Moment certificates for homogeneous stable count laws

Scott Schubr
Luminary Oddities

Abstract

Let P be a probability law on nonnegative integer count vectors of total size n, with a homogeneous real-stable probability generating polynomial and positive mean vector t. We compare P with $M = \mathrm{Mult}(n, t/n)$ using the weighted covariance deficit

$$W = \frac{1}{2}\left(m - 1 - \sum_{i=1}^{m} \frac{\mathrm{Var}(A_i)}{t_i}\right).$$

We prove $\mathrm{TV}(P, M) \le \min\{1, CW\}$ with $C = 20/(2\sqrt{10} - 1) < 3.757$, independently of degree, category count and the mean vector. The argument transfers the centered elementary-symmetric-polynomial inequality of Han and Niles-Weed through Banach's theorem and an orthogonal multinomial density expansion, then uses positivity of the likelihood ratio. A classical hyperbolic-slice reduction gives finite exact coefficient certificates. A standalone rational-arithmetic verification proves the stronger bound $\mathrm{TV}(P, M) \le \min\{1, 2W\}$ for every $1 \le n \le 50$. Every all-degree linear coefficient must be at least two. The results include nonfactorizable stable generating polynomials.

1 Setting and relation to prior work

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

$$p(x) = \sum_{|a|=n} P(a)x^a, \qquad P(a) \ge 0, \qquad p(\mathbf{1}) = 1$$

be homogeneous and real stable. Here $|a| = \sum_i a_i$ and A has law P. Put $t_i = \mathbb{E}A_i > 0$, $q = t/n$, $L(x) = q \cdot x$, and $M = \mathrm{Mult}(n, q)$. Total variation is $\mathrm{TV}(P, M) = \frac{1}{2}\sum_a |P(a) - M(a)|$. Define

$$B = \frac{1}{2}\left(\mathrm{diag}(t) - \frac{tt^\top}{n} - \mathrm{Cov}(A)\right), \qquad W = \mathrm{Tr}(\mathrm{diag}(t)^{-1}B). \tag{1}$$

We will prove $B \succeq 0$. The identity $\sum_i t_i = n$ gives the formula for W in the abstract.

Multinomial approximation for dependent random variables is classical; Loh's Stein-method treatment [7] is a central predecessor. Ehm's binomial approximation [3] already treats the bivariate stable subclass, whose generating polynomials factor into linear forms. Roos [9] develops approximation of convolution products, including independent categorical sums.

The closest methodological predecessor is Han and Niles-Weed [5]. Their analysis of permutation mixtures combines centered elementary symmetric polynomials, Banach transfer and orthogonal chi-square expansions. Their Theorem 4.3(2), equation (27), supplies the coefficient inequality used below. Related Banach and coefficient methods occur in Roos [10]; later centered-vector and matrix

inequalities appear in [6]. Our argument adapts these tools to the centered hyperbolic expansion of an arbitrary homogeneous stable count law. For categorical observations, a single permutation mixture of independent coordinates has the same count generating polynomial as the underlying product of categorical linear forms. A general exchangeable mixture representation instead yields a mixture of independent reference laws; it does not directly give our one mean-matched multinomial or the deficit in (1).

The contribution here is this full-stable-class transfer, its explicit covariance certificate, and a reproducible finite-degree verification of coefficient two. Banach's theorem, the centered ESP inequality, hyperbolic-slice extremality, and the orthogonal-expansion architecture are credited inputs rather than new general mechanisms. Fixed degree and category count are separate parameters throughout.

2 Centered expansion and the exact density identity

Degree one is exactly multinomial. In the lemmas below we therefore assume $n \geq 2$.

Lemma 1. *There are homogeneous polynomials R_k of degree k, $2 \leq k \leq n$, invariant under $x \mapsto x + a\mathbf{1}$, such that*

$$p(x) = L(x)^n + \sum_{k=2}^{n} R_k(x) L(x)^{n-k}. \tag{2}$$

Moreover $B \succeq 0$, $B\mathbf{1} = 0$, and $R_2(x) = -x^\top Bx$.

Proof. For real x, set $y = x - L(x)\mathbf{1}$. The polynomial $p(z\mathbf{1} + y)$ is monic of degree n. Every root is real: an upper-half-plane root contradicts stability, and a lower-half-plane root has an upper conjugate. Its coefficient of z^{n-1} is $t \cdot y = 0$. Write its roots as $\lambda_1, \ldots, \lambda_n$, so $\sum_j \lambda_j = 0$, and let $R_k(x)$ be its coefficient of z^{n-k}. This defines homogeneous polynomials, invariant under addition of $a\mathbf{1}$ because y is unchanged. Substitution $z = L(x)$ gives (2).

Write $H = \nabla^2 p(\mathbf{1})$. The coefficient of z^{n-2} is $\frac{1}{2} y^\top H y$. Euler's identity gives $H\mathbf{1} = (n-1)t$ and $\mathbf{1}^\top H \mathbf{1} = n(n-1)$, so

$$R_2(x) = \tfrac{1}{2} x^\top \left(H - \frac{n-1}{n} t t^\top \right) x.$$

Since $H = \mathrm{Cov}(A) + tt^\top - \mathrm{diag}(t)$, this is $-x^\top Bx$. On the other hand the root identity gives

$$R_2(x) = -\tfrac{1}{2} \sum_j \lambda_j^2, \qquad \sum_j \lambda_j^2 = 2x^\top Bx. \tag{3}$$

Thus $B \succeq 0$; the displayed formula or translation invariance gives $B\mathbf{1} = 0$. □

If $W = 0$, positive t and $B \succeq 0$ imply $B = 0$. Equation (3) then makes every centered root zero for every real x, so all R_k vanish and $P = M$. This proves rigidity without dividing by W. The cases $n = 1$ or $m = 1$ are also exact multinomial laws. Zero-mean categories can be deleted because their nonnegative counts vanish almost surely.

For a homogeneous polynomial R_k, let T_k denote its symmetric k-linear form, with $R_k(x) = T_k(x, \ldots, x)$. Our ordered-index convention means that a monomial coefficient at α is $(k!/\alpha!)T_{k,\alpha}$.

Lemma 2. *Let ν_k be the signed coefficient measure of $R_k L^{n-k}$. There exist mutually orthogonal mean-zero densities $U_k = d\nu_k/dM$ satisfying*

$$\frac{P}{M} - 1 = \sum_{k=2}^{n} U_k, \qquad \chi^2(P\|M) = \sum_{k=2}^{n} \mathbb{E}_M U_k^2, \tag{4}$$

and

$$\mathbb{E}_M U_k^2 = \binom{n}{k}^{-1} \sum_{i_1,\dots,i_k} \frac{T_k(e_{i_1},\dots,e_{i_k})^2}{q_{i_1}\cdots q_{i_k}}. \tag{5}$$

Proof. Translation invariance of R_k gives $T_k(\mathbf{1}, v_2, \dots, v_k) = 0$: first differentiate the diagonal polynomial and then polarize its remaining arguments. Let $X_1, \dots, X_n$ be independent categorical variables of law q and set

$$H_k(i_1,\dots,i_k) = \frac{T_k(e_{i_1},\dots,e_{i_k})}{q_{i_1}\cdots q_{i_k}}, \qquad U_k = \binom{n}{k}^{-1} \sum_{|S|=k} H_k(X_S).$$

Symmetry makes U_k a function of the count vector. Summing any one argument of H_k against q gives zero by the contraction identity. Hence the kernel is completely conditionally degenerate. For each fixed subset S,

$$\mathbb{E}[H_k(X_S) x^{\mathrm{counts}(X)}] = R_k(x) L(x)^{n-k}.$$

Coefficient comparison after symmetrization proves $\nu_k(a) = M(a)U_k(a)$. All $M(a)$ on $|a| = n$ are positive. In a product of two kernels associated with different subsets, an index in the symmetric difference occurs in only one kernel; integrate that independent categorical observation to obtain zero. This applies both within an order and between distinct orders. Thus only identical subsets contribute to the square norm, giving (5), while (2) gives (4). All operations are finite sums. □

3 Hilbert tensor transfer and a universal certificate

We use Banach's classical real Hilbert-space theorem [1, 2]: the norm of a symmetric multilinear form on arbitrary unit-vector tuples equals its norm on diagonal tuples. It is also the input in [5, Lemma 4.2].

Lemma 3. *If $|R_k(x)| \le c(x^\top B x)^{k/2}$ for all real x, then*

$$\mathbb{E}_M U_k^2 \le \frac{c^2 (nW)^k}{\binom{n}{k}}. \tag{6}$$

Proof. For $v \in \ker B$, the polynomial $R_k(x + sv)$ is bounded independently of real s, hence is constant. Differentiation and polarization give $T_k(v, v_2, \dots, v_k) = 0$. Thus T_k descends to the real Hilbert quotient by $\ker B$, with squared norm $x^\top B x$. Banach's theorem there implies

$$|T_k(v_1,\dots,v_k)| \le c \prod_j \sqrt{v_j^\top B v_j}.$$

Apply this to coordinate vectors in (5) and sum the products to obtain

$$\mathbb{E}_M U_k^2 \le \binom{n}{k}^{-1} c^2 \left(\sum_i \frac{B_{ii}}{q_i} \right)^k = \binom{n}{k}^{-1} c^2 (nW)^k.$$

No inverse of B is used. □

Han and Niles-Weed [5, Theorem 4.3(2)] prove that a real centered vector z with $\sum_j z_j^2 = n$ satisfies $|e_k(z)|^2 \le 10\binom{n}{k}$. Scaling the centered roots in (3) therefore yields, with $u = x^\top Bx$,

$$|R_k(x)|^2 \le 10\binom{n}{k}(2u/n)^k.$$

The zero case $u = 0$ follows directly from the roots. Lemmas 2 and 3 now give

$$\mathbb{E}_M U_k^2 \le 10(2W)^k, \qquad \chi^2(P\|M) \le \frac{40W^2}{1-2W} \quad (2W < 1). \tag{7}$$

The next elementary bound retains positivity of the actual density. It is the upper-distance branch of the classical sharp relation between chi-square divergence and total variation; see Guntuboyina, Saha and Schiebinger [4, Corollary 5.3], which credits Gilardoni's Proposition 2.1. We give the short proof needed here.

Lemma 4. *For probability measures $P \ll M$, write $v = \chi^2(P\|M)$ and $\delta = \mathrm{TV}(P, M)$. If $\delta > 1/2$, then $v \ge \delta/(1-\delta)$.*

Proof. Let $Y = dP/dM - 1$ and $a = M(Y < 0)$. Then $\mathbb{E}Y_+ = \mathbb{E}Y_- = \delta$ and $Y_- \le 1$, so $a \ge \delta$. Cauchy–Schwarz on the two parts of the partition gives

$$v \ge \frac{\delta^2}{a} + \frac{\delta^2}{1-a} = \frac{\delta^2}{a(1-a)}.$$

For $a \ge \delta > 1/2$, $a(1-a) \le \delta(1-\delta)$. The degenerate partition cases force $\delta = 0$ and do not occur here. □

Theorem 5. *For every law in our setting,*

$$\mathrm{TV}(P, M) \le \min\{1, C_* W\}, \qquad C_* = \frac{20}{2\sqrt{10}-1} = \frac{20+40\sqrt{10}}{39} < 3.757.$$

Proof. Put $C = C_*$. For $0 < W \le 1/(2C)$, (7) and ordinary Cauchy–Schwarz give

$$\frac{\delta}{W} \le \frac{\sqrt{10}}{\sqrt{1-2W}} \le \frac{\sqrt{10}}{\sqrt{1-1/C}} \le C.$$

The last inequality holds because $C^2 - C - 10 = (C+10)/39 > 0$, using $39C^2 - 40C - 400 = 0$.

For $1/(2C) < W < 1/C$, suppose $\delta > CW$. Then $\delta > 1/2$, so Lemma 4 gives $v > CW/(1-CW)$. But (7) gives the opposite weak inequality because

$$\frac{40W^2}{1-2W} \le \frac{CW}{1-CW} \quad \Longleftrightarrow \quad 40CW^2 - (40+2C)W + C \ge 0.$$

Both denominators are positive since $C > 2$. The quadratic on the right has positive leading coefficient and discriminant $(40+2C)^2 - 160C^2 = 0$, by the definition of C. This contradiction proves the claim in this range. For $W \ge 1/C$, use $\delta \le 1$. The case $W = 0$ was settled by rigidity. □

4 Exact finite-degree certificates

The coefficient extremum reduction used next is classical: see Riener and Schabert [8, Theorem 2.8 and Corollary 2.10], as well as the real-vector reduction in [5]. We include the needed special proof.

Lemma 6. *For fixed $n \geq 3$ and $3 \leq k \leq n$, the maximum of $|e_k(\lambda)|$ on $\sum_i \lambda_i = 0$, $\sum_i \lambda_i^2 = 2$ is attained at a vector with two distinct values.*

Proof. The slice is nonempty and compact. Maximize either sign of e_k, and among maximizers select one with the largest number d of distinct roots of $f(z) = \prod_i (z - \lambda_i)$. If $3 \leq d < n$, write $f = hg$, where g contains each distinct linear factor once and h retains multiplicity minus one. For either sufficiently small sign of ε, $g + \varepsilon$ has d simple real roots near the old roots. Thus $f + \varepsilon h = h(g + \varepsilon)$ is real-rooted and preserves the first two coefficients, since $\deg h = n - d \leq n - 3$. If the target coefficient of h is nonzero, one sign improves the objective. Otherwise the objective is constant. Every repeated old root remains through h and also supplies a new nearby root through $g + \varepsilon$; no new root equals an old one. Taking disjoint neighborhoods of distinct old roots shows the distinct-root count strictly increases, a contradiction. If $d = n$, perturbing by εz^{n-k} preserves the first two coefficients and real-rootedness for both small signs, but changes the objective. Hence $d \leq 2$. The positive square sum rules out $d = 1$. Apply this argument to both signs to obtain the absolute-value assertion. □

For $r + s = n$, a normalized two-valued vector has values a repeated r times and $b = -ra/s$ repeated s times, with $a^2 = 2s/(nr)$. Thus define the exact rational constants

$$c_{n,k}^2 = \max_{1 \leq r < n} \left(\frac{2s}{nr}\right)^k \left[\sum_{j=\max(0,k-s)}^{\min(k,r)} \binom{r}{j}\binom{s}{k-j}(-r/s)^{k-j}\right]^2, \tag{8}$$

$$A_{n,k} = \frac{n^k c_{n,k}^2}{\binom{n}{k}}, \qquad V_n(w) = \sum_{k=2}^{n} A_{n,k} w^k.$$

For $k = 2$, $e_2 = -1$ on the normalized slice, so the same formula applies. Homogeneity, Lemma 6 and the tensor transfer prove

$$\chi^2(P\|M) \leq V_n(W). \tag{9}$$

These constants come from proved extrema and finite arithmetic.

Proposition 7. *If*

$$S_n = \sum_{k=2}^{n} A_{n,k} 4^{-(k-2)} \leq 16, \qquad D_n = \sum_{k=2}^{n} A_{n,k} \frac{(k-1)^{k-1}}{2^{k-1} k^k} \leq 2,$$

then every law of degree n in our setting satisfies $\mathrm{TV}(P, M) \leq \min\{1, 2W\}$.

Proof. For $0 < W \leq 1/4$, nonnegative coefficients give $V_n(W) \leq S_n W^2 \leq 16W^2$, and $\delta \leq \sqrt{v}/2$ proves $\delta \leq 2W$. For $1/4 < W < 1/2$, differentiation gives

$$\max_{0 \leq w \leq 1/2} w^{k-1}(1 - 2w) = \frac{(k-1)^{k-1}}{2^{k-1} k^k}.$$

Consequently $(1 - 2W)V_n(W)/W \leq D_n \leq 2$, hence $v \leq 2W/(1 - 2W)$. If $\delta \leq 1/2$ the conclusion is immediate. Otherwise Lemma 4 gives $\delta \leq v/(1 + v) \leq 2W$. For $W \geq 1/2$ use $\delta \leq 1$, and for $W = 0$ use rigidity. □

Theorem 8. *For every degree $1 \le n \le 50$, every category count, and every positive mean vector in the stable setting,*

$$\mathrm{TV}(P, \mathrm{Mult}(n, t/n)) \le \min\{1, 2W\}.$$

Proof. Degree one is exact. For $2 \le n \le 50$, the accompanying standalone program `verify_coefficients.py` evaluates every candidate in (8) with integer and rational arithmetic, forms S_n, D_n, and checks their exact inequalities against 16 and 2. All 49 degrees pass. The largest S_n is $981/200$ at $n = 6$, and all $D_n < 141/100 < 2$. Proposition 7 proves the result.

For reproducibility, the program uses integer roots s repeated r times and $-r$ repeated s times. Their squared sum is nrs; their degree-k elementary coefficient is

$$b_{r,s,k} = \sum_j \binom{r}{j}\binom{s}{k-j} s^j (-r)^{k-j}.$$

The normalized squared candidate is $b_{r,s,k}^2/(nrs/2)^k$. This is algebraically identical to (8), avoids square roots entirely, and is the full finite computation required for the theorem. The program enumerates all $r = 1, \dots, n-1$ and all $k = 2, \dots, n$, using Python's standard-library arbitrary-precision integers and `Fraction`. It checks the stated maximum of S_n and the stronger common rational bound $D_n < 141/100$. There is no floating-point decision in the certificate. The analytic reduction and finite code are both parts of this proof; no inference to untested degrees is made. □

For illustration, at $n = 8$ the coefficients $A_{8,2}, \dots, A_{8,8}$ are

$$\frac{16}{7},\ \frac{256}{49},\ \frac{288}{35},\ \frac{4096}{189},\ \frac{256}{7},\ \frac{12288}{125},\ 256.$$

Here $S_8 = 12558191/2646000$ and $D_8 = 15114877380793361149/15738151816396800000$. Although $V_8(1/2)/(1/2)^2 > 16$, the positive-density criterion succeeds. Thus discarding positivity at the last TV step would leave an avoidable obstruction.

5 Limitations and the coefficient-two question

Every universal coefficient is at least two. For even n, take the stable monomial $p(x_1, x_2) = x_1^{n/2} x_2^{n/2}$. Then $W = 1/2$ and

$$\mathrm{TV}(P, \mathrm{Mult}(n, 1/2)) = 1 - 2^{-n}\binom{n}{n/2} \longrightarrow 1.$$

This elementary binomial example gives the lower barrier. Consequently the optimal all-degree coefficient satisfies

$$2 \le C_{\mathrm{opt}} \le \frac{20}{2\sqrt{10} - 1}.$$

The number fifty is a verified computational cutoff.

The mean vector and covariance are treated as known inputs. We prove no sample-complexity result for estimating W, no practical testing guarantee, no sampler construction and no computational speedup. Stability is a substantive hypothesis; matching a few moments without it does not certify the full law. The finite result and universal theorem allow singular covariance deficit, repeated roots and zero coefficients. Signed centered-expansion terms are not themselves probability laws, so the positivity lemma is applied only to the full density P/M.

Computational and AI assistance

The finite certificate uses only the accompanying standard-library Python source, which prints exact rational quantities and terminates with an explicit failure if a required inequality fails. No random sampling or external numerical package is used. Generative AI systems assisted with mathematical exploration, literature discovery, proof drafting, exact-code development and independent adversarial proof checks. These checks are internal AI assistance. Mathematical claims are presented with their proofs and reproducible computation so that they can be assessed independently; the author retains responsibility for the manuscript. Classical inputs and the close Han–Niles-Weed methodology are explicitly attributed.

References

[1] S. Banach, Über homogene Polynome in L^2, *Studia Mathematica* **7** (1938), 36–44. doi:10.4064/sm-7-1-36-44.

[2] D. Carando and J. T. Rodríguez, Symmetric multilinear forms on Hilbert spaces: where do they attain their norm? (2018). arXiv:1810.09373.

[3] W. Ehm, Binomial approximation to the Poisson binomial distribution, *Statistics & Probability Letters* **11**(1) (1991), 7–16. doi:10.1016/0167-7152(91)90170-V.

[4] A. Guntuboyina, S. Saha and G. Schiebinger, Sharp inequalities for f-divergences (2013). arXiv:1302.0336.

[5] Y. Han and J. Niles-Weed, Approximate independence of permutation mixtures, *The Annals of Statistics* **54**(3) (2026), 1594–1621. doi:10.1214/26-AOS2619. arXiv:2408.09341; the coefficient result used here is Theorem 4.3(2) in version 3.

[6] Y. Han and J. Niles-Weed, Elementary symmetric polynomial inequalities for centered vectors and matrices (2026). arXiv:2607.23836.

[7] W.-L. Loh, Stein's method and multinomial approximation, *The Annals of Applied Probability* **2**(3) (1992), 536–554. doi:10.1214/aoap/1177005648.

[8] C. Riener and R. Schabert, Linear slices of hyperbolic polynomials and positivity of symmetric polynomial functions (2023). arXiv:2203.08727, version 2.

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

[10] B. Roos, On Bobkov's approximate de Finetti representation via approximation of permanents of complex rectangular matrices (2014). arXiv:1203.2871, version 2.