
The Quadratic Cost of Replicable Distribution Estimation

Pahan Dewasurendra
Johns Hopkins University

Abstract

Estimating a distribution on k symbols to expected total-variation error α needs $\Theta(k/\alpha^2)$ samples. Under algorithmic replicability, the best known upper bound is $O(k^2 \log(1/\rho)/(\alpha^2 \rho^2))$. Whether the quadratic alphabet dependence is necessary was left open by Bun et al. We prove that it is. Every ρ -replicable estimator with expected error at most α requires

$$\Omega\left(\frac{k^2}{\alpha^2 \rho^2}\right)$$

samples, for all sufficiently small universal α and ρ . The lower bound holds for arbitrary randomized estimators and matches every polynomial dependence in the known upper bound. Our proof introduces an average-distortion partition principle. Fixing the public randomness turns a replicable estimator into canonical output cells. Expected accuracy does not bound every cell, so bounded-diameter partition arguments do not apply. Instead, we show that cells with low average distortion cover constant measure, that each such cell has exponentially small measure, and that Gaussian-profile isoperimetry forces total boundary $\Omega(\sqrt{k})$. The hard instances form a smooth toroidal family of categorical distributions. Nearby parameters generate statistically indistinguishable samples at scale $\sqrt{k}/(\alpha\sqrt{m})$. Thickening all canonical cells to this scale converts their boundary into nonreplicability, yielding the result by coarea.

1 Introduction

Suppose two laboratories run the same randomized analysis with the same published random seed, but collect independent data from the same population. Algorithmic replicability asks that they return exactly the

same answer with high probability (Impagliazzo et al., 2022). Formally, an estimator A using m samples is ρ -replicable if, for every data distribution p ,

$$\Pr_{r, S, S' \sim p^m} \{A(S; r) = A(S'; r)\} \geq 1 - \rho. \quad (1)$$

The random string r is shared and the samples S, S' are independent. Exact agreement makes this requirement much stronger than ordinary concentration.

We study the most basic nonparametric estimation problem. Given samples from an arbitrary distribution p on $[k]$, output $\hat{p}$ satisfying

$$\mathbb{E}_{r, S} d_{\text{TV}}(\hat{p}, p) \leq \alpha. \quad (2)$$

Without stability, the minimax sample complexity is $\Theta(k/\alpha^2)$ (Devroye and Lugosi, 2001). Bun et al. (2023) obtained the first replicable estimator by converting a private histogram estimator (Diakonikolas et al., 2015; Acharya et al., 2021). Their bound is

$$O\left(\frac{k^2 \log(1/\rho)}{\alpha^2 \rho^2}\right), \quad (3)$$

after a constant rescaling of α . They explicitly left open whether linear dependence on k is possible.

Our main result closes this question.

Theorem 1 (Main result). *There are universal constants $c, \alpha_0, \rho_0 > 0$ and $k_0 \in \mathbb{N}$ such that the following holds. For $k \geq k_0$, $0 < \alpha \leq \alpha_0$, and $0 < \rho \leq \rho_0$, every estimator satisfying (1) and (2) uses*

$$m \geq c \frac{k^2}{\alpha^2 \rho^2} \quad (4)$$

i.i.d. samples in the worst case.

Thus the known upper bound is optimal in all three polynomial dependences and is separated only by $\log(1/\rho)$. At constant α and ρ , replicability changes the alphabet dependence from linear to quadratic. The lower bound is information theoretic. It allows unbounded computation, a continuous output space, and arbitrary internal randomness.

Why existing lower bounds do not settle this problem. High-dimensional replicable mean estimation is governed by low-surface-area partitions (Hopkins et al., 2024). That work also treats coordinate sampling, but its lower bounds require high-probability Euclidean or simultaneous coordinatewise accuracy. Our categorical experiment is a uniform-coordinate experiment whose expected total-variation loss controls only average ℓ_1 error across coordinates. Converting this guarantee to their ℓ_2 or ℓ_∞ criteria loses the factor needed to distinguish k from k^2 . There is a second obstacle. Equation (2) controls expected loss, not a high-probability correctness event. Consequently, a fixed public seed can induce cells with long, inaccurate tendrils, so no uniform diameter bound is available.

Our approach. We prove a reusable partition principle based on *average distortion*. For a fixed public seed, let F_y contain parameters on which y is output with probability above $3/4$. Replicability makes these canonical cells cover almost all parameter space. Rather than truncate every cell to an accuracy ball, we retain the cells whose average loss is $O(\alpha)$. Expected accuracy ensures that they still cover constant measure. If every low-loss basin has exponentially small measure, Gaussian-profile isoperimetry gives total perimeter $\Omega(\sqrt{d})$ in dimension d .

The hard family is parameterized by the flat torus $\mathbb{T}^d$ and has $4d$ symbols. The torus avoids the external boundary term that defeats a cube argument. Its coordinates modulate sine and cosine pairs in a categorical distribution. An arbitrary accurate output determines an angular center, and concentration shows that its low-loss basin has measure $\exp(-\Omega(d))$. Meanwhile, m samples cannot distinguish parameters within distance $R = \Theta(\sqrt{d}/(\alpha\sqrt{m}))$. The R -thickenings of distinct canonical cells remain disjoint. Every thickened boundary is noncanonical and therefore incurs constant disagreement. Coarea gives $R\sqrt{d} = O(\rho)$, which rearranges to (4).

Relation to distribution testing. Recent work obtains sharper and sometimes nearly linear $1/\rho$ costs for uniformity and closeness testing (Liu and Ye, 2024; Diakonikolas et al., 2025; Aamand et al., 2025). Those tasks return one bit and need not reconstruct the distribution. Theorem 1 shows a contrasting phenomenon for estimation: even a constant-error histogram has a quadratic joint cost in alphabet size and $1/\rho$.

2 Related work

Impagliazzo et al. (2022) introduced the shared-randomness definition in (1), under the original name

reproducibility, and proved the sharp $\Theta(1/(\alpha^2\rho^2))$ scalar statistical-query rate at constant confidence. Bun et al. (2023) connected replicability with differential privacy and perfect generalization. Their generic conversion produced the histogram upper bound (3) and the open question resolved here. List and certificate replicability quantify different relaxations of exact shared-seed agreement (Dixon et al., 2023). Our lower bound concerns the original pointwise definition and therefore does not assert a lower bound for those relaxations.

The closest geometric work is Hopkins et al. (2024). It relates replicable high-dimensional mean estimation to isoperimetric tilings, proves optimal Euclidean mean-estimation rates, and treats the N -coin problem under vector and coordinate sampling. Its estimator-to-partition reduction uses cells with a Euclidean radius guarantee, obtained from high-probability correctness. Our Proposition 2 has a different input and conclusion. It starts from expected loss, allows individual canonical cells to be inaccurate on subsets of their volume, and combines an average-distortion budget with small-basin isoperimetry. The categorical experiment (13) can be viewed as uniform coordinate sampling. The key difference from their N -coin lower bounds is the guarantee: expected total variation permits a canonical output to be inaccurate on a subset of coordinates and parameters, while their high-probability ℓ_2 and ℓ_∞ guarantees produce bounded-radius cells.

Replicable distribution testing was developed for uniformity, identity, closeness, and independence (Liu and Ye, 2024; Diakonikolas et al., 2025). The later structural theory of Aamand et al. (2025) shows that binary testers can be reduced to canonical random-threshold form and develops chaining lower bounds. These testing results can beat quadratic $1/\rho$ overhead in some regimes. They do not estimate an arbitrary histogram or bound its expected total-variation loss. To our knowledge, no work after Bun et al. (2023) gave a lower bound with superlinear alphabet dependence for that problem.

3 Preliminaries

Write Δ_k for the simplex on $[k]$ and $d_{\text{TV}}(p, q) = \frac{1}{2} \sum_x |p(x) - q(x)|$. An estimator is a randomized map $A : [k]^m \times \mathcal{R} \rightarrow \Delta_k$. We impose no restriction on $\mathcal{R}$ or on the range of A . Once r is fixed, however, its range has size at most k^m .

For a compact Riemannian manifold M with probability measure μ , let

$$E_t = \{x \in M : \text{dist}(x, E) < t\}$$

and define the outer Minkowski perimeter

$$\text{Per}_\mu(E) = \liminf_{t \downarrow 0} \frac{\mu(E_t) - \mu(E)}{t}.$$

We use the standard coarea identity for distance functions. In particular, for measurable E and $R > 0$,

$$\int_0^R \text{Per}_\mu(E_t) dt \leq \mu(E_R \setminus E), \quad (5)$$

with equality for regular sets and almost every level. The inequality is all that we need.

4 An average-distortion partition principle

We first isolate the proof mechanism. Let $\{P_\theta : \theta \in M\}$ be a family of data distributions and let $\ell(y, \theta) \geq 0$ be the loss of output y . Assume that $\theta \mapsto P_\theta^m$ is continuous in total variation. The following proposition turns three geometric and statistical properties into a replicability lower bound.

Proposition 2 (Distortion to boundary). *Fix $0 < \eta < 1/4$, $K \geq 32$, and $R, \lambda > 0$. Suppose:*

1. $d_{\text{TV}}(P_\theta^m, P_\phi^m) \leq \eta$ whenever $\text{dist}(\theta, \phi) \leq R$.
2. For every output y ,

$$\mu\{\theta : \ell(y, \theta) \leq 2K\alpha\} \leq s. \quad (6)$$

3. $\text{Per}_\mu(E) \geq \lambda\mu(E)$ whenever $\mu(E) \leq 2s$.

If A has expected loss at most α for every θ and is ρ -replicable for every θ , then

$$R\lambda \leq C\rho \quad (7)$$

for a universal C , provided ρ is below a universal constant.

The proof is included because the expected-loss step is central. Let

$$L_r = \int_M \mathbb{E}_{S \sim P_\theta^m} \ell(A(S; r), \theta) d\mu(\theta)$$

and let D_r be the analogous integrated probability that two runs disagree. Averaging and Markov's inequality give one seed r with

$$L_r \leq 3\alpha, \quad D_r \leq 3\rho. \quad (8)$$

For this seed, write $h_y(\theta) = \Pr_S\{A(S; r) = y\}$ and

$$F_y = \{\theta : h_y(\theta) > 3/4\}.$$

If θ lies outside every F_y , then $\sum_y h_y(\theta)^2 \leq 5/8$. It follows from (8) that

$$\mu\left(\bigcup_y F_y\right) \geq 1 - 8\rho. \quad (9)$$

Let $G_{y,t} = (F_y)_t$ for $0 \leq t \leq R$. Data processing and assumption 1 give

$$h_y(\theta) \geq \beta := 3/4 - \eta > 1/2 \quad \text{on } G_{y,t}. \quad (10)$$

Hence the closures of distinct $G_{y,t}$ are disjoint. Since the cells are disjoint, (8) and (10) imply

$$\sum_y \int_{G_{y,t}} \ell(y, \theta) d\mu(\theta) \leq \frac{3\alpha}{\beta}. \quad (11)$$

Call a cell good if its average loss is at most $K\alpha$. Equations (9) and (11) show that good cells cover at least $1 - 8\rho - 3/(\beta K)$ measure. For every good cell, Markov's inequality and (6) give $\mu(G_{y,t}) \leq 2s$. Assumptions 2 and 3 now yield

$$\sum_y \text{Per}_\mu(G_{y,t}) \geq c\lambda \quad (12)$$

for almost every $t \in [0, R]$.

A boundary point of $G_{y,t}$ belongs to no F_z . Indeed, membership in F_z would contradict the disjointness following (10). Its fixed-seed disagreement probability is therefore at least $3/8$. The union of all distance levels has measure at most 8ρ by (8). Integrating (12) and applying coarea gives $cR\lambda \leq 8\rho$. This proves Proposition 2. The appendix records the measurability details and a finite-range reduction.

Comparison with bounded-radius partitions.

The argument never claims that every F_y has small diameter. It uses the integrated distortion budget (11) to keep enough small-measure cells. This is what permits the expected guarantee (2) without inserting an accuracy failure probability into the final rate.

5 A toroidal categorical family

Let $d = \lfloor k/4 \rfloor$ and identify $\mathbb{T}^d = (\mathbb{R}/2\pi\mathbb{Z})^d$ with normalized Haar measure μ . Choose an amplitude $a = C_0\alpha$, where C_0 is a sufficiently large universal constant and $a \leq 1/2$. For $\theta \in \mathbb{T}^d$, define a distribution on $[d] \times [4]$ by

$$p_\theta(i, 1) = \frac{1 + a \cos \theta_i}{4d}, \quad p_\theta(i, 2) = \frac{1 - a \cos \theta_i}{4d}, \\ p_\theta(i, 3) = \frac{1 + a \sin \theta_i}{4d}, \quad p_\theta(i, 4) = \frac{1 - a \sin \theta_i}{4d}. \quad (13)$$

Remaining symbols, if any, receive mass zero. Every coordinate block has mass $1/d$.

5.1 Nearby parameters are indistinguishable

Let $\Delta_i \in [-\pi, \pi]$ be the shortest signed difference between θ_i and ϕ_i .

Lemma 3 (Sample continuity). *For the family (13),*

$$D_{\text{KL}}(p_\theta \| p_\phi) \leq \frac{a^2}{d} \sum_{i=1}^d \Delta_i^2. \quad (14)$$

Consequently, for every event B of m samples,

$$|p_\theta^m(B) - p_\phi^m(B)| < \frac{1}{15} \quad (15)$$

whenever $\text{dist}(\theta, \phi) \leq c\sqrt{d}/(a\sqrt{m})$.

The atom probabilities are at least $1/(8d)$. Bounding KL by chi-square, pairing the two cosine atoms and the two sine atoms, and using chord length at most arc length proves (14). Tensorization and Pinsker's inequality prove (15).

5.2 Accurate basins are exponentially small

The next lemma applies to every $q \in \Delta_k$, including outputs outside the hard family.

Lemma 4 (Small loss basins). *There are universal constants $c_0, C_0 > 0$ such that, with $a = C_0\alpha \leq 1/2$,*

$$\mu\{\theta : d_{\text{TV}}(q, p_\theta) \leq 2K\alpha\} \leq e^{-c_0d} \quad (16)$$

for every $q \in \Delta_k$ and the fixed universal K in Proposition 2.

To see the mechanism, define

$$u_i = \frac{2d}{a}(q(i, 1) - q(i, 2), q(i, 3) - q(i, 4)).$$

Choose φ_i as the direction of u_i , arbitrarily when $u_i = 0$, and set $Z_i(\theta) = \min\{1, \text{dist}_{\mathbb{T}}(\theta_i, \varphi_i)\}$. Radial projection onto the unit circle gives

$$d_{\text{TV}}(q, p_\theta) \geq \frac{a}{4\pi d} \sum_{i=1}^d Z_i(\theta). \quad (17)$$

Under Haar measure the Z_i are independent and $\mathbb{E}Z_i = 1 - 1/(2\pi)$. For sufficiently large C_0 , the event in (16) forces their empirical mean below half its expectation. Hoeffding's inequality completes the proof.

5.3 Torus isoperimetry

Let Φ and ϕ denote the standard Gaussian cdf and density, and set $I_\gamma(u) = \phi(\Phi^{-1}(u))$.

Lemma 5 (Gaussian profile on the torus). *For every measurable $E \subseteq \mathbb{T}^d$,*

$$\text{Per}_\mu(E) \geq \frac{I_\gamma(\mu(E))}{\sqrt{2\pi}}. \quad (18)$$

In particular, if $\mu(E) \leq 2e^{-c_0d}$, then

$$\text{Per}_\mu(E) \geq c_1\sqrt{d}\mu(E) \quad (19)$$

for all sufficiently large d .

There is a short transport proof. The coordinatewise map

$$T(x)_i = 2\pi\Phi(x_i) \pmod{2\pi}$$

pushes standard Gaussian measure on $\mathbb{R}^d$ to μ and is $\sqrt{2\pi}$ -Lipschitz. If $A = T^{-1}(E)$, then $A_t \subseteq T^{-1}(E_{\sqrt{2\pi}t})$. Gaussian isoperimetry (Borell, 1975; Sudakov and Tsirelson, 1978) and differentiation at zero give (18). The standard Gaussian-tail estimate $I_\gamma(u) \geq cu\sqrt{\log(1/u)}$ gives (19).

6 Proof of the main theorem

Apply Proposition 2 with $K = 32$ and $\eta = 1/15$ on $M = \mathbb{T}^d$, with $P_\theta = p_\theta$ and loss $\ell(q, \theta) = d_{\text{TV}}(q, p_\theta)$. Lemma 4 supplies assumption 2 with $s = e^{-c_0d}$. Lemma 5 supplies assumption 3 with $\lambda = c_1\sqrt{d}$. Lemma 3 supplies assumption 1 at radius

$$R = \min\left\{1, \frac{c_2\sqrt{d}}{a\sqrt{m}}\right\}. \quad (20)$$

Proposition 2 yields

$$\min\left\{1, \frac{c_2\sqrt{d}}{a\sqrt{m}}\right\} \sqrt{d} \leq C\rho. \quad (21)$$

For d above a universal constant and ρ below a universal constant, the first branch in (21) is impossible. Therefore

$$m \geq c \frac{d^2}{a^2\rho^2} = c' \frac{d^2}{\alpha^2\rho^2}.$$

Since $d = \lfloor k/4 \rfloor$, this proves Theorem 1.

Corollary 6 (Minimax complexity). *Let $m_{\text{rep}}^*(k, \alpha, \rho)$ be the minimum fixed sample size of a ρ -replicable estimator with worst-case expected TV loss at most α . For the parameter range of Theorem 1,*

$$c \frac{k^2}{\alpha^2\rho^2} \leq m_{\text{rep}}^*(k, \alpha, \rho) \leq C \frac{k^2 \log(1/\rho)}{\alpha^2\rho^2}.$$

The upper bound is Theorem 6.14 of Bun et al. (2023), with constants rescaled to match expected TV loss. The remaining logarithm is the only unresolved factor.

7 Discussion

The lower bound identifies two independent sources of the quadratic cost. A categorical observation selects one of d angular coordinates uniformly, which makes m -sample laws smooth at radius $\sqrt{d}/(\alpha\sqrt{m})$. At the same time, accurate canonical outputs partition a d -dimensional parameter space into exponentially small cells, forcing boundary density $\sqrt{d}$. Their product is $d/(\alpha\sqrt{m})$.

The toroidal construction is not cosmetic. If one works in a parameter cube and applies Euclidean isoperimetry to truncated cells, the external cube boundary has normalized area of order d and can dominate the desired $\sqrt{d}/\alpha$ internal boundary. Periodic sine-cosine encoding removes this loss while keeping every atom positive and the KL calculation elementary.

Our result concerns the standard fixed-sample definition used for discrete distribution estimation. Adaptive expected-sample procedures can behave differently in replicability problems (Hopkins et al., 2024). It is also open whether the logarithm in Corollary 6 is necessary. The average-distortion principle may be useful for other estimation losses where expected accuracy does not produce bounded-radius cells.

References

- Aamand, A., Aliakbarpour, M., Chen, J. Y., Narayanan, S., and Silwal, S. (2025). On the structure of replicable hypothesis testers. *arXiv preprint arXiv:2507.02842*.
- Acharya, J., Sun, Z., and Zhang, H. (2021). Differentially private assouad, fano, and le cam. In *Proceedings of the 32nd International Conference on Algorithmic Learning Theory*, volume 132 of *Proceedings of Machine Learning Research*, pages 48–78.
- Borell, C. (1975). The Brunn–Minkowski inequality in gauss space. *Inventiones Mathematicae*, 30(2):207–216.
- Bun, M., Gaboardi, M., Hopkins, M., Impagliazzo, R., Lei, R., Pitassi, T., Sivakumar, S., and Sorrell, J. (2023). Stability is stable: Connections between replicability, privacy, and adaptive generalization. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 520–527.
- Devroye, L. and Lugosi, G. (2001). *Combinatorial Methods in Density Estimation*. Springer.
- Diakonikolas, I., Gao, J., Kane, D., Liu, S., and Ye, C. (2025). Replicable distribution testing. In *Advances in Neural Information Processing Systems*, volume 38.
- Diakonikolas, I., Hardt, M., and Schmidt, L. (2015). Differentially private learning of structured discrete distributions. In *Advances in Neural Information Processing Systems*, volume 28.
- Dixon, P., Pavan, A., Vander Woude, J., and Vinodchandran, N. V. (2023). List and certificate complexities in replicable learning. In *Advances in Neural Information Processing Systems*, volume 36, pages 30784–30806.
- Hopkins, M., Impagliazzo, R., Kane, D., Liu, S., and Ye, C. (2024). Replicability in high dimensional statistics. *arXiv preprint arXiv:2406.02628*.
- Impagliazzo, R., Lei, R., Pitassi, T., and Sorrell, J. (2022). Reproducibility in learning. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 818–831.
- Liu, S. and Ye, C. (2024). Replicable uniformity testing. In *Advances in Neural Information Processing Systems*, volume 37.
- Sudakov, V. N. and Tsirelson, B. S. (1978). Extremal properties of half-spaces for spherically invariant measures. *Journal of Soviet Mathematics*, 9:9–18.

A Full proof of the partition principle

We prove Proposition 2 with all quantifiers explicit. The argument is stated for a compact Riemannian manifold with normalized volume measure. The same proof applies to any metric probability space that has the required isoperimetric and coarea properties.

For a fixed public random string r , the map $S \mapsto A(S; r)$ has finite range because the sample space is finite. We denote this range by $\mathcal{Y}_r$. For $y \in \mathcal{Y}_r$, let

$$h_{r,y}(\theta) = \Pr_{S \sim P_\theta^m} \{A(S; r) = y\}.$$

Assumption 1 and data processing imply

$$|h_{r,y}(\theta) - h_{r,y}(\phi)| \leq d_{\text{TV}}(P_\theta^m, P_\phi^m) \leq \eta \quad (22)$$

whenever $\text{dist}(\theta, \phi) \leq R$. For the categorical family used in the paper, $h_{r,y}$ is continuous. Continuity also follows locally from the vanishing right side of the KL bound as $\text{dist}(\theta, \phi) \downarrow 0$.

Define

$$L_r = \int_M \sum_{y \in \mathcal{Y}_r} h_{r,y}(\theta) \ell(y, \theta) d\mu(\theta),$$

$$D_r = \int_M \left(1 - \sum_{y \in \mathcal{Y}_r} h_{r,y}(\theta)^2 \right) d\mu(\theta).$$

The hypotheses of Proposition 2 hold pointwise in θ . Fubini’s theorem therefore gives

$$\mathbb{E}_r L_r \leq \alpha, \quad \mathbb{E}_r D_r \leq \rho.$$

By Markov's inequality,

$$\Pr_r\{L_r > 3\alpha\} < \frac{1}{3}, \quad \Pr_r\{D_r > 3\rho\} < \frac{1}{3}.$$

There is consequently a fixed seed r for which both inequalities in (8) hold. We fix it below and suppress r from the notation.

A.1 Canonical cells and coverage

For every $y \in \mathcal{Y}_r$, define the open canonical cell

$$F_y = \{\theta : h_y(\theta) > 3/4\}.$$

Distinct canonical cells are disjoint. If $\theta \notin \cup_y F_y$, every coordinate of the probability vector $(h_y(\theta))_y$ is at most $3/4$. Among probability vectors with this constraint, the squared ℓ_2 norm is maximized by $(3/4, 1/4, 0, \dots)$. Thus

$$1 - \sum_y h_y(\theta)^2 \geq 1 - \frac{10}{16} = \frac{3}{8}.$$

Since $D_r \leq 3\rho$,

$$\mu(N) \leq 8\rho, \quad N := M \setminus \bigcup_y F_y. \quad (23)$$

This proves (9).

For $t \in [0, R]$, put

$$G_{y,t} = \{\theta : \text{dist}(\theta, F_y) < t\}$$

for $t > 0$, and take $G_{y,0} = F_y$. If $\theta \in \overline{G_{y,t}}$, choose a sequence $\phi_n \in F_y$ with $\text{dist}(\theta, \phi_n) \leq t + o(1)$. Equation (22) and continuity give

$$h_y(\theta) \geq 3/4 - \eta = \beta > 1/2. \quad (24)$$

It follows that the closures of $G_{y,t}$ and $G_{z,t}$ are disjoint when $y \neq z$.

A.2 The distortion budget

For any $t \in [0, R]$, disjointness and (24) give

$$\begin{aligned} 3\alpha &\geq \int_M \sum_y h_y(\theta) \ell(y, \theta) d\mu(\theta) \\ &\geq \beta \sum_y \int_{G_{y,t}} \ell(y, \theta) d\mu(\theta). \end{aligned} \quad (25)$$

Let $\mathcal{G}_t$ be the indices satisfying

$$\frac{1}{\mu(G_{y,t})} \int_{G_{y,t}} \ell(y, \theta) d\mu(\theta) \leq K\alpha.$$

Empty cells are omitted. Equation (25) implies

$$\sum_{y \notin \mathcal{G}_t} \mu(G_{y,t}) \leq \frac{3}{\beta K}. \quad (26)$$

The thickened cells contain the original canonical cells, so (23) and (26) yield

$$\sum_{y \in \mathcal{G}_t} \mu(G_{y,t}) \geq 1 - 8\rho - \frac{3}{\beta K}. \quad (27)$$

For $K \geq 32$, $\eta < 1/4$, and sufficiently small universal ρ , the right side is at least a positive universal constant c_* .

Fix $y \in \mathcal{G}_t$. Markov's inequality within the cell gives

$$\mu\{\theta \in G_{y,t} : \ell(y, \theta) \leq 2K\alpha\} \geq \frac{1}{2} \mu(G_{y,t}).$$

Assumption 2 of Proposition 2 therefore implies

$$\mu(G_{y,t}) \leq 2s. \quad (28)$$

Assumption 3, (27), and (28) now show that for almost every t ,

$$\sum_y \text{Per}_\mu(G_{y,t}) \geq \lambda \sum_{y \in \mathcal{G}_t} \mu(G_{y,t}) \geq c_* \lambda. \quad (29)$$

A.3 Boundary points are noncanonical

Fix $t \in (0, R]$ and $\theta \in \partial G_{y,t}$. Then $\theta \notin F_y$, since every point of the open set F_y is interior to $G_{y,t}$. If $\theta \in F_z$ for some $z \neq y$, then $h_z(\theta) > 3/4$ while (24) gives $h_y(\theta) \geq \beta > 1/2$. This is impossible. Thus $\theta \in N$. More generally, every point of the shell $G_{y,R} \setminus F_y$ is on the boundary of $G_{y,t}$ at $t = \text{dist}(\theta, F_y)$ and belongs to N by the same argument. The shells for distinct y are disjoint because the closures of $G_{y,R}$ are disjoint. Consequently,

$$\sum_y \mu(G_{y,R} \setminus F_y) \leq \mu(N) \leq 8\rho. \quad (30)$$

The distance function to a set has metric gradient one almost everywhere outside its closure. The coarea formula, first for regular distance levels and then by approximation, gives

$$\int_0^R \text{Per}_\mu(G_{y,t}) dt \leq \mu(G_{y,R} \setminus F_y). \quad (31)$$

Summing (31), using (30), and integrating (29) gives

$$c_* R \lambda \leq \int_0^R \sum_y \text{Per}_\mu(G_{y,t}) dt \leq 8\rho.$$

This is (7) and completes the proof of Proposition 2.

B Continuity of the categorical experiment

We prove Lemma 3. For brevity, write $c_i = \cos \theta_i$, $c'_i = \cos \phi_i$, and similarly s_i, s'_i . Since $a \leq 1/2$, every

nonzero atom of p_ϕ is at least $1/(8d)$. The inequality $D_{\text{KL}}(p\|q) \leq \chi^2(p, q)$ gives

$$\begin{aligned} D_{\text{KL}}(p_\theta\|p_\phi) &\leq \sum_{i,j} \frac{(p_\theta(i,j) - p_\phi(i,j))^2}{p_\phi(i,j)} \\ &\leq 8d \sum_{i=1}^d \frac{2a^2(c_i - c'_i)^2 + 2a^2(s_i - s'_i)^2}{16d^2} \\ &= \frac{a^2}{d} \sum_{i=1}^d ((c_i - c'_i)^2 + (s_i - s'_i)^2). \end{aligned}$$

The expression in parentheses is the squared chord length between two points on the unit circle. It is at most Δ_i^2 , proving (14).

KL tensorizes over independent samples. Pinsker's inequality therefore gives

$$d_{\text{TV}}(p_\theta^m, p_\phi^m) \leq a \sqrt{\frac{m}{2d}} \text{dist}(\theta, \phi). \quad (32)$$

Choose $c < \sqrt{2}/15$ in the radius of Lemma 3. Then the right side of (32) is below $1/15$. The probability of any event, including the event that a fixed-seed estimator returns a specified output, changes by at most this total variation distance. This proves the lemma.

C The small-basin lemma

We prove Lemma 4. Fix $q \in \Delta_k$ and define u_i as in the main text. Let $v(\theta_i) = (\cos \theta_i, \sin \theta_i)$. If

$$L_i(\theta) = \sum_{j=1}^4 |q(i, j) - p_\theta(i, j)|,$$

then the triangle inequality in each signed pair gives

$$\|u_i - v(\theta_i)\|_2 \leq \frac{2d}{a} L_i(\theta). \quad (33)$$

We next record the radial projection fact used in (17). For any $u \in \mathbb{R}^2$, there is an angle φ such that, for every θ ,

$$\min\{1, \text{dist}_{\mathbb{T}}(\theta, \varphi)\} \leq \pi \|u - v(\theta)\|_2. \quad (34)$$

If $\|u\|_2 < 1/2$, the right side is at least $\pi/2$ and the claim is immediate. Otherwise choose $v(\varphi) = u/\|u\|_2$. Then

$$\|v(\theta) - v(\varphi)\|_2 \leq 2 \|v(\theta) - u\|_2.$$

For an angular separation $x \in [0, \pi]$, the chord length is $2 \sin(x/2) \geq 2x/\pi$. Equation (34) follows.

Combining (33) and (34), and summing over blocks,

yields

$$\begin{aligned} d_{\text{TV}}(q, p_\theta) &= \frac{1}{2} \sum_{i,j} |q(i, j) - p_\theta(i, j)| \\ &\geq \frac{a}{4\pi d} \sum_{i=1}^d \min\{1, \text{dist}_{\mathbb{T}}(\theta_i, \varphi_i)\}. \end{aligned}$$

Mass of q on unused symbols only increases the left side, so the inequality holds for all $q \in \Delta_k$.

For independent uniform angles Θ_i , the circular distance to a fixed φ_i is uniform on $[0, \pi]$. Thus, with $Z_i = \min\{1, \text{dist}_{\mathbb{T}}(\Theta_i, \varphi_i)\}$,

$$\mu_0 := \mathbb{E}Z_i = \frac{1}{\pi} \left(\int_0^1 x \, dx + \int_1^\pi 1 \, dx \right) = 1 - \frac{1}{2\pi}.$$

If $d_{\text{TV}}(q, p_\Theta) \leq 2K\alpha$ and $a = C_0\alpha$, then

$$\frac{1}{d} \sum_i Z_i \leq \frac{8\pi K}{C_0}.$$

Choose $C_0 \geq 16\pi K/\mu_0$. Hoeffding's inequality gives

$$\Pr \left\{ \frac{1}{d} \sum_i Z_i \leq \frac{\mu_0}{2} \right\} \leq \exp \left(-\frac{\mu_0^2 d}{2} \right).$$

Taking $c_0 = \mu_0^2/2$ proves (16).

D Isoperimetry on the flat torus

We prove Lemma 5. Let γ_d be standard Gaussian measure on $\mathbb{R}^d$ and define

$$T(x)_i = 2\pi \Phi(x_i) \pmod{2\pi}.$$

Each coordinate is uniform on the circle, so $T_{\#}\gamma_d = \mu$. Moreover,

$$\sup_x \left| \frac{d}{dx} 2\pi \Phi(x) \right| = \sqrt{2\pi}.$$

Taking shortest circular distances can only reduce distance, hence T is $L = \sqrt{2\pi}$ -Lipschitz in the product Euclidean metrics.

Fix measurable $E \subseteq \mathbb{T}^d$ and set $A = T^{-1}(E)$. Lipschitzness implies

$$A_t \subseteq T^{-1}(E_{Lt}).$$

The Gaussian isoperimetric inequality (Borell, 1975; Sudakov and Tsirelson, 1978) states

$$\gamma_d(A_t) \geq \Phi(\Phi^{-1}(\gamma_d(A)) + t).$$

Since $\gamma_d(A) = \mu(E)$, we obtain

$$\mu(E_{Lt}) \geq \Phi(\Phi^{-1}(\mu(E)) + t).$$

Subtracting $\mu(E)$, dividing by Lt , and taking a lower limit at zero proves

$$\text{Per}_\mu(E) \geq \frac{\phi(\Phi^{-1}(\mu(E)))}{\sqrt{2\pi}}.$$

For $0 < u \leq 1/2$, standard Gaussian tail bounds imply

$$\phi(\Phi^{-1}(u)) \geq cu\sqrt{\log(1/u)} \quad (35)$$

for a universal $c > 0$. If $u \leq 2e^{-c_0 d}$ and d is above a universal constant, then $\log(1/u) \geq c_0 d/2$. Substitution into (35) proves (19).

E Completion of Theorem 1

The application of Proposition 2 in the main text uses

$$R = \min \left\{ 1, \frac{c_2 \sqrt{d}}{a\sqrt{m}} \right\}.$$

Lemma 3 verifies sample continuity at this radius after reducing c_2 if needed. When the second term exceeds one, Equation (32) verifies continuity throughout the unit radius because $a\sqrt{m/d} < c_2$.

Proposition 2 and Lemma 5 give

$$R\sqrt{d} \leq C\rho.$$

Choose d_0 and ρ_0 so that $\sqrt{d} > C\rho$ for $d \geq d_0$ and $\rho \leq \rho_0$. Then $R \neq 1$, so

$$\frac{c_2 d}{a\sqrt{m}} \leq C\rho.$$

Rearranging and using $a = C_0\alpha$ gives

$$m \geq \frac{c_2^2}{C^2 C_0^2} \frac{d^2}{\alpha^2 \rho^2}.$$

Finally, for $k \geq 8$, $d = \lfloor k/4 \rfloor \geq k/8$. Increasing k_0 to ensure $d \geq d_0$ proves Theorem 1.