hexagon:2611.00708v1

The exact Poisson–binomial comparison in every dimension

Christopher D. Long
galizur@gmail.com

September 13, 2026

Abstract

For coordinatewise ordered Bernoulli parameters with a prescribed total mean gap, the maximum difference of count tails is minimized by homogeneous parameters. For a nondegenerate gap $0 < \Delta < n$, in odd dimension $n = 2m + 1 \geq 3$ the two minimizers are the homogeneous pairs whose likelihood ratios equal one at count m or $m + 1$; in even dimension $n \geq 4$ the unique minimizer is the complementary homogeneous pair. We prove the comparison and these equality statements by a global stationary-point reduction, with all boundary faces included. A normalized switch-mass identity orders the homogeneous switching values. The same extremizers give the exact product total-variation minimum. The exceptional two-dimensional equality families are classified in the appendix.

1 Statement

For $p \in [0,1]^n$, let S_p be a sum of independent Bernoulli variables with parameters p_i. Put

$$D_k(p,q) = \mathbb{P}(S_p \geq k) - \mathbb{P}(S_q \geq k), \qquad T(p,q) = \max_{1 \leq k \leq n} D_k(p,q), \qquad \Delta = \sum_i (p_i - q_i).$$

For probability measures on a finite set, we use $\mathrm{TV}(P,Q) = \frac{1}{2}\sum_x |P(x) - Q(x)|$. We assume $p \geq q$ coordinatewise. For $n \geq 2$, $1 \leq j \leq n-1$, and $0 < \gamma < 1$, there is a unique pair $0 < b < a < 1$ satisfying

$$a - b = \gamma, \qquad a^j (1-a)^{n-j} = b^j (1-b)^{n-j}. \tag{1.1}$$

Define

$$F_{n,j}(\gamma) = \mathbb{P}(\mathrm{Bin}(n,a) \geq j) - \mathbb{P}(\mathrm{Bin}(n,b) \geq j), \qquad \mathcal{B}_n(\Delta) = F_{n,\lfloor n/2 \rfloor}(\Delta/n).$$

Set $\mathcal{B}_n(0) = 0$ and $\mathcal{B}_n(n) = 1$.

Theorem 1.1. *For every $n \geq 2$ and every $p, q \in [0,1]^n$ with $p \geq q$,*

$$T(p,q) \geq \mathcal{B}_n(\Delta). \tag{1.2}$$

For $n \geq 3$ and $0 < \Delta < n$, equality forces both vectors to be homogeneous. More precisely:

1. *If n is even, equality holds exactly when $p_i = (1 + \Delta/n)/2$ and $q_i = (1 - \Delta/n)/2$ for all i.*
2. *If $n = 2m+1$ is odd, let (a,b) solve (1.1) with $j = m$ and $\gamma = \Delta/n$. Equality holds exactly for*

$$(p,q) = (a\mathbf{1}, b\mathbf{1}) \quad \text{or} \quad (p,q) = ((1-b)\mathbf{1}, (1-a)\mathbf{1}).$$

The exact minimum of $\mathrm{TV}(\bigotimes_i \mathrm{Ber}(p_i), \bigotimes_i \mathrm{Ber}(q_i))$ *under the same constraint is also* $\mathcal{B}_n(\Delta)$*, with the same equality cases for* $n \geq 3$ *and* $0 < \Delta < n$*.*

The exceptional equality cases for $n = 2$ are classified in Proposition A.1 of Appendix A.

For odd $n = 2m + 1$, the benchmark can also be written

$$\mathcal{B}_{2m+1}(\Delta) = (2m+1)\binom{2m}{m}\int_b^a [u(1-u)]^m \, \mathrm{d}u, \tag{1.3}$$

where (a, b) is the switch with $j = m$. Equivalently, let $z > 1$ be the unique solution of

$$\frac{(z^m - 1)(z^{m+1} - 1)}{z^{2m+1} - 1} = \frac{\Delta}{2m+1}.$$

Then

$$b = \frac{z^m - 1}{z^{2m+1} - 1}, \qquad a = \frac{z^{m+1}(z^m - 1)}{z^{2m+1} - 1}.$$

These formulas follow directly from the equal-height equation.

The even-dimensional comparison and related product optimization were posed by Kontorovich [3, 2]. The discussion of [3] includes Steinke's randomized symmetrization and Sawin's fixed-mean tail-extremization approach. General lower bounds for product total variation in terms of marginal distances are developed in [4]. More directly, [5, Theorem 1.5] compares Bernoulli products with their mean-preserving homogenizations up to a universal constant. Here only the total gap is fixed, and we determine the exact minimum and all equality cases. The proof combines the deletion-mixture bound in Lemma 2.2, the switch identities in Lemma 3.1, and a global extremizer argument. The even-dimensional benchmark is recovered as

$$\mathcal{B}_n(\Delta) = 2^{-n}\binom{n}{n/2}\int_0^{\Delta} (1 - u^2/n^2)^{n/2-1} \, \mathrm{d}u.$$

For $n = 3$, (1.3) becomes

$$\mathcal{B}_3(\Delta) = \frac{2\Delta}{9}\left(1 + \sqrt{1 - \Delta^2/12}\right).$$

The proof proceeds by induction from $n = 2$.

2 Bernoulli facts and the randomized test

We use the likelihood-ratio order only for Bernoulli sums. Their mass functions have interval support and are strictly log-concave at positive interior triples. A deterministic shift preserves these facts. To see strict log-concavity, remove the deterministic variables and write the remaining probability generating polynomial as $\prod_{i=1}^{M}(1 - t_i + t_i z)$, where $0 < t_i < 1$. Its zeros are real and negative. If its coefficients are $a_0, \ldots, a_M$, Newton's inequality (see [6]) gives, for $1 \leq j \leq M - 1$,

$$a_j^2 \geq a_{j-1}a_{j+1}\frac{(j+1)(M-j+1)}{j(M-j)} > a_{j-1}a_{j+1}.$$

Here and below a comparison of ratios is understood by cross multiplication when an endpoint mass is zero.

Lemma 2.1 (Increasing a parameter)**.** *Increasing Bernoulli parameters, or adjoining a Bernoulli variable, increases the law in likelihood-ratio order. If at least one parameter increases strictly, the adjacent likelihood-ratio comparison is strict when both adjacent masses are positive in both laws. In particular, if* $\mathbb{P}(Z = k-1) = \mathbb{P}(Z = k) > 0$*, and* Z' *is obtained by a strict parameter increase, then*

$$\mathbb{P}(Z' = k) > 0 \quad \Longrightarrow \quad \mathbb{P}(Z' = k-1) < \mathbb{P}(Z' = k).$$

If the mass at k *disappears because the lower endpoint of the support moves past* k*, both displayed masses of* Z' *are zero instead. On a parameter interval where both adjacent masses remain positive, their difference can cross zero only from positive to negative.*

Proof. With all other variables summed into R, changing a parameter from u to $v > u$ changes the mass at j from $(1-u)r_j + ur_{j-1}$ to $(1-v)r_j + vr_{j-1}$. Their ratio increases with r_{j-1}/r_j, which increases strictly in j on the positive support by strict log-concavity. Equivalently, the relevant adjacent determinant is a positive multiple of $r_{j-1}^2 - r_{j-2}r_j$. Successive coordinate changes prove the assertion. Adjoining a variable is the change of an appended parameter from zero to its new value. □

For ordered p, q with $\Delta > 0$ and $T(p,q) < 1$, Lemma 2.1 shows that the maximizing thresholds consist of one threshold or two adjacent thresholds. Indeed the mass difference changes sign once, and can vanish at at most one count in the overlapping support. Disjoint supports would give $T = 1$.

For $1 \le k \le n-1$ and $0 \le \lambda \le 1$, define

$$H(x) = \lambda \mathbb{P}(S_x \ge k) + (1-\lambda)\mathbb{P}(S_x \ge k+1).$$

An independent $E \sim \mathrm{Ber}(\lambda)$ gives $H(x) = \mathbb{P}(S_x + E \ge k+1)$. Write $H_i = \partial_i H$ and $C_{ij} = \partial_i \partial_j H$. Direct differentiation gives

$$H_i(x) = \mathbb{P}(S_{x,-i} + E = k), \tag{2.1}$$

$$C_{ij}(x) = \mathbb{P}(S_{x,-ij} + E = k-1) - \mathbb{P}(S_{x,-ij} + E = k), \tag{2.2}$$

$$H_i(x) - H_j(x) = (x_j - x_i)C_{ij}(x). \tag{2.3}$$

A pure threshold is included by taking $\lambda = 0$ or 1. For fixed other coordinates, splitting two equal coordinates u, u into $u+t, u-t$ changes H by exactly $-C_{ij}t^2$.

Lemma 2.2 (A maximal-atom bound)**.** *Put*

$$\kappa_n = \binom{n}{\lfloor n/2 \rfloor} \left(\frac{\lfloor n/2 \rfloor}{n}\right)^{\lfloor n/2 \rfloor} \left(\frac{\lceil n/2 \rceil}{n}\right)^{\lceil n/2 \rceil}.$$

Every $(n-1)$*-trial Bernoulli sum has a mass at least* κ_n*. If at most two of the gaps* $p_i - q_i$ *are positive, then*

$$T(p,q) \ge \kappa_n \Delta. \tag{2.4}$$

Proof. First, an n-trial Bernoulli sum of integer mean has probability at that integer at least κ_n. To prove this, minimize that probability at fixed mean. The fixed-mean reduction to equal interior parameters is classical; see Hoeffding [1]. Here is the two-coordinate argument needed in this case. Among the minimizers choose one with least sum of squared parameters. If two interior parameters $u \ne v$ occur, then, with the others fixed, the objective has the form $A + B(u+v) + Cuv$. The transfer $(u,v) \mapsto (u+t, v-t)$ is feasible for both signs of sufficiently small t, so minimality gives $C(v-u) = 0$. Thus $C = 0$, and replacing u, v by their average preserves the objective while strictly

decreasing the sum of squares, a contradiction. Consequently an extremizer is a deterministic shift plus $\mathrm{Bin}(M, \ell/M)$ with integer ℓ and $M \le n$. For

$$b_{M,\ell} = \binom{M}{\ell} (\ell/M)^{\ell} (1 - \ell/M)^{M-\ell}, \qquad h(x) = x \log(1 + 1/x), \quad h(0) = 0,$$

ratio calculations give

$$\log \frac{b_{M,\ell}}{b_{M,\ell+1}} = h(M - \ell - 1) - h(\ell), \qquad \log \frac{b_{M+1,\ell}}{b_{M,\ell}} = h(M - \ell) - h(M).$$

The increasing function h shows that the least n-trial integer-mean mass is κ_n, and that increasing M with ℓ fixed can only decrease the mass. Deterministic cases give mass one.

If $\mathbb{E}Z = k + f$ with $0 \le f \le 1$, adjoin $\mathrm{Ber}(1 - f)$. Its mass at the integer mean $k + 1$ is $(1 - f)\mathbb{P}(Z = k) + f\mathbb{P}(Z = k + 1) \ge \kappa_n$, proving the first assertion. This extra Bernoulli variable explains the index n in the bound for an $(n - 1)$-trial sum.

For the second assertion, the case $\Delta = 0$ is immediate; assume $\Delta > 0$. Put $r_i = (p_i + q_i)/2$, $d_i = p_i - q_i$ and $w_i = d_i/\Delta$. When at most two d_i are nonzero, symmetric expansion of the generating polynomials gives exactly

$$\frac{D_{j+1}}{\Delta} = \sum_i w_i \mathbb{P}(S_{r,-i} = j).$$

This deletion mixture is an $(n - 1)$-trial Bernoulli law. For completeness, its generating polynomial is $G(z) = \sum_i w_i \prod_{\ell \ne i}(1 - r_\ell + r_\ell z)$. For distinct interior r_i and positive weights, factor out the full product: the remaining rational function has positive residues at its negative poles and therefore one zero between each successive pair of poles. Thus G has only negative real zeros, and $G(1) = 1$ factors it into Bernoulli generating factors. Continuity covers all degeneracies. The first assertion now proves (2.4). □

We will repeatedly use one more elementary consequence of Lemma 2.1. If W is a fixed Bernoulli sum and

$$g_M(t) = \mathbb{P}(\mathrm{Bin}(M, t) + W = k), \tag{2.5}$$

then

$$g'_M(t) = M\{\mathbb{P}(\mathrm{Bin}(M - 1, t) + W = k - 1) - \mathbb{P}(\mathrm{Bin}(M - 1, t) + W = k)\}.$$

The derivative formula is used for $M \ge 1$; g_0 is simply a constant. On $0 < t < 1$ the support is fixed. Thus the derivative changes sign at most once, from positive to negative. An interior critical point of positive height is its maximum; it is strict when $M \ge 2$. For $M = 1$ an interior critical point means that g_1 is constant.

We also need the following adjoining rule, with its positivity condition explicit: if a Bernoulli sum Z satisfies

$$\mathbb{P}(Z = k - 1) \le \mathbb{P}(Z = k), \qquad \mathbb{P}(Z = k) > 0,$$

then adjoining further independent Bernoulli variables cannot increase its mass at k. Indeed, write $a = \mathbb{P}(Z = k - 2)$, $b = \mathbb{P}(Z = k - 1)$ and $c = \mathbb{P}(Z = k) > 0$. Log-concavity and interval support give $a \le b \le c$. After adjoining $\mathrm{Ber}(t)$ the relevant masses are $(1 - t)b + ta$ and $(1 - t)c + tb$, so their order persists and the mass at k does not increase. If that mass becomes zero, the support has moved strictly above k, and further adjoining cannot recreate it. If $b = c > 0$ and $t > 0$, strict log-concavity gives $a < b$; the first addition makes the adjacent difference strictly negative, and a second addition with positive parameter strictly decreases the mass at k. All uses of this rule below start with a positive mass at k.

Corollary 2.3 (Derivative criterion for adjoining)**.** *Let g_M be as in* (2.5)*, with $M \geq 1$ and $0 < t < 1$. If $g_M(t) > 0$ and $g'_M(t) \leq 0$, then*

$$\mathbb{P}(\mathrm{Bin}(M,t) + W = k-1) \leq \mathbb{P}(\mathrm{Bin}(M,t) + W = k).$$

Consequently adjoining further independent Bernoulli variables cannot increase the mass at k.

Proof. Write $Y = \mathrm{Bin}(M-1,t) + W$ and $y_j = \mathbb{P}(Y = j)$. The derivative condition gives $y_{k-1} \leq y_k$, and $g_M(t) > 0$ then implies $y_k > 0$. Log-concavity and interval support give $y_{k-2} \leq y_{k-1} \leq y_k$. For $Z = Y + \mathrm{Ber}(t)$, it follows that

$$\mathbb{P}(Z = k-1) = (1-t)y_{k-1} + ty_{k-2} \leq (1-t)y_k + ty_{k-1} = \mathbb{P}(Z = k).$$

The adjoining rule applies. □

3 The homogeneous problem in both parities

Suppress the first index in $F_{n,j}$, and write $c = j/n$. Let (a,b) satisfy (1.1), and set

$$v = c(1-c), \quad x = a - c, \quad y = c - b, \quad R = \frac{xy}{v}, \quad f_j(\gamma) = \binom{n}{j} a^j (1-a)^{n-j}.$$

Here $x, y > 0$, and

$$v(1-R) = c(1-a)(1-b) + (1-c)ab > 0. \tag{3.1}$$

The binomial masses at count j agree, so thresholds $j, j+1$ tie and both equal $F_j(\gamma)$.

Lemma 3.1 (Switch identities)**.** *Primes denote differentiation in γ. Then*

$$F'_j = \frac{nf_j}{1-R}, \qquad \frac{f'_j}{f_j} = -\frac{nR}{\gamma(1-R)}. \tag{3.2}$$

Consequently,

$$F_j(\gamma) = (n+1)\int_0^\gamma f_j(u)\,\mathrm{d}u - \gamma f_j(\gamma), \qquad \int_0^1 f_j(u)\,\mathrm{d}u = \frac{1}{n+1}. \tag{3.3}$$

For fixed $\gamma \in (0,1)$, R is strictly decreasing in c on $(0,1/2)$. Also $R' \leq 2R/\gamma$.

Proof. The equation defining b has derivative

$$-\gamma\left\{\frac{c}{ab} + \frac{1-c}{(1-a)(1-b)}\right\} < 0$$

in b, with opposite infinite endpoint limits. This proves existence and uniqueness of the switch. Differentiating the equal-height equation and $a - b = \gamma$, and using (3.1), gives

$$a' = \frac{ya(1-a)}{\gamma v(1-R)}, \qquad b' = -\frac{xb(1-b)}{\gamma v(1-R)}.$$

Differentiation of the binomial tail proves (3.2). In particular $F'_j = nf_j - \gamma f'_j$. Integration, followed by $F_j(1) = 1$ and $f_j(1) = 0$, proves (3.3).

We give the geometric monotonicity explicitly. Suppose $c \le 1/2$, and put $z = 1 - 2c$ and $e = a + b - 2c = x - y$. Let $A = \operatorname{logit}(a)$, $B = \operatorname{logit}(b)$, $M = (A + B)/2$, $s = (A - B)/2$, and $\sigma(t) = e^t/(1 + e^t)$. The equal-height equation says

$$c = \frac{1}{2s} \int_{M-s}^{M+s} \sigma(t)\, \mathrm{d}t.$$

Thus $M \le 0$. The identity

$$\sigma(M + u) + \sigma(M - u) = 1 + \frac{\sinh M}{\cosh M + \cosh u}$$

shows that the paired average increases with $u \ge 0$. Hence $\sigma(M) \le c \le (a + b)/2$, strictly if $c < 1/2$. The first inequality is equivalent to

$$c^2(1 - a)(1 - b) - (1 - c)^2 ab = v(zR - e) \ge 0,$$

and the second gives $e \ge 0$.

Now keep γ fixed and differentiate in the continuous parameter c. The derivative $t = \partial_c a = \partial_c b$ is positive, since the derivative of the switch equation in c is $A - B > 0$ whereas its derivative in b is negative. Therefore

$$\partial_c R = -\frac{(zR - e) + et}{v} < 0 \qquad (0 < c < 1/2). \tag{3.4}$$

Finally set $\theta = x/\gamma$. Differentiation gives $\theta' = (zR - e)/[\gamma^2(1 - R)] \ge 0$, and $e \ge 0$ gives $\theta \ge 1/2$. Since $R/\gamma^2 = \theta(1 - \theta)/v$, its derivative is nonpositive. Reflection covers $c > 1/2$ and proves the last assertion. □

Proposition 3.2 (Strict ordering of switches)**.** *For $1 \le i < j \le n/2$ and $0 < \gamma < 1$,*

$$F_{n,i}(\gamma) > F_{n,j}(\gamma).$$

Consequently the homogeneous minimum is $\mathcal{B}_n(\Delta)$, with exactly the central switch when n is even and the two reflected central switches when n is odd. For $n \ge 3$, $\mathcal{B}_n$ is strictly concave on $(0, n)$ and

$$0 < \mathcal{B}_n(\Delta) < \min\{\kappa_n \Delta, 1\} \qquad (0 < \Delta < n). \tag{3.5}$$

Proof. By (3.4), $R_i > R_j$. Thus

$$\left(\log \frac{f_i}{f_j}\right)' = -\frac{n}{\gamma}\left(\frac{R_i}{1 - R_i} - \frac{R_j}{1 - R_j}\right) < 0.$$

The equal integrals in (3.3) imply that $h = f_i - f_j$ changes sign exactly once, from positive to negative. Where $h > 0$, writing $A_\ell = nR_\ell/[\gamma(1 - R_\ell)]$ gives $h' = -A_i h - (A_i - A_j) f_j < 0$. Before the crossing, $\int_0^\gamma h \ge \gamma h(\gamma) > 0$, and afterwards $\int_0^\gamma h = -\int_\gamma^1 h > 0$. In both cases

$$F_i - F_j = (n + 1) \int_0^\gamma h - \gamma h(\gamma) > 0.$$

Reflection gives $F_{n,j} = F_{n,n-j}$.

It remains to justify that switches exhaust the homogeneous minima. Define

$$\mathcal{D}_h(a, b) = D_h(a\mathbf{1}, b\mathbf{1}), \qquad M_\gamma(b) = \max_{1 \le h \le n} \mathcal{D}_h(b + \gamma, b).$$

At fixed γ, each threshold difference has the form

$$\mathcal{D}_h(b+\gamma, b) = n \int_b^{b+\gamma} \binom{n-1}{h-1} u^{h-1}(1-u)^{n-h}\,\mathrm{d}u.$$

Its derivative has the sign of the ratio of the two endpoint densities minus one. Strict log-concavity makes that ratio strictly decreasing in b, so the individual threshold difference has no interior local minimum. Suppose that $b_0 \in (0, 1-\gamma)$ is a local minimum of M_γ. If exactly one threshold is active at b_0, its strict separation from the other finitely many thresholds persists in a neighborhood. There M_γ equals that threshold difference, a contradiction. Thus at least two thresholds are active. Here $0 < b_0 < b_0 + \gamma < 1$, so both count laws have full support and $T < 1$. The active-threshold property in Section 2 therefore shows that they are exactly two adjacent thresholds $k, k+1$. Their equality is equivalent to

$$\mathbb{P}(\mathrm{Bin}(n, b_0+\gamma) = k) = \mathbb{P}(\mathrm{Bin}(n, b_0) = k),$$

so b_0 is a switch. The homogeneous minimum is therefore attained at a switch or at $b = 0, 1-\gamma$. At an endpoint its value is $1-(1-\gamma)^n$, strictly larger than a central switch value: in the coordinatewise uniform coupling of a central homogeneous pair, the event of at least one discrepancy has this probability, and the positive-probability event of exactly one discrepancy with all other coordinates zero does not cross threshold $\lfloor n/2 \rfloor + 1$.

Finally (3.2) and $R' \le 2R/\gamma$ give

$$(\log F_j')' \le -\frac{(n-2)R}{\gamma(1-R)} < 0 \quad (n \ge 3).$$

For the central index $j = \lfloor n/2 \rfloor$, as $\gamma \downarrow 0$ we have $a, b \to j/n$ and $R \to 0$. Since $\Delta = n\gamma$,

$$\mathcal{B}_n'(0+) = \lim_{\gamma \downarrow 0} \frac{1}{n} F_{n,j}'(\gamma) = \lim_{\gamma \downarrow 0} \frac{f_j(\gamma)}{1-R(\gamma)} = \binom{n}{j} \left(\frac{j}{n}\right)^j \left(1-\frac{j}{n}\right)^{n-j} = \kappa_n.$$

Strict concavity, together with $\mathcal{B}_n(0) = 0$ and $\mathcal{B}_n(n) = 1$, proves (3.5). □

4 Removing a common deterministic coordinate

The following strict dimension improvement supplies the boundary step in the induction.

Lemma 4.1. *Let $N \ge 2$. For every $0 < \Delta \le N$, there is an ordered $(N+1)$-dimensional pair of gap Δ with tail objective strictly less than $\mathcal{B}_N(\Delta)$.*

Proof. For $\Delta = N$, any strictly interior ordered pair of that gap has $T < 1 = \mathcal{B}_N(N)$. Suppose $0 < \Delta < N$ and take a homogeneous central switch pair $(a\mathbf{1}, b\mathbf{1})$ in dimension N, with switch index j, common mass $f > 0$, and geometric quantity $R > 0$ from Lemma 3.1. Set

$$\rho = \frac{(N-j)ab}{(N-j)ab + j(1-a)(1-b)} \in (0,1).$$

The density

$$g(u) = \rho\mathbb{P}(\mathrm{Bin}(N-1, u) = j-1) + (1-\rho)\mathbb{P}(\mathrm{Bin}(N-1, u) = j)$$

satisfies, by substitution,

$$g(a) = g(b) = \frac{f}{1-R}. \tag{4.1}$$

Adjoin the common parameter ρ. To identify the active threshold, write D_h for the old tail differences and set $D_0 = D_{N+1} = 0$. The new differences satisfy

$$\widetilde{D}_h = (1-\rho)D_h + \rho D_{h-1}, \qquad 1 \le h \le N+1.$$

Exactly the old thresholds $j, j+1$ attain $\mathcal{B}_N(\Delta)$. Since $0 < \rho < 1$, the new threshold $j+1$ still has this value, whereas every other new threshold has a strictly smaller value: its convex combination includes at least one old value below the maximum. Thus $j+1$ is the unique maximizing threshold. For small $\varepsilon > 0$ replace the parameters by

$$p_\varepsilon = \left((a - \varepsilon/(2N))^N, \rho + \varepsilon/2\right), \qquad q_\varepsilon = \left((b + \varepsilon/(2N))^N, \rho - \varepsilon/2\right).$$

Here a superscript on a coordinate denotes repetition. The gap remains Δ and the pair is feasible. By (4.1), the unique active tail has derivative at zero

$$f - \frac{g(a)+g(b)}{2} = f - \frac{f}{1-R} < 0.$$

It remains the maximizing tail for sufficiently small ε. This proves the strict improvement. □

5 The global minimizer

We prove Theorem 1.1 by induction. For $n = 2$, $D_1 + D_2 = \Delta$, so $T \ge \Delta/2 = \mathcal{B}_2(\Delta)$, attained by the complementary homogeneous pair. Fix $n \ge 3$, assume the theorem's minimum-value assertion in dimension $n-1$, and let (p, q) be a global minimizer at a fixed gap $0 < \Delta < n$. Compactness gives its existence. Its value is at most the homogeneous candidate $\mathcal{B}_n(\Delta)$.

By Lemma 2.2 and (3.5), at least three gaps are positive. By Lemma 4.1, no coordinate can be common deterministic. Such a coordinate requires $\Delta \le n-1$; deleting it would give $T \ge \mathcal{B}_{n-1}(\Delta)$, whereas a strict improvement is available. Thus $p_i > 0$ and $q_i < 1$ for every i. Also p cannot be identically one, nor q identically zero. For example, if $p = \mathbf{1}$, then

$$T = 1 - \prod_i q_i \ge 1 - (1 - \Delta/n)^n > \mathcal{B}_n(\Delta),$$

where the strict endpoint comparison was proved above.

Write

$$I = \{i : p_i > q_i\}, \qquad C = \{i : p_i = q_i\}, \qquad r = |I| \ge 3, \quad s = |C|.$$

All common parameters, when present, are strictly interior.

5.1 Stationarity and common coordinates

There are one or two adjacent active thresholds. Choose their convex combination H as in (2.1) so that the first-order optimality conditions hold. Existence follows from the separating hyperplane theorem applied to the feasible tangent cone of the linear constraints and the active gradients. With a multiplier c for the total gap, these conditions read, for $i \in I$,

$$p_i < 1 \Longrightarrow H_i(p) = c, \qquad p_i = 1 \Longrightarrow H_i(p) \le c, \tag{5.1}$$

$$q_i > 0 \Longrightarrow H_i(q) = c, \qquad q_i = 0 \Longrightarrow H_i(q) \le c, \tag{5.2}$$

and, for $i \in C$,

$$H_i(p) = H_i(q) = c + \nu_i, \qquad \nu_i \ge 0. \tag{5.3}$$

The signs follow, for example, by using $H(p) - H(q) - c(\sum_i (p_i - q_i) - \Delta) - \sum_{i \in C} \nu_i (p_i - q_i)$ as the Lagrangian, together with the endpoint constraints.

We first establish, without any nonconstancy assumption, that

$$c > 0. \tag{5.4}$$

Write the supports of S_p, S_q as $[L, n]$ and $[0, U]$. Since $T < 1$, we have $L \le U$. For every changing coordinate i, the derivative of the pure threshold h at p is positive whenever $L + 1 \le h \le n$, and its derivative at q is positive whenever $1 \le h \le U$. This follows from the deletion supports and includes $p_i = 1$ and $q_i = 0$. Every threshold belongs to at least one of these two ranges. Therefore their active convex combination satisfies

$$0 < H_i(p) + H_i(q) \le 2c \qquad (i \in I),$$

where the upper bound follows from (5.1)–(5.2). This proves (5.4).

If there is only one active threshold h and a common coordinate t, stationarity in t gives $\mathbb{P}(S_{p,-i} = h - 1) = \mathbb{P}(S_{q,-i} = h - 1)$. These are positive masses. Indeed, write the original supports as $[L, n]$ and $[0, U]$. Deleting the common interior coordinate gives supports $[L, n-1]$ and $[0, U-1]$. Both masses at $h - 1$ could vanish only if $h \le L$ and $h \ge U + 1$, forcing $L > U$ and hence $T = 1$. Likelihood-ratio ordering then says that the two deletion tails at $h - 1, h$ both maximize their difference. Adjoining any common Bernoulli parameter preserves this maximum. Taking that parameter to zero creates a common deterministic coordinate at the same minimum, a contradiction.

Consequently, if $C \neq \varnothing$, there are two active thresholds $k, k + 1$. For a common coordinate t, put $u_j = \mathbb{P}(S_{p,-i} = j) - \mathbb{P}(S_{q,-i} = j)$. The active tie and (5.3) give

$$t u_{k-1} + (1 - t) u_k = 0, \qquad \lambda u_{k-1} + (1 - \lambda) u_k = 0.$$

The tied count lies in the overlap, so $L \le k \le U$. The deletion supports are $[L, n-1]$ and $[0, U-1]$; the upper deletion mass at k and the lower deletion mass at $k - 1$ are positive. If $u_{k-1} = u_k = 0$, equality would make both adjacent masses positive in both deletion laws. At least one changing coordinate remains, so strict likelihood-ratio ordering would contradict the two equalities. Thus the two adjacent differences cannot both vanish, and hence

$$t = \lambda \in (0, 1) \quad \text{for every common coordinate.} \tag{5.5}$$

In particular all common coordinates agree. Their derivative is exactly the original mass at the tied count. That mass is positive because S_p has support $[L, n]$ and $L \le k \le n - 1$. Thus

$$H_i(p) = H_i(q) = f := \mathbb{P}(S_p = k) = \mathbb{P}(S_q = k) > 0 \quad (i \in C), \qquad f = c + \nu, \quad \nu \ge 0. \tag{5.6}$$

5.2 One changing vector is homogeneous

For $i, j \in I$, (5.1) and (2.3) show

$$p_i \neq p_j \quad \Longrightarrow \quad C_{ij}(p) \ge 0;$$

indeed distinct interior entries have equal gradients, and a larger entry equal to one has no larger gradient. Likewise

$$q_i \neq q_j \quad \Longrightarrow \quad C_{ij}(q) \le 0.$$

Suppose that both $p_i \neq p_j$ and $q_i \neq q_j$. The laws in (2.2) are ordered in likelihood ratio, and at least one of the other changing coordinates increases strictly because $r \ge 3$. The upper deletion

law has positive mass at $k-1$: this follows from a free gradient equal to $c>0$ and its nonnegative adjacent difference. Similarly the lower deletion law has positive mass at k. Their ordered interval supports therefore make both adjacent masses positive in both laws. If A_p, B_p denote the upper deletion masses at $k-1, k$, and A_q, B_q the corresponding lower deletion masses, then the coefficient signs and Lemma 2.1 would give

$$1 \le \frac{A_p}{B_p} < \frac{A_q}{B_q} \le 1,$$

a contradiction. Thus every pair satisfies $p_i = p_j$ or $q_i = q_j$. The elementary consequence is that one of p_I, q_I is constant: two different values in one vector force all entries of the other to agree. Reflection lets us assume

$$p_i = a \quad (i \in I). \tag{5.7}$$

Here reflection means $(p,q) \mapsto (\mathbf{1}-q, \mathbf{1}-p)$; it preserves Δ and T, with D_h sent to D_{n-h+1}.

If $a=1$, then $s>0$, since the all-one vector was excluded. By (5.6) and (5.1), for a common coordinate i and a changing coordinate j,

$$(1-\lambda)C_{ij}(p) = H_i(p) - H_j(p) \ge 0.$$

Consider $h(t_1,\dots,t_r) = \mathbb{P}(\sum_{j=1}^r \mathrm{Ber}(t_j) + \mathrm{Bin}(s,\lambda) = k)$. Its partial derivatives at $t=\mathbf{1}$ equal $C_{ij}(p) \ge 0$. Along the path from q_I to $\mathbf{1}$, the deletion laws are strictly smaller in likelihood ratio than the endpoint deletion law $(r-1) + \mathrm{Bin}(s,\lambda)$. Its two relevant masses are positive because $r \le k \le n-1$, so $k-r$ and $k-r+1$ both lie in $\{0,\dots,s\}$. All partial derivatives are therefore positive in the interior of this path. It follows that $h(\mathbf{1}) > h(q_I)$, contradicting the active mass equality. Hence

$$0 < a < 1. \tag{5.8}$$

5.3 The other changing vector is homogeneous

Fix the common coordinates and set

$$W \sim \mathrm{Bin}(s+1,\lambda), \qquad g_M(t) = \mathbb{P}(\mathrm{Bin}(M,t) + W = k).$$

This notation includes deterministic W when $s=0$ and λ is an endpoint. By (5.7), (5.8), and (5.1),

$$c = g_{r-1}(a) > 0. \tag{5.9}$$

Every positive q_i, $i \in I$, is less than a and is a free coordinate. Its gradient is c.

There are at most two distinct positive values among these q_i. Indeed, suppose three such coordinates have distinct values t_1, t_2, t_3, and let r_j be the masses of the law with those three coordinates deleted and E adjoined. Put $A = r_{k-2} - r_{k-1}$ and $B = r_{k-1} - r_k$. The pair coefficients vanish by (2.3), giving $(1-t_\ell)B + t_\ell A = 0$ for $\ell = 1,2,3$. Subtracting two equations yields $A=B$, and then $A=B=0$. Thus $r_{k-2} = r_{k-1} = r_k$. Their common value equals a free gradient, which is a convex combination of these three masses, and hence equals $c>0$. This contradicts strict log-concavity.

If two positive values $u<v$ occur, the smaller value occurs only once. Otherwise split two u coordinates. The coefficient for a pair u,v is zero, so its deletion law Z_{uv}, with E adjoined, has equal masses at $k-1, k$. The free-gradient identity

$$c = (1-v)\mathbb{P}(Z_{uv} = k) + v\mathbb{P}(Z_{uv} = k-1)$$

shows that both equal masses are $c > 0$. The deletion law for two u coordinates replaces the remaining u by v. Since $0 < u < v < a < 1$, this replacement preserves the support and both adjacent masses stay positive. Strict likelihood-ratio ordering therefore gives $C_{uu} < 0$. The objective $H(p) - H(q)$ then has negative second variation $C_{uu}t^2$ along the split. For a single active threshold this is an immediate contradiction. For two active thresholds keep their tie exactly: an unequal u, v transfer has nonzero first derivative of $G = \mathbb{P}(S_p = k) - \mathbb{P}(S_q = k)$, whereas the equal-coordinate split has zero first derivative. To check nonvanishing, write the coefficient as $\lambda A + (1 - \lambda)B = 0$, where

$$A = \mathbb{P}(R = k - 2) - \mathbb{P}(R = k - 1), \qquad B = \mathbb{P}(R = k - 1) - \mathbb{P}(R = k)$$

and R excludes those two coordinates, without E. If $A - B = 0$, then $A = B = 0$, so the three consecutive masses of R are equal. Their common value is $H_i(q) = c > 0$ for the coordinate with value u: adjoining $\mathrm{Ber}(v)$ and E takes a convex combination of those three masses. Strict log-concavity gives a contradiction. In fact, along the transfer that increases a u coordinate and decreases a v coordinate,

$$\frac{\mathrm{d}G}{\mathrm{d}s} = -(v - u)(A - B) \neq 0.$$

Let v_0 denote the equal-coordinate split direction and w_0 this unequal transfer direction, both acting on q alone. For $q(t, s) = q + tv_0 + sw_0$, we have $G_t(0, 0) = 0$ and $G_s(0, 0) \neq 0$. The implicit-function theorem gives $s(t) = O(t^2)$ with $G(t, s(t)) = 0$. Both directions have coordinate sum zero. Moreover, for $J(t, s) = H(p) - H(q(t, s))$, the free-gradient equalities give $J_t(0, 0) = J_s(0, 0) = 0$. Consequently

$$J(t, s(t)) = J(0, 0) + C_{uu}t^2 + O(t^3) < J(0, 0)$$

for sufficiently small nonzero t. All other thresholds remain strictly below the two tied active thresholds by continuity. On this curve $T = J$, so this is a genuine feasible decrease of the original maximum, not merely of a supporting test. All moves stay strictly between zero and a, contradicting minimality.

Suppose now that the positive entries are one $u < v$ and M copies of v, with b zero entries; thus $r = M + b + 1$. Deleting u gives $c = g_M(v)$, while the zero u, v coefficient gives $g'_M(v) = 0$. Consequently $g_M(v)$ is its maximum. If $M \geq 2$, then $a > v$ gives $g_M(a) < g_M(v)$ and $g'_M(a) < 0$. The mass $g_M(a)$ is positive because the support is fixed on $0 < t < 1$ and $g_M(v) = c > 0$. Corollary 2.3 therefore gives

$$g_{M+b}(a) \leq g_M(a) < g_M(v) = c,$$

contradicting (5.9). If $M = 1$, the function is constant and $b \geq 1$ because $r \geq 3$. The law W has equal positive masses at $k - 1, k$. Adding one $\mathrm{Ber}(a)$ preserves the mass c at k but makes the adjacent difference strictly negative. The second addition therefore gives $g_2(a) < g_1(a) = c$, and further additions cannot increase the mass. Thus there cannot be two positive values.

Finally suppose that there are M copies of a single positive value v and $b \geq 1$ zeros, so $r = M + b$. A positive-coordinate derivative and a zero-coordinate derivative give

$$c = g_{M-1}(v), \qquad g_M(v) \leq c.$$

For $R_0 = \mathrm{Bin}(M - 1, v) + W$, the identity

$$g_M(v) - g_{M-1}(v) = v\{\mathbb{P}(R_0 = k - 1) - \mathbb{P}(R_0 = k)\} \leq 0$$

shows that $\mathbb{P}(R_0 = k - 1) \leq \mathbb{P}(R_0 = k) = c > 0$, since $v > 0$. Adjoin b Bernoulli variables with parameter a and then raise the original $M - 1$ parameters from v to a. Every step can only decrease

the mass at k. For the raising steps, its deletion law dominates R_0: one appended $a > v$ replaces the deleted v, and any remaining appended variables are extra. The decrease is strict. If the initial adjacent masses are unequal, the first addition is strict. If they tie, two additions are strict in aggregate; when $b = 1$, there is a parameter to raise because $r \geq 3$, and that raising step is strict. Hence again $g_{r-1}(a) < c$, a contradiction.

We have proved that $q_I = b\mathbf{1}$ for some $0 \leq b < a$. The case $b = 0$ is excluded either by the all-zero-vector endpoint comparison when $s = 0$, or by reflection of the $a = 1$ argument when $s > 0$. Therefore

$$0 < b < a < 1. \tag{5.10}$$

5.4 No unchanged coordinate remains

Suppose $s > 0$. Recalling (5.5) and (5.10), we have

$$p = (a^r, \lambda^s), \qquad q = (b^r, \lambda^s), \qquad 0 < b < a < 1.$$

Set

$$\begin{aligned} g(t) &= \mathbb{P}(\mathrm{Bin}(r-1,t) + \mathrm{Bin}(s+1,\lambda) = k), \\ h(t) &= \mathbb{P}(\mathrm{Bin}(r-1,t) + \mathrm{Bin}(s,\lambda) = k-1) - \mathbb{P}(\mathrm{Bin}(r-1,t) + \mathrm{Bin}(s,\lambda) = k). \end{aligned}$$

The free gradients and common-coordinate gradients give

$$g(a) = g(b) = c, \qquad \nu = (a-\lambda)h(a) = (b-\lambda)h(b) \geq 0. \tag{5.11}$$

Since $r - 1 \geq 2$, g has a unique mode $u \in (b, a)$. At that point the law $\mathrm{Bin}(r-2,u) + \mathrm{Bin}(s+1,\lambda)$ has equal positive masses at $k-1, k$. All laws compared below have $n-1$ strictly interior Bernoulli parameters, including those obtained by intermediate replacements. They therefore have positive masses at both $k-1$ and k.

If $\lambda \leq b$, then $u > \lambda$. Replacing one λ parameter by u increases this deletion law strictly in likelihood ratio, so $h(u) < 0$ and hence $h(a) < 0$. This makes $(a-\lambda)h(a) < 0$. If $\lambda \geq a$, the reflected argument gives $h(b) > 0$ and $(b-\lambda)h(b) < 0$. In the remaining case $b < \lambda < a$, nonnegativity in (5.11) would force $h(b) \leq 0 \leq h(a)$, impossible because increasing t crosses the adjacent-mass equality only from positive to negative, strictly. Every case contradicts (5.11). Thus $s = 0$.

The global minimizer is therefore fully homogeneous, with strictly interior parameters. Proposition 3.2 identifies it as precisely the central switch or the two reflected central switches. This proves (1.2) and all equality cases for $n \geq 3$, and completes the induction.

6 Product total variation and endpoints

For the threshold event $E_k = \{x : \sum_i x_i \geq k\}$,

$$\mathrm{TV}\left(\bigotimes_i \mathrm{Ber}(p_i), \bigotimes_i \mathrm{Ber}(q_i)\right) \geq D_k(p,q).$$

At a homogeneous pair the product likelihood ratio depends only on the count and increases strictly with it, so product total variation equals the maximizing tail difference. The sharp minimum is therefore $\mathcal{B}_n(\Delta)$, and equality for $n \geq 3$ forces equality in the tail comparison, giving the same classification. At $\Delta = 0$, ordering forces $p = q$, and every such pair minimizes. At $\Delta = n$, feasibility forces $p = \mathbf{1}, q = 0$. The nonunique two-dimensional equality families are proved in Appendix A.

A Two-dimensional equality families

For $x \in [0,1]^2$, write $P_x = \mathrm{Ber}(x_1) \otimes \mathrm{Ber}(x_2)$.

Proposition A.1. *Let $p, q \in [0,1]^2$ satisfy $p \geq q$, and put $d_i = p_i - q_i$, $\Delta = d_1 + d_2$, and $A = p_1 p_2 - q_1 q_2$. Then*

$$T(p,q) = \frac{\Delta}{2} + \left| A - \frac{\Delta}{2} \right|, \tag{A.1}$$

$$\mathrm{TV}(P_p, P_q) = \frac{\Delta}{2} + \frac{1}{2}(|d_1 - A| + |d_2 - A|). \tag{A.2}$$

For $0 < \Delta < 2$, both minima are $\Delta/2$. Tail equality holds exactly when $A = \Delta/2$. Product equality holds exactly for

$$p = (r + d, 1 - r), \qquad q = (r, 1 - d - r), \qquad d = \Delta/2, \quad 0 \leq r \leq 1 - d. \tag{A.3}$$

The product-minimizing family is a proper subset of the tail-minimizing family for every $0 < \Delta < 2$.

Proof. Ordering gives $0 \leq A = d_1 p_2 + d_2 q_1 \leq \Delta$. The two tail differences are $D_1 = \Delta - A$ and $D_2 = A$, proving (A.1). The signed product mass differences at $00, 10, 01, 11$ are

$$A - \Delta, \qquad d_1 - A, \qquad d_2 - A, \qquad A.$$

Taking half the sum of their absolute values proves (A.2). The lower bound $\Delta/2$ in both formulas is attained by the complementary homogeneous pair. The tail equality condition follows immediately. Product equality is equivalent to $d_1 = d_2 = A = d$. Since $d > 0$ and

$$A = (q_1 + d)(q_2 + d) - q_1 q_2 = d(q_1 + q_2 + d),$$

this is equivalent to $q_1 + q_2 = 1 - d$. Writing $q_1 = r$ gives (A.3), with the displayed interval exactly expressing feasibility. Conversely every member of that family attains equality; its homogeneous member is $r = (1 - d)/2$.

For strict containment, choose $0 < \varepsilon < \min\{d, 1 - d\}$, set $d_1 = d + \varepsilon$, $d_2 = d - \varepsilon$, and take $p_i = (1 + d_i)/2$, $q_i = (1 - d_i)/2$. Then $A = d$, so $T(p,q) = d$ but $\mathrm{TV}(P_p, P_q) = d + \varepsilon > d$. At $\Delta = 0$, ordering forces $p = q$, and every such pair minimizes both objectives. At $\Delta = 2$, the only feasible pair is $p = (1,1)$, $q = (0,0)$. □

AI assistance

GPT-6 Astra and Claude Opus 5 assisted with proof discovery, auditing, and editorial suggestions. The author takes full responsibility for the correctness of this work.

References

[1] W. Hoeffding, *On the distribution of the number of successes in independent trials*, Ann. Math. Statist. **27** (1956), no. 3, 713–721. doi:10.1214/aoms/1177728178.

[2] A. Kontorovich, *Minimizing total variation under constraint*, MathOverflow, question 476847, August 13, 2024. `https://mathoverflow.net/q/476847`.

[3] A. Kontorovich, *Poisson binomial conjecture*, MathOverflow, question 477402, August 22, 2024, with answers by T. Steinke (August 22, 2024) and W. Sawin (September 24, 2024). `https://mathoverflow.net/q/477402`.

[4] A. Kontorovich, *On the tensorization of the variational distance*, Electron. Commun. Probab. **30** (2025), paper no. 32, 10 pp. doi:10.1214/25-ECP680.

[5] A. Kontorovich, *TV homogenization inequalities*, preprint (2026), arXiv:2601.04079v3 [math.PR].

[6] J. Pitman, *Probabilistic bounds on the coefficients of polynomials with only real zeros*, J. Combin. Theory Ser. A **77** (1997), no. 2, 279–303. doi:10.1006/jcta.1997.2747.