Proportionally Fair Clustering of Distributions

Debmalya Mandal
University of Warwick
Debmalya.Mandal@warwick.ac.uk

Nisarg Shah
University of Toronto
nisarg@cs.toronto.edu

Ziqi Yu
University of Toronto
ziqiyu@cs.toronto.edu

Abstract

Choosing cluster centers to minimize average distance from data points can leave a substantial part of the training data poorly represented. Proportional fairness asks that no group of data points large enough to deserve a center to represent them unanimously prefer a new center. Our main contribution is to expand this framework from clustering a finite set of data points to clustering a given probability distribution. A natural question is whether every distribution admits a proportionally fair clustering once the number of centers is large enough; we answer this negatively. We also identify broad families of distributions which admit a proportionally fair clustering when the number of centers is sufficiently large, and those that admit one for any number of centers. We also establish deep connections between proportional fairness for a distribution and that for a finite set of i.i.d. samples drawn from the distribution, allowing us to transfer approximation guarantees (and nonexistence results) known for the finite sample case with sample-efficiency.

1 Introduction

Suppose a city plans to open k service centers, with the goal of minimizing the average distance its n residents would have to travel to go to their nearest centers. Such an objective, though natural, can overlook certain neighborhoods whose residents all have to travel far. A neighborhood would have a particularly compelling complaint if its population is *sufficiently large* and they can propose a new location that *each of its residents prefers*. The literature on proportionally fair clustering formalizes this idea. It posits that any group of at least n/k residents is entitled to have a center representing them and should not be able to find a new location that is closer to every member than their respective closest centers. Such a clustering is said to be in the *(exact) core*. Note that this imposes a condition for every possible group of size at least n/k, not just a small number of prespecified groups (e.g., those identified by demographic attributes). Unfortunately, counterexamples similar to those presented by Chen et al. (2019) show that for every $k \geqslant 5$, there exist sufficiently large instances for which no k-clustering is in the core.

In machine learning applications, the input to the clustering problem often involves n training data points drawn i.i.d. from an underlying distribution $\mathcal{D}$ and the ultimate goal is that the chosen k centers represent the true distribution $\mathcal{D}$ ("low true risk"), not just the training data ("low empirical risk"). Known generalization bounds connect the empirical risk to the true risk for traditional accuracy objectives. Our goal is to chart a similar trajectory for the core, a proportional fairness objective.

Modulo minor technical caveats, which we explain later, the definition of the core extends naturally to a distribution: no subset of its domain with probability mass more than $1/k$ should admit a new location for a center that is closer to all points in the region than their respective nearest centers. The aforementioned counterexamples imply that for each $k \geqslant 5$, there exists a distribution that does not admit a k-clustering in the core.

What if we fix a distribution $\mathcal{D}$ first and ask whether the (exact) core would begin to exist once k is large enough? In case of clustering n data points, the answer is trivially positive once $k \geqslant n$. For distributions, this turns out to be a non-trivial question. More broadly, we addresses the following research questions.

Which distributions admit a k-clustering in the (exact) core when k is sufficiently large and which admit it for every k? How does the (exact or approximate) core for a distribution relate to that for n i.i.d. samples from the distribution?

1.1 Our Results

In Section 3, we begin by establishing connections between proportional fairness for a distribution $\mathcal{D}$ and that for n i.i.d. samples from $\mathcal{D}$. Specifically, we consider a bicriteria approximation, (α, β)-core, which demands that no group of greater than $\beta \cdot n/k$ points in the finite case and region with more than β/k probability mass in the distributional case can propose a new location that improves the distance to each member by a multiplicative factor more than α. We show that for all distributions on a proper metric space and all values of k, a k-clustering in the (α, β)-core exists:

- when $(\alpha, \beta) = (2, 1)$, importing a recent approximation result for the finite case by Cookson et al. (2026a) for the case where each data point is a feasible cluster center location (which is true in our distributional case), and
- when $(\alpha, \beta) = (1, 4.911)$, importing a result from a more general setup in social choice theory by Song et al. (2026) and Charikar et al. (2026).

For distributions over the Euclidean space $\mathbb{R}^D$, we show that these approximations transfer with sample-efficiency. Specifically, $n = O(Dk^2 \log^2 k)$ i.i.d. samples from a distribution $\mathcal{D}$ suffice to ensure, with high probability, that every clustering in (α, β)-core with respect to the n samples is in $(\alpha, O(\beta))$-core with respect to $\mathcal{D}$, and vice-versa (Theorem 3); we also show that the quadratic dependence on k is unavoidable (Theorem 4).

For (exact) $(1, 1)$-core, counterexamples similar to those presented by Chen et al. (2019) show that for each $k \geqslant 5$, there is a distribution which does not admit a k-clustering in the exact core, leaving open whether each distribution may admit a k-clustering in the exact core once k is sufficiently large. In Section 4, we answer this negatively: we exhibit a distribution supported on the unit square in the plane for which a k-clustering in the exact core does not exist for $k = 5 \cdot 9^t$ for each $t \geqslant 0$ (Theorem 5). Next, we prove that for the following broad families of distributions, a k-clustering in the core exists for sufficiently large k (Theorem 6):

- Absolutely continuous measures on $\mathbb{R}^D$ with $D \leqslant 5$.
- Arbitrary measures on $\mathbb{R}^D$, for any D, with a mass greater than $3911/4911$ concentrated on a countable support (and the rest distributed arbitrarily).

For the following more structured families of distributions, we improve the guarantee to show that a k-clustering in the core exists for all k (Theorem 7):

- Arbitrary measures on a line in $\mathbb{R}^1$.
- Uniform measure over rectangles and disks in $\mathbb{R}^2$.

We also show that all Gaussian distributions in any finite dimension admit a k-clustering in the core for $k \leqslant 4$, and conjecture that this should in fact hold for all k.

1.2 Related Work

Proportional fairness in finite clustering. Chen et al. (2019) introduced proportionally fair clustering and gave the GreedyCapture algorithm, which successively serves sufficiently large nearby groups. The general metric guarantee is a cost factor of $1 + \sqrt{2}$, and Micha & Shah (2020) proved a factor of 2 in Euclidean space, as well as exact existence for finite populations on trees. Cookson et al. (2026b) strengthened the general metric lower bound to 2.1508 in the model with a prescribed candidate set that need not contain the voter locations. When every voter location is feasible, Cookson et al. (2026a) proved existence with cost factor 2 under the stronger finite Droop quota, which allows deviations by groups larger than a $1/(k+1)$ fraction of the population. Thus the feasible center domain matters when comparing constants. Our transfer arguments preserve finite approximation guarantees for distributions; our exact existence theorems use the geometry of a fixed distribution as the budget grows.

Other proportional representation criteria. Aziz et al. (2026) introduce proportionally representative fairness, which links the number of centers near a group to both its size and its diameter, and give efficient algorithms for unrestricted and discrete center domains. Their axiom addresses the representation of groups entitled to several centers. Li et al. (2021) study a transferable-utility core: a coalition can deviate whenever doing so lowers its total distance, even if some members individually lose. They consider relaxations in both improvement and coalition size. Our two approximation parameters have analogous roles, but our blocking condition requires every member to improve. Kellerhals & Peters (2024) connect proportional clustering to multiwinner voting, prove approximation implications between proportional and individual fairness, and study deviations to multiple centers. These results situate our single-alternative core among broader representation requirements; exact existence for our condition does not assert exact satisfaction of those stronger requirements.

Population clustering and sampling. Pollard (1981) established strong consistency of empirical k-means centers under suitable assumptions. For proportional fairness, Chen et al. (2019) gave sampling guarantees for finite candidate domains, and Micha & Shah (2020) extended the analysis to unrestricted Euclidean centers. Our uniform transfer theorem uses relative approximations for geometric range spaces (Har-Peled & Sharir, 2011) to control error at the shrinking mass scale $1/k$, and the planar lower bound shows why a quadratic dependence on k is necessary when every empirical core must transfer. This is a guarantee about deviating groups, separate from convergence of average clustering cost. The latter follows here through Wasserstein distance, whose empirical convergence rates are studied by Fournier & Guillin (2015).

Committee selection and sortition. Choosing centers can also be viewed as choosing representatives: each point ranks candidate centers by distance. Jiang et al. (2020) establish constant-factor approximate stability for committees with general monotone preferences, allowing coalitions to propose alternative committees of proportional weight. The ranked-preference theorem of Song et al. (2026), with a proof exposition by Charikar et al. (2026), gives the constant used in our 4.911 mass approximation. Its input has finitely many voters and candidates; Section A supplies the limits needed for continuum center domains and population measures. In sortition, Ebadian & Micha (2025) combine equal selection probabilities for individuals with a constant-factor core guarantee for each panel produced by their randomized algorithm. Their centers must be selected citizens, and the selection probabilities are an additional fairness constraint. Our distributional results concern unrestricted locations and the existence of a center set.

Our work is also related to demographic and individual fairness, non-centroid clustering, and Geometric hitting sets. Due to limited space, we discuss these connections in Appendix H.

2 Preliminaries

A population is a Borel probability measure μ on a metric space (M, d). Unless explicitly stated otherwise, every point of M is a feasible center. A budget k is a positive integer, and a clustering is a nonempty set $C \subseteq M$ of at most k centers. A point x is served by its nearest center, at distance $d(x, C) = \min_{c \in C} d(x, c)$. Using fewer than k centers is harmless: adding centers can only improve the guarantees considered here. The principal setting is $M = \mathbb{R}^D$ with Euclidean distance, with no restriction that a center belong to the support of μ.

Definition 1 (Approximate core)**.** For factors $\alpha, \beta \geqslant 1$, a clustering C and an alternative center y have deviation region $D_\alpha(C, y) = \{x \in M : \alpha d(x, y) < d(x, C)\}$. The clustering is an (α, β)-*core* for budget k if $\mu(D_\alpha(C, y)) \leqslant \beta/k$ for every $y \in M$. An *exact core*, or simply a *core*, is a $(1, 1)$-core. We write $D(C, y) = D_1(C, y)$.

The factor α approximates the individual cost: a point must improve its distance by more than this factor to join a deviation. The factor β approximates the deviating mass: only groups of mass strictly larger than β/k are forbidden. Distances improve strictly, while the upper bound on captured mass is non-strict. The regions are open and hence measurable. If $\beta \geqslant k$, the mass condition imposes no restriction.

Remark 1 (Quota convention). For n equally weighted points, our convention forbids a deviation by $\lfloor n\beta/k \rfloor + 1$ points. The common finite convention instead forbids $\lceil n\beta/k \rceil$ points, and differs only when $n\beta/k$ is an integer. All finite theorems we inherit give guarantees at least as strong as the ones used here. A bound of $1/(k+1)$ corresponds to the finite Droop quota $\lfloor n/(k+1) \rfloor + 1$. Assigning zero mass to every individual point does not in general make the strict and non-strict conventions equivalent; we fix the convention in Definition 1 throughout.

To compare distributions and constructions, it is useful to ask for the smallest possible size of the largest deviating group, without fixing a quota in advance.

Definition 2 (Largest deviating mass)**.** For a Borel probability measure μ and a budget k, define

$$\delta(\mu, k) = \inf_{\substack{C \subseteq M \\ 1 \leqslant |C| \leqslant k}} \sup_{y \in M} \mu(D(C, y)). \tag{1}$$

An exact core is a center set for which the inner supremum is at most $1/k$. In particular, $\delta(\mu, k) < 1/k$ implies existence. A bound $\delta(\mu, k) \leqslant 1/k$ alone requires care if the infimum is not attained. All asymptotic existence results below establish a strict margin and therefore do not rely on attainment.

For a sample $P = (x_1, \ldots, x_n)$, repetitions are allowed, and its empirical measure $\widehat{\mu}_P$ assigns mass $1/n$ to each occurrence. For the distinction between distribution families, a point x is an *atom* if $\mu(\{x\}) > 0$. A probability measure has at most countably many such points; it need not place any mass on them. A measure is *nonatomic* if it has none, and is *absolutely continuous* in $\mathbb{R}^D$ if it has a density with respect to D-dimensional Lebesgue measure. Nonatomic measures need not have densities, as the example in Theorem 5 will demonstrate.

Example 1 (A uniformly populated interval)**.** Let μ be uniform on $[0, 1]$ and choose $C = \{1/6, 1/2, 5/6\}$ for a budget of three. For an alternative y between consecutive centers $a < b$, the captured interval has endpoints $(a + y)/2$ and $(b + y)/2$, so its length is $(b - a)/2 = 1/6$. If $y < 1/6$, its captured portion of $[0, 1]$ has length at most $1/6$; the right endpoint is symmetric. Choosing an incumbent center captures nothing. Thus every alternative captures at most $1/6$, below the core quota $1/3$. For instance, $y = 2/3$ captures $(7/12, 3/4)$, as shown in Figure 1.

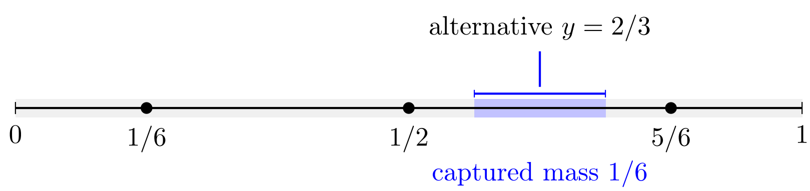

Figure 1: Three centers on a uniformly populated interval. The proposed new center is better precisely between its bisectors with the two adjacent centers. This interval has half the mass represented by one equally spaced center.

Two elementary bounds locate the scale of Equation (1). A measure assigning equal mass to $k + 1$ distinct points has $\delta(\mu, k) = 1/(k + 1)$: every clustering omits a support point, and choosing all but one support point attains this bound. A different lower bound holds for every nonatomic Euclidean distribution, regardless of its shape.

Lemma 1. *For any nonatomic Borel probability measure μ on $\mathbb{R}^D$ with Euclidean distance, and any nonempty set C of at most k centers, $\sup_y \mu(D(C, y)) \geqslant 1/(2k)$.*

Proof. Break ties between nearest centers measurably, so some cell V assigned to a center c has mass at least $1/k$. A direction chosen uniformly on the unit sphere almost surely has $\mu(V \cap \{x : \langle x - c, v\rangle = 0\}) = 0$: for each $x \neq c$, the orthogonal directions have surface measure zero, and Fubini's theorem applies. This also holds in dimension one, where the only excluded point is c. Reverse a suitable direction if necessary so that the positive side has at least half of $\mu(V)$. The alternative $c + tv$, for $t > 0$, captures every $x \in V$ with $\langle x - c, v\rangle > t/2$, since squaring the distance comparison gives exactly this inequality. Letting t decrease to zero proves the bound by continuity from below. □

We will again encounter the distinction between these bounds later in the paper. The equal point masses describe a worst case that can change with k. For all absolutely continuous probability measures on $\mathbb{R}^D$, the limit of $k\delta(\mu, k)$ as $k \to \infty$ is the same, depending only on the dimension D.

3 Inheriting Approximations From the Finite Case

We first establish what carries over from finite populations: approximate-core existence for general probability distributions, finite-sample upper and lower bounds for transferring cores between empirical and population distributions, and smooth counterexamples to exact-core existence at fixed budgets. The last result rules out existence for every k under smoothness and convex support alone, while leaving open whether an exact core exists for all sufficiently large k for a fixed distribution.

Approximation guarantees. Finite results provide two relaxations of the core: relaxing to stricter cost improvements or larger deviating groups. Both extend to distributions without increasing the approximation factors. Here, a metric space is *proper* if every closed bounded subset is compact, and a Borel probability measure μ is *tight* if, for every $\varepsilon > 0$, some compact set K satisfies $\mu(K) > 1 - \varepsilon$.

Theorem 1. *Every tight Borel probability measure on a metric space admits a $(2, 1)$-core for every positive integer k, when every point of the space is a feasible center. If the measure has compact support, or if the metric space is proper, the centers can moreover be chosen so that $\mu(D_2(C, y)) \leqslant 1/(k + 1)$ for every y.*

Proof of Theorem 1. Suppose first that μ is supported on a compact set K. Approximate μ weakly by empirical measures supported on K. The finite theorem selects all centers from K, so a subsequence of the resulting k-tuples converges in K^k. If a challenger attracted mass greater than $1/(k + 1)$ from the limiting clustering, some positive improvement margin would already hold on a set of mass greater than $1/(k + 1)$. The same set would then violate the finite guarantee for all sufficiently accurate approximations. The full limit argument is given in Lemma 4.

For a tight measure and $k \geqslant 2$, choose a compact set K whose omitted mass $\varepsilon = \mu(M \setminus K)$ is less than $1/k^2$. Apply the compact case to the conditional probability measure on K. For the resulting centers and every challenger y,

$$\mu(D_2(C, y)) \leqslant \frac{1 - \varepsilon}{k + 1} + \varepsilon = \frac{1 + k\varepsilon}{k + 1} < \frac{1}{k}.$$

For $k = 1$, any center satisfies the required bound of 1. Finally, in a proper space, Lemma 3 applies directly with $\alpha = 2$ and $q = 1/(k + 1)$, including when the support is unbounded. □

The cost guarantee follows from the finite theorem of the work of Cookson et al. (2026a) who show when every data point is a feasible center, one can select at most k data points so that a group of mass greater than $1/(k + 1)$ cannot improve its cost by a factor greater than 2. The main idea is to successively approximate a measure through a limit of finite samples. In order for the limiting argument on a compact support to work, we need to ensure that the centers selected for successive finite approximations remain in a compact set. Fortunately, the result of Cookson et al. (2026a) allows us to choose centers at data points themselves and keeps them in that compact support. Details appear in Section A.

Theorem 2. *Every Borel probability measure on a proper metric space admits a $(1, 4.911)$-core for every positive integer k. In particular, this holds on $\mathbb{R}^D$ under any norm.*

Proof of Theorem 2. For a finite empirical population and a finite set of candidate centers, rank the candidates in increasing order of distance, breaking ties arbitrarily. A voter who strictly prefers an outsider in distance also ranks it above every selected center, regardless of the tie-breaking rule. The finite committee theorem therefore verifies the hypothesis of Lemma 3 with $\alpha = 1$ and $q = 4.9108/k$. When $q < 1$, that lemma gives actual centers with the same bound for μ. When $q \geqslant 1$, any center suffices. Increasing the bound to $4.911/k$ proves the statement. □

The mass guarantee follows from the works of Song et al. (2026) and Charikar et al. (2026) by ranking centers according to distance. At a high level, we approximate the measure by finite empirical populations and pass to a limit of their center sets. After padding each solution to a k-tuple, we extract subsequences one center at a time. In a proper space, each center either has a convergent subsequence or escapes every bounded set, in which case it can be discarded. For the cost guarantee on tight measures, we first restrict attention to a compact set, using that the corresponding finite theorem selects centers from the population. Details appear in Section A.

Finite samples. The preceding arguments establish existence by taking limits of finite populations. However, an application instead observes a sample of a fixed size drawn from the distribution. We now ask how much the two approximation factors can change when the same centers are evaluated according the underlying distribution. The next theorem gives a distribution-free sample-complexity bound for transferring approximate-core guarantees between the empirical measure and the underlying distribution.

Theorem 3. *Fix $\alpha \geqslant 1$, $1 \leqslant \beta < k$, and $\eta, \zeta \in (0,1)$. There is an absolute constant $c > 0$ such that, for any Borel probability measure μ on $\mathbb{R}^D$, an independent sample of size*

$$n \geqslant c\frac{k}{\beta\eta^2}\left(Dk\log(k+1)\log\frac{2k}{\beta} + \log\frac{1}{\zeta}\right) \tag{2}$$

satisfies the following simultaneously with probability at least $1 - \zeta$:

- *Every empirical (α, β)-core is a population $(\alpha, \beta/(1-\eta))$-core.*
- *Every population (α, β)-core is an empirical $(\alpha, (1+\eta)\beta)$-core.*

Distance is Euclidean, and the first assertion allows arbitrary sample-dependent centers.

Each deviation region is an intersection of at most k halfspaces when $\alpha = 1$, or balls when $\alpha > 1$. Their VC dimension is $O(Dk\log(k+1))$, and we use relative approximation bounds (Har-Peled & Sharir, 2011) to control their masses uniformly at scale β/k, yielding the theorem. For fixed dimension, approximation factors, accuracy, and confidence, $O(k^2\log^2 k)$ observations suffice. The next result shows that the quadratic dependence on k is necessary for a guarantee covering every empirical core.

Theorem 4 (A lower bound for uniform transfer)**.** *For every sufficiently large integer k and every integer n with*

$$200k\log k \leqslant n \leqslant k^2/100,$$

there exist a finitely supported probability measure μ on $\mathbb{R}^2$ and a measurable rule assigning exactly k centers C_P to each sample of size n such that, with probability $1 - o_k(1)$,

$$\sup_y \widehat{\mu}_P(D(C_P, y)) \leqslant \frac{9}{10k}, \qquad \mu(D(C_P, 0)) \geqslant \frac{4}{3k}.$$

The probability bound is uniform over the stated range of n. Thus an empirical exact core can fail to be a population $(1, 5/4)$-core with probability tending to one.

The construction places one heavy point outside a circle and many rare points on the circle. Sample-dependent centers keep every empirical deviation small while allowing the origin to attract all unobserved rare atoms. A point other than the origin cannot attract too many sampled points, because any such points lie in a semicircle. The origin, however, attracts all unobserved rare points, and their total mass is large enough to violate population proportionality. The theorem implies a sample threshold of order at least k^2 for a distribution-free guarantee for the transfer of *empirical* core. It does not rule out smaller samples for an algorithm that selects a particular empirical core. Proofs of both sampling bounds, including an explicit probability bound for Theorem 4, appear in Section B.

Exact core impossibility. If we require an exact core instead of an approximate core, we cannot expect every probability measure to admit one for every k: the case of finite samples is already a special case of probability measures as it can be represented by atomic measures, and for each $k \geqslant 5$, counterexamples similar to those presented by Chen et al. (2019) show that there exists a finite sample instance that does not admit an exact core. It is therefore natural to ask what happens for "really continuous" measures. Could the finite obstructions disappear as soon as there is a smooth density and a convex support?

However, a positive margin in a finite counterexample rules this out. For example, place equal mass at the vertices of three small equilateral triangles that are far apart, and allow five centers anywhere in the plane. One triangle has at most one nearby center. Two of its vertices can then strictly improve by moving to a common alternative, so a mass of $2/9 > 1/5$ can deviate. For $k > 5$, adding $k - 5$ distant atoms of mass just above $1/k$ forces the additional centers and preserves the obstruction. The geometric argument in Lemma 7 gives a uniform positive distance margin around both vertices. Replacing each point mass by a smooth probability density in a sufficiently small ball preserves the deviation for every possible placement of five centers. Finally, mixing these nine clouds with a sufficiently small amount of a smooth density supported on a rectangle and positive throughout its interior gives us a continuous and smooth measure with convex support. The latter density can be obtained by taking the product of two normalized copies of $\exp(-1/(t(1-t)))$ on $(0, 1)$ and setting it to zero outside. Therefore, this construction rules out the possibility that every smooth probability measure with convex support and smooth density admits an exact core for every k.

This argument leaves open the following interesting question: if we fixed a distribution and increase k, would an exact core always begin to exist for all sufficiently large k? For the finite case of n data points (equivalently, atomic measures with finite support), this holds trivially whenever $k \geqslant n$ since one can place a center at every data point. But the question turns out to be non-trivial for general (nonatomic) distributions.

4 Exact Core Existence for Distributions

We answer this question negatively by designing a nonatomic distribution supported on the unit square for which the exact core is empty at infinitely many values of k.

Theorem 5. *There is a nonatomic Borel probability measure μ with support $[0,1]^2$ which admits no k-clustering in the exact core whenever $k = 5 \cdot 9^t$ for any integer $t \geqslant 0$.*

The measure is constructed roughly as follows. It takes a counterexample to the finite case with $n = 9$ and $k = 5$ and nests it recursively. All the point masses are infinitely replaced by smaller copies of the counterexample. A uniform component gives positive mass to every open subset of the square. This makes the measure nonatomic, and yet, we show that the exact core nonexistence carries through. The complete proof is in Section C.1.

The impossibility of Theorem 5 raises an interesting question: Which distributions admit an exact core once k is sufficiently large?

4.1 Exact Core For Sufficiently Large k

One observation is that the measure in Theorem 5, while nonatomic, was quite artificially constructed. Could we get a positive result if the measure was "truly continuous"? On the other extreme, if the answer is trivially positive for the finite case, could that extend to measures that are "sufficiently atomic"? We show positive answers in both these cases.

Theorem 6 (Exact cores for sufficiently many centers)**.** *Let μ be any Borel probability measure on $\mathbb{R}^D$, equipped with Euclidean distance, and allow centers anywhere in $\mathbb{R}^D$. In the following two cases, there is an integer K such that there exists a k-clustering in the exact core for every $k \geqslant K$.*

- *The measure is absolutely continuous with respect to D-dimensional Lebesgue measure and $D \leqslant 5$.*
- *More than $3911/4911$ of its mass lies on a countable set; the remaining mass is arbitrary.*

The proof begins with a stronger statement about ordinary densities. Recall that $\delta(\mu, k)$ is the smallest possible largest deviating mass. We show that $\lim_{k\to\infty} k \cdot \delta(\mu, k) = b_D$, a constant that depends only on the dimension D and not on the measure μ. Thus, the shape of the distribution affects the value of K, after which an exact core may begin to exist, but not this constant. We then prove that $b_D < 1$ for $D \leqslant 5$. The dimension four and five proofs require computer-assisted checks implemented via exact rational arithmetic.

Lemma 2 (The asymptotic constant for densities)**.** *For each integer $D \geqslant 1$, there exists a constant $b_D \in [1/2, 4911/1000]$ such that every absolutely continuous Borel probability measure μ on $\mathbb{R}^D$ satisfies $\lim_{k\to\infty} k\delta(\mu, k) = b_D$. Moreover,*

$$b_1 = \frac{1}{2}, \qquad b_2 \leqslant \frac{9}{16}, \qquad b_3 \leqslant \frac{7}{8}, \qquad b_4 < \frac{19}{20}, \qquad b_5 < \frac{99}{100}. \tag{3}$$

The relevant threshold is one: a limit below one implies $\delta(\mu, k) < 1/k$ for all sufficiently large k, and hence the existence of an actual core. The constants in Equation (3) are consequently strong enough to prove exact existence, whereas the universal bound $4.911/k$ does not. We emphasize that these are upper bounds on b_D; we do not identify b_D exactly for any $D \geqslant 2$.

Here is the high-level idea behind the upper bound $\lim_{k\to\infty} k\delta(\mu, k) \leqslant b_D$. We observe that for a uniform measure, placing the centers in a repeating pattern lets us approach the bound b_D/k. Since our measure is non-uniform, we need to make this construction adaptive to the density, placing a coarse pattern in regions where the density is small, and a fine pattern in regions where the density is large. Finitely many such patterns placed over separated cubes can cover almost all probability mass of the measure. The remaining mass is served using a small fraction of the centers and the universal approximation from Theorem 2.

The matching lower bound $\lim_{k\to\infty} k\delta(\mu, k) \geqslant b_D$ reverses this construction. We show that varying the density cannot reduce the asymptotic constant below that of the uniform cube. We first establish the limiting constant for the cube by repeating its center arrangement across space, adding a negligible fraction of extra centers to control deviations near the boundaries. We then consider a density that is constant on each of several separated cubes. Any clustering must serve each cube, so we can apply the uniform-cube lower bound separately to each cube and combine the resulting bounds. Finally, we approximate an arbitrary density by densities of this form and show that the approximation error does not affect the limit. Together with the upper bound, this proves that all densities in the same dimension share the same asymptotic constant b_D.

The proof of Theorem 6 is given in Appendix C.

4.2 Exact Core For All k

While establishing the existence of exact core for sufficiently large k is helpful, it leaves one wondering whether, at least for certain classical distributions, one can expect a stronger result: exact core existence for all values of k.

Note that the strength of Lemma 2 now disappears. This question depends not only on the dimension, but on the exact shape of the distribution. While we believe that many classical distributions should admit exact core for all values of k, proving this has turned out to be surprisingly difficult, except in the case of measures in 1D, for which a simple argument works. In 2D, we are able to establish the result for uniform measures on planar rectangles and disks.

Theorem 7 (Exact cores for every number of centers)**.** *For Euclidean distance and unrestricted centers, each of the following probability measures admits an exact core for every positive integer k:*

- *every Borel probability measure supported on a line;*
- *the uniform area measure on any planar rectangle of positive area;*
- *the uniform area measure on any planar disk of positive radius.*

The proof of Theorem 7 appears in Appendix D. For rectangles, the construction combines rows of evenly spaced centers with rectangular grids. When the rectangle is narrow compared to k, then choosing rows suffice. Once the width of the rectangle is large enough so that centers are available to populate both directions, the grid spacing can be chosen in a way that each possible alternative captures less area than the proportional share. The proof tracks the integer number of rows and columns, which is necessary to cover all values of k rather than just a subsequence of square budgets.

For disks, a fine grid proves the claim once the budget is large enough, while smaller budgets either use explicit arrangements adapted to the boundary, or checked through satisfaction of a system of inequalities covering the entire continuum of alternative centers. Lemma 18 highlights these computations.

A much more interesting class of distributions than those covered in Theorem 7 is that of Gaussian distributions. We conjecture that they should admit an exact core for all k as well.

Conjecture 1. *Every Gaussian probability measure on a finite-dimensional Euclidean space admits a k-clustering in the exact core for every positive integer k.*

As evidence towards this conjecture, we prove that the exact core indeed exists for $k \leqslant 4$.

Theorem 8. *Every Gaussian probability measure on a finite-dimensional Euclidean space admits a k-clustering in the exact core for $k \in \{1, 2, 3, 4\}$.*

The proof appears in Section E. GPT-6 Astra also proved exact core existence for spherical planar Gaussians when $k \leqslant 11$ and when $k \geqslant 10000$, which provides further evidence for Conjecture 1, but we have chosen to omit this partial result from our paper.

5 Discussion

In this work, we initiate the study of proportionally fair clustering of distributions. In contrast to clustering finitely many data points, where only approximate core can be guaranteed, we identify several families of distributions which admit exact core when the number of centers k is sufficiently large, and some that even admit it for all values of k. Pushing the boundary of these existence results remains open.

For exact core with sufficiently large k, the counterexample in Theorem 5 suggests one needs more structure on the probability mass than merely ruling out out atoms or disconnected support. Absolute continuity is one such structure, for which Lemma 2 already reduces the question across all measures to bounding a dimension-dependent constant b_D. Proving $b_D < 1$ for some D would yield exact core existence for all absolutely continuous measures in $\mathbb{R}^D$ for all sufficiently large k. On the other hand, establishing $b_D > 1$ for some D would imply that every absolutely continuous measure in $\mathbb{R}^D$ in fact has empty core for every sufficiently large k. If $b_D = 1$, the existence of exact core would require a finer analysis. This leaves the following tantalizing open question.

Open Question 1. Does every absolutely continuous probability measure on $\mathbb{R}^D$, in every finite dimension D, admit an exact core for all sufficiently large k? More strongly, is $b_D < 1$ for every D?

For the existence of the exact core at every k, a natural candidate would be log-concavity. A log-concave density satisfies $f((1-t)x + ty) \geqslant f(x)^{1-t} f(y)^t$ for $t \in (0, 1)$; its mass cannot be concentrated in several separated peaks with deep valleys between them. This family includes Gaussian distributions, uniform measures on convex bodies, and many distributions used in statistics.

Open Question 2. Does every log-concave probability density on $\mathbb{R}^D$ admit an exact core for every positive integer k? Is this at least true for the uniform measure on any convex body?

AI Use Statement

GPT-5.6 Sol and GPT-6 Astra were used to formulate and prove some of the mathematical claims presented in this work, and in writing the first draft of this manuscript and during its iterative revisions. The authors retain full responsibility for the correctness of the mathematical arguments presented in this paper.

References

Haris Aziz, Barton E. Lee, Sean Morota Chu, and Jeremy Vollen. Proportionally representative clustering. In *Web and Internet Economics*, volume 15534 of *Lecture Notes in Computer Science*, pp. 155–171. Springer, 2026. doi: 10.1007/978-3-032-08560-3_9. URL https://link.springer.com/chapter/10.1007/978-3-032-08560-3_9. WINE 2024; proceedings published in 2026.

Ioannis Caragiannis, Evi Micha, and Nisarg Shah. Proportional fairness in non-centroid clustering. *Advances in Neural Information Processing Systems*, 37:19139–19166, 2024.

Moses Charikar, Prasanna Ramakrishnan, and Kangning Wang. An exposition of five candidates suffice for a majority. *arXiv preprint arXiv:2606.14666*, 2026.

Xingyu Chen, Brandon Fain, Liang Lyu, and Kamesh Munagala. Proportionally fair clustering. In *International conference on machine learning*, pp. 1032–1041. PMLR, 2019.

Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, and Sergei Vassilvitskii. Fair clustering through fairlets. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/hash/978fce5bcc4eccc88ad48ce3914124a2-Abstract.html.

Benjamin Cookson, Nisarg Shah, and Ziqi Yu. Unifying proportional fairness in centroid and non-centroid clustering. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/32b640528f5b67975562210f00c131ed-Abstract-Conference.html.

Benjamin Cookson, Eva Deltl, and Yeeseok Oh. Optimally selecting representative agents from a metric space. *arXiv preprint arXiv:2608.29097*, 2026a. URL https://arxiv.org/abs/2608.29097.

Benjamin Cookson, Eva Deltl, and Yeeseok Oh. Improved lower bounds for proportionally fair clustering. *arXiv preprint arXiv:2606.07285*, 2026b.

Soroush Ebadian and Evi Micha. Boosting sortition via proportional representation. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems*, pp. 667–675. International Foundation for Autonomous Agents and Multiagent Systems, 2025. URL https://ifaamas.csc.liv.ac.uk/Proceedings/aamas2025/pdfs/p667.pdf.

Nicolas Fournier and Arnaud Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability theory and related fields*, 162(3):707–738, 2015.

Sariel Har-Peled and Micha Sharir. Relative (p, ε)-approximations in geometry. *Discrete & Computational Geometry*, 45(3):462–496, 2011. doi: 10.1007/s00454-010-9248-1.

Zhihao Jiang, Kamesh Munagala, and Kangning Wang. Approximately stable committee selection. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 463–472, 2020.

Christopher Jung, Sampath Kannan, and Neil Lutz. Service in your neighborhood: Fairness in center location. In *1st Symposium on Foundations of Responsible Computing (FORC 2020)*, volume 156 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 5:1–5:15. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2020. doi: 10.4230/LIPIcs.FORC.2020.5. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.FORC.2020.5.

Leon Kellerhals and Jannik Peters. Proportional fairness in clustering: A social choice perspective. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/c981fd12b1d5703f19bd8289da9fc996-Abstract-Conference.html.

Bo Li, Lijun Li, Ankang Sun, Chenhao Wang, and Yingfan Wang. Approximate group fairness for clustering. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 6381–6391. PMLR, 2021. URL https://proceedings.mlr.press/v139/li21j.html.

Evi Micha and Nisarg Shah. Proportionally fair clustering revisited. In *47th International Colloquium on Automata, Languages, and Programming (ICALP 2020)*, pp. 85:1–85:16. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2020.

David Pollard. Strong consistency of K-means clustering. *The Annals of Statistics*, 9(1):135–140, 1981. URL https://www.stat.yale.edu/~pollard/Papers/Pollard81AS.pdf.

Richard Rado. A theorem on general measure. *Journal of the London Mathematical Society*, 21(4):291–300, 1946. doi: 10.1112/jlms/s1-21.4.291.

Natan Rubin. An improved bound for weak epsilon-nets in the plane. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 224–235. IEEE, 2018.

Natan Rubin. Stronger bounds for weak epsilon-nets in higher dimensions. *arXiv preprint arXiv:2104.12654, version 2*, 2023. URL https://arxiv.org/abs/2104.12654v2. Corrected and revised version of the STOC 2021 paper.

Haoyu Song, Thành Nguyen, and Young-San Lin. A few good choices. In *Proceedings of the 2026 Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 4861–4874. SIAM, 2026. doi: 10.1137/1.9781611978971.175.

Appendix

Contents

A Passing from Finite Populations to Measures

This appendix justifies the limit arguments used for the approximation results. The main point is to preserve the approximation constants while allowing centers to move with the finite population. We first give the argument for proper metric spaces, then record the compact version needed for arbitrary tight measures.

Lemma 3. *Let (M, d) be a nonempty proper metric space, let $k \geqslant 1$ and $\alpha \geqslant 1$, and let $0 \leqslant q < 1$. Suppose that, for every finite empirical probability measure ν on M and every finite set $A \subseteq M$, there exists a nonempty set $C \subseteq M$ of at most k centers such that $\nu(D_\alpha(C, y)) \leqslant q$ for every $y \in A$. Then every Borel probability measure μ on M admits a nonempty set $C \subseteq M$ of at most k centers such that $\mu(D_\alpha(C, y)) \leqslant q$ for every $y \in M$.*

Proof. Fix $o \in M$. The closed balls of integer radius centered at o are compact, so M is separable and complete. Choose a countable dense sequence $(y_j)_{j\geqslant 1}$ in M and finite empirical measures ν_n converging weakly to μ. Such empirical approximations can be obtained by assigning rational approximations to the masses of successively finer finite partitions of larger compact sets. For $A_n = \{y_1, \dots, y_n\}$, choose C_n as in the hypothesis and repeat entries to write $C_n = (c_{n,1}, \dots, c_{n,k})$.

Pass to a subsequence for each of the k coordinates. Properness ensures that each coordinate either converges to some $c_i \in M$ or satisfies $d(o, c_{n,i}) \to \infty$. Let I be the set of coordinates of the first kind. We claim that I is nonempty. Otherwise, for every fixed $R > 0$ and all sufficiently large n, every point of the open ball $B(o, R)$ would satisfy $\alpha d(x, y_1) < d(x, C_n)$. Choose R with $\mu(B(o, R)) > q$. The Portmanteau theorem gives $\liminf_n \nu_n(B(o, R)) \geqslant \mu(B(o, R)) > q$, contradicting the guarantee for y_1.

Set $C = \{c_i : i \in I\}$. On each bounded ball, the functions $d(\cdot, C_n)$ converge uniformly to $d(\cdot, C)$. To see this, distances to the convergent coordinates converge uniformly by the triangle inequality. Distances to every other coordinate diverge uniformly on the ball, whereas distance to any one convergent coordinate remains uniformly bounded there. Thus the latter coordinates eventually determine the minimum.

Suppose that $\mu(D_\alpha(C, y)) > q$ for some $y \in M$. Since this deviation set is open and is the increasing union, over positive integers s, of

$$U_s = \{x \in B(o, s) : d(x, C) > \alpha d(x, y) + 3/s\},$$

some s satisfies $\mu(U_s) > q$. Choose y_j with $\alpha d(y_j, y) < 1/s$. For all sufficiently large n, uniform convergence on $B(o, s)$ gives $|d(x, C_n) - d(x, C)| < 1/s$ there, and hence $U_s \subseteq D_\alpha(C_n, y_j)$. The Portmanteau theorem gives $\nu_n(U_s) > q$ for all sufficiently large n. Taking also $n \geqslant j$ contradicts the defining guarantee of C_n. □

Lemma 4. *Let K and L be compact subsets of a metric space M, let $k \geqslant 1$ and $\alpha \geqslant 1$, and let $q \geqslant 0$. Suppose that every finite empirical probability measure ν supported on K admits a nonempty set $C \subseteq L$ of at most k centers satisfying $\nu(D_\alpha(C, y)) \leqslant q$ for every $y \in M$. Then every Borel probability measure supported on K admits such a center set in L.*

Proof. Let ν_n be empirical probability measures on K converging weakly to μ, and choose corresponding center sets $C_n \subseteq L$. Pad them to k-tuples. Compactness of L^k gives a subsequence converging coordinatewise to a tuple with underlying set C. The triangle inequality gives uniform convergence of $d(\cdot, C_n)$ to $d(\cdot, C)$ on M. If $\mu(D_\alpha(C, y)) > q$, some $t > 0$ satisfies $\mu(\{x : d(x, C) > \alpha d(x, y) + t\}) > q$. This is an open set contained in $D_\alpha(C_n, y)$ for all sufficiently large n. The Portmanteau theorem gives a contradiction. □

The finite guarantees needed for Lemma 3 require no uniform convergence theorem for an entire range family. For GreedyCapture in an arbitrary proper metric, use the finite candidate set A itself, adding one arbitrary candidate if it is empty, and apply Chen et al. (2019). For its Euclidean factor of 2, use the theorem allowing centers anywhere in Euclidean space of Micha & Shah (2020). Both give the hypothesis with $q = 1/k$ when $k \geqslant 2$; the case $k = 1$ is immediate under Definition 1. For the stronger existential factor of 2, Cookson et al. (2026a) allow an arbitrary candidate set containing the finite population and give the hypothesis with $q = 1/(k + 1)$. Moreover, their centers belong to the population itself, which allows Lemma 4 with $L = K$.

For completeness, we retain the more general geometric condition used in the earlier approximation argument. It can be useful when bounded sets are not compact.

Lemma 5. *Let M be a nonempty metric space such that, for every compact $K \subseteq M$, there is a nonempty compact $L \subseteq M$ with the following property: for every $y \in M$, some $z \in L$ satisfies $d(x,z) \leqslant d(x,y)$ for all $x \in K$. Then every tight Borel probability measure on M admits a $(1, 4.911)$-core for every positive integer k.*

Proof. Write $a = 4.9108$ and $b = 4.911$. First suppose that the measure is supported on a compact set K, and choose L as in the hypothesis. For a finite empirical population on K, take a finite $1/r$-net A_r of L and apply the finite committee theorem to distance rankings on A_r. It returns centers $C_r \subseteq L$ for which no candidate in A_r attracts more than an a/k fraction of the population. Extract a convergent subsequence in L^k, with limiting center set C.

If a challenger $y \in M$ attracted a fraction greater than a/k from C, its replacement $z \in L$ would attract at least the same voters. There are finitely many such voters, so their strict distance improvements have a positive minimum margin. A net point approaching z would preserve all these improvements for the corresponding center sets C_r once r is sufficiently large, a contradiction. Thus every finite empirical population on K has centers in L with the bound a/k against every challenger in M. Lemma 4 gives the same bound for the measure supported on K.

Now let μ be tight. If $k \leqslant b$, any center suffices. Otherwise choose a compact set K whose omitted mass ε is less than $(b-a)/(k-a)$, and apply the compact case to the conditional probability measure on K. For the resulting centers and every $y \in M$,

$$\mu(D_1(C,y)) \leqslant (1-\varepsilon)\frac{a}{k} + \varepsilon = \frac{a}{k} + \varepsilon\left(1 - \frac{a}{k}\right) < \frac{b}{k}.$$

The strict difference between the cited constant and its rounded value is what permits this last step without assuming properness. ☐

B Details for the Sampling Guarantees

The upper and lower bounds concern the same issue: a center set is selected using observations and then evaluated against all alternatives in the population. The upper bound controls the entire family of possible comparisons. The lower bound uses the freedom to choose a particular empirical core after observing the sample.

B.1 The Geometry and Complexity of Deviation Regions

Definition 3 (Family of deviation regions)**.** For a fixed cost factor $\alpha \geqslant 1$ and budget k, let $\mathcal{D}_{\alpha,k}$ be the family of all sets $D_\alpha(C,y)$ with $1 \leqslant |C| \leqslant k$ and $C \cup \{y\} \subseteq \mathbb{R}^D$.

Lemma 6. *The VC dimension of $\mathcal{D}_{\alpha,k}$ is $O(Dk\log(k+1))$. For $\alpha = 1$, the family consists exactly of the empty set and all intersections of between one and k open halfspaces.*

Proof of Lemma 6. For $c \neq y$, squaring $\|x-y\|_2 < \|x-c\|_2$ gives the open halfspace $2\langle c-y, x\rangle < \|c\|_2^2 - \|y\|_2^2$. Thus every nonempty exact deviation region is an intersection of between one and k open halfspaces. Conversely, take a nonempty intersection $P = \bigcap_{i=1}^r \{x : \langle a_i, x\rangle < b_i\}$ with $1 \leqslant r \leqslant k$ and $a_i \neq 0$. Choose $y \in P$ and reflect it across each bounding hyperplane, obtaining

$$c_i = y + \frac{2(b_i - \langle a_i, y\rangle)}{\|a_i\|_2^2} a_i.$$

Expanding the comparison with c_i recovers exactly the ith halfspace. Therefore $D(\{c_1, \ldots, c_r\}, y) = P$. The empty set is obtained by taking $y \in C$. An intersection with no halfspaces would be the entire space and is not included.

When $\alpha > 1$ and $c \neq y$, completing the square shows that $\{x : \alpha\|x-y\|_2 < \|x-c\|_2\}$ is the open ball with center $(\alpha^2 y - c)/(\alpha^2 - 1)$ and radius $\alpha\|c-y\|_2/(\alpha^2-1)$. For $c = y$ it is empty. Open halfspaces and open balls have VC dimension at most $D+2$. If a base family of VC dimension v is intersected at most k times, its number of traces on $N \geqslant v$ points is at most $(eN/v)^{vk}$, by the Sauer–Shelah bound and padding intersections by repetitions. Shattering N points would imply $2^N \leqslant (eN/v)^{vk}$, which forces $N = O(vk\log(k+1))$. Taking $v = D+2$ proves the claim. □

The relative approximation theorem used in Equation (4) states that a class of VC dimension v satisfies that inequality simultaneously for all its members with probability at least $1-\zeta$ when the sample size is at least a constant times $(v\log(2/p) + \log(1/\zeta))/(\eta^2 p)$ (Har-Peled & Sharir, 2011). Applying it to the class just described, with $p = \beta/k$, yields Equation (2). This use of a uniform theorem is essential: concentration for one fixed center set would not justify choosing the centers from the same sample.

Proof of Theorem 3. Set $p = \beta/k$. The relative approximation theorem and Lemma 6 imply that, with the stated probability,

$$\left|\widehat{\mu}_P(A) - \mu(A)\right| \leqslant \eta \max\{\mu(A), p\} \qquad \text{for every } A \in \mathcal{D}_{\alpha,k}. \tag{4}$$

If some deviation from an empirical (α, β)-core had population mass greater than $p/(1-\eta)$, its empirical mass would exceed p by Equation (4), a contradiction. Conversely, a deviation of population mass at most p has empirical mass at most $(1+\eta)p$. The event is uniform over the entire family, so both implications hold for every permitted clustering. □

B.2 A Planar Lower Bound at a Fixed Error

The construction places a very small amount of mass at each of many rare locations. The centers are adjusted outward only at locations absent from the sample. The adjustment is small enough that all sampled locations captured by any one alternative still lie in a semicircle, but large enough that the origin captures every unseen location. We give both geometric statements before estimating their probabilities.

Proof of Theorem 4. Let $m = 2\left\lfloor \frac{k-1}{2} \right\rfloor$. We prove the stronger statement that the two conclusions hold with probability at least

$$1 - m\exp\left(-\frac{9n}{640k}\right) - \left(\frac{3e}{20}\right)^{k/10}.$$

Let $\xi = 1/(100m^2)$ and put $r = 1+\xi$. For $j = 0, \dots, m-1$, set $u_j = (\cos(2\pi j/m), \sin(2\pi j/m))$ and $e_j = ru_j$. Give each e_j mass $q = 3/(2km)$, and give $b = (10, 0)$ the remaining mass $1 - 3/(2k)$. The integer m is even, $k - 2 \leqslant m \leqslant k-1$, and $m \geqslant 9k/10$ for $k \geqslant 20$.

Let T be the set of indices not observed in the sample. Choose b as a center and, for each j, choose

$$c_j = \begin{cases} (2+4\xi)u_j, & j \in T, \\ 2u_j, & j \notin T. \end{cases}$$

If k is even, add the center $(20, 0)$; otherwise no additional center is needed. These are exactly k distinct centers. The rule depends only on which of finitely many sample locations occur, so it is measurable.

Consider any alternative y. The heavy point b cannot deviate because it is itself a center. At a sampled rare location e_j, the designated center $2u_j$ is at distance $1-\xi$. Therefore, if e_j deviates, then

$$\|ru_j - y\|_2^2 < (1-\xi)^2, \qquad \text{and hence} \qquad 2r\langle u_j, y\rangle > \|y\|_2^2 + 4\xi > 0. \tag{5}$$

For nonzero y, all captured sampled directions consequently lie in an open semicircle. Such a semicircle contains at most $m/2$ consecutive directions, and its directions are contained in one of the m cyclic blocks of $m/2$ consecutive indices. Each block has population mass $p = 3/(4k)$. Bernstein's inequality for its empirical mass, with excess $t = 3/(20k)$, gives

$$\Pr\left[\widehat{\mu}_P(\text{block}) > \frac{9}{10k}\right] \leqslant \exp\left(-\frac{nt^2}{2(p+t/3)}\right) = \exp\left(-\frac{9n}{640k}\right).$$

A union bound over the m blocks shows that, except with probability $m\exp(-9n/(640k))$, every alternative captures empirical mass at most $9/(10k)$. This also covers $y = 0$, which captures no sampled rare point by Equation (5).

We next check that 0 captures every unseen point e_j. For its own center, $2 + 4\xi > 2r$, so $\|e_j - c_j\|_2 > r = \|e_j\|_2$. Every other center associated with a rare point has the form ρu_i with $\rho \geqslant 2$. For $i \neq j$, the largest possible inner product $\langle u_i, u_j\rangle$ is $\cos(2\pi/m)$. Since $\sin(\pi/m) \geqslant 2/m$,

$$1 - \cos(2\pi/m) = 2\sin^2(\pi/m) \geqslant \frac{8}{m^2}, \qquad r\cos(2\pi/m) < 1.$$

It follows that $\|ru_j - \rho u_i\|_2^2 - r^2 = \rho(\rho - 2r\langle u_i, u_j\rangle) > 0$. The centers at $(10, 0)$ and, when present, $(20, 0)$ are farther than r by the triangle inequality. Thus all centers are farther from e_j than the origin is. A sampled e_j does not prefer the origin because its designated center is at distance $1 - \xi < r$. Consequently, $\mu(D(C_P, 0)) = q|T|$.

Let N_R be the number of sample observations at rare locations, counted with multiplicity. This is binomial with mean $3n/(2k) \leqslant 3k/200$. The usual Chernoff bound gives

$$\Pr[N_R > k/10] \leqslant \left(\frac{e(3k/200)}{k/10}\right)^{k/10} = \left(\frac{3e}{20}\right)^{k/10}.$$

If $N_R \leqslant k/10$, at most that many rare locations were observed. Hence

$$q|T| \geqslant \frac{3}{2k}\left(1 - \frac{k}{10m}\right) \geqslant \frac{3}{2k}\left(1 - \frac{1}{9}\right) = \frac{4}{3k}.$$

Combining the two events proves the stated probability. The interval of permitted sample sizes is nonempty for all sufficiently large k. At its lower endpoint, $m\exp(-9n/(640k)) \leqslant k^{-29/16}$, so the success probability tends to one uniformly throughout that interval. $\square$

The measure in this proof has at most k support points. Choosing all its support points would therefore give zero population cost and an exact core. The result concerns the transfer of *every* empirical core, not the impossibility of choosing a good one. For any proposed sufficient sample threshold $o(k^2)$ with a guarantee for all larger sample sizes, taking a sample size near $k^2/100$ contradicts Theorem 4. This is the precise sense in which a quadratic threshold is necessary.

B.3 Transferring Clustering Cost

Fairness is separate from the sum of distances, but the latter can be transferred at the same time. For a measure ν with a finite first moment, write

$$\mathrm{cost}_\nu(C) = \int d(x, C)\, d\nu(x).$$

The Wasserstein distance $W_1(\mu, \nu)$ is the infimum, over couplings of the two measures, of the expected distance between coupled points. Since $x \mapsto d(x, C)$ is 1-Lipschitz, every coupling gives

$$|\mathrm{cost}_\mu(C) - \mathrm{cost}_\nu(C)| \leqslant W_1(\mu, \nu)$$

after taking the infimum.

Proposition 1. *Suppose μ has a finite first moment and a sample-dependent clustering C_P has total empirical distance at most ρ times the best total empirical distance achievable with at most k centers, where $\rho \geqslant 1$. Then*

$$\mathrm{cost}_\mu(C_P) \leqslant \rho \inf_{1 \leqslant |C| \leqslant k} \mathrm{cost}_\mu(C) + (\rho + 1) W_1(\mu, \widehat{\mu}_P).$$

Proof. Apply the Lipschitz bound first to C_P and then to an arbitrary population comparator with at most k centers. The two errors are $W_1(\mu, \widehat{\mu}_P)$ and $\rho W_1(\mu, \widehat{\mu}_P)$, respectively. Take the infimum over comparators. $\square$

The empirical Wasserstein distance tends to zero almost surely under the finite first-moment assumption. Its quantitative rate needs additional tail assumptions; precise bounds are given by Fournier & Guillin (2015). Thus a sample-based construction may simultaneously enjoy the fairness guarantee in Theorem 3 and the cost guarantee in Proposition 1, with their different assumptions made explicit.

C Existence for Sufficiently Many Centers

This appendix proves Theorems 5 and 6. For the positive result, we first show that every density in a given dimension has the same asymptotic deviation constant. We then bound that constant in dimensions up to five. All distances in this appendix are Euclidean, and centers are allowed throughout the ambient space. For vectors, $|\cdot|$ denotes the Euclidean norm.

C.1 One Nonatomic Measure With Infinitely Many Empty Cores

We prove Theorem 5 by repeating a robust finite obstruction at every scale. The local geometric lemma must allow arbitrary centers outside the region under consideration; otherwise, applying it inside a smaller copy would not be justified. After proving that lemma, we construct a singular measure with no atoms and add a uniform component to give full convex support.

Lemma 7. *Let T be the three vertices of an equilateral triangle of side $s > 0$ in the plane, and let $U = \bigcup_{v \in T} \overline{B}(v, s)$. Put $\tau = (2/\sqrt{3} - 1)s/4$. If a nonempty finite center set $C \subseteq \mathbb{R}^2$ has at most one point in U, then, for every $\varepsilon > 0$ with $2\varepsilon < \tau$, there are two vertices $u, v \in T$ and an alternative center y for which $B(u, \varepsilon) \cup B(v, \varepsilon) \subseteq D(C, y)$.*

Proof. Write $r_v = d(v, C)$. We first find two vertices with both radii at least $s/2$ and sum at least $2s/\sqrt{3}$. If two radii exceed s, they work. If exactly one exceeds s, the other two nearest centers must coincide with the unique center in U. Their distances sum to at least s, so one of them is at least $s/2$; pair that vertex with the one whose radius exceeds s. Their sum exceeds $3s/2 > 2s/\sqrt{3}$.

If all three radii are at most s, they are distances to a common center $z \in U$. The second-largest radius is at least $s/2$: otherwise the two smallest radii would sum to less than s, contradicting the triangle inequality between their vertices. Their sum is at least two thirds of the sum of all three radii. The latter sum is at least $\sqrt{3}s$: average z over the three rotations about the triangle's centroid and apply convexity of the sum of distances. The average is the centroid, where the sum is $\sqrt{3}s$. This proves the claimed bounds in every case.

For the resulting u, v, the open balls $B(u, r_u - \tau)$ and $B(v, r_v - \tau)$ have positive radii, and their radii sum to at least $2s/\sqrt{3} - 2\tau > s$. They intersect, so choose y in their intersection. For $x \in B(u, \varepsilon)$,

$$\|x - y\|_2 < r_u - \tau + \varepsilon < r_u - \varepsilon \leqslant d(x, C).$$

The same argument applies at v. No restriction was placed on the centers outside U. $\square$

Proof of Theorem 5. Let $Q = [0, 1]^2$ and fix $s = 1/100$. Take the triangle with vertices $(1/5, 1/5)$, $(1/5+s, 1/5)$, and $(1/5 + s/2, 1/5 + \sqrt{3}s/2)$. Translate it by $(1/2, 0)$ and by $(1/4, 1/2)$ to obtain two more triangles. Denote their nine vertices by $p_1, \ldots, p_9$. For each of the three triangles let U_j be the union of the closed radius-s balls about its vertices. The displayed coordinates give three pairwise disjoint sets U_j lying in the interior of Q.

Take τ from Lemma 7, and set $\varepsilon = \tau/4$ and $\lambda = \varepsilon/(4\sqrt{2})$. Define contractions $f_i(x) = p_i + \lambda x$ for $i = 1, \ldots, 9$. Each square $f_i(Q)$ is contained in Q and in $B(p_i, \varepsilon)$, because its points are at distance at most $\sqrt{2}\lambda = \varepsilon/4$ from p_i. Distinct vertices are at least distance s apart, whereas $\varepsilon/4 < s/4$, so these nine squares are disjoint.

Let $I_1, I_2, \ldots$ be independent uniform indices in $\{1, \ldots, 9\}$, and let ν be the distribution of the almost surely convergent series $\sum_{j \geqslant 1} \lambda^{j-1} p_{I_j}$. Its compact support S satisfies $S = \bigcup_i f_i(S)$. For a word $w = (i_1, \ldots, i_t)$, write $f_w = f_{i_1} \circ \cdots \circ f_{i_t}$. Disjointness implies that $\nu(f_w(S)) = 9^{-t}$ and that the sets at each fixed level are disjoint. Their diameters tend to zero; a point therefore has mass at most 9^{-t} for every t, so ν has no atoms. Also, S is covered by 9^t squares of area λ^{2t}. Since $9\lambda^2 < 1$, it has area zero and ν is singular.

Fix $t \geqslant 0$ and let C be any nonempty set of at most $5 \cdot 9^t$ centers. The $3 \cdot 9^t$ sets $f_w(U_j)$, over words of length t and the three triangles, are pairwise disjoint: within a word they inherit the disjointness of the U_j, and across words they lie in disjoint squares $f_w(Q)$. At least one of these neighborhoods has at most one center, since otherwise at least $6 \cdot 9^t$ centers would be needed.

Apply Lemma 7 to that scaled triangle, with side $s\lambda^t$ and ball radius $\varepsilon\lambda^t$. A common challenger captures the balls about two of its vertices. Each ball contains a child copy $f_w(f_i(S))$ of mass $9^{-(t+1)}$. Thus the challenger captures ν-mass at least $2 \cdot 9^{-(t+1)}$. The lemma applies to the full global center set, including every center outside the selected neighborhood.

Let $\mu = (19/20)\nu + (1/20)\lambda_Q$, where λ_Q is the uniform probability measure on Q. This measure has no atoms and has support exactly Q. At budget $k = 5 \cdot 9^t$, every clustering admits a deviation of mass at least $(19/20)2 \cdot 9^{-(t+1)} = 19/(18k) > 1/k$, so the exact core is empty. ☐

The same obstruction can be placed in any convex body of affine dimension at least two. Embed a sufficiently small square in a two-dimensional affine plane through the body's relative interior, put a copy of the singular measure there, and mix it with normalized intrinsic volume of the body. For any ambient center set, orthogonal projection onto the plane weakly decreases every distance from a point in the singular component. A deviation that beats all projected centers therefore also beats all original centers. The population mass calculation is unchanged, and the volume component gives full support on the chosen body.

C.2 The Constant Shared by All Densities

It is useful to perform the constructions on parts of a probability measure without renormalizing their masses. The following definition extends Definition 2 accordingly and describes the repeated center arrangements used in the proof.

Definition 4 (Finite measures and periodic arrangements)**.** For a finite nonnegative Borel measure ν on $\mathbb{R}^D$, let

$$\Delta(\nu, k) = \inf_{1 \leqslant |C| \leqslant k} \sup_{y \in \mathbb{R}^D} \nu(D_1(C, y)).$$

A full-rank lattice has the form $\Lambda = B\mathbb{Z}^D$ for an invertible matrix B; the volume of a fundamental cell is $\det \Lambda = |\det B|$. A periodic point set is a set $P = \bigcup_{i=1}^{q} (\Lambda + p_i)$, where Λ is a full-rank lattice and the p_i are distinct modulo Λ. Its number of points per unit volume is $\lambda(P) = q/\det \Lambda$. Define

$$\gamma(P) = \lambda(P) \sup_{y \in \mathbb{R}^D} \mathrm{vol}_D\{x : |x - y| < d(x, P)\}. \tag{6}$$

Write $A = 4911/1000$ for the universal mass bound from Theorem 2.

Thus Δ scales linearly with the mass of its first argument, and $\gamma(P)$ compares the largest volume captured by one additional center with the volume available per existing center. For a lattice, this is the constant used in Lemma 14. A periodic point set has finite covering radius and positive minimum separation. Consequently its captured regions are bounded, and all quantities in Equation (6) are finite.

Two elementary facts will be used repeatedly. Measure domination implies monotonicity of Δ. In addition, if $\nu \leqslant \rho + \sigma$, then

$$\Delta(\nu, n + m) \leqslant \Delta(\rho, n) + \Delta(\sigma, m). \tag{7}$$

Indeed, the union of two center sets only decreases their respective deviation regions. Apply the domination to that union, bound each summand by the corresponding largest deviation, and then use arbitrarily close approximations to the two infima. No infimum needs to be attained. The mass guarantee in Theorem 2 gives $\Delta(\rho, n) \leqslant A\rho(\mathbb{R}^D)/n$.

Lemma 8 (Placing periodic arrangements according to the density)**.** *For every finite absolutely continuous measure ν on $\mathbb{R}^D$ and every periodic point set P, one has* $\limsup_{k\to\infty} k\Delta(\nu, k) \leqslant \gamma(P)\nu(\mathbb{R}^D)$.

Proof. First suppose ν has density $\sum_{j=1}^{J} a_j \mathbf{1}_{Q_j}$, where $a_j > 0$ and the Q_j are finitely many closed cubes at positive pairwise distance. Write $M = \nu(\mathbb{R}^D) > 0$. For a parameter $t > 0$, scale P on cube Q_j by a factor $h_j = (\lambda(P)t/a_j)^{1/D}$. The resulting arrangement has a_j/t points per unit volume and captured volume at most $\gamma(P)t/a_j$. Select every point of this arrangement whose closed Voronoi cell meets Q_j. A nearest point of the full arrangement is then selected at every point of the cube.

The covering radius on each cube is $O(t^{1/D})$. For small enough t, a challenger cannot capture points from two cubes: two such points would be closer than twice the largest covering radius, contradicting the fixed separation of the cubes. Every deviation therefore has ν-mass at most $\gamma(P)t$.

The number of selected points on cube Q_j is $a_j \operatorname{vol}_D(Q_j)/t + O(t^{-(D-1)/D})$. To justify this estimate, each selected point is within one covering radius of the cube; count separately in the finitely many translates of the period lattice, using their fundamental cells. The union of those fundamental cells is contained in a neighborhood of the cube of width $O(h_j)$ and contains the cube apart from such a neighborhood of its boundary. The difference of these volumes is $O(h_j)$, which gives the stated estimate after division by the cell volume. Summing over the fixed collection of cubes gives $M/t + o(1/t)$ points. For every fixed $\eta > 0$, the choice $t = (1+\eta)M/k$ consequently uses at most k points for all sufficiently large k. Letting η tend to zero proves the lemma for these step densities.

Now let ν have arbitrary integrable density f and positive total mass M. For $0 < \varepsilon < \min\{1/4, M\}$, choose a nonnegative step function g supported on finitely many positively separated closed cubes such that $\|f - g\|_1 < \varepsilon$. Such functions are dense in the nonnegative part of L^1: first approximate by a compactly supported continuous function, approximate that function on a fine finite grid, and finally shrink the finitely many grid cubes slightly. The shrinking changes the integral by an arbitrarily small amount and separates the cubes. Let $d\sigma = g\,dx$ and $d\rho = (f-g)_+\,dx$. Then $\nu \leqslant \sigma + \rho$, $\sigma(\mathbb{R}^D) \leqslant M + \varepsilon$, and $\rho(\mathbb{R}^D) < \varepsilon$.

Give $\lfloor (1-\sqrt{\varepsilon})k \rfloor$ centers to σ and the remaining centers to ρ. The result for step densities, Equation (7), and the universal bound imply

$$\limsup_{k\to\infty} k\Delta(\nu, k) \leqslant \frac{\gamma(P)(M+\varepsilon)}{1-\sqrt{\varepsilon}} + A\sqrt{\varepsilon}.$$

Letting ε tend to zero proves the claim. The zero measure is immediate. □

For each positive integer D, let $b_D = \liminf_{k\to\infty} k\delta(\mu, k)$ when μ is the uniform probability measure on $[0,1]^D$ and distance is Euclidean.

We next show that the infimum over periodic arrangements is attained asymptotically by a uniform cube. The purpose of the additional points in the proof is to prevent one challenger from combining mass across several copies of the cube.

Lemma 9 (The uniform cube)**.** *Let λ_Q be uniform probability on the unit cube $Q = [0,1]^D$. Then* $\lim_{k\to\infty} k\delta(\lambda_Q, k) = b_D = \inf_{P \text{ periodic}} \gamma(P)$*, and* $1/2 \leqslant b_D \leqslant A$.

Proof. Let $\ell = \liminf_n n\delta(\lambda_Q, n)$. The universal bound and Lemma 1 give $1/2 \leqslant \ell \leqslant A$. Along a subsequence realizing this limit inferior, choose a set C_n of at most n points whose largest deviation has mass at most $s_n = \delta(\lambda_Q, n) + n^{-2}$. Project its points onto Q; this only reduces distances on Q. In particular, $s_n = \Theta(1/n)$ along the chosen subsequence.

Add the midpoint grid with $L_n = \lceil n^{1/(2D)} \rceil$ positions per coordinate. It uses $O(\sqrt{n}) = o(n)$ points and covers Q at radius $R_n = \sqrt{D}/(2L_n)$. Repeat the enlarged configuration in every translate of Q by $\mathbb{Z}^D$. This gives a periodic point set with at most $n + o(n)$ points per unit volume and covering radius $R_n \to 0$.

Let S be the union of the hyperplanes on which at least one coordinate is an integer, and put $U_n = \{x : d(x, S) \leqslant 2R_n\}$. Choose the integer $H_n = \lceil (D^D/s_n)^{1/D} \rceil$. Add every point of $H_n^{-1}\mathbb{Z}^D$ whose closed Voronoi cube meets U_n. This selection is periodic with period lattice $\mathbb{Z}^D$. The full fine lattice has covering radius $\sqrt{D}/(2H_n)$, so each of its captured regions is contained in a cube of volume $D^{D/2}H_n^{-D} \leqslant s_n$.

The added points have number per unit volume $O(R_n H_n^D + H_n^{D-1})$. Indeed, every selected point lies within $2R_n + \sqrt{D}/H_n$ of a coordinate face of a unit cube, and counting grid points in these finitely many strips gives the bound. Since $H_n^D = \Theta(n)$, this number is $o(n)$. Denote the complete periodic arrangement by P_n.

Every point captured by y lies within distance R_n of y. If $d(y, S) > R_n$, all captured points lie in one open unit cube, where the translated set C_n bounds their volume by s_n. If $d(y, S) \leqslant R_n$, all captured points lie in U_n. At such a point, some nearest point of the full fine lattice was selected. Hence the entire captured region is contained in a captured region of that full lattice and again has volume at most s_n. The boundaries of the cubes have zero volume. Therefore $\gamma(P_n) \leqslant (n + o(n))s_n \to \ell$, and $\inf_{P \text{ periodic}} \gamma(P) \leqslant \ell$.

On the other hand, Lemma 8, followed by the infimum over periodic P, gives $\limsup_n n\delta(\lambda_Q, n) \leqslant \inf_{P \text{ periodic}} \gamma(P)$. By construction, $b_D = \ell$. The two inequalities prove the limit, the periodic characterization, and the claimed bounds. ☐

Lemma 10 (Separated cubes)**.** *If ν has density $\sum_{j=1}^{J} a_j \mathbf{1}_{Q_j}$ on positively separated closed cubes, where $a_j > 0$, then $\lim_{k\to\infty} k\Delta(\nu, k) = b_D \nu(\mathbb{R}^D)$.*

Proof. The upper bound follows from Lemma 8. For the lower bound, put $p_j = a_j \operatorname{vol}_D(Q_j)$ and choose center sets with largest deviation at most $s_k = \Delta(\nu, k) + k^{-2} = O(1/k)$. Add to each cube a midpoint grid with $\lceil k^{1/(2D)} \rceil$ positions per coordinate. These grids use $o(k)$ points in total and provide a common covering radius $R_k \to 0$ on the cubes.

For all sufficiently large k, the closed R_k-neighborhoods of the cubes are disjoint. Let $C_{j,k}$ consist of the enlarged center set's points in the neighborhood of Q_j, and let $n_{j,k} = |C_{j,k}|$. At every point of Q_j, a nearest center lies in $C_{j,k}$, so its distances to the local and global center sets agree. It follows that $p_j \delta(\lambda_Q, n_{j,k}) \leqslant s_k$. Moreover, $\sum_j n_{j,k} \leqslant k + o(k)$.

By Lemma 1, $n_{j,k} \geqslant p_j/(2s_k) \to \infty$. Thus Lemma 9 implies that, for every fixed $\eta > 0$ and all sufficiently large k, $(b_D - \eta)p_j \leqslant s_k n_{j,k}$ for every j. Summing gives $(b_D - \eta)\nu(\mathbb{R}^D) \leqslant s_k(k + o(k))$. Let k tend to infinity and then let η tend to zero. ☐

Proof of the asymptotic assertion in Lemma 2. The upper bound follows from Lemma 8. For the lower bound, write $d\mu = f\,dx$ and, for $0 < \varepsilon < 1/4$, choose a separated cubical step density g with $\|f - g\|_1 < \varepsilon$. Let $d\nu = g\,dx$, $M = \nu(\mathbb{R}^D)$, and $d\rho = (g - f)_+\,dx$. Then $M \geqslant 1 - \varepsilon$, $\rho(\mathbb{R}^D) < \varepsilon$, and $\nu \leqslant \mu + \rho$.

Set $m_k = \lceil \sqrt{\varepsilon}\, k/(1 - \sqrt{\varepsilon}) \rceil$. By Equation (7), $\Delta(\nu, k + m_k) \leqslant \delta(\mu, k) + A\varepsilon/m_k$. Multiply by k and use Lemma 10 to obtain

$$\liminf_{k\to\infty} k\delta(\mu, k) \geqslant (1 - \sqrt{\varepsilon})\big(b_D(1 - \varepsilon) - A\sqrt{\varepsilon}\big).$$

Letting ε tend to zero proves the equality. The numerical bounds in Equation (3) are proved next. ☐

C.3 Bounds in Dimensions Up to Five

The asymptotic result reduces exact existence to finding a repeated arrangement for which the volume captured by one additional center is smaller than the volume per existing center. The integer lattice has this property in the first five dimensions. The first three bounds are analytic; the last two use an exact finite calculation over all possible locations of the additional center.

Lemma 11 (Integer lattices in low dimension)**.** *The integer lattices satisfy $\gamma(\mathbb{Z}) = 1/2$, $\gamma(\mathbb{Z}^2) = 9/16$, and $\gamma(\mathbb{Z}^3) \leqslant 7/8$.*

Proof. An additional point $a \in (0, 1)$ on the line captures precisely the interval $(a/2, (a + 1)/2)$, of length $1/2$; a lattice point captures nothing. The planar claim is Lemma 14 at aspect ratio one.

For dimension three, use the functions $\phi_a(t) = d(t, \mathbb{Z})^2 - (t - a)^2$ and $J_a(s) = \int_{\mathbb{R}} (\phi_a(t) + s)_+\,dt$ from Lemma 13. By reflections, translations, and a permutation of coordinates, a challenger has coordinates

$(a,b,c) \in [0,1/2]^3$ with $u = a(1-a) \leqslant v = b(1-b) \leqslant w = c(1-c)$. If $w = 0$, it is a lattice point. Otherwise, integrate first in the third coordinate and use the width bound in Lemma 13 to get

$$\mathrm{vol}_3\{x : |x-(a,b,c)| < d(x,\mathbb{Z}^3)\} \leqslant \frac{1}{2w}\int_{\mathbb{R}^2} (w + \phi_a(s) + \phi_b(t))_+ \, ds\, dt. \tag{8}$$

Since $w + \phi_a(s) \leqslant w + u \leqslant 1/2$, the last inequality of Lemma 13 bounds the double integral by $J_a(w) + J_a(w+v)/4$. Its sharp bound on J_a applies because $u \leqslant w \leqslant 1/4$ and $u \leqslant w + v \leqslant 2w \leqslant 1/2$. Consequently the double integral is at most $w + w^2/2 + (2w + 2w^2)/4 = 3w/2 + w^2$. Substitution into Equation (8) gives at most $3/4 + w/2 \leqslant 7/8$. ☐

Proposition 2 (Exact calculation for dimensions four and five)**.** *One has $\gamma(\mathbb{Z}^4) < 19/20$ and $\gamma(\mathbb{Z}^5) < 99/100$. These inequalities are certified by exact rational calculations that cover every real challenger location.*

Proof. Translations, sign changes, and permutations reduce the challenger to a point of $[0,1/2]^D$ with ordered coordinates. For a coordinate interval $[l,u] \subseteq [0,1/2]$, where $l < u$, consider the following common upper bound for $\phi_a(t)$ over all $a \in [l,u]$:

$$\psi_{l,u}(t) = \begin{cases} \phi_l(t), & t \leqslant l, \\ (l+u)t - lu, & l \leqslant t \leqslant u, \\ \phi_u(t), & t \geqslant u. \end{cases} \tag{9}$$

Outside $[l,u]$, the nearest endpoint maximizes $\phi_a(t)$ as a function of a. Inside this interval, the exact maximum is t^2 and the displayed chord is at least t^2, because $(t-l)(u-t) \geqslant 0$. Therefore a box $\prod_i [l_i, u_i]$ of challenger locations is bounded by the single volume $\mathrm{vol}_D\{x : \sum_i \psi_{l_i,u_i}(x_i) > 0\}$.

We describe how that volume is evaluated exactly. The maximum of $\psi_{l,u}$ is $U = u(1-u)$, at $t = 1/2$. Push Lebesgue measure on the line forward under the map $t \mapsto U - \psi_{l,u}(t)$. Up to height two, its continuous part is given by the following intervals and densities, omitting intervals of zero length:

Height interval	Density
$[0, 2(1-u)]$	$1/(2(1-u))$
$[2(1-u), 2)$	$1/(2(2-u))$
$[0, u(1-2u)]$	$1/(2u)$
$[u(1-2u), U - l^2]$	$1/(l+u)$
$[U - l^2, U + l + l^2]$, when $l > 0$	$1/(2l)$
$[U + l + l^2, 2)$	$1/(2(1+l))$

When $l = 0$, the fifth row is replaced by an atom of mass $1/2$ at height U. The rows are added: the first two describe the right side of the maximum, and the remaining rows describe the left side. These formulas follow by taking the reciprocal absolute slopes of the affine pieces in Equation (9). The next omitted breakpoints are at height at least two, whereas the sum of the coordinate maxima is at most $D/4 \leqslant 5/4$.

Represent each bounded density interval as the difference of two constant densities on rays, and retain the atoms separately. In each term of the product over coordinates, let r be the number of ray densities, let c_i be its signed coefficients, and let a_i be the ray starting heights or atomic heights. With $T = \sum_i u_i(1-u_i)$, this term contributes

$$\left(\prod_i c_i\right) \frac{(T - \sum_i a_i)_+^r}{r!} \tag{10}$$

when $r \geqslant 1$. For $r = 0$, the contribution is $\prod_i c_i$ if $\sum_i a_i < T$, and zero otherwise. The strict inequality handles the atoms without changing the strict preference convention. Summing the finitely many terms gives the box's exact rational upper bound.

The initial coordinate partition has endpoints 0, 1/500, 1/200, 1/100, 1/50, 1/25, 7/100, 1/10, 3/20, 1/5, 1/4, 3/10, 7/20, 2/5, 9/20, and 1/2. Coordinate permutations reduce its boxes to combinations with repetition. Every one of these boxes in dimension four has the bound in Equation (10) strictly below 19/20. In dimension five, bisecting a widest coordinate interval as needed gives a finite refinement whose every box has bound strictly below 99/100. Each bisection covers its parent, and sorting coordinates only applies a symmetry, so no real challenger is omitted.

One can programmatically verify every terminal inequality with rational arithmetic using Equation (10). The resulting covers have 3060 boxes in dimension four and 15331 boxes in dimension five. Floating point calculations choose which boxes to refine, but every accepted box is rechecked exactly. The finite cover and its strict rational inequalities prove both claimed bounds. □

Since $b_D \leqslant \gamma(\mathbb{Z}^D)$ by Lemma 9, Lemma 11 and Proposition 2 prove all upper bounds in Equation (3). The lower bound $b_1 \geqslant 1/2$ from Lemma 9 makes the first one an equality.

Proof of Theorem 6. By Lemma 2, an absolutely continuous measure on $\mathbb{R}^D$ with $D \leqslant 5$ satisfies $\lim_{k\to\infty} k\delta(\mu, k) = b_D < 1$, proving the first item.

For the second item, write $\mu = \mu_{\mathrm{at}} + \rho$, where μ_{at} is supported on a countable set and $\rho(\mathbb{R}^D) = a < 1000/4911$. Select finitely many atoms whose omitted mass is less than ε and place centers at them. If this uses r centers, Theorem 2 bounds the infimum of the largest remaining deviation by $4911(a+\varepsilon)/(1000(k-r))$. Therefore $\limsup_k k\delta(\mu, k) \leqslant 4911a/1000 < 1$.

In each case, choose a constant strictly between the resulting limit superior and one. For all sufficiently large k, the infimum defining $\delta(\mu, k)$ is below this constant divided by k. Some finite center set therefore has every deviating mass strictly below $1/k$, proving the claim without requiring that the infimum be attained. □

D Exact Cores for Every Number of Centers

We prove the three cases of Theorem 7. Rectangles admit an analytic construction that combines a single row of centers with rectangular grids. Disks use the same grids except for two small budgets, which are covered by exact rational certificates. Throughout this section, distance is Euclidean and area is ordinary, unnormalized Lebesgue measure.

D.1 Measures Supported on a Line

The idea is to place quantiles so that no gap between consecutive centers contains too much mass. A deviating interval cannot cross a chosen center.

Lemma 12. *For every Borel probability measure μ supported on a line in $\mathbb{R}^D$ and every positive integer k, there are at most k centers with $\mu(D(C, y)) \leqslant 1/(k+1)$ for every $y \in \mathbb{R}^D$. Consequently, $\delta(\mu, k) \leqslant 1/(k+1)$, and this bound is sharp across measures.*

Proof. For a measure supported on a line, projecting any proposed alternative center onto that line weakly decreases its distance to every population point. We may therefore identify the line with $\mathbb{R}$ and restrict alternative centers to it.

Let F be the cumulative distribution function. Choose the lower quantiles $c_j = \inf\{t : F(t) \geqslant j/(k+1)\}$ for $j = 1, \ldots, k$, and remove repetitions. Each quantile is finite, $F(c_j) \geqslant j/(k+1)$, and $F(c_j^-) \leqslant j/(k+1)$. The two unbounded components of $\mathbb{R} \setminus C$ therefore have mass at most $1/(k+1)$. For consecutive distinct centers $a < b$, let r be the largest index with $c_r = a$. Then $c_{r+1} = b$, and

$$\mu((a, b)) = F(b^-) - F(a) \leqslant \frac{r+1}{k+1} - \frac{r}{k+1} = \frac{1}{k+1}.$$

Every deviation region is an interval disjoint from C, so it lies in a single such component. The equal measure on $k+1$ points discussed in Section 2 attains the bound. □

The uniform interval from Example 1 permits an even smaller deviating mass. With centers $(2j-1)/(2k)$ for $j = 1, \ldots, k$, every captured interval has length at most $1/(2k)$. Together with Lemma 1, this gives $\delta(\mu, k) = 1/(2k)$ for that measure.

D.2 A bound for rectangular lattices

Definition 5 (Distance comparisons on a lattice)**.** For a lattice $\Lambda \subseteq \mathbb{R}^2$, write $I_\Lambda(y) = \{x : \|x-y\|_2 < d(x,\Lambda)\}$. The following calculation bounds its area. For $a \in [0,1/2]$, put $\phi_a(x) = d(x,\mathbb{Z})^2 - (x-a)^2$, $u = a(1-a)$, and $J_a(s) = \int_{\mathbb{R}} (\phi_a(x)+s)_+ \, dx$, where $z_+ = \max\{z, 0\}$.

Lemma 13. *For $a \in [0,1/2]$, the preceding functions satisfy*

$$\begin{aligned}
\mathrm{vol}_1\{x : \phi_a(x) + q > 0\} &\leqslant \frac{(u+q)_+}{2u} && \text{if } a > 0,\\
J_a(s) &\leqslant s + \frac{s^2}{2} && \text{if } u \leqslant s \leqslant \frac{3}{4},\\
J_a(s) &\leqslant s + \frac{s^2}{2} + \frac{1}{24} && \text{if } u \leqslant s \leqslant 1,\\
J_a(q) &\leqslant q_+ + \frac{(q+u)_+}{4} && \text{if } q \leqslant \frac{1}{2}.
\end{aligned}$$

Proof. On $[j-1/2, j+1/2]$, the profile is $\phi_a(x) = 2(a-j)x + j^2 - a^2$. Equivalently, it is the minimum of these affine functions over $j \in \mathbb{Z}$. For $a > 0$, the two functions for $j = 0, 1$ form a tent of height u at $x = 1/2$, with slopes $2a$ and $-2(1-a)$. The tent majorizes ϕ_a and proves the width bound. Its integral of the positive part after adding s is $(u+s)_+^2/(4u)$.

When $a = 0$, direct integration over the central interval and its two neighboring intervals gives $J_0(s) = s + s^2/2$ for $0 \leqslant s \leqslant 2$, whereas $J_0(q) = 0$ for $q \leqslant 0$. All assertions follow. Suppose henceforth that $a > 0$, and put $A = a(1+a)$. For $u \leqslant s \leqslant A$, the tent bound suffices: the difference $s + s^2/2 - (u+s)^2/(4u)$ is concave in s, and its endpoint values are $u^2/2$ and $a^2(1-a-a^2-a^3)/(2(1-a))$, respectively. Both are positive because $a \leqslant 1/2$.

For $s > A$, retain the three affine functions for $j = -1, 0, 1$. Their minimum still majorizes the profile. Integrating its positive part yields

$$J_a(s) \leqslant \frac{(u+s)^2}{4u} - \frac{(s-A)^2}{4a(1+a)} = s + \frac{s^2}{2} + \frac{a^2(s^2+a^2-1)}{2(1-a^2)}.$$

The last term is nonpositive when $s \leqslant 3/4$, and is at most $a^4/(2(1-a^2)) \leqslant 1/24$ when $s \leqslant 1$. This proves the second and third bounds. Using a majorant makes the calculation valid even when its support extends into further lattice intervals.

For the last bound, $q \leqslant -u$ is immediate. If $-u \leqslant q \leqslant 0$, the tent gives $J_a(q) \leqslant (q+u)^2/(4u) \leqslant (q+u)/4$. If $0 \leqslant q \leqslant A$, it gives

$$q + \frac{q+u}{4} - J_a(q) \geqslant \frac{q(3u-q)}{4u} \geqslant 0,$$

where $A \leqslant 3u$ follows from $a \leqslant 1/2$. Finally, if $A < q \leqslant 1/2$, the second bound gives $J_a(q) \leqslant q + q^2/2 \leqslant q + q/4 \leqslant q + (q+u)/4$. $\square$

Lemma 14. *For $1 \leqslant r \leqslant \sqrt{3}$ and $\Lambda_r = \mathbb{Z} \times r\mathbb{Z}$,*

$$\sup_{y \in \mathbb{R}^2} \frac{\mathrm{area}(I_{\Lambda_r}(y))}{r} = \frac{1}{2} + \frac{r^2}{16}.$$

In particular, every insertion into a square lattice of spacing h captures area at most $9h^2/16$.

Proof. By lattice symmetries, write $y = (a, rb)$ with $a, b \in [0,1/2]$, and put $u = a(1-a)$ and $v = r^2 b(1-b)$. At $x = (z, rw)$, the difference of squared distances is $\phi_a(z) + r^2\phi_b(w)$. If $v \geqslant u$ and $v > 0$, the width bound in Lemma 13, applied first in w, gives

$$\frac{\mathrm{area}(I_{\Lambda_r}(y))}{r} \leqslant \frac{J_a(v)}{2v} \leqslant \frac{1}{2} + \frac{v}{4} \leqslant \frac{1}{2} + \frac{r^2}{16},$$

because $u \leqslant v \leqslant r^2/4 \leqslant 3/4$. If $v < u$, put $s = u/r^2$. Reversing the order of integration gives the bound $J_b(s)/(2s) \leqslant 1/2 + s/4 \leqslant 9/16$, since $b(1-b) < s \leqslant 1/4$. If $u = v = 0$, the challenger is a lattice point and captures nothing. These cases prove the upper bound. At $a = 0, b = 1/2$, the vertical width estimate is exact throughout the relevant region, so the normalized area is $J_0(r^2/4)/(2(r^2/4)) = 1/2 + r^2/16$. $\square$

D.3 Rectangles with any aspect ratio

For rectangles, the construction combines rows of evenly spaced centers with rectangular grids. Rows suffice when the rectangle is narrow relative to the available budget. Once enough centers are available to populate both directions, the grid spacing can be chosen so that each possible alternative captures less area than one proportional share. The proof tracks the integer number of rows and columns, which is necessary to cover every budget rather than just a subsequence of square budgets.

Projection of either a challenger or a selected center onto a closed convex support weakly decreases its distances from points in that support. Thus challengers may be restricted to the support, and selected centers may be projected into it after a construction. A midpoint grid on a rectangle also retains a nearest site of its full infinite lattice for every population point. Consequently Lemma 14 applies to finite midpoint grids without a boundary loss.

An elongated rectangle is more conveniently handled by a single row. The next bound leaves enough slack to combine that row with grids having an integer number of rows and columns.

Lemma 15. *For the row $h\mathbb{Z} \times \{0\}$ and the strip $\mathbb{R} \times [-W/2, W/2]$, where $0 < W \leqslant 2h$, every challenger in the strip captures area strictly below $7hW/12$.*

Proof. Scale so that $h = 1$. Translation and reflection reduce the challenger to (a, b) with $0 \leqslant a \leqslant 1/2$ and $0 \leqslant b \leqslant W/2$. If $b = 0$, each horizontal insertion interval has length at most $1/2$, so the captured area is at most $W/2$. Suppose $b > 0$, and put $t = b/W$, $q = b(W-b) = W^2 t(1-t)$, and $u = a(1-a)$. Here $t \leqslant 1/2$ and $q \leqslant 1$. Integrating the inequality $\phi_a(x) + 2bz - b^2 > 0$ first in z gives captured area

$$\frac{J_a(q) - J_a(-b(W+b))}{2b} \leqslant \frac{J_a(q)}{2b}.$$

If $u \leqslant q \leqslant 3/4$, Lemma 13 implies that the captured area divided by W is at most

$$\frac{1-t}{2} + \frac{W^2 t(1-t)^2}{4} \leqslant \frac{1}{2} + \frac{t}{2} - 2t^2 + t^3 \leqslant \frac{1}{2} + \frac{t}{2} - \frac{3}{2}t^2 \leqslant \frac{13}{24} < \frac{7}{12}.$$

If $q > 3/4$, then $q \leqslant 4t(1-t)$ forces $t > 1/4$. The third bound in Lemma 13 instead gives the strict upper bound

$$(1-t)\left(\frac{1}{2} + \frac{q}{4} + \frac{1}{48q}\right) < \frac{3}{4}\left(\frac{1}{2} + \frac{1}{4} + \frac{1}{36}\right) = \frac{7}{12}.$$

It remains to consider $q < u$, which implies $a > 0$. The tent width at height z is at most $(u + 2bz - b^2)_+/(2u)$. If $b(W+b) \leqslant u$, integration gives area at most $W(u-b^2)/(2u) \leqslant W/2$. Otherwise, put $v = u/(bW)$; then $1 - t < v < 1 + t$. Direct integration gives the upper bound $(v + 1 - t)^2/(8v)$ on the area divided by W. As a convex function of v, this is at most the larger endpoint value, namely $\max\{(1-t)/2, 1/(2(1+t))\} \leqslant 1/2$. Every case is strictly below $7/12$. $\square$

Lemma 16. *Uniform area on every nondegenerate Euclidean rectangle admits an exact core for every positive integer k.*

Proof. Normalize the rectangle to $[0, R] \times [-1/2, 1/2]$, where $R \geqslant 1$. Put $m = \lceil 7k/12 \rceil$. If $R \geqslant m/2$, place m equally spaced midpoint centers on the midline parallel to the longer side. Their spacing is $h = R/m \geqslant 1/2$. Every population point retains a nearest site of the infinite row, so Lemma 15 bounds every captured probability strictly below $7/(12m) \leqslant 1/k$. This proves the assertion for $k = 1, 2, 3$, since their values of m are $1, 2, 2$.

Suppose $k \geqslant 4$, initially excluding $k = 8$. For each $2 \leqslant n \leqslant N = \lfloor\sqrt{k}\rfloor$, put $m_n = \lfloor k/n\rfloor$ and $a_n = m_n/n$. Use the midpoint grid with m_n columns and n rows whenever $a_n/\sqrt{3} \leqslant R \leqslant \sqrt{3}a_n$. Its cell aspect ratio is at most $\sqrt{3}$, so Lemma 14 bounds the captured probability by $11/(16nm_n)$. Writing $k = nm_n + r$, where $0 \leqslant r \leqslant n-1$ and $m_n \geqslant n$, gives

$$\frac{nm_n}{k} \geqslant \frac{n^2}{n^2+n-1} \geqslant \frac{4}{5},$$

the last inequality being equivalent to $(n-2)^2 \geqslant 0$. Thus the captured probability is at most $55/(64k) < 1/k$.

These aspect intervals overlap. Indeed, for $n+1 \leqslant N$, one has $m_{n+1} \geqslant n+1$ and $m_n < ((n+1)/n)(m_{n+1}+1)$, whence

$$\frac{a_n}{a_{n+1}} < \frac{(n+1)^2}{n^2}\left(1+\frac{1}{m_{n+1}}\right) \leqslant \frac{(n+1)(n+2)}{n^2} \leqslant 3.$$

Their lowest endpoint is below 1: for $N \geqslant 3$, the inequality $k \leqslant N^2 + 2N$ gives $a_N \leqslant 1 + 2/N \leqslant 5/3 < \sqrt{3}$; for $N = 2$ and $k \neq 8$, directly $a_N \leqslant 3/2$. Their highest endpoint, $\sqrt{3}\lfloor k/2\rfloor/2$, reaches the row regime. For $k \geqslant 7$, this follows from

$$\left\lceil\frac{7k}{12}\right\rceil \leqslant \frac{7k+11}{12} \leqslant \frac{5(k-1)}{6} < \frac{\sqrt{3}(k-1)}{2} \leqslant \sqrt{3}\left\lfloor\frac{k}{2}\right\rfloor.$$

For $k = 4, 5, 6$, it follows respectively from $3 < 2\sqrt{3}$, $3 < 2\sqrt{3}$, and $4 < 3\sqrt{3}$. The grids and row construction therefore cover every $R \geqslant 1$.

Finally, for $k = 8$ and $1 \leqslant R \leqslant 5/2$, use three columns and two rows. The cell aspect ratio is at most $5/3 < \sqrt{3}$, and the captured probability is at most $11/96 < 1/8$. For $R \geqslant 5/2$, the row of five centers applies. □

D.4 Disks

For disks, a fine grid proves the claim once the budget is large enough, while smaller budgets use explicit arrangements adapted to the boundary. The remaining cases $k = 3$ and $k = 8$ are certified by exact inequalities covering the entire continuum of alternative centers. Lemma 18 separates the geometric reduction from these two computations.

Only two budgets require a separate construction. We first state precisely what their certificates establish and then explain why the finite computations cover all real challenger locations.

Lemma 17 (Computer-assisted)**.** *In the unit disk, each of the following sets of k centers makes every deviation region have area strictly below* $157/(50k)$*:*

$$\begin{aligned}
k = 3: \quad & C = \{(3/5, 0), (-3/10, 13/25), (-3/10, -13/25)\};\\
k = 8: \quad & C = \{(0,0), (13/20, 0)\}\\
& \qquad \cup \{(81/200, \pm 127/250), (-29/200, \pm 317/500), (-293/500, \pm 141/500)\}.
\end{aligned}$$

Proof. Let B be the closed unit disk. Restrict challengers to B by projection. Let P be the rational polygon specified by $|x_1|, |x_2| \leqslant 1$ and, for every choice of signs, $\pm x_1 \pm x_2 \leqslant 10/7$, $\pm 3x_1 \pm 4x_2 \leqslant 5$, and $\pm 4x_1 \pm 3x_2 \leqslant 5$. Cauchy–Schwarz gives $B \subseteq P$.

For a challenger box $q + [-h, h]^2$, comparison with a center c implies the relaxed inequality

$$2(c-q)\cdot x \leqslant \|c\|_2^2 - \|q\|_2^2 + 2h(\|q\|_1 + 10/7). \tag{11}$$

Indeed, writing $y = q + e$, we have $\|e\|_\infty \leqslant h$ and $\|x\|_1 \leqslant \sqrt{2} < 10/7$, and expansion bounds the change in the comparison of squared distances by $2h(\|q\|_1 + 10/7)$. Intersecting P with Equation (11) for every $c \in C$ therefore contains every captured point for every challenger in the box. The polygon's area is computed exactly by rational clipping and the shoelace formula.

Near a center c_i, omit its comparison and call the resulting polygon Q. Choose a rational $r_i > 0$ with $\|c_i\|_2 + r_i \leqslant 1$. Put $\varepsilon = \|q - c_i\|_1 + 2h$, and check $(2r_i + \varepsilon)^2 \leqslant \min_{j \neq i} \|c_i - c_j\|_2^2$. For every y in the box and $x \in B(c_i, r_i)$, the triangle inequality gives $\|x - y\|_2 \leqslant r_i + \varepsilon \leqslant \|c_i - c_j\|_2 - r_i \leqslant \|x - c_j\|_2$ for all $j \neq i$. Thus $B(c_i, r_i) \subseteq Q$. When $y \neq c_i$, strict improvement over c_i requires $(x - c_i) \cdot (y - c_i) > \|y - c_i\|_2^2/2$, excluding at least half of this ball. Hence captured area is at most $\operatorname{area}(Q) - 3r_i^2/2$, using $\pi > 3$. If $y = c_i$, capture is empty.

The verification procedure starts with $[-1, 1]^2$. It discards a box only when its minimum squared distance from the origin exceeds 1, when the ordinary polygon bound is strictly below $157/(50k)$, or when the preceding bound that omits one comparison is strictly below that threshold. Otherwise it replaces the box by its four closed dyadic children. All comparisons and polygon areas use rational arithmetic. The auxiliary radii are chosen among positive multiples of $1/1000$ satisfying $\|c_i\|_2 + 10r_i/9 \leqslant 1$ and $20r_i/9 \leqslant \min_{j \neq i} \|c_i - c_j\|_2$; these conditions are checked by squaring nonnegative quantities.

One can programmatically verify that, for each displayed configuration, every leaf passes. Thus this is a cover of the full challenger domain, including its boundary and the center locations. Finally, $\pi > 157/50$ follows from $\pi/4 = \arctan(1/2) + \arctan(1/3)$ and the four-term alternating lower bounds for the two arctangents: these give $\pi > 1538665/489888 > 157/50$. $\square$

Lemma 18. *Uniform area on every Euclidean disk admits an exact core for every positive integer k.*

Proof. Normalize to the unit disk. For $k = 1, 2$, the origin suffices: a nonzero challenger captures a circular cap strictly smaller than a semicircle. The two cases $k = 3, 8$ follow from Lemma 17 and $157/50 < \pi$.

For the next cases, use midpoint grids in $[-1, 1]^2$. Every point in the disk has a nearest site of the full rectangular lattice among the selected grid points. By Lemma 14, the following bounds hold; the middle entry specifies the number of columns and rows, and the last is ordinary area:

Budgets	Grid	Captured area
$4, 5$	2×2	$9/16 < \pi/5$
$6, 7$	2×3	$41/96 < \pi/7$
$9, 10, 11, 12$	3×3	$1/4 < \pi/12$
$13, 14$	3×4	$11/54 < \pi/14$

Projecting any exterior grid centers into the disk preserves these upper bounds.

For $k \geqslant 15$, take a square lattice of spacing h and retain the sites whose closed square Voronoi cells meet the disk. Averaging over its translation gives expected count $(\pi + 4h + h^2)/h^2$: the eligible sites lie in the Minkowski sum of the disk with $[-h/2, h/2]^2$, whose area is $\pi + 4h + h^2$. At $h_0^2 = 16\pi/(9k)$, this expectation is $9k/16 + 3\sqrt{k/\pi} + 1 < k + 1$. The inequality is equivalent to $k > 2304/(49\pi)$, which holds for $k \geqslant 15$ since $\pi > 157/50$. Choose $h < h_0$ sufficiently close that the expectation remains below $k + 1$. Some translation then retains at most k sites. Every population point retains a nearest lattice site, so Lemma 14 bounds captured area by $9h^2/16 < \pi/k$. Projection places all centers inside the disk. $\square$

Proof of Theorem 7. Lemmas 12, 16 and 18 establish the three cases, respectively. $\square$

E Exact Cores for Gaussians with At Most Four Centers

This appendix proves Theorem 8.

Definition 6 (Standard Gaussian notation)**.** The standard normal density is $\varphi(t) = (2\pi)^{-1/2} e^{-t^2/2}$, and its distribution function is $\Phi(t) = \int_{-\infty}^{t} \varphi(u)\, du$. We write $\overline{\Phi}(t) = 1 - \Phi(t)$ for its upper tail.

Put $a = \sqrt{2/\pi}$, the mean absolute value of a standard normal variable. This value cancels the first moment in the boundary comparison below. We first prove that the three centers $(-a, 0), (0, 0), (a, 0)$ leave every challenger at most one quarter of the standard bivariate Gaussian population. The proof treats challengers outside the circle of radius a using just the two outer centers, and then excludes an interior maximum for the three-center construction.

Lemma 19. *For independent standard normal coordinates (X, Z) and centers $(-a, 0), (a, 0)$, every challenger (s, t) with $s^2 + t^2 \geqslant a^2$ captures probability at most $1/4$.*

Proof. Reflect so that $t \geqslant 0$. When $t > 0$, put $r^2 = s^2 + t^2$, $A = (a^2 - r^2)/(2t)$, and $m_\pm = (a \mp s)/t$. The captured probability is

$$F(s,t) = \int_0^\infty \varphi(x)\{\Phi(A - m_+x) + \Phi(A - m_-x)\}\,dx.$$

Write $I_\pm = \int_0^\infty \varphi(x)\varphi(A - m_\pm x)\,dx$, $J_\pm = \int_0^\infty x\varphi(x)\varphi(A - m_\pm x)\,dx$, and $I = I_+ + I_-$. At a stationary point, differentiation gives $J_+ - J_- = sI$ and $(a-s)J_+ + (a+s)J_- = (a^2 - s^2 + t^2)I/2$. Thus $J_\pm = ((a \pm s)^2 + t^2)I/(4a)$. Integration by parts also gives

$$((a \mp s)^2 + t^2)J_\pm = \frac{(a \mp s)(a^2 - r^2)}{2} I_\pm + t^2\varphi(0)\varphi(A).$$

The two left sides agree. If $r \neq a$, it follows that $(a - s)I_+ = (a + s)I_-$. This is impossible for $|s| \geqslant a$. For $0 < s < a$, it contradicts strict increase in $m > 0$ of

$$m \int_0^\infty \varphi(x)\varphi(A - mx)\,dx = \int_0^\infty \varphi(u/m)\varphi(A - u)\,du.$$

Reflection handles $s < 0$. Hence every stationary point with $r > a$ lies on the perpendicular axis $s = 0$.

We show that $F(0,t)$ strictly decreases for $t \geqslant a$. Put $I_0 = \int_0^\infty \varphi(x)\varphi(A - ax/t)\,dx$ and $B = \varphi(0)\varphi(A)$. Differentiation and the preceding integration by parts give

$$\frac{d}{dt}F(0,t) = \frac{2aB - (3a^2 + t^2)I_0}{a^2 + t^2}, \qquad \frac{B}{I_0} = \frac{\sqrt{a^2 + t^2}}{t}\frac{\varphi(z)}{\Phi(z)}, \qquad z = \frac{a(a^2 - t^2)}{2t\sqrt{a^2 + t^2}}.$$

For $t \geqslant a$, we have $-a/2 \leqslant z \leqslant 0$. The ratio $M(z) = \varphi(z)/\Phi(z)$ decreases: its derivative is $-M(z)(z + M(z)) < 0$, where $z\Phi(z) + \varphi(z) = \int_{-\infty}^z (z - u)\varphi(u)\,du > 0$. Using $\varphi(0) = a/2$, $e^{-v} \leqslant (1 + v)^{-1}$, and $\Phi(-a/2) \geqslant 1/2 - a\varphi(0)/2$, we obtain

$$M(z) \leqslant M(-a/2) \leqslant a\frac{4\pi^2}{(4\pi + 1)(\pi - 1)} < \frac{18a}{13} < \sqrt{2}a \leqslant \frac{t(3a^2 + t^2)}{2a\sqrt{a^2 + t^2}}.$$

The rational function of π decreases for $\pi > 1$, so $\pi > 3$ proves the strict bound by $18a/13$. The final expression increases for $t \geqslant a$ and equals $\sqrt{2}a$ at $t = a$. This proves the required derivative sign.

On the circle $r = a$, the two comparisons have zero thresholds and orthogonal normal vectors, so their intersection has Gaussian probability $1/4$, except at an incumbent where it is empty. At infinity, capture tends uniformly to zero because a captured point satisfies $(sX + tZ)/r > (r^2 - a^2)/(2r)$. It remains to control approach to an incumbent from outside the circle. At $(a, 0)$, every limiting displacement direction has nonnegative first coordinate. The limiting event is an outward halfspace through $(a, 0)$ intersected with $\{X > 0\}$.

To bound these halfspaces, define $w(v) = \varphi(a+v) - \mathbf{1}_{v<a}\varphi(a-v)$ for $v > 0$. It is negative before a, positive after a, and $\int_0^\infty vw(v)\,dv = \varphi(0) - a/2 = 0$. For $h \geqslant 0$, the ratio $(\Phi(hv) - 1/2)/v$ decreases in v. Comparing it with its value at a on the two sign intervals of w gives $\int_0^\infty (\Phi(hv) - 1/2)w(v)\,dv \leqslant 0$. Equivalently,

$$\int_0^\infty \varphi(x)\Phi(h(x - a))\,dx \leqslant \frac{1}{4}.$$

This bounds every outward direction; purely axial directions follow by limits. It also bounds the capture of axis challengers beyond the outer centers, whose capture is a subset of $\{X > a\}$ or $\{X < -a\}$. The other incumbent is symmetric.

If the exterior supremum exceeded $1/4$, a maximizing sequence could neither escape to infinity nor approach the circle, the axis, or an incumbent. A subsequence would attain an interior maximum with $t > 0$ and $r > a$. The calculation at stationary points forces $s = 0$, where strict decrease gives $F(0,t) < F(0,a) = 1/4$, a contradiction. □

Lemma 20. *The centers $(-a,0)$, $(0,0)$, and $(a,0)$ give captured probability at most $1/4$ for the standard bivariate Gaussian.*

Proof. By Lemma 19, only challengers (s,t) with $s^2+t^2 < a^2$ need consideration. Reflect so that $t \geqslant 0$. For $t > 0$, let $c_j \in \{-a, 0, a\}$ and let V_j be its one-dimensional Voronoi interval, with boundaries $-a/2, a/2$. Put $r^2 = s^2 + t^2$ and

$$q_j(x) = \frac{c_j^2 - r^2 + 2(s-c_j)x}{2t}, \qquad I_j = \int_{V_j} \varphi(x)\varphi(q_j(x))\,dx, \qquad J_j = \int_{V_j} x\varphi(x)\varphi(q_j(x))\,dx.$$

The captured probability is $F(s,t) = \sum_j \int_{V_j} \varphi(x)\Phi(q_j(x))\,dx$. Write $I = \sum_j I_j$. At a stationary point, differentiation gives $\sum_j J_j = sI$ and $\sum_j c_j J_j = (r^2 I + \sum_j c_j^2 I_j)/2$. Let D_j be the left endpoint value minus the right endpoint value of $\varphi(x)\varphi(q_j(x))$ on V_j, with zero at infinite endpoints. Integration by parts gives

$$((s-c_j)^2 + t^2)J_j = \frac{(c_j - s)(c_j^2 - r^2)}{2} I_j + t^2 D_j.$$

Adjacent q_j agree at their common boundary, so $\sum_j D_j = 0$. Summing the identities, substituting the two stationary equations, and using $c_j^3 = a^2 c_j$ yields $(a^2 - r^2)K = a^2(sI_0 - 2J_0)$, where $K = \sum_j (c_j - s)I_j$ and the subscript 0 denotes the central interval. Its individual identity gives $r^2 J_0 = sr^2 I_0/2 + t^2 D_0$. Therefore

$$K = -\frac{2a^2t^2D_0}{r^2(a^2-r^2)}, \qquad D_0 = \varphi(a/2)\left\{\varphi\left(\frac{r^2+as}{2t}\right) - \varphi\left(\frac{r^2-as}{2t}\right)\right\}.$$

If $s > 0$, then $D_0 < 0$, so $K > 0$. But another integration by parts, with the boundary terms canceling, gives

$$-\frac{K}{t} = \sum_j \int_{V_j} x\varphi(x)\Phi(q_j(x))\,dx > 0.$$

The strict sign follows by pairing x with $-x$: the old squared distance to the nearest existing center is even, while the conditional capture probability at $x > 0$ is strictly larger than at $-x$ when $s > 0$. This contradiction, and reflection for $s < 0$, show that every interior stationary point lies on $s = 0$.

We next bound that perpendicular profile. Put $p_0 = 2\Phi(a/2) - 1$. Splitting the central and outer intervals and using $\varphi(a/2+u) \leqslant \varphi(a/2)e^{-au/2}$ gives

$$\begin{aligned} F(0,t) &= p_0\Phi(-t/2) + 2\int_0^\infty \varphi(a/2+u)\Phi(-t/2 - au/t)\,du \\ &\leqslant p_0\Phi(-t/2) + \frac{4\varphi(a/2)}{a}\{\Phi(-t/2) - e^{3t^2/8}\Phi(-t)\}. \end{aligned}$$

The second line follows by integration by parts; its braces are nonnegative by the preceding integral representation. The bounds $25/8 < \pi < 22/7$ and $e^{-v} \leqslant 1 - v + v^2/2$ give

$$p_0 \leqslant \frac{1}{\pi}\left(1 - \frac{a^2}{24} + \frac{a^4}{640}\right) < \frac{8}{25}\left(1 - \frac{7}{264} + \frac{2}{3125}\right) < \frac{5}{16}, \qquad \frac{4\varphi(a/2)}{a} = 2e^{-1/(4\pi)} < \frac{37}{20}.$$

Consequently $F(0,t) \leqslant B(t)$, where $B(t) = (5/16 + 37/20)\Phi(-t/2) - (37/20)e^{3t^2/8}\Phi(-t)$. We verify $B(t) < 1/4$ for $0 \leqslant t \leqslant 4/5$, which includes $0 \leqslant t \leqslant a$.

The elementary estimates $157/50 < \pi < 22/7$ imply $997/2500 < \varphi(0) < 3991/10000$. Alternating bounds for the exponential therefore give

$$\Phi(-t/2) \leqslant U(t) := \frac{1}{2} - \frac{997}{2500}\left(\frac{t}{2} - \frac{t^3}{48}\right),$$
$$\Phi(-t) \geqslant L(t) := \frac{1}{2} - \frac{3991}{10000}\left(t - \frac{t^3}{6} + \frac{t^5}{40}\right) > 0.$$

Let $E(t) = \sum_{j=0}^{4}(3t^2/8)^j/j!$. Then $1/4 - B(t) \geqslant P(t) := 1/4 - (5/16 + 37/20)U(t) + (37/20)E(t)L(t)$. This is an explicit degree-thirteen polynomial. To verify positivity without numerical integration, express it in the degree-thirteen Bernstein basis on $[0, 4/5]$. If $P(t) = \sum_{j=0}^{13} p_j t^j$, its coefficient of $\binom{13}{i}(5t/4)^i(1-5t/4)^{13-i}$ is $\sum_{j=0}^{i} p_j (4/5)^j \binom{i}{j}/\binom{13}{j}$. Exact expansion bounds these fourteen coefficients below by the following integers divided by 10^6:

$$(93750, 74849, 58795, 45279, 34032, 24814, 17411, 11634, 7309, 4276, 2382, 1475, 1392, 1935).$$

Every coefficient is positive, and the basis polynomials are nonnegative and sum to one. This proves $F(0,t) < 1/4$ in the required interval. The lower bound on π was proved in Lemma 17; the upper bound follows, for example, from $\int_0^1 x^4(1-x)^4/(1+x^2)\,dx = 22/7 - \pi > 0$.

We finish by checking possible boundary limits of a maximizing sequence. An axis challenger between neighboring centers captures an interval of length $a/2$, hence probability at most $a\varphi(0)/2 = 1/(2\pi) < 1/4$. On approach to the origin, the limiting capture is a halfspace through the center of the symmetric strip $|X| < a/2$, of mass $p_0/2 < 1/4$. On approach to $(a, 0)$, the limiting capture is a halfspace through $(a, 0)$ intersected with $X > a/2$. Outward directions are bounded by Lemma 19. For inward directions, put $w(v) = \varphi(a+v) - \mathbf{1}_{v<a/2}\varphi(a-v)$. This function is negative before $a/2$ and positive after it, and

$$\int_0^\infty w(v)\,dv = 1 - 2\Phi(a) + \Phi(a/2) \geqslant \frac{1}{2} - \frac{3}{2\pi} > 0.$$

The function $g_h(v) = \Phi(hv) - 1/2$ increases. Comparing it with its value at $a/2$ on the two sign intervals gives $\int_0^\infty g_h(v)w(v)\,dv \geqslant 0$. An inward halfspace consequently has mass at most $(1 - \Phi(a/2))/2 < 1/4$. Axial and transverse directions follow by limits; the other incumbent is symmetric. These limiting descriptions follow by dominated convergence off the finitely many boundary lines. Capture tends to zero at infinity. A supremum exceeding $1/4$ would therefore be attained with $t > 0, r < a$, where the stationary calculation and the perpendicular bound give a contradiction. $\square$

Proof of Theorem 8. A Gaussian of rank zero is a point mass and needs one center. Otherwise translate its mean to zero and scale its largest covariance eigenvalue to one. Choose a corresponding eigenvector as the first coordinate. Then X_1 is standard normal, independent of the perpendicular coordinates, whose covariance matrix $\Sigma_\perp$ has eigenvalues at most one. Use centers $\{-ae_1, 0, ae_1\}$. For a challenger (s, b), put $t^2 = \|b\|_2^2$ and $t'^2 = b^\top \Sigma_\perp b \leqslant t^2$. Its capture event can be written

$$2sX_1 + 2t'Z - s^2 - t^2 > \max_{c\in\{-a,0,a\}}(2cX_1 - c^2),$$

where Z is an independent standard normal; when $t' = 0$, its contribution is zero. Replacing t^2 by t'^2 enlarges the event and gives exactly the bivariate comparison in Lemma 20. These three centers therefore work for $k = 3, 4$, including singular covariance. For $k = 1, 2$, the mean alone suffices: every nonempty deviation is a noncentral Gaussian halfspace of mass strictly below $1/2$. $\square$

F One Center

For one center, existence of an exact core is immediate under our quota convention: no group has mass greater than one, and $\delta(\mu, 1) \leqslant 1$. Euclidean geometry nevertheless gives a useful stronger bound. A point can be chosen so that every closed halfspace containing it has substantial population mass. An alternative center then leaves that much mass behind.

Definition 7 (Tukey depth)**.** The Tukey depth of a point $c \in \mathbb{R}^D$ under a probability measure μ is the infimum of $\mu(H)$ over all closed halfspaces H containing c.

The centerpoint theorem guarantees a point of depth at least $1/(D+1)$ for every Borel probability measure on $\mathbb{R}^D$ (Rado, 1946). This is a distributional version of a median: in one dimension, neither closed halfline containing a median has mass below one half.

Lemma 21. *For Euclidean distance, every Borel probability measure on $\mathbb{R}^D$ has a center c such that $\mu(D(\{c\}, y)) \leqslant D/(D+1)$ for every y. In particular, $\delta(\mu, 1) \leqslant D/(D+1)$.*

Proof. Choose a point c of depth at least $1/(D+1)$. For $y \neq c$, the points closer to y than to c form an open halfspace whose complementary closed halfspace contains c. The complement has mass at least $1/(D+1)$. For $y = c$, the deviation region is empty. □

G A Two-Way Connection with Weak ε-Nets

The exact core asks a set of centers to meet every sufficiently large group that could share a better alternative. In Euclidean space, those groups arise from convex regions. This connects proportional fairness to a geometric question: how many points are needed to intersect every convex region of a prescribed mass? The connection gives another way to obtain upper bounds and also describes a geometric consequence of belonging to the core.

Definition 8 (Weak ε-net)**.** For $\varepsilon \in (0, 1)$ and a Borel probability measure μ on $\mathbb{R}^D$, a finite set $N \subseteq \mathbb{R}^D$ is a weak ε-net if it intersects every Borel convex set K with $\mu(K) \geqslant \varepsilon$.

The word "weak" means that the selected points need not belong to the population's support. This is the appropriate version when centers are unrestricted. We first explain why finite results about these nets also apply to arbitrary probability measures. A direct appeal to uniform sampling over *all* convex sets would be invalid; that class does not have bounded VC dimension in dimensions at least two.

Lemma 22. *Suppose that every finite point set in $\mathbb{R}^D$ has a weak t-net of size at most $N_D(t)$. Then every Borel probability measure on $\mathbb{R}^D$ has a weak ε-net of size at most $N_D(\varepsilon/2)$.*

Proof. Put $t = \varepsilon/2$ and $N = N_D(t)$. Consider the family of compact convex sets K with $\mu(K) > t$. Any finite subfamily $K_1, \ldots, K_s$ admits an empirical measure assigning mass greater than t to each member, by the law of large numbers for this finite collection of indicator functions.

The empirical observations may coincide. For each positive integer ℓ, replace them by distinct points at distance at most $1/\ell$ from their original locations. Every expanded set $K_i + \overline{B}(0, 1/\ell)$ then contains more than a t fraction of these points. A finite weak t-net of size at most N hits every expanded set. Pad its centers to an N-tuple and pass to a convergent subsequence in the N-fold product of the one-point compactification $\mathbb{R}^D \cup \{\infty\}$. Each K_i is bounded and closed, so at least one finite coordinate of the limiting tuple lies in K_i. This produces a tuple hitting the entire finite subfamily.

For each compact convex K under consideration, the condition that at least one of the N coordinates belong to K is closed in this compact product space. The preceding paragraph establishes the finite intersection property for these conditions. Compactness therefore gives one tuple hitting all such K simultaneously. Discard its coordinates equal to ∞.

Finally, if a Borel convex set A has $\mu(A) \geqslant \varepsilon$, inner regularity gives a compact subset $L \subseteq A$ with $\mu(L) > t$. Its convex hull is compact, is contained in A, and has mass greater than t. The constructed tuple hits this hull and hence A. At least one finite coordinate remains, since some compact ball has mass greater than t. □

The implication for core clustering now follows directly from convexity of distance comparisons.

Theorem 9. *If a probability measure μ on $\mathbb{R}^D$ has a weak ε-net C with at most k points, then $\mu(D(C, y)) < \varepsilon$ for every $y \in \mathbb{R}^D$, and hence $\delta(\mu, k) \leqslant \varepsilon$.*

Proof. The region $D(C, y)$ is an intersection of open halfspaces and is disjoint from C. If its mass were at least ε, it would be a Borel convex set missed by the weak net, a contradiction. $\square$

The quantitative consequences are weaker than the $4.911/k$ guarantee in Theorem 2, but show what follows from the general theory of convex sets alone. For every fixed $\tau > 0$, Rubin (2018) give planar weak nets of size $O(\varepsilon^{-(3/2+\tau)})$. The corrected higher-dimensional bounds of Rubin (2023) give $O(\varepsilon^{-(a_D+\tau)})$, with $a_3 = 2.558$, $a_4 = 3.48$, and $a_D = (D + \sqrt{D^2 - 2D})/2$ for $D \geqslant 5$. Constants in this paragraph can depend on D and τ.

Corollary 1. *For every $D \geqslant 2$ and $\tau > 0$, every Borel probability measure μ on $\mathbb{R}^D$ satisfies $\delta(\mu, k) = O(k^{-1/(a_D+\tau)})$, where $a_2 = 3/2$ and the other exponents are as above.*

Proof. Apply Lemma 22 to the cited finite bounds, and choose ε to make the resulting net size at most k. Then apply Theorem 9. The finitely many smaller budgets can be absorbed in the constant using $\delta(\mu, k) \leqslant 1$. $\square$

For the converse, a core need not hit every convex set of large mass. A region may be very thin, so its points need not agree on any better center inside it. A ball well inside a missed region does create such a common deviation.

Theorem 10. *Let C be an (α, β)-core of a Borel probability measure μ on $\mathbb{R}^D$. Let $K \subsetneq \mathbb{R}^D$ be a nonempty closed set, and let $p \in K$ have $r = d(p, \mathbb{R}^D \setminus K)$. If $\mu(B(p, r/(\alpha + 1))) > \beta/k$, then $C \cap K \neq \varnothing$.*

Proof. If C missed K, every x in the displayed open ball would satisfy

$$d(x, C) \geqslant d(x, \mathbb{R}^D \setminus K) \geqslant r - \|x - p\|_2 > \alpha \|x - p\|_2.$$

Thus the entire ball would belong to $D_\alpha(C, p)$, contradicting the core bound. The strict mass inequality is required by the non-strict quota convention in Definition 1. $\square$

This converse does not require K to be convex. In the exact case, an open ball of radius half the distance from p to the complement is sufficient. The two implications therefore compare distinct geometric demands: weak nets hit all heavy convex sets, while the core necessarily hits sets whose interior contains enough mass to support one common deviation.

H Additional Related Work

Demographic and individual fairness. Chierichetti et al. (2017) require protected groups to be appropriately represented within each cluster and introduce fairlets to obtain approximation algorithms for fair k-center and k-median objectives. Such constraints use demographic labels, whereas our groups are determined by shared preferences over center locations. A closer geometric comparison is the individual fairness criterion of Jung et al. (2020): each point should have a center within a constant times the smallest radius of a ball centered at that point containing at least n/k individuals. Their factor-2 guarantee adapts to local population concentration. The approximation connections of Kellerhals & Peters (2024) relate this radius condition to proportional fairness, although it is a different exact requirement.

Beyond centroid losses. Caragiannis et al. (2024) study proportional fairness in non-centroid clustering, where a point's loss depends on the other points assigned to its cluster. They analyze both the core and its relaxation, fully justified representation. Cookson et al. (2025) introduce semi-centroid clustering, combining distance to a chosen center with loss from the other cluster members, and obtain polynomial-time constant-factor core approximations even when the two losses use different metrics. These models extend the preferences and deviations under consideration. Here, nearest-center distance alone determines each point's preferences, which makes a deviation describable by a geometric region and permits our distributional analysis.

Geometric hitting sets. A weak ε-net intersects every convex region containing at least an ε fraction of a population. Since the points preferring one Euclidean center to all chosen centers form a convex region, such a net bounds every deviating group's mass. Section G gives this implication and a converse for regions containing a sufficiently massive ball, together with the quantitative bounds of Rubin (2018, 2023). The connection explains what geometric hitting alone yields, even though the inherited $4.911/k$ guarantee is stronger asymptotically.