the $3 \times 3 \times 3$ model has maximum likelihood degree 511.

1. Introduction

The no-three-way interaction model for an $I \times J \times K$ contingency table consists of the distributions proportional to $a_{ij} b_{ik} c_{jk}$, and their limits: each pair of the three variables interacts, and there is no three-factor interaction [3, 4]. Its maximum likelihood estimate has no closed form and is computed iteratively, for instance by iterative proportional fitting [11, 10]. The algebraic complexity of the estimate is measured by the maximum likelihood degree: the number of complex critical points of the likelihood function on the model for generic data [13, 2]. Only when one variable is binary is it known in closed form. Améndola et al. conjectured that the maximum likelihood degree of the $2 \times 2 \times K$ model is $2^K - 1$ [2, Conj. 36], and Brysiewicz and Maraj proved this and gave a closed formula for every $2 \times J \times K$ model by identifying it with a Lawrence lifting and expressing its maximum likelihood degree as an evaluation of a Tutte polynomial [5, Thm. 23, Cor. 32, Cor. 38]. They noted that the degree of the toric variety $\mathcal{M}_{3,3,3}$ is 2457 and left the models with no binary variable to future work [5, Rem. 39]. In the binary case, the maximum likelihood estimate is also given by explicit determinants: for $2 \times 2 \times K$ tables it is determined by a root of the determinant of a $2^K \times 2^K$ matrix, which gives an elementary proof that the maximum likelihood degree is $2^K - 1$, and for $2 \times J \times K$ tables the likelihood equations form a multiparameter eigenvalue problem [14].

We give a formula for every $I \times J \times K$ model (Theorem 2.1):

$$\mathrm{mldeg}(\mathcal{M}_{I,J,K}) = (-1)^{(I-1)(J-1)} \int \beta(M_A)^K \, d\chi(A),$$

where A runs over the $I \times J$ matrices with nonzero entries modulo row and column scaling, M_A is the matroid of the columns of $[\mathrm{Id}_I \mid A]$, β is Crapo's beta invariant [9], and the integral is with respect to the Euler characteristic [15]. By a theorem of Clarke, Hoşten, Kushnerchuk, and Oldekop [8], $\beta(M_A)$ is the maximum likelihood degree of the independence model of an $I \times J$ table scaled by A, so the maximum likelihood degree of the three-way model is a signed Euler-characteristic sum of the K-th powers of those of the scaled two-way models. For fixed I and J it is an exponential polynomial in K, whose bases are the beta invariants that occur and whose coefficients are Euler characteristics of matroid strata.

The proof applies Huh's theorem that the maximum likelihood degree of a smooth very affine variety is its signed Euler characteristic [13] to the model with its coordinate hyperplanes and the hyperplane $\sum p = 0$ removed, together with the observation that for a fixed interaction A the model decomposes into K independent slices whose contributions multiply. Theorem 2.1 is stated for the model X_T scaled by any tensor T, with $\beta(M_A)^K$ replaced by $\prod_k \beta(M_{T_k \circ A})$; for generic T this gives the degree of the toric variety, and in particular an expression for the degree 2457 of $\mathcal{M}_{3,3,3}$ (Corollary 4.1). The formula recovers the result of Brysiewicz and Maraj (Corollary 4.3): for $I = 2$ the strata are indexed by set partitions of $[J]$, and the integral becomes a Stirling-number sum equivalent to their Tutte-polynomial evaluation. Since the model is symmetric in its three variables, the formula also yields an identity between integrals over matrices of different shapes (Corollary 4.2).

Since $\beta(M_A)$ depends only on the matroid, the integral is a finite sum of Euler characteristics of matroid strata. For 3×3 matrices we compute them exactly, by fibering over the second column of A: the fibers are strata of arrangements of lines through pairs of at most five points in the plane, whose combinatorics depends only on which three of these points are collinear. This gives the maximum likelihood degree of every $3 \times 3 \times K$ model (Proposition 5.1); in particular the $3 \times 3 \times 3$ model has maximum likelihood degree 511. For 3×4 matrices we evaluate the integral by counting points over finite fields; the counts are polynomial in q for all primes tested, and their values at $q = 1$ give a formula for the $3 \times 4 \times K$ models (Section 5). Up to sign, the coefficient of the largest base $\binom{I+J-2}{I-1}$ is the Euler characteristic of the moduli space $X(I, I+J)$ of points in linearly general position, whose likelihood equations are the scattering equations of Cachazo, Early, Guevara, and Mizera [6, 1]; hence the leading coefficient of the maximum likelihood degree in K is the number of solutions of these equations (Proposition 6.1).

2. Statement

Fix integers $I, J, K \geq 2$. The no-three-way interaction model $\mathcal{M}_{I,J,K}$ is the toric variety $X \subseteq \mathbb{P}^{IJK-1}$, the closure of the image of

$$(a, b, c) \longmapsto \big(a_{ij} b_{ik} c_{jk}\big)_{1 \leq i \leq I,\, 1 \leq j \leq J,\, 1 \leq k \leq K}, \qquad (a, b, c) \in (\mathbb{C}^*)^{IJ} \times (\mathbb{C}^*)^{IK} \times (\mathbb{C}^*)^{JK}.$$

More generally, for $T = (t_{ijk}) \in (\mathbb{C}^*)^{IJK}$ let X_T be the closure of the image of $(a, b, c) \mapsto (t_{ijk} a_{ij} b_{ik} c_{jk})$, the model scaled by T, so that $X = X_{\mathbf{1}}$, and write $T_k = (t_{ijk})_{i,j}$ for its slices. Let $\mathcal{H} \subseteq \mathbb{P}^{IJK-1}$ be the union of the coordinate hyperplanes and the hyperplane $\sum_{ijk} p_{ijk} = 0$. The maximum likelihood degree $\mathrm{mldeg}(X_T)$ is the number of complex critical points of the log-likelihood function $\sum_{ijk} u_{ijk} \log p_{ijk} - u_{+++} \log \sum_{ijk} p_{ijk}$, where $u_{+++} = \sum_{ijk} u_{ijk}$, on the smooth locus of $X_T \setminus \mathcal{H}$ for generic data u.

For an $I \times J$ matrix $A = (a_{ij})$ with nonzero entries, let M_A be the matroid on $I+J$ elements represented by the columns of $[\mathrm{Id}_I \mid A]$, and let $\beta(M)$ be the beta invariant of Crapo [9],

$$\beta(M) = (-1)^{\operatorname{rk} M} \sum_{S \subseteq E} (-1)^{|S|} \operatorname{rk} S.$$

The torus $G = \big((\mathbb{C}^*)^I \times (\mathbb{C}^*)^J\big)/\mathbb{C}^*$ acts freely on $(\mathbb{C}^*)^{I \times J}$ by row and column scaling, and M_A is constant on orbits; the quotient is identified with the torus $T_{I,J} = (\mathbb{C}^*)^{(I-1)(J-1)}$ of matrices with $a_{1j} = a_{i1} = 1$. The function $A \mapsto \beta(M_A)$ is constructible, since the matroid strata are defined by the vanishing and nonvanishing of minors, and $\beta(M_A) \geq 1$, since M_A is connected on $I + J \geq 4$ elements: each column of A has full support, so $e_1, \ldots, e_I$ and any column of A form a circuit, and β of a connected matroid on at least two elements is positive [9].

For a constructible function φ on a complex algebraic variety Y that takes finitely many values, $\int_Y \varphi \, d\chi = \sum_t t\, \chi(\varphi^{-1}(t))$, where χ is the topological Euler characteristic [15].

Theorem 2.1. *For all $I, J, K \geq 2$ and $T \in (\mathbb{C}^*)^{IJK}$,*

$$\operatorname{mldeg}(X_T) = (-1)^{(I-1)(J-1)} \int_{T_{I,J}} \prod_{k=1}^{K} \beta(M_{T_k \circ A})\, d\chi(A),$$

where $\circ$ is the entrywise product. In particular,

$$\operatorname{mldeg}(\mathcal{M}_{I,J,K}) = (-1)^{(I-1)(J-1)} \int_{T_{I,J}} \beta(M_A)^K \, d\chi(A).$$

When $T = \mathbf{1}$, $\beta(M_A)$ is the maximum likelihood degree of the independence model of an $I \times J$ table scaled by A [8, Thm. 1.3] (this also follows from Lemma 3.5 below and Huh's theorem), so the integrand is its K-th power. Grouping matrices by the value of β, Theorem 2.1 reads

$$\operatorname{mldeg}(\mathcal{M}_{I,J,K}) = (-1)^{(I-1)(J-1)} \sum_{s \geq 1} W_{I,J}(s)\, s^K, \qquad W_{I,J}(s) = \chi\big(\{A \in T_{I,J} : \beta(M_A) = s\}\big),$$

an exponential polynomial in K whose bases are the beta invariants that occur.

3. Proof of Theorem 2.1

We use two properties of the Euler characteristic of complex algebraic varieties: it is additive over finite partitions into locally closed subvarieties, and for a morphism $\pi\colon Y \to B$,

$$\chi(Y) = \int_B \chi\big(\pi^{-1}(x)\big)\, d\chi(x), \tag{1}$$

where $x \mapsto \chi(\pi^{-1}(x))$ is constructible [15]. Viro's calculus uses the Euler characteristic with compact supports, which for complex algebraic varieties coincides with χ (see [7] and the references there). In particular χ is multiplicative on products, and $\chi(\mathbb{C}^*) = 0$.

The proof proceeds in four steps. We first identify $X_T \setminus \mathcal{H}$ with a smooth hypersurface $V = \{f = 1\}$ in a torus T' (Lemma 3.1) and apply Huh's theorem. Projecting V onto $A \in T_{I,J}$ reduces $\chi(V)$ to the Euler characteristic of a level set $\{g_1 + \cdots + g_K = 1\}$ of a sum of K homogeneous functions, which factors into single-slice contributions (Lemmas 3.2 and 3.3); each slice then evaluates to a signed beta invariant via hyperplane arrangement complements (Lemmas 3.4 and 3.5).

Lemma 3.1. *Let $d = (I-1)(J-1) + (I-1)K + JK$ and let $T' = (\mathbb{C}^*)^d$ have coordinates a_{ij} $(i,j \geq 2)$, b_{ik} $(i \geq 2)$, and c_{jk}, and put $a_{1j} = a_{i1} = b_{1k} = 1$. Then $X_T \setminus \mathcal{H}$ is isomorphic to the hypersurface*

$$V = \Big\{ x \in T' : f(x) = 1 \Big\}, \qquad f = \sum_{k=1}^{K} \sum_{i,j} t_{ijk} a_{ij} b_{ik} c_{jk} = \sum_{k=1}^{K} b_{\cdot k}^{\top} (T_k \circ A)\, c_{\cdot k},$$

which is smooth, of dimension $d-1$.

Proof. The map $\psi \colon T' \to (\mathbb{C}^*)^{IJK}$, $x \mapsto (t_{ijk} a_{ij} b_{ik} c_{jk})$, is injective with an inverse given by monomials on its image: with $q_{ijk} = p_{ijk}/t_{ijk}$, $c_{jk} = q_{1jk}$, $b_{ik} = q_{i1k}/q_{11k}$, and $a_{ij} = q_{ij1} q_{111} / (q_{i11} q_{1j1})$. Every (a,b,c) in the full torus can be rescaled by $a_{ij} \mapsto a_{ij} \lambda_i \mu_j$, $b_{ik} \mapsto b_{ik} \lambda_i^{-1} \nu_k$, $c_{jk} \mapsto c_{jk} \mu_j^{-1} \nu_k^{-1}$, which preserves every product $a_{ij} b_{ik} c_{jk}$, so as to satisfy $a_{1j} = a_{i1} = b_{1k} = 1$. Hence $\psi(T')$ is the image of the full torus, a translate of a closed subtorus of $(\mathbb{C}^*)^{IJK}$ that is stable under scaling (by scaling c); it is therefore $\widehat{X}_T \cap (\mathbb{C}^*)^{IJK}$, where $\widehat{X}_T \subseteq \mathbb{C}^{IJK}$ is the affine cone over X_T, and ψ is an isomorphism onto it. A point of $X_T \setminus \mathcal{H}$ has a unique representative p in this cone with $\sum p = 1$, and $\sum \psi(x) = f(x)$. The function f is homogeneous of degree one under $c \mapsto \lambda c$, so $\sum_{jk} c_{jk} \partial f / \partial c_{jk} = f = 1$ on V, and $df \neq 0$ there. ☐

By Lemma 3.1, $X_T \setminus \mathcal{H} \cong V$ is smooth, so it coincides with the set $(X_T)_{\mathrm{reg}} \setminus \mathcal{H}$ on which critical points are counted, and through the coordinates p it is a closed subvariety of $(\mathbb{C}^*)^{IJK}$, namely the torus coset $\widehat{X}_T \cap (\mathbb{C}^*)^{IJK}$ intersected with $\{\sum p = 1\}$. On V the term $u_{+++} \log \sum p$ vanishes, and since the log-likelihood is invariant under scaling p, its critical points on $X_T \setminus \mathcal{H}$ correspond to the critical points on V of the master function $\prod p_{ijk}^{u_{ijk}}$. By Huh's theorem [13, Thm. 1] (see also [8, Thm. 2.3] for this form), for generic u

$$\mathrm{mldeg}(X_T) = (-1)^{d-1} \chi(V). \tag{2}$$

Lemma 3.2. *For $m \geq 1$ let $H_m = \{v \in (\mathbb{C}^*)^m : v_1 + \cdots + v_m = 1\}$. Then $\chi(H_m) = (-1)^{m-1}$.*

Proof. For $m = 1$, H_1 is a point. For $m \geq 2$ project to v_m. Over $v_m \in \mathbb{C}^* \setminus \{1\}$ the fiber is $\{v' \in (\mathbb{C}^*)^{m-1} : \sum v' = 1 - v_m\} \cong H_{m-1}$, and $(v', v_m) \mapsto (v'/(1-v_m), v_m)$ trivializes this family over $\mathbb{C} \setminus \{0,1\}$. Over $v_m = 1$ the fiber $\{v' \in (\mathbb{C}^*)^{m-1} : \sum v' = 0\}$ is empty for $m = 2$ and isomorphic to $\mathbb{C}^* \times H_{m-2}$ for $m \geq 3$, by $v' \mapsto (v'_{m-1}, -(v'_1, \ldots, v'_{m-2})/v'_{m-1})$, so its Euler characteristic is 0. Hence $\chi(H_m) = \chi(\mathbb{C} \setminus \{0,1\}) \chi(H_{m-1}) = -\chi(H_{m-1})$. ☐

Lemma 3.3. *For $k = 1, \ldots, K$ let Y_k be a variety with an algebraic $\mathbb{C}^*$-action and $g_k \colon Y_k \to \mathbb{C}$ a regular function with $g_k(\lambda \cdot y) = \lambda g_k(y)$. Put $z_k = \chi(g_k^{-1}(0))$ and $e_k = \chi(g_k^{-1}(1))$. Then $z_k = \chi(Y_k)$ and*

$$\chi\Big(\Big\{ y \in \prod_k Y_k : \sum_k g_k(y_k) = 1 \Big\} \Big) = \prod_k z_k - \prod_k (z_k - e_k).$$

Proof. The map $y \mapsto (g_k(y), g_k(y)^{-1} \cdot y)$ is an isomorphism $Y_k \setminus g_k^{-1}(0) \to \mathbb{C}^* \times g_k^{-1}(1)$, so $\chi(Y_k) = z_k$. Partition the set on the left by $P = \{k : g_k(y_k) \neq 0\}$, which is nonempty. The part with a given P is isomorphic to $\prod_{k \notin P} g_k^{-1}(0) \times \prod_{k \in P} g_k^{-1}(1) \times H_{|P|}$, by $y \mapsto \big((y_k)_{k \notin P}, (g_k(y_k)^{-1} \cdot y_k)_{k \in P}, (g_k(y_k))_{k \in P} \big)$, with inverse $(y', e, v) \mapsto y$, where $y_k = y'_k$ for $k \notin P$ and $y_k = v_k \cdot e_k$ for $k \in P$. By Lemma 3.2 the left side is

$$\sum_{\emptyset \neq P \subseteq [K]} (-1)^{|P|-1} \prod_{k \notin P} z_k \prod_{k \in P} e_k = \prod_k z_k - \prod_k (z_k - e_k). \qquad ☐$$

Lemma 3.4. *Let $v_1, \dots, v_n \in \mathbb{C}^r$ span $\mathbb{C}^r$, let M be their matroid, and let $U_M \subseteq \mathbb{P}^{r-1}$ be the complement of the hyperplanes $\{x : v_e \cdot x = 0\}$. Then $\chi(U_M) = (-1)^{r-1}\beta(M)$.*

Proof. We use the following properties of β [9]: $\beta(M) = \beta(M \setminus e) + \beta(M/e)$ if e is neither a loop nor a coloop; $\beta(M) = 0$ if M has a loop, or a coloop and at least two elements; and $\beta(M) = 1$ if M is loopless of rank one. If M has a loop, U_M is empty. If $r = 1$ and M is loopless, $U_M = \mathbb{P}^0$ is a point. For $r \geq 2$ induct on n, for all ranks simultaneously. If every element is a coloop, $U_M \cong (\mathbb{C}^*)^{r-1}$ and $\beta(M) = 0$. Otherwise choose e neither a loop nor a coloop; then $M \setminus e$ has rank r and M/e has rank $r - 1 \geq 1$. Then $U_{M\setminus e}$ is the disjoint union of U_M and the complement in the hyperplane $H_e \cong \mathbb{P}^{r-2}$ of the traces of the remaining hyperplanes, which is $U_{M/e}$, since the trace of $\{v_f \cdot x = 0\}$ on H_e is defined by the image of v_f in $\mathbb{C}^r/\langle v_e\rangle$. Hence $\chi(U_M) = \chi(U_{M\setminus e}) - \chi(U_{M/e}) = (-1)^{r-1}\big(\beta(M \setminus e) + \beta(M/e)\big)$. □

Lemma 3.5. *For an $I \times J$ matrix A with nonzero entries let $Y = \{y \in (\mathbb{C}^*)^I : y_1 = 1\} \times (\mathbb{C}^*)^J$ and $g(y, \gamma) = y^\top A\gamma$. Then $\chi(g^{-1}(1)) = (-1)^{I+J}\beta(M_A)$.*

Proof. Project $g^{-1}(1)$ to y and put $w = A^\top y$. If $Z = \{j : w_j = 0\}$, the fiber is $\{\gamma \in (\mathbb{C}^*)^J : \sum_{j\notin Z} w_j\gamma_j = 1\}$, which is isomorphic to $(\mathbb{C}^*)^Z \times H_{J-|Z|}$ by $\gamma_j \mapsto w_j\gamma_j$ for $j \notin Z$ and the identity on γ_j for $j \in Z$ when $Z \neq [J]$, and is empty when $Z = [J]$. Its Euler characteristic is $(-1)^{J-1}$ if $Z = \emptyset$ and 0 otherwise, and it is constant on the locally closed sets of y with a given Z. By (1), $\chi(g^{-1}(1)) = (-1)^{J-1}\chi(U_A)$, where

$$U_A = \{y \in (\mathbb{C}^*)^I : y_1 = 1,\ (A^\top y)_j \neq 0 \text{ for all } j\}$$

is the complement in $\mathbb{P}^{I-1}$ of the hyperplanes whose normal vectors are the columns of $[\mathrm{Id}_I \mid A]$. By Lemma 3.4, $\chi(U_A) = (-1)^{I-1}\beta(M_A)$. □

Proof of Theorem 2.1. Project V to $A \in T_{I,J}$. The fiber over A is $\{(b, c) : \sum_k g_k(b_{\cdot k}, c_{\cdot k}) = 1\}$, where g_k is the function of Lemma 3.5 for the matrix $T_k \circ A$ on $Y_k = Y$, with $\mathbb{C}^*$ acting on the γ factor. Since Y is a torus of positive dimension, $z_k = \chi(Y) = 0$, and Lemmas 3.3 and 3.5 give the fiber Euler characteristic

$$-\prod_k(-e_k) = -(-1)^{(I+J-1)K}\prod_k \beta(M_{T_k\circ A}).$$

By (1) and (2),

$$\mathrm{mldeg}(X_T) = (-1)^d(-1)^{(I+J-1)K}\int_{T_{I,J}}\prod_k \beta(M_{T_k\circ A})\, d\chi(A),$$

and $d + (I + J - 1)K = (I - 1)(J - 1) + 2(I + J - 1)K \equiv (I - 1)(J - 1) \pmod 2$. □

4. Consequences

Corollary 4.1 (Degree)**.** *For generic $T \in (\mathbb{C}^*)^{IJK}$,*

$$\deg(\mathcal{M}_{I,J,K}) = (-1)^{(I-1)(J-1)}\int_{T_{I,J}}\prod_{k=1}^{K}\beta(M_{T_k\circ A})\, d\chi(A).$$

Proof. For a generic scaling the maximum likelihood degree of a toric variety equals its degree [2, Cor. 8]; apply Theorem 2.1. □

Brysiewicz and Maraj noted that they have no combinatorial interpretation of the degree 2457 of $\mathcal{M}_{3,3,3}$ [5, Rem. 39]; Corollary 4.1 expresses it as an Euler-characteristic integral of products of beta invariants over the common refinement of the matroid stratifications of

the matrices $T_k \circ A$. For example, for $I = J = 2$ write A in terms of $\psi = a_{22}$ and let $\rho_k = t_{11k}t_{22k}/(t_{12k}t_{21k})$ be the odds ratio of T_k. Then $\beta(M_{T_k \circ A}) = 2$ unless $\psi\rho_k = 1$, where it is 1. For generic T the ρ_k are distinct, the stratum $\mathbb{C}^* \setminus \{\rho_1^{-1}, \dots, \rho_K^{-1}\}$ has Euler characteristic $-K$, and the integral is $-K\,2^K + K\,2^{K-1} = -K\,2^{K-1}$, so $\deg(\mathcal{M}_{2,2,K}) = K\,2^{K-1}$ [2, Prop. 34].

Corollary 4.2 (Symmetry)**.** *For all $I, J, K \geq 2$,*

$$(-1)^{(I-1)(J-1)} \int_{(\mathbb{C}^*)^{(I-1)(J-1)}} \beta(M_A)^K \, d\chi(A) = (-1)^{(I-1)(K-1)} \int_{(\mathbb{C}^*)^{(I-1)(K-1)}} \beta(M_B)^J \, d\chi(B),$$

where A ranges over $I \times J$ and B over $I \times K$ matrices with first row and column equal to 1.

Proof. Both sides equal $\mathrm{mldeg}(\mathcal{M}_{I,J,K})$, by Theorem 2.1 applied to $\mathcal{M}_{I,J,K}$ and to the isomorphic model $\mathcal{M}_{I,K,J}$. □

The symmetry in I and J is visible directly: the matrix $[-A^\top \mid \mathrm{Id}_J]$ represents the dual of M_A with the same labels, and after scaling the first block by -1 and exchanging the two blocks it becomes $[\mathrm{Id}_J \mid A^\top]$, which represents $M_{A^\top}$; transposition maps the $I \times J$ chart onto the $J \times I$ chart, and $\beta(M^*) = \beta(M)$ for matroids on at least two elements [9]. The symmetry between J and K is not apparent from the integral.

Corollary 4.3 (Brysiewicz and Maraj [5, Cor. 32 and Cor. 38], in an equivalent form)**.**

$$\mathrm{mldeg}(\mathcal{M}_{2,J,K}) = (-1)^{J-1} \sum_{m=1}^{J} (-1)^{m-1}(m-1)!\, S(J,m)\, m^K,$$

where $S(J,m)$ is a Stirling number of the second kind.

Proof. For $I = 2$, A is determined by $x_j = a_{2j}$ with $x_1 = 1$, and M_A consists of the points $[1:0]$, $[0:1]$, and $[1:x_j]$ of $\mathbb{P}^1$. If the x_j take m distinct values, M_A is a rank-2 matroid whose simplification is the uniform matroid on $m+2$ elements, so $\beta(M_A) = m$ (deleting a parallel element in a loopless matroid leaves β unchanged, since contraction produces a loop). The matrices inducing a given partition of $[J]$ into m blocks are parametrized by the values of the $m-1$ blocks not containing 1: distinct points of $\mathbb{C} \setminus \{0,1\}$, whose configuration space has Euler characteristic $(-1)(-2)\cdots(-(m-1))$. There are $S(J,m)$ partitions, and the sign is $(-1)^{(2-1)(J-1)}$. □

5. Three-level variables

For $I = 3$, the matroid M_A is that of $J+3$ points in $\mathbb{P}^2$: the coordinate points and the columns of A. When the columns are distinct and no three of the $J+3$ points are collinear, M_A is uniform of rank three and $\beta(M_A) = \binom{J+1}{2}$; this is the largest value that occurs, since $\beta(M_A)$ is the maximum likelihood degree of the scaled independence model [8, Thm. 1.3], which is at most the degree $\binom{J+1}{2}$ of $\mathbb{P}^2 \times \mathbb{P}^{J-1}$ [2, Cor. 8].

We use two facts about β. If M is a simple matroid of rank three on n elements represented over $\mathbb{C}$, then

$$\beta(M) = \binom{n-2}{2} - \sum_L \binom{|L|-1}{2}, \tag{3}$$

the sum over the lines L of M with at least three elements. By Lemma 3.4, $\beta(M)$ is the Euler characteristic of the complement in $\mathbb{P}^2$ of the n lines dual to the elements, which is $3 - 2n + \sum_L (|L|-1)$, the sum now over all rank-two flats, the points where $|L|$ of the lines meet; and $\sum_L (|L|-1) = \binom{n}{2} - \sum_{|L|\geq 3} \binom{|L|-1}{2}$. Second, if e and f are parallel elements of a

matroid M without loops, then $\beta(M) = \beta(M \setminus e)$, since e is not a loop, is not a coloop because f spans it, and M/e has the loop f.

Proposition 5.1. *The weights* $W_{3,3}(1), \dots, W_{3,3}(6)$ *are* $1, -9, 6, 36, -60, 26$, *and* $W_{3,3}(s) = 0$ *for* $s > 6$. *Consequently, for all* $K \geq 2$,

$$\mathrm{mldeg}(\mathcal{M}_{3,3,K}) = 26 \cdot 6^K - 60 \cdot 5^K + 36 \cdot 4^K + 6 \cdot 3^K - 9 \cdot 2^K + 1;$$

in particular $\mathrm{mldeg}(\mathcal{M}_{3,3,3}) = 511$.

Proof. Write $A = [u \mid v \mid w]$ with $u = (1,1,1)^\top$ and $v = (1, v_2, v_3)^\top$, and regard e_1, e_2, e_3, u, v, w as points of $\mathbb{P}^2$; M_A is their matroid. We apply (1) to each set $\{A : \beta(M_A) = s\}$ and the projection $A \mapsto v$; the fibers of the projection are copies of the torus $(\mathbb{C}^*)^2$ of points w. For each v let $\varphi_v(s)$ be the Euler characteristic of the set of w in the torus with $\beta(M_A) = s$. We show below that φ_v depends only on which of three types v belongs to; then $W_{3,3}(s)$ is the sum over the types of the Euler characteristic of the set of such v times $\varphi_v(s)$. We record φ_v via the exponential polynomial $\Phi_v(K) = \sum_s \varphi_v(s) s^K$.

Let $P_v = \{e_1, e_2, e_3, u, v\}$. Since u and v have no zero coordinate, three points of P_v are collinear only if they are e_i, u, v for some i, that is, if v lies on one of the lines $v_2 = v_3$, $v_3 = 1$, $v_2 = 1$ through u and e_1, e_2, e_3, respectively, which meet only at u. This gives three types of v: the three sets S_i of $v \neq u$ on the line through e_i and u, each isomorphic to $\mathbb{C}^* \setminus \{1\}$, with $\chi(S_i) = -1$; the point $v = u$; and the complement S_0 of the three lines, with $\chi(S_0) = \chi((\mathbb{C}^*)^2) - 3\chi(S_1) - 1 = 2$.

For fixed v, let $\mathcal{A}_v$ be the arrangement of lines spanned by pairs of points of P_v. The matroid M_A is determined by the set of lines of $\mathcal{A}_v$ that contain w, since w equals a point a of P_v exactly when it lies on two lines through a. Two lines of $\mathcal{A}_v$ share at most one point of P_v; if three lines passed through a point outside P_v, the sets of points of P_v on them would be pairwise disjoint, which needs six points. Hence the points lying on at least two lines of $\mathcal{A}_v$, which we call intersection points, are the points of P_v and, outside P_v, the points where two lines containing disjoint sets of points of P_v meet, each of which lies on exactly two lines. By additivity, the open part of a line ℓ, the set of points on ℓ and on no other line, has Euler characteristic 2 minus the number of intersection points on ℓ, and the complement of $\mathcal{A}_v$ has Euler characteristic 3 minus the sum of the Euler characteristics of the open parts minus the total number of intersection points. The point w lies in the torus exactly when it avoids the three coordinate lines e_ie_j; in particular, since each intersection point outside P_v lies on exactly two lines, it lies in the torus unless one of its two lines is a coordinate line. All the counts below are combinatorial, so φ_v is constant on each type of v.

If $v \in S_0$, no three points of P_v are collinear, $\mathcal{A}_v$ has 10 lines, and there are 15 intersection points outside P_v, so that each line contains five intersection points, two of them in P_v; nine of them lie on the coordinate lines, each of which meets the three lines spanned by the other three points. By (3) with the parallel-element rule, the strata of $w \in (\mathbb{C}^*)^2$ are:

- the complement of $\mathcal{A}_v$ ($\chi = 3 + 30 - 20 = 13$, $\beta = 6$);
- the open parts of the seven noncoordinate lines ($\chi = -3$ each, $\beta = 5$);
- the six intersection points in the torus ($\chi = 6$, $\beta = 4$);
- the two points $w \in \{u, v\}$ ($\chi = 2$, $\beta = 3$, for five points with no three collinear).

Thus $\Phi_v(K) = 13 \cdot 6^K - 21 \cdot 5^K + 6 \cdot 4^K + 2 \cdot 3^K$.

If $v \in S_i$, permuting the coordinates of $\mathbb{P}^2$ we may take $i = 1$. Then $\mathcal{A}_v$ consists of the line m through e_1, u, v, the three coordinate lines, and the four lines e_2u, e_2v, e_3u, e_3v. The pairs of lines with disjoint sets of points of P_v give seven intersection points outside P_v: $m \cap e_2e_3$, the four points where e_1e_2 meets e_3u, e_3v and e_1e_3 meets e_2u, e_2v, and the two points $e_2u \cap e_3v$

and $e_2v \cap e_3u$, which lie in the torus. The line e_2e_3 contains three intersection points and every other line four. By (3) with the parallel-element rule, the strata of $w \in (\mathbb{C}^*)^2$ are:

- the complement of $\mathcal{A}_v$ ($\chi = 3 + 15 - 12 = 6$, $\beta = 5$);
- the open parts of e_2u, e_2v, e_3u, e_3v ($\chi = -2$ each, $\beta = 4$);
- the open part of m ($\chi = -2$, $\beta = 3$) and the two intersection points in the torus ($\chi = 2$, $\beta = 3$);
- the two points $w \in \{u, v\}$ ($\chi = 2$, $\beta = 2$).

Thus $\Phi_v(K) = 6 \cdot 5^K - 8 \cdot 4^K + 2 \cdot 2^K$, the terms in 3^K canceling.

If $v = u$, then $\beta(M_A) = \beta(M_{[u|w]})$ by the parallel-element rule, so $\Phi_u(K)$ is $\sum_s W_{3,2}(s)s^K = 2 \cdot 3^K - 3 \cdot 2^K + 1$, by Corollary 4.3 and the symmetry $W_{3,2} = W_{2,3}$ noted after Corollary 4.2.

Adding 2 times the first of these, -3 times the second, and the third gives the stated weights, and Theorem 2.1 with $(-1)^{(I-1)(J-1)} = 1$ gives the formula. □

For $J = 4$ we evaluate the weights by counting points over finite fields. The numbers of points over $\mathbb{F}_q$ of the strata $\{A \in T_{3,4} : \beta(M_A) = s\}$, with $\beta(M_A)$ computed from the ranks of submatrices over $\mathbb{F}_q$, agree for each s with a polynomial in q of degree at most 6 for all primes $7 \le q \le 37$, including at least two primes beyond those needed to determine it. If the numbers of points are given by these polynomials for all sufficiently large finite fields, the value of the polynomial at 1 is the Euler characteristic of the stratum, by a theorem of Katz [12, Thm. 6.1.2.3; see Thm. 2.1.8]. The entries for $J = 4$ in Table 1, and the formula and values derived from them below, rest on this polynomiality, which we have verified for the primes listed.

TABLE 1. The weights $W_{3,J}(s)$, the Euler characteristics of the sets of A with $\beta(M_A) = s$. For $J = 2$ they are those of Corollary 4.3, using $W_{3,2} = W_{2,3}$; for $J = 3$ they are given by Proposition 5.1; for $J = 4$ they are the values at $q = 1$ of the point counts.

s	1	2	3	4	5	6	7	8	9	10
$W_{3,2}(s)$	1	−3	2							
$W_{3,3}(s)$	1	−9	6	36	−60	26				
$W_{3,4}(s)$	1	−21	14	204	−288	252	−2088	4878	−4224	1272

With these weights, Theorem 2.1 gives

$$\begin{aligned}\operatorname{mldeg}(\mathcal{M}_{3,4,K}) = {} & 1272 \cdot 10^K - 4224 \cdot 9^K + 4878 \cdot 8^K - 2088 \cdot 7^K + 252 \cdot 6^K \\ & - 288 \cdot 5^K + 204 \cdot 4^K + 14 \cdot 3^K - 21 \cdot 2^K + 1.\end{aligned}$$

Table 2 lists the first values of this formula and of Proposition 5.1.

TABLE 2. Maximum likelihood degrees of the $3 \times 3 \times K$ and $3 \times 4 \times K$ models.

K	2	3	4	5	6
$3 \times 3 \times K$	31	511	5755	52711	426811
$3 \times 4 \times K$	115	5755	172951	3797491	68382295

These values are consistent with the symmetry of the model. For $K = 2$ they agree with the $2 \times 3 \times 3$ and $2 \times 3 \times 4$ values of Corollary 4.3, and the $3 \times 4 \times 3$ and $3 \times 3 \times 4$ models, computed

from different weights, both give 5755. The weights $W_{4,3}$, computed in the same way, coincide with $W_{3,4}$, as the remark following Corollary 4.2 requires. For $K = 1$ every formula gives 1, the maximum likelihood degree of the saturated model.

6. The leading term

The leading coefficient is an Euler characteristic that appears in the theory of generalized scattering amplitudes. Let $X(I, n)$ be the moduli space of n points in linearly general position in $\mathbb{P}^{I-1}$, the quotient by $(\mathbb{C}^*)^n$ of the open subset of the Grassmannian $\mathrm{Gr}(I, n)$ where all Plücker coordinates are nonzero [6]. It is a smooth very affine variety of dimension $(I-1)(n-I-1)$.

Proposition 6.1. *For every $A \in T_{I,J}$, $\beta(M_A) \leq \binom{I+J-2}{I-1}$, with equality if and only if every minor of A is nonzero, and the set of such A is isomorphic to $X(I, I+J)$. Consequently, as $K \to \infty$,*

$$\mathrm{mldeg}(\mathcal{M}_{I,J,K}) = \mathrm{mldeg}\big(X(I, I+J)\big)\binom{I+J-2}{I-1}^K + O\Big(\Big(\binom{I+J-2}{I-1} - 1\Big)^K\Big).$$

Proof. By [8, Thm. 1.3], $\beta(M_A)$ is the maximum likelihood degree of the independence model of an $I \times J$ table scaled by A, whose degree is $\binom{I+J-2}{I-1}$. By [2, Thm. 2] the maximum likelihood degree is smaller than the degree exactly when the scaling is a zero of the principal $\mathcal{A}$-determinant, which for $\mathbb{P}^{I-1} \times \mathbb{P}^{J-1}$ is the product of all minors of the scaling matrix [8, Prop. 2.1]. A point of $\mathrm{Gr}(I, I+J)$ with all Plücker coordinates nonzero is the row space of a unique matrix $[\mathrm{Id}_I \mid A]$, whose maximal minors are, up to sign, the minors of A; scaling the columns of $[\mathrm{Id}_I \mid A]$ and then restoring the identity block by scaling rows acts on A by row and column scaling. Hence the set of $A \in T_{I,J}$ with all minors nonzero is $X(I, I+J)$, and its Euler characteristic is the coefficient of the largest base in Theorem 2.1. Since $\dim X(I, I+J) = (I-1)(J-1)$, Huh's theorem gives $(-1)^{(I-1)(J-1)}\chi(X(I, I+J)) = \mathrm{mldeg}(X(I, I+J))$. The other bases in Theorem 2.1 are integers smaller than $\binom{I+J-2}{I-1}$, which gives the error term. □

The leading coefficient $\mathrm{mldeg}(X(I, I+J))$ is positive in every case in which it is known. For $I = 2$, $X(2, J+2)$ is the moduli space $\mathcal{M}_{0,J+2}$ of $J+2$ points on the line, whose maximum likelihood degree is $(J-1)!$, the top term of Corollary 4.3. For $I = 3$, the maximum likelihood degrees of $X(3, n)$ for $n = 5, \ldots, 9$ are 2, 26, 1272, 188112, and 74570400 [6, 1]; they are the numbers of solutions of the scattering equations on $X(3, n)$. The first three are the last entries of the rows of Table 1, and the next two are the leading coefficients of $\mathrm{mldeg}(\mathcal{M}_{3,5,K})$ and $\mathrm{mldeg}(\mathcal{M}_{3,6,K})$. In particular, a closed formula for the maximum likelihood degree of the $3 \times J \times K$ model would give one for the Euler characteristic of $X(3, J+3)$. The duality $X(I, n) \cong X(n-I, n)$ [6] identifies $X(I, I+J)$ with $X(J, I+J)$, in accordance with the symmetry of the model in I and J; the first case not covered by $I \leq 3$ is the self-dual $X(4, 8)$, which governs the $4 \times 4 \times K$ models, and whose maximum likelihood degree was computed numerically to be 5211816, with a certified lower bound of 5211598 [1, Thm. 6.1].

References

[1] D. Agostini, T. Brysiewicz, C. Fevola, L. Kühne, B. Sturmfels, and S. Telen, *Likelihood degenerations*, with an appendix by T. Lam, Adv. Math. **414** (2023), Paper No. 108863, doi:10.1016/j.aim.2023.108863.

[2] C. Améndola, N. Bliss, I. Burke, C. R. Gibbons, M. Helmer, S. Hoşten, E. D. Nash, J. I. Rodriguez, and D. Smolkin, *The maximum likelihood degree of toric varieties*, J. Symbolic Comput. **92** (2019), 222–242, doi:10.1016/j.jsc.2018.04.016; numbering as in arXiv:1703.02251.

[3] M. W. Birch, *Maximum likelihood in three-way contingency tables*, J. Roy. Statist. Soc. Ser. B **25** (1963), no. 1, 220–233, doi:10.1111/j.2517-6161.1963.tb00504.x.

[4] Y. M. M. Bishop, S. E. Fienberg, and P. W. Holland, *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, MA, 1975; reprinted by Springer, New York, 2007, doi:10.1007/978-0-387-72806-3.
[5] T. Brysiewicz and A. Maraj, *Lawrence lifts, matroids, and maximum likelihood degrees*, Algebr. Stat. **16** (2025), no. 2, 217–242, doi:10.2140/astat.2025.16.217; numbering as in arXiv:2310.13064.
[6] F. Cachazo, N. Early, A. Guevara, and S. Mizera, *Scattering equations: from projective spaces to tropical Grassmannians*, J. High Energy Phys. (2019), no. 6, 039, doi:10.1007/JHEP06(2019)039.
[7] S. E. Cappell, L. Maxim, and J. L. Shaneson, *Euler characteristics of algebraic varieties*, Comm. Pure Appl. Math. **61** (2008), no. 3, 409–421, doi:10.1002/cpa.20201; arXiv:math/0606654.
[8] O. Clarke, S. Hoşten, N. Kushnerchuk, and J. Oldekop, *Matroid stratification of ML degrees of independence models*, Algebr. Stat. **15** (2024), no. 2, 199–223, doi:10.2140/astat.2024.15.199; numbering as in arXiv:2312.10010.
[9] H. H. Crapo, *A higher invariant for matroids*, J. Combinatorial Theory **2** (1967), 406–417, doi:10.1016/S0021-9800(67)80051-6.
[10] I. Csiszár, *I-divergence geometry of probability distributions and minimization problems*, Ann. Probab. **3** (1975), no. 1, 146–158, doi:10.1214/aop/1176996454.
[11] W. E. Deming and F. F. Stephan, *On a least squares adjustment of a sampled frequency table when the expected marginal totals are known*, Ann. Math. Statist. **11** (1940), no. 4, 427–444, doi:10.1214/aoms/1177731829.
[12] T. Hausel and F. Rodriguez-Villegas, *Mixed Hodge polynomials of character varieties*, with an appendix by N. M. Katz, Invent. Math. **174** (2008), no. 3, 555–624, doi:10.1007/s00222-008-0142-x.
[13] J. Huh, *The maximum likelihood degree of a very affine variety*, Compos. Math. **149** (2013), no. 8, 1245–1266, doi:10.1112/S0010437X13007057.
[14] S. Sivarajan, *An explicit determinantal formula for limits of iterative proportional fitting with a binary variable*, hexagon:2611.00705 (2026).
[15] O. Ya. Viro, *Some integral calculus based on Euler characteristic*, in: Topology and Geometry, Rohlin Seminar, Lecture Notes in Math. **1346**, Springer, Berlin, 1988, 127–138, doi:10.1007/BFb0082775.