Covariance confidence bounds and parity certificates for stable count laws

Scott Schubr
Luminary Oddities, La Grange Highlands, Illinois, USA
scottschubr@luminaryoddities.com
ORCID: 0009-0009-7357-4439

October 23, 2026
Archive edition 2

Abstract

We study the normalized covariance deficit of a homogeneous real-stable count law on m categories with total count k. This deficit is a rescaling of an established canonical dependence matrix for two balls drawn without replacement. Its positive semidefiniteness gives a diagonal domination property that controls rare categories and normalization error. We use this structure to construct an upper confidence bound for its operator norm without a minimum category-probability assumption and without knowing the mean. Independent batches of full count observations suffice: the mean stage uses $O(m(k-1)\varepsilon^{-2}\log(m/\delta)/k)$ observations and the second stage uses $O(m(k-1)^2\varepsilon^{-2}\log(m/\delta))$. A stable degree-two subclass gives a matching linear category and inverse-square accuracy lower bound up to logarithms. Real-rooted one-dimensional projections then turn a covariance bound and a reference Fourier margin into simultaneous positive-parity certificates. We give a sharp positivity boundary on a binary subclass and stable laws with identical means and covariance but opposite parity. The constants in the confidence bound are conservative, and the observation model is classical labeled counts.

Keywords. covariance estimation; count distributions; negative dependence; parity; real-stable polynomials.

1 Introduction and relation to previous work

Let $A = (A_1, \dots, A_m)$ be a nonnegative integer count vector with $\sum_i A_i = k$. A homogeneous real-stable generating polynomial imposes substantial structure on its covariance and one-dimensional projections. The question here is whether this structure permits a uniform spectral confidence bound when categories may have arbitrarily small unknown probabilities, and whether the resulting bound certifies parity after a known binary labeling of categories.

The normalized object is familiar. If J is the joint law of two ordered distinct balls sampled from A, with common marginal $q = \mathbb{E}A/k$, its centered canonical dependence matrix is

$$Q^{-1/2}(J - qq^T)Q^{-1/2} = -\frac{D}{k-1}, \qquad Q = \operatorname{diag}(q).$$

Its operator norm is the Hirschfeld–Gebelein–Rényi maximal correlation of the two ball labels. Spectral dependence estimation and empirical marginal normalization are established topics. Huang, Makur, Wornell and Zheng [1], Section 6.2, particularly Subsections 6.2.1–6.2.5 and Proposition 6.7, treat singular-value calibration and empirical normalization; their displayed concentration bound involves a minimum marginal. Xu,

Wang and Huang [2], Section V.B, explicitly study estimated marginal normalization. We use the established dependence measure and unknown-margin estimation framework.

Our additional hypothesis makes the centered dependence matrix negative semidefinite. Equivalently, $D \succeq 0$. The underlying negative-spectrum Hessian structure is classical: Anari, Liu, Oveis Gharan and Vinzant [3], Lemma 2.7 and Proposition 4.1, relate one-positive-eigenvalue nonnegative Hessians to normalized kernels with nonpositive nontrivial spectrum. We use this structural identification to obtain statistical rare-category control. Combined with the nonnegativity of J, this implies $D_{ii} \le (k-1)q_i$. It pays for omitted categories by their total mass and controls changes of normalization through an aggregate chi-square error, rather than uniform relative accuracy in every rare category. The main result estimates the norm of the original D. Regularized normalized-graph concentration is important related methodology [4]; its independent-edge experiment differs from the independent count blocks considered here. Covariance estimation for Poisson multinomial distributions also has an extensive literature, including [5], Lemma 3.4 and Appendix A, which control covariance error relative to $\mathrm{Cov}(A) + I$. Our count model contains these product categorical laws but need not admit a joint independent categorical factorization.

The stable-polynomial background is classical [6]. The covariance deficit also appears in moment approximation for homogeneous stable count laws [7]; the present note develops its statistical calibration and parity application rather than the higher-order approximation estimates of that work. The concentration tool is matrix Bernstein [8]. The parity argument uses classical real-rooted factorization and the Laguerre–Samuelson variance bound [9]; these ingredients are proved in the normalizations needed here. Our contribution is their combination with rare-category control, an explicit unknown-mean observation contract, and sharp limitations.

2 Stable counts and the pair representation

A real polynomial is *real stable* if it has no zero when every variable has positive imaginary part. Let

$$p(z) = \mathbb{E} \prod_{i=1}^{m} z_i^{A_i}$$

be real stable, homogeneous of degree $k \ge 2$, with nonnegative coefficients and $p(\mathbf{1}) = 1$. Write

$$q_i = \frac{\mathbb{E}A_i}{k}, \quad Q = \mathrm{diag}(q), \quad u_i = \sqrt{q_i}, \quad \kappa = k-1, \quad D = I - uu^T - \frac{1}{k} Q^{-1/2} \mathrm{Cov}(A) Q^{-1/2}. \tag{1}$$

We first work on the positive support of q. If the named alphabet includes zero-mean categories, their counts are zero almost surely; define D on positive support and embed it with zero rows and columns. All statements below use this convention. The empirical smoothing estimator can retain such categories without knowing the support.

Lemma 2.1 (Covariance domination)**.** *For the stable count law above,*

$$0 \preceq D \preceq I - uu^T, \qquad Du = 0.$$

Moreover, the count $A_S = \sum_{i \in S} A_i$ of any category subset has a Poisson-binomial law with k Bernoulli parameters, some of which may be zero or one.

Proof. For real z, the degree-k polynomial $s \mapsto p(z + s\mathbf{1})$ has only real roots: stability excludes roots in the upper half-plane, and real coefficients exclude roots in the lower half-plane. By homogeneity its reciprocal gives

$$p(\mathbf{1} + tz) = \prod_{j=1}^{k} (1 + t\lambda_j), \qquad \lambda_j \in \mathbb{R},$$

where zero λ_j pad any degree drop. Differentiating at zero gives

$$\sum_j \lambda_j = kq^T z, \qquad z^T \operatorname{Cov}(A) z = k \sum_i q_i z_i^2 - \sum_j \lambda_j^2.$$

Cauchy–Schwarz therefore implies $\operatorname{Cov}(A) \preceq k(Q - qq^T)$, proving $D \succeq 0$. Covariance positivity proves the other inequality. The fixed total count gives $\operatorname{Cov}(A)\mathbf{1} = 0$, hence $Du = 0$.

Specialize variables in S to z and the remaining variables to one. Boundary specialization preserves stability by a limiting argument, and the specialization is nonzero since its value at one is one. Thus its univariate generating polynomial has only real roots. Nonnegative coefficients exclude positive roots. Normalization then writes it as $\prod_{j=1}^{k}(1 - p_j + p_j z)$ with $p_j \in [0, 1]$, padding by $p_j = 0$ if necessary. This is the required Poisson-binomial representation. This elementary factorization is also standard in the Poisson-binomial literature [10]. □

One may sample an ordered pair (I, J) of distinct balls from a count vector. Its unconditional joint probability matrix, denoted by the same letter J, is

$$J_{ij} = \frac{\mathbb{E}[A_i(A_j - \mathbf{1}_{i=j})]}{k\kappa}. \tag{2}$$

It is symmetric and has row and column margins q.

Lemma 2.2 (Pair structure)**.** *Let $K = D/\kappa$. Then*

$$K = uu^T - Q^{-1/2} J Q^{-1/2} \succeq 0, \qquad K_{ii} \le q_i, \tag{3}$$

$$J_{ii} \le q_i^2, \qquad 0 \le J_{ij} \le 2q_i q_j. \tag{4}$$

In particular, $D_{ii} \le \kappa q_i$, $\operatorname{tr} D \le \kappa$, and

$$\|D\|_F^2 = \kappa^2 \chi^2(J \| q \otimes q). \tag{5}$$

Proof. The identity follows by substituting $\mathbb{E}AA^T = kQ + k\kappa J$ into (1). Since $J_{ii} \ge 0, 0 \le K_{ii} \le q_i$. Positive semidefinite Cauchy–Schwarz gives $|K_{ij}| \le \sqrt{q_i q_j}$, which yields (4). The trace bound and (5) follow by summing diagonal entries and squared entries, respectively. □

These conclusions require the covariance inequalities and the pair representation, rather than full real stability. Stability is used separately for Poisson-binomial marginals and parity transfer. A finite sample does not verify the structural promise.

3 Unknown-mean calibration

The two data batches below consist of independent observations from the same count law, with independence between batches. No ordered latent decomposition of a count vector is observed or assumed.

Theorem 3.1 (A uniform original-target confidence bound)**.** *Fix $0 < \varepsilon \le 1$ and $0 < \delta < 1$, and put $L = \log(4m/\delta)$. Let*

$$N_0 \ge \frac{10752 m\kappa L}{k\varepsilon^2}, \qquad N_1 \ge \frac{1216 m\kappa^2 L}{\varepsilon^2}. \tag{6}$$

From the first batch define

$$\widehat{q}_i = \frac{1}{kN_0} \sum_{a=1}^{N_0} A_i^{(a)}, \quad t = \frac{L}{kN_0}, \quad b = 2t, \quad c = 1 + mb, \quad r_i = \frac{\widehat{q}_i + b}{c}.$$

Write $R = \mathrm{diag}(r)$, $v = \sqrt{r}$, $P = I - vv^T$, and set

$$\tau = \frac{\varepsilon}{64\kappa}, \qquad H = \{i : r_i \geq \tau/m\}, \quad H^c = \{1, \ldots, m\} \setminus H, \quad w_r = \sum_{i \notin H} r_i.$$

From the second batch form the full-count pair estimator

$$\widehat{J}_{ij} = \frac{1}{N_1} \sum_{a=1}^{N_1} \frac{A_i^{(a)}(A_j^{(a)} - \mathbf{1}_{i=j})}{k\kappa}, \qquad \widehat{C}_H = -\kappa\big(PR^{-1/2}\widehat{J}R^{-1/2}P\big)_H. \tag{7}$$

A subscript H denotes principal compression. Define

$$U = \max\{0, \lambda_{\max}(\widehat{C}_H)\} + \frac{\varepsilon}{2} + 16\kappa w_r. \tag{8}$$

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

$$\|D\| \leq U \leq \|D\| + \frac{5}{4}\varepsilon. \tag{9}$$

The guarantee is independent of the smallest positive q_i. Numerical eigenvalue enclosures of width ξ add at most ξ to the upper slack.

The constants are explicit and conservative. The theorem gives a sample-order guarantee, not affordable absolute budgets or an efficient dense rational eigensolver for large m.

3.1 Smoothed mean control

Lemma 3.2. *For the first batch in Theorem 3.1, with probability at least $1 - \delta/2$,*

$$\frac{q_i}{r_i} \leq 4 \quad \textit{for every } i, \qquad h := \chi^2(q\|r) \leq 42mt \leq \frac{\varepsilon^2}{256\kappa}. \tag{10}$$

Proof. By Lemma 2.1, each marginal count is Poisson-binomial. The sum of its N_0 independent copies is therefore a sum of kN_0 independent Bernoulli variables, not necessarily identically distributed. Scalar Bernstein, with variance bounded by kN_0q_i, and a union bound imply

$$|\widehat{q}_i - q_i| \leq \sqrt{2q_it} + \frac{2}{3}t \quad \text{for all } i \tag{11}$$

except on an event of probability at most $2me^{-L} = \delta/2$. The argument needs only marginal representations; the counts for different categories need not be independent.

Since $\sqrt{2q_it} \leq q_i/2 + t$, (11) yields $q_i \leq 2\widehat{q}_i + 10t/3 \leq 2(\widehat{q}_i + 2t)$. Put $h_i = \widehat{q}_i + 2t$, so $q_i/h_i \leq 2$ and $h_i \geq 2t$. Also $mt \leq 1/2$ under (6); hence $c \leq 2$ and $q_i/r_i \leq 4$. Squaring (11) gives

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

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

$$\chi^2(q\|r) = c\sum_i \frac{q_i^2}{h_i} - 1 = cE_0 - (c-1)^2 \leq 42mt.$$

The first sample-size bound gives the final inequality in (10). These arguments include $q_i = 0$ without division by q_i. □

3.2 Centered weighted perturbation and discarded coordinates

Lemma 3.3. *Suppose r is a positive probability vector, $h = \chi^2(q\|r)$, and $q_i/r_i \le 4$. Define*

$$T = \operatorname{diag}\sqrt{q/r}, \quad e = (q-r)/\sqrt{r}, \quad B = TDT, \quad C = -\kappa P R^{-1/2} J R^{-1/2} P.$$

Then

$$C = B - \kappa e e^T, \tag{12}$$

$$\|C - D\| \le 2\sqrt{\kappa h} + 2\kappa h =: E_h, \tag{13}$$

$$\|D\| \le \lambda_{\max}(C_H) + E_h + 16\kappa w_r. \tag{14}$$

For zero-mean coordinates take $T_{ii} = 0$ and the zero embedding of D.

Proof. Let $a = R^{-1/2}q = v + e$. Since $v^T e = 0$, $v^T T = u^T$, and $u^T D = 0$, we have $PB = BP = B$. Also $R^{-1/2} J R^{-1/2} = aa^T - B/\kappa$. Projection gives (12).

The diagonal bound $D_{ii} \le \kappa q_i$ implies

$$\|(T-I)D^{1/2}\|_F^2 \le \kappa \sum_i q_i\big(\sqrt{q_i/r_i} - 1\big)^2 \le \kappa \sum_i \frac{(q_i - r_i)^2}{r_i} = \kappa h.$$

The scalar inequality in the last step follows by canceling $(\sqrt{q_i} - \sqrt{r_i})^2$; the remaining ratio is $q_i/(\sqrt{q_i} + \sqrt{r_i})^2 \le 1$. Since $\|D\| \le 1$, expansion of $TDT - D$ yields

$$\|B - D\| \le 2\sqrt{\kappa h} + \kappa h.$$

Equation (12) and $\|e\|^2 = h$ prove (13).

For any positive semidefinite block matrix B,

$$\|B\| \le \|B_H\| + \operatorname{tr} B_{H^c}. \tag{15}$$

Indeed, block Cauchy–Schwarz bounds its quadratic form at $x \oplus y$ by $(\sqrt{\|B_H\|}\|x\| + \sqrt{\|B_{H^c}\|}\|y\|)^2$, which is at most $(\|B_H\| + \|B_{H^c}\|)(\|x\|^2 + \|y\|^2)$. Here $B_{ii} \le \kappa q_i^2/r_i \le 16\kappa r_i$, so its omitted trace is at most $16\kappa w_r$. Moreover $B_H = C_H + \kappa e_H e_H^T$, whence $\|B_H\| \le \lambda_{\max}(C_H) + \kappa h$ by Weyl, even though C_H need not be positive semidefinite. Combining these bounds with the bound on $\|B - D\|$ gives (14). □

3.3 Conditional matrix concentration

Proof of Theorem 3.1. Condition on the first-stage event of Lemma 3.2, so r, H are fixed independently of the second batch. For a random ordered pair (i, j) of distinct balls, put

$$Y = \frac{e_i e_j^T + e_j e_i^T}{2\sqrt{r_i r_j}}, \qquad Z = (PYP)_H.$$

Here the e_i in the definition of Y are coordinate basis vectors, distinct from the vector e in Lemma 3.3. The formula includes $i = j$ as $e_i e_i^T/r_i$. The full matrix $\mathbb{E}Y^2$ is diagonal, with ith entry

$$\frac{J_{ii}}{r_i^2} + \sum_{j \ne i} \frac{J_{ij}}{2r_i r_j} \le 16m,$$

by (4) and $q_i/r_i \le 4$. Since P is an orthogonal projection, $(PYP)^2 \le PY^2P$. Principal compression has the same inequality, so $\mathbb{E}Z^2 \le 16mI$. The mean of the unprojected Y has norm at most four: it equals

$T(uu^T - D/\kappa)T$, whose middle factor has norm at most one and $\|T\| \le 2$. Consequently $\|\mathbb{E}Z\| \le 4$ and the centered variance is at most $16m$.

The range of Z does not involve inverse probabilities of omitted categories. Let $b_H = (Yv)_H$. Since $v^T Y v = 1$,

$$Z = Y_H - v_H b_H^T - b_H v_H^T + v_H v_H^T.$$

We have $\|Y_H\| \le m/\tau$ and $\|b_H\| \le \sqrt{m/\tau}$; only retained endpoints contribute to either. Therefore

$$\|Z\| \le m/\tau + 2\sqrt{m/\tau} + 1 \le 4m/\tau, \qquad \|Z - \mathbb{E}Z\| \le 4m/\tau + 4.$$

For a whole count observation, replace Z by its conditional expectation given that count. This is exactly the summand in (7) after whitening, projection and compression. Its range is no greater, and operator Jensen gives no larger second moment. Independence is between count observations, not between the $k\kappa$ ordered pairs within one count.

The two-sided bounded matrix Bernstein inequality [8, Theorem 6.1], at $z = \varepsilon/(4\kappa) = 16\tau$, now gives

$$\begin{aligned}\mathbb{P}\left(\|\widehat{C}_H - C_H\| > \varepsilon/4 \mid \text{first batch}\right) &\le 2m \exp\left\{-\frac{N_1 z^2}{2[16m + (4m/\tau + 4)z/3]}\right\} \\ &\le 2m \exp\left\{-\frac{N_1 \varepsilon^2}{1216 m \kappa^2}\right\} \le \delta/2.\end{aligned}$$

The second inequality uses $\tau \le 1/64$, which makes the denominator before z^2 substitution at most $76m$.

On both good events,

$$E_h \le \frac{\varepsilon}{8} + \frac{\varepsilon^2}{128} \le \frac{17\varepsilon}{128} < \frac{\varepsilon}{4}, \qquad w_r \le \tau.$$

The set H is nonempty because $w_r \le \tau < 1$. Lemma 3.3 and Weyl give $\|D\| \le U$. Conversely,

$$\max\{0, \lambda_{\max}(\widehat{C}_H)\} \le \|D\| + E_h + \varepsilon/4 \le \|D\| + \varepsilon/2.$$

Adding $\varepsilon/2 + 16\kappa w_r \le 3\varepsilon/4$ proves (9). The union of the two failure events has probability at most δ. □

4 From covariance to all parity characters

Assign a known label $\ell_i \in \mathbb{F}_2^n$ to each category and put $Z = \sum_i A_i \ell_i$. All label sums below are in $\mathbb{F}_2^n$.

Theorem 4.1 (A stable-slice parity certificate)**.** *Suppose $0 < a < 1$ is a lower bound for all nonzero one-draw characters:*

$$g_v = \sum_i q_i (-1)^{v \cdot \ell_i} \ge a \qquad (v \ne 0).$$

If $\rho = \|D\| < a^2/[\kappa(1 - a^2)]$, then simultaneously

$$\mathbb{E}(-1)^{v \cdot Z} \ge \left[a - \sqrt{\kappa\rho(1 - a^2)}\right]^k > 0 \qquad (v \ne 0). \tag{16}$$

For $a = 1$ every character is identically one.

Proof. For $S = \{i : v \cdot \ell_i = 1\}$, put $q_S = \sum_{i \in S} q_i$ and $g = 1 - 2q_S \ge a$. By Lemma 2.1, A_S is Poisson-binomial with parameters $p_1, \dots, p_k$. Let $\delta_j = p_j - q_S$ and

$$s = \sum_j \delta_j^2 = k q_S (1 - q_S) - \mathrm{Var}(A_S).$$

Writing $w_i = \sqrt{q_i}\mathbf{1}_{i\in S}$ gives $s = kw^T Dw$. Since $Du = 0$ and $\|w - (w^T u)u\|^2 = q_S(1 - q_S)$,

$$0 \le s \le k\rho q_S(1 - q_S).$$

The zero-sum condition $\sum_j \delta_j = 0$ implies $\delta_j^2 \le (k-1)s/k$ by Cauchy–Schwarz applied to the other $k-1$ terms. This is the classical Laguerre–Samuelson bound in this normalization. Hence every factor satisfies

$$1 - 2p_j \ge g - \sqrt{\kappa\rho(1-g^2)} \ge a - \sqrt{\kappa\rho(1-a^2)} > 0.$$

The second inequality follows by monotonicity in $g \in [a, 1]$. Multiplying the k Bernoulli parity factors proves the result. The factorization may depend on S; no joint independent categorical decomposition is assumed. □

Corollary 4.2 (An unknown-mean, all-character confidence certificate)**.** *Under Theorem 3.1, let r_0 be total r-mass of zero-labelled categories and set $a = 2r_0 - 1 - \varepsilon/16$. If $0 < a < 1$ and*

$$U < \frac{a^2}{\kappa(1-a^2)},$$

then, except on an event of probability at most δ, every nonzero character has expectation at least $[a - \sqrt{\kappa U(1-a^2)}]^k > 0$.

Proof. On the first-stage good event, $\mathrm{TV}(q, r) \le \sqrt{h}/2 \le \varepsilon/32$, since $\kappa \ge 1$. Thus true zero-label mass is at least $r_0 - \varepsilon/32$, which implies $g_v \ge 2r_0 - 1 - \varepsilon/16 = a$ for every v. Apply Theorems 3.1 and 4.1 on their simultaneous good event. Data-dependent success of the displayed test introduces no additional union bound over characters. □

This is one efficient sufficient reference condition, not a necessary condition for positive Fourier coefficients. It avoids enumerating 2^n characters. Also, whenever $a > 1 - 1/k$, the simpler mean-only bound

$$\mathbb{E}(-1)^{A_S} \ge 1 - 2\mathbb{E}A_S \ge 1 - k(1-a) > 0$$

already proves positivity. Covariance estimation is unnecessary in that regime.

Proposition 4.3 (Sharp positivity boundary on a binary subclass)**.** *For a single binary label and $q \in [1/(2k), 1/2)$, the threshold in Theorem 4.1 cannot be increased uniformly at reference margin $a = 1 - 2q$.*

Proof. Take the stable product of binary categorical factors with

$$p_1 = \frac{1}{2}, \qquad p_2 = \cdots = p_k = \frac{kq - 1/2}{k-1}.$$

Its parity is zero. With $g = 1 - 2q$, its centered deviations have squared sum $kg^2/(4\kappa)$; the sole nonzero eigenvalue of D is $g^2/[\kappa(1-g^2)]$. For $q > 1/(2k)$, increasing p_1 slightly and reducing the others equally gives negative parity while retaining the mean. If $q < 1/(2k)$, the mean alone forces all $p_j < 1/2$ and positive parity. Sharpness of the positivity boundary does not imply sharpness of the numerical lower floor in (16). □

Proposition 4.4 (A sharper cubic interval)**.** *For $k = 3$, a grouped count with $g = 1 - 2q \ge 0$ and $s = 3q(1-q) - \mathrm{Var}(A_S)$ satisfies*

$$g^3 - 2gs - \frac{8s^{3/2}}{3\sqrt{6}} \le \mathbb{E}(-1)^{A_S} \le g^3 - 2gs + \frac{8s^{3/2}}{3\sqrt{6}}.$$

The lower endpoint equals $(g - 4h)(g + 2h)^2$, $h = \sqrt{s/6}$, and is attained if $(q + 2h, q - h, q - h) \in [0, 1]^3$; the upper endpoint has the analogous feasibility condition.

Proof. For three zero-sum deviations of squared sum s, $|\delta_1\delta_2\delta_3| \le s^{3/2}/(3\sqrt{6})$. For $s > 0$, a compact constrained extremum has two equal deviations: Lagrange multipliers make all three solve one common quadratic. The extrema are $(2h, -h, -h)$ and its negative. The case $s = 0$ is immediate. Expand $\prod_j (g - 2\delta_j) = g^3 - 2gs - 8\delta_1\delta_2\delta_3$. □

5 Limitations of moment and dependence information

Proposition 5.1 (Identical stable moments, opposite parity)**.** *There are strictly positive-coefficient homogeneous stable degree-three binary count laws with identical full mean and covariance and opposite parity signs, both having* $\|D\| = 3/100$.

Proof. Use the products of binary categorical factors with success probabilities

$$(7/25, 23/50, 23/50), \qquad (17/50, 17/50, 13/25).$$

In both cases $\sum p_i = 6/5$, $\sum p_i^2 = 627/1250$, and the success-count variance is $873/1250$. Thus the vectors $(3 - A_S, A_S)$ have the same mean and covariance. Their parities are respectively $44/15625$ and $-64/15625$. At mean $q = 2/5$, the normalized deficit is $3/100$. Each factor is a positive stable linear form, so both laws meet the hypothesis. □

A vanishing-deficit version takes $q = 1/2 - 3r/2$, $0 < r < 1/7$, and parameters $(q-2r, q+r, q+r)$ versus $(q-r, q-r, q+2r)$. They share the first two moments and have deficit $2r^2/[q(1-q)] \to 0$, but parities $7r^3$ and $-25r^3$. Thus no absolute positive deficit threshold, independent of the reference margin, suffices even within the stable class. These obstructions require a sufficient certificate to abstain on some moment pairs; they do not contradict Theorem 4.1.

The structural promise is also indispensable for a uniform rare-category dependence estimate. For arbitrary symmetric pair laws with $q = (1 - s, s)$, compare $J_0 = qq^T$ with $J_1 = \mathrm{diag}(q)$. Their maximal correlations are zero and one, while $\mathrm{TV}(J_0, J_1) = 2s(1 - s)$. The total variation distance of N-sample laws is at most $2Ns(1 - s)$ and tends to zero for fixed N as $s \downarrow 0$. Here $J_1 - qq^T$ is positive semidefinite, violating the sign promise of this note. General dependence estimators cannot simply drop that promise and retain a uniform minimum-mass-free guarantee.

6 A stable-subclass sampling lower bound

Theorem 6.1. *Let m be even and $0 < \rho < 1$. Even with the exact mean $q_i = 1/m$ supplied, distinguishing a stable count law with $D = 0$ from all stable laws with $\|D\| = \rho$ cannot have both type-I and worst-case type-II error at most $1/3$ if*

$$N \le \frac{m}{100\rho^2}.$$

Consequently estimating $\|D\|$ to additive accuracy less than $\rho/3$ with error probability at most $1/3$ requires $\Omega(m/\rho^2)$ observations in a degree-two stable subclass.

Proof. For every balanced sign vector $s \in \{-1, 1\}^m$, use the stable degree-two product generating function

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

Both factors are positive categorical laws. Their common count mean is $2q$, and direct substitution gives $D_s = \rho vv^T$ with $v = s/\sqrt{m}$. A randomly ordered pair has law

$$J_s(i,j) = \frac{1 - \rho s_i s_j}{m^2}.$$

The null is $p_0(x) = (m^{-1}\sum_i x_i)^2$. A degree-two count and this ordered pair are equivalent experiments for testing: the order of two distinct labels, conditionally on the count, is a uniformly random ancillary bit under both hypotheses.

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

$$\chi^2(\mathcal{M}_N \| \mathcal{P}_0^{\otimes N}) = \mathbb{E}_{s,t}(1 + \rho^2 T^2)^N - 1 \le \mathbb{E}\exp(N\rho^2 T^2) - 1.$$

To bound T, fix s and generate uniform balanced t by a uniform perfect matching and independent choices of which endpoint of each pair has sign +1. Conditional on the matching, $s^T t$ is a sum of independent Rademachers with coefficients in $\{-2, 0, 2\}$. Since $\cosh x \le e^{x^2/2}$, its moment generating function 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

$$\mathbb{E}e^{bT^2} - 1 \le \int_0^\infty 2be^{by}e^{-my/4}\,dy = \frac{8N\rho^2}{m - 4N\rho^2}.$$

At the asserted sample size this is at most $1/12$. Thus total variation between null and mixture is at most $\frac{1}{2}\sqrt{1/12} < 1/3$. A test with both errors at most $1/3$ would produce a probability gap of at least $1/3$, a contradiction. Disjoint accuracy intervals about zero and ρ give the estimation conclusion. □

This is a standard hidden-sign mixture and second-moment argument specialized to a stable product subclass. It matches the category and accuracy order up to logarithms at fixed degree two.

7 Implementation and scope

The accompanying exact-arithmetic implementation uses count histograms from two batches. It replaces L by the integer b_0 satisfying $2^{b_0} \ge 4m/\delta$, so $L \le b_0$ and all budgets and smoothing parameters are rational. It encloses the positive part of the largest eigenvalue by exact positive-semidefinite tests and bisection. The implementation uses an eigenvalue enclosure width $\xi = 10^{-6}$ and reports it with the resulting upper slack $5\varepsilon/4 + \xi$, as allowed in Theorem 3.1. In particular, writing $d = \widehat{J}\mathbf{1}$, congruence by $R_H^{1/2}$ gives

$$G_H = -\kappa[\widehat{J}_H - r_H d_H^T - d_H r_H^T + r_H r_H^T],$$

so the eigenvalue upper test is the rational matrix inequality $\lambda R_H - G_H \succeq 0$. Here d is the second-stage empirical margin, not r. Exact Schur complements give a transparent small-dimensional implementation. Adaptive rational upper enclosures for the square root then produce a positive rational parity floor whenever the strict test succeeds.

For illustration, with $m = 2$, $k = 3$, $\varepsilon = 1/4$ and $\delta = .05$, this conservative integer-log implementation requires 1,835,008 mean-stage and 1,245,184 pair-stage counts. A synthetic independent $\mathrm{Bin}(3, 1/5)$ experiment with labels $0, 1$ passes the covariance-based sufficient test; its actual parity is $.6^3 = .216$. The

mean-only positivity bound is unavailable at the resulting reference margin near .585. The millions of observations illustrate the large constants rather than an economical experimental protocol. Rational arithmetic and dense matrix processing add costs beyond the number of observations.

The statistical contract assumes a stable count model, known category labels, and classical iid count observations in those coordinates. Real stability is a structural model promise. The intended operational use is classical calibration in known coordinates.

Disclosure of AI assistance

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

Statements and Declarations

Funding. This research received no external grant funding.

Competing interests. The author owns Luminary Oddities, which operates the Research Archive hosting the preprint.

Data and materials availability. The results are analytic. The accompanying source package contains the manuscript and the verification code and outputs supporting its reproducible calculations.

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

References

[1] 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.

[2] X. Xu, W. Wang and S.-L. Huang, On the sample complexity of estimating small singular modes, *2020 IEEE International Symposium on Information Theory*. Author manuscript.

[3] 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*. STOC, 2019. arXiv:1811.01816v3.

[4] C. M. Le, E. Levina and R. Vershynin, Concentration and regularization of random graphs, *Random Structures Algorithms* **51** (2017), 538–561. arXiv:1506.00669.

[5] I. Diakonikolas, D. M. Kane and A. Stewart, The Fourier transform of Poisson multinomial distributions and its algorithmic applications, arXiv:1511.03592v2 (2016).

[6] J. Borcea, P. Brändén and T. M. Liggett, Negative dependence and the geometry of polynomials, *J. Amer. Math. Soc.* **22** (2009), 521–567. arXiv:0707.2340.

[7] S. Schubr, *Moment certificates for homogeneous stable count laws*, Luminary Research Archive (2026). Archive edition 6.

[8] J. A. Tropp, User-friendly tail bounds for sums of random matrices, *Found. Comput. Math.* **12** (2012), 389–434. doi:10.1007/s10208-011-9099-z.

[9] P. A. Samuelson, How deviant can you be?, *J. Amer. Statist. Assoc.* **63** (1968), 1522–1525. doi:10.1080/01621459.1968.10480944.

[10] W. Tang and F. Tang, The Poisson binomial distribution—old & new, *Statistical Science* **38** (2023), 108–119. arXiv:1908.10024.