Degree-uniform covariance-deficit estimation for Lorentzian count laws

Scott Schubr
Luminary Oddities

November 26, 2026 (edition 6)

Abstract

We estimate the normalized covariance deficit of a homogeneous Lorentzian count law from independent full count vectors, with unknown mean and no lower bound on the category probabilities. 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 with an exact one-ball increment and a diagonal constraint for rare categories. Independent mean smoothing and a classical robust matrix-mean estimator then give an operator-norm confidence bound from $O(m\varepsilon^{-2}\log(m/\delta))$ count observations, uniformly over total counts; a degree-two stable-subclass lower bound matches the category and accuracy dependence up to logarithms. The trace of the deficit also controls the distance to the matching multinomial law. The previous edition bounded the relative entropy by $k/2$ times the trace, while the author's covariance-spectrum certificates give a degree-free linear total-variation bound for real-stable laws; comparing the two poses the question of the correct degree dependence of the relative-entropy coefficient for Lorentzian laws. Applying the Poincaré inequality to successive uniform subsamples bounds the predictive label covariances and gives relative entropy at most h_k times the trace, where $h_k = k(H_k - 1)/(k-1)$, with H_k the kth harmonic number, is of order $\log k$, and total variation at most $2h_k - 1$ times the trace. A deterministic binary count shows that the logarithmic order is necessary for relative entropy. The estimator uses each count once and needs no simulation of the down-up chain.

Keywords. covariance estimation; Lorentzian polynomials; count distributions; rare categories; spectral gap; robust matrix mean; relative entropy; multinomial approximation.
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. The two results measure covariance error in different metrics.

Normalized-dependence estimation and empirical marginal normalization also have established antecedents. Makur, Wornell and Zheng [9, Section III] give finite-sample singular-mode bounds using known true margins and a minimum-marginal parameter. Huang, Makur, Wornell and Zheng [6, 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.

Normalization by the unknown mean 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 [3] 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 [10] 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, including laws without an independent categorical factorization. Real-stable laws form a subclass of the model, and the proofs use only the Lorentzian structure.

Distance to the matching multinomial law. Write $\operatorname{tr} D$ for the trace of the deficit. Edition 5 of this article [13] bounded the relative entropy of a Lorentzian count law from $\mathrm{Mult}(k, q)$ by $(k/2)\operatorname{tr} D$, a coefficient linear in the degree. For the real-stable subclass, the author's covariance-spectrum certificates [12, Theorem 1.3] give $\mathrm{TV}(\mathcal{L}(A), \mathrm{Mult}(k, q)) \le \min\{1, \operatorname{tr} D\}$, with a coefficient that is optimal and independent of the degree. Comparing these two results poses the following question.

> *Question.* How must the coefficient of $\operatorname{tr} D$ in relative-entropy bounds grow with the degree k for homogeneous Lorentzian count laws, and do these laws admit a total-variation bound linear in $\operatorname{tr} D$?

This edition replaces Section 7 of edition 5 by Theorem 7.1, which determines the order $\log k$ of the relative-entropy coefficient and gives a linear total-variation bound. The relative entropy is at most $h_k \operatorname{tr} D$ with $h_k = k(H_k - 1)/(k - 1) = \log k + O(1)$, where H_k is the kth harmonic number, and Proposition 7.4 shows that the order $\log k$ is necessary. Total variation is at most $\min\{1, (2h_k - 1)\operatorname{tr} D\}$. For example, every sequence of Lorentzian count laws with $\operatorname{tr} D \log k \to 0$ converges to its matching multinomial laws in

relative entropy, whereas the coefficient $k/2$ required $k \operatorname{tr} D \to 0$. The proof applies the Poincaré inequality of Lemma 4.1 to successive uniform subsamples, which bounds the covariance of each predictive label law by a remaining-population multiple of D (Proposition 7.3); the chain rule then gives the entropy bound, and a retained-prefix Taylor expansion with a smoothing estimate of Roos [11] gives the linear total-variation bound.

2 Model and estimator

We use the homogeneous Lorentzian class of Brändén and Huh [3]. 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 [3, 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

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

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

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

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. All constants are explicit. 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 [3] 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* $\|D\| = \operatorname{tr} D = 3/7$*. This illustrates a conditioned count law covered by Theorems 2.1 and 7.1 without an independent categorical factorization.*

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 $D \le P_q$, 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

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

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 [3, Proposition 3.1]; the resulting homogeneous multiaffine polynomial is strongly log-concave [3, 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

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

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*

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

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,

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

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

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

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

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

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 $\operatorname{diag} A$ throughout the one-ball increment. An uncorrected whitened covariance can have divergent rare-coordinate fourth moments even when (13) holds. The down-up chain enters only through this population inequality; the estimator processes the observations directly.

6 Unknown means and robust concentration

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

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

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

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

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

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

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*

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

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 $\|B\| \le 2$, $v^T b = 0$, $\|b\|^2 = \varrho$, 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 $P_r T D^{1/2} = TD^{1/2}$ and hence $BDB^T = TDT^T$. With all operators zero-extended as appropriate, let $F = (T-I)D^{1/2}$. By (7),

$$\begin{aligned}\|F\|^2 \le \|F\|_F^2 &= \sum_{i:q_i>0} (\sqrt{q_i/r_i} - 1)^2 D_{ii} \\ &\le \kappa \sum_i \frac{(q_i - r_i)^2}{r_i} \frac{q_i}{(\sqrt{q_i} + \sqrt{r_i})^2} \le \kappa\varrho.\end{aligned}$$

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

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

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

where $k\varrho \le k\varepsilon^2/(256\kappa) \le 1/128$. Also $\|C^2\| = \kappa^2\varrho^2 \le 1/65536$. 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 [10, 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 $D \succeq 0$, 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 Logarithmic covariance-deficit bounds for Lorentzian count laws

Write KL for relative entropy with natural logarithms and use probability total variation, $\mathrm{TV}(\mu, \nu) = \frac{1}{2}\|\mu - \nu\|_1$. On the positive mean support put

$$\Sigma_q = Q - qq^T, \qquad G = Q^{1/2} D Q^{1/2} = \Sigma_q - \frac{1}{k}\operatorname{Cov}(A),$$

so that $\operatorname{tr} D = \operatorname{tr}(Q^{-1}G)$. With the harmonic numbers $H_k = \sum_{j=1}^{k} 1/j$ define

$$h_k = \frac{k}{k-1}(H_k - 1) = \frac{1}{k-1}\sum_{r=0}^{k-1} \frac{r}{k-r}. \tag{23}$$

The two expressions agree because $\sum_{r=0}^{k-1} r/(k-r) = \sum_{s=1}^{k}(k/s - 1) = k(H_k - 1)$. Thus $h_2 = 1$, $h_k \le k/2$ for every $k \ge 2$, and $h_k = \log k + \gamma - 1 + o(1)$, where γ is Euler's constant.

Theorem 7.1 (Logarithmic deficit bounds)**.** *For every homogeneous Lorentzian count law of degree $k \ge 2$,*

$$\mathrm{KL}(\mathcal{L}(A) \| \operatorname{Mult}(k, q)) \le h_k \operatorname{tr} D, \qquad \mathrm{TV}(\mathcal{L}(A), \operatorname{Mult}(k, q)) \le \min\Big\{1,\ (2h_k - 1)\operatorname{tr} D,\ \sqrt{h_k \operatorname{tr} D/2}\Big\}. \tag{24}$$

For any strictly positive probability vector r,

$$\mathrm{KL}(\mathcal{L}(A) \| \operatorname{Mult}(k, r)) \le h_k \operatorname{tr} D + k\,\mathrm{KL}(q\|r). \tag{25}$$

The coefficients are independent of the number of categories and of the smallest positive q_i. Proposition 7.4 shows that the logarithmic order of h_k is necessary for relative entropy. For the real-stable subclass, Theorem 1.3 of [12] gives $\mathrm{TV}(\mathcal{L}(A), \operatorname{Mult}(k, q)) \le \min\{1, \operatorname{tr} D\}$; in the notation of that article $\operatorname{tr} D = 2W$. Its coefficient one is optimal uniformly in the degree, by the balanced deterministic binary count credited there to Roos [11, Remark 2.2(a)]. For Lorentzian count laws the linear coefficient in (24) is $2h_k - 1$, which is also one at $k = 2$.

Relation to entropy contraction. Anari, Jain, Koehler, Pham and Vuong [1] introduced entropic independence and proved relative-entropy contraction under downsampling for fractionally log-concave measures. For exchangeable words with Lorentzian count polynomials, Lee and Seo [8, Lemmas 3.5 and 3.7] prove a one-coordinate relative-entropy contraction and a strengthened delete-one entropy inequality, and Kortchemski [7, Lemma 3.1] deduces that the relative entropy of the first l of N exchangeable coordinates from the product of their marginal is at most $l(l-1)/(N(N-1))$ times that of all N coordinates. Ercan [5, Theorem 1.1] proves that retaining a uniform l of r particles contracts entropy by the factor l/r for every homogeneous Lorentzian law. These results compare a law or its subsamples with structured references through entropies already present. Theorem 7.1 instead bounds the distance of the full count law from its matching multinomial law by the covariance deficit at the original mean. Its proof uses the variance form of the subsampling contraction (Lemma 7.2) to control predictive covariances, and a smoothing estimate of Roos [11] for total variation.

7.1 Uniform subsamples and predictive covariances

Give the balls of A a uniform ordering $X_1, \dots, X_k$, and let $A_{[j]}$ be the histogram of $X_1, \dots, X_j$, so that $A_{[k]} = A$. Conditional on A, $A_{[j]}$ is a uniform j-ball subsample, and its generating polynomial is

$$p_j(z) = \frac{(\sum_i \partial_i)^{k-j} p(z)}{(k)_{k-j}}, \qquad (k)_s = k(k-1)\cdots(k-s+1). \tag{26}$$

Expanding the derivative on each monomial gives the multivariate hypergeometric subsampling law, and homogeneity gives $p_j(\mathbf{1}) = 1$. Nonnegative directional derivatives preserve the Lorentzian class [3, Corollary 2.11], so each $A_{[j]}$ with $j \geq 1$ is a homogeneous Lorentzian count of degree j. Its mean fraction is q, and for $j \geq 2$ its ordered-pair matrix is the same J as that of A. Equation (9) applied at degree j therefore gives

$$\operatorname{Cov}(A_{[j]}) = j\Sigma_q - \frac{j(j-1)}{k-1} G, \qquad 1 \leq j \leq k. \tag{27}$$

Lemma 7.2 (Subsample variance contraction)**.** *For every real function f on the support of A and $0 \leq r \leq k$,*

$$\operatorname{Var}\big(\mathbb{E}[f(A) \mid A_{[r]}]\big) \leq \frac{r}{k} \operatorname{Var}(f(A)). \tag{28}$$

For vector-valued f the same inequality holds in the Loewner order of covariance matrices.

Proof. Put $F_j = \mathbb{E}[f(A) \mid A_{[j]}]$. Given $A_{[j]}$, the label X_j is a uniform ball of $A_{[j]}$, independent of A. Hence $A \to A_{[j]} \to A_{[j-1]}$ is a Markov chain and $\mathbb{E}[F_j \mid A_{[j-1]}] = F_{j-1}$; the σ-fields of successive histograms need not be nested, and this conditional independence is the tower identity used. The count $A_{[j-1]}$ is a uniform one-ball deletion of the degree-j Lorentzian count $A_{[j]}$, so Lemma 4.1 at degree j and the law of total variance give

$$\operatorname{Var}(F_j) \leq j\, \mathbb{E} \operatorname{Var}(F_j \mid A_{[j-1]}) = j\big[\operatorname{Var}(F_j) - \operatorname{Var}(F_{j-1})\big]$$

for $j \geq 2$. Thus $\operatorname{Var}(F_{j-1}) \leq (j-1)\operatorname{Var}(F_j)/j$ for $j \geq 2$, and the same inequality holds at $j = 1$ because F_0 is constant. Iteration from $j = k$ down to $j = r+1$ proves (28). The vector statement follows by applying the scalar one to every linear functional. □

Linearizing the entropy contraction of Ercan [5, Theorem 1.1] at a constant density gives the same inequality; the derivation above uses only Lemma 4.1.

For $0 \leq r \leq k-1$ let $\pi_r = \pi_r(A_{[r]})$ be the conditional law of X_{r+1} given $A_{[r]}$. By exchangeability, the conditional law of the next label given the ordered prefix depends on the prefix only through its histogram, and $\mathbb{E}\pi_r = q$. For $r = 1$ and $A_{[1]} = e_i$, π_1 is the residual mean fraction $t^{(i)}$ of Lemma 3.2.

Proposition 7.3 (Predictive covariance)**.** *On the positive mean support, for* $0 \le r \le k-1$*,*

$$Q^{-1/2}\operatorname{Cov}(\pi_r)Q^{-1/2} \preceq \frac{r}{(k-1)(k-r)}D. \tag{29}$$

Consequently

$$\mathbb{E}\,\mathrm{KL}(\pi_r\|q) \le \mathbb{E}\chi^2(\pi_r\|q) = \operatorname{tr}\big(Q^{-1}\operatorname{Cov}(\pi_r)\big) \le \frac{r\operatorname{tr}D}{(k-1)(k-r)}. \tag{30}$$

Proof. Write $n = k - r$. Conditional exchangeability gives $\mathbb{E}[A \mid A_{[r]}] = A_{[r]} + n\pi_r$, and uniform subsampling gives $\mathbb{E}[A_{[r]} \mid A] = rA/k$, hence $\operatorname{Cov}(A_{[r]}, A) = r\operatorname{Cov}(A)/k$. Since $\operatorname{Cov}(A_{[r]}, \mathbb{E}[A \mid A_{[r]}]) = \operatorname{Cov}(A_{[r]}, A)$, expanding the covariance of $A_{[r]} + n\pi_r$ gives exactly

$$\operatorname{Cov}(\mathbb{E}[A \mid A_{[r]}]) = n^2\operatorname{Cov}(\pi_r) + \frac{2r}{k}\operatorname{Cov}(A) - \operatorname{Cov}(A_{[r]}).$$

By Lemma 7.2, the left side is at most $r\operatorname{Cov}(A)/k$ in the Loewner order. Therefore, by (27) and $\operatorname{Cov}(A) = k(\Sigma_q - G)$,

$$n^2\operatorname{Cov}(\pi_r) \preceq \operatorname{Cov}(A_{[r]}) - \frac{r}{k}\operatorname{Cov}(A) = \frac{r(k-r)}{k-1}G.$$

Divide by n^2 and conjugate by $Q^{-1/2}$ to obtain (29); for $r = 0$ both sides vanish. All inverse weights are the original positive means. For (30), the inequality $\log x \le x - 1$ gives $\mathrm{KL}(\pi\|q) \le \chi^2(\pi\|q)$ for every probability vector π, and $\mathbb{E}\chi^2(\pi_r\|q)$ is the trace of the left side of (29) because $\mathbb{E}\pi_r = q$. □

Proof of the relative-entropy bound in Theorem 7.1. The chain rule for relative entropy and (30) give

$$\mathrm{KL}(\mathcal{L}(X_1,\dots,X_k)\|q^{\otimes k}) = \sum_{r=0}^{k-1}\mathbb{E}\,\mathrm{KL}(\pi_r\|q) \le \frac{\operatorname{tr}D}{k-1}\sum_{r=0}^{k-1}\frac{r}{k-r} = h_k\operatorname{tr}D.$$

For a histogram a, the $k!/\prod_i a_i!$ words with histogram a are equiprobable under both the uniformly ordered law and $q^{\otimes k}$. Their likelihood ratio is therefore the count likelihood ratio against $\mathrm{Mult}(k, q)$, and the ordered and count relative entropies are equal. This proves the first bound in (24), and Pinsker's inequality gives the square-root term.

Direct expansion of the reference likelihoods gives the exact identity

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

because $\mathbb{E}A_i = kq_i$; this proves (25). All arguments are on the positive mean support; deleted coordinates are identically zero, and the identity with positive r holds with $0\log 0 = 0$. A single positive category gives zero divergence directly. □

Proposition 7.4 (Logarithmic order for relative entropy)**.** *For $k = 2m$ let $A = (m, m)$ almost surely, the count law of the real-stable Lorentzian monomial $x^m y^m$. Then* $\operatorname{tr}D = 1$ *and*

$$\mathrm{KL}(\mathcal{L}(A)\|\,\mathrm{Mult}(k,q)) = \log\frac{4^m}{\binom{2m}{m}} \ge \frac{H_m}{2} \ge \frac{1}{2}\log(m+1).$$

Hence, for even k, every coefficient c_k such that $\mathrm{KL}(\mathcal{L}(A)\|\,\mathrm{Mult}(k,q)) \le c_k\operatorname{tr}D$ *for all Lorentzian count laws of degree k satisfies $c_k \ge \frac{1}{2}\log(k/2+1)$; in particular $h_k \le (2+o(1))c_k$.*

Proof. Here $q = (1/2, 1/2)$ and $\operatorname{Cov}(A) = 0$, so $D = I - uu^T$ has trace one. The relative entropy of a point mass is $-\log\mathrm{Mult}(k,q)(\{(m,m)\})$, and $\binom{2m}{m}4^{-m} = \prod_{j=1}^{m}(1-1/(2j)) \le \mathrm{e}^{-H_m/2}$. Finally $H_m \ge \log(m+1)$ and $h_k = \log k + O(1)$. □

The same count is the optimality example for the real-stable total-variation coefficient in [12], where it is credited to Roos [11, Remark 2.2(a)].

7.2 Retained-prefix expansion and total variation

We view a homogeneous polynomial with real coefficients as a signed measure on count vectors, and multiplication as convolution. For a signed measure of total mass zero we use the total-variation norm $\|\cdot\|_{\mathrm{TV}}$, one half of its coefficient ℓ^1 norm, which is the probability total variation of a difference of two laws.

Lemma 7.5 (Quadratic smoothing)**.** *Let $\Gamma \geq 0$ satisfy $\Gamma\mathbf{1} = 0$, and let q be a positive probability vector. For every integer $l \geq 0$,*

$$\left\|(z^T\Gamma z)\langle q, z\rangle^l\right\|_{\mathrm{TV}} \leq \frac{\mathrm{tr}(Q^{-1}\Gamma)}{\sqrt{2(l+1)(l+2)}} \leq \frac{\mathrm{tr}(Q^{-1}\Gamma)}{l+1}. \tag{31}$$

Convolution with a probability law does not increase this norm.

Proof. This is the case $k = 2$ of Roos's inequality [11, Lemma 4.2, inequality (35)]. For completeness, here is a direct proof. First let $\Gamma = ww^T$ with $\sum_i w_i = 0$, and put $W = \sum_i w_i^2/q_i$ and $f(i) = w_i/q_i$. With $l+2$ independent labels X'_s of law q, the coefficient of z^a in $(w^Tz)^2\langle q, z\rangle^l$ is $\mathbb{E}[f(X'_1)f(X'_2);$ the histogram of X' is $a]$. By exchangeability it equals $\mathbb{E}[T;$ histogram $a]$ with

$$T = \frac{1}{(l+2)(l+1)} \sum_{s\neq s'} f(X'_s)f(X'_{s'}).$$

Since $\mathbb{E}f(X'_s) = 0$, only identical or reversed index pairs survive in $\mathbb{E}T^2$, so $\mathbb{E}T^2 = 2W^2/((l+2)(l+1))$. The coefficient ℓ^1 norm is at most $\mathbb{E}|T| \leq (\mathbb{E}T^2)^{1/2}$, which gives the first inequality after halving, and the second follows from $l+1 \leq 2(l+2)$. For general Γ, use a spectral decomposition into terms ww^T with $w \perp \mathbf{1}$ and the triangle inequality; the values of W add up to $\mathrm{tr}(Q^{-1}\Gamma)$. The coefficient ℓ^1 norm contracts under convolution with a probability law. □

Put $\Pi = I - \mathbf{1}q^T$, so $\Pi z = z - \langle q, z\rangle\mathbf{1}$, and $z_\sigma = \langle q, z\rangle\mathbf{1} + \sigma\Pi z$ for $0 \leq \sigma \leq 1$. Then $z_1 = z$, $z_0 = \langle q, z\rangle\mathbf{1}$ and $p(z_0) = \langle q, z\rangle^k$. The derivative of $\sigma \mapsto p(z_\sigma)$ at $\sigma = 0$ is $\langle q, z\rangle^{k-1}k\, q^T\Pi z = 0$. Taylor's formula with integral remainder therefore gives the polynomial identity

$$p(z) - \langle q, z\rangle^k = \int_0^1 (1-\sigma)\,(\Pi z)^T\nabla^2 p(z_\sigma)\,\Pi z\,\mathrm{d}\sigma. \tag{32}$$

For $0 \leq r \leq k-2$ and a histogram c of total r, let J_c be the conditional law of the ordered pair (X_{r+1}, X_{r+2}) given $A_{[r]} = c$. The ordered-pair expansion of the Hessian is

$$\nabla^2 p(z) = k(k-1)\,\mathbb{E}\left[z^{A_{[k-2]}}J_{A_{[k-2]}}\right],$$

since $\partial_i\partial_j p(z) = \mathbb{E}[A_i(A_j - \mathbf{1}_{i=j})z^{A-e_i-e_j}]$ and the last two labels are (i, j) with conditional probability $A_i(A_j-\mathbf{1}_{i=j})/(k(k-1))$ given A. Substitute z_σ: each of the $k-2$ prefix labels contributes $\sigma z_x+(1-\sigma)\langle q, z\rangle$, that is, it is retained with probability σ or replaced by a fresh label of law q. If r labels are retained, their histogram is distributed as the uniform prefix $A_{[r]}$ jointly with the last two labels, by exchangeability, and the fresh labels contribute $\langle q, z\rangle^{k-2-r}$. Summing over the discarded labels replaces $J_{A_{[k-2]}}$ by $J_{A_{[r]}}$, so, exactly,

$$\nabla^2 p(z_\sigma) = k(k-1)\sum_{r=0}^{k-2}\binom{k-2}{r}\sigma^r(1-\sigma)^{k-2-r}\langle q, z\rangle^{k-2-r}\,\mathbb{E}\left[z^{A_{[r]}}J_{A_{[r]}}\right]. \tag{33}$$

This grouping uses random subsampling of an exchangeable ordering, not independence of the original labels.

Conditional on $A_{[r]} = c$, the remaining $k - r \geq 2$ balls have a normalized partial derivative of p as generating polynomial, which is Lorentzian [3, Corollary 2.11]. Its mean fraction is $\pi_r(c)$ and its ordered-pair law is J_c. Lemma 3.2 at degree $k - r$ shows that $K_c = \pi_r(c)\pi_r(c)^T - J_c$ is positive semidefinite with $K_c\mathbf{1} = 0$. Since $J_c\mathbf{1} = \pi_r(c)$ and $\mathbf{1}^T J_c \mathbf{1} = 1$,

$$\Pi^T J_c \Pi = (\pi_r(c) - q)(\pi_r(c) - q)^T - K_c, \tag{34}$$

and both positive semidefinite terms on the right annihilate $\mathbf{1}$. Averaging $J_{A_{[r]}}$ returns the pair law $J = qq^T - G/(k-1)$ of (9), so $\mathbb{E}K_{A_{[r]}} = \mathrm{Cov}(\pi_r) + G/(k-1)$. Proposition 7.3 consequently yields

$$\mathbb{E}\,\mathrm{tr}\Big[Q^{-1}\big((\pi_r - q)(\pi_r - q)^T + K_{A_{[r]}}\big)\Big] \leq \Big[\frac{2r}{(k-1)(k-r)} + \frac{1}{k-1}\Big]\,\mathrm{tr}\,D. \tag{35}$$

Proof of the linear total-variation bound in Theorem 7.1. Insert (33) and (34) into (32). The difference $\mathcal{L}(A) - \mathrm{Mult}(k, q)$ is the signed measure $p(z) - \langle q, z\rangle^k$, which becomes an average over σ, r and $A_{[r]}$ of convolutions of the point mass $z^{A_{[r]}}$ with the quadratic signed measures $(z^T\Gamma z)\langle q, z\rangle^{k-2-r}$, for the two positive semidefinite matrices Γ on the right of (34). Apply Lemma 7.5 to each, with $l = k - 2 - r$, and then (35). The remaining scalar integral is the beta integral

$$\int_0^1 (1-\sigma)\binom{k-2}{r}\frac{\sigma^r(1-\sigma)^{k-2-r}}{k-1-r}\,\mathrm{d}\sigma = \frac{(k-2)!}{k!} = \frac{1}{k(k-1)},$$

in which the factor $l + 1 = k - 1 - r$ cancels against the factorial. The Hessian factor $k(k-1)$ cancels as well, and

$$\mathrm{TV}(\mathcal{L}(A), \mathrm{Mult}(k, q)) \leq \mathrm{tr}\,D \sum_{r=0}^{k-2}\Big[\frac{2r}{(k-1)(k-r)} + \frac{1}{k-1}\Big] = (2h_k - 1)\,\mathrm{tr}\,D,$$

because $\sum_{r=0}^{k-2} r/(k-r) = (k-1)h_k - (k-1)$. All quadratic signed measures used here have total mass zero, so their norms are probability total-variation norms. The trivial bound one completes (24). □

Corollary 7.6 (An estimated-reference confidence bound)**.** *On the same probability-*$1-\delta$ *event used to prove Theorem 2.1,*

$$\mathrm{KL}(\mathcal{L}(A) \| \mathrm{Mult}(k, r)) \leq h_k(m-1)U + \frac{\varepsilon^2}{128},$$
$$\mathrm{TV}(\mathcal{L}(A), \mathrm{Mult}(k, r)) \leq \min\Big\{1,\ (2h_k - 1)(m-1)U + \frac{\varepsilon}{16},\ \sqrt{\frac{h_k}{2}(m-1)U + \frac{\varepsilon^2}{256}}\Big\}.$$

Proof. The proof event includes (19) and $\|D\| \leq U$. Since $D \geq 0$, $Du = 0$, and its positive support has size at most m, $\mathrm{tr}\,D \leq (m-1)U$. Also $\mathrm{KL}(q\|r) \leq \varrho$ and $k\varrho \leq \varepsilon^2/128$. Theorem 7.1 and Pinsker's inequality give the relative-entropy bound and the square-root term. For the linear term, the triangle inequality and Pinsker's inequality give $\mathrm{TV}(\mathrm{Mult}(k, q), \mathrm{Mult}(k, r)) \leq \sqrt{k\,\mathrm{KL}(q\|r)/2} \leq \varepsilon/16$. This is a deterministic comparison for each realized reference r on the already budgeted joint event; no additional failure probability is spent. □

8 Category and accuracy lower bound

The following stable-subclass construction supplies the sample-order comparison for a guarantee uniform over total counts.

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 the category and accuracy order, already at degree two.

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 works directly with the observations; the polarized measure and its down-up chain enter only the proofs. Ordinary dense spectral functional calculus costs $O(m^3)$ per observation, with $O(m^2)$ memory for accumulation. Theorem 2.1 is an exact statistical statement about the estimator computed in exact arithmetic. 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. The proofs use neither real-rooted slices, nor negative association, nor an independent categorical decomposition. The distributional bounds of Section 7 use the same Poincaré inequality, applied to successive uniform subsamples.

Statements and Declarations

Funding. This research was self-funded.

Competing interests. The author owns Luminary Oddities and has no other relevant competing interests.

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.

AI assistance. OpenAI Codex assisted with derivations, proof development, source checks and independent internal review of each part. Anthropic Claude edited, consolidated and cross-referenced the parts and checked computations. Scott Schubr directed the research and is responsible for the content.

References

[1] N. Anari, V. Jain, F. Koehler, H. T. Pham and T.-D. Vuong, Entropic independence I: modified log-Sobolev inequalities for fractionally log-concave distributions and high-temperature Ising models (2021). arXiv:2106.04105, version 2.

[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**(1) (2024), 259–299. doi:10.4007/annals.2024.199.1.4. Theorem numbering refers to arXiv:1811.01816v3 (2019).

[3] P. Brändén and J. Huh, Lorentzian polynomials, *Annals of Mathematics* **192**(3) (2020), 821–891. doi:10.4007/annals.2020.192.3.4. Numbering refers to arXiv:1902.03719v8.

[4] C. Daskalakis, G. Kamath and C. Tzamos, On the structure, covering, and learning of Poisson multinomial distributions, *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science (FOCS)* (2015). doi:10.1109/FOCS.2015.77. The cited lemma numbering refers to arXiv:1504.08363v3.

[5] E. Ercan, Lorentzian entropy, subgaussian matrix concentration, and sharp functional inequalities for all matroids (2026). arXiv:2611.00626, version 1.

[6] 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**(1–2) (2024), 1–299. doi:10.1561/0100000107.

[7] I. Kortchemski, A total variation threshold for the first coordinates of uniform parking functions via Lorentzian polynomials (2026). arXiv:2610.40082, version 1.

[8] J. Lee and J. Seo, Antiferromagnetic models are clique-minimizing (2026). arXiv:2608.17920, version 1.

[9] A. Makur, G. W. Wornell and L. Zheng, On estimation of modal decompositions, *2020 IEEE International Symposium on Information Theory* (2020). doi:10.1109/ISIT44484.2020.9174093; author manuscript.

[10] S. Minsker, Sub-Gaussian estimators of the mean of a random matrix with heavy-tailed entries, *Annals of Statistics* **46**(6A) (2018), 2871–2903. doi:10.1214/17-AOS1642; arXiv:1605.07129v5.

[11] B. Roos, Closeness of convolutions of probability measures, *Bernoulli* **16**(1) (2010), 23–50. doi:10.3150/08-BEJ171. Numbering refers to arXiv:0811.0622v1.

[12] S. Schubr, Covariance-spectrum certificates for homogeneous stable count laws, edition 9. Luminary Research Archive, November 2026. `https://doi.org/10.5281/zenodo.24472412`.

[13] S. Schubr, Degree-uniform covariance-deficit estimation for Lorentzian count laws, edition 5. Luminary Research Archive, November 2026. `https://doi.org/10.5281/zenodo.24472488`.