RANDOM SETS ARE CLOSE TO LOW-DISCREPANCY SETS

GLEB SMIRNOV AND ROMAN VERSHYNIN

Abstract. We show that a random sample from an arbitrary probability measure on $\mathbb{R}^d$ is close to a low-discrepancy point set. Namely, after moving only a small fraction of the sample points in expectation, one obtains an n-point set with star discrepancy $\operatorname{polylog}(n)/n$ with respect to the original measure.

1. Main results

A fundamental limitation of random sampling is that its typical estimation error is of order $n^{-1/2}$. We ask whether this barrier can be overcome by modifying only a small fraction of the sample, assuming the underlying distribution is known.

1.1. Empirical processes and star discrepancy.

Let $X_1, \ldots, X_n$ be iid random variables taking values in $\mathbb{R}^d$. Let μ denote their common distribution, and let μ_n^X denote the empirical measure:

$$\mu(B) = \mathbb{P}\{X_i \in B\}, \qquad \mu_n^X(B) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{X_i \in B\}},$$

for every Borel set $B \subset \mathbb{R}^d$.

The classical VC theory (see e.g. [17, Theorem 8.3.15]) implies that the empirical measure approximates the population measure uniformly over lower orthants:

$$\mathbb{E}\sup_B \left|\mu_n^X(B) - \mu(B)\right| \leqslant C\sqrt{\frac{d}{n}}. \tag{1.1}$$

Here the supremum is over all lower orthants

$$B = (-\infty, t_1] \times \cdots \times (-\infty, t_d], \qquad (t_1, \ldots, t_d) \in \mathbb{R}^d. \tag{1.2}$$

The quantity $\sup_B \left|\mu_n^X(B) - \mu(B)\right|$ has two classical interpretations, one in probability and the other in discrete geometry. For random points X_i, it is the supremum of the empirical process indexed by lower orthants; see [6, 9, 10, 12, 16]. For deterministic point sets, it is precisely the star discrepancy.

It is known that there exist n-point sets with star discrepancy as low as $\operatorname{polylog}(n)/n$; see [1, 2, 4, 7, 11]. This is substantially smaller than the $n^{-1/2}$ bound in (1.1), showing that low-discrepancy point sets are considerably more regular than random samples.

1.2. Main result.

Our main result shows that a random sample is close to a low-discrepancy point set. By modifying only a prescribed expected number of sample points, one can reduce its discrepancy to nearly the optimal order.

R.V. acknowledges support from NSF Grant DMS-2451011 and U.S. Air Force Grant FA9550-25-1-0294.

Theorem 1.1 (A random sample is close to a low-discrepancy set)**.** *Let $X_1,\dots,X_n$ be iid random variables taking values in $\mathbb{R}^d$, and let μ denote their common distribution. For every $m \in (0,n]$, there exists a randomized algorithm, possibly depending on μ, that maps points $X_1,\dots,X_n$ to points $Y_1,\dots,Y_n \in \mathbb{R}^d$ such that*
$$\mathbb{E}\,\#\{i : Y_i \neq X_i\} \leqslant m,$$
and
$$\mathbb{E}\sup_B \left|\mu_n^Y(B) - \mu(B)\right| \leqslant \frac{C^d \log^{3d/2+1}(2n)}{m}.$$
Here μ_n^Y is the empirical measure of $Y_1,\dots,Y_n$, the supremum is over all lower orthants B, and $C_d > 0$ depends only on d. The expectations are taken over both the sample and the internal randomness of the algorithm.

In particular, if we choose $m = 0.01n$, Theorem 1.1 implies that moving just 1% of the sample points can reduce the star discrepancy to $\operatorname{polylog}(n)/n$.

Although the exponent of the logarithm in Theorem 1.1 is certainly not optimal, the rate $1/m$ is optimal even in dimension $d = 1$ [13].

1.3. **Prior work.** Several partial cases of Theorem 1.1 are known. The case $m = n$ corresponds to finding any n-point set in $\mathbb{R}^d$ with small star discrepancy, and the best known constructions give $O_d(\log^{d-1/2}(2n)/n)$ [1]. In dimension $d = 1$, Theorem 1.1 was proved in [13] with the optimal bound $O(1/m)$. Moreover, still in dimension $d = 1$, it is known how to *remove* (rather than move) m points to achieve discrepancy $O(\log(2n)/m)$, which is optimal if only removals are allowed [5]. Perhaps the closest result to ours is [8], which proves a version of Theorem 1.1 in the particular case of the uniform measure on the unit cube $[0,1]^d$. They show how to remove m points to achieve star discrepancy $O(\log^{2d}(2n)/m)$. The proof in [8] relies on the orthogonality of Haar wavelets, and is therefore specific to the product structure of the measure μ. Our argument is completely different and it applies to arbitrary Borel measures.

1.4. **Sparse families.** The star discrepancy measures the uniform approximation over a specific family of sets – the lower orthants in $\mathbb{R}^d$. Let's now be more ambitious and try to consider general families of subsets.

We will now state an abstract result, from which Theorem 1.1 will be deduced in § 4 by standard tricks – VC theory, dyadic decomposition, and approximation.

A family $A_1,\dots,A_D$ of subsets of $[N] = \{1,\dots,N\}$ is called *s-sparse* if each point is covered by at most s subsets, i.e.
$$\#\{j : i \in A_j\} \leqslant s \quad \text{for each } i \in [N].$$

Theorem 1.2 (Sparse families)**.** *Fix $0 < \delta < 1$. Let $A_1,\dots,A_D$ be an s-sparse family of subsets of $[N]$. Assume that N has the form $N = n2^k$ for some positive integers n and k. Let $S \subset [N]$ be a random n-element subset, chosen uniformly without replacement. Then there is a randomized algorithm that produces an n-element subset $Q \subset [N]$ such that*
$$\mathbb{E}|Q \cap S| \geqslant n(1-\delta) - 2\sqrt{n}$$
and
$$\left|\frac{|A_j \cap Q|}{n} - \frac{|A_j|}{N}\right| \leqslant C\frac{\sqrt{s}}{n}\left(\frac{k}{\delta} + \log(2n)\right) \qquad \text{for each } j.$$
Here $C > 0$ is an absolute constant. The expectation is taken over both the sample S and the internal randomness of the algorithm.

1.5. **A word on the proof.** In § 2 and § 3, we deduce Theorem 1.2 by combining our earlier work on two-color discrepancy [14,15] with some randomization of Beck's transference principle [1,3]. The randomized transference principle is developed in the proof of Proposition 2.3; we believe this simple argument can be useful in many other applications.

Then, in § 4, we deduce Theorem 1.1 by approximation based on VC theory.

2. Iterated thinning

The proof of Theorem 1.2 rests on an iterated thinning procedure. We first record the one-step input.

Proposition 2.1 (Partial coloring [14])**.** *Let $v_1, \ldots, v_m \in \mathbb{R}^D$ be any vectors, and let $\varepsilon_1, \ldots, \varepsilon_m$ be independent fair signs. For every $T > 0$ there is a randomized algorithm that discards some of the terms $\varepsilon_i v_i$ so that*

$$\max_{r \leqslant m} \Big\| \sum_{i \leqslant r \text{ accepted}} \varepsilon_i v_i \Big\|_\infty \leqslant T, \tag{2.1}$$

and

$$\mathbb{P}\{v_i \text{ is discarded}\} \leqslant \frac{\pi}{T} \|v_i\|_2 \qquad \text{for each } i. \tag{2.2}$$

Iterating gives a full-coloring form:

Lemma 2.2 (Full coloring)**.** *Let $v_1, \ldots, v_m \in \mathbb{R}^D$ be any vectors satisfying $\|v_i\|_2 \leqslant 1$, and let $\varepsilon_1, \ldots, \varepsilon_m$ be independent fair signs. For every $T > 0$ there is a randomized algorithm producing signs $\varepsilon'_1, \ldots, \varepsilon'_m$ such that*

$$\Big\| \sum_{i=1}^m \varepsilon'_i v_i \Big\|_\infty \leqslant T + C \log(2m), \tag{2.3}$$

and

$$\mathbb{P}\{\varepsilon'_i \neq \varepsilon_i\} \leqslant \frac{\pi}{T} \qquad \text{for each } i. \tag{2.4}$$

Proof. Apply Proposition 2.1. Keep the original signs on the accepted vectors. Color the discarded vectors by iterating the same procedure with threshold 2π. At each iteration the expected number of discarded vectors is at most half the number entering that iteration, so there is a deterministic choice of the auxiliary signs for which at most half survive. Summing the contributions over the $O(\log(2m))$ iterations gives $C \log(2m)$. Together with (2.1), this proves (2.3). A sign can change only if the corresponding vector was discarded in the first application, and (2.4) follows from (2.2). □

We will now convert coloring to sampling; our argument can be seen as a randomized version of Beck's transference principle [1,3].

Proposition 2.3 (Iterated thinning)**.** *Fix $0 < \delta < 1$. Let $N = n2^k$ and let $v_1, \ldots, v_N \in \mathbb{R}^D$ satisfy $\|v_i\|_2 \leqslant 1$. Let $S \subset [N]$ be obtained by selecting every index independently with probability 2^{-k}. Then there is a randomized algorithm producing a subset $Q \subset [N]$ such that*

$$\mathbb{E}|Q \cap S| \geqslant n(1-\delta) \tag{2.5}$$

and

$$\Big\| \frac{1}{n} \sum_{i \in Q} v_i - \frac{1}{N} \sum_{i=1}^N v_i \Big\|_\infty \leqslant \frac{C}{n} \left(\frac{k}{\delta} + \log(2n) \right). \tag{2.6}$$

Proof. For each vector v_i, choose its independent "fate"
$$\big(\varepsilon_i(1),\dots,\varepsilon_i(k)\big)\in\{\pm1\}^k$$
uniformly at random, and realize
$$S=\big\{i:\varepsilon_i(1)=\dots=\varepsilon_i(k)=+1\big\}.$$
Start with $I(0):=[N]$. In round $r\in[k]$, apply Lemma 2.2 to the vectors $(v_i)_{i\in I(r-1)}$ with input signs $\varepsilon_i(r)$ and threshold
$$T:=\frac{2\pi k}{\delta}.$$
Let $I(r)$ consist of the indices from $I(r-1)$ that receive output sign $\varepsilon_i'(r)=+1$. Then, for each round r,
$$\sum_{i\in I(r-1)}v_i-2\sum_{i\in I(r)}v_i=-\sum_{i\in I(r-1)}\varepsilon_i'(r)v_i.$$
Hence:
$$\Bigg\|\sum_{i\in I(r-1)}v_i-2\sum_{i\in I(r)}v_i\Bigg\|_\infty=\Bigg\|\sum_{i\in I(r-1)}\varepsilon_i'(r)v_i\Bigg\|_\infty\leqslant T+C\log(2N). \tag{2.7}$$
Telescoping gives
$$\sum_{r=1}^k2^{r-1}\left(\sum_{i\in I(r-1)}v_i-2\sum_{i\in I(r)}v_i\right)=\sum_{i=1}^Nv_i-2^k\sum_{i\in I(k)}v_i,$$
so we obtain from (2.7):
$$\Bigg\|\sum_{i=1}^Nv_i-2^k\sum_{i\in I(k)}v_i\Bigg\|_\infty\leqslant\sum_{r=1}^k2^{r-1}\big(T+C\log(2N)\big)\leqslant2^k\big(T+C\log(2N)\big).$$
Put
$$Q:=I(k).$$
Then, dividing by N, we conclude:
$$\Bigg\|\frac1N\sum_{i=1}^Nv_i-\frac1n\sum_{i\in Q}v_i\Bigg\|_\infty\leqslant\frac1n\big(T+C\log(2N)\big),$$
and (2.6) follows after adjusting the constant C.

It remains to prove (2.5). By construction,
$$Q\cap S=\big\{i:\varepsilon_i'(r)=\varepsilon_i(r)=1\text{ for all }r\in[k]\big\}.$$

To bound the expected cardinality of $Q\cap S$, let us address each round r separately. Condition on the outcomes of all past rounds (i.e. any configuration of the input and output signs in the rounds $1,\dots,r-1$ that leave $i\in I(r-1)$). Then, using Lemma 2.2 and the fairness of the sign $\varepsilon_i(r)$, we get
$$\mathbb{P}\big\{\varepsilon_i'(r)\neq\varepsilon_i(r)\,\big|\,\text{past rounds and }\varepsilon_i(r)=+1\big\}\leqslant2\,\mathbb{P}\{\varepsilon_i'(r)\neq\varepsilon_i(r)\,|\,\text{past rounds}\}\leqslant\frac{2\pi}{T}=\frac{\delta}{k}.$$
Equivalently,
$$\mathbb{P}\big\{\varepsilon_i'(r)=\varepsilon_i(r)\,\big|\,\text{past rounds and }\varepsilon_i(r)=+1\big\}\geqslant1-\frac{\delta}{k}.$$

From this, using the fairness of the sign again, we obtain

$$\mathbb{P}\{\varepsilon_i'(r) = \varepsilon_i(r) = 1 \,\big|\, \text{past rounds}\} \geqslant \left(1 - \frac{\delta}{k}\right) \cdot \frac{1}{2}.$$

Hence, telescoping the conditional probability we conclude that

$$\mathbb{P}\{\varepsilon_i'(r) = \varepsilon_i(r) = 1 \text{ for all } r \in [k]\} \geqslant \left(1 - \frac{\delta}{k}\right)^k \cdot \frac{1}{2^k} \geqslant 2^{-k}(1-\delta).$$

Now sum over $i \in [N]$ to get

$$\mathbb{E}|Q \cap S| \geqslant N2^{-k}(1-\delta) = n(1-\delta).$$

□

3. Sparse families

We now pass from Proposition 2.3 to Theorem 1.2. The proof has two parts: first we prove the result in the Bernoulli model, then we use coupling to pass to sampling without replacement.

Lemma 3.1 (Sparse families: Bernoulli selection)**.** *Fix $0 < \delta < 1$. Let $A_1, \dots, A_D$ be an s-sparse family of subsets of $[N]$. Assume that N has the form $N = n2^k$ for some positive integers n and k. Let $S \subset [N]$ be a random subset obtained by selecting each point independently with probability 2^{-k}. Then there is a randomized algorithm that produces an n-element subset $Q \subset [N]$ such that*

$$\mathbb{E}|Q \cap S| \geqslant n(1-\delta) - \sqrt{n} \tag{3.1}$$

and

$$\left|\frac{|A_j \cap Q|}{n} - \frac{|A_j|}{N}\right| \leqslant C\frac{\sqrt{s}}{n}\left(\frac{k}{\delta} + \log(2n)\right) \qquad \text{for each } j. \tag{3.2}$$

Proof. Adjoin $A_0 = [N]$ to the family and associate to each $i \in [N]$ the incidence vector

$$v_i = \left(\mathbf{1}_{\{i \in A_0\}}, \dots, \mathbf{1}_{\{i \in A_D\}}\right) \in \mathbb{R}^{D+1}.$$

Because the original family is s-sparse,

$$\|v_i\|_2 \leqslant \sqrt{s+1} \leqslant \sqrt{2s}.$$

Applying Proposition 2.3 for the scaled vectors $v_i/\sqrt{2s}$, we obtain a subset Q_0 satisfying

$$\left|\frac{|A_j \cap Q_0|}{n} - \frac{|A_j|}{N}\right| \leqslant C\frac{\sqrt{s}}{n}\left(\frac{k}{\delta} + \log(2n)\right) \quad \text{for each } j = 0, \dots, D. \tag{3.3}$$

For $j = 0$, this implies

$$\big||Q_0| - n\big| \leqslant C\sqrt{s}\left(\frac{k}{\delta} + \log(2n)\right). \tag{3.4}$$

Moreover, (2.5) yields

$$\mathbb{E}\,|Q_0 \cap S| \geqslant n(1-\delta). \tag{3.5}$$

Modify Q_0 to have cardinality exactly n: add arbitrary points when $|Q_0| < n$; when $|Q_0| > n$, remove points outside S first and then, only if necessary, remove points of S. Call the resulting set Q. By (3.4), this adjustment changes every normalized count by at most the right-hand side of (3.3), so (3.2) follows after changing C.

The prescribed removal rule gives

$$|Q \cap S| \geqslant |Q_0 \cap S| - (|S| - n)_+.$$

Since $|S| \sim \mathrm{Bin}(N, 2^{-k})$ has mean $N2^{-k} = n$ and variance at most n,

$$\mathbb{E}(|S|-n)_+ \leqslant \mathbb{E}\big||S|-n\big| \leqslant \sqrt{n}.$$

Combining this with (3.5) proves (3.1). □

Lemma 3.2 (Coupling Bernoulli and fixed-size sampling)**.** *Let S be a uniformly distributed n-element subset of $[N]$, and let $\xi \sim \mathrm{Bin}(N, 2^{-k})$ be independent of S. Construct S' by adding $\xi - n$ uniformly chosen points of $[N] \setminus S$ when $\xi > n$, and by removing $n - \xi$ uniformly chosen points of S when $\xi < n$. Then S' is a Bernoulli subset with inclusion probability 2^{-k}, and*

$$\mathbb{E}\,|S \triangle S'| = \mathbb{E}\,|\xi - n| \leqslant \sqrt{n}. \tag{3.6}$$

Proof. Conditional on $|S'| = r$, the construction is invariant under every permutation of $[N]$, hence S' is uniform among the r-element subsets. Since $|S'| = \xi \sim \mathrm{Bin}(N, 2^{-k})$, it follows that S' has the Bernoulli law, as claimed. The construction changes exactly $|\xi - n|$ points, and Cauchy–Schwarz gives

$$\mathbb{E}\,|\xi - n| \leqslant \sqrt{\mathrm{Var}(\xi)} \leqslant \sqrt{n}.$$ □

Proof of Theorem 1.2. Couple the given uniform n-element set S with a Bernoulli set S' as in Lemma 3.2. Apply Lemma 3.1 to S' and let Q be the resulting n-element set. The discrepancy estimate is exactly (3.2). Furthermore,

$$|Q \cap S| \geqslant |Q \cap S'| - |S \triangle S'|.$$

Taking expectations and using (3.1) and (3.6), we obtain

$$\mathbb{E}\,|Q \cap S| \geqslant n(1-\delta) - 2\sqrt{n}.$$

This proves the theorem. □

4. Proof of Theorem 1.1

We will deduce the main theorem from Theorem 1.2.

Let's start with a few reductions as in [15]. First, we can assume that the cumulative distribution functions of all marginals of μ are strictly increasing. This follows by mixing the original distribution with a Gaussian: sample from the original distribution with probability p and from the standard Gaussian distribution with probability $1-p$; then take $p \to 1$.

Moreover, it is enough to treat measures on $[0,1]^d$. To see this, consider a fixed increasing homeomorphism from $\mathbb{R}$ onto $(0,1)$, applied coordinatewise.

Finally, the following observation will allow us to focus on measures with uniform marginals:

Lemma 4.1 (Uniformization)**.** *Let X be a random vector with distribution μ on $[0,1]^d$. Let F_k denote the cumulative distribution function of the kth marginal of μ. Assume all F_k are strictly increasing on $[0,1]$. Then there exists a random vector $\widehat{X}$ whose marginals are uniform on $[0,1]$ and such that, simultaneously for every anchored box*

$$B = [0,t_1] \times \cdots \times [0,t_d], \qquad (t_1, \ldots, t_d) \in \mathbb{R}^d, \tag{4.1}$$

we have

$$X \in B \Longleftrightarrow \widehat{X} \in \widehat{B},$$

where $\widehat{B} = [0, F_1(t_1)] \times \cdots \times [0, F_d(t_d)]$.

Proof. The argument is the same as in [15, Section 3], and we include it here for completeness. Let $U_1, \dots, U_d$ be independent Unif$[0,1]$ random variables, independent of X. Define the (randomized) integral transform

$$\widehat{X}(k) = U_k \cdot F_k(X(k)) + (1-U_k) \cdot F_k(X(k)^-), \qquad k = 1, \dots, d, \tag{4.2}$$

where $F_k(a^-) = \lim_{x \to a^-} F_k(x)$. Then $\widehat{X}(k)$ is uniform on $[0,1]$ for each coordinate k.

If $X \in B$ then $\widehat{X} \in \widehat{B}$, since $\widehat{X}(k) \leqslant F_k(X(k))$ by construction. If $X \notin B$, then $X(k) > t_k$ for some k. Hence, there is y with $t_k < y < X(k)$. Then

$$F_k(t_k) < F_k(y) \leqslant F_k(X(k)^-) \leqslant \widehat{X}(k),$$

where the first two bounds follow by strict monotonicity of F_k, and the last one follows from (4.2) if one replaces $F_k(X(k))$ with the smaller value $F_k(X(k)^-)$. Thus, $\widehat{X} \notin \widehat{B}$. □

Now we will prove a preliminary version of Theorem 1.1, where the ground measure μ is replaced with a very large iid sample from μ.

Proposition 4.2 (Sampling from a large sample)**.** *Fix $0 < \delta < 1$. Assume that N has the form $N = n2^k$ for some positive integers n and k. Let $Z_1, \dots, Z_N$ be iid random variables taking values in $[0,1]^d$. Let $S \subset [N]$ be a random n-element subset, chosen uniformly without replacement, independently of the sample Z_i. Then there is a randomized algorithm that produces an n-element subset $Q \subset [N]$ such that*

$$\mathbb{E}|Q \cap S| \geqslant n(1-\delta) - 2\sqrt{n}, \tag{4.3}$$

and

$$\mathbb{E} \sup_B \left| \frac{1}{n} \sum_{i \in Q} \mathbf{1}_{\{Z_i \in B\}} - \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\{Z_i \in B\}} \right| \leqslant \frac{C^d \log^{3d/2}(2n)}{n} \left(\frac{k}{\delta} + \log(2n) \right), \tag{4.4}$$

where the supremum is over anchored boxes (4.1).

Proof. Let μ denote the law of Z_i. Applying the uniformization (4.2) independently to the sample points Z_i, we see that it is enough to consider the case in which every marginal of μ is uniform.

Step 1. Dyadic boxes. Fix the resolution parameter

$$L = \lceil \log_2(2n) \rceil. \tag{4.5}$$

A dyadic interval is an interval in [0,1] of the form

$$\left[j2^{-\ell}, (j+1)2^{-\ell} \right] \quad \text{for some } \ell \in \{0, 1, \dots, L-1\},\ j \in \{0, 1, \dots, 2^\ell - 1\}.$$

A dyadic box is a product of d dyadic intervals.

Almost surely no sample point Z_i lies on the boundary of any dyadic box. Each point belongs to at most L^d dyadic boxes, so the traces

$$A_B := \{ i \in [N] : \ Z_i \in B \}, \qquad B \text{ is a dyadic box},$$

form an L^d-sparse family on $[N]$.

Apply Theorem 1.2 to this family and to S. We obtain Q satisfying (4.3) and the discrepancy bound

$$\operatorname{disc}(B) \leqslant C \frac{L^{d/2}}{n} \left(\frac{k}{\delta} + \log(2n) \right) \quad \text{for any dyadic box } B, \tag{4.6}$$

where

$$\operatorname{disc}(B) := \left| \frac{1}{n} \sum_{i \in Q} \mathbf{1}_{\{Z_i \in B\}} - \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\{Z_i \in B\}} \right|. \tag{4.7}$$

Step 2. Lattice boxes. A lattice interval is an interval of the type

$$\left[0, j2^{-L+1}\right] \quad \text{for some } j \in \{0, 1, \ldots, 2^{L-1}\}.$$

A lattice box is the product of d lattice intervals.

Any lattice interval can be partitioned into at most L dyadic intervals. Thus, any lattice box can be partitioned into at most L^d dyadic boxes. Therefore, summing L^d bounds (4.6) by triangle inequality, we conclude

$$\operatorname{disc}(B) \leqslant C \frac{L^{3d/2}}{n} \left(\frac{k}{\delta} + \log(2n) \right) \quad \text{for any lattice box } B. \tag{4.8}$$

Step 3. Slices. We are about to approximate a general box by a dyadic box, with the difference between the two covered by a union of d thin slices. To control the approximation error, we will need to bound the number of points Z_i that can fall into a thin slice. Let's do that first.

A slice is the product of one interval of the type

$$\left[j2^{-L+1}, (j+1)2^{-L+1}\right] \quad \text{for some } j \in \{0, 1, \ldots, 2^{L-1} - 1\}, \tag{4.9}$$

and $d-1$ intervals $[0,1]$; the interval can appear as any of the d factors.

Since the sample points Z_i have uniform marginals on $[0,1]$, the number of points falling into any slice S,

$$\sum_{i=1}^{N} \mathbf{1}_{\{Z_i \in S\}},$$

has binomial distribution $\operatorname{Bin}(N, 2^{-L+1})$. Using the standard exponential moment method one can check that the expected maximum of m random variables with distribution $\operatorname{Bin}(n,p)$ is bounded by $C(np + \log m)$; see e.g. [17, Second proof of Proposition 2.7.6]. In our case, there are a total of $d2^{L-1}$ slices, so

$$\mathbb{E} \max_{S:\text{ slice}} \sum_{i=1}^{N} \mathbf{1}_{\{Z_i \in S\}} \leqslant C\Big(N2^{-L} + \log(d2^L)\Big). \tag{4.10}$$

Step 4. Approximation. Any interval $B = [0,t] \subset [0,1]$ can be approximated from below by some lattice interval $B' \subset B$ by rounding t down to the nearest multiple of 2^{-L+1}; this way $B \setminus B'$ lies in some dyadic interval of the type (4.9). Therefore, a general box

$$B = [0, t_1] \times \cdots \times [0, t_d]$$

can be approximated from below by a lattice box $B' \subset B$ so that $B \setminus B'$ lies in a union of d slices.

By definition of discrepancy (4.7) and triangle inequality, we have

$$\operatorname{disc}(B) \leqslant \operatorname{disc}(B') + \operatorname{disc}(B \setminus B'), \tag{4.11}$$

and

$$\begin{aligned}
\operatorname{disc}(B\setminus B') &\leqslant \frac{1}{n}\sum_{i\in Q}\mathbf{1}_{\{Z_i\in B\setminus B'\}}+\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}_{\{Z_i\in B\setminus B'\}}\\
&\leqslant d\max_{S:\text{ slice}}\left(\frac{1}{n}\sum_{i\in Q}\mathbf{1}_{\{Z_i\in S\}}+\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}_{\{Z_i\in S\}}\right)\quad(\text{since } B\setminus B' \text{ is covered by } d \text{ slices})\\
&\leqslant d\max_{S:\text{ slice}}\left(\frac{2}{N}\sum_{i=1}^{N}\mathbf{1}_{\{Z_i\in S\}}+\operatorname{disc}(S)\right)\quad(\text{by definition of discrepancy}).
\end{aligned}$$

Putting this into (4.11) and taking maximum over all anchored boxes, and then taking expectation, we obtain:

$$\mathbb{E}\max_{B:\text{ box}}\operatorname{disc}(B)\leqslant \underbrace{\mathbb{E}\max_{B':\text{ lattice box}}\operatorname{disc}(B')}_{E_1}+\underbrace{\frac{2d}{N}\mathbb{E}\max_{S:\text{ slice}}\sum_{i=1}^{N}\mathbf{1}_{\{Z_i\in S\}}}_{E_2}+\underbrace{d\mathbb{E}\max_{S:\text{ slice}}\operatorname{disc}(S)}_{E_3}.$$

We can use (4.8) to bound E_1, (4.10) to bound E_2, and, since a slice is a dyadic box, (4.6) can be used to bound E_3.

Recalling the choice of the resolution parameter $L=\lceil\log_2(2n)\rceil$ we made in (4.5) and taking into account that $n>1$ (otherwise the conclusion is trivial), one can check that the bound for E_1 dominates both bounds for E_2 and E_3 (up to a factor C^d), and it gives the right hand side of (4.4). □

Proof of Theorem 1.1. As we mentioned in the beginning of this section, we can assume that the measure μ is supported on $[0,1]^d$. Set

$$\delta := \frac{m}{3n}.$$

If $\delta\leqslant n^{-1/2}$, leave the sample unchanged. Then (1.1) with $N=n$ gives

$$\mathbb{E}\sup_B\left|\mu_n^X(B)-\mu(B)\right|\leqslant C\sqrt{\frac{d}{n}}\leqslant\frac{C\sqrt{d}}{\delta n}\leqslant\frac{3C\sqrt{d}}{m}.$$

So, in this case we obtain a stronger conclusion than claimed in Theorem 1.1.

Assume now that $\delta>n^{-1/2}$. Let

$$k:=\lceil\log_2(2n)\rceil,\qquad N:=n2^k. \tag{4.12}$$

Let $Z_1,\dots,Z_N$ be iid random variables with law μ. Realize $X_1,\dots,X_n$ as a random subsample of Z_i, i.e. set

$$\{X_1,\dots,X_n\}:=\{Z_i:\ i\in S\}, \tag{4.13}$$

where $S\subset[N]$ is a uniformly random n-element subset independent of the full sample.

Apply Proposition 4.2 and obtain an n-element set Q. Set[1]

$$\{Y_1,\dots,Y_n\}:=\{Z_i:\ i\in Q\}. \tag{4.14}$$

[1]Formally, we fix the canonical order of the X-sample in (4.13) and choose the order in the Y-sample (4.14) to ensure that the overlap between S and Q translates to overlap between the samples. For example, if $S=\{5,7,8\}$ and $Q=\{7,8,9\}$, then we set $(X_1,X_2,X_3)=(Z_5,Z_7,Z_8)$ and $(Y_1,Y_2,Y_3)=(Z_9,Z_7,Z_8)$.

By (4.3) and $\delta > n^{-1/2}$,

$$\mathbb{E}\,\#\{i : Y_i \neq X_i\} = \mathbb{E}|S \setminus Q| = n - \mathbb{E}|S \cap Q| \leqslant n\delta + 2\sqrt{n} \leqslant 3n\delta = m.$$

Finally, by the triangle inequality, Proposition 4.2, and (1.1) for the full sample,

$$\begin{aligned}\mathbb{E}\sup_B \left|\mu_n^Y(B) - \mu(B)\right| &\leqslant \mathbb{E}\sup_B \left|\mu_n^Y(B) - \mu_N^Z(B)\right| + \mathbb{E}\sup_B \left|\mu_N^Z(B) - \mu(B)\right| \\ &\leqslant \frac{C^d \log^{3d/2}(2n)}{n}\left(\frac{k}{\delta} + \log(2n)\right) + \sqrt{\frac{d}{N}}.\end{aligned}$$

Due to the choice of k and N in (4.12), we have $N \geqslant n^2$, and thus we conclude that

$$\mathbb{E}\sup_B \left|\mu_n^Y(B) - \mu(B)\right| \leqslant \frac{C_1^d \log^{3d/2+1}(2n)}{m}$$

by adjusting an absolute constant C_1. □

References

[1] C. Aistleitner, D. Bilyk, and A. Nikolov. Tusnády's problem, the transference principle, and non-uniform QMC-sampling. In *Monte Carlo and Quasi-Monte Carlo Methods 2016*, pages 169–180.

[2] C. Aistleitner and J. Dick. Low-discrepancy point sets for non-uniform measures. *Acta Arith.*, 163(4):345–369, 2014.

[3] J. Beck. Balanced two-colorings of finite sets in the square I. *Combinatorica*, 1(4):327–335, 1981.

[4] J. Beck and W. Chen. *Irregularities of Distribution*. Cambridge University Press, 1987.

[5] D. Bilyk and S. Steinerberger. Regularizing random points by deleting a few. *Journal of Geometric Analysis*, 35(11), 2025.

[6] J Blum, J. Kiefer, and M. Rosenblatt. Distribution free tests of independence based on the sample distribution function. *Ann. Math. Stat.*, 32(2):485–498, 1961.

[7] B. Chazelle. *The Discrepancy Method*. Cambridge University Press, 2000.

[8] R. Dwivedi, O. Feldheim, O. Gurel-Gurevich, and A. Ramdas. The power of online thinning in reducing discrepancy. *Prob. Th. Relat. Fields*, 174(1–2):103–131, 2018.

[9] J. Kiefer and J. Wolfowitz. On the deviations of the empiric distribution function of vector chance variables. *Trans. Amer. Math. Soc.*, 87(1):173–186, 1958.

[10] P. Massart. Strong approximation for multivariate empirical and related processes, via KMT constructions. *Ann. Proba.*, 17(1):266–291, 1989.

[11] J. Matoušek. *Geometric Discrepancy*. Springer-Verlag, 1999.

[12] P. Revesz. On strong approximation of the multidimensional empirical process. *Ann. Proba.*, 4(5), 1976.

[13] G. Smirnov and R. Vershynin. Improving discrepancy by moving a few points. *Proc. Amer. Math. Soc.*, 154(7):2765–2774, 2025.

[14] G. Smirnov and R. Vershynin. Discrepancy and Fisher information. 2026.

[15] G. Smirnov and R. Vershynin. Thinning to improve two-sample discrepancy. *Rand. Struct. Alg.*, 68(2), 2026.

[16] W. Stute. The oscillation behavior of empirical processes: the multivariate case. *Ann. Proba.*, 12(2), 1984.

[17] R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2 edition, 2026.

Mathematical Sciences Institute, Australian National University, Canberra, Australia
Email address: gleb.smirnov@anu.edu.au

Department of Mathematics, University of California, Irvine, US
Email address: rvershyn@uci.edu