hexagon:2611.00705v1

AN EXPLICIT DETERMINANTAL FORMULA FOR LIMITS OF ITERATIVE PROPORTIONAL FITTING WITH A BINARY VARIABLE

SHANKAR SIVARAJAN

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

Abstract. Iterative proportional fitting rescales a table cyclically until it has prescribed margins. For a three-way table and its three two-way margins, the limit is in general an algebraic function of the data of degree greater than one, and from the uniform seed table it is the maximum likelihood estimate under the no-three-way-interaction model. We study tables in which one variable is binary, a setting that includes logistic regression of a binary response on two categorical factors and stratified 2×2 tables with a common odds ratio. For $2 \times 2 \times K$ tables, we show that the limit is determined by a root of the determinant of an explicit $2^K \times 2^K$ matrix, built from one 2×2 companion pencil per stratum, and we identify all of its roots. For the uniform seed, this exhibits the unconditional maximum likelihood estimate of the common odds ratio as a root of a polynomial of degree $2^K - 1$, and gives an elementary proof that this is the maximum likelihood degree of the model. For $2 \times J \times K$ tables and generic data, the fitting equations form a multiparameter eigenvalue problem, which yields a matrix pencil of size $J^{K-1}K^{J-1}$ whose eigenvalues are the values of one cell at all complex solutions. Finally, every complex solution is a fixed point of the iteration; we characterize the real ones that are linearly attracting, and show that for $2 \times 2 \times K$ tables and generic margins only the limit is attracting.

1. Introduction

Iterative proportional fitting adjusts a table to prescribed margins by rescaling it cyclically to match each margin in turn. It was introduced by Deming and Stephan [11], its convergence was proved in general by Csiszár [9], and it is used wherever a table must be reconciled with known totals, from sample-survey adjustment to the analysis of contingency tables [1, 6]. For a three-way table and its three two-way margins, started from a positive seed table $T = (t_{ijk})$, the limit is the unique positive table of the form $P_{ijk} = t_{ijk}a_{ij}b_{ik}c_{jk}$ with the given margins; for the uniform seed $T = \mathbf{1}$ it is the maximum likelihood estimate under the no-three-way-interaction model [5]. Unlike the estimate under the independence model, which is a product of margins, this limit is in general an algebraic function of the data of degree greater than one, and it is computed by the iteration itself. Our aim is to describe it exactly: we show that, when one variable is binary, a coordinate of the limit is a root of an explicit determinant, and we identify the other roots as the remaining complex solutions of the fitting equations. The determinants are a structural description of the limit, not a faster algorithm for computing it.

Suppose that one of the three variables is binary. The no-three-way-interaction model for a $2 \times J \times K$ table is equivalent to the logistic regression of a binary response on two categorical factors without interaction. For $J = 2$ it is the model of homogeneous association for K stratified 2×2 tables, as in stratified case–control studies [17], under which the odds ratio is the same in every stratum; the margins n_{ij+}, n_{i+k}, and n_{+jk} are sufficient, and the common odds ratio $\hat{\psi}$ of the fitted table is its unconditional maximum likelihood estimate, as distinct from the

conditional estimate obtained by conditioning on the stratum margins. For $K = 2$, Bartlett [4] reduced the fit to a cubic equation (see also [21]). Norton [18] gave a scheme of successive approximation for $2\times2\times K$ tables and their generalizations, Roy and Kastenbaum [19] wrote the likelihood equations of an $r\times s\times t$ table as $(r-1)(s-1)(t-1)$ simultaneous equations and solved them numerically, and Birch [5] proved that the likelihood equations have a unique positive solution, at which the likelihood is maximized. For general K the estimate is characterized by the equality of the observed and fitted totals of the cell $(1,1)$, $\sum_k n_{11k} = \sum_k \xi_k(\psi)$, where the fitted value $\xi_k(\psi)$ is the admissible root of a quadratic equation in each stratum, and it is computed iteratively [7]. The Mantel–Haenszel estimator [17] is explicit, but it is a different estimator. In algebraic terms, the $2 \times J \times K$ model is the Lawrence lifting of the vertex-edge incidence matrix of the complete bipartite graph $K_{J,K}$, the design matrix of the independence model of a $J \times K$ table [8].

We show that $\hat{\psi}$ is a root of an explicit polynomial, given as a single determinant of size 2^K, and we identify all of its roots (Theorems 3.2 and 3.5). The determinant eliminates the stratum quadratics from the equation $\sum_k \xi_k(\psi) = n_{11+}$: each stratum contributes a quadratic in its own entry P_{11k}, with coefficients linear in ψ, and the coupling is a Kronecker sum of the companion pencils of these quadratics. For $T = \mathbf{1}$ and $n_{11+} \neq \sum_k n_{1+k}n_{+1k}/n_{++k}$, the determinant is $(\psi-1)^{K2^{K-1}-2^K+1}F(\psi)$, where F has degree 2^K-1; its roots are the common odds ratios of all complex critical points of the likelihood, and for generic margins they are simple. The number of complex critical points of the likelihood for generic data is the maximum likelihood degree of the model; the value 2^K-1 was conjectured in [2] and proved in [8], and our argument gives an elementary proof of it together with an explicit polynomial. The estimate $\hat{\psi}$ is the unique root that belongs to a positive solution (Proposition 3.4). For every positive seed the determinant has degree $K2^{K-1}$, and up to a factor that depends only on the seed, its leading and constant coefficients are products of linear forms in the margins; when the stratum odds ratios of the seed are distinct, this degree is the number of complex solutions for generic margins.

For $2 \times J \times K$ tables we write the fitted odds as $s_{jk}\alpha_j\beta_k$, where $s_{jk} = t_{1jk}/t_{2jk}$ are the odds of the seed. Each column margin is then a secular equation, one whose terms are rational functions with simple poles, and after inverting the row parameters each margin equation becomes the singularity of a matrix depending linearly on $(1/\alpha_1, \ldots, 1/\alpha_J, \beta_1, \ldots, \beta_K)$: the fit is a multiparameter eigenvalue problem in the sense of Atkinson [3]. We prove that for a generic seed table and generic margins the commuting operators of this problem, restricted to an explicit invariant subspace, give a matrix pencil of size $J^{K-1}K^{J-1}$ whose eigenvalues are exactly the values of the cell $(1,1,1)$ at all complex solutions (Theorem 4.7). For the maximum likelihood case the corresponding pencil is in general singular, and we show that the maximum likelihood eliminant divides the limit of these eliminants along a generic line of seeds $T = \mathbf{1} + \eta T'$ as $\eta \to 0$, for generic margins (Theorem 4.9). We conjecture a description of the remaining eigenvalues of the operators and an exact description of the maximum likelihood eliminant within the singular pencil (Conjectures 4.8 and 4.10).

For three-way tables of any format, every complex solution is a fixed point of iterative proportional fitting, although from a positive seed the iteration stays positive and reaches only $\hat{P}$. We compute the derivative of one cycle at each complex solution; at $\hat{P}$ its spectral radius is the local rate of convergence of the iteration. For a solution S with real entries, the relevant quadratic form is $H_S(z) = \sum_{ijk} S_{ijk}z_{ijk}^2$, the quadratic form of the derivative of the margins with respect to the log-linear parameters, restricted to the tangent space of the model: the solution is linearly attracting if the form is positive definite and linearly unstable if the form is nondegenerate and not positive definite (Theorem 5.2), and for generic data one of the two

holds. For $2 \times 2 \times K$ tables and generic margins, the fitted table is the only solution with real entries that is linearly attracting (Theorem 5.3), while complex solutions can be.

For matrix and tensor scaling with one-dimensional margins, each entry of the limit is a root of a multilinear resultant that is a single Sylvester determinant [20]. For two-way margins, as here, each entry of the limit is a root of a sparse resultant on the toric variety of the model [20], for which general formulas are quotients of determinants of Macaulay type [10]; the constructions of the present paper use instead the slice structure of the model and the theory of multiparameter eigenvalue problems.

2. Setting

Let $n = (n_{ijk})$ be a positive $2 \times J \times K$ table and $T = (t_{ijk})$ a positive seed table of the same format. The model $\mathcal{M}_T$ consists of the positive tables $P_{ijk} = t_{ijk} a_{ij} b_{ik} c_{jk}$. A table $P \in \mathcal{M}_T$ with $P_{ij+} = n_{ij+}$, $P_{i+k} = n_{i+k}$, and $P_{+jk} = n_{+jk}$ exists and is unique [9, Thms. 2.1, 3.1, Cor. 3.1] (for $T = \mathbf{1}$ see also [5]), and iterative proportional fitting started at T converges to it [9, Thm. 3.2]; we call it the fitted table $\hat{P}$. For $T = \mathbf{1}$ it is the maximum likelihood estimate under the no-three-way-interaction model.

Write $m_{jk} = n_{+jk}$, $R_j = n_{1j+}$, $C_k = n_{1+k}$, $s_{jk} = t_{1jk}/t_{2jk}$, and $X_{jk} = P_{1jk}$. A positive table with margins m lies in $\mathcal{M}_T$ exactly when its odds have the form

$$\frac{X_{jk}}{m_{jk} - X_{jk}} = s_{jk}\, \alpha_j \beta_k, \tag{1}$$

and then it has the prescribed margins exactly when

$$\sum_k X_{jk} = R_j \quad (1 \le j \le J), \qquad \sum_j X_{jk} = C_k \quad (1 \le k \le K). \tag{2}$$

The parameters (α, β) are determined up to $(\alpha, \beta) \mapsto (\lambda\alpha, \lambda^{-1}\beta)$. A *complex solution* is a pair $(\alpha, \beta) \in (\mathbb{C}^*)^J \times (\mathbb{C}^*)^K$, modulo this scaling, with $1 + s_{jk}\alpha_j\beta_k \neq 0$ for all j, k, such that $X_{jk} = m_{jk} s_{jk} \alpha_j \beta_k / (1 + s_{jk}\alpha_j\beta_k)$ satisfies (2); equivalently, a table $P = T \circ (a_{ij} b_{ik} c_{jk})$ with complex factors and nonzero entries that has the margins of n. The fitted table corresponds to the unique positive complex solution. For $T = \mathbf{1}$ the complex solutions are exactly the critical points, with all coordinates nonzero, of the log-likelihood on the model, so for generic data their number is the maximum likelihood degree of the model. By the *eliminant* of a cell, or of a parameter, we mean a polynomial whose roots are its values at the complex solutions; the cell $(1, 1, 1)$ is the *corner entry*.

A property holds for *generic* data if it holds outside a proper Zariski-closed set, that is, for all data outside the zero set of some nonzero polynomial in the entries of n and T, or in the margins and T, as the context indicates; when T is fixed, the polynomial is in the data that vary. In particular, it holds with probability one for data drawn from any distribution with a density with respect to Lebesgue measure.

3. The common odds ratio of a $2 \times 2 \times K$ table

Let $J = 2$. For each stratum k write

$$r_k = n_{1+k}, \qquad r'_k = n_{2+k}, \qquad c_k = n_{+1k}, \qquad N_k = n_{++k}, \qquad \rho_k = \frac{t_{11k} t_{22k}}{t_{12k} t_{21k}},$$

and $N = n_{11+}$; thus $r_k = C_k$ in the notation of Section 2. A table with the prescribed margins of stratum k is determined by $x = P_{11k}$, with entries x, $r_k - x$, $c_k - x$, and $r'_k - c_k + x$. By (1), a

positive table with the prescribed stratum margins lies in $\mathcal{M}_T$ exactly when its stratum odds ratios satisfy

$$\frac{P_{11k}P_{22k}}{P_{12k}P_{21k}} = \psi\rho_k \qquad (1 \le k \le K)$$

for a common $\psi > 0$, namely $\psi = \alpha_1/\alpha_2$; for $T = \mathbf{1}$ it is the common odds ratio. Clearing denominators, $x = P_{11k}$ is a root of

$$q_k(x;\psi) = u_k(\psi)x^2 + v_k(\psi)x + w_k(\psi), \tag{3}$$

$$u_k = 1 - \psi\rho_k, \qquad v_k = r'_k - c_k + \psi\rho_k(r_k + c_k), \qquad w_k = -\psi\rho_k r_k c_k,$$

and the remaining margins P_{ij+} hold exactly when $\sum_k P_{11k} = N$. Accordingly, a complex solution is a tuple $(\psi, x_1, \dots, x_K)$ with $\psi \neq 0$, $q_k(x_k;\psi) = 0$ and $x_k(r_k - x_k)(c_k - x_k)(r'_k - c_k + x_k) \neq 0$ for every k, and $\sum_k x_k = N$. The last condition on the entries is automatic.

Lemma 3.1. *If $\psi \neq 0$ and $q_k(x;\psi) = 0$, then the four entries x, $r_k - x$, $c_k - x$, and $r'_k - c_k + x$ are nonzero.*

Proof. At the four values of x where an entry vanishes, $q_k(0) = -\psi\rho_k r_k c_k$, $q_k(r_k) = r_k n_{+2k}$, $q_k(c_k) = r'_k c_k$, and $q_k(c_k - r'_k) = -\psi\rho_k r'_k n_{+2k}$, all nonzero. □

The quadratic (3) is $\det(A_k - xB_k)$ for

$$A_k = \begin{pmatrix} 0 & -w_k \\ 1 & -v_k \end{pmatrix}, \qquad B_k = \begin{pmatrix} 1 & 0 \\ 0 & u_k \end{pmatrix}.$$

For a 2×2 matrix Y and $1 \le k \le K$, write $Y^{(k)} = B_1 \otimes \cdots \otimes B_{k-1} \otimes Y \otimes B_{k+1} \otimes \cdots \otimes B_K$, a $2^K \times 2^K$ matrix. For ψ with $u_k(\psi) \neq 0$ let $x_k^{\pm}(\psi)$ be the two roots of $q_k(\,\cdot\,;\psi)$, and for $\varepsilon \in \{+,-\}^K$ let $\Sigma^{\varepsilon}(\psi) = \sum_k x_k^{\varepsilon_k}(\psi)$ be the branch sums, which are branches of an algebraic function of ψ.

Theorem 3.2. *Define*

$$E(\psi) = \det\Big(N\,B_1 \otimes \cdots \otimes B_K - \sum_{k=1}^{K} A_k^{(k)}\Big).$$

(1) *E is a polynomial in ψ, and at every ψ with $u_1(\psi)\cdots u_K(\psi) \neq 0$,*

$$E(\psi) = \prod_{k=1}^{K} u_k(\psi)^{2^{K-1}} \prod_{\varepsilon \in \{+,-\}^K} \big(N - \Sigma^{\varepsilon}(\psi)\big).$$

(2) *The coefficient of $\psi^{K2^{K-1}}$ in E is*

$$\prod_{k=1}^{K} (-\rho_k)^{2^{K-1}} \prod_{\mathcal{A} \subseteq \{1,\dots,K\}} \Big(N - \sum_{k \in \mathcal{A}} r_k - \sum_{k \notin \mathcal{A}} c_k\Big),$$

and the constant term is $\prod_{\mathcal{A}}\big(N - \sum_{k\in\mathcal{A}}(c_k - r'_k)\big)$. Both are nonzero, so $\deg E = K2^{K-1}$ and $E(0) \neq 0$.

(3) *If ψ_0 is the parameter of a complex solution and $u_k(\psi_0) \neq 0$ for all k, then $E(\psi_0) = 0$. Conversely, if $E(\psi_0) = 0$ with $u_k(\psi_0) \neq 0$ for all k, then ψ_0 is the parameter of a complex solution. If the ρ_k are distinct, the same holds at $\psi_0 = 1/\rho_k$, where $E(1/\rho_k) = 0$ exactly when a complex solution has parameter $1/\rho_k$.*

(4) *Suppose the ρ_k are distinct. For generic margins, E has $K2^{K-1}$ simple roots, and there are exactly $K2^{K-1}$ complex solutions, one for each root.*

Proof. (1) Where $u_k \neq 0$ for all k, $B = B_1 \otimes \cdots \otimes B_K$ is invertible and

$$NB - \sum_k A_k^{(k)} = B\Big(N - \sum_k I \otimes \cdots \otimes B_k^{-1} A_k \otimes \cdots \otimes I\Big).$$

The eigenvalues of $B_k^{-1}A_k$ are the roots $x_k^{\pm}$ of $\det(A_k - xB_k) = q_k(x)$, the eigenvalues of a Kronecker sum are the sums of eigenvalues, one from each summand, and $\det B = \prod_k u_k^{2^{K-1}}$. The determinant is a polynomial in the entries, which are polynomials in ψ.

(2) As $\psi \to \infty$, $q_k/(\psi\rho_k) \to -(x - r_k)(x - c_k)$, so $\{x_k^+, x_k^-\} \to \{r_k, c_k\}$, while $u_k^{2^{K-1}} \sim (-\rho_k\psi)^{2^{K-1}}$. Hence $E(\psi)\psi^{-K2^{K-1}}$ tends to the stated product. At $\psi = 0$, $u_k = 1$ and $q_k = x(x + r'_k - c_k)$, with roots 0 and $c_k - r'_k$. Since $r_k = n_{11k} + n_{12k}$, $c_k = n_{11k} + n_{21k}$, and $c_k - r'_k = n_{11k} - n_{22k}$, the factors are $-\sum_{k\in\mathcal{A}} n_{12k} - \sum_{k\notin\mathcal{A}} n_{21k} < 0$ and $\sum_{k\notin\mathcal{A}} n_{11k} + \sum_{k\in\mathcal{A}} n_{22k} > 0$.

(3) If $u_k(\psi_0) \neq 0$ for all k, the factor $\prod u_k^{2^{K-1}}$ is nonzero and finite, so $E(\psi_0) = 0$ exactly when some branch sum equals N, that is, when there are roots x_k of $q_k(\,\cdot\,;\psi_0)$ with $\sum_k x_k = N$. Since $E(0) \neq 0$, $\psi_0 \neq 0$, and by Lemma 3.1 these roots give a complex solution. At $\psi_0 = 1/\rho_k$ the quadratic q_k drops degree, $v_k(1/\rho_k) = r_k + r'_k = N_k \neq 0$, and x_k^+ has a simple pole while $x_k^- \to r_k c_k/N_k$. The 2^{K-1} branches with $\varepsilon_k = +$ each contribute a simple pole, which is cancelled by $u_k^{2^{K-1}}$; since the ρ_k are distinct, the other factors are finite. So $E(1/\rho_k)$ is a nonzero constant times $\prod_{\varepsilon_k=-}(N - \Sigma^{\varepsilon}(1/\rho_k))$, which vanishes exactly when a branch through the finite root $x_k^-(1/\rho_k)$ sums to N, that is, when there is a complex solution with parameter $1/\rho_k$.

(4) The branch sums through the finite root $x_k^-(1/\rho_k)$ take finitely many values at $\psi = 1/\rho_k$, so for all but finitely many N no complex solution has parameter in $\{1/\rho_1, \ldots, 1/\rho_K\}$, and by (3) $E(1/\rho_k) \neq 0$. Apply Lemma 3.3 below with $\Psi = \{1/\rho_1, \ldots, 1/\rho_K\}$. At every root ψ_0 of E exactly one factor $N - \Sigma^{\varepsilon}$ of the product in (1) vanishes, and it vanishes simply, while the other factors are finite and nonzero; so ψ_0 is a simple root of E and carries exactly one complex solution. By (2) there are $K2^{K-1}$ roots, and by (3) every complex solution has one of them as its parameter. □

Lemma 3.3. *Fix positive stratum margins r_k, r'_k, c_k and a finite set Ψ of values of ψ, and suppose that the 2^K numbers $\sum_k(r_k \text{ or } c_k)$ are distinct. For all but finitely many values of N, every $\psi_0 \notin \Psi$ with $u_1(\psi_0)\cdots u_K(\psi_0) \neq 0$ and $N = \Sigma^{\varepsilon}(\psi_0)$ for some branch ε is not a branch point of any $x_k^{\pm}$, lies on exactly one branch, and is a simple zero of $N - \Sigma^{\varepsilon}$.*

Proof. The branch points of the $x_k^{\pm}$ form a finite set Ψ_0, independent of N, and the finite values of the branch sums at the points of Ψ_0 form a finite set. Each branch sum is nonconstant: it tends to $\sum_k(0 \text{ or } c_k - r'_k)$ as $\psi \to 0$ and to $\sum_k(r_k \text{ or } c_k)$ as $\psi \to \infty$, and $c_k - r'_k = n_{11k} - n_{22k} < \min(r_k, c_k)$ and $0 < \min(r_k, c_k)$, so the second limit exceeds the first. Hence each branch sum has finitely many critical values. Moreover, regard N in Theorem 3.2 as an indeterminate, so that E is a polynomial in ψ and N whose roots in N are the branch sums. Its discriminant with respect to N is a polynomial in ψ that does not vanish identically, because as $\psi \to \infty$ the 2^K roots in N tend to the distinct limits $\sum_k(r_k \text{ or } c_k)$. Thus two distinct branches agree at only finitely many points $\psi \notin \Psi_0$, giving finitely many further values of N. Excluding these finitely many values of N proves the lemma. □

The set of margins at which the conclusion fails is constructible, being the projection of an algebraic set, and its fiber over each choice of stratum margins satisfying the hypothesis is finite. Since the hypothesis on the numbers $\sum_k(r_k \text{ or } c_k)$ is a polynomial inequality and N can

be varied independently of the stratum margins, this set is contained in a proper Zariski-closed set of margins; that is, the conclusion holds for generic margins.

Proposition 3.4. *Let $L_k = \max(0, c_k - r'_k)$ and $U_k = \min(r_k, c_k)$. For every $\psi > 0$, each $q_k(\,\cdot\,;\psi)$ has exactly one root $\xi_k(\psi)$ in (L_k, U_k), and ξ_k is continuous and strictly increasing, with limits L_k and U_k as $\psi \to 0$ and $\psi \to \infty$. There is exactly one $\hat\psi > 0$ with $\sum_k \xi_k(\hat\psi) = N$; it is the parameter of the fitted table, $\hat P_{11k} = \xi_k(\hat\psi)$, and $E(\hat\psi) = 0$. A null vector of the matrix in Theorem 3.2 at $\hat\psi$ is $\bigotimes_k \bigl(-w_k/\xi_k, 1\bigr)^\top$, whose entries are positive.*

Proof. On (L_k, U_k) the four entries x, $r_k - x$, $c_k - x$, and $r'_k - c_k + x$ are positive, and

$$\frac{d}{dx}\log\frac{x(r'_k - c_k + x)}{(r_k - x)(c_k - x)} = \frac{1}{x} + \frac{1}{r'_k - c_k + x} + \frac{1}{r_k - x} + \frac{1}{c_k - x} > 0.$$

The odds ratio increases from 0 to ∞ on the interval, so it equals $\psi\rho_k$ at exactly one point $\xi_k(\psi)$, which depends continuously and increasingly on ψ. The sum $\sum_k \xi_k(\psi)$ therefore increases from $\sum_k L_k$ to $\sum_k U_k$. Since n is positive, $n_{11k} \in (L_k, U_k)$, so $N \in (\sum L_k, \sum U_k)$ and $\hat\psi$ exists and is unique. The table with $P_{11k} = \xi_k(\hat\psi)$ is positive, has the prescribed margins, and has stratum odds ratios $\hat\psi\rho_k$, so it lies in $\mathcal{M}_T$ and equals $\hat P$. For the null vector, $(A_k - \xi B_k)(y, 1)^\top = 0$ reads $-\xi y - w_k = 0$ and $y - v_k - \xi u_k = 0$, and $y = -w_k/\xi = \psi\rho_k r_k c_k/\xi > 0$. Writing y_k for these vectors, $A_k y_k = \xi_k B_k y_k$, so $\bigl(NB - \sum_k A_k^{(k)}\bigr) \bigotimes_k y_k = \bigl(N - \sum_k \xi_k\bigr) \bigotimes_k B_k y_k = 0$. □

The maximum likelihood case. For $T = \mathbf{1}$ all ρ_k equal 1, and the poles in the proof of Theorem 3.2(3) coincide at $\psi = 1$.

Theorem 3.5. *Let $T = \mathbf{1}$ and suppose $N \neq \sum_k r_k c_k/N_k$. Then*

$$E(\psi) = (\psi - 1)^{K2^{K-1} - 2^K + 1} F(\psi), \qquad \deg F = 2^K - 1, \qquad F(1) \neq 0.$$

No complex solution has parameter 1, and the roots of F are exactly the common odds ratios of the complex solutions; the maximum likelihood estimate $\hat\psi$ is the root of F described in Proposition 3.4. For generic margins, F has $2^K - 1$ simple roots and there are exactly $2^K - 1$ complex solutions, one for each root.

Proof. Near $\psi = 1$, $u_k = 1 - \psi$ and $v_k \to r'_k + r_k = N_k$, so $x_k^+ = -N_k/(1 - \psi) + O(1)$ and $x_k^- \to r_k c_k/N_k$. For a branch ε let $\mathcal{A} = \{k : \varepsilon_k = +\}$. If $\mathcal{A} \neq \emptyset$, then $N - \Sigma^\varepsilon = (1 - \psi)^{-1} \sum_{k \in \mathcal{A}} N_k + O(1)$, a simple pole with nonzero residue. The branch $\mathcal{A} = \emptyset$ tends to $N - \sum_k r_k c_k/N_k \neq 0$. The prefactor $(1 - \psi)^{K2^{K-1}}$ therefore leaves a zero of order exactly $K2^{K-1} - (2^K - 1)$, and $F(1) \neq 0$; with Theorem 3.2(2), $\deg F = 2^K - 1$. At $\psi = 1$ each q_k is linear with root $r_k c_k/N_k$, so a complex solution with parameter 1 would give $N = \sum_k r_k c_k/N_k$. The roots of F are the roots of E other than 1, at which all $u_k = 1 - \psi$ are nonzero, so by the first part of Theorem 3.2(3) they are the common odds ratios of the complex solutions. The last statement follows from Lemma 3.3 with $\Psi = \{1\}$, as in the proof of Theorem 3.2(4). □

The quantity $\sum_k r_k c_k/N_k$ is the expected value of n_{11+} under conditional independence, which appears in the Mantel–Haenszel statistic [17]; the hypothesis of Theorem 3.5 excludes exactly the case $\hat\psi = 1$. The count $2^K - 1$ is the maximum likelihood degree of the model, conjectured in [2, Conj. 36] and proved in [8, Cor. 38]; Theorem 3.5 gives an elementary proof together with an explicit polynomial. Similarly, Theorem 3.2(4) recovers the number of complex solutions for a generic seed table, which is the degree of the model [2, Cor. 8], $K2^{K-1}$ [2, Prop. 34].

The polynomial F can have several real roots, each the common odds ratio of a solution with real entries (Example A.7); among them $\hat\psi$ is singled out by positivity (Proposition 3.4).

4. $2 \times J \times K$ TABLES

Return to general J, with the notation of Section 2. Set $\theta = (\alpha'_1, \dots, \alpha'_J, \beta_1, \dots, \beta_K)$ with $\alpha'_j = 1/\alpha_j$, a point of $\mathbb{P}^{J+K-1}$, and write $\omega_{jk} = s_{jk}\beta_k/\alpha'_j$, so that $X_{jk} = m_{jk}\omega_{jk}/(1+\omega_{jk})$. Let $\mathbf{e}_j$ denote the standard basis vectors, $\mathbf{1}$ the all-ones vector, and

$$G_k = -I_J + \tfrac{1}{C_k}\, m_{\cdot k}\mathbf{1}^\top, \qquad G'_j = -I_K + \tfrac{1}{R_j}\, m_{j\cdot}\mathbf{1}^\top.$$

Define the pencils

$$W_k(\theta) = \beta_k G_k - \sum_{j=1}^{J} \frac{\alpha'_j}{s_{jk}} \mathbf{e}_j\mathbf{e}_j^\top \quad \text{on } \mathbb{C}^J \qquad (1 \le k \le K),$$

$$V_j(\theta) = \sum_{k=1}^{K} \beta_k s_{jk} \mathbf{e}_k\mathbf{e}_k^\top G'_j - \alpha'_j I_K \quad \text{on } \mathbb{C}^K \qquad (1 \le j \le J-1).$$

There are $J+K-1$ pencils in $J+K$ homogeneous parameters.

Lemma 4.1. *Let θ have all coordinates nonzero and $\alpha'_j + s_{jk}\beta_k \neq 0$ for all j, k. Then all $W_k(\theta)$ and $V_j(\theta)$ are singular if and only if (α, β) is a complex solution. The kernels are then spanned by $X_{\cdot k}$ and $X_{j\cdot}$.*

Proof. $W_k(\theta)x = 0$ reads $x_j(\alpha'_j/s_{jk} + \beta_k) = \beta_k m_{jk}(\mathbf{1}^\top x)/C_k$ for all j. A nonzero solution has $\mathbf{1}^\top x \neq 0$, so up to scaling $x_j = m_{jk}\omega_{jk}/(1+\omega_{jk})$, and $\mathbf{1}^\top x = C_k$ holds exactly when $\sum_j X_{jk} = C_k$. The same computation for V_j gives the row margins for $j < J$, and the row margin for $j = J$ follows from the others. □

Number the pencils $W_1, \dots, W_K, V_1, \dots, V_{J-1}$ as $Y_1, \dots, Y_{J+K-1}$, number the parameters $\theta_1, \dots, \theta_{J+K}$ in the order $\alpha'_1, \dots, \alpha'_J, \beta_1, \dots, \beta_K$, and write $Y_i(\theta) = \sum_q \theta_q A_{iq}$. Following Atkinson [3], the operator determinant

$$\Delta_p = (-1)^{p-1} \sum_{\pi} \operatorname{sgn}(\pi)\, A_{1\pi(1)} \otimes A_{2\pi(2)} \otimes \cdots \otimes A_{J+K-1,\pi(J+K-1)}$$

is the sum over the bijections π from $\{1, \dots, J+K-1\}$ to $\{1, \dots, J+K\} \setminus \{p\}$, with the sign of π as a permutation after the order-preserving identification of $\{1, \dots, J+K\} \setminus \{p\}$ with $\{1, \dots, J+K-1\}$. It acts on $\mathcal{H} = (\mathbb{C}^J)^{\otimes K} \otimes (\mathbb{C}^K)^{\otimes (J-1)}$, of dimension $J^K K^{J-1}$. We use three facts. First, Δ_p is singular if and only if there is a nonzero θ with $\theta_p = 0$ at which all pencils are singular [3, Thm. 8.7.1]. Second, if Δ_p is invertible, the operators $\Gamma_q = \Delta_p^{-1}\Delta_q$ commute, their joint eigenvalues are the points θ, normalized by $\theta_p = 1$, at which all pencils are singular, and the joint eigenspace at such a point is the tensor product of the kernels of the pencils [3, Thms. 6.7.2, 6.8.1, 6.9.1], [16, Thm. 2]. Third, suppose that at such a point λ every pencil has a one-dimensional kernel, spanned by x_i, with left kernel spanned by y_i, and that the matrix $\mathbf{B}_0 = (y_i^\top A_{iq} x_i)_{i,\, q \neq p}$ is nonsingular. Then λ is an algebraically simple joint eigenvalue. Indeed, by [16, Lemma 3] the joint second root subspace $\bigcap_{q,q'} \ker(\Gamma_q - \lambda_q)(\Gamma_{q'} - \lambda_{q'})$ equals the joint eigenspace, which is one-dimensional. If the joint generalized eigenspace G at λ were larger, let $\mathfrak{A}$ be the algebra of nilpotent operators on G generated by the $\Gamma_q - \lambda_q$, and let $r \ge 1$ be largest with $\mathfrak{A}^r G \neq 0$; a vector $v \in \mathfrak{A}^{r-1}G$ with $\mathfrak{A}v \neq 0$ lies in the joint second root subspace but not in the joint eigenspace.

Definition 4.2. The ratios s_{jk} are *cycle-generic* if $s_{j_1k_1}s^{-1}_{j_2k_1}s_{j_2k_2}s^{-1}_{j_3k_2}\cdots s_{j_rk_r}s^{-1}_{j_1k_r} \neq 1$ for every cycle $j_1, k_1, j_2, k_2, \dots, j_r, k_r$ in the complete bipartite graph on rows and columns. The

margins are *admissible* if

$$\sum_{j\notin\sigma} R_j \neq \sum_{k\in\mathcal{B}}\Big(C_k - \sum_{j\in\sigma} m_{jk}\Big)$$

for every set σ of rows containing J and every nonempty set $\mathcal{B}$ of columns. Both conditions hold outside proper Zariski-closed sets, so generic seed tables and margins are in particular cycle-generic and admissible.

Lemma 4.3. *Let the s_{jk} be cycle-generic, let θ be a point at which all pencils are singular, and let $\mathcal{D}$ be the set of cells (j,k) with $\beta_k \neq 0$ and $\alpha'_j + s_{jk}\beta_k = 0$. Then $\mathcal{D} = \emptyset$.*

Proof. If $(j,k) \in \mathcal{D}$, the equation of $W_k(\theta)x = 0$ in row j reads $0 = \beta_k m_{jk}(\mathbf{1}^\top x)/C_k$, so $\mathbf{1}^\top x = 0$. Every other row i then gives $x_i(\alpha'_i/s_{ik} + \beta_k) = 0$, so x is supported on the cells of $\mathcal{D}$ in column k and sums to zero, and column k has at least two cells in $\mathcal{D}$. The same argument with V_j, which forces $y_k = 0$ whenever $\beta_k = 0$ because $\alpha'_j = -s_{jk'}\beta_{k'} \neq 0$ for a cell $(j,k') \in \mathcal{D}$, shows that every row $j \leq J-1$ meeting $\mathcal{D}$ has at least two cells in $\mathcal{D}$. A cycle in $\mathcal{D}$ would give $\prod_i s_{j_ik_i}/s_{j_{i+1}k_i} = 1$. So $\mathcal{D}$ is a forest in which every vertex other than row J has degree at least two. A nonempty finite forest has at least two leaves, so $\mathcal{D} = \emptyset$. □

Lemma 4.4. *Let the s_{jk} be cycle-generic and the margins admissible. Then no point with $\alpha'_J = 0$ makes all pencils singular. Consequently $\Delta_{\alpha'_J}$ is invertible.*

Proof. Suppose all pencils are singular at $\theta \neq 0$ with $\alpha'_J = 0$. Let $\sigma = \{j : \alpha'_j = 0\} \ni J$ and $\mathcal{B} = \{k : \beta_k \neq 0\}$. If $\mathcal{B} = \emptyset$, then $\sigma \neq \{1,\dots,J\}$ and $V_j = -\alpha'_j I$ is invertible for any $j \notin \sigma$, a contradiction. So $\mathcal{B} \neq \emptyset$. For $k \in \mathcal{B}$ and $j \in \sigma$, the equation of $W_k(\theta)x = 0$ in row j gives $x_j = m_{jk}(\mathbf{1}^\top x)/C_k$. If σ is all of $\{1,\dots,J\}$, summing gives $\mathbf{1}^\top x = (m_{+k}/C_k)\mathbf{1}^\top x$, so $\mathbf{1}^\top x = 0$ and $x = 0$, since $C_k < m_{+k}$; a contradiction. Otherwise, by Lemma 4.3, $\alpha'_j + s_{jk}\beta_k \neq 0$ for $j \notin \sigma$ and $k \in \mathcal{B}$. As in Lemma 4.1, the kernel of $W_k(\theta)$, $k \in \mathcal{B}$, normalized to sum C_k, has entries $X_{jk} = m_{jk}$ for $j \in \sigma$ and $X_{jk} = m_{jk}\omega_{jk}/(1+\omega_{jk})$ for $j \notin \sigma$. For $j \notin \sigma$, the equation of $V_j(\theta)y = 0$ forces $y_k = 0$ for $k \notin \mathcal{B}$, and normalizing to sum R_j gives $\sum_{k\in\mathcal{B}} X_{jk} = R_j$ with the same X_{jk}. Summing these row equations over $j \notin \sigma$ and the column equations $\sum_j X_{jk} = C_k$ over $k \in \mathcal{B}$ gives $\sum_{j\notin\sigma} R_j = \sum_{k\in\mathcal{B}}(C_k - \sum_{j\in\sigma} m_{jk})$, which admissibility excludes. The last statement is the first fact above. □

Lemma 4.5. *Under the hypotheses of Lemma 4.4, if all pencils are singular at θ and $\alpha'_j \neq 0$ for $1 \leq j \leq J-1$, then θ satisfies the hypotheses of Lemma 4.1, and hence is a complex solution.*

Proof. By Lemma 4.4, $\alpha'_J \neq 0$. If $\beta_k = 0$, then $W_k(\theta) = -\sum_j(\alpha'_j/s_{jk})\mathbf{e}_j\mathbf{e}_j^\top$ is invertible, so all coordinates are nonzero, and Lemma 4.3 gives $\alpha'_j + s_{jk}\beta_k \neq 0$ for all j,k. □

Lemma 4.6. *For generic T and margins, every complex solution is nondegenerate, that is, the Jacobian of (2) with respect to $(\alpha'_1,\dots,\alpha'_{J-1},\beta_1,\dots,\beta_K)$ at $\alpha'_J = 1$ is nonsingular, and every complex solution is an algebraically simple joint eigenvalue of the $\Gamma_q = \Delta_{\alpha'_J}^{-1}\Delta_q$. Nondegeneracy also holds for every fixed positive T, in particular $T = \mathbf{1}$, and generic margins.*

Proof. Consider the morphism that sends $(T, m, \alpha'_1,\dots,\alpha'_{J-1},\beta)$, with $\alpha'_J = 1$ and in the open set where all coordinates and all $1+\omega_{jk}$ are nonzero, to $\big(T, m, (\sum_j X_{jk})_k, (\sum_k X_{jk})_{j<J}\big)$. Its source is an open subset of an affine space, hence nonsingular, and its target has the same dimension, where $R_{<J} = (R_1,\dots,R_{J-1})$ determines $R_J = \sum_k C_k - \sum_{j<J} R_j$. Moreover, the morphism is dominant, since its image contains all positive $(T, m, C, R_{<J})$ arising from positive tables n, a Euclidean-open and hence Zariski-dense subset of the target. By generic

smoothness [13, Cor. III.10.7] it is smooth over a dense open set of $(T, m, C, R_{<J})$, and since its differential is block triangular with identity blocks for (T, m), smoothness means that the Jacobian of the margins with respect to $(\alpha'_{<J}, \beta)$ is invertible at every point of the fiber. For a fixed T the same argument applies to the morphism with T omitted, which is dominant and smooth over a dense open set of $(m, C, R_{<J})$. At a complex solution each pencil has a one-dimensional kernel by Lemma 4.1, and by the matrix determinant lemma

$$\det W_k(\theta) = (-1)^J \prod_j \Big(\frac{\alpha'_j}{s_{jk}} + \beta_k\Big)\Big(1 - \frac{1}{C_k}\sum_j X_{jk}(\theta)\Big),$$
$$\det V_j(\theta) = (-1)^K \prod_k (\alpha'_j + s_{jk}\beta_k)\Big(1 - \frac{1}{R_j}\sum_k X_{jk}(\theta)\Big).$$

If a matrix W has a one-dimensional kernel spanned by x and left kernel spanned by y, then $\operatorname{adj} W = c\, xy^\top$ with $c \neq 0$, and $\partial_q \det W = \operatorname{tr}(\operatorname{adj} W\, \partial_q W) = c\, y^\top (\partial_q W) x$. Hence the matrix $\mathbf{B}_0$, in the chart $\alpha'_J = 1$, is an invertible diagonal matrix times the Jacobian of the residuals $1 - C_k^{-1}\sum_j X_{jk}$ and $1 - R_j^{-1}\sum_k X_{jk}$ $(j < J)$, which is nonsingular. The third fact above completes the proof. □

Theorem 4.7. *Let T and the margins be generic, and let $\Gamma_q = \Delta_{\alpha'_J}^{-1}\Delta_q$. Let $\Gamma_* = \Gamma_{\alpha'_1}\cdots\Gamma_{\alpha'_{J-1}}$ and let $U = \operatorname{range}(\Gamma_*^{\dim \mathcal{H}})$, an invariant subspace of all Γ_q. Then $\dim U = J^{K-1}K^{J-1}$, the joint eigenvalues of the $\Gamma_q|_U$ are exactly the complex solutions, each of multiplicity one, and*

$$E_{11}(x) = \det\Big(x\big(\Gamma_{\alpha'_1} + s_{11}\Gamma_{\beta_1}\big)\big|_U - m_{11}s_{11}\Gamma_{\beta_1}\big|_U\Big) = c\prod_\xi (x - X_{11}(\xi)), \qquad c \neq 0,$$

the product over all complex solutions ξ. In particular $\hat{P}_{111}$, the value at the unique positive solution, is a root of E_{11}.

Proof. By Lemma 4.4 the Γ_q are defined and commute, so $\mathcal{H}$ splits into joint generalized eigenspaces, and on the one for θ the operator Γ_* has the single eigenvalue $\prod_{j<J}\alpha'_j/\alpha'_J$. Thus U is the sum of the generalized eigenspaces with $\alpha'_j \neq 0$ for all $j < J$. By Lemmas 4.5 and 4.1 these are exactly the complex solutions, and by Lemma 4.6 each is one-dimensional, so $\dim U$ is the number of complex solutions. For generic T this number is the degree of the toric variety of the model [2, Cor. 8], which is $J^{K-1}K^{J-1}$ [8, Cor. 2]. On the eigenspace of θ the operator in the determinant acts by $\big(x(\alpha'_1 + s_{11}\beta_1) - m_{11}s_{11}\beta_1\big)/\alpha'_J$, which vanishes exactly at $x = X_{11}$, with nonzero coefficient of x. □

Conjecture 4.8. *For generic T and margins, the joint eigenvalues outside U are the complex solutions of the subproblems in which a nonempty set $\sigma \subseteq \{1, \dots, J-1\}$ of rows is saturated $(X_{jk} = m_{jk})$ on a set $\mathcal{B}$ of columns and the columns outside $\mathcal{B}$ are dropped $(\beta_k = 0)$, each with multiplicity $(K - |\mathcal{B}|)^{|\sigma|}|\sigma|^{K-|\mathcal{B}|}$, where $0^0 = 1$ and $\mathcal{B} = \emptyset$ occurs only for $\sigma = \{1, \dots, J-1\}$.*

Since the generalized eigenspaces fill $\mathcal{H}$, Conjecture 4.8 implies the identity

$$J^K K^{J-1} = \sum_{\mu,\nu} \binom{J-1}{\mu}\binom{K}{\nu}(K-\nu)^\mu\, \mu^{K-\nu}\, f(J-\mu, \nu), \qquad f(j,\nu) = j^{\nu-1}\nu^{j-1},\ f(j,0) = 1,$$

summed over $0 \le \mu \le J-1$ and $0 \le \nu \le K$, where $0^0 = 1$ and the terms with $\nu = 0$ are included only for $\mu = J-1$.

The maximum likelihood case. For $T = \mathbf{1}$ all $s_{jk} = 1$, the ratios are not cycle-generic, and Theorem 4.7 does not apply. At a complex solution ξ, the tensor product of the kernel vectors of Lemma 4.1 lies in the kernel of the pencil

$$L(x) = x\big(\Delta_{\alpha'_1} + \Delta_{\beta_1}\big) - m_{11}\Delta_{\beta_1}$$

at $x = X_{11}(\xi)$, but L is in general singular, so this does not by itself make $X_{11}(\xi)$ an eigenvalue. The finite regular eigenvalues of L are the points where its rank drops below its normal rank; let $\mathcal{R}(x)$ be the monic polynomial whose roots are these eigenvalues, with their algebraic multiplicities in the Kronecker canonical form of L. Let $d = \dim \mathcal{H}$ and let κ be the difference between d and the normal rank of L. For generic $Y, Z \in \mathbb{C}^{d\times\kappa}$ and generic diagonal $\Lambda_1, \Lambda_2 \in \mathbb{C}^{\kappa\times\kappa}$, the regular part of $L(x) + Y(x\Lambda_1 - \Lambda_2)Z^\top$ has the same Kronecker canonical form as that of L, and its remaining eigenvalues depend on the perturbation [14, Summary 4.7]. Hence $\det\big(L(x) + Y(x\Lambda_1 - \Lambda_2)Z^\top\big)$ is $\mathcal{R}(x)$ times a factor whose roots are not regular eigenvalues of L. At such a root x_0, $L(x_0)$ has the normal rank of L, so a second generic perturbation has nonzero determinant at x_0; hence for two generic perturbations the greatest common divisor of the two determinants is $\mathcal{R}(x)$. The polynomials $\mathcal{R}$ in Examples A.4 and A.5 were computed in this way, with random perturbations modulo primes, and reconstructed over $\mathbb{Q}$; the statements made about them there hold with probability one in the choice of perturbations, rather than being certified.

The maximum likelihood eliminant is also a limit of the eliminants of Theorem 4.7. Let the margins and a direction T' be rational, and for $\eta \in \mathbb{C}$ let $T(\eta) = \mathbf{1} + \eta T'$. Let $E_\eta(x)$ be the monic polynomial E_{11}/c of Theorem 4.7 for the seed table $T(\eta)$. The construction of Theorem 4.7 is algebraic and does not use positivity of T. Fixing a set of pivot columns of $\Gamma_*^{\dim\mathcal{H}}$ that is independent for generic η, the coefficients of E_η are computed from the Δ_q by rational operations, so they are rational functions of η with rational coefficients, and for all η outside a finite set $E_\eta(x) = \prod_\xi (x - X_{11}(\xi))$, the product over the complex solutions for $T(\eta)$. Multiplying by an element of $\mathbb{Q}(\eta)^\times$, we obtain $E(x,\eta) \in \mathbb{Q}[\eta][x]$ whose coefficients have no common factor η, and we set $E_0(x) = E(x,0) \neq 0$.

Theorem 4.9. *Let the margins be generic and rational, and let T' be rational and generic. Let $E_{\mathrm{ML}}(x) = \prod_\xi (x - X_{11}(\xi))$, the product over the complex solutions for $T = \mathbf{1}$, a polynomial with rational coefficients whose degree is the number of these solutions, the maximum likelihood degree of the model* [8, Cor. 32]*. Then E_0 is a nonzero polynomial of degree at most $J^{K-1}K^{J-1}$, and E_{ML} divides E_0. In particular $\hat{P}_{111}$ is a root of E_0.*

Proof. The conditions of Theorem 4.7 fail on a proper Zariski-closed set of seed tables, which contains $\mathbf{1}$. A line through $\mathbf{1}$ in a generic direction is not contained in this closed set and therefore meets it in only finitely many points; hence E_η is the eliminant of Theorem 4.7 for all η outside a finite set. By Lemma 4.6, each complex solution ξ for $T = \mathbf{1}$ is nondegenerate, so by the implicit function theorem there is a holomorphic family $\xi(\eta)$ of zeros of the margin equations for $T(\eta)$ with $\xi(0) = \xi$; for small η its coordinates and the quantities $1 + \omega_{jk}$ stay nonzero, so $\xi(\eta)$ is a complex solution for $T(\eta)$, and distinct ξ give distinct families. Hence, for small $\eta \neq 0$ outside the finite set, $E(x,\eta) = g(\eta)\, M_\eta(x)\, Q_\eta(x)$ with $g(\eta) \in \mathbb{C}$, $M_\eta(x) = \prod_\xi (x - X_{11}(\xi(\eta)))$ monic, and Q_η a polynomial. As $\eta \to 0$, $M_\eta \to E_{\mathrm{ML}}$ and $E(x,\eta) \to E_0$. Division with remainder by a monic polynomial is continuous in the coefficients, and the remainder of $E(x,\eta)$ on division by M_η is zero, so the remainder of E_0 on division by E_{ML} is zero. The complex solutions for $T = \mathbf{1}$ are nondegenerate, hence isolated, hence finite in number, and the set they form is defined over $\mathbb{Q}$, so $E_{\mathrm{ML}} \in \mathbb{Q}[x]$; its degree is the maximum likelihood degree [8, Cor. 32]. □

Conjecture 4.10. *Let $T = \mathbf{1}$. For generic margins, E_{ML} divides $\mathcal{R}$ with multiplicity one. With the margins as indeterminates (subject to $\sum_j R_j = \sum_k C_k$), E_{ML} is irreducible over $\mathbb{Q}(m, C, R)$.*

Irreducibility over $\mathbb{Q}$ for particular rational margins can fail on a dense set of margins: if n itself lies in $\mathcal{M}_{\mathbf{1}}$, then $\hat{P} = n$, and E_{ML} has the rational root n_{111}.

5. Fixed points of iterative proportional fitting

In this section n and T are positive tables of any format $I \times J \times K$, and a complex solution is a table $T \circ (a_{ij}b_{ik}c_{jk})$ with complex factors and nonzero entries that has the margins of n. One cycle of iterative proportional fitting is the map $\Phi = \Phi_3 \circ \Phi_2 \circ \Phi_1$, where

$$\Phi_1(P)_{ijk} = P_{ijk}\frac{n_{ij+}}{P_{ij+}}, \qquad \Phi_2(P)_{ijk} = P_{ijk}\frac{n_{i+k}}{P_{i+k}}, \qquad \Phi_3(P)_{ijk} = P_{ijk}\frac{n_{+jk}}{P_{+jk}},$$

defined for complex tables whose margins are nonzero. Each Φ_m multiplies P by a function of two of the three indices, so Φ preserves the set of complex tables $T \circ (a_{ij}b_{ik}c_{jk})$, and a table with the margins of n is fixed by Φ. In particular, every complex solution is a fixed point of Φ; for $2 \times 2 \times K$ tables with distinct ρ_k, or with $T = \mathbf{1}$, and generic margins, the complex solutions correspond bijectively to the roots of E and of F (Theorems 3.2 and 3.5). Iterative proportional fitting started at T converges to the fitted table $\hat{P}$ [9, Thm. 3.2].

Let S be a complex solution and write tables near S as $S \circ \exp(z)$. The tangent space at S of the set above is $\mathcal{V} = \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3$, where $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$ consist of the tables z that depend only on (i, j), on (i, k), and on (j, k), respectively. Let Π_1 be the operator of weighted averaging over the fibers of the first margin,

$$(\Pi_1 z)_{ijk} = \frac{1}{S_{ij+}} \sum_l S_{ijl} z_{ijl},$$

which is defined because $S_{ij+} = n_{ij+} \neq 0$, and define Π_2 and Π_3 in the same way. Each Π_m is idempotent with image $\mathcal{L}_m$. Since S has the margins of n, the derivative of $\log \Phi_m$ at S in the coordinates z is $I - \Pi_m$, and the derivative of Φ at S is

$$D_S = (I - \Pi_3)(I - \Pi_2)(I - \Pi_1),$$

which maps $\mathcal{V}$ into itself. We call S *linearly attracting* if the spectral radius of D_S on $\mathcal{V}$ is less than one, and *linearly unstable* if it is greater than one; in the first case iterative proportional fitting converges to S geometrically from every starting table $T \circ (a_{ij}b_{ik}c_{jk})$ sufficiently close to S.

Proposition 5.1. *The fitted table $\hat{P}$ is linearly attracting.*

Proof. The Hermitian form $\langle u, v\rangle = \sum_{ijk} \hat{P}_{ijk} u_{ijk} \bar{v}_{ijk}$ is an inner product. Each Π_m is the orthogonal projection onto $\mathcal{L}_m$ for it, since $\langle z - \Pi_m z, w\rangle = 0$ for $w \in \mathcal{L}_m$. Hence $\|D_{\hat{P}} z\| \leq \|z\|$, with equality only if $(I - \Pi_m)z = z$ for $m = 1, 2, 3$, that is, only if z is orthogonal to $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$. For $z \in \mathcal{V}$ this means $z \perp \mathcal{V}$, so $z = 0$. Since the unit sphere of $\mathcal{V}$ is compact, the operator norm of $D_{\hat{P}}$ on $\mathcal{V}$ is less than one, and so is its spectral radius. □

Geometric convergence of iterative proportional fitting was proved by Fienberg [12] for two-way tables, by showing that one cycle is a contraction, and for the scaling of a matrix to a doubly stochastic matrix P the asymptotic rate is $\sigma_2(P)^2$, the square of its second largest singular value [15]. The derivative D_S above is the corresponding object at every complex

solution of the three-margin problem. At the solutions other than $\hat{P}$ the weights S_{ijk} are not all positive. For a solution with real entries the quadratic form

$$H_S(z) = \sum_{ijk} S_{ijk} z_{ijk}^2 \qquad (z \in \mathcal{V} \text{ real})$$

decides whether S attracts. Let $\mathbf{A}$ be the design matrix of the model, whose rows are the indicator functions of the cells of the three margins, so that $\mathcal{V}$ is the row space of $\mathbf{A}$, and write $z = \mathbf{A}^\top \vartheta$ with log-linear parameters ϑ. The margins of $S \circ \exp(\mathbf{A}^\top \vartheta)$ are $\mathbf{A}\big(S \circ \exp(\mathbf{A}^\top \vartheta)\big)$, whose derivative at $\vartheta = 0$ is $\mathbf{A}\,\mathrm{diag}(S)\mathbf{A}^\top$, and $H_S(\mathbf{A}^\top \vartheta) = \vartheta^\top \mathbf{A}\,\mathrm{diag}(S)\mathbf{A}^\top \vartheta$. We call S *nondegenerate* if this derivative, as a map from ϑ modulo $\ker \mathbf{A}^\top$ to the span of the columns of $\mathbf{A}$, is invertible; equivalently, H_S is nondegenerate on $\mathcal{V}$. For fixed T and generic n every complex solution is nondegenerate: the map from the parameters modulo $\ker \mathbf{A}^\top$, on the torus, to the space $\mathrm{im}\,\mathbf{A}$ of consistent margins is a dominant morphism between nonsingular varieties of the same dimension (positive tables again provide a Zariski-dense subset of the target), and generic smoothness [13, Cor. III.10.7] applies as in Lemma 4.6. The form H_S is also the Hessian at $z = 0$ of $f_S(z) = \sum_{ijk} S_{ijk}(e^{z_{ijk}} - z_{ijk})$ on $\mathcal{V}$, whose critical points with real z are the real solutions near S.

Theorem 5.2. *Let S be a complex solution with real entries. If H_S is positive definite on $\mathcal{V}$, then S is linearly attracting. If H_S is nondegenerate on $\mathcal{V}$ and not positive definite, then S is linearly unstable. In particular, for generic data a solution with real entries is linearly attracting if and only if it is a nondegenerate local minimum of f_S, and linearly unstable otherwise.*

Proof. On $\mathcal{L}_1$ the form is $H_S(w) = \sum_{ij} S_{ij+} w_{ij}^2 = \sum_{ij} n_{ij+} w_{ij}^2$, which is positive definite, and the same holds on $\mathcal{L}_2$ and $\mathcal{L}_3$. Let $h(z) = \sum_{ijk} S_{ijk} |z_{ijk}|^2$ be the Hermitian extension of H_S. For $w \in \mathcal{L}_1$,

$$\sum_{ijk} S_{ijk}\big(z_{ijk} - (\Pi_1 z)_{ij}\big)\bar{w}_{ij} = \sum_{ij} \bar{w}_{ij}\Big(\sum_k S_{ijk} z_{ijk} - S_{ij+}(\Pi_1 z)_{ij}\Big) = 0,$$

so Π_1 is the projection onto $\mathcal{L}_1$ that is orthogonal for h, and similarly for Π_2 and Π_3. Hence $h(z) = h((I - \Pi_m)z) + h(\Pi_m z) \geq h((I - \Pi_m)z)$, with equality only if $\Pi_m z = 0$, and so $h(D_S z) \leq h(z)$ for all $z \in \mathcal{V}$. If H_S is positive definite, the proof of Proposition 5.1 applies with h in place of the inner product defined by $\hat{P}$. Suppose that H_S is nondegenerate on $\mathcal{V}$ and not positive definite. If $D_S z = \lambda z$ with $|\lambda| = 1$ and $z \neq 0$, then $h(D_S z) = h(z)$ forces $\Pi_1 z = 0$, then $\Pi_2 z = 0$, then $\Pi_3 z = 0$, so $\lambda = 1$ and z is h-orthogonal to $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$, hence to $\mathcal{V}$, contradicting nondegeneracy. So D_S has no eigenvalue of modulus one on $\mathcal{V}$. If its spectral radius were less than one, then $D_S^k z \to 0$ and $H_S(z) = h(z) \geq h(D_S^k z) \to 0$ for every real $z \in \mathcal{V}$, so H_S would be positive semidefinite, hence positive definite; a contradiction. Hence the spectral radius is greater than one. The last statement combines this with the preceding paragraph. □

For $2 \times 2 \times K$ tables the form H_S is not positive definite at any real solution other than the fitted table.

Theorem 5.3. *Let $I = J = 2$, and let $S \neq \hat{P}$ be a complex solution with real entries such that $\sum_{ij} S_{ijk}^{-1} \neq 0$ for every k. Then $H_S(z) < 0$ for some $z \in \mathcal{V}$. For generic margins this condition holds at every complex solution, so $\hat{P}$ is the only complex solution with real entries that is linearly attracting, and every other complex solution with real entries is linearly unstable.*

Proof. Write $S_{ij} = S_{ijk}$ for a fixed stratum k, with positive margins S_{i+} and S_{+j}, those of n. For real x_1, x_2, y_1, y_2, the table equal to $x_i + y_j$ in stratum k and to 0 elsewhere lies in $\mathcal{V}$, and H_S takes the value $Q_k(x,y) = \sum_{ij} S_{ij}(x_i + y_j)^2$ on it. Taking $x = (1, -1)$ and $y_j = -(S_{1j} - S_{2j})/S_{+j}$,

$$Q_k = S_{1+} + S_{2+} - \sum_j \frac{(S_{1j} - S_{2j})^2}{S_{+j}} = 4\sum_j \frac{S_{1j}S_{2j}}{S_{+j}} = \frac{4\prod_{ij} S_{ij}}{S_{+1}S_{+2}} \sum_{ij} \frac{1}{S_{ij}},$$

where the last equality is verified by clearing denominators. The product $\prod_{ij} S_{ij}$ has the sign of the odds ratio $S_{11}S_{22}/(S_{12}S_{21}) = \psi\rho_k$, that is, the sign of ψ.

Suppose $\psi > 0$. Since $S \neq \hat{P}$, the uniqueness of the positive solution (Proposition 3.4) gives a stratum k with a negative entry. The odds ratio is positive and the margins are positive, so the negative entries of that stratum are exactly S_{11}, S_{22} or exactly S_{12}, S_{21}. In the first case $S_{12} > -S_{11} > 0$ and $S_{21} > -S_{22} > 0$, because $S_{1+}, S_{2+} > 0$, so $1/S_{11} + 1/S_{12} < 0$ and $1/S_{21} + 1/S_{22} < 0$; the second case is symmetric. Thus $\sum_{ij} 1/S_{ij} < 0 < \prod_{ij} S_{ij}$, and $Q_k < 0$.

Suppose $\psi < 0$, so that $\prod_{ij} S_{ijk} < 0$ for every k. If $\sum_{ij} 1/S_{ijk} > 0$ for some k, then $Q_k < 0$. Otherwise $\sum_{ij} 1/S_{ijk} < 0$ for every k. Let $\tau_k = (\sum_{ij} 1/S_{ijk})^{-1}$, $\chi = \left(\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right)$, and $z_{ijk} = \tau_k \chi_{ij}/S_{ijk}$. In each stratum the interaction contrast $z_{11k} - z_{12k} - z_{21k} + z_{22k}$ equals 1, so $z_{ijk} - \delta_{(i,j),(1,1)}$ is additive in each stratum, that is, of the form $x_i^{(k)} + y_j^{(k)}$, which lies in $\mathcal{L}_2 + \mathcal{L}_3$; since $\delta_{(i,j),(1,1)}$ lies in $\mathcal{L}_1$, z lies in $\mathcal{V}$. Hence

$$H_S(z) = \sum_k \sum_{ij} S_{ijk}\Big(\frac{\tau_k \chi_{ij}}{S_{ijk}}\Big)^2 = \sum_k \Big(\sum_{ij} \frac{1}{S_{ijk}}\Big)^{-1} < 0.$$

For the genericity statement, note that $\frac{d}{dx}\log$ of the stratum odds ratio at $x = S_{11k}$ is $\sum_{ij} 1/S_{ijk}$, so $\sum_{ij} 1/S_{ijk} = 0$ means that S_{11k} is a double root of $q_k(\,\cdot\,;\psi)$, that is, ψ is a branch point of $x_k^{\pm}$. For generic margins this does not happen at a complex solution: at parameters with all $u_k \neq 0$ by Lemma 3.3, and at the finitely many $\psi \in \{1/\rho_1, \ldots, 1/\rho_K\}$ because the branch sums there take finitely many values, so generic N avoids them. The rest follows from Theorem 5.2, since for generic margins H_S is nondegenerate. □

Complex solutions other than $\hat{P}$ can be linearly attracting (Example A.6), so Theorem 5.3 does not extend to them.

Appendix A. Examples

Example A.1. Let $K = 3$ and

$$(t_{ij1}) = \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}, \quad (t_{ij2}) = \begin{pmatrix} 1 & 4 \\ 3 & 1 \end{pmatrix}, \quad (t_{ij3}) = \begin{pmatrix} 3 & 1 \\ 2 & 2 \end{pmatrix},$$

$$(n_{ij1}) = \begin{pmatrix} 12 & 5 \\ 8 & 10 \end{pmatrix}, \quad (n_{ij2}) = \begin{pmatrix} 7 & 14 \\ 11 & 6 \end{pmatrix}, \quad (n_{ij3}) = \begin{pmatrix} 9 & 6 \\ 4 & 13 \end{pmatrix}.$$

The polynomial E of Theorem 3.2 has degree 12 and is irreducible over $\mathbb{Q}$. Its real roots are 0.1734984020, 0.1779707973, 0.9717681410, 1.2197185992, 1.9477252408, and 2.8146073184, and the root given by Proposition 3.4 is $\hat{\psi} = 1.2197185992$.

Example A.2. With the counts of Example A.1 and $T = \mathbf{1}$, $E(\psi)$ is a constant multiple of $(\psi-1)^5 F(\psi)$, where

$$F(\psi) = 4236232000\,\psi^7 + 17421462240\,\psi^6 + 27735604391\,\psi^5 + 5810133454\,\psi^4 \\ - 45268267044\,\psi^3 - 75784685929\,\psi^2 - 54554310492\,\psi - 16963914240$$

is irreducible over $\mathbb{Q}$ and has the single real root $\hat{\psi} = 1.4210239993$, the maximum likelihood estimate of the common odds ratio.

Example A.3. Let $J = K = 3$, with rows indexed by j and columns by k,

$$(t_{1jk}) = \begin{pmatrix} 3 & 3 & 4 \\ 5 & 1 & 1 \\ 5 & 5 & 2 \end{pmatrix}, \quad (t_{2jk}) = \begin{pmatrix} 2 & 5 & 3 \\ 2 & 5 & 2 \\ 3 & 4 & 3 \end{pmatrix}, \quad (n_{1jk}) = \begin{pmatrix} 1 & 1 & 8 \\ 7 & 8 & 5 \\ 8 & 3 & 5 \end{pmatrix}, \quad (n_{2jk}) = \begin{pmatrix} 8 & 2 & 3 \\ 2 & 5 & 9 \\ 2 & 4 & 4 \end{pmatrix}.$$

The polynomial E_{11} of Theorem 4.7 has degree 81 and is irreducible over $\mathbb{Q}$. It has 33 real roots, of which five lie in the interval $(0, m_{11}) = (0, 9)$: 2.2887616252, 2.7354972215, 3.2838372433, 4.2591282778, and 4.5510375640. The entry of the fitted table is $\hat{P}_{111} = 3.2838372433$.

Example A.4. With the counts of Example A.3 and $T = \mathbf{1}$, the polynomial $\mathcal{R}(x)$ has degree 89. Its factor E_{ML} has degree 31, is irreducible over $\mathbb{Q}$, occurs with multiplicity one, and has the single real root $\hat{P}_{111} = 4.1615312803$.

Example A.5. Let $J = 3$, $K = 4$, $T = \mathbf{1}$, and

$$(n_{1jk}) = \begin{pmatrix} 8 & 3 & 5 & 8 \\ 2 & 3 & 2 & 5 \\ 9 & 2 & 4 & 4 \end{pmatrix}, \qquad (n_{2jk}) = \begin{pmatrix} 9 & 2 & 5 & 3 \\ 1 & 7 & 1 & 3 \\ 5 & 5 & 2 & 9 \end{pmatrix}.$$

The polynomial $\mathcal{R}(x)$ has degree 477. Its factor E_{ML} has degree 115, is irreducible over $\mathbb{Q}$, occurs with multiplicity one, and has the single real root $\hat{P}_{111} = 9.9706390967$.

Example A.6. Let $K = 3$ and

$$(t_{ij1}) = \begin{pmatrix} 197 & 53 \\ 98 & 102 \end{pmatrix}, \quad (t_{ij2}) = \begin{pmatrix} 23 & 5 \\ 23 & 57 \end{pmatrix}, \quad (t_{ij3}) = \begin{pmatrix} 142 & 79 \\ 85 & 48 \end{pmatrix},$$

$$(n_{ij1}) = \begin{pmatrix} 132 & 430 \\ 663 & 34 \end{pmatrix}, \quad (n_{ij2}) = \begin{pmatrix} 77 & 439 \\ 472 & 34 \end{pmatrix}, \quad (n_{ij3}) = \begin{pmatrix} 107 & 518 \\ 785 & 3493 \end{pmatrix}.$$

The polynomial E has degree 12, and $\hat{\psi} = 0.0103229485$. At the pair of complex conjugate solutions with $\psi = -0.0429312847 \pm 0.0067384684i$, the spectral radius of D_S on $\mathcal{V}$ is 0.5347, so these solutions are linearly attracting; at the fitted table it is 0.7310.

Example A.7. Let $K = 2$, $T = \mathbf{1}$, and

$$(n_{ij1}) = \begin{pmatrix} 10 & 1 \\ 3 & 1 \end{pmatrix}, \qquad (n_{ij2}) = \begin{pmatrix} 1 & 1 \\ 50 & 10 \end{pmatrix}.$$

Then $N = 11 \neq \sum_k r_k c_k / N_k$, and $E(\psi)$ is a constant multiple of $(\psi-1)F(\psi)$ with

$$F(\psi) = 10812\,\psi^3 + 3686\,\psi^2 - 5687\,\psi - 2420,$$

irreducible over $\mathbb{Q}$, with the three real roots -0.5882641171, -0.5054368472, and 0.7527834652. Each is the common odds ratio of a complex solution with real entries, and $\hat{\psi} = 0.7527834652$.

References

[1] A. Agresti, *Categorical Data Analysis*, 3rd ed., Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, NJ, 2013.

[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] F. V. Atkinson, *Multiparameter Eigenvalue Problems: Matrices and Compact Operators*, Mathematics in Science and Engineering, vol. 82, Academic Press, New York, 1972.

[4] M. S. Bartlett, *Contingency table interactions*, Suppl. J. Roy. Statist. Soc. **2** (1935), no. 2, 248–252, doi:10.2307/2983639.

[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] N. E. Breslow and N. E. Day, *Statistical Methods in Cancer Research, Vol. I: The Analysis of Case-Control Studies*, IARC Scientific Publications, no. 32, International Agency for Research on Cancer, Lyon, 1980.

[8] 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.

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

[10] C. D'Andrea, *Macaulay style formulas for sparse resultants*, Trans. Amer. Math. Soc. **354** (2002), 2595–2629, doi:10.1090/S0002-9947-02-02910-0.

[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] S. E. Fienberg, *An iterative procedure for estimation in contingency tables*, Ann. Math. Statist. **41** (1970), no. 3, 907–917, doi:10.1214/aoms/1177696968.

[13] R. Hartshorne, *Algebraic Geometry*, Graduate Texts in Mathematics, vol. 52, Springer, New York, 1977, doi:10.1007/978-1-4757-3849-0.

[14] M. E. Hochstenbach, C. Mehl, and B. Plestenjak, *Solving singular generalized eigenvalue problems by a rank-completing perturbation*, SIAM J. Matrix Anal. Appl. **40** (2019), no. 3, 1022–1046, doi:10.1137/18M1188628; numbering as in arXiv:1805.07657.

[15] P. A. Knight, *The Sinkhorn–Knopp algorithm: convergence and applications*, SIAM J. Matrix Anal. Appl. **30** (2008), no. 1, 261–275, doi:10.1137/060659624.

[16] T. Košir, *Finite-dimensional multiparameter spectral theory: the nonderogatory case*, Linear Algebra Appl. **212–213** (1994), 45–70, doi:10.1016/0024-3795(94)90396-4.

[17] N. Mantel and W. Haenszel, *Statistical aspects of the analysis of data from retrospective studies of disease*, J. Natl. Cancer Inst. **22** (1959), no. 4, 719–748, doi:10.1093/jnci/22.4.719.

[18] H. W. Norton, *Calculation of chi-square for complex contingency tables*, J. Amer. Statist. Assoc. **40** (1945), no. 230, 251–258, doi:10.2307/2280138.

[19] S. N. Roy and M. A. Kastenbaum, *On the hypothesis of no "interaction" in a multi-way contingency table*, Ann. Math. Statist. **27** (1956), no. 3, 749–757, doi:10.1214/aoms/1177728181.

[20] S. Sivarajan, *An explicit determinantal formula for Sinkhorn limits of matrices and tensors*, hexagon: 2611.00433 (2026).

[21] F. Teuscher, *The quantification of Simpson's paradox and other contributions to contingency table theory*, PLoS ONE **17** (2022), no. 2, e0262502, doi:10.1371/journal.pone.0262502.