Degree-uniform covariance-deficit estimation for Lorentzian count laws

Scott Schubr
Luminary Oddities

Abstract

We estimate the normalized covariance deficit of a homogeneous Lorentzian count law from independent full count vectors, without knowing the mean or imposing a minimum category-probability assumption. For total count k and m categories, we prove an operator second-moment bound of $10m$ for the full-count factorial-pair matrix, uniformly in k. The proof combines the classical down-up Poincaré inequality for strongly log-concave measures with an exact one-ball increment and a diagonal constraint that pays for rare categories. Independent mean smoothing and a classical robust matrix-mean estimator then give an operator-norm confidence bound using $O(m\varepsilon^{-2}\log(m/\delta))$ count observations. The category and accuracy dependence, for a guarantee uniform over total counts, matches a degree-two stable-subclass lower bound up to logarithms. We give explicit constants and prove the normalization error bound on the full category space, including unknown zero-mean categories. We also derive a relative-entropy bound of $(k/2)\operatorname{tr} D$ to the matching multinomial and give a confidence bound relative to the estimated mean. The statistical result concerns sample complexity; it does not require running a down-up chain and does not assert near-linear computation.

Keywords. covariance estimation; Lorentzian polynomials; count distributions; rare categories; spectral gap; robust matrix mean.
2020 Mathematics Subject Classification. Primary 62H12; Secondary 60E15.

1 Introduction

Let $A = (A_1, \dots, A_m)$ be a nonnegative integer vector with fixed total $\sum_i A_i = k$. Write $q = \mathbb{E}A/k$, $Q = \operatorname{diag}(q)$, and $u = \sqrt{q}$. The object studied here is the normalized covariance deficit

$$D = I - uu^T - \frac{1}{k} Q^{-1/2} \operatorname{Cov}(A) Q^{-1/2}.$$

This formula is defined on the positive support of q and extended by zero on zero-mean categories. For a homogeneous Lorentzian generating polynomial, D is positive semidefinite and has norm at most one. Thus D measures a reduction of covariance relative to the matching multinomial law.

The same matrix is a rescaled canonical dependence matrix. If J is the joint law of two ordered distinct balls sampled uniformly from one count, then

$$Q^{-1/2}(J - qq^T)Q^{-1/2} = -D/(k-1).$$

The statistical target is therefore an established normalized dependence operator. Unknown means are a serious issue when some categories are rare: entrywise inverse-probability normalization can be large even though the target has bounded norm.

Degree-independent covariance estimation for Poisson multinomial distributions predates this work. Daskalakis, Kamath and Tzamos [4, Lemma 8 and Appendix D.1] estimate their mean and covariance using $O(m^4\varepsilon^{-2})$ samples, independent of the number of categorical summands; the displayed multiplicative covariance guarantee assumes a positive lower spectral bound. Their model is a sum of independent categorical vectors. Thus independence of the total count is not, by itself, the distinction here. Our target is the normalized deficit, the category dependence is linear up to logarithms, and the theorem allows arbitrary Lorentzian coefficient laws with unknown means and no minimum category mass. We do not claim that these two covariance-error metrics are interchangeable.

Normalized-dependence estimation and empirical marginal normalization also have established antecedents. Makur, Wornell and Zheng [6, Section III] give finite-sample singular-mode bounds using known true margins and a minimum-marginal parameter. Huang, Makur, Wornell and Zheng [5, Section 6.2] study the same canonical dependence framework: their finite-sample concentration retains a minimum-marginal parameter, and they separately analyze empirical normalization. We use this established target and concentration setting. The present guarantee uses the Lorentzian sign and derivative structure to control the full factorial-pair matrix and normalization error without such a parameter.

Unknown normalization is controlled by a rare-category diagonal bound and an aggregate mean-smoothing estimate, both proved below. The degree-uniform variance bound comes from a direct analysis of the full count. The diagonal factor $A_i(A_i - 1)$ cancels the self-pair contribution of a singleton count. Keeping this cancellation through a one-ball deletion gives a variance bound with no degree factor.

The structural input is classical. Brändén and Huh [1] prove that polarization preserves Lorentzian polynomials and identify the homogeneous Lorentzian and strong-log-concavity conditions. Anari, Liu, Oveis Gharan and Vinzant [2] prove the reciprocal-degree spectral-gap bound for the down-up walk on a strongly log-concave set measure. We transfer that Poincaré inequality to count vectors and apply it to the factorial-pair matrix. Minsker's robust matrix mean [3] then supplies concentration from an uncentered second-moment bound, without an almost-sure matrix envelope.

The contribution is the resulting cancellation-preserving operator bound, its full-space unknown-mean normalization, and their explicit statistical combination for all Lorentzian count laws. No joint independent categorical factorization is assumed. The distributional consequence is proved here through an additional posterior and thinning argument. Real-stable laws form a subclass of the model considered here, but the proof does not use real-rooted one-dimensional slices.

2 Model and estimator

We use the homogeneous Lorentzian class of Brändén and Huh [1]. Equivalently, a homogeneous polynomial with nonnegative coefficients is Lorentzian precisely when, for every multiindex $\alpha \geq 0$ including $\alpha = 0$, the derivative $\partial^\alpha p$ is identically zero or log-concave on the positive orthant [1, Theorem 2.30]. Let

$$p(z) = \mathbb{E}z^A = \sum_{|a|=k} p_a z^a, \qquad p(\mathbf{1}) = 1,$$

be such a polynomial of degree $k \geq 2$. The structural promise is part of the model, rather than a property verified by the observations. Put

$$\kappa = k - 1, \quad q = \mathbb{E}A/k, \quad Q = \operatorname{diag}(q), \quad u = \sqrt{q}, \quad P_q = I - uu^T, \qquad D = P_q - k^{-1}Q^{-1/2}\operatorname{Cov}(A)Q^{-1/2}.$$

If $q_i = 0$, then $A_i = 0$ almost surely. Define D on the positive support and embed it with zero rows and columns in $\mathbb{R}^m$. In particular, the zero-mean coordinates do not carry an additional identity block. All norms below are operator norms unless a Frobenius norm is indicated.

Define the scalar function

$$\psi(x) = \begin{cases} \log(1 + x + x^2/2), & x \geq 0, \\ -\log(1 - x + x^2/2), & x < 0. \end{cases} \tag{1}$$

For a real symmetric matrix, ψ denotes spectral functional calculus.

Theorem 2.1 (Degree-uniform unknown-mean estimation)**.** *Fix $0 < \varepsilon \leq 1$, $0 < \delta < 1$, and $L = \log(4m/\delta)$. Take two independent batches of iid observations from the count law, with integer sizes satisfying*

$$N_0 \geq \frac{10752 m \kappa L}{k\varepsilon^2}, \qquad N_1 \geq \frac{15392 m L}{\varepsilon^2}. \tag{2}$$

From the first batch construct

$$\widehat{q}_i = \frac{1}{kN_0} \sum_{a=1}^{N_0} A_i^{(a)}, \quad t = \frac{L}{kN_0}, \quad r_i = \frac{\widehat{q}_i + 2t}{1 + 2mt}, \quad R = \operatorname{diag}(r), \quad v = \sqrt{r}, \quad P_r = I - vv^T.$$

For every second-stage observation form the full symmetric matrix

$$Z_r(A) = -\frac{1}{k} P_r R^{-1/2} (AA^T - \operatorname{diag} A) R^{-1/2} P_r. \tag{3}$$

Set $V = 481m$, $a_1 = 2L$, $\theta = \sqrt{a_1/(N_1 V)}$, and

$$\widehat{D} = \frac{1}{N_1 \theta} \sum_{a=1}^{N_1} \psi(\theta Z_r(A^{(a)})). \tag{4}$$

Then, with probability at least $1 - \delta$,

$$\|\widehat{D} - D\| < \varepsilon/2. \tag{5}$$

In particular, for

$$U = \max\{0, \lambda_{\max}(\widehat{D})\} + \varepsilon/2, \tag{6}$$

the same event gives

$$\|D\| \leq U \leq \|D\| + \varepsilon.$$

The guarantee is uniform in k and independent of the smallest positive q_i. No knowledge of the positive mean support is required.

The total count k, named alphabet and observations are known. Within a count, balls can be dependent; independence is required only between count observations and between the two batches. The constants are conservative. For $m = 1$, the target is identically zero and can be returned without data.

Remark 2.2 (A conditioned count model beyond real stability)**.** *Condition a* $\mathrm{Bin}(4, 1/2)$ *count to lie in* $\{1, 2, 3\}$*, and record its success and failure counts. Its normalized generating polynomial is*

$$p(x, y) = \frac{(x + y)^4 - x^4 - y^4}{14} = \frac{xy(4x^2 + 6xy + 4y^2)}{14}.$$

Its binomially normalized coefficients are $(0, 1/14, 1/14, 1/14, 0)$*, a log-concave sequence with interval support; the binary characterization in [1] makes it Lorentzian. The quadratic factor has negative discriminant, so this polynomial is not real stable. Its mean fraction is* $(1/2, 1/2)$*, its success variance is* $4/7$*, and* $\|D\| = 3/7$*. This illustrates a conditioned count law covered by Theorem 2.1 without an independent categorical factorization. It is an example of the model class, not a new distributional construction or evidence of economical sample constants.*

3 Covariance and derivative structure

Lemma 3.1 (Covariance domination and marginal concentration)**.** *For the model above,*

$$0 \le D \le P_q, \quad Du = 0, \quad D_{ii} \le \kappa q_i. \tag{7}$$

On the positive orthant, $p(z) \le (q^T z)^k$*. Consequently, for every real* s*,*

$$\log \mathbb{E} e^{s(A_i - kq_i)} \le kq_i(e^s - 1 - s). \tag{8}$$

Proof. A homogeneous log-concave polynomial of degree k has a concave degree root. One can see the required Hessian inequality directly. At a positive point z, decompose a direction as $w = az + w_0$, where $\nabla p(z)^T w_0 = 0$. Euler's identities give $\nabla p(z)^T z = kp(z)$ and $\nabla^2 p(z) z = (k-1)\nabla p(z)$. Log concavity implies $w_0^T \nabla^2 p(z) w_0 \le 0$, so

$$w^T \nabla^2 p(z) w \le a^2 k\kappa p(z) = \frac{\kappa}{kp(z)} (w^T \nabla p(z))^2.$$

This is precisely the Hessian condition for concavity of $p^{1/k}$. Its tangent at $\mathbf{1}$ is $q^T z$, which proves the tangent bound.

At $\mathbf{1}$ the same Hessian inequality gives $\nabla^2 p(\mathbf{1}) \le k\kappa q q^T$. Using

$$\mathrm{Cov}(A) = \nabla^2 p(\mathbf{1}) + kQ - k^2 q q^T$$

proves $D \succeq 0$. Covariance positivity gives $D \preceq P_q$, and the fixed total gives $Du = 0$. Integer-valuedness gives $\mathrm{Var}(A_i) \ge kq_i - k^2 q_i^2$, hence $D_{ii} = 1 - q_i - \mathrm{Var}(A_i)/(kq_i) \le \kappa q_i$. For the moment generating function, set one coordinate of z to e^s and the rest to one in the tangent bound, and use $\log(1+x) \le x$. Zero means follow by the prescribed embedding. □

Let I be the label of a uniformly selected ball. Conditional on $I = i$, write $B = A - e_i$. Its generating polynomial is

$$p_i(z) = \frac{\partial_i p(z)}{kq_i},$$

which is Lorentzian of degree κ. It may have additional zero-mean coordinates. For ordered distinct ball labels define

$$J_{ij} = \frac{\mathbb{E}[A_i(A_j - \mathbf{1}_{i=j})]}{k\kappa}.$$

Lemma 3.2 (Pair and residual means)**.** *On the parent-positive support,*

$$J = qq^T - Q^{1/2} D Q^{1/2}/\kappa, \qquad 0 \le J_{ij} \le 2q_i q_j. \tag{9}$$

The mean fraction of the residual conditional on $I = i$ *is*

$$t_j^{(i)} = J_{ij}/q_i, \qquad 0 \le t_j^{(i)} \le 2q_j. \tag{10}$$

Proof. The identity follows from $\mathbb{E}AA^T = kQ + k\kappa J$ and the definition of D. Since $D \succeq 0$, $|D_{ij}| \le \sqrt{D_{ii} D_{jj}} \le \kappa\sqrt{q_i q_j}$; substitution yields the upper bound, and J is nonnegative by definition. The conditional mean formula follows by size biasing with A_i. □

4 A count-space Poincaré inequality

Lemma 4.1 (Down-up inequality)**.** *Let B be obtained from a stationary count A by deleting one uniformly selected ball. For every real-valued function f on the count support,*

$$\operatorname{Var}(f(A)) \le k\mathbb{E}\operatorname{Var}(f(A) \mid B). \tag{11}$$

For vector-valued f, the same statement holds with $\operatorname{Var}(f) = \mathbb{E}\|f - \mathbb{E}f\|^2$.

Proof. Polarize each variable into k clones. A clone set S of size k whose block counts are a has probability

$$\mu(S) = p_a \Big/ \prod_i \binom{k}{a_i}.$$

Its count marginal is the original law. Polarization preserves the Lorentzian property [1, Proposition 3.1]; the resulting homogeneous multiaffine polynomial is strongly log-concave [1, Theorem 2.30]. The set down-up chain, which removes a uniform element and then chooses a containing set with weight proportional to μ, has spectral gap at least $1/k$ [2, Theorem 1.1].

For a residual clone set with counts b, the combined up-weight for adding a clone from block i is

$$\frac{(k-b_i)p_{b+e_i}}{\binom{k}{b_i+1}\prod_{j\neq i}\binom{k}{b_j}} = \frac{(b_i+1)p_{b+e_i}}{\prod_j \binom{k}{b_j}}.$$

Thus the count up-step selects i with probability proportional to $(b_i+1)p_{b+e_i}$. This is exactly the conditional law of $A = B + e_i$ given B after uniform deletion, since their stationary joint probability is $(b_i+1)p_{b+e_i}/k$. The clone chain therefore lumps to the count down-up chain.

Lift f to the clone space by its block counts. Its variance is unchanged. Given a residual set, old and new parents are independent conditional draws. Hence the Dirichlet form, one half the expected squared change, equals $\mathbb{E}\operatorname{Var}(f(A) \mid B)$, with no extra factor two. The spectral-gap inequality proves (11). Sum the scalar inequality over coordinates for vector-valued f. □

5 The full-count operator second moment

Theorem 5.1 (Degree-uniform factorial-pair variance)**.** *On positive mean support, define*

$$Z_q(A) = -\frac{1}{k}P_q Q^{-1/2}(AA^T - \operatorname{diag} A)Q^{-1/2}P_q. \tag{12}$$

Then $\mathbb{E}Z_q = D$ and

$$\mathbb{E}Z_q^2 \le 10mI. \tag{13}$$

The same bound holds after zero extension.

Proof. Let the positive support size be $s \le m$, put $d = s - 1$, and work on $\mathcal{H} = u^\perp \subset \mathbb{R}^s$. Define

$$\xi_i = e_i/\sqrt{q_i} - u.$$

For a uniform ordering $I_1, \dots, I_k$ of the observed multiset,

$$Z_q = -\frac{1}{k}\sum_{a\neq b}\xi_{I_a}\xi_{I_b}^T.$$

The labels are exchangeable, not independent. The pair identity (9) gives $\mathbb{E}Z_q = D$. If $S = \sum_a \xi_{I_a}$, then

$$\mathbb{E}S = 0, \qquad \mathbb{E}SS^T = k(P_q - D) \le kP_q. \tag{14}$$

For a residual B of size κ, set $S_B = \sum_i B_i \xi_i$ and

$$Z_{\text{base}}(B) = -\frac{1}{k} \sum_{a \ne b \le \kappa} \xi_{I_a} \xi_{I_b}^T.$$

The denominator in this reference matrix is still k. Adding a ball of label i gives the exact increment

$$Z_q(B + e_i) - Z_{\text{base}}(B) = -\frac{1}{k}(\xi_i S_B^T + S_B \xi_i^T). \tag{15}$$

For a unit $x \in \mathcal{H}$, apply Lemma 4.1 to $f(A) = Z_q(A)x$. Conditional variance is no larger than the squared distance from the B-measurable reference $Z_{\text{base}}(B)x$. Hence

$$\mathbb{E}\|Z_q x - Dx\|^2 \le \frac{2}{k}(T_1 + T_2), \tag{16}$$

where, in the joint deletion experiment $A = B + e_I$,

$$T_1 = \mathbb{E}\big[\|\xi_I\|^2 (x^T S_B)^2\big], \qquad T_2 = \mathbb{E}\big[\|S_B\|^2 (x^T \xi_I)^2\big].$$

Condition now on the removed label $I = i$. Lemma 3.2 gives

$$\mathbb{E}[S_B \mid I = i] = \kappa Q^{-1/2}(t^{(i)} - q) = -D\xi_i. \tag{17}$$

The residual is Lorentzian, so Lemma 3.1 applied to it gives

$$\begin{aligned} \operatorname{Cov}(S_B \mid I = i) &= Q^{-1/2} \operatorname{Cov}(B \mid I = i) Q^{-1/2} \\ &\le \kappa Q^{-1/2}[\operatorname{diag}(t^{(i)}) - t^{(i)} t^{(i)T}] Q^{-1/2} \le 2\kappa I_{\mathcal{H}}. \end{aligned} \tag{18}$$

The final inequality is on $\mathcal{H}$, using $t_j^{(i)} \le 2q_j$. Its trace is at most $2\kappa d$. Degree-one residuals satisfy the same covariance bound directly; zero residual means require no inversion.

The categorical identities $\mathbb{E}\xi_I \xi_I^T = P_q$, $\mathbb{E}\|\xi_I\|^2 = d$, and $\|\xi_i\|^2 = 1/q_i - 1$ now imply

$$\begin{aligned} T_1 &\le 2\kappa d + \sum_i (1/q_i - 1)(Dx)_i^2 \\ &\le 2\kappa d + \|DQ^{-1}D\| \le 2\kappa d + \operatorname{tr}(DQ^{-1}D) \le \kappa(2d + s). \end{aligned}$$

For the last step, $D^2 \le D$ and $D_{ii} \le \kappa q_i$ give

$$\operatorname{tr}(DQ^{-1}D) = \sum_i \frac{(D^2)_{ii}}{q_i} \le \sum_i \frac{D_{ii}}{q_i} \le \kappa s.$$

Similarly, $\mathbb{E}(x^T \xi_I)^2 = 1$ and the pointwise bound

$$\|D\xi_i\|^2 = \frac{(D^2)_{ii}}{q_i} \le \frac{D_{ii}}{q_i} \le \kappa$$

give

$$T_2 \le 2\kappa d + \kappa.$$

Consequently $T_1 + T_2 \le \kappa(5s - 3)$. By (16),

$$x^T \mathbb{E} Z_q^2 x = \mathbb{E}\|Z_q x - Dx\|^2 + \|Dx\|^2 \le \frac{2\kappa}{k}(5s - 3) + 1 \le 10s.$$

Both Z_q and D vanish on u. For $s = 1$ they vanish identically. This proves (13) in all cases. □

The proof retains the factorial correction diag A throughout the one-ball increment. An uncorrected whitened covariance can have divergent rare-coordinate fourth moments even when (13) holds. The down-up chain is used only in the proof of a population inequality; it is not run to produce or process the observations.

6 Unknown means and robust concentration

Lemma 6.1 (Smoothed mean control)**.** *For the first batch in Theorem 2.1, with probability at least* $1 - \delta/2$*,*

$$q_i/r_i \le 4 \quad \textit{for all } i, \qquad h := \chi^2(q\|r) \le 42mt \le \frac{\varepsilon^2}{256\kappa}. \tag{19}$$

Proof. The marginal mgf (8) and independence across counts give the usual two-sided scalar Bernstein bound. A union bound yields

$$|\widehat{q}_i - q_i| \le \sqrt{2q_i t} + 2t/3 \quad \text{for all } i$$

except with probability $2me^{-L} = \delta/2$. This follows directly by Chernoff optimization from $e^s - 1 - s \le s^2/[2(1 - s/3)]$ for $0 < s < 3$; the lower-tail bound is at least as strong. No independence across categories is needed.

Put $h_i = \widehat{q}_i + 2t$ and $c = 1 + 2mt$. Since $\sqrt{2q_i t} \le q_i/2 + t$, we have $q_i/h_i \le 2$; also $h_i \ge 2t$. The budget gives $mt \le 1/2$, so $c \le 2$ and $q_i/r_i \le 4$. Squaring the coordinate estimate gives

$$(q_i - \widehat{q}_i)^2 \le 4q_i t + 8t^2/9, \qquad \sum_i \frac{(q_i - \widehat{q}_i)^2}{h_i} \le \frac{76}{9}mt.$$

Thus $E_0 := \sum_i (q_i - h_i)^2/h_i \le (188/9)mt < 21mt$. Since $\sum_i h_i = c$,

$$\chi^2(q\|r) = c\sum_i q_i^2/h_i - 1 = cE_0 - (c-1)^2 \le 42mt.$$

The first budget in (2) proves the last assertion. This proof includes unknown zero-mean categories. □

Lemma 6.2 (Full-space whitening and second moment)**.** *Suppose* r *satisfies* (19)*, and let* $C_r = \mathbb{E}Z_r(A)$ *with* r *fixed. Then*

$$\|C_r - D\| \le 2\sqrt{\kappa h} + 2\kappa h \le 17\varepsilon/128 < \varepsilon/4, \qquad \mathbb{E}Z_r(A)^2 \le 481mI. \tag{20}$$

Proof. Let T map the parent-positive support to the full category space by the diagonal factors $\sqrt{q_i/r_i}$ and zero extension. Put $B = P_r T|_{\mathcal{H}}$ and $b = (q - r)/\sqrt{r}$. Then $\|B\| \le 2$, $v^T b = 0$, $\|b\|^2 = h$, and for every positive category,

$$P_r e_i/\sqrt{r_i} = B\xi_i + b.$$

Writing $S = \sum_a \xi_{I_a}$ as in (14), direct expansion of the ordered-pair sum gives

$$Z_r = BZ_q B^T - \frac{\kappa}{k}(BSb^T + bS^T B^T) - \kappa bb^T, \qquad C_r = BDB^T - \kappa bb^T. \tag{21}$$

Since $T^T v = u$ and $D^{1/2}u = 0$, one has $P_r TD^{1/2} = TD^{1/2}$ and hence $BDB^T = TDT^T$. With all operators zero-extended as appropriate, let $F = (T - I)D^{1/2}$. By (7),

$$\|F\|^2 \le \|F\|_F^2 = \sum_{i:q_i>0} (\sqrt{q_i/r_i} - 1)^2 D_{ii}$$

$$\le \kappa \sum_i \frac{(q_i - r_i)^2}{r_i} \frac{q_i}{(\sqrt{q_i} + \sqrt{r_i})^2} \le \kappa h.$$

Expansion of $TDT^T - D$ and $\|D^{1/2}\| \le 1$ give $\|TDT^T - D\| \le 2\sqrt{\kappa h} + \kappa h$. Equation (21) proves the perturbation bound, including zero means. Its numerical upper bound uses $2\sqrt{\kappa h} \le \varepsilon/8$ and $2\kappa h \le \varepsilon^2/128$.

For the second moment, denote the three terms in (21) by X, Y, and C, respectively. Theorem 5.1 gives

$$\mathbb{E}X^2 \le \|B\|^2 B\mathbb{E}Z_q^2 B^T \le 160mI.$$

For $z = BS$, (14) gives $\mathbb{E}zz^T \le 4kI$ and $\mathbb{E}\|z\|^2 \le 4km$. Pointwise,

$$(zb^T + bz^T)^2 \le 2hzz^T + 2\|z\|^2 bb^T.$$

Therefore

$$\mathbb{E}Y^2 \le 8\kappa^2 h(1+m)I/k \le 16khmI \le mI/8,$$

where $kh \le k\varepsilon^2/(256\kappa) \le 1/128$. Also $\|C^2\| = \kappa^2 h^2 \le 1/65536$. The Hermitian square inequality $(X+Y+C)^2 \le 3(X^2+Y^2+C^2)$ now yields

$$\mathbb{E}Z_r^2 \le 3(160m + m/8 + 1/65536)I < 481mI.$$

□

Lemma 6.3 (Classical robust matrix mean)**.** *Let $Y_1, \dots, Y_N$ be iid real symmetric m by m matrices, and let $V > 0$ satisfy $V \ge \|\mathbb{E}Y_1^2\|$. For $a > 0$, put $\theta = \sqrt{a/(NV)}$ and $\widehat{M} = (N\theta)^{-1}\sum_j \psi(\theta Y_j)$. Then*

$$\mathbb{P}\{\|\widehat{M} - \mathbb{E}Y_1\| \ge \sqrt{Va/N}\} \le 2me^{-a/2}. \tag{22}$$

Proof. This is Minsker's second-moment matrix-mean bound [3, Theorem 3.1 and Remark 2], with the function (1). The scalar logarithmic sandwich for ψ and the matrix trace-mgf argument give the two-sided tail $2m\exp\{-N\theta\eta + N\theta^2 V/2\}$ at radius η. Substitute $\eta = \sqrt{Va/N}$ and the displayed θ. The input is an uncentered second-moment bound; no range assumption is used. □

Proof of Theorem 2.1. Condition on the first batch and the event of Lemma 6.1. The second-stage matrices are iid with fixed r. By Lemma 6.2, their mean is C_r and their second moment is bounded by $481mI$. Apply Lemma 6.3 with $V = 481m$ and $a = 2L$. The second budget makes $\sqrt{Va/N_1} \le \varepsilon/4$ and the failure probability at most $2me^{-L} = \delta/2$. On this event,

$$\|\widehat{D} - D\| \le \|\widehat{D} - C_r\| + \|C_r - D\| < \varepsilon/4 + \varepsilon/4 = \varepsilon/2.$$

Union with the first-stage failure probability proves (5). Since $D \ge 0$, Weyl's inequality and truncation at zero show that $\max\{0, \lambda_{\max}(\widehat{D})\}$ differs from $\|D\|$ by at most $\varepsilon/2$. Equation (6) follows. □

7 An entropy and multinomial-approximation consequence

Write KL for relative entropy with natural logarithms and use probability total variation, $\mathrm{TV}(\mu,\nu) = \frac{1}{2}\|\mu - \nu\|_1$. The same count-space Poincaré inequality controls a conditional label posterior and yields a distributional consequence.

Theorem 7.1. *For every homogeneous Lorentzian count law of degree $k \ge 2$,*

$$\mathrm{KL}(\mathcal{L}(A)\|\,\mathrm{Mult}(k,q)) \le \frac{k}{2}\operatorname{tr} D, \qquad \mathrm{TV}(\mathcal{L}(A), \mathrm{Mult}(k,q)) \le \min\{1, \sqrt{k\operatorname{tr} D/4}\}. \tag{23}$$

For any strictly positive probability vector r,

$$\mathrm{KL}(\mathcal{L}(A)\|\,\mathrm{Mult}(k,r)) \le \frac{k}{2}\operatorname{tr} D + k\,\mathrm{KL}(q\|r). \tag{24}$$

Proof. First consider an arbitrary degree-$n \geq 2$ Lorentzian count law with mean fraction q and deficit D_n. Give its counts a uniform ordering $X_1, \ldots, X_n$, and let B be the histogram of the first $n-1$ labels. Write $t_B(j) = \mathbb{P}(X_n = j \mid B)$ and $M = Q^{-1/2} \operatorname{Cov}(t_B) Q^{-1/2}$. For $f(A) = x^T Q^{-1/2} A$, the two sides of the count Poincaré inequality are

$$\operatorname{Var} f(A) = n x^T (P_q - D_n) x, \qquad n\mathbb{E} \operatorname{Var}(f(A) \mid B) = n x^T (P_q - M) x.$$

Here $\mathbb{E} t_B = q$, and conditional on B only the last label varies. Lemma 4.1 therefore gives $M \preceq D_n$, whence

$$\mathbb{E} \operatorname{KL}(t_B \| q) \leq \mathbb{E} \chi^2(t_B \| q) = \operatorname{tr} M \leq \operatorname{tr} D_n. \tag{25}$$

The first inequality follows from $\log x \leq x - 1$.

Return to the degree-k law. For $1 \leq l \leq k$, the histogram of $X_1, \ldots, X_l$ has generating polynomial

$$p_l(z) = \frac{(\sum_i \partial_i)^{k-l} p(z)}{(k)_{k-l}}, \qquad (k)_r = k(k-1)\cdots(k-r+1).$$

This identity follows by expanding the derivative on each monomial: the remaining l counts have the multivariate hypergeometric law of a uniform l-ball subsample. Homogeneity gives $p_l(\mathbf{1}) = 1$. Nonnegative directional derivatives preserve the Lorentzian class [1]. Its mean fraction is q, and for $l \geq 2$ its normalized ordered-pair matrix is the same J as that of the original count. Equation (9) consequently gives the exact deficit scaling

$$D_l = \frac{l-1}{k-1} D.$$

Apply (25) to the degree-l marginal. Its final-label posterior is exactly the step-l posterior in the original ordering; conditional exchangeability makes the posterior depend on the prefix only through its histogram. The $l = 1$ contribution is zero. The entropy chain rule thus yields

$$\begin{aligned} \operatorname{KL}(\mathcal{L}(X_1, \ldots, X_k) \| q^{\otimes k}) &= \sum_{l=1}^{k} \mathbb{E} \operatorname{KL}(\mathcal{L}(X_l \mid X_1, \ldots, X_{l-1}) \| q) \\ &\leq \sum_{l=1}^{k} \frac{l-1}{k-1} \operatorname{tr} D = \frac{k}{2} \operatorname{tr} D. \end{aligned}$$

For a histogram a, the $k!/\prod_i a_i!$ words are equiprobable under both the canonical ordering and the product law, conditional on that histogram. Their likelihood ratio is exactly the count likelihood ratio against $\operatorname{Mult}(k, q)$. The ordered and count relative entropies are therefore equal, proving the first bound in (23). Pinsker's inequality proves the TV statement.

Finally, direct expansion of the reference likelihood gives the exact identity

$$\operatorname{KL}(\mathcal{L}(A) \| \operatorname{Mult}(k, r)) = \operatorname{KL}(\mathcal{L}(A) \| \operatorname{Mult}(k, q)) + k \operatorname{KL}(q \| r),$$

because $\mathbb{E} A_i = k q_i$. This proves (24). All arguments are on positive mean support. Deleted coordinates are identically zero, and the identity with positive r remains valid with $0 \log 0 = 0$. A one-category support, or degree one, gives zero divergence to the matching multinomial directly. □

Corollary 7.2 (An estimated-reference confidence bound)**.** *On the same probability-$1-\delta$ event used to prove Theorem 2.1,*

$$\begin{aligned} \operatorname{KL}(\mathcal{L}(A) \| \operatorname{Mult}(k, r)) &\leq \frac{k}{2}(m-1)U + \frac{\varepsilon^2}{128}, \\ \operatorname{TV}(\mathcal{L}(A), \operatorname{Mult}(k, r)) &\leq \min\left\{1, \sqrt{\frac{k}{4}(m-1)U + \frac{\varepsilon^2}{256}}\right\}. \end{aligned}$$

Proof. The proof event includes (19) and $\|D\| \le U$. Since $D \succeq 0$, $Du = 0$, and its positive support has size at most m, $\operatorname{tr} D \le (m-1)U$. Also $\mathrm{KL}(q\|r) \le h$ and $kh \le \varepsilon^2/128$. Apply Theorem 7.1 and Pinsker. This is a deterministic comparison for each realized reference r on the already budgeted joint event; no additional failure probability is spent. □

The entropy coefficient is not asserted to be optimal. The square-root TV bound need not dominate linear-in-deficit approximation bounds near zero. Its purpose here is to supply a distributional confidence consequence from the same observations; it does not identify a count law from its first two moments.

8 Category and accuracy lower bound

The following stable-subclass construction supplies the sample-order comparison for a guarantee uniform over total counts. This lower bound is included as a supporting comparison, rather than a separate contribution.

Proposition 8.1. *Let m be even and $0 < \rho < 1$. Even with $q_i = 1/m$ supplied, no test distinguishes $D = 0$ from all homogeneous stable degree-two laws with $\|D\| = \rho$, with both type-I and worst-case type-II error at most $1/3$, when $N \le m/(100\rho^2)$. Thus estimating the deficit norm to accuracy less than $\rho/3$ requires $\Omega(m/\rho^2)$ full counts.*

Proof. For a balanced sign vector $s \in \{-1,1\}^m$, take the stable product

$$p_s(z) = \left[\sum_i \frac{1+\sqrt{\rho}s_i}{m} z_i\right]\left[\sum_i \frac{1-\sqrt{\rho}s_i}{m} z_i\right].$$

It has $D_s = \rho v v^T$, $v = s/\sqrt{m}$, and a uniformly ordered pair has law $J_s(i,j) = (1-\rho s_i s_j)/m^2$. The null is $p_0(z) = (m^{-1}\sum_i z_i)^2$. A degree-two count and its uniformly ordered pair have identical testing information: conditional order is ancillary under every hypothesis.

Mix alternatives uniformly over balanced signs. For independent balanced s, t, write $T = s^T t/m$. The null inner product of their one-sample likelihoods is $1 + \rho^2 T^2$, so the N-sample mixture satisfies

$$\chi^2(\mathcal{M}_N \| \mathcal{P}_0^{\otimes N}) \le \mathbb{E} e^{N\rho^2 T^2} - 1.$$

Fix s and generate t by a uniform perfect matching, assigning opposite signs in each pair by independent fair choices. This gives a uniform balanced t. Conditional on the matching, $s^T t$ is a Rademacher sum with coefficients in $\{-2, 0, 2\}$, whose mgf is at most $e^{m\lambda^2}$. Hence $\mathbb{P}(|T| \ge z) \le 2e^{-mz^2/4}$. For $b = N\rho^2 < m/4$, tail integration gives

$$\mathbb{E} e^{bT^2} - 1 \le \frac{8N\rho^2}{m - 4N\rho^2} \le \frac{1}{12}$$

at the stated budget. Thus null–mixture TV is at most $\frac{1}{2}\sqrt{1/12} < 1/3$, whereas a test with both errors at most $1/3$ would give a probability gap at least $1/3$. This contradiction proves the testing claim. Disjoint norm-accuracy intervals give the estimation consequence. □

This lower bound concerns category and accuracy order, already at degree two. It does not show that every logarithm in Theorem 2.1 is necessary, or identify an optimal leading constant.

9 Computation and scope

The observable matrix (3) uses the full count once, not independent samples from its $k(k-1)$ ordered pairs. The estimator does not simulate the polarized measure or run its down-up chain. Ordinary dense spectral functional calculus costs $O(m^3)$ per observation naively, with $O(m^2)$ memory for accumulation. Large inverse weights may require numerical care; the theorem is an exact statistical statement, not a floating-point error analysis. A certified upper eigenvalue enclosure with excess ξ adds at most ξ to the upper slack for U.

The mean-smoothing and deterministic whitening arguments need only concavity of the degree root, covariance domination and integer counts. The category-linear operator variance additionally uses the down-up Poincaré inequality, obtained here from the full Lorentzian structure. Root concavity alone is not substituted for that hypothesis. Conversely, the proof does not require real-rooted slice factorization, negative association, or an independent categorical decomposition. The distributional consequence above uses a further posterior and thinning argument. Parity certification requires additional structural arguments; a covariance estimate alone does not determine higher moments.

Disclosure of AI assistance

Generative AI tools, including OpenAI Codex, assisted with mathematical exploration, literature searches, proof checking and manuscript preparation. The author takes responsibility for the mathematical claims and references.

Statements and Declarations

Funding. This research received no external grant funding.

Competing interests. The author owns Luminary Oddities.

Data and materials availability. The results are analytic; all arguments needed for the stated bounds are given in the text or in the explicitly cited classical results.

Author contributions. Scott Schubr is the sole author and is responsible for the research and final manuscript. AI assistance is disclosed above.

References

[1] P. Brändén and J. Huh, Lorentzian polynomials, *Annals of Mathematics* **192** (2020), 821–891. arXiv:1902.03719.

[2] N. Anari, K. Liu, S. Oveis Gharan and C. Vinzant, Log-concave polynomials II: high-dimensional walks and an FPRAS for counting bases of a matroid, *Annals of Mathematics* **199** (2024), 259–299. Theorem numbering here refers to arXiv:1811.01816v3 (2019).

[3] S. Minsker, Sub-Gaussian estimators of the mean of a random matrix with heavy-tailed entries, *Annals of Statistics* **46** (2018), 2871–2903. arXiv:1605.07129v5.

[4] C. Daskalakis, G. Kamath and C. Tzamos, On the structure, covering, and learning of Poisson multinomial distributions, *Proceedings of FOCS* (2015). The cited lemma numbering refers to arXiv:1504.08363v3.

[5] S.-L. Huang, A. Makur, G. W. Wornell and L. Zheng, Universal features for high-dimensional learning and inference, *Foundations and Trends in Communications and Information Theory* **21** (2024), 1–299. doi:10.1561/0100000107.

[6] A. Makur, G. W. Wornell and L. Zheng, On estimation of modal decompositions, *2020 IEEE International Symposium on Information Theory* (2020). Author manuscript.