hexagon:2611.00734v2

# 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, and the formula extends to tables with structural zeros under a connectivity condition on the pattern of zeros. We evaluate the integral for $3 \times J$ matrices with $J = 3, 4, 5, 6$, which gives the maximum likelihood degree of the $3 \times J \times K$ model as an explicit exponential polynomial in $K$; for instance, the $3 \times 3 \times 3$ model has maximum likelihood degree 511. The case $J = 3$ is proved; the cases $J = 4, 5, 6$ rest on point counts over finite fields and on numerical values of Euler characteristics of matroid strata.

## 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 [5, 6]. Its maximum likelihood estimate has no closed form and is computed iteratively, for instance by iterative proportional fitting [15, 14]. 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 [18, 3]. 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 $2^K - 1$ [3, 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 [7, 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 [7, 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 [19].

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

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

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 $[\mathrm{Id}_I \mid A]$, $\beta$ is Crapo's beta invariant [13], and the integral is with respect to the Euler characteristic [20]. By a theorem of Clarke, Hoşten, Kushnerchuk, and Oldekop [10], $\beta(M_A)$ 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 [18] 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 5.1). The formula recovers the result of Brysiewicz and Maraj (Corollary 5.3): 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 5.2). The formula extends to tables with structural zeros, under a connectivity condition on the pattern of zeros (Theorem 4.1); for a $3 \times 3 \times K$ table with the cells $(1, 1, k)$ empty, it gives $6 \cdot 5^K - 11 \cdot 4^K + 4 \cdot 3^K + 2 \cdot 2^K - 1$, assuming that the relevant point counts are polynomial.

Since $\beta(M_A)$ depends only on the matroid, the integral is a finite sum of Euler characteristics of matroid strata. For $3 \times 3$ 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 6.1); in particular the $3 \times 3 \times 3$ model has maximum likelihood degree 511. For $3 \times 4$ and $3 \times 5$ matrices we evaluate the integral by counting points over finite fields, which gives formulas for the $3 \times 4 \times K$ and $3 \times 5 \times K$ models (Section 6). For $3 \times 4$ matrices the counts are polynomial in $q$ for all primes tested. For $3 \times 5$ matrices they are not: they depend on $q$ modulo 3, through configurations of eight points with seven of the eight lines of the Möbius–Kantor configuration, whose realization spaces are open subsets of nodal cubic curves with branches at the node defined over $\mathbb{Q}(\sqrt{-3})$; assuming that the counts have this form for all large fields, the Euler characteristic is the value at $q = 1$ of the count for $q \equiv 1 \pmod 3$. For $3 \times 6$ matrices we build on the Euler characteristics of the strata of nine points in the plane computed by Agostini, Brysiewicz, Fevola, Kühne, Sturmfels, and Telen [1], together with symbolic computations for some of the strata, and obtain the maximum likelihood degree of every $3 \times 6 \times K$ model (Section 6.4). 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 [8, 1]; hence the leading coefficient of the maximum likelihood degree in $K$ is the number of solutions of these equations (Proposition 7.1). At the other end, for each fixed $s$ the weight of $s^K$ is an exponential polynomial in $I$ and $J$ given by a double sum of Stirling numbers, which we evaluate for $s \leq 3$ in every format (Proposition 8.1).

## 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 [13],

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

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 $\beta$ of a connected matroid on at least two elements is positive [13].

For a constructible function $\varphi$ on a complex algebraic variety $Y$ that takes finitely many values, $\int_Y \varphi\, d\chi = \sum_t t\, \chi(\varphi^{-1}(t))$, where $\chi$ is the topological Euler characteristic [20].

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

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

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

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

When $T = \mathbf{1}$, $\beta(M_A)$ is the maximum likelihood degree of the independence model of an $I \times J$ table scaled by $A$ [10, 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 $\beta$, Theorem 2.1 reads

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

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 [20]. Viro's calculus uses the Euler characteristic with compact supports, which for complex algebraic varieties coincides with $\chi$ (see [9] and the references there). In particular $\chi$ 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 $\psi$ 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 [18, Thm. 1] (see also [10, 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 $(y', w, v) \mapsto y$, where $y_k = y'_k$ for $k \notin P$ and $y_k = v_k \cdot w_k$ for $k \in P$. By Lemma 3.2 the left side is

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

**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 $\beta$ [13]: $\beta(M) = \beta(M \setminus e) + \beta(M/e)$ 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 $r - 1 \geq 1$. Moreover, $U_{M\setminus e}$ 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 $\chi(U_M) = \chi(U_{M\setminus e}) - \chi(U_{M/e}) = (-1)^{r-1}\big(\beta(M \setminus e) + \beta(M/e)\big)$. $\square$

**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 $\gamma_j$ 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, $\chi(U_A) = (-1)^{I-1}\beta(M_A)$. $\square$

*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 $\gamma$ 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 $d + (I + J - 1)K = (I - 1)(J - 1) + 2(I + J - 1)K \equiv (I - 1)(J - 1) \pmod 2$. $\square$

## 4. Structural zeros

In an incomplete contingency table some cells are empty by design, and the log-linear model is restricted to the remaining cells [6, Ch. 5]. For two-way tables these are the quasi-independence models; Coons and Sullivant characterized those with maximum likelihood degree one [11]. For the no-three-way interaction model with one binary variable and the same structural zeros in both of its levels, Brysiewicz and Maraj expressed the maximum likelihood degree as an evaluation of a Tutte polynomial [7, Cor. 41]. Theorem 2.1 extends to arbitrary patterns of structural zeros, subject to a connectivity condition.

Let $Z \subseteq [I]\times[J]\times[K]$ be a set of cells and $C$ its complement. For $T \in (\mathbb{C}^*)^C$ let $X_T^Z \subseteq \mathbb{P}^{|C|-1}$ be the closure of the image of

$$(a,b,c) \longmapsto \big(t_{ijk}a_{ij}b_{ik}c_{jk}\big)_{(i,j,k)\in C},$$

where $a$ runs over the coordinates $a_{ij}$ with $(i,j,k) \in C$ for some $k$, and $b$ and $c$ likewise. The maximum likelihood degree $\mathrm{mldeg}(X_T^Z)$ is defined as in Section 2, with the coordinate hyperplanes and the hyperplane $\sum_{c\in C} p_c = 0$ of $\mathbb{P}^{|C|-1}$ removed. Extend $T$ by $t_{ijk} = 0$ for $(i,j,k) \in Z$, so that its slices $T_k$ are $I \times J$ matrices that may have zero entries. For each $k$ let $\Gamma_k$ be the bipartite graph on the vertex set $[I] \sqcup [J]$ with an edge $\{i,j\}$ when $t_{ijk} \neq 0$, and let $S \subseteq [I] \times [J]$ be the set of pairs that are edges of some $\Gamma_k$. We assume:

(C) every $\Gamma_k$ is connected, and the graph on $[K]$ is itself connected, where $k$ and $l$ are adjacent when the graph with edge set $\Gamma_k \cap \Gamma_l$ on all of $[I] \sqcup [J]$ is connected.

For instance, (C) holds when every $\Gamma_k$ is connected and either the zeros are the same in every slice or some slice has no zeros.

The torus $(\mathbb{C}^*)^I \times (\mathbb{C}^*)^J$ of row and column scalings acts on $(\mathbb{C}^*)^S$, the matrices supported on $S$, with a one-dimensional stabilizer, since the bipartite graph with edge set $S$ is connected. Fix a spanning tree $\tau \subseteq S$ of this graph. Every orbit meets the set of matrices with $a_{ij} = 1$ for $(i,j) \in \tau$ in exactly one point, and we identify the quotient with this torus $T_S \cong (\mathbb{C}^*)^{|S|-I-J+1}$. For $Z = \emptyset$ and $\tau$ the first row and column, $T_S = T_{I,J}$. As before, $M_{T_k\circ A}$ is the matroid of the columns of $[\mathrm{Id}_I \mid T_k \circ A]$, and $A \mapsto \prod_k \beta(M_{T_k\circ A})$ is constructible on $T_S$.

**Theorem 4.1.** *Under condition* (C)*,*

$$\mathrm{mldeg}(X_T^Z) = (-1)^{|S|-I-J+1} \int_{T_S} \prod_{k=1}^{K} \beta(M_{T_k\circ A})\, d\chi(A).$$

*Proof.* Let $T'$ be the torus with coordinates $a_{ij}$ for $(i,j) \in S \setminus \tau$, $b_{ik}$ for $i \geq 2$, and $c_{jk}$ for all $j,k$; put $a_{ij} = 1$ for $(i,j) \in \tau$ and $b_{1k} = 1$. By (C), every $b_{ik}$ and $c_{jk}$ occurs in some coordinate of the parametrization, since $\Gamma_k$ is connected and so has no isolated vertex. We show that $\psi\colon T' \to (\mathbb{C}^*)^C$, $x \mapsto (t_{ijk}a_{ij}b_{ik}c_{jk})$, is injective. Suppose $\psi(x) = \psi(x')$, and put $\xi_{ij} = a'_{ij}/a_{ij}$, $\eta_{ik} = b'_{ik}/b_{ik}$, $\zeta_{jk} = c'_{jk}/c_{jk}$, so that $\xi_{ij}\eta_{ik}\zeta_{jk} = 1$ for $(i,j,k) \in C$, $\xi = 1$ on $\tau$, and $\eta_{1k} = 1$. For fixed $k$, $\xi_{ij} = \lambda_i^{(k)}\mu_j^{(k)}$ on the edges of $\Gamma_k$, with $\lambda_i^{(k)} = \eta_{ik}^{-1}$ and $\mu_j^{(k)} = \zeta_{jk}^{-1}$. Since $\Gamma_k$ is connected and contains every vertex, the restriction of $\xi$ to its edges determines $(\lambda^{(k)}, \mu^{(k)})$ up to $(s\lambda^{(k)}, s^{-1}\mu^{(k)})$ with $s \in \mathbb{C}^*$. If $k$ and $l$ are adjacent in the graph of (C), the same argument on $\Gamma_k \cap \Gamma_l$ shows that $(\lambda^{(l)}, \mu^{(l)}) = (s\lambda^{(k)}, s^{-1}\mu^{(k)})$ for some $s$. By the connectivity in (C) there is therefore a single pair $(\lambda, \mu)$ with $\xi_{ij} = \lambda_i\mu_j$ for all $(i,j) \in S$. Since $\xi = 1$ on the spanning tree $\tau$, $\lambda$ and $\mu^{-1}$ are constant and $\xi = 1$. Then $\eta_{ik}\zeta_{jk} = 1$ on the edges of each $\Gamma_k$, so by connectivity $\eta_{ik} = \nu_k$ and $\zeta_{jk} = \nu_k^{-1}$, and $\eta_{1k} = 1$ gives $\nu_k = 1$. Hence $x = x'$.

The map $\psi$ is the composition of a homomorphism of tori with translation by $T$; a homomorphism of complex algebraic tori that is injective on points has trivial kernel, so it is

an isomorphism onto a closed subtorus. Every $(a, b, c)$ 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 the coordinates of the parametrization, so as to satisfy $a = 1$ on $\tau$ and $b_{1k} = 1$. The proof of Lemma 3.1 then applies verbatim: $X_T^Z \setminus \mathcal{H}$ is isomorphic to the smooth hypersurface $V = \{f = 1\}$ in $T'$, where

$$f = \sum_{k=1}^{K} b_{\cdot k}^{\top}(T_k \circ A)\, c_{\cdot k},$$

of dimension $d-1$ with $d = |S| - I - J + 1 + (I-1)K + JK$, and (2) holds with $X_T$ replaced by $X_T^Z$.

The proof of Lemma 3.5 uses only that the columns of $[\mathrm{Id}_I \mid A]$ span $\mathbb{C}^I$, so it holds for every $I \times J$ matrix $A$, with $M_A$ the matroid of these columns. Projecting $V$ to $A \in T_S$, the fiber over $A$ is $\{(b, c) : \sum_k g_k(b_{\cdot k}, c_{\cdot k}) = 1\}$, with $g_k$ the function of Lemma 3.5 for $T_k \circ A$. As in the proof of Theorem 2.1, Lemmas 3.3 and 3.5 give the fiber Euler characteristic $-(-1)^{(I+J-1)K} \prod_k \beta(M_{T_k \circ A})$, and by (1) and (2),

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

with $d + (I+J-1)K \equiv |S| - I - J + 1 \pmod 2$. □

Condition (C) cannot be omitted. For positive $T$ the maximum likelihood degree of $X_T^Z$ is positive: for positive data the log-likelihood tends to $-\infty$ at the boundary of the closure in the probability simplex of the positive part of $X_T^Z$, so it attains a maximum, and hence has a critical point, in the interior. On the other hand, let $I = J = 3$ and $K = 2$. For $Z = \{(1,2,1), (1,3,1), (2,1,1), (3,1,1)\}$ the graph $\Gamma_1$ is disconnected, so every $M_{T_1 \circ A}$ is disconnected and has $\beta = 0$, and the integral vanishes. For $Z = \{(1,1,1), (1,2,1), (2,3,2), (3,3,2)\}$ both graphs are connected but $\Gamma_1 \cap \Gamma_2$ is not, and the integral vanishes again: the entry $a_{13}$ is the only nonzero entry of the first row of $T_1 \circ A$ and of the third column of $T_2 \circ A$, so scaling it alone is a row scaling of $T_1 \circ A$ and a column scaling of $T_2 \circ A$, which preserves both matroids; with a spanning tree $\tau$ avoiding $(1,3)$, this is a free action of $\mathbb{C}^*$ on $T_S$ preserving the integrand, and every level set of the integrand has Euler characteristic 0.

**Example 4.2.** Let $I = J = 3$ and $Z = \{(1,1,k) : k \in [K]\}$, a line of structural zeros, with $T = \mathbf{1}$ on $C$. Then (C) holds, $S$ consists of the pairs other than $(1,1)$, and $T_S$ is three-dimensional; with $\tau = \{(1,2), (1,3), (2,1), (3,1), (2,2)\}$,

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & y \\ 1 & x & z \end{pmatrix}, \qquad (x, y, z) \in (\mathbb{C}^*)^3,$$

and the integrand is $\beta(M_A)^K$. Let $W'(s)$ be the Euler characteristic of the set of such $A$ with $\beta(M_A) = s$. For each $s$, the number of points over $\mathbb{F}_q$ of this set, with $\beta$ computed over $\mathbb{F}_q$, agrees with

$$1, \quad 2q-4, \quad 4q-8, \quad 6q^2-29q+34, \quad (q-2)(q-3)(q-4)$$

for $s = 1, \ldots, 5$ and every prime $5 \le q \le 23$, and no other values occur. If these counts are polynomial for all large fields, Katz's theorem gives $W' = (1, -2, -4, 11, -6)$, and Theorem 4.1 gives

$$\mathrm{mldeg}(X_{\mathbf{1}}^Z) = 6 \cdot 5^K - 11 \cdot 4^K + 4 \cdot 3^K + 2 \cdot 2^K - 1,$$

which is 17, 169, 1289, and 8521 for $K = 2, 3, 4, 5$. For $K = 2$ the model is the Lawrence lifting of the quasi-independence model of the complete bipartite graph $K_{3,3}$ with one edge removed, and 17 agrees with [7, Cor. 41].

## 5. Consequences

**Corollary 5.1** (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 [3, Cor. 8]; apply Theorem 2.1. □

Brysiewicz and Maraj noted that they have no combinatorial interpretation of the degree 2457 of $\mathcal{M}_{3,3,3}$ [7, Rem. 39]; Corollary 5.1 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 $\rho_k$ are distinct, the stratum $\mathbb{C}^* \setminus \{\rho_1^{-1}, \ldots, \rho_K^{-1}\}$ has Euler characteristic $-K$, and the integral is $-K\,2^K + K\,2^{K-1} = -K\,2^{K-1}$, so $\deg(\mathcal{M}_{2,2,K}) = K\,2^{K-1}$ [3, Prop. 34].

**Corollary 5.2** (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 [13]. The symmetry between $J$ and $K$ is not apparent from the integral.

**Corollary 5.3** (Brysiewicz and Maraj [7, 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 $\beta$ 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 $(-1)^{(2-1)(J-1)}$. □

## 6. 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 [10, Thm. 1.3], which is at most the degree $\binom{J+1}{2}$ of $\mathbb{P}^2 \times \mathbb{P}^{J-1}$ [3, Cor. 8].

Table 1 lists the weights $W_{3,J}(s)$ for $J \leq 5$. For $J = 2$ they are given by Corollary 5.3; for $J = 3$ they are proved in Section 6.1; for $J = 4, 5$ they are obtained from point counts over finite fields in Sections 6.2 and 6.3. With Theorem 2.1 they give, for all $K \geq 2$ and, for $J = 4, 5$, under the hypotheses of Sections 6.2 and 6.3,

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

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

$$\begin{aligned}\text{mldeg}(\mathcal{M}_{3,5,K}) = {} & 188112 \cdot 15^K - 849600 \cdot 14^K + 1506240 \cdot 13^K - 1286040 \cdot 12^K \\ & + 487440 \cdot 11^K + 480 \cdot 10^K - 88440 \cdot 9^K + 63180 \cdot 8^K - 23040 \cdot 7^K \\ & + 1610 \cdot 6^K - 708 \cdot 5^K + 780 \cdot 4^K + 30 \cdot 3^K - 45 \cdot 2^K + 1,\end{aligned} \tag{5}$$

whose first values are listed in Table 2. For $K = 1$ every formula gives 1, the maximum likelihood degree of the saturated model. For $J = 6$ the weights are assembled from Euler characteristics of matroid strata in Section 6.4, which gives (9) and the last row of Table 2.

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 5.3, using $W_{3,2} = W_{2,3}$; for $J = 3$ they are given by Proposition 6.1; for $J = 4, 5$ they are obtained from the point counts of Sections 6.2 and 6.3.

| $s$ | $W_{3,2}(s)$ | $W_{3,3}(s)$ | $W_{3,4}(s)$ | $W_{3,5}(s)$ |
|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 |
| 2 | −3 | −9 | −21 | −45 |
| 3 | 2 | 6 | 14 | 30 |
| 4 | | 36 | 204 | 780 |
| 5 | | −60 | −288 | −708 |
| 6 | | 26 | 252 | 1610 |
| 7 | | | −2088 | −23040 |
| 8 | | | 4878 | 63180 |
| 9 | | | −4224 | −88440 |
| 10 | | | 1272 | 480 |
| 11 | | | | 487440 |
| 12 | | | | −1286040 |
| 13 | | | | 1506240 |
| 14 | | | | −849600 |
| 15 | | | | 188112 |

TABLE 2. Maximum likelihood degrees of the $3 \times J \times K$ models for $J = 3, \ldots, 6$. The rows for $J = 4, 5, 6$ rest on the hypotheses of Sections 6.2–6.4.

| $K$ | 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 |
| $3 \times 5 \times K$ | 391 | 52711 | 3797491 | 183308911 | 6785333491 |
| $3 \times 6 \times K$ | 1267 | 426811 | 68382295 | 6785333491 | 486261533527 |

6.1. **The $3 \times 3$ case.** We use two facts about $\beta$. If $M$ is a simple matroid of rank three on $n$ elements represented over $\mathbb{C}$, then

$$\beta(M) = \binom{n-2}{2} - \sum_{L} \binom{|L|-1}{2}, \tag{6}$$

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 6.1.** *The weights* $W_{3,3}(1), \ldots, W_{3,3}(6)$ *are* $1, -9, 6, 36, -60, 26$, *and* $W_{3,3}(s) = 0$ *for* $s > 6$. *Consequently,* (3) *holds for all* $K \geq 2$*; in particular* $\operatorname{mldeg}(\mathcal{M}_{3,3,3}) = 511$.

*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 $\varphi_v$ 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 $\varphi_v$ 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 requires 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 $\ell$, the set of points on $\ell$ and on no other line, has Euler characteristic 2 minus the number of intersection points on $\ell$, 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 $\varphi_v$ 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 the 15 lie on the coordinate lines, each of which meets the three lines spanned by the other three points. By (6) 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 (6) 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 5.3 and the symmetry $W_{3,2} = W_{2,3}$ noted after Corollary 5.2.

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. □

6.2. **The $3 \times 4$ case.** 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; these polynomials are listed in Appendix A. 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 [17, Thm. 6.1.2.3; see Thm. 2.1.8]. The entries for $J = 4$ in Table 1, and hence (4) and the $3 \times 4 \times K$ row of Table 2, rest on this polynomiality, which we have verified for the primes listed.

6.3. **The $3 \times 5$ case.** For $J = 5$ the numbers of points are not polynomial in $q$. Let $\rho(q)$ be the number of roots of $x^2 + x + 1$ in $\mathbb{F}_q$, which is 2, 0, or 1 according as $q \equiv 1$, 2, or 0 (mod 3). For each $s$, the number of points over $\mathbb{F}_q$ of $\{A \in T_{3,5} : \beta(M_A) = s\}$ equals $P_s(q) + c_s\,\rho(q)$, with $P_s \in \mathbb{Z}[q]$ of degree at most 8 and $c_s \in \mathbb{Z}$ listed in Appendix A, for every odd prime power $q \le 31$ and for $q = 37$. The ten fields of characteristic at least 5 with $q \le 31$ determine $P_s$ and $c_s$, and the three fields of characteristic 3 and the field with 37 elements agree with them. The constants $c_s$ vanish for $s \le 7$.

Each stratum is a finite union of locally closed subschemes of the torus over $\mathbb{Z}$, defined by the vanishing and nonvanishing of minors, and its points over $\mathbb{F}_q$ are the $A \in T_{3,5}(\mathbb{F}_q)$ with

$\beta(M_A) = s$ computed over $\mathbb{F}_q$. Let $\omega$ be a primitive cube root of unity. Every finite field receiving a homomorphism from $\mathbb{Z}[\omega, 1/3]$ has $q \equiv 1 \pmod 3$, and for $q = 7$, 13, 19, 25, 31, and 37 the count is $P_s(q)+2c_s$. If this holds for all sufficiently large such fields, Katz's theorem, applied over $\mathbb{Z}[\omega, 1/N]$ for a suitable $N$, gives the Euler characteristic $W_{3,5}(s) = P_s(1) + 2c_s$. The entries for $J = 5$ in Table 1, and hence (5) and the $3 \times 5 \times K$ row of Table 2, rest on this hypothesis.

The term $c_s\,\rho(q)$ comes from configurations of the following kind. Write the columns of $A$ as $u, c_2, \dots, c_5$. Among the matroids with $\beta = 8$ are those with seven three-point lines that, together with the noncollinear triple $\{e_1, e_2, e_3\}$, form the incidence structure of the Möbius–Kantor configuration $(8_3)$ [16]. For the one with lines $e_1uc_4$, $e_1c_3c_5$, $e_2c_2c_5$, $e_2c_3c_4$, $e_3uc_5$, $e_3c_2c_4$, and $uc_2c_3$, the matrices are given by $c_2 = (1, a, b)^\top$, $c_3 = (1, a/b, a)^\top$, $c_4 = (1, a, a)^\top$, and $c_5 = (1, 1, b)^\top$, with $(a, b)$ on the cubic $b^2 + (a^2 - 3a)b + a = 0$. This curve is rational, and its only singular point is $(1, 1)$, where all columns coincide; there its tangent cone is $x^2 - xy + y^2 = 0$ in the coordinates $x = a - 1$, $y = b - 1$, so that the singular point is a node over $\mathbb{C}$ (a cusp in characteristic 3) and has $\rho(q)$ preimages over $\mathbb{F}_q$ in the normalization. The stratum is the normalization $\mathbb{P}^1$ with these preimages and three further rational points removed, so it has $q - 2 - \rho(q)$ points over $\mathbb{F}_q$. There are 120 such labeled matroids, and for the fields computed each of their strata has $q - 2 - \rho(q)$ points, which accounts for $c_8 = -120$.

The values in Table 2 are consistent with the symmetry of the model. For $K = 2$ they agree with the $2 \times 3 \times 3$, $2 \times 3 \times 4$, and $2 \times 3 \times 5$ values of Corollary 5.3; the $3 \times 4 \times 3$ and $3 \times 3 \times 4$ models, computed from different weights, both give 5755; and the $3 \times 5 \times 3$ and $3 \times 5 \times 4$ models give the values 52711 and 3797491 of the $3 \times 3 \times 5$ and $3 \times 4 \times 5$ models. The numbers of points for $4 \times 3$ matrices coincide with those for $3 \times 4$ matrices at every prime listed, so that $W_{4,3} = W_{3,4}$, as the remark following Corollary 5.2 requires. The coefficient 188112 of $15^K$ in (5) is the maximum likelihood degree of $X(3, 8)$, as Proposition 7.1 requires. The other identities in this paragraph hold with $c_s$ replaced by any multiple of it, since $\sum_s c_s s^K = 0$ for $K \le 4$; the coefficient 188112 is the one that tests the factor 2 in $P_s(1) + 2c_s$.

6.4. **The $3 \times 6$ case.** For $J = 6$ we assemble the weights from Euler characteristics of matroid strata instead of counting points.

Write $T^\circ_{3,m} \subseteq T_{3,m}$ for the matrices with pairwise distinct columns, and let $V_m(s)$ be the Euler characteristic of $\{A \in T^\circ_{3,m} : \beta(M_A) = s\}$. Grouping the columns of $A \in T_{3,J}$ into classes of equal columns, and using that $\beta$ is unchanged by deleting a parallel element (Section 6.1), gives

$$W_{3,J}(s) = \sum_{m=1}^{J} S(J, m)\, V_m(s), \tag{7}$$

where $S(J, m)$ is the Stirling number of the second kind. Summing over $s$, $\sum_s V_m(s) = (-1)^{m-1}(m-1)!$ is the Euler characteristic of the space of $m-1$ distinct points in $(\mathbb{C}^*)^2 \backslash \{(1, 1)\}$.

For $A \in T^\circ_{3,m}$, the points $e_1, e_2, e_3$ and the columns of $A$ are $m + 3$ distinct points of $\mathbb{P}^2$, of which $e_1, e_2, e_3$ and the first column $(1, 1, 1)^\top$ form a projective frame, and no column lies on a line $e_ie_j$. Hence $T^\circ_{3,m}$ is the disjoint union, over the simple matroids $M$ of rank three on these labeled points satisfying these two conditions, of the realization spaces of $M$ with the frame fixed, which are the strata $X(M)$ of the moduli space of $m + 3$ points in $\mathbb{P}^2$ studied by Agostini, Brysiewicz, Fevola, Kühne, Sturmfels, and Telen [1]. Grouping the labeled matroids

by isomorphism type,

$$V_m(s) = \sum_{\beta(M)=s} \lambda(M)\,\chi(X(M)), \tag{8}$$

the sum over isomorphism types of simple rank-three matroids on $m+3$ elements, where $\lambda(M)$ is the number of admissible labelings divided by the order of the automorphism group of $M$. Since the realization spaces of rank-three matroids on at most eleven elements realizable over $\mathbb{C}$ are smooth [12], $\chi(X(M)) = (-1)^{\dim X(M)}\,\mathrm{mldeg}(X(M))$ by Huh's theorem [18], and Agostini et al. computed these maximum likelihood degrees numerically for all matroids on at most nine elements [1, Thm. 4.2] and [2]. For $m \le 5$, (8) with their values agrees with the weights $V_m$ obtained from (7) and Table 1.

For $m = 6$ we use these values together with symbolic computations for the thirteen strata on nine elements listed in Table 3. For each, an elimination order of the nine points (a frame, then each point either free, on one line through two earlier points, or at the intersection of two such lines) writes $X(M)$, with all coordinates units, either as the complement of a hypersurface in $\mathbb{C}^d$ or, for four of them, as an open subset of the zero set of one polynomial $g$ in $\mathbb{C}^{d+1}$: a linear form for matroid 81, and $x^2 - x + 1$ in one coordinate $x$ for matroids 107, 108, and 238, which therefore have two components, conjugate over $\mathbb{Q}(\sqrt{-3})$. In each case $X(M)$ is smooth and very affine (for the four, $g$ and its gradient have no common zero on $X(M)$), and its maximum likelihood degree is the number of solutions of the critical equations $z_i f_i = 1$, $\sum_i w_i z_i\,\partial f_i/\partial x_j = 0$, where the $f_i$ are the irreducible factors of the nonvanishing $3\times 3$ minors and the $w_i$ are random, with a Lagrange multiplier for $g$ when present. We counted the solutions modulo a large prime with msolve [4], with two choices of weights and primes for the three strata of dimension at least four; such a count equals the count over $\mathbb{C}$ for generic weights and all but finitely many primes. As a check, the same computation gives $\chi(X(M)) = 8$ for the stratum with nine three-point lines $145, 246, 368, 357, 123, 279, 189, 478, 569$, in agreement with its point counts, and reproduces the values of [2] for matroids 107, 108, 238, and 338. For matroid 19, the number of points of $X(M)$ over $\mathbb{F}_p$ equals

$$p^5 - 26p^4 + 265p^3 - 1317p^2 + 3183p - 3009 - \Big(\frac{-1}{p}\Big) + (p-25)\Big(\frac{-3}{p}\Big) - 10\Big(\frac{5}{p}\Big)$$

for all primes $7 \le p \le 211$, with Legendre symbols $(d/p)$; at $p = 1$ with all three Legendre symbols equal to 1, this gives $\chi(X(M)) = -938$, in agreement with Table 3.

The two isomorphism types with $\beta = 19$, those with two three-point lines, have maximum likelihood degrees of order $10^5$. We determine $V_6(19)$ from $\sum_s V_6(s) = -120$. The resulting weights $V_6(s)$ and $W_{3,6}(s)$ are listed in Table 4, and Theorem 2.1 gives

$$\begin{aligned}\mathrm{mldeg}(\mathcal{M}_{3,6,K}) = {} & 74570400\cdot 21^K - 438750000\cdot 20^K + 1082211840\cdot 19^K \\ & - 1432087920\cdot 18^K + 1062738000\cdot 17^K - 392337360\cdot 16^K \\ & + 15302880\cdot 15^K + 31727520\cdot 14^K + 7858080\cdot 13^K - 17962320\cdot 12^K \\ & + 7459920\cdot 11^K - 149520\cdot 10^K - 957480\cdot 9^K + 533070\cdot 8^K \\ & - 168120\cdot 7^K + 8220\cdot 6^K + 360\cdot 5^K + 2460\cdot 4^K + 62\cdot 3^K - 93\cdot 2^K \\ & + 1. \end{aligned} \tag{9}$$

For $K = 1, \ldots, 5$ this gives 1, 1267, 426811, 68382295, and 6785333491, the maximum likelihood degrees of the saturated model and of the $2\times3\times6$, $3\times3\times6$, $3\times4\times6$, and $3\times5\times6$ models from Corollary 5.3 and (3)–(5). These five identities are independent of the condition $\sum_s V_6(s) = -120$ used to determine $V_6(19)$; for $K = 6$, (9) gives $\mathrm{mldeg}(\mathcal{M}_{3,6,6}) = 486261533527$. Two

TABLE 3. Euler characteristics of strata on nine points computed with msolve modulo a large prime for $V_6$. Matroids are numbered as in [2] and given by their three- and four-point lines.

| No. | lines | dim | $\chi(X(M))$ | $\lambda(M)$ | $\beta(M)$ |
|---|---|---|---|---|---|
| 12 | $123, 145, 267, 389, 468, 579$ | 4 | 96 | 360 | 15 |
| 19 | $123, 145, 246, 378, 579$ | 5 | $-938$ | 25920 | 16 |
| 23 | $123, 145, 178, 246, 379$ | 5 | $-1312$ | 32400 | 16 |
| 81 | $123, 145, 169, 246, 258, 347, 389, 579, 678$ | 1 | $-2$ | 360 | 12 |
| 106 | $123, 145, 169, 246, 258, 347, 359, 368, 567$ | 1 | $-5$ | 2880 | 12 |
| 107 | $123, 145, 178, 246, 258, 347, 368, 567$ | 2 | 36 | 360 | 13 |
| 108 | $123, 145, 178, 246, 258, 347, 359, 368, 567$ | 1 | $-6$ | 1080 | 12 |
| 117 | $123, 145, 178, 246, 279, 356, 589$ | 3 | $-40$ | 8640 | 14 |
| 119 | $123, 145, 178, 246, 279, 356, 389$ | 3 | $-39$ | 1080 | 14 |
| 226 | $1234, 156, 189, 257, 269, 358, 379$ | 2 | 18 | 3600 | 12 |
| 238 | $1234, 156, 189, 257, 269, 358, 379, 678$ | 1 | $-6$ | 720 | 11 |
| 337 | $1234, 1567, 189, 258, 359, 368, 469, 478$ | 0 | 1 | 2160 | 9 |
| 338 | $1234, 1567, 258, 279, 359, 368, 469, 478$ | 0 | 2 | 360 | 9 |

further checks are independent of the strata data. A line contains at most two of the points $e_1, e_2, e_3, (1,1,1)^\top$, and if it contains two it is a line $e_i e_j$, which contains no column, or a line through $(1,1,1)^\top$ and some $e_i$; with (6) it follows that $\beta(M_A) \geq 6$ for $A \in T^\circ_{3,6}$, with equality exactly when all six columns lie on one of the three lines through $(1,1,1)^\top$ and some $e_i$. Hence $V_6(s) = 0$ for $s < 6$, and $V_6(6) = -360$, since each of these lines meets the torus in a set that, minus $(1,1,1)^\top$, has Euler characteristic $-1$, so that $V_6(6) = 3 \cdot (-1)(-2)(-3)(-4)(-5)$. The leading coefficient 74570400 is the maximum likelihood degree of $X(3,9)$, as Proposition 7.1 requires. Formula (9) uses the values of [1, Thm. 4.2] for the strata not in Table 3.

## 7. 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 [8]. It is a smooth very affine variety of dimension $(I-1)(n-I-1)$.

**Proposition 7.1.** *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 [10, 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 [3, 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 [10, 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

TABLE 4. The weights $V_6(s)$ of six distinct columns and $W_{3,6}(s)$, related by (7).

| $s$ | $V_6(s)$ | $W_{3,6}(s)$ |
|---|---|---|
| 1 | | 1 |
| 2 | | −93 |
| 3 | | 62 |
| 4 | | 2460 |
| 5 | | 360 |
| 6 | −360 | 8220 |
| 7 | | −168120 |
| 8 | | 533070 |
| 9 | 10080 | −957480 |
| 10 | −48600 | −149520 |
| 11 | 148320 | 7459920 |
| 12 | 1328280 | −17962320 |
| 13 | −14735520 | 7858080 |
| 14 | 44471520 | 31727520 |
| 15 | 12481200 | 15302880 |
| 16 | −392337360 | −392337360 |
| 17 | 1062738000 | 1062738000 |
| 18 | −1432087920 | −1432087920 |
| 19 | 1082211840 | 1082211840 |
| 20 | −438750000 | −438750000 |
| 21 | 74570400 | 74570400 |

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 5.3. For $I = 3$, the maximum likelihood degrees of $X(3, n)$ for $n = 5, \ldots, 9$ are 2, 26, 1272, 188112, and 74570400 [8, 1]; they are the numbers of solutions of the scattering equations on $X(3, n)$. The first four are the weights $W_{3,J}\left(\binom{J+1}{2}\right)$ of Table 1 for $J = 2, \ldots, 5$, and the last is the leading coefficient of $\mathrm{mldeg}(\mathcal{M}_{3,6,K})$. 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 $X(I, n) \cong X(n-I, n)$ [8] 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].

## 8. Small bases and open questions

The leading coefficient $\mathrm{mldeg}(X(I, I+J))$ and the arithmetic strata of Sections 6.3 and 6.4 suggest that the weights $W_{I,J}(s)$ have no elementary closed form in general. The weights at small $s$, however, are given for all $I$ and $J$ by a double sum of Stirling numbers.

Write $T^{\circ}_{I,m} \subseteq T_{I,m}$ for the matrices with pairwise distinct columns, and $T^{\circ\circ}_{\ell,m} \subseteq T_{\ell,m}$ for those with pairwise distinct rows and pairwise distinct columns; set

$$U_{\ell,m}(s) = \chi\big(\{A \in T^{\circ\circ}_{\ell,m} : \beta(M_A) = s\}\big),$$

and note that $U_{1,1}(1) = 1$ and $U_{1,m}(s) = U_{\ell,1}(s) = 0$ whenever $\ell, m \geq 2$ or $s > 1$.

**Proposition 8.1.** *For all $I, J \geq 2$ and $s \geq 1$,*

$$W_{I,J}(s) = \sum_{\ell=1}^{s} \sum_{m=1}^{s} S(I,\ell)\, S(J,m)\, U_{\ell,m}(s), \tag{10}$$

*where $U_{\ell,m}(s) = U_{m,\ell}(s)$, and $U_{\ell,m}(s) = 0$ whenever $\ell > s$ or $m > s$. In particular, for every fixed $s \geq 1$, $W_{I,J}(s)$ is a finite exponential polynomial in $I$ and $J$ with bases at most $s$, and for all $I, J \geq 2$:*

$$\begin{aligned} W_{I,J}(1) &= 1, \\ W_{I,J}(2) &= -S(I,2)\, S(J,2) = -(2^{I-1}-1)(2^{J-1}-1), \\ W_{I,J}(3) &= 2\, S(I,2)\, S(J,3) + 2\, S(I,3)\, S(J,2) - 6\, S(I,3)\, S(J,3). \end{aligned}$$

*Proof.* Grouping the columns of $A \in T_{I,J}$ into $m$ classes of equal columns and then the rows into $\ell$ classes of equal rows preserves $\beta(M_A)$ by the parallel-element rule (Section 6.1) and duality $\beta(M_A) = \beta(M_{A^\top})$ (Section 5), giving

$$W_{I,J}(s) = \sum_{\ell=1}^{I} \sum_{m=1}^{J} S(I,\ell)\, S(J,m)\, U_{\ell,m}(s),$$

with $U_{\ell,m}(s) = U_{m,\ell}(s)$.

We show by induction on $m \geq 2$ that $\beta(M_A) \geq m$ for every $A \in T^{\circ}_{I,m}$ ($I \geq 2$). For $m = 2$, $A = [u \mid v]$ with $u = \mathbf{1}$ and $v \neq \mathbf{1}$; by duality, $\beta(M_A) = \beta(M_{A^\top})$ is the number of distinct entries of $v$ (proof of Corollary 5.3), which is at least 2. For $m \geq 3$, let $c_m$ be the last column of $A$ and let $A' \in T^{\circ}_{I,m-1}$ be obtained by deleting $c_m$. Since $c_m$ has full support and $e_1, \ldots, e_I$ span $\mathbb{C}^I$, $c_m$ is neither a loop nor a coloop of $M_A$, so $\beta(M_A) = \beta(M_{A'}) + \beta(M_A/c_m)$. In $M_A/c_m$, the vectors $\bar{e}_1, \ldots, \bar{e}_I$ form a circuit (as $c_m = \sum_{i=1}^{I} a_{im} e_i$ with every $a_{im} \neq 0$) and span $\mathbb{C}^I/\langle c_m \rangle$. Every remaining column $c_j$ ($j < m$) is distinct from $c_m$ with first entry 1, so $\bar{c}_j \neq 0$ and $\bar{c}_j$ lies in a circuit of $M_A/c_m$ with a nonempty subset of $\{\bar{e}_1, \ldots, \bar{e}_I\}$. Thus $M_A/c_m$ is connected on $I + m - 1 \geq 4$ elements, so $\beta(M_A/c_m) \geq 1$ [13] and $\beta(M_A) \geq (m-1) + 1 = m$.

By duality, $\beta(M_A) \geq \max(\ell, m)$ on $T^{\circ\circ}_{\ell,m}$ for $\ell, m \geq 2$. Hence $U_{\ell,m}(s) = 0$ whenever $\max(\ell, m) > s$, which truncates the sum to (10). Since $S(I,1) = 1$, this gives $W_{I,J}(1) = U_{1,1}(1) = 1$. For $s = 2$, $U_{2,2}(2) = \chi(\mathbb{C} \setminus \{0,1\}) = -1$ and $U_{1,m}(2) = U_{\ell,1}(2) = 0$, giving $W_{I,J}(2) = -S(I,2)\, S(J,2)$. For $s = 3$, Corollary 5.3 gives $U_{2,2}(3) = 0$ and $U_{2,3}(3) = U_{3,2}(3) = 2$, and Proposition 6.1 gives $6 = W_{3,3}(3) = 3 \cdot 1 \cdot 2 + 1 \cdot 3 \cdot 2 + U_{3,3}(3)$, so $U_{3,3}(3) = -6$, yielding the formula for $W_{I,J}(3)$. □

For $I = 2$ the proof of Corollary 5.3 gives $U_{2,m}(s) = 0$ for $s \neq m$ and $U_{2,m}(m) = (-1)^{m-1}(m-1)!$, the Euler characteristic of the space of $m-1$ distinct points of $\mathbb{C}^* \setminus \{1\}$, so that $W_{2,J}(s) = (-1)^{s-1}(s-1)!\, S(J,s)$. At $I = 3$ the $S(J,3)$ terms in $W_{I,J}(3)$ cancel, leaving

$2\,S(J,2) = 2(2^{J-1}-1)$. Under the point-count and strata hypotheses of Sections 6.2–6.4, the weights in Tables 1 and 4 similarly determine $W_{3,J}(s)$ for $s = 4,5,6$ and all $J \geq 2$:

$$\begin{aligned} W_{3,J}(4) &= 36\,S(J,3) - 12\,S(J,4), \\ W_{3,J}(5) &= -60\,S(J,3) + 72\,S(J,4) + 72\,S(J,5), \\ W_{3,J}(6) &= 26\,S(J,3) + 96\,S(J,4) - 360\,S(J,6). \end{aligned}$$

Inverting $W_{3,J}(s) = \sum_{\ell=1}^{3}\sum_{m=1}^{J} S(3,\ell)\,S(J,m)\,U_{\ell,m}(s)$ gives $U_{3,4}(4) = 6$, while $U_{3,5}(s) = 0$ for $s \leq 6$ and $U_{3,6}(s) = 0$ for $s \leq 8$, so that the contribution of $U_{3,m}$ to $W_{3,J}(s)$ vanishes for $m$ well below the bound $m \leq s$ of Proposition 8.1. In addition, $\sum_s U_{\ell,m}(s) = (-1)^{\ell+m-1}(\ell-1)!(m-1)!$ for all $\ell, m \geq 2$. To see this, note that the Euler characteristic of $T^{\circ}_{I,m}$, the space of $m-1$ distinct points in $(\mathbb{C}^*)^{I-1}\setminus\{\mathbf{1}\}$, is $(-1)^{m-1}(m-1)!$, and grouping its rows as in the proof gives $(-1)^{m-1}(m-1)! = \sum_\ell S(I,\ell)\sum_s U_{\ell,m}(s)$ for every $I \geq 2$. These equations determine the sums recursively in $I$, and the stated values satisfy them because $\sum_\ell (-1)^{\ell-1}(\ell-1)!\,S(I,\ell) = 0$ for $I \geq 2$, the coefficient of $t^I/I!$ in $\log(1+(e^t-1)) = t$.

The computed weights suggest the following structural properties across all $I, J \geq 2$:

**Conjecture 8.2.** *For all $I, J \geq 2$:*

(a) (Full support.) *Every integer $1 \leq s \leq \binom{I+J-2}{I-1}$ occurs with nonzero weight $W_{I,J}(s) \neq 0$.*

(b) (Lower bound for three rows.) *Every $A \in T^{\circ\circ}_{3,m}$ with $m \geq 5$ satisfies $\beta(M_A) \geq 2m-3$. In particular $U_{3,m}(s) = 0$ whenever $m \geq 5$ and $s < 2m-3$, so that, apart from the term $3\,U_{2,s}(s)\,S(J,s)$, the weight $W_{3,J}(s)$ involves only the $S(J,m)$ with $m \leq \max\bigl(4, \lfloor (s+3)/2 \rfloor\bigr)$.*

(c) (Signs near the top.) *With $s_{\max} = \binom{I+J-2}{I-1}$, the weight $W_{I,J}(s_{\max}-d)$ has sign $(-1)^{(I-1)(J-1)+d}$ for $0 \leq d \leq \min(2, s_{\max}-1)$; equivalently, the coefficients of $s_{\max}^K$, $(s_{\max}-1)^K$, and $(s_{\max}-2)^K$ in $\operatorname{mldeg}(\mathcal{M}_{I,J,K})$ are positive, negative, and positive, for those of these bases that are at least 1.*

Part (a) holds for $I = 2$, by the formula for $W_{2,J}$ above, and for $I = 3$ and $J \leq 6$, by Tables 1 and 4, for $J = 4,5,6$ under the hypotheses of Sections 6.2–6.4. The vanishing in part (b) agrees with the values $U_{3,5}(s) = 0$ for $s \leq 6$ and $U_{3,6}(s) = 0$ for $s \leq 8$ above, although a vanishing Euler characteristic does not by itself show that the strata are empty, and the bound fails for $m = 4$, since $U_{3,4}(4) = 6 \neq 0$. For $I = 3$ and $m$ columns not all on a line through $(1,1,1)^\top$ and some $e_i$, the bound $2m-3$ corresponds via (6) to placing $m-1$ columns on one such line and the last column on two three-point lines. Part (c) holds for $I = 2$ and every $d$, by the formula for $W_{2,J}$, and for $I = 3$ and $J \leq 6$, under the same hypotheses; for $J = 4,5,6$ the alternation persists up to $d = 6,4,6$, while for $J = 3$ it fails at $d = 3$, and in the interior of $[1, s_{\max}]$ the signs do not alternate in general.

At the bottom, by (10), $W_{I,J}(4)$ is determined for all $I, J \geq 2$ by $U_{2,4}(4) = -6$, $U_{3,3}(4) = 36$, $U_{3,4}(4) = 6$ (the last under the point-count hypothesis of Section 6.2), and the single remaining value $U_{4,4}(4)$, since $U_{2,2}(4) = U_{2,3}(4) = 0$.

## Acknowledgments

The author expresses gratitude for the contributions of Claude Opus 5.5.

## Appendix A. Point counts

For $J = 4$ and $s = 1,\dots,10$ let $N_s(q)$ be the number of $A \in T_{3,4}(\mathbb{F}_q)$ with $\beta(M_A) = s$, where $M_A$ is the matroid of the columns of $[\mathrm{Id}_3 \mid A]$ over $\mathbb{F}_q$. Table 5 lists, for each $s$, the polynomial

that agrees with $N_s(q)$ for $q = 7$, 11, 13, 17, 19, 23, 29, 31, and 37, and its value at $q = 1$, the weight $W_{3,4}(s)$ of Table 1. As they must, the polynomials sum to $(q-1)^6$, the number of points of $T_{3,4}(\mathbb{F}_q)$.

TABLE 5. The numbers of points $N_s(q)$ of the strata $\{A \in T_{3,4} : \beta(M_A) = s\}$ for the primes listed, as polynomials in $q$, and their values at $q = 1$.

| $s$ | $N_s(q)$ | value at $q = 1$ |
|---|---|---|
| 1 | $1$ | 1 |
| 2 | $21(q-2)$ | $-21$ |
| 3 | $(q-2)(25q-39)$ | 14 |
| 4 | $3(q^3+27q^2-154q+194)$ | 204 |
| 5 | $12(q-3)(5q^2-24q+31)$ | $-288$ |
| 6 | $6(q-3)(q^3+q^2-8q-15)$ | 252 |
| 7 | $(q-3)(13q^3+121q^2-1282q+2192)$ | $-2088$ |
| 8 | $9(q-3)(15q^3-178q^2+674q-782)$ | 4878 |
| 9 | $22(q-3)(q-5)(q^3-13q^2+54q-66)$ | $-4224$ |
| 10 | $(q-3)(q-5)(q^4-20q^3+148q^2-468q+498)$ | 1272 |

For $J = 5$, Table 6 lists the polynomials $P_s$ and the constants $c_s$ of Section 6.3; the number of points of the stratum $\{A \in T_{3,5} : \beta(M_A) = s\}$ over $\mathbb{F}_q$ is $P_s(q) + c_s\,\rho(q)$ for every odd prime power $q \le 31$ and for $q = 37$. As they must, the polynomials sum to $(q-1)^8$ and the constants sum to 0.

TABLE 6. The polynomials $P_s(q)$ and constants $c_s$ for the strata $\{A \in T_{3,5} : \beta(M_A) = s\}$, and the weights $W_{3,5}(s) = P_s(1) + 2c_s$.

| $s$ | $P_s(q)$ | $c_s$ | $W_{3,5}(s)$ |
|---|---|---|---|
| 1 | $1$ | 0 | 1 |
| 2 | $45(q-2)$ | 0 | $-45$ |
| 3 | $30(q-2)(3q-4)$ | 0 | 30 |
| 4 | $30(q^3+6q^2-49q+68)$ | 0 | 780 |
| 5 | $(q-3)(3q^3+217q^2-666q+800)$ | 0 | $-708$ |
| 6 | $5(q-3)(5q^3+89q^2-369q+114)$ | 0 | 1610 |
| 7 | $10(q-3)(13q^3+139q^2-1372q+2372)$ | 0 | $-23040$ |
| 8 | $30(49q^4-713q^3+3890q^2-9306q+8194)$ | $-120$ | 63180 |
| 9 | $2(118q^5-2425q^4+21950q^3-100625q^2+222282q-186720)$ | 1200 | $-88440$ |
| 10 | $10(q^6-28q^5+611q^4-5696q^3+23382q^2-41172q+23958)$ | $-5040$ | 480 |
| 11 | $60(10q^5-93q^4-265q^3+5496q^2-19996q+22592)$ | 11400 | 487440 |
| 12 | $5(q-5)(7q^5+444q^4-9700q^3+68193q^2-197430q+201288)$ | $-15000$ | $-1286040$ |
| 13 | $60(q-5)(9q^5-232q^4+2343q^3-11452q^2+26678q-23526)$ | 11520 | 1506240 |
| 14 | $40(q-5)(q^6-34q^5+476q^4-3485q^3+13914q^2-28296q+22674)$ | $-4800$ | $-849600$ |
| 15 | $(q-5)(q^7-43q^6+788q^5-7937q^4+47097q^3-162834q^2+299280q-222960)$ | 840 | 188112 |

## 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] D. Agostini, T. Brysiewicz, C. Fevola, L. Kühne, B. Sturmfels, and S. Telen, *Certifying maximum likelihood degrees of matroid strata*, dataset, Zenodo, 2022, doi:10.5281/zenodo.7454826.

[3] 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.

[4] J. Berthomieu, C. Eder, and M. Safey El Din, *msolve: A library for solving polynomial systems*, in Proceedings of the 2021 International Symposium on Symbolic and Algebraic Computation (ISSAC '21), ACM, 2021, pp. 51–58, doi:10.1145/3452143.3465545.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] J. I. Coons and S. Sullivant, *Quasi-independence models with rational maximum likelihood estimator*, J. Symbolic Comput. **104** (2021), 917–941, doi:10.1016/j.jsc.2020.10.006.

[12] D. Corey and D. Luber, *Singular matroid realization spaces*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5), published online 2025, doi:10.2422/2036-2145.202311_001.

[13] H. H. Crapo, *A higher invariant for matroids*, J. Combinatorial Theory **2** (1967), 406–417, doi:10.1016/S0021-9800(67)80051-6.

[14] I. Csiszár, *I-divergence geometry of probability distributions and minimization problems*, Ann. Probab. **3** (1975), no. 1, 146–158, doi:10.1214/aop/1176996454.

[15] 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.

[16] B. Grünbaum, *Configurations of Points and Lines*, Grad. Stud. Math. **103**, American Mathematical Society, Providence, RI, 2009, doi:10.1090/gsm/103.

[17] 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.

[18] J. Huh, *The maximum likelihood degree of a very affine variety*, Compos. Math. **149** (2013), no. 8, 1245–1266, doi:10.1112/S0010437X13007057.

[19] S. Sivarajan, *An explicit determinantal formula for limits of iterative proportional fitting with a binary variable*, hexagon:2611.00705 (2026).

[20] 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.