An Exact-Radius Convex Characterization of Discrete Identity Testing

Zhijie Zheng
Peking University

Abstract

Identity testing asks how many samples are needed to distinguish a known distribution q from every distribution p with $\mathrm{TV}(p, q) > \varepsilon$. Existing instance-dependent bounds can change the radius by constant factors, which can leave unbounded multiplicative gaps in the sample count. We characterize the sample complexity up to universal constant factors for every finite q at the specified radius ε, answering the finite-domain form of Canonne's question on a succinct instance-optimal characterization. The characterization uses an explicit convex program with $O(k)$ variables and second-order-cone constraints, where k is the domain size. The sample threshold of this program is combined with the $1/\varepsilon$ baseline. For rational explicit inputs, the resulting functional is computable within constant factors in polynomial bit time. Finite coarsenings give an exact statistical completion to countable domains.

The proof connects both sides of testing to the same moment program: separating gradients yield tests, while low-cost moments yield hard priors. The lower bound requires converting scalar moment witnesses into probability vectors while preserving the requested radius. We construct normalizations that preserve the relevant excess or deficit. In the deficit-preserving construction, a strict anticoncentration bound gives constant prior mass beyond the radius.

1 Introduction

Identity testing asks how many samples are needed to distinguish a known distribution from all distributions farther than a prescribed total-variation radius. The answer can depend strongly on the reference distribution. A small effective support, a dominant atom, and the amount of mass in a light tail can each change the difficulty of the testing problem. An instance-optimal characterization must capture these features while keeping the requested radius fixed.

Keeping the radius fixed is essential to instance-optimal sample complexity. Valiant and Valiant [VV17] express reference-dependent bounds through a truncated 2/3-quasinorm, with different constant multiples of the radius in the two bounds. Blais, Canonne, and Gur [BCG19] obtain related bounds through an inverse K-functional. Canonne [Can24] explains that radius rescaling can leave an arbitrarily large multiplicative gap in sample complexity, and asks for a succinct functional that characterizes the optimum. Balakrishnan and Wasserman [BW19] and Chhor and Carpentier [CC22] study local minimax separation radii for a given sample size. Constant-factor bounds on that radius need not invert to constant-factor sample complexity at one specified radius.

We give such a characterization for every finite explicit reference distribution. An explicit convex program with $O(k)$ variables and second-order-cone constraints determines the optimal sample count up to universal constant factors at the given radius. The constants apply uniformly over the reference distribution, radius, and domain size. The resulting functional answers the finite-domain form of Canonne's Question 1 and is computable within constant factors in polynomial bit time for rational inputs. Appendix A extends the statistical characterization to countable domains through finite coarsenings.

The characterization connects an instance's geometry to both sides of the testing problem. Moments with sufficiently small objective value produce a prior that obstructs testing; once the objective is large enough, a separating gradient produces a test. Thus the same finite convex description determines when the prescribed discrepancy becomes statistically detectable. The moment constraints retain the exact radius throughout these two constructions.

A boundary instance. Consider a reference with one atom of mass $1-\alpha$ and k equal tail atoms of mass α/k, where $0 < \alpha \le 1/4$. At the radius $\varepsilon = \alpha$, our characterization gives sample complexity $\Theta(k^{2/5}/\alpha)$. The calculation balances a change in the tail means against random fluctuations within the tail; Appendix C.4 derives it directly from the program. At this boundary, the tail mass and its internal structure jointly determine the sample count.

The convex description. Each coordinate is represented by the mean, squared displacement, excess, and deficit of a scalar random variable. These quantities admit an exact convex description with a constant number of variables per coordinate. Two costs then measure how visible a candidate perturbation is in Poisson samples: a triangular cost for changes in the means and a quadratic cost for random squared displacements. The resulting program has two forms. With no large reference atom, all coordinates carry moments and their means sum to one. With a large reference atom, only the tail carries quadratic moments, and the head supplies the remaining mean. This distinction accounts for the dependence imposed by probability normalization.

From moments to tests and lower bounds. Convex separation at an optimizer produces a linear statistic and a weighted quadratic statistic. A third statistic monitors the variance of the linear statistic, giving power even when an alternative has much larger variance than the reference. The lower bound addresses the converse task: realizing a small-cost moment witness as a hard prior

of probability vectors. Independent scalar realizations have tractable observation laws and require normalization of their fluctuating total mass. Our main normalization preserves each coordinate deficit and hence the requested radius. A strict anticoncentration inequality supplies positive prior mass beyond the requested radius, and a fixed amount of confidence amplification converts that mass into a sample lower bound. For small radii, a reservoir formed from light coordinates provides an exact reduction to the large-atom case.

Relation to optimization-based testing. Goldenshluger, Juditsky, and Nemirovski [GJN15] construct tests by convex optimization for hypotheses represented by convex compact parameter sets, and combine these tests to handle more general hypotheses. In our program, convexity arises from the coordinate moments of random alternatives. Explicit normalization then connects these moment witnesses to the set $\mathrm{TV}(p, q) > \varepsilon$, yielding a characterization with $O(k)$ variables and universal constants at the specified radius.

Dang, McKelvie, Valiant, and Wang [DMVW23] develop an optimization framework for semilinear tests. Their Theorem 4 gives a finite-instance Chernoff upper bound and a matching asymptotic error-exponent lower bound under domain subdivision. Our characterization concerns fixed-confidence sample complexity on each original instance, with universal factors at the given radius. The new ingredients are the finite convex moment description and the normalization arguments that establish this equivalence.

Dewasurendra [Dew26] gives an exact-radius characterization within a fixed Zipfian envelope, in a nondegenerate tail-quantile range and for $0 < \varepsilon \le 1/3$, using graded thinning and reservoir normalization. That characterization allows its comparison constants to depend on the envelope. Our characterization uses universal constants for arbitrary finite reference vectors, and the reservoir reduction preserves the radius for these general instances.

Organization. Section 2 defines the programs and states the theorem. Sections 3–4 give the scalar cone description and upper bound. The scalar laws in Section 5 support three lower-bound constructions: a large reference atom in Section 6, a small-radius reservoir reduction in Section 7, and deficit-preserving normalization in Section 9. Section 10 completes the sample-root argument. The appendices treat countable domains, rational computation, and detailed estimates.

2 Statistical conventions and the characterization

Let $q = (q_1, \dots, q_k)$ be a known probability vector on a known finite domain. Samples are independent draws from an unknown probability vector p on the same domain. Zero reference coordinates are allowed. Write

$$\mathrm{TV}(p, q) = \frac{1}{2} \sum_i |p_i - q_i|.$$

Let $N^*(q, \varepsilon)$ be the least fixed sample size permitting type-I error at most $1/3$ under $p = q$ and type-II error at most $1/3$ for every p with $\mathrm{TV}(p, q) > \varepsilon$. We assume $\varepsilon > 0$. The strict alternative is empty exactly when $\varepsilon \ge 1 - \min_i q_i$; in that case $N^* = 0$. This convention includes $k = 1$.

All formulas below concern nonempty strict alternatives. For a random variable $U \in [0, 1]$, define its scalar moments relative to q by

$$r = \mathbb{E}U, \qquad w = \mathbb{E}(U - q)^2, \qquad e = \mathbb{E}(U - q)_+, \qquad d = \mathbb{E}(q - U)_+ = q - r + e.$$

Section 3 gives an exact convex description of their feasible set. Set

$$T_q(r) = \frac{(r-q)^2}{r+q}, \qquad \Phi_q(w) = \frac{w^2}{q^2+w},$$

with value zero when both numerator and denominator vanish, as given by the closed extension.

2.1 Diffuse references

If $\max_i q_i < 1/8$, let

$$A_n(q,\varepsilon) = \min \max \left\{ n \sum_i T_{q_i}(r_i),\ n^2 \sum_i \Phi_{q_i}(w_i) \right\},$$

where the minimum is over scalar moment tuples on every coordinate satisfying

$$\sum_i r_i = 1, \qquad \sum_i d_i \geq \varepsilon.$$

Since $\sum_i r_i = \sum_i q_i$, expected excess and deficit have equal total mass. The last constraint is therefore equivalent to $\sum_i \mathbb{E}|U_i - q_i| \geq 2\varepsilon$. These variables describe scalar moment witnesses. The lower-bound proof constructs normalized probability vectors from these witnesses.

2.2 A large reference atom

If $\max_i q_i \geq 1/8$, relabel a largest atom as $q_0 = h$ and put $\alpha = 1 - h$. For $\alpha > 0$, retain scalar moment tuples only on tail coordinates $i \geq 1$. Require $\beta = \sum_{\text{tail}} r_i \leq 1$ and set $r_0 = 1 - \beta$. Define

$$J_n^+(q,\varepsilon) = \min \max \left\{ n \sum_{i \geq 0} T_{q_i}(r_i),\ n^2 \sum_{i \geq 1} \Phi_{q_i}(w_i) \right\},$$

subject to $\sum_{\text{tail}} e_i \geq \varepsilon$. Coordinate 0 carries only its mean. The companion deficit program J_n^- replaces the excess constraint by $\sum_{\text{tail}} d_i \geq \varepsilon$. It enters the upper-bound proof, while the characterization uses J_n^+.

Let F_n denote A_n in the diffuse case and J_n^+ in the large-atom case, and define

$$R(q,\varepsilon) = \inf\{n \geq \varepsilon^{-1} : F_n(q,\varepsilon) \geq 1\}. \tag{1}$$

For a point mass, set $R(q,\varepsilon) = 1/\varepsilon$ directly.

Theorem 2.1 (Exact-radius sample complexity)**.** *There are universal positive constants c, C such that, for every finite q and every nonempty strict radius ε,*

$$cR(q,\varepsilon) \leq N^*(q,\varepsilon) \leq CR(q,\varepsilon).$$

The convex programs have $O(k)$ variables and constraints and explicit second-order-cone representations. The characterization preserves the exact ε. The constants are independent of q, ε, and k.

3 Exact scalar moment descriptions

For $q > 0$, introduce $z \in [0,1]$ and put $s = w + 2qr - q^2$, $A = e + qz$, and $B = r - e - qz$. The exact scalar moment region is

$$\begin{aligned} qz \le A \le z, \qquad & 0 \le B \le q(1-z), \\ s \ge \frac{A^2}{z} + \frac{B^2}{1-z}, \qquad & s \le qr + e. \end{aligned} \tag{2}$$

The perspective functions are interpreted by their closed extensions at $z = 0, 1$. Interpret z as the probability assigned to $[q, 1]$, with first moment A on that interval and B on $[0, q]$. Conditional Jensen's inequality gives the lower second moment. The maximum conditional second moments are attained at interval endpoints and sum to

$$(q+1)A - qz + qB = qr + e.$$

Mixing distributions with the same interval masses and first moments realizes every intermediate second moment. Thus (2) is exactly the convex hull of $(u, (u-q)^2, (u-q)_+)$ for $u \in [0,1]$. Each lower perspective is a rotated second-order cone. For $q = 0$, the region is $e = r$ and $r^2 \le w \le r$. The case $q = 1$ follows by reflection or directly from the same description.

Both T_q and Φ_q are convex, so each program minimizes a convex objective over a compact moment set. For a nonempty alternative their values are finite, continuous in $n > 0$, and satisfy

$$\begin{aligned} c^2 F_n \le F_{cn} \le c F_n \quad & (0 < c \le 1), \\ C F_n \le F_{Cn} \le C^2 F_n \quad & (C \ge 1). \end{aligned} \tag{3}$$

At positive ε, zero cost is infeasible. Hence the root is finite and positive.

We will also use a universal baseline. On a domain with at least two labels, some vertex of the probability simplex has distance at least $1/2$ from q. If $\varepsilon \le 1/4$, interpolate toward that vertex to obtain $\mathrm{TV}(p, q) = 2\varepsilon$. Then $\mathrm{TV}(p^{\otimes m}, q^{\otimes m}) \le 2m\varepsilon$, proving $N^* = \Omega(1/\varepsilon)$. If $\varepsilon > 1/4$ and the alternative is nonempty, $N^* \ge 1$ gives the same bound with another universal constant. Thus

$$N^*(q, \varepsilon) = \Omega(\varepsilon^{-1}). \tag{4}$$

For a point mass, observing any different label gives the matching $O(1/\varepsilon)$ upper bound.

4 Upper bounds from separating moments

The convex program gives two directions of statistical separation, leading to linear and quadratic statistics. A third statistic controls the variance of the linear statistic under an alternative. Together they give a test with controlled type-I and type-II errors.

The argument applies to the all-coordinate program and to either nonempty head-deleted branch. Let I be the coordinates carrying quadratic moments: all coordinates in the first case, and the tail in the second. The full mean vector r is normalized in both cases. Write

$$M(r) = \sum_i T_{q_i}(r_i), \qquad C_I(w) = \sum_{i \in I} \Phi_{q_i}(w_i), \qquad J = \min_K \max\{nM(r), n^2 C_I(w)\},$$

where K is the relevant convex feasible moment set. At an optimizer (r^*, w^*), convex first-order optimality gives $\theta \in [0,1]$, supported on the active costs, such that every feasible (r, w) satisfies

$$\theta n a \cdot (r - r^*) + (1-\theta) n^2 b \cdot (w - w^*) \ge 0,$$

where

$$a_i = 1 - \frac{4q_i^2}{(r_i^* + q_i)^2}, \qquad b_i = \frac{w_i^*(2q_i^2 + w_i^*)}{(q_i^2 + w_i^*)^2} \quad (i \in I),$$

and $b_i = 0$ outside I. At $q_i = 0$, the corresponding nonzero-cost derivatives are $a_i = b_i = 1$. We have $-3 \le a_i \le 1$, $0 \le b_i \le 1$, and

$$\begin{aligned} n\sum_i q_i a_i^2 \le 9J, \qquad n^2 \sum_i q_i^2 b_i^2 \le 4J, \\ a \cdot (r^* - q) \ge M(r^*), \qquad b \cdot w^* \ge C_I(w^*). \end{aligned} \tag{5}$$

The finite-max subgradient rule and the Fermat rule in the relative affine hull give the separation even on a degenerate moment face. Appendix C.2 verifies the derivative inequalities.

A deterministic alternative in the branch is feasible with $r = p$ and $w_i = (p_i - q_i)^2$. Since θ is supported on the active costs, their weighted sum at the optimizer equals J. Separation and (5) yield

$$\theta n \sum_i a_i(p_i - q_i) + (1-\theta)n^2 \sum_i b_i (p_i - q_i)^2 \ge \theta n M(r^*) + (1-\theta)n^2 C_I(w^*) = J.$$

The left side is a convex combination of the linear and quadratic signals. Consequently at least one of

$$n\sum_i a_i(p_i - q_i) \ge J \quad \text{or} \quad n^2 \sum_i b_i (p_i - q_i)^2 \ge J \tag{6}$$

holds. The all-coordinate program covers every $\mathrm{TV} \ge \varepsilon$ alternative. For the head-deleted program, the identity

$$\mathrm{TV}(p, q) = \max\left\{ \sum_{i \ge 1} (q_i - p_i)_+, \ \sum_{i \ge 1} (p_i - q_i)_+ \right\} \tag{7}$$

places every alternative in at least one branch.

Under independent counts $X_i \sim \mathrm{Poi}(np_i)$, use

$$L = \sum_i a_i (X_i - nq_i), \qquad Q = \sum_i b_i [(X_i - nq_i)^2 - X_i], \qquad V = \sum_i a_i^2 X_i.$$

Reject if $L > J/2$, $Q > J/2$, or $V > AJ$, for a sufficiently large fixed A. The last statistic monitors the variance of the first. Under q,

$$\mathrm{Var}_q\, L \le 9J, \qquad \mathrm{Var}_q\, Q \le 8J, \qquad \mathbb{E}_q V \le 9J.$$

Chebyshev's and Markov's inequalities give null error at most $68/J + 9/A$.

For the quadratic statistic, put $E = \mathbb{E}_p Q = n^2 \sum_i b_i (p_i - q_i)^2$. Direct Poisson moments yield

$$\begin{aligned} \mathrm{Var}_p\, Q &= 2n^2 \sum_i b_i^2 p_i^2 + 4n^3 \sum_i b_i^2 p_i (p_i - q_i)^2 \\ &\le 16J + 4E + 8\sqrt{J}\, E + 4E^{3/2}. \end{aligned}$$

Here $\max_i n b_i q_i \le 2\sqrt{J}$, $b_i \le 1$, and the cubic norm inequality give the second line. If $E \ge J$, the probability $Q \le J/2$ tends uniformly to zero as J grows.

A large linear expectation guarantees power when its variance is controlled. The statistic V handles the remaining alternatives because its expectation equals that variance, while its own variance grows only linearly with the same quantity.

More precisely, let $v = \operatorname{Var}_p L = \mathbb{E}_p V$; then $\operatorname{Var}_p V \le 9v$. If $v \ge 2AJ$, the statistic V detects the increased variance, whereas if $v < 2AJ$ and $\mathbb{E}_p L \ge J$, Chebyshev's inequality bounds the failure probability of the linear test based on L. The respective failure bounds are at most $18/(AJ)$ and $8A/J$. Thus a sufficiently large universal threshold J_{up} gives both errors at most $1/10$. Tightening the constants allows both head-deleted branches to be run with a union bound.

For strict alternatives, omit the deficit branch when $\varepsilon \ge \alpha$, its maximum attainable deficit. Omit the excess branch when $\varepsilon \ge 1 - \min_{i \ge 1} q_i$. If both branches are empty, there are no alternatives. The remaining branches use closed moment constraints at ε, which contain every deterministic alternative in the branch. Finally, simulate Poisson sampling using a fixed block of $O(n+1)$ observations, returning an arbitrary answer on the exponentially unlikely count overflow. This establishes the upper bounds at the requested radius.

5 Scalar laws and Poisson-mixture comparisons

The lower bounds begin with independent scalar laws. Their variances determine how close their Poisson mixtures are to the null; the subsequent sections restore exact probability normalization.

For $q > 0$, a nonnegative scalar law with mean r, deficit d, and excess e satisfies $r = q - d + e$. Its minimum variance at these moments is

$$v = \begin{cases} 4de, & d + e \le q, \\ r^2 d/(q-d), & d + e \ge q. \end{cases} \tag{8}$$

In the first branch the law is supported on $q - (d+e)$ and $q + (d+e)$, with respective probabilities $d/(d+e)$ and $e/(d+e)$. In the second it is supported on zero and $rq/(q-d)$, with probability $(q-d)/q$ assigned to the positive value. Endpoint degeneracies are interpreted directly. Splitting at q, applying conditional Jensen's inequality, and minimizing over the upper-part probability proves the formula; details appear in Appendix C.1. In particular, $v \ge 4de$.

We call the first branch of (8) the *local law* and the second the *sparse law*; at the common boundary either representation applies. Small local fluctuations are controlled through χ^2 divergence, while sparse laws also use the total-variation estimate below.

When $q \le 1/2$ and the original moments come from $[0,1]$, these extremizers also lie in $[0,1]$. The local support is at most $2q$, and the sparse support is at most one because $e \le r(1-q)$. For larger q, the nonnegative extremizer remains available as an intermediate Poisson intensity before probability normalization.

For fixed deficit d, decreasing a mean $r > q$ to q decreases variance, expected squared displacement, triangular mean cost, and positive support. For fixed excess e, increasing a mean $r < q$ to q has the analogous property. The local formulas give these facts directly. In the sparse fixed-e branch, differentiating $r^2(q - r + e)/(r - e)$ gives a nonpositive derivative when $r < 2e$.

If U has mean r and variance v, then

$$\mathrm{TV}(\mathbb{E}\operatorname{Poi}(nU), \operatorname{Poi}(nr)) \le n^2 v. \tag{9}$$

The Poisson kernel's second derivative has ℓ^1 norm at most four; Taylor expansion around the mean cancels the first derivative. If the centered support has maximum absolute value R and $nR^2/r \le c$, the exact identity, with U' an independent copy of U,

$$1 + \chi^2(\mathbb{E}\operatorname{Poi}(nU), \operatorname{Poi}(nr)) = \mathbb{E}\exp\left(\frac{n(U-r)(U'-r)}{r}\right)$$

and cancellation of the first-order term give

$$\chi^2(\mathbb{E}\,\mathrm{Poi}(nU), \mathrm{Poi}(nr)) \le C_c \frac{n^2 v^2}{r^2}. \tag{10}$$

Indeed, for $m \ge 2$, bound $|\mathbb{E}(U-r)^m|$ by $R^{m-2}v$ and sum the exponential series. For independent scalar mixtures, use the product formula for χ-square divergence or sum the total-variation bounds. To compare their mean law to the reference, use

$$\sum_i (\sqrt{r_i} - \sqrt{q_i})^2 \le \sum_i T_{q_i}(r_i). \tag{11}$$

6 A lower bound with a positive reference atom

This section applies whenever a largest atom $q_0 = h$ satisfies $h \ge h_* > 0$, with constants depending only on h_*. We use $h_* = 1/8$. Put $\alpha = 1 - h$. For $n\alpha \ge 1$ and sufficiently small J_n^+, we construct a normalized prior supported entirely on strict TV $> \varepsilon$ alternatives whose Poisson-n mixture is within $1/4$ total variation of the null.

The construction proceeds in three stages. First, scalar projections arrange nonnegative tail mean shifts while leaving a fixed amount of head mass available. Next, we increase the expected tail excess enough to make a strict alternative occur with high probability. Finally, we bound the change in the observation law and use the head coordinate to normalize each draw. The excess constraint is preserved through these operations, so the strictness margin can be checked before normalization.

6.1 Positive means, reflection, and reservoir control

Choose scalar minimum-variance tail laws. For an excess witness, project every $r_i < q_i$ upward to q_i, preserving e_i. Thus every tail mean is at least its reference value. For a deficit witness, first project every $r_i > q_i$ down to q_i, then exchange d_i and e_i and replace r_i by $2q_i - r_i$. The target deficit becomes an excess. Local variances are unchanged. In the sparse case, with $x = d/q \ge y = e/q$ and $x + y \ge 1$,

$$v - v' = \frac{q^2(x-y)(x+y-1)^2}{(1-x)(1-y)} \ge 0.$$

The reflected positive support is no larger, since

$$(1 + x - y)(1 - x) \le (1 + y - x)(1 - y).$$

Every tail atom is at most $\min(h, 1-h) \le 1/2$, so the scalar supports are valid. These operations also decrease expected squared displacement.

Let M be the original summed tail triangular mean cost. With $\Delta_+ = \sum_i (r_i - q_i)_+$ and $\Delta_- = \sum_i (q_i - r_i)_+$,

$$\Delta_+ \le \sqrt{2\alpha M} + M, \qquad \Delta_-^2 \le 2\alpha M.$$

For the negative shifts, Cauchy–Schwarz bounds the square of their sum by the product of their summed triangular cost and their summed denominators; those denominators total at most 2α. For the positive shifts, the same calculation gives

$$\Delta_+^2 \le M \sum_{r_i > q_i} (r_i + q_i) \le M(2\alpha + \Delta_+).$$

Solving this quadratic inequality yields the stated bound on Δ_+.

Under $nM \le j$ and $n\alpha \ge 1$, all projected total tail means are $\alpha + O(\alpha\sqrt{j})$. The corresponding head mean cost is $O_{h_*}(\alpha M + M^2)$, so the total mean objective remains $O_{h_*}(j)$. For small j, the tail mean stays below $1 - h/2$. This also proves the small-objective comparison $J_n^+ \le K_{h_*} J_n^-$.

For the projected laws put $V = \sum_i \mathrm{Var}(U_i)$. Splitting raw moments at $w_i = q_i^2$ and using Cauchy–Schwarz gives

$$V \le \frac{\alpha\sqrt{2j}}{n} + \frac{2j}{n^2}. \tag{12}$$

Hence $nV/\alpha = O(\sqrt{j})$ and $\sqrt{V}/\alpha = O(j^{1/4})$. Coordinates with $q_i = 0$ are replaced by their deterministic means: their contribution to the mean objective is n times their summed mean, and they add no variance.

6.2 Creating a strict excess margin

The total excess has variance at most the total coordinate variance. Increasing its mean by a fixed large multiple of the standard deviation therefore makes failure of the strict excess condition unlikely. We implement this increase in one coordinate when a tail atom has substantial mass, and distribute it proportionally to the reference masses otherwise. This choice keeps the increase in the moment costs controlled in both cases.

Fix a large constant L, then $\eta \le 1/(100L^2)$. If $V = 0$, add an arbitrarily small positive tail mass; the remaining head mass is bounded below by a positive constant and the costs change continuously.

Suppose first that some tail coordinate has $q_j \ge \eta\alpha$. Replace U_j by the deterministic value $q_j + e_j$. This preserves its expected excess and raises its mean by d_j. The bounds $v_j \ge 4d_j e_j \ge 4d_j^2$ and (12) imply $nd_j^2/q_j = O(\sqrt{j})$; when $w_j > q_j^2$, use $d_j \le q_j$ and $nq_j \le \sqrt{2j}$. Thus replacing this coordinate by its deterministic value costs only $O(\sqrt{j})$ in the mean objective and $O(j)$ in the collision objective.

Let V_{rest} be the remaining variance. If it is zero, use an arbitrarily small positive deterministic augmentation that preserves the reserved head mass and cost budgets. Otherwise add $\delta = L\sqrt{V_{\mathrm{rest}}}$ deterministically at coordinate j. The total expected excess rises by δ and its variance is at most V_{rest}. Its probability of being at most ε is therefore at most L^{-2}. The additional objectives are $O(L^2\sqrt{j}/\eta)$ and $O(L^4 j/\eta^2)$.

Now suppose every tail atom satisfies $q_i < \eta\alpha$. Set $\delta = L\sqrt{V}$ and $h_i = q_i\delta/\alpha$ on positive-q coordinates. For a nondegenerate local law, raise only its upper support point by h_i divided by the probability of that point, which is at least one half. Raise deterministic coordinates at or above q_i by h_i. For sparse laws, this scales the positive support point by $1 + h_i/r_i$. Means and expected excesses each increase by h_i.

Writing $s = \delta/\alpha$, the local variances satisfy

$$v_i' \le v_i + 2h_i\sqrt{v_i} + h_i^2,$$

and the sparse variances satisfy $v_i' \le (1+s)^2 v_i$. Since $\sum_i q_i^2 \le \eta\alpha^2$,

$$V' \le [(1+s)^2 + 2L\sqrt{\eta} + L^2\eta]V \le 2V.$$

The positive-part map is one-Lipschitz in the independent-copy variance identity. Thus excess variance is at most V', and the probability of excess at most ε is at most $2/L^2$. Raw moments obey $w_i' \le 2w_i + 4h_i^2$; for sparse laws use $v_i' \le 2v_i$ and the squared mean-shift bound. The new collision objective is $O_L(j + n^2\delta^4/\alpha^2) = O_L(j)$ and the mean objective is $O_L(\sqrt{j})$. All means leave at least a fixed fraction of the head mass available.

6.3 Observation closeness and exact normalization

We compare the augmented scalar mixture with the product Poisson law at the same deterministic means, and then compare that mean law with the null. Write $\beta' = \sum_{i\geq 1} \mathbb{E}U_i$ for the augmented tail mean, and give the independent head count mean $n(1-\beta')$ during these comparisons. We then normalize the draws using the head coordinate.

For the original local coordinates, $w_i = a_i^2 \leq q_i^2$ and $na_i^2/r_i \leq \sqrt{2j}$. The support increase preserves a bounded Taylor exponent. To see this in the diffuse-tail branch, its contribution is controlled by

$$\frac{nh_i^2}{q_i} = \frac{q_i}{\alpha}\frac{n\delta^2}{\alpha} = O_L(\sqrt{j}),$$

using (12); the probability of the upper support point is at least one half, so the new Taylor exponent is bounded by a constant multiple of this term and the original exponent. In the other branch, the augmented coordinate is deterministic and the remaining local supports are unchanged. The raw local moments stay within a fixed multiple of q_i^2, so their summed variance-squared terms in (10) are controlled by the new collision objective. For product laws, one plus the χ-square divergence equals the product of one plus the corresponding coordinate divergences. The resulting product χ-square divergence is $O_L(j)$, which bounds the total-variation distance by $O_L(\sqrt{j})$.

For sparse laws, the size of the deficit determines the relevant objective. Substantial deficits allow the collision cost to control the total-variation mixture bound; smaller deficits force a mean shift through which the mean objective controls the total intensity.

For sparse coordinates, retain their original deficit d_i. If $d_i \geq q_i/4$, their positive probability is at most $3/4$ and after the increase $v_i' \geq q_i^2/3$. Consequently $n^2\sum_i v_i'$ is bounded by a constant times the new collision objective, so (9) sums to $O_L(j)$. If $d_i < q_i/4$, sparsity implies $r_i > 3q_i/2$. The triangular mean objective then bounds $\sum_i nr_i' = O_L(\sqrt{j})$, and $v_i' \leq (r_i')^2/3$. Equation (9) again sums to $O_L(j)$. Coordinates replaced by deterministic values contribute no mixture cost. By (11), the normalized deterministic mean law differs from q by $O_{h_*}(j^{1/4})$ in sample total variation.

Let $S = \sum_{\text{tail}} U_i$. On $S \leq 1$, set $p_0 = 1 - S$ and $p_i = U_i$; on $S > 1$, use any fixed probability vector. Since the mean tail total β' leaves a positive head mass,

$$\mathbb{P}(S > 1) \leq \frac{V'}{(1-\beta')^2}.$$

Changing the independent head mean from $n(1-\beta')$ to $n(1-S)$ costs at most $\sqrt{nV'/(1-\beta')}$ in total variation. Both errors vanish with j. On valid draws, tail excess greater than ε implies $\mathrm{TV}(p,q) > \varepsilon$ by (7). Condition on this event, whose complement has probability at most $2/L^2 + o_j(1)$. Choosing L, then η, then j gives final null-mixture total variation below $1/4$, with every draw from the prior in the strict alternative.

6.4 All radii and the sample root

If $\varepsilon > 2\alpha$, every excess witness has $\beta \geq \varepsilon$ and

$$\sum_{i\geq 1} T_{q_i}(r_i) \geq \frac{(\beta-\alpha)^2}{\beta+\alpha} \geq \frac{\varepsilon}{6}.$$

Thus the root is $O(1/\varepsilon)$ and the baseline gives the lower bound. The tail count gives an $O(1/\varepsilon)$ upper test because its null mass α is below $\varepsilon/2$.

Otherwise $\varepsilon \leq 2\alpha$. When R is less than a sufficiently large constant times $1/\varepsilon$, the baseline handles the lower bound. In the remaining case choose $n = cR$ so that $n\alpha \geq 1$ and J_n^+ is below the

small threshold. The comparison factor in the baseline case is chosen after this reduction factor, ensuring the tail-mass condition in the remaining case. Since the root is then above the baseline, continuity makes the objective equal to one there; scaling gives the small objective required by the construction. The constructed prior gives $\Omega(R)$.

For the upper bound compare branches at $2R$, where $n\alpha \geq 1$ and $J_n^+ \geq 1$. A deficit value below $\min(j_{\text{small}}, 1/K_{h*})$ would contradict the small-objective reflection. Both nonempty branch values are therefore bounded below by a fixed constant. Scaling samples by a further constant reaches the upper-test threshold. This proves the large-atom case at every nonempty strict radius.

7 Diffuse references at small radii: an exact reservoir reduction

Assume $\max_i q_i < 1/8$ and $\varepsilon \leq 1/256$. We will combine a set of coordinates into a head with constant reference mass. Combining them discards their individual deficits, so we choose a set with small deficit per unit mass and replace the lost deficit on the remaining coordinates. The construction must preserve the radius exactly and increase both costs by only constant factors; the final refinement of the head label then transfers the lower bound back to the original domain.

Start with an all-coordinate witness of objective at most j. Mix each scalar law with its point mass at q_i, using a common mixing proportion, so that $\sum_i d_i = \varepsilon$. This preserves the summed mean and decreases both costs by convexity. Replace each scalar law by its minimum-variance extremizer, then project every $r_i > q_i$ down to q_i while preserving d_i. Coordinates with $q_i = 0$ become zero. The resulting means satisfy $r_i \leq q_i$, with the same total deficit and smaller costs.

Let $H = \{i : d_i \geq q_i/8,\ q_i > 0\}$, and let L contain the remaining positive-q coordinates. Then $q(H) \leq 8\varepsilon \leq 1/32$ and $q(L) \geq 31/32$. Order L by increasing d_i/q_i and take the first prefix R with reference mass at least $1/4$. Since every atom is below $1/8$,

$$\frac{1}{4} \leq q(R) < \frac{3}{8}, \qquad d(R) \leq \frac{q(R)}{q(L)} d(L) \leq \frac{1}{2} d(L).$$

The last inequality is strict when $d(L) > 0$. In that case put $K = d(L)/(d(L) - d(R)) < 2$; otherwise put $K = 1$. Collapse R to one head of reference mass $Q_0 = q(R)$. Leave H unchanged. On every coordinate in $L \setminus R$, use the unbiased local law

$$U_i' = q_i \pm 2K d_i.$$

Its amplitude is below $q_i/2$, so its support lies in $[0, 1]$. The total tail deficit is exactly

$$d(H) + K[d(L) - d(R)] = \varepsilon. \tag{13}$$

The expected squared displacement is at most $16 d_i^2 \leq 16 w_i$, so the collision cost grows by at most a factor of 256.

The new tail means are q_i on $L \setminus R$ and the old projected means on H. Give the collapsed head mean

$$r_0' = Q_0 + \Delta, \qquad \Delta = \sum_{i \in H} (q_i - r_i).$$

This lies in $[0, 1]$ and normalizes the full mean. Cauchy–Schwarz gives $\Delta^2 \leq 2M_{\text{original}}$, so the head triangular cost is at most $\Delta^2/Q_0 \leq 8M_{\text{original}}$. Including the tail, the full mean cost is at most nine times its original value.

We have constructed a head-deleted deficit witness at the identical radius, with objective at most $256j$, for a reference whose largest atom lies in $[1/4, 3/8]$. The tail mass is at least $5/8$, so

$n \geq 2$ ensures $n\alpha \geq 1$. Section 6, including its deficit-to-excess reflection, gives a normalized prior on collapsed laws in the strict alternative with sample-mixture total variation below $1/4$.

Lift a collapsed law P to the original domain by setting $p_i = P_0 q_i / Q_0$ on R and $p_i = P_i$ outside R. Its absolute discrepancy inside R is $|P_0 - Q_0|$, hence $\mathrm{TV}(p, q) = \mathrm{TV}(P, Q)$. Lift observations by independently refining each head label according to q_i/Q_0. This fixed Markov kernel applies to the null and every alternative, so data processing preserves sample-mixture closeness. Both the statistical experiment and the radius are therefore preserved by the lift. The lower bound for the sample root follows from (3), (4), and the fixed-sample conversion in Section 10.

8 Strict weighted-Bernoulli anticoncentration

The remaining lower bound normalizes an independent scalar construction while keeping its total deficit. We need positive probability of a deficit strictly above its expectation, including when the expectation is an atom of the distribution.

Lemma 8.1. *Let B_i be independent Bernoulli variables with probabilities $p_i \leq 3/4$, let $a_i > 0$, and suppose $\mathrm{Var}(\sum_i a_i B_i) > 0$. Then*

$$\mathbb{P}\left(\sum_i a_i B_i < \sum_i a_i p_i\right) \geq \frac{1}{152}. \tag{14}$$

Equivalently, a nonconstant positive weighted sum of Bernoulli variables with success probabilities at least $1/4$ strictly exceeds its mean with probability at least $1/152$.

Proof. We first remove weights that are large compared with the remaining standard deviation. Their total success probability stays small, so all removed variables can simultaneously be zero with constant probability. On this event, the removed variables contribute a nonpositive shift to the centered sum. We control the residual sum by its variance or fourth moment to obtain the strict sign event.

Remove deterministic coordinates and order the remaining weights nonincreasingly. Maintain a removed prefix R, its mean $A = \sum_R a_i p_i$, and the variance σ^2 of the remaining centered sum Y. Stop if $\sigma = 0$, $A \geq \sigma$, or the largest remaining weight is at most 4σ; otherwise remove the largest weight and repeat.

At the last removal j, the previous removed mean was less than $\sigma_{\mathrm{old}} < a_j/4$. All previously removed weights are at least a_j, so their success probabilities sum to less than $1/4$. Also $p_j(1-p_j) < 1/16$ and $1 - p_j \geq 1/4$, giving $p_j < 1/4$. Thus the total removed success probability is below $1/2$, and the independent event that all removed variables are zero has probability at least $1/2$. On that event, the full centered sum is $Y - A$. If no coordinates are removed, the same event has probability one.

If $\sigma = 0$, then $A > 0$ and the sum is strictly negative. If $A \geq \sigma > 0$, Cantelli's inequality gives $\mathbb{P}(Y < A) \geq 1/2$. In the remaining case, every residual weight is at most 4σ and independence yields

$$\mathbb{E}Y^4 \leq 19\sigma^4.$$

Indeed, for the remaining coordinates the fourth-cumulant identity is

$$\mathbb{E}Y^4 = 3\sigma^4 + \sum_i a_i^4 p_i(1-p_i)(1 - 6p_i(1-p_i)).$$

The sum on the right is at most $\max_i a_i^2 \sum_i a_i^2 p_i(1-p_i) \leq 16\sigma^4$. Interpolation gives $\mathbb{E}|Y| \geq (\mathbb{E}Y^2)^{3/2}/(\mathbb{E}Y^4)^{1/2} \geq \sigma/\sqrt{19}$. Since $\mathbb{E}Y = 0$, we have $\mathbb{E}Y_- = \mathbb{E}|Y|/2$. Cauchy–Schwarz gives

$\mathbb{P}(Y < 0) \geq 1/76$. In every branch the desired strict event has conditional probability at least $1/76$; multiplying by $1/2$ proves (14). □

The fourth-moment argument is related to Alon, Emek, Feldman, and Tennenholtz [AEFT13]. The removal step establishes the strict form required here, with atoms at the mean allowed.

9 Diffuse references at radii bounded below

Fix $\varepsilon_0 = 1/256$ and assume $\varepsilon \geq \varepsilon_0$. The argument applies to every reference, although the theorem needs it only for diffuse q. Let an all-coordinate moment witness have value at most j, with $n \geq 1$.

As in Section 7, arrange total deficit exactly ε, choose nonnegative minimum-variance scalar laws, and project all means down to $r_i \leq q_i$ while preserving deficits. For $q_i > 1/2$, the intermediate extremizer can exceed one; it still decreases the raw second moment and remains a valid nonnegative Poisson intensity. Probability normalization follows below. Zero-q coordinates become zero.

9.1 A strict alternative event and an indistinguishable product law

A local law has deficit a_i times a Bernoulli variable with success probability $d_i/a_i \geq 1/2$. A sparse law has deficit q_i times a Bernoulli variable with success probability $d_i/q_i \geq 1/2$. Therefore

$$D = \sum_i (q_i - U_i)_+$$

is a positive weighted Bernoulli sum plus a deterministic constant. When $\mathrm{Var}(D) > 0$, Lemma 8.1 gives

$$\mathbb{P}(D > \varepsilon) \geq \gamma, \qquad \gamma = \frac{1}{152}. \tag{15}$$

If $\mathrm{Var}(D) = 0$, every prepared law is deterministic and at most q_i, so $\beta = \sum_i r_i = 1 - \varepsilon$. Cauchy–Schwarz then gives mean cost at least $\varepsilon^2/2$, contradicting $nM \leq j$ when $j < \varepsilon_0^2/2$. Hence (15) applies.

For observation closeness, split local coordinates at $w_i = q_i^2/4$. On the smaller side $r_i \geq q_i/2$, the centered support is at most $2\sqrt{w_i}$, and

$$\frac{n \max |U_i - r_i|^2}{r_i} \leq \frac{8nw_i}{q_i} \leq 9\sqrt{j}.$$

Equation (10) sums to $O(j)$ in product χ-square. Every other local coordinate and every sparse coordinate satisfies $w_i \geq q_i^2/4$; sparse laws even satisfy $w_i \geq q_i^2$ by Jensen and $d_i + e_i \geq q_i$. Thus $w_i \leq 5\Phi_{q_i}(w_i)$, and (9) sums to at most $5j$. The independent product mixture is within $O(\sqrt{j})$ total variation of the product Poisson law at its deterministic mean intensities. By (11), that law is $\sqrt{j}$-close to the null, including when its total intensity is below n.

9.2 Normalization that preserves every deficit

Put $Z = \sum_i U_i$ and $\delta = 1 - \beta$. We have $\delta^2 \leq 2M \leq 2j/n$. Splitting raw moments at q_i^2 gives

$$V = \sum_i \mathrm{Var}(U_i) \leq \frac{\sqrt{2j}}{n} + \frac{2j}{n^2}, \qquad n\mathbb{E}(Z-1)^2 \leq \sqrt{2j} + 4j. \tag{16}$$

Also $\mathrm{Var}(D) \le V$ by the positive-part Lipschitz inequality. Let $E = \sum_i (U_i - q_i)_+$, so that $D - E = 1 - Z$, and put $G = \{D \ge \varepsilon/2,\ E > 0\}$. Since $\mathbb{E}D = \varepsilon$, the event $D < \varepsilon/2$ is controlled by $\mathrm{Var}(D)$. If $D \ge \varepsilon/2$ but $E = 0$, then $1 - Z = D \ge \varepsilon/2$. These two observations give

$$\mathbb{P}(G^c) \le \frac{4V}{\varepsilon^2} + \frac{4\mathbb{E}(Z-1)^2}{\varepsilon^2} \le \frac{C\sqrt{j}}{\varepsilon_0^2}. \tag{17}$$

On G, the clipped mass $\sum_i \min(U_i, q_i)$ equals $1 - D$, while the total excess is E. We keep the first part fixed and rescale only the excess so that it contributes exactly D. Thus define

$$p_i = \min(U_i, q_i) + \frac{D}{E}(U_i - q_i)_+; \tag{18}$$

on G^c, put $p = q$. The sum in (18) is $(1 - D) + D = 1$. Every coordinate below its reference value is unchanged, and every coordinate above its reference value remains above it. Thus p is a probability vector and

$$\mathrm{TV}(p, q) = D \qquad \text{on } G. \tag{19}$$

Preserving the full deficit is what retains the exact radius through normalization.

For nonnegative b, a, c,

$$(\sqrt{b + ca} - \sqrt{b + a})^2 \le a(\sqrt{c} - 1)^2.$$

Summing this shared-background inequality in (18) gives, on G,

$$\sum_i (\sqrt{p_i} - \sqrt{U_i})^2 \le (\sqrt{D} - \sqrt{E})^2 \le \frac{(D-E)^2}{D} \le \frac{2(Z-1)^2}{\varepsilon}.$$

Use the Poisson Hellinger bound to compare the product laws for each draw of the scalar variables, charging at most one unit of total variation outside the normalization event. Averaging these comparisons and applying Cauchy–Schwarz therefore bound the change in the sample mixture by

$$\mathbb{P}(G^c) + \sqrt{\frac{2n\mathbb{E}(Z-1)^2}{\varepsilon}} \le \frac{C\sqrt{j}}{\varepsilon_0^2} + \frac{Cj^{1/4}}{\sqrt{\varepsilon_0}}. \tag{20}$$

Choose the universal j so small that all errors before and during normalization sum to less than $\gamma/4$ and $\mathbb{P}(G^c) < \gamma/2$. Equations (15) and (19) show that the normalized prior puts mass at least $\gamma/2$ on strict TV $> \varepsilon$ alternatives. Every draw from the prior is a probability vector. Section 10 uses the entire prior, including draws inside the radius, and converts its constant far mass into a lower bound by confidence amplification.

10 From Poisson priors to the sample-root characterization

The preceding constructions produce two types of prior. The head and reservoir constructions are supported entirely on strict alternatives. The deficit-preserving construction assigns a fixed positive fraction of its mass to strict alternatives while retaining all other draws. We treat the first type directly and the second by amplifying the success probability of a hypothetical tester.

First consider a prior supported entirely on alternatives whose Poisson-n mixture is within $1/4$ total variation of the null. A fixed-m tester with $m \le n/32$ would induce a Poisson tester using the first m observations whenever available. For $n \ge 32$, the underflow probability is uniformly small, contradicting the sum-of-errors bound from mixture total variation. For $n < 32$, the baseline $N^* \ge 1$ suffices. Hence these priors give $N^* = \Omega(n)$.

More generally, suppose a normalized prior puts mass at least $\rho > 0$ on alternatives and its Poisson-n mixture is within $\rho/2$ total variation of the null. Amplify a hypothetical fixed-m test on a fixed number B of independent blocks until both errors are below $\rho/10$, with B depending only on ρ. Under the null, rejection has probability at most $\rho/10$; averaged over the prior, it has probability at least $\rho(1 - \rho/10)$. The gap is larger than $\rho/2$. Simulating the Bm samples from $\mathrm{Poi}(n)$, with Bm a sufficiently small constant fraction of n, gives a contradiction after absorbing count underflow in the gap. Thus $N^* = \Omega_\rho(n)$. Section 9 has fixed $\rho = \gamma/2$; its bound $\gamma/4 = \rho/2$ can be made strict by taking j smaller.

The values scale according to (3). For a diffuse reference, if R is within a fixed factor of $1/\varepsilon$, the baseline proves the lower bound. Otherwise $n = cR$ has moment value below the small threshold and is at least two. Section 7 applies for $\varepsilon \le \varepsilon_0$, and Section 9 for $\varepsilon \ge \varepsilon_0$. For the upper bound, evaluate at CR so that the value exceeds J_{up}. Section 6.4 gives the corresponding sample-root argument for a large reference atom. This proves Theorem 2.1.

11 Computation and strict endpoints

The scalar region in Section 3 and the quadratic-over-linear costs give $O(k)$-size second-order-cone programs for fixed n. Their sample root is determined by a monotone threshold in one variable. Appendix B provides a rational relative-Slater construction and a direct root bracket. Together with the standard weak-optimization theorem stated there, these yield polynomial bit complexity at fixed objective accuracy for explicit rational q and ε. The computation returns the sample-complexity functional to constant factors while retaining the input radius in every moment constraint.

We first test whether $\varepsilon \ge 1 - \min_i q_i$, which characterizes the empty strict alternative. At such endpoints, a closed moment constraint can remain feasible even though the strict simplex alternative is empty. The lower bounds account for strictness through an explicit excess margin or the strict Bernoulli event. The finite program uses a number of moment and cone variables linear in the domain size, including zero reference entries. Appendix A gives a countable-domain extension as a supremum of these finite programs; the finite-input computational guarantee applies to explicit finite references.

A motivating boundary instance is $q = (1 - \alpha, \alpha/k, \dots, \alpha/k)$ at $\varepsilon = \alpha$, with $0 < \alpha \le 1/4$, where the sample complexity is $\Theta(k^{2/5}/\alpha)$. Appendix C.4 derives this rate from the convex program. It illustrates the dependence on the precise radius at a tail-mass boundary.

A An exact completion to countable domains

Let $X = \{1, 2, \dots\}$ be a known countable domain, including zero-probability labels whenever alternatives may use them. Fix a reference law q and radius $\varepsilon > 0$. Let $N^*(q, \varepsilon)$ be the least fixed sample count with each error at most $1/3$ against the strict alternative $\mathrm{TV}(p, q) > \varepsilon$, and set it to zero when that alternative is empty.

For each m, let π_m partition the domain into singleton labels $1, \dots, m$ and the remaining tail. Let $q^{(m)}$ be the resulting $(m+1)$-dimensional reference. Then

$$N^*(q, \varepsilon) = \sup_m N^*(q^{(m)}, \varepsilon). \tag{21}$$

Lower bound. Given a law P on the coarsened domain, refine its tail label according to q restricted to $\{m+1, m+2, \dots\}$, normalized by its total mass. If this mass is zero, use any fixed

probability law on the nonempty tail. This fixed Markov kernel maps $q^{(m)}$ to q and maps every P to a law p satisfying

$$\mathrm{TV}(p, q) = \mathrm{TV}(P, q^{(m)}).$$

All tail discrepancies have the same sign because they are proportional to q there, or because the reference tail mass is zero. An N-sample tester for q therefore induces one for $q^{(m)}$ by independent refinement of tail observations. Hence $N^*(q, \varepsilon) \geq N^*(q^{(m)}, \varepsilon)$ for every m.

Upper bound. If the right side of (21) is infinite, the lower bound proves the equality. Otherwise let it be the integer N. For each m, choose an N-sample randomized tester for $q^{(m)}$, adding unused observations if necessary. Composing with π_m gives a rejection-probability function $f_m : X^N \to [0, 1]$. The space X^N is countable, so diagonal compactness gives a subsequence $m_j \to \infty$ such that $f_{m_j}(x)$ converges for every sample tuple x. Its limit $f(x)$ defines a randomized N-sample test. The subsequence is selected on sample tuples alone, before fixing an alternative, so the same limit test is used for every alternative.

Under the null, dominated convergence gives $\mathbb{E}_{q^N} f = \lim_j \mathbb{E}_{q^N} f_{m_j} \leq 1/3$. For any fixed p with $\mathrm{TV}(p, q) > \varepsilon$, refinement gives

$$\mathrm{TV}(p^{(m)}, q^{(m)}) \uparrow \mathrm{TV}(p, q).$$

Thus $\mathrm{TV}(p^{(m)}, q^{(m)}) > \varepsilon$ for every sufficiently large m. These tests have $\mathbb{E}_{p^N}(1 - f_m) \leq 1/3$, and dominated convergence gives $\mathbb{E}_{p^N}(1 - f) \leq 1/3$. The conclusion holds for every strict alternative, proving $N^*(q, \varepsilon) \leq N$ at the original radius.

The functional. Write $\Phi_{\mathrm{fin}}(q, \varepsilon)$ for the finite-domain functional, with universal c, C such that $c\Phi_{\mathrm{fin}} \leq N^* \leq C\Phi_{\mathrm{fin}}$, and set it to zero when the finite strict alternative is empty. Define

$$\Phi_{\mathrm{ctbl}}(q, \varepsilon) = \sup_m \Phi_{\mathrm{fin}}(q^{(m)}, \varepsilon).$$

Taking suprema in the finite-domain bounds and using (21) gives

$$c\Phi_{\mathrm{ctbl}} \leq N^*(q, \varepsilon) \leq C\Phi_{\mathrm{ctbl}}.$$

This is a supremum of explicit finite convex programs. Its value can depend on the enumeration within the same universal factors. A permutation-invariant version takes the supremum over all finite partitions: the refinement lower bound holds for every partition, and the nested singleton-tail partitions already give the upper bound. Efficient finite-input computation concerns explicit finite q, as in the computational desideratum of [Can24].

The sample complexity $N^*(q, \varepsilon)$ is finite for every positive radius with a nonempty countable-domain alternative. Choose m with $q(\mathrm{tail}) \leq \varepsilon/2$. Directly,

$$\mathrm{TV}(p, q) - \mathrm{TV}(p^{(m)}, q^{(m)}) \leq \min\{p(\mathrm{tail}), q(\mathrm{tail})\} \leq q(\mathrm{tail}).$$

Every strict ε-far law therefore remains strictly $\varepsilon/2$-far after this fixed coarsening, and a standard finite-domain tester proves finiteness. This observation establishes finiteness; the functional comparison above uses the exact radius throughout.

B Computation for rational explicit inputs

We establish weak optimization in polynomial bit time for explicit finite rational q and rational ε. The numerical guarantee requires quantitative inner and outer bounds on the feasible set. We first construct a scalar moment tuple in its relative interior, then mix it with a sufficiently far simplex vertex to satisfy the radius constraint strictly. These steps provide quantitative slack in both the scalar cones and the radius constraint. Let B be the total binary encoding length of q, ε, and the domain. Handle point masses and empty strict alternatives directly. For the other instances, put $g = 1 - \min_i q_i - \varepsilon > 0$. Each positive coordinate used in a scalar moment cone satisfies $q_i < 1$; define

$$a_i = \frac{q_i(1-q_i)}{8}, \qquad \lambda = \min\left\{\frac{1}{4},\ \min_{0<q_i<1} \frac{a_i}{4}\right\}.$$

The inner minimum is nonempty outside the point-mass case. These rationals have $O(B)$ or polynomial-in-B bit length, and positive g, λ, a_i are at least $2^{-\operatorname{poly}(B)}$.

B.1 A rational scalar witness in the relative interior

Set $r_i^B = (1-\lambda)q_i + \lambda/k$. All means lie strictly between zero and one, and $\sum_i r_i^B = 1$. For positive q_i, put $\mu_i = r_i^B - q_i$ and use the four support values

$$q_i - 2a_i, \qquad q_i - a_i, \qquad q_i + a_i, \qquad q_i + 2a_i.$$

Give the middle two values probability $1/4$ each. Give the extreme negative and positive values probabilities $1/4 - \mu_i/(4a_i)$ and $1/4 + \mu_i/(4a_i)$. They are at least $3/16$, since $|\mu_i| \le \lambda \le a_i/4$. The law has mean r_i^B and two distinct support values on each side of q_i, all strictly between zero and one.

For $q_i = 0$, eliminate the affine equality $e_i = r_i$ and use the reduced cone $r_i^2 \le w_i \le r_i$. At $r_i = r_i^B$, choose $w_i = ((r_i^B)^2 + r_i^B)/2$, strictly between both bounds. This treats zero-reference coordinates in their relative affine space.

For positive q_i, compute z_i, A_i, B_i, s_i from the four-point law. Every linear split-moment inequality is strict. Both conditional laws have positive variance, giving strict slack in the lower second-moment inequality. Quantitatively, support probabilities are at least $3/16$, within-side separation is a_i, and the support values stay a fixed fraction of q_i from zero and of $1 - q_i$ from one. The lower perspective slack is at least ca_i^2. The upper second-moment slack and every nontrivial linear slack are bounded below by a positive polynomial in q_i and $1 - q_i$. All scalar cone and linear slacks are therefore at least $2^{-\operatorname{poly}(B)}$.

The denominators z_i and $1 - z_i$ are bounded away from zero. Zero-q cone slacks are at least a positive polynomial in λ/k. Consequently the product moment set contains a relative Euclidean ball of radius $2^{-\operatorname{poly}(B)}$ centered at the exact tuple. This follows by bounding the gradients of the perspective inequalities in a neighborhood where their denominators remain bounded away from zero. The cone representation gives the same conclusion by placing the perspective epigraph variables strictly between their lower bounds and the available summed second-moment slack.

Impose the full mean equality in the relative affine space $\sum_i r_i = 1$. For a head-deleted program, retain only tail tuples and set $r_0 = 1 - \sum_{\text{tail}} r_i$. The base head mean is $r_0^B > 0$, with polynomial-bit lower bound, so the tail-mean upper bound is strict.

B.2 Strict feasibility at the requested radius

Let P be a simplex vertex on a label of minimum reference mass. Its full deficit is $d_{\max} = 1 - \min_i q_i$. For the head-deleted program, choose such a minimum label in the tail; this is possible since the

pivot is a largest atom. Its tail excess also equals $d_{\max}$. The deterministic tuple is feasible, possibly on scalar faces.

Mix this tuple coordinatewise with the base tuple from Section B.1, giving the base tuple weight $t = g/4$. Means remain normalized, and the relevant total deficit or excess is at least

$$(1-t)d_{\max} \geq d_{\max} - \frac{g}{4} = \varepsilon + \frac{3g}{4},$$

since $d_{\max} \leq 1$. Thus the exact radius inequality has margin at least $3g/4$.

For a convex set containing a ball $B(x, r)$ and a feasible point y, the mixture $tx + (1-t)y$ is the center of a ball of radius tr contained in that set and in the same relative affine space: mix each point in $B(x, r)$ with y. The scalar interior radius remains at least $2^{-\operatorname{poly}(B)}$. Intersecting with the radius halfspace loses another polynomial-bit factor because its coefficients have polynomial magnitude. This gives a rational relative-Slater witness with bit length and logarithmic inverse radius polynomial in B.

For the companion deficit branch, use the vertex at the head, whose tail deficit is α. Whenever $\varepsilon < \alpha$, the same argument applies with the rational gap $\alpha - \varepsilon$. Empty branches are omitted before optimization.

B.3 The objective epigraph and weak optimization

For rational sample parameter n, introduce a cost epigraph variable for each T_q and Φ_q, and a common objective u. At the strict moment witness, set every coordinate cost epigraph to 2 and choose

$$u = 4k(n + n^2) + 1.$$

All objective epigraph inequalities are strict. Scalar moments lie in $[0, 1]$. Cost epigraphs can be restricted to a known bounded box containing both an optimizer and the strict witness, and u to twice the displayed bound, $2[4k(n + n^2) + 1]$, leaving the witness strictly inside the box. The logarithms of the outer bounding radius and inverse relative-interior radius are polynomial in $B + \log(1 + n)$. Zero-q affine equalities have already been eliminated; rational equalities such as $\sum_i r_i = 1$ are treated in their explicit affine hull.

Standard rational weak convex optimization with a separation oracle, or weak second-order-cone optimization under these inner and outer radius bounds [BTN01, LVBL98], evaluates the optimal objective to additive accuracy η in time polynomial in

$$B + \log(1 + n) + \log(1/\eta).$$

The separation predicates are the explicit linear and second-order-cone constraints. Weak optimization supplies numerical objective accuracy even when the optimum lies on a cone boundary; the interior witness maintains feasibility at the exact input radius.

B.4 Computing the root within constant factors

Use objective accuracy $\eta = 1/8$. Start from $n = 1/\varepsilon$ and repeatedly double n until an estimate of F_n is at least one. When the procedure stops, the true value is at least $7/8$; at the previous point it was below $9/8$. The scaling inequalities place this n between universal multiples of the exact root: for example,

$$\frac{7}{8}R \leq n \leq \frac{9}{4}R,$$

with the baseline-endpoint case interpreted directly.

A direct optimization bracket bounds the number of iterations. Every feasible witness satisfies $\sum_i \sqrt{w_i} \geq \varepsilon$, hence $W = \sum_i w_i \geq \varepsilon^2/k$. Since $U_i \in [0,1]$ and the summed mean is at most one,

$$W = \sum_i \mathbb{E}(U_i - q_i)^2 \leq \sum_i (r_i + q_i^2) \leq 2.$$

Cauchy–Schwarz then gives

$$C = \sum_i \frac{w_i^2}{q_i^2 + w_i} \geq \frac{W^2}{\sum_i q_i^2 + W} \geq \frac{\varepsilon^4}{3k^2}.$$

Thus $F_n \geq n^2\varepsilon^4/(3k^2)$ and $R \leq \sqrt{3}k/\varepsilon^2$, directly from the program. Doubling takes $O(\log k + \log(1/\varepsilon) + 1)$ iterations, with polynomial-bit sample parameters.

The explicit cone representation and standard weak-optimization theorem therefore compute the sample-complexity functional within constant factors in Turing-polynomial time. This guarantee concerns the functional; implementing the exact-dual statistical test in polynomial bit time is a further computational question.

C Details of estimates and constant choices

C.1 Scalar minimization and projections

Put $\pi = \mathbb{P}(U > q)$, with a consistent allocation of any atom at q. Conditional Jensen's inequality gives

$$\mathbb{E}(U - q)^2 \geq \frac{e^2}{\pi} + \frac{d^2}{1 - \pi}, \qquad 0 \leq \pi \leq 1 - \frac{d}{q}.$$

The upper constraint on π enforces nonnegativity of the lower conditional mean. If $d = 0$, the deterministic law $U = q + e$ attains variance zero, including when $e > q$. For $d > 0$, the unconstrained minimizer is $\pi = e/(d + e)$, feasible exactly when $d + e \leq q$. Subtracting $(e - d)^2$ gives variance $4de$ in this case. Otherwise the constraint binds, giving lower support zero, upper support $rq/(q - d)$, and variance $r^2 d/(q - d)$. At $d = q$, any finite-mean law has $r = e = 0$, so the extremizer is the deterministic zero variable. At $d = e = 0$, use the deterministic variable q.

For the sparse fixed-excess projection, write

$$f(r) = \frac{r^2(q - r + e)}{r - e}.$$

For $e > 0$, its derivative, after removing the positive factor $r/(r - e)^2$, has numerator

$$e^2 \left[\left(\frac{r}{e} - 2 \right) \frac{q}{e} - 2 \left(\frac{r}{e} - 1 \right)^2 \right].$$

The sparse branch has $r < 2e$, so this numerator is nonpositive. Increasing $r < q$ to q while preserving e decreases the variance, $(r - q)^2$, and the upper support $rq/(r - e)$. The formulas agree continuously at a branch crossing. The fixed-deficit projection follows from the monotonicity of $4d(d + r - q)$ and $r^2 d/(q - d)$. Subtracting these rational expressions gives the reflection identity in Section 6.1; deterministic endpoints are handled directly.

C.2 Derivative inequalities for the upper certificate

For $q > 0$, put $t = r/q$. Then

$$T_q'(r) = \frac{(t-1)(t+3)}{(t+1)^2}, \qquad T_q(r) = q\frac{(t-1)^2}{t+1}.$$

Consequently

$$\frac{q[T_q'(r)]^2}{T_q(r)} = \frac{(t+3)^2}{(t+1)^3} \le 9, \qquad T_q'(r)(r-q) \ge T_q(r).$$

For $x = w/q^2$,

$$\Phi_q'(w) = \frac{x(x+2)}{(1+x)^2}, \qquad \frac{q^2[\Phi_q'(w)]^2}{\Phi_q(w)} = \frac{(x+2)^2}{(x+1)^3} \le 4, \qquad \Phi_q'(w)w \ge \Phi_q(w).$$

At $q = 0$, the costs are linear and the stated derivative conventions apply.

For the alternative quadratic variance, let $\delta_i = p_i - q_i$ and $E = n^2 \sum_i b_i \delta_i^2$. The first term satisfies

$$2n^2 \sum_i b_i^2 p_i^2 \le 4n^2 \sum_i b_i^2 q_i^2 + 4n^2 \sum_i b_i^2 \delta_i^2 \le 16J + 4E.$$

In the second term, $p_i \le q_i + |\delta_i|$ and $\max_i n b_i q_i \le 2\sqrt{J}$ imply

$$4n^3 \sum_i b_i^2 q_i \delta_i^2 \le 8\sqrt{J}\, E.$$

Writing $t_i = n\sqrt{b_i}|\delta_i|$ gives

$$4n^3 \sum_i b_i^2 |\delta_i|^3 = 4 \sum_i \sqrt{b_i}\, t_i^3 \le 4 \left(\sum_i t_i^2 \right)^{3/2} = 4E^{3/2}.$$

The Poisson variance formula follows by expanding factorial moments $\mathbb{E}(X)_\ell = (np)^\ell$ for $\ell \le 4$.

C.3 Summed moments and strict-margin estimates

For coordinates of total reference mass α, split at $w_i = q_i^2$. On the lower part, $\Phi_{q_i}(w_i) \ge w_i^2/(2q_i^2)$, and Cauchy–Schwarz gives

$$\sum_{w_i \le q_i^2} w_i \le \sqrt{\sum_i q_i^2} \sqrt{2 \sum_i \Phi_{q_i}(w_i)} \le \alpha\sqrt{2C}.$$

On the upper part, $\Phi_{q_i}(w_i) \ge w_i/2$. Hence $\sum_i w_i \le \alpha\sqrt{2C} + 2C$, proving the variance bounds used above.

For a local-law increment, let $\pi \ge 1/2$ be the probability of the upper support point. The increment is $(h/\pi)B$ on the same Bernoulli variable B. Its variance is at most h^2, so

$$v' \le v + 2h\sqrt{v} + h^2.$$

For a sparse increment, $U' = (1 + h/r)U$, so $v' = v(1 + h/r)^2$. If $h/r \le s \le 0.1$, then $v' \le 2v$, while $(r + h - q)^2 \le 2(r-q)^2 + 2h^2$; thus $w' \le 2w + 2h^2$. Locally, direct squaring and $\pi \ge 1/2$ give $w' \le 2w + 4h^2$.

The collision-cost estimates use

$$\Phi_q(a+b) \le 2\Phi_q(a) + 2\Phi_q(b), \qquad \Phi_q(ca) \le c^2\Phi_q(a) \quad (c \ge 1).$$

For the first inequality, use $\Phi_q(a+b) \le 2a^2/(q^2+a+b) + 2b^2/(q^2+a+b)$. With $h_i = q_i\delta/\alpha$,

$$\sum_i \Phi_{q_i}(h_i^2) \le \sum_i \frac{h_i^4}{q_i^2} = \frac{\delta^4}{\alpha^4}\sum_i q_i^2 \le \frac{\delta^4}{\alpha^2}.$$

For the deterministic heavy-coordinate increment, use its mass lower bound $q_j \ge \eta\alpha$ in place of α. For $r \ge q > 0$ and $h \ge 0$, the precise triangular bound is $T_q(r+h) \le 2T_q(r) + h^2/q$, absorbing the linear cross term into twice the old cost. Bound the head cost by the squared total shift divided by its fixed positive reference mass.

Finally, a one-Lipschitz function f and an independent copy U' satisfy

$$\mathrm{Var}(f(U)) = \frac{1}{2}\mathbb{E}[f(U) - f(U')]^2 \le \frac{1}{2}\mathbb{E}(U-U')^2 = \mathrm{Var}(U).$$

This proves the excess and deficit variance claims, including degenerate laws.

C.4 A tail-mass boundary example

Let the reference have head mass $1-\alpha$ and k equal tail masses $a = \alpha/k$, where $0 < \alpha \le 1/4$, and set $\varepsilon = \alpha$. Averaging a feasible witness over tail permutations preserves feasibility and cannot increase either convex cost. Mixing its scalar laws with the reference point masses then reduces the total excess to exactly α without increasing the objective. We may therefore take the same moments on every tail coordinate, with $e = a$. Write $d = a(1-t)$, so $r = a(1+t)$ and $0 < t \le 1$. The minimum-variance law from (8) is supported on zero and $a(1+t)/t$, with positive-support probability t. Its expected squared displacement is

$$w = a^2(t^{-1} + 1 - t).$$

Feasibility in $[0,1]$ is equivalent to $t \ge a/(1-a)$. The head mean is $1 - \alpha(1+t) \ge 1/2$. Uniformly on this interval, the full triangular mean cost is $\Theta(\alpha t^2)$, while

$$\sum_{i \ge 1} \Phi_a(w) = \Theta\left(\frac{\alpha^2}{kt}\right).$$

For the mean cost, the tail contributes exactly

$$kT_a(a(1+t)) = \frac{\alpha t^2}{2+t},$$

and the head contributes $(\alpha t)^2/[2(1-\alpha) - \alpha t]$, at most a constant times the tail contribution in the stated range. For the collision cost, $a^2/t \le w \le 2a^2/t$ and $w \ge a^2$ imply $w/2 \le \Phi_a(w) \le w$. Summing over the k tail coordinates gives the second estimate. Consequently

$$J_n^+ = \Theta\left(\min_{a/(1-a) \le t \le 1} \max\left\{n\alpha t^2, \frac{n^2\alpha^2}{kt}\right\}\right).$$

At $n = k^{2/5}/\alpha$, the choice $t = k^{-1/5}$ is feasible and makes both displayed terms one. Every $t \ge k^{-1/5}$ makes the first term at least one, and every $t \le k^{-1/5}$ makes the second at least one. Thus the program has constant value at this sample scale. Its scaling inequalities, the baseline $n \ge 1/\alpha$, and Theorem 2.1 give $N^*(q,\alpha) = \Theta(k^{2/5}/\alpha)$.

C.5 Ordering the universal constants

Choose constants in the following order. Fix $h_* = 1/8$, $\varepsilon_0 = 1/256$, and $\gamma = 1/152$. In the upper test, take $A = 900$ and $J_{\mathrm{up}} = 10^8$, increasing the latter fixed threshold if needed when running two branches. In the head lower bound choose $L \geq 100$, then $\eta \leq 1/(100L^2)$, then a small objective threshold such that $s \leq 0.1$, all modified means retain at least $h_*/3$ head mass, and observation and conditioning errors sum to less than $1/4$. Every error tends to zero as this threshold tends to zero except $2/L^2$, which was fixed first. Reflection costs a fixed factor; divide the small threshold by that factor and by 256 for the reservoir reduction.

In the constant-radius lower bound, choose the threshold so that the bad-event probability is below $\gamma/2$ and total sample-mixture error below $\gamma/8$. The bounds are fixed multiples of $\sqrt{j}$ and $j^{1/4}$ once ε_0 is fixed. For prior far mass $\rho = \gamma/2$, majority voting over

$$B = \lceil 18 \log(100/\rho) \rceil$$

repetitions of a putative 2/3-accurate test has error below $\rho/100$. Run these repetitions within a fixed fraction of the Poisson sample budget. The probability of a Poisson count below the required sample budget is exponentially small above a fixed numerical budget; the one-sample baseline handles smaller budgets. This establishes universal fixed-sample lower constants and a consistent choice of all constants in Theorem 2.1.

AI methodology

OpenAI Codex assisted with organizing the exposition, revising the English, developing and checking proof details, deriving the boundary example, reconstructing and checking the LaTeX, and verifying selected primary references. The author retains responsibility for all claims and the final manuscript.

References

[AEFT13] Noga Alon, Yuval Emek, Michal Feldman, and Moshe Tennenholtz. Adversarial leakage in games. *SIAM Journal on Discrete Mathematics*, 27:363–385, 2013. https://web.math.princeton.edu/~nalon/PDFS/leakage7.pdf.

[BW19] Sivaraman Balakrishnan and Larry Wasserman. Hypothesis testing for densities and high-dimensional multinomials: Sharp local minimax rates. *The Annals of Statistics*, 47(4):1893–1927, 2019. doi:10.1214/18-AOS1729.

[BTN01] Aharon Ben-Tal and Arkadi Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. SIAM, 2001. doi:10.1137/1.9780898718829.

[BCG19] Eric Blais, Clément L. Canonne, and Tom Gur. Distribution testing lower bounds via reductions from communication complexity. *ACM Transactions on Computation Theory*, 11(2), article 6, 2019. Conference version: CCC 2017. https://cs.uwaterloo.ca/~eblais/assets/pdf/distribution_testing.pdf.

[Can24] Clément L. Canonne. Open problem: Tight characterization of instance-optimal identity testing. In *Proceedings of the Thirty Seventh Conference on Learning Theory*, PMLR 247:5312–5316, 2024. https://proceedings.mlr.press/v247/canonne24a.html.

[CC22] Julien Chhor and Alexandra Carpentier. Sharp local minimax rates for goodness-of-fit testing in multivariate binomial and Poisson families and in multinomials. *Mathematical Statistics and Learning*, 5(1–2):1–54, 2022. https://ems.press/content/serial-article-files/28933.

[DMVW23] Trung Dang, Walter McKelvie, Paul Valiant, and Hongao Wang. Improving Pearson's chi-squared test: hypothesis testing of distributions—optimally. Preprint, arXiv:2310.09408, 2023. https://arxiv.org/abs/2310.09408.

[Dew26] Pahan Dewasurendra. Closing the harmonic gap: Exact identity testing for Zipfian distributions. Public preprint, 2026. https://andotheror.github.io/pdf/closing-the-harmonic-gap-exact-identity-testing-for-zipfian-distributi.pdf.

[GJN15] Alexander Goldenshluger, Anatoli Juditsky, and Arkadi Nemirovski. Hypothesis testing by convex optimization. *Electronic Journal of Statistics*, 9(2):1645–1712, 2015. doi:10.1214/15-EJS1054.

[LVBL98] Miguel Sousa Lobo, Lieven Vandenberghe, Stephen Boyd, and Hervé Lebret. Applications of second-order cone programming. *Linear Algebra and its Applications*, 284:193–228, 1998. https://stanford.edu/~boyd/papers/socp.html.

[VV17] Gregory Valiant and Paul Valiant. An automatic inequality prover and instance optimal identity testing. *SIAM Journal on Computing*, 46(1):429–455, 2017. Conference version: FOCS 2014. https://www.cs.purdue.edu/homes/pvaliant/instanceoptimalidentitysiam.pdf.