



**HAL**  
open science

# A Sharp Discrepancy Bound for Jittered Sampling

Benjamin Doerr

► **To cite this version:**

| Benjamin Doerr. A Sharp Discrepancy Bound for Jittered Sampling. 2021. hal-03203695v2

**HAL Id: hal-03203695**

**<https://hal.science/hal-03203695v2>**

Preprint submitted on 12 Nov 2021 (v2), last revised 30 Dec 2021 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Sharp Discrepancy Bound for Jittered Sampling

Benjamin Doerr  
 Laboratoire d'Informatique (LIX)  
 CNRS  
 École Polytechnique  
 Institut Polytechnique de Paris  
 Palaiseau  
 France

November 11, 2021

## Abstract

For  $m, d \in \mathbb{N}$ , a jittered (or stratified) sampling point set  $P$  having  $N = m^d$  points in  $[0, 1)^d$  is constructed by partitioning the unit cube  $[0, 1)^d$  into  $m^d$  axis-aligned cubes of equal size and then placing one point independently and uniformly at random in each cube. We show that there are constants  $c > 0$  and  $C$  such that for all  $d$  and all  $m \geq d$  the expected non-normalized star discrepancy of a jittered sampling point set satisfies

$$c dm^{\frac{d-1}{2}} \sqrt{1 + \log\left(\frac{m}{d}\right)} \leq \mathbb{E}D^*(P) \leq C dm^{\frac{d-1}{2}} \sqrt{1 + \log\left(\frac{m}{d}\right)}.$$

This discrepancy is thus smaller by a factor of  $\Theta\left(\left(\frac{1+\log(m/d)}{m/d}\right)^{1/2}\right)$  than the one of a uniformly distributed random point set (Monte Carlo point set) of cardinality  $m^d$ . This result improves both the upper and the lower bound for the discrepancy of jittered sampling given by Pausinger and Steinerberger (Journal of Complexity (2016)). It also removes the asymptotic requirement that  $m$  is sufficiently large compared to  $d$ .

# 1 Introduction

## 1.1 Star Discrepancy

The *star discrepancy*  $D^*(P)$  of a set  $P$  of  $N$  points in the  $d$ -dimensional unit cube  $[0, 1]^d$  is a measure for the uniformity of the distribution of these points. It is defined by

$$D^*(P) := \sup_{B \in \mathcal{B}} | |P \cap B| - N\lambda(B) |, \quad (1)$$

where  $\lambda(\cdot)$  denotes the Lebesgue measure and  $\mathcal{B}$  is the set of all axis-parallel rectangles  $[0, x) := \prod_{i=1}^d [0, x_i)$ ,  $x = (x_1, \dots, x_d) \in [0, 1]^d$ . We call  $[0, x)$  a rectangle anchored in the origin or a *box*. The star discrepancy thus is a worst-case measure for how well  $P$  satisfies the target of having the fair number  $N\lambda(B)$  of points in each box  $B$ .

Evenly distributed points sets and the star discrepancy have found applications in various areas such as machine learning [ASYM16], heuristic search [KM05, OT05, TG07], statistics [FW93], and computer graphics [Owe03]. Most prominent is its role in numerical integration, where the *Koksma-Hlawka inequality* [Kok43, Hla61] bounds the integration error in terms of the star discrepancy: For all functions  $f : [0, 1]^d \rightarrow \mathbb{R}$  having variation in the sense of Hardy and Krause bounded by 1, we have

$$\left| \int_{[0,1]^d} f(x) dx - \frac{1}{|P|} \sum_{p \in P} f(p) \right| \leq \frac{1}{|P|} D^*(P).$$

We remark that in this work we use the non-normalized version of the star discrepancy as defined above, which is more common in those areas of discrepancy theory that aim at a unified view on discrepancies and exploit connections between different discrepancy notions, e.g., between geometric discrepancies like the star discrepancy and combinatorial discrepancies like hypergraph discrepancies. See the books of Matoušek [Mat99] and Chazelle [Cha00] for an introduction to this field. In contexts closer related to numerical integration, motivated by results like the Koksma-Hlawka inequality, a normalized version of the star discrepancy is more common. The normalized star discrepancy is exactly the notion of (1) multiplied by a factor of  $\frac{1}{|P|}$ . There is little risk of confusion since the normalized version is always at most one and usually far less than one, whereas our notion is at least  $1/2$  and usually much larger than 1.

## 1.2 Estimates for the Star Discrepancy

The interest in discrepancies from various research communities has led to a huge body of research (see, e.g., [Cha00, DP10, DT97, Mat99, Nie92]), which we cannot

fully review here. The classic view on geometric discrepancies is to treat the dimension  $d$  as a constant and investigate the asymptotic behavior for growing numbers  $N$  of points. In this view, a large number of constructions of point sets has been exhibited that have a discrepancy of  $D^*(P) = O(\log(N)^{d-1})$ . It is also known that such a polylogarithmic discrepancy cannot be avoided, though the optimal exponent is not known and finding it is a famous open problem, see, e.g., [BL13]. As a side remark, we note that axis-parallel regular grids have a discrepancy of order  $\Theta(N^{(d-1)/d})$  and uniformly distributed random point sets (Monte Carlo point sets) have a discrepancy of order  $\Theta(\sqrt{N})$ , both in expectation and with high probability. Thus both are not competitive when treating  $d$  as a constant and looking for asymptotic discrepancy guarantees in terms on  $N$ .

From the viewpoint of numerical integration in high dimension, a behavior exponential in  $d$  like  $(\log N)^{d-1}$  is not very desirable, since such bounds become interesting often only for numbers  $N$  of points that are far beyond any practical meaning. For this reason, Heinrich, Novak, Wasilkowski, and Woźniakowski [HNWW01] started the quest for discrepancy bounds that both make the dependence on  $d$  fully explicit and that give reasonable discrepancy guarantees also when  $N$  is only of moderate size compared to  $d$ . Interestingly, this brought random constructions back on stage, and in fact, they are at the moment the best constructions in the regime where  $N$  is not very large (say exponential) compared to  $d$ .

In [HNWW01], Heinrich et al. prove that random point sets have an expected discrepancy of order  $\sqrt{dN}$ , that is, there is a constant  $C$  such that for all  $d$  and  $N$  a set  $P$  of  $N$  points chosen independently and uniformly at random in  $[0, 1)^d$  satisfies  $\mathbb{E}D^*(P) \leq C\sqrt{dN}$ . This is asymptotically tight [Doe14] in the sense that there is a constant  $c > 0$  such that for all  $d$  and all  $N \geq d$  the corresponding random point set satisfies  $\mathbb{E}D^*(P) \geq c\sqrt{dN}$ .

Determining the leading constant  $C$  remains a major open problem. For the upper bound, the original proof of Heinrich et al. does not easily reveal information on  $C$ . Aistleitner [Ais11] gave an alternative, more direct proof that also shows that with positive probability,  $D^*(P) \leq 10\sqrt{dN}$ . The currently strongest estimate, lowering the 10 to 2.525, is due to Gnewuch and Hebbinghaus [GH21]. For the lower bound, the elementary proof of [Doe14] clearly can be made more precise and then give a reasonable constant, but this has not been done so far.

### 1.3 Jittered Sampling

Given the success of random point sets, it is natural to think of constructions that employ randomness in a more clever way than just by taking all decisions independently and uniformly at random. The two most prominent dependent randomized constructions are *Latin hypercube samplings* [MBC79] and *jittered sampling* (also called *stratified sampling*) [Bel81, CPC84]. While the discrepancy of Latin hyper-

cube samples was analyzed only recently [DDG18, GH21] in the paradigm of not treating  $d$  as a constant, the first such analysis for jittered sampling was conducted already in 2004.

Assume that we can write  $N = m^d$  for some integer  $m$ . To obtain a random  $N$ -point set  $P$  via jittered sampling, we partition the unit cube  $[0, 1)^d$  into  $m^d$  axis-parallel cubes of identical size and, independently, place a uniformly distributed random point in each cube.

We formulate all results in the following in terms of  $m$  and  $d$ , and recall that the number  $N$  of points is  $N = m^d$ . To ease comparing the different results, we write all bounds with an explicit term  $\sqrt{dm^d}$  (or  $\sqrt{m^d}$  when  $d$  is treated as a constant), which is the order of magnitude of the discrepancy of a uniform random point set. For constant  $d$ , Beck [Bec87] showed that the expected discrepancy of such a point set  $P$  satisfies

$$\mathbb{E}D^*(P) = O\left(\sqrt{m^d}\sqrt{\frac{\log m}{m}}\right).$$

In the technical report [DGS04] (also described in [Gne12, (28)]), an upper bound of

$$O\left(\sqrt{dm^d}\sqrt{\frac{\log N}{m/d}}\right) = O\left(\sqrt{dm^d}\sqrt{d\frac{\log m}{m/d}}\right)$$

was shown. This bound extends the result of Beck [Bec87] to non-constant  $d$ , but the dependence of the discrepancy on  $d$  is weak due to use of the (as we know from [Gne08]) non-optimal  $\delta$ -covers from [DGS05] and the absence of Aistleitner's [Ais11] dyadic chaining method.

A significant improvement of the dependence on  $d$  and the first lower bound was presented by Pausinger and Steinerberger [PS16], who proved that for all  $d \in \mathbb{N}$  and  $m$  sufficiently large relative to  $d$ ,

$$\frac{1}{10}\sqrt{dm^d}\sqrt{\frac{1}{m/d}} \leq \mathbb{E}D^*(P) \leq \sqrt{dm^d}\sqrt{\frac{\log m}{m/d}}. \quad (2)$$

It is not stated in the paper for which values of  $m$  and  $d$  this bound is valid, that is, what “ $m$  sufficiently large compared to  $d$ ” precisely means, and the authors state it as an open problem to overcome this asymptotic requirement.

Concerning the lower bound, Pausinger and Steinerberger conjecture that it is not tight and speculate that a lower bound of

$$\Omega\left(\sqrt{dm^d}\frac{1 + \sqrt{\log(m)/d}}{\sqrt{m/d}}\right) \quad (3)$$

“might actually be very close to the truth.” Note that this bound is asymptotically stronger than the previous lower bound only for  $m$  superexponential in  $d$ , that is,  $m = 2^{\omega(d)}$ .

## 1.4 Our Result: A Tight Discrepancy Estimate for Jittered Sampling

We show that both the conjectured lower bound (3), and thus also the proven lower bound in (2), and the proven upper bound in (2) are not tight, but that instead the true order of magnitude is

$$\Theta \left( \sqrt{dm^d} \sqrt{\frac{1 + \log(m/d)}{m/d}} \right)$$

for all  $m$  and  $d$  such that  $m \geq d$ .

Our result shows that it is the ratio of  $m$  and  $d$  that describes by how much better jittered sampling is compared to uniform random sampling. For  $m = \Theta(d)$ , the expected star discrepancies are asymptotically the same, namely  $\Theta(\sqrt{dm^d})$ . When  $m = \omega(d)$ , jittered sampling is superior, leading to discrepancies smaller by a factor of  $\Theta\left(\sqrt{\frac{\log(m/d)}{m/d}}\right)$ . Note that the upper bound of [PS16] shows an advantage of jittered sampling only for  $m = \omega(d \log d)$  and “ $m$  large enough compared to  $d$ ”.

In this article, we did not aim at making the leading constant or any lower order terms precise, though in principle we do not see any obstacles for obtaining reasonable absolute bounds for both the lower bound (mainly relying on an estimate on the maximum of independent binomial random variables) and the upper bound (mainly relying on Aistleitner’s dyadic chaining method).

Our result suggests (but we do not prove this) that for  $m < d$ , jittered sampling does not lead to discrepancies of asymptotic order smaller than those of independent random point sets. Since there is good reason to believe that jittered sampling nevertheless has some (small, but gratuitous) advantage, we discuss briefly in Section 5 how to use jittered sampling also for smaller number of points (including less than  $2^d$ ) and what can be said about the discrepancy of such point sets.

## 2 Notation and Preliminaries

Throughout this work, we use the following notation: Given an  $N$ -point set  $P \subseteq [0, 1)^d$ , we denote for each Lebesgue-measurable set  $A$  by  $\lambda(A)$  its Lebesgue measure and by

$$\text{disc}(A) := \text{disc}_P(A) := |P \cap A| - N\lambda(A)$$

the *signed non-normalized discrepancy* of the set  $A$ .

The (non-normalized) *star discrepancy* of the point set  $P$  is  $D^*(P) := \sup_B |\text{disc}(B)|$ , where  $B$  runs over all *axis-aligned rectangles* (boxes) with lower left corner in the origin, that is, all sets  $[0, x) := \prod_{i=1}^d [0, x_i)$ ,  $x \in [0, 1)^d$ .

For a positive integer  $n$ , we use  $[n] := \{1, \dots, n\}$  as shorthand for the set of the first  $n$  positive integers. We also write  $[a..b] := \{z \in \mathbb{Z} \mid a \leq z \leq b\}$ .

For given  $m$  and (usually suppressed)  $d$ , we call  $G_m := \{0, \frac{1}{m}, \dots, \frac{m-1}{m}\}^d$  the *m-grid* (in dimension  $d$ ). For  $x \in G_m$ , we call  $C_x := [x, x + \frac{1}{m}\mathbf{1}_m) = \prod_{i=1}^d [x_i, x_i + \frac{1}{m})$  an *m-cube* (or *cube*) in dimension  $d$ . A *jittered sampling point set* of  $N = m^d$  points is obtained from taking independently and uniformly at random one point from each *m-cube*.

We now prove an elementary lower bound for tail probabilities of the binomial distribution in the special case  $p = 1/2$ . Such estimates can be proven via normal approximations and the Berry-Esseen theorem, see, e.g., [Fel71, XVI.5]. To avoid such deep methods in this otherwise elementary combinatorial paper, we now show the following estimate.

**Lemma 1.** *Let  $k \in \mathbb{N}$ . Let  $X_1, \dots, X_k$  be independent random variables each having a binomial distribution with parameters  $n$  and  $\frac{1}{2}$ . We denote the maximum of these by  $X_{\max} := \max\{X_i \mid i \in [k]\}$ . If  $c \in \mathbb{R}$  is such that*

$$\alpha(c) := \sqrt{\frac{n(\ln k - \frac{1}{2} \ln \ln k - c)}{2(1 + \sqrt{2 \ln(k)/n})}}$$

*is at least  $\sqrt{n}$  and small enough to satisfy  $\alpha(c) + \frac{n}{\alpha(c)} \leq \frac{n}{2}$ , then*

$$\Pr[X_{\max} \geq \frac{n}{2} + \alpha(c)] \geq 1 - \exp\left(-\frac{1}{1.5e^{169/6}\sqrt{\pi}}e^c\right).$$

*Consequently, for  $e^6 \leq k \leq e^{n/2}$ , we have*

$$\mathbb{E}[\max\{0, X_{\max} - \frac{n}{2}\}] \geq \sqrt{\frac{n(\ln k - \ln \ln k)}{2(1 + \sqrt{2 \ln(k)/n})}} \left(1 - \exp\left(-\frac{\sqrt{\ln k}}{1.5e^{169/6}\sqrt{\pi}}\right)\right).$$

*Proof.* Let  $X$  be a random variable having a binomial distribution with parameters  $n$  and  $\frac{1}{2}$ . Let  $\alpha \geq 0$  be such that  $\frac{n}{2} + \alpha \in \mathbb{N}$ . We first give a lower bound for the probability that  $X$  exceeds its expectation by exactly  $\alpha$ . Using Stirling's approximation in the version

$$\sqrt{2\pi n}^{n+\frac{1}{2}} e^{-n} e^{\frac{1}{12n+1}} < n! < \sqrt{2\pi n}^{n+\frac{1}{2}} e^{-n} e^{\frac{1}{12n}}$$

for all  $n \geq 1$  due to Robbins [Rob55] and the elementary estimate  $1 + x \leq e^x$  valid for all  $x \in \mathbb{R}$ , we compute for  $\alpha < \frac{n}{2}$  that

$$\begin{aligned}
\Pr[X = \frac{n}{2} + \alpha] &= 2^{-n} \binom{n}{\frac{n}{2} + \alpha} \\
&= 2^{-n} \frac{n!}{(\frac{n}{2} + \alpha)! (\frac{n}{2} - \alpha)!} \\
&\geq \frac{1}{\sqrt{2\pi} \sqrt{n} (1 + \frac{2\alpha}{n})^{n/2 + \alpha + 1/2} (1 - \frac{2\alpha}{n})^{n/2 - \alpha + 1/2}} \frac{\exp(\frac{1}{12n+1})}{\exp(\frac{1}{12(\frac{n}{2} + \alpha)}) \exp(\frac{1}{12(\frac{n}{2} - \alpha)})} \\
&\geq \frac{1}{\sqrt{2\pi} \sqrt{n} (1 + \frac{2\alpha}{n})^{2\alpha} (1 - (\frac{2\alpha}{n})^2)^{n/2 - \alpha + 1/2}} \frac{1}{\exp(\frac{1}{12}) \exp(\frac{1}{12})} \\
&\geq \frac{1}{e^{1/6} \sqrt{2\pi} \sqrt{n} \exp((2\alpha)^2/n) \exp(-(2\alpha/n)^2(n/2 - \alpha + 1/2))} \\
&\geq \frac{1}{e^{1/6} \sqrt{2\pi} \sqrt{n}} e^{-2\alpha^2/n - 4\alpha^3/n^2}.
\end{aligned}$$

Note that the lower bound in the last line is also valid for  $\alpha = \frac{n}{2}$ , simply because it is less than  $2^{-n}$ .

Let  $\alpha \geq \sqrt{n}$  with  $\alpha + \frac{n}{\alpha} \leq \frac{n}{2}$ . We now estimate the probability that  $X$  exceeds its expectation by at least  $\alpha$  via the probability that  $X \in [\mathbb{E}X + \alpha, \mathbb{E}X + \alpha + \frac{n}{\alpha}]$ . Using that  $\binom{n}{\frac{n}{2} + \alpha}$  and thus  $\Pr[X = \frac{n}{2} + \alpha]$  is decreasing in  $\alpha$  for integral  $\alpha \geq 0$ , we compute

$$\begin{aligned}
\Pr[X \geq \frac{n}{2} + \alpha] &\geq \Pr[\frac{n}{2} + \alpha \leq X \leq \frac{n}{2} + \alpha + \frac{n}{\alpha}] \\
&\geq \lfloor \frac{n}{\alpha} \rfloor \Pr[X = \lfloor \frac{n}{2} + \alpha + \frac{n}{\alpha} \rfloor] \\
&\geq \frac{1}{e^{1/6} \sqrt{2\pi} \sqrt{n}} \lfloor \frac{n}{\alpha} \rfloor e^{-2(\alpha + n/\alpha)^2/n - 4(\alpha + n/\alpha)^3/n^2} \\
&\geq \frac{1}{1.5e^{169/6} \sqrt{2\pi}} \frac{\sqrt{n}}{\alpha} e^{-2\alpha^2/n - 4\alpha^3/n^2},
\end{aligned} \tag{4}$$

where the last inequality uses  $\alpha \geq \sqrt{n}$ ,  $\alpha \leq \frac{n}{2}$ , and  $\alpha \geq 1$ , which give the estimates  $2(\alpha + n/\alpha)^2/n = 2\frac{\alpha^2}{n} + 4 + 2\frac{n}{\alpha^2} \leq 2\frac{\alpha^2}{n} + 6$  and  $4(\alpha + n/\alpha)^3/n^2 = 4\frac{\alpha^3}{n^2} + 12\frac{\alpha}{n} + 12\frac{1}{\alpha} + 4\frac{n}{\alpha^3} \leq 4\frac{\alpha^3}{n^2} + 6 + 12 + \frac{4}{\alpha} \leq 4\frac{\alpha^3}{n^2} + 22$ .

Consequently, the probability that all of  $X_1, \dots, X_k$  are below  $\frac{n}{2} + \alpha$  is

$$\begin{aligned}
\Pr[X_{\max} < \frac{n}{2} + \alpha] &= \Pr[\forall i \in [k] : X_i < \frac{n}{2} + \alpha] \\
&\leq \left( 1 - \frac{1}{1.5e^{169/6} \sqrt{2\pi}} \frac{\sqrt{n}}{\alpha} e^{-2\alpha^2/n - 4\alpha^3/n^2} \right)^k \\
&\leq \exp \left( - \frac{k}{1.5e^{169/6} \sqrt{2\pi}} \frac{\sqrt{n}}{\alpha} e^{-2\alpha^2/n - 4\alpha^3/n^2} \right).
\end{aligned}$$



Let  $c \in \mathbb{R}$  be such that

$$\alpha = \alpha(c) = \sqrt{\frac{\frac{1}{2}n(\ln k - \frac{1}{2} \ln \ln k - c)}{1 + \sqrt{2 \ln(k)/n}}}$$

satisfies  $\alpha \geq \sqrt{n}$  and  $\alpha + \frac{n}{\alpha} \leq \frac{n}{2}$ . With  $\alpha \leq \sqrt{\frac{1}{2}n \ln k}$ , we continue the previous estimate and compute

$$\begin{aligned} & \exp\left(-\frac{k}{1.5e^{169/6}\sqrt{2\pi}} \frac{\sqrt{n}}{\alpha} e^{-(2\alpha^2/n)(1+2\alpha/n)}\right) \\ & \leq \exp\left(-\frac{k}{1.5e^{169/6}\sqrt{2\pi}} \frac{\sqrt{n}}{\sqrt{\frac{1}{2}n \ln k}} \exp\left(-\frac{(\ln k - \frac{1}{2} \ln \ln k - c)(1 + 2\alpha/n)}{1 + \sqrt{2 \ln(k)/n}}\right)\right) \\ & \leq \exp\left(-\frac{k}{1.5e^{169/6}\sqrt{\pi}} \frac{1}{\sqrt{\ln k}} \exp(-(\ln k - \frac{1}{2} \ln \ln k - c))\right) \\ & = \exp\left(-\frac{1}{1.5e^{169/6}\sqrt{\pi}} e^c\right). \end{aligned}$$

To prove the second claim, where  $e^6 \leq k \leq e^{n/2}$ , let  $c_k = \frac{1}{2} \ln \ln k$  and

$$\alpha = \alpha_k = \alpha(c_k) = \sqrt{\frac{\frac{1}{2}n(\ln k - \ln \ln k)}{1 + \sqrt{2 \ln(k)/n}}}.$$

We note that for  $k \in [e^6, e^{n/2}]$ , we have  $\alpha_k \geq \sqrt{\frac{\frac{1}{2}n(\ln e^6 - \ln \ln e^6)}{1 + \sqrt{2 \ln(e^{n/2})/n}}} \geq \sqrt{n}$ ; here we used that  $k \mapsto \ln k - \ln \ln k$  is increasing for  $k \geq e$ . For  $\alpha \geq \sqrt{n}$ , the expression  $\alpha + \frac{n}{\alpha}$  is increasing in  $\alpha$ . Hence noting that  $\alpha_k \leq \sqrt{\frac{\frac{1}{2}n \ln k}{1 + \sqrt{2 \ln(k)/n}}} = \frac{n}{2} \sqrt{\frac{2 \ln(k)/n}{1 + \sqrt{2 \ln(k)/n}}} \leq \frac{n}{2} \sqrt{\frac{2 \ln(e^{n/2})/n}{1 + \sqrt{2 \ln(e^{n/2})/n}}} \leq \frac{n}{\sqrt{8}}$  when  $k \leq e^{n/2}$  – here we used the fact that  $x \mapsto \frac{x}{1 + \sqrt{x}}$  is increasing in  $\mathbb{R}_{\geq 0}$  –, we estimate  $\alpha_k + \frac{n}{\alpha_k} \leq \frac{n}{\sqrt{8}} + \frac{n}{n/\sqrt{8}} \leq \frac{n}{\sqrt{8}} + \sqrt{8}$ , which is at most  $\frac{n}{2}$  when  $n \geq 20$ . Hence for  $n \geq 20$  and any  $k \in [e^6, e^{n/2}]$ , we can use the first claim of this lemma and compute

$$\begin{aligned} \mathbb{E}[\max\{0, X_{\max} - \frac{n}{2}\}] & \geq \alpha_k \cdot \Pr[X_{\max} \geq \frac{n}{2} + \alpha_k] \\ & \geq \sqrt{\frac{n(\ln k - \ln \ln k)}{2(1 + \sqrt{2 \ln(k)/n})}} \left(1 - \exp\left(-\frac{\sqrt{\ln k}}{1.5e^{169/6}\sqrt{\pi}}\right)\right). \end{aligned}$$

For  $n < 20$ , the second claim is trivially fulfilled as the following two estimates show (where the latter again uses  $e^x \geq 1 + x$ , valid for all  $x \in \mathbb{R}$ ).

$$E[\max\{0, X_{\max} - \frac{n}{2}\}] \geq E[\max\{0, X_1 - \frac{n}{2}\}] \geq \frac{n}{2} \Pr[X_1 = n] = 2^{-n} \frac{n}{2} \geq 2^{-19} \frac{n}{2}.$$

$$\begin{aligned}
& \sqrt{\frac{n(\ln k - \ln \ln k)}{2(1 + \sqrt{2 \ln(k)/n})}} \left( 1 - \exp \left( - \frac{\sqrt{\ln k}}{1.5e^{169/6} \sqrt{\pi}} \right) \right) \\
& \leq \sqrt{\frac{n \ln k}{2}} \left( 1 - 1 - \left( \frac{\sqrt{\ln k}}{1.5e^{169/6} \sqrt{\pi}} \right) \right) \\
& \leq \sqrt{\frac{1}{2} n \ln e^{n/2}} \frac{\sqrt{\ln e^{n/2}}}{1.5e^{169/6} \sqrt{\pi}} < \frac{n}{2} \sqrt{8} e^{-28} < 2^{-19} \frac{n}{2}.
\end{aligned}$$

□

### 3 Proof of the Lower Bound

In this section, we prove that for all  $m \geq d$  the discrepancy of a jittered sampling point set having  $m^d$  point in  $[0, 1]^d$  is at least of order  $\sqrt{dm^d} \sqrt{\frac{1 + \log(m/d)}{m/d}}$ .

**Theorem 2.** *There is a constant  $C > 0$  such that for all  $m, d \in \mathbb{N}_{\geq 2}$  with  $m \geq d$ , the expected discrepancy of a jittered sampling point set  $P \subset [0, 1]^d$ ,  $|P| = m^d$ , is at least*

$$\mathbb{E}D^*(P) \geq C \sqrt{dm^d} \sqrt{\frac{1 + \log(m/d)}{m/d}}.$$

To ease the presentation, we treat the “small” case that  $m$  is at most a constant factor larger than  $d$  separately in Lemma 6. The more interesting case, naturally, is that  $m$  is of larger order than  $d$ . For this, we prove the following result, which we state in a non-asymptotic fashion, noting again that in this work we did not optimize the leading constant or the lower order terms.

**Lemma 3.** *Let  $m, d \in \mathbb{N}_{\geq 2}$  with  $\lfloor \frac{m}{d} \rfloor \geq e^6$ , and  $N = m^d$ . Let  $P$  be a random  $N$ -point set in  $[0, 1]^d$  obtained from jittered sampling. Then*

$$\begin{aligned}
\mathbb{E}D^*(P) & \geq (2e)^{-\frac{1}{2}} dm^{\frac{d-1}{2}} \sqrt{\ln(\lfloor \frac{m}{d} \rfloor) - \ln \ln(\lfloor \frac{m}{d} \rfloor)} \\
& \left( 1 + \sqrt{\frac{2 \ln(\lfloor \frac{m}{d} \rfloor)}{(m - \lfloor \frac{m}{d} \rfloor)^{d-1}} \right)^{-\frac{1}{2}} \left( 1 - \exp \left( - \frac{\sqrt{\ln \lfloor \frac{m}{d} \rfloor}}{1.5e^{169/6} \sqrt{\pi}} \right) \right).
\end{aligned}$$

*In particular, there is a  $C > 0$  such that for all  $m, d \in \mathbb{N}_{\geq 2}$  with  $\lfloor \frac{m}{d} \rfloor \geq e^6$  a jittered sampling point set  $P$  with  $N = m^d$  points satisfies  $\mathbb{E}D^*(P) \geq C \sqrt{dm^d} \sqrt{\frac{1 + \log(m/d)}{m/d}}$ .*

While we did not optimize for the leading constant, our result can be written as  $\mathbb{E}D^*(P) \geq (1 - f(\frac{m}{d}))(2e)^{-\frac{1}{2}} \sqrt{dm^d} \sqrt{\frac{\ln(m/d)}{m/d}}$ , where  $f$  is a function tending to

zero when the argument tends to infinity. Consequently, our leading constant of  $(2e)^{-\frac{1}{2}} \geq 0.4288$  is not too bad. We note that Pausinger and Steinerberger [PS16] state a constant of  $1/10$  in their theorem. An inspection of their proof shows that they actually prove their result with a leading constant of  $0.5\sqrt{\pi/2}\ln(2) = 0.4343\dots$

In the proof of Lemma 3, we use the elementary observation that any measurable set has expected signed discrepancy zero when all cubes intersecting it contain exactly one random point distributed uniformly in the cube.

**Lemma 4.** *Let  $A \subseteq [0, 1]^d$  be a measurable set. Let  $P$  be a set of  $N = m^d$  random points such that each  $m$ -cube having non-empty intersection with  $A$  contains exactly one point of  $P$  and this point is uniformly distributed in this cube. Then  $\mathbb{E} \text{disc}(A) = 0$ .*

*Proof.* Let  $\mathcal{Q}$  be the set of all cubes and  $\mathcal{Q}_A$  the set of cubes having non-empty intersection with  $A$ . Then

$$\text{disc}(A) = \sum_{Q \in \mathcal{Q}} \text{disc}(A \cap Q) = \sum_{Q \in \mathcal{Q}_A} \text{disc}(A \cap Q).$$

Let  $Q \in \mathcal{Q}_A$ . With probability  $\lambda(A \cap Q)/\lambda(Q)$ , the random point in  $Q$  lies also in  $A$  and we have  $\text{disc}(A \cap Q) = 1 - N\lambda(A \cap Q)$ . Otherwise,  $P \cap (A \cap Q)$  is empty, giving  $\text{disc}(A \cap Q) = -N\lambda(A \cap Q)$ . Consequently,

$$\begin{aligned} \mathbb{E} \text{disc}(A \cap Q) &= \frac{\lambda(A \cap Q)}{\lambda(Q)}(1 - N\lambda(A \cap Q)) - \left(1 - \frac{\lambda(A \cap Q)}{\lambda(Q)}\right)N\lambda(A \cap Q) \\ &= \frac{\lambda(A \cap Q)}{\lambda(Q)} - N\lambda(A \cap Q) = 0, \end{aligned}$$

where the last equality follows from  $\lambda(Q) = 1/N$ . By linearity of expectation,  $\mathbb{E} \text{disc}(A) = \sum_{Q \in \mathcal{Q}_A} \mathbb{E} \text{disc}(A \cap Q) = 0$ .  $\square$

We shall use the above observation to combine certain rectangles with known signed discrepancy into an anchored box with expected discrepancy signed equal to the sum of the discrepancies of these rectangles.

**Lemma 5.** *Let  $r_1, \dots, r_d \in [0, 1]$  be integer multiples of  $1/m$ . Let  $B_0 := [0, r]$ , where  $r = (r_1, \dots, r_d)$ . Let  $P$  be a random set of  $m^d$  points in  $[0, 1]^d$  obtained from jittered sampling. For each  $i \in [d]$ , let  $S_i = S_i(P)$  be a random variable taking values in  $[r_i, 1)$  and let  $R_i = \prod_{j=1}^{i-1} [0, r_j] \times [r_i, S_i) \times \prod_{j=i+1}^d [0, r_j]$ . Assume that each  $R_i$ , i.e., the number  $S_i$ , is independent of the position of all points outside  $\bar{R}_i := \prod_{j=1}^{i-1} [0, r_j] \times [r_i, 1) \times \prod_{j=i+1}^d [0, r_j]$ . Let  $B = \prod_{i=1}^d [0, S_i)$  be the smallest anchored box containing the  $R_i$ . Then*

$$\mathbb{E} D^*(P) \geq \mathbb{E} \text{disc}(B) = \sum_{i=1}^d \mathbb{E} \text{disc}(R_i).$$

Note that the lemma in particular covers the case that the  $S_i$  are chosen as to maximize the signed discrepancy of the  $R_i$ . Consequently, with this lemma we can construct a box with large (expected) discrepancy by finding  $R_i$  with large signed discrepancy. We speculate that this *construction principle* can be useful in other lower bound proofs for jittered sampling as well.

*Proof of Lemma 5.* Let us first condition on a fixed outcome of the  $S_i$ , that is, let  $s_i \in [r_i, 1)$  for all  $i \in [1..d]$  and we condition on  $S_i = s_i$  for all  $i \in [1..d]$ . In this conditional probability space, by construction, all points in cubes not contained in  $\bar{R} := \bigcup_{i \in [d]} \bar{R}_i$  are uniformly distributed in their cube. Also, no cube intersects both  $A := B \setminus \bigcup_{i \in [d]} R_i$  and  $\bar{R}$ . Hence all points in cubes with non-empty intersection with  $A$  are distributed uniformly in their cube. By Lemma 4 we have  $\mathbb{E}[\text{disc}(A) \mid S_1 = s_1, \dots, S_d = s_d] = 0$ . We thus have

$$\begin{aligned} & \mathbb{E}[\text{disc}(B) \mid S_1 = s_1, \dots, S_d = s_d] \\ &= \mathbb{E} \left[ \text{disc}(A) + \sum_{i=1}^d \text{disc}(R_i) \mid S_1 = s_1, \dots, S_d = s_d \right] \\ &= \sum_{i=1}^d \mathbb{E}[\text{disc}(R_i) \mid S_1 = s_1, \dots, S_d = s_d]. \end{aligned}$$

Hence the law of total expectation gives

$$\begin{aligned} \mathbb{E} \text{disc}(B) &= \mathbb{E}[\mathbb{E}[\text{disc}(B) \mid S_1 = s_1, \dots, S_d = s_d]] \\ &= \mathbb{E} \left[ \sum_{i=1}^d \mathbb{E}[\text{disc}(R_i) \mid S_1 = s_1, \dots, S_d = s_d] \right] \\ &= \sum_{i=1}^d \mathbb{E}[\mathbb{E}[\text{disc}(R_i) \mid S_1 = s_1, \dots, S_d = s_d]] \\ &= \sum_{i=1}^d \mathbb{E} \text{disc}(R_i). \end{aligned}$$

We thus have  $\mathbb{E}D^*(P) \geq \mathbb{E}|\text{disc}(B)| \geq \mathbb{E} \text{disc}(B) = \sum_{i=1}^d \mathbb{E} \text{disc}(R_i)$ . □

Before giving the precise proof of Lemma 3, let us give a brief outline of the main ideas and compare them to the proof of Pausinger and Steinerberger [PS16]. The main argument of [PS16] is the following. Let  $x_1, \dots, x_d \in [\frac{m-1}{m}, 1)$  and  $x = (x_1, \dots, x_d)$ . If  $m$  is sufficiently large compared to  $d$ , then the discrepancy of the box  $B = [0, x)$  is very close to the discrepancy of the union  $R := R_1 \cup \dots \cup R_d$  of the slices  $R_i := R_i(x_i) := [0, \frac{m-1}{m})^{i-1} \times [\frac{m-1}{m}, x_i) \times [0, \frac{m-1}{m})^{d-i}$ . Note that here

the above lemma would have directly shown that the expected star discrepancy of  $P$  is at least the discrepancy of  $R$ . By construction, the discrepancy of  $R_i(x_i)$  has the same distribution as the discrepancy of the interval  $[0, m(x_i - \frac{m-1}{m}))$  in a one-dimensional uniformly distributed random set of  $(m-1)^{d-1}$  points in  $[0, 1)$ . This one-dimensional discrepancy problem can be analyzed quite well, in particular, the expected maximum discrepancy (over all choices of  $x_i$ ) can be determined. Consequently, the expected maximum discrepancy of a suitable choice of  $B$  is at least  $d/2$  times this number (the factor of  $1/2$  stems from the fact that we need the discrepancies of the  $R_i$  to have the same sign).

From a broader perspective, the main idea of [PS16] is to regard all boxes  $B = [0, x)$  with  $x$  lying in the upper right sub-cube  $C^+ := [\frac{m-1}{m}, 1)^d$  – and only these – and to then exploit that for these the discrepancy is well described by one-dimensional discrepancies, which can be analyzed very precisely. While the reduction to the one-dimensional discrepancy problem allow a very precise analysis of the maximum discrepancy of a box  $[0, x)$ ,  $x \in C^+$ , this approach carries the risk that the restricted choice of boxes underestimates the star discrepancy significantly.

For this reason, we follow a different road. We do not restrict ourselves to boxes  $[0, x)$ ,  $x \in C^+$ , but overcome the increased complexity of the larger range for  $x$  by restricting ourselves to a suitable *discrete* set of choices for  $x$ . Taking  $r = 1 - \frac{1}{m} \lfloor \frac{m}{d} \rfloor \approx 1 - \frac{1}{d}$ , our  $x$  will be such that all  $x_i$  are in  $[r, 1)$  and are integral multiples of  $\frac{1}{2m}$ . Clearly, with this relatively small discrete set of boxes, our bounds will necessarily be off the truth by constant factors. However, the more diverse set of boxes together with the right (not very difficult) combinatorial way of selecting a large-discrepancy box among them will enable us to prove the stronger (and in fact asymptotically tight) lower bound.

The main combinatorial observation is that for all  $j = 0, \dots, \lfloor \frac{m}{d} \rfloor - 1$ , the rectangle  $R'_1 = [r + \frac{j}{m}, r + \frac{j}{m} + \frac{1}{2m}) \times [0, r)^{d-1}$  has the same discrepancy distribution, which is the deviation of a binomial random variable with parameters  $N' \geq \frac{N}{em}$  and  $p = \frac{1}{2}$  from its expectation. Consequently, by elementary properties of the maximum of  $\lfloor \frac{m}{d} \rfloor$  independent binomial random variables (Lemma 1), with good probability there is a choice for  $j$  such that the signed discrepancy of  $R'_1$  satisfies  $\text{disc}(R'_1) \geq C \sqrt{N \log(\frac{m}{d})}$  for some absolute constant  $C$ . By construction, the discrepancy of  $R'_1$  is identical to the one of  $R_1 := [r, r + \frac{j}{m} + \frac{1}{2m}) \times [0, r)^{d-1}$ . Repeating this argument in each dimension and taking as  $B$  the smallest anchored box that contains all  $R_i$ , with Lemma 5 again we obtain a box with  $\mathbb{E} \text{disc}(B) \geq Cd \sqrt{N \log(\frac{m}{d})}$  as desired.

*Proof of Lemma 3.* Let  $k := \lfloor \frac{m}{d} \rfloor$  and  $r = \frac{m-k}{m} \geq 1 - \frac{1}{d}$ . For each  $i \in [d]$  let  $S_i \in [r_i, 1)$  be (maximal, in case of ambiguity) such that  $R_i := [0, r)^{i-1} \times [r, S_i) \times [0, r)^{d-i}$  has maximum signed discrepancy  $\text{disc}(R_i) := |P \cap R_i| - N \lambda(R_i)$ . Note that  $\text{disc}(R_i)$  can only be maximal when  $S_i$  coincides with the  $i$ -th coordinate of some point of

$P$ , which shows that  $S_i$  is a well-defined random variable. Note further that  $\text{disc}(R_i) \geq 0$  since taking  $S_i = 1$  would give a rectangle containing only full cubes, and hence, having discrepancy zero.

Let  $B := \prod_{i \in [d]} [0, S_i]$  be the smallest box containing all  $R_i$ . Note that the value of the  $S_i$  depends only on the position of the points in the cubes in  $\bar{R}_i := [0, r)^{i-1} \times [r, 1) \times [0, r)^{d-i}$ . By Lemma 5, we have

$$\mathbb{E}D^*(P) \geq \sum_{i=1}^d \mathbb{E} \text{disc}(R_i) = d \mathbb{E} \text{disc}(R_1), \quad (5)$$

where the last equality exploits the symmetry between the  $R_i$ .

So it suffices to analyze  $\text{disc}(R_1)$ . For  $j = 0, \dots, k-1$ , let  $y_j := r + \frac{j}{m}$  and  $z_j := y_j + \frac{1}{2m}$ . Let  $U_j := [r, z_j) \times [0, r)^{d-1}$  and  $T_j = [y_j, z_j) \times [0, r)^{d-1}$ . Since  $U_j \setminus T_j$  can be written as union of cubes, we have  $\text{disc}(U_j) = \text{disc}(T_j)$ . Now each  $T_j$  is composed of the “left” halves of exactly  $N' = (m-k)^{d-1} \geq (1-\frac{1}{d})^{d-1} m^{d-1} \geq \frac{1}{e} m^{d-1}$  cubes. Consequently,  $|P \cap T_j|$  follows a binomial distribution with parameters  $N'$  and  $p = \frac{1}{2}$ . Since no cube intersects non-trivially two different  $T_j$ , the discrepancies of the  $T_j$  (and thus of the  $U_j$ ) are independent. Let  $X_j$ ,  $j = 0, \dots, k-1$ , be independent random variables with binomial distribution with parameters  $N'$  and  $p = \frac{1}{2}$ . Then  $\text{disc}(R_1) \geq \max\{0, \text{disc}(U_j) \mid j = 0, \dots, k-1\}$  and the latter is distributed as  $\max\{0, X_j - \mathbb{E}X_j \mid j = 0, \dots, k-1\}$ .

Since  $k \geq e^6$  by assumption and further  $k \leq m-k \leq N' \leq e^{N'/2}$ , by the second part of Lemma 1 we have

$$\begin{aligned} \mathbb{E} \text{disc}(R_1) &\geq \sqrt{\frac{N'(\ln k - \ln \ln k)}{2(1 + \sqrt{2 \ln(k)/N'})}} \left(1 - \exp\left(-\frac{\sqrt{\ln k}}{1.5e^{169/6}\sqrt{\pi}}\right)\right) \\ &\geq (2e)^{-\frac{1}{2}} m^{\frac{d-1}{2}} \sqrt{\ln(\lfloor \frac{m}{d} \rfloor) - \ln \ln(\lfloor \frac{m}{d} \rfloor)} \left(1 + \sqrt{\frac{2 \ln(\lfloor \frac{m}{d} \rfloor)}{(m - \lfloor \frac{m}{d} \rfloor)^{d-1}}}\right)^{-\frac{1}{2}} \\ &\quad \left(1 - \exp\left(-\frac{\sqrt{\ln(\lfloor \frac{m}{d} \rfloor)}}{1.5e^{169/6}\sqrt{\pi}}\right)\right). \end{aligned}$$

Together with (5), this proves the main claim. The last claim follows from the following three estimates.

- We note that  $f : x \mapsto \sqrt{x} - \ln x$  has its global minimum (in  $\mathbb{R}_{>0}$ ) at  $x = 4$ . Since  $f(4) > 0$ , we have  $\sqrt{x} > \ln(x)$  for all  $x \in \mathbb{R}_{>0}$ . Consequently,  $\ln \ln(\lfloor \frac{m}{d} \rfloor) \leq \frac{1}{2} \ln(\lfloor \frac{m}{d} \rfloor)$  and thus  $\sqrt{\ln(\lfloor \frac{m}{d} \rfloor) - \ln \ln(\lfloor \frac{m}{d} \rfloor)} \geq \sqrt{\frac{1}{2} \ln(\lfloor \frac{m}{d} \rfloor)}$ . Since  $\lfloor \frac{m}{d} \rfloor \geq e^6$ , we can continue with  $\sqrt{\frac{1}{2} \ln(\lfloor \frac{m}{d} \rfloor)} = \sqrt{\frac{1}{4} \ln(\lfloor \frac{m}{d} \rfloor)^2} \geq \frac{1}{2} \sqrt{1 + \ln(\frac{m}{d})}$ .

- From  $d \geq 2$  and  $\lfloor \frac{m}{d} \rfloor \geq e^6$ , we conclude

$$\begin{aligned}
1 + \sqrt{\frac{2 \ln(\lfloor \frac{m}{d} \rfloor)}{(m - \lfloor \frac{m}{d} \rfloor)^{d-1}}} &\leq 1 + \sqrt{\frac{2 \ln(\lfloor \frac{m}{d} \rfloor)}{(2\lfloor \frac{m}{d} \rfloor - \lfloor \frac{m}{d} \rfloor)^{d-1}}} \\
&\leq 1 + \sqrt{\frac{2 \ln(\lfloor \frac{m}{d} \rfloor)}{\lfloor \frac{m}{d} \rfloor}} \\
&\leq 1 + \sqrt{\frac{2 \ln(e^6)}{e^6}} \leq 1.18.
\end{aligned}$$

Here we used that  $x \mapsto \frac{\ln(x)}{x}$  is decreasing for  $x \geq e$ .

- Finally,  $\lfloor \frac{m}{d} \rfloor \geq e^6$  implies that

$$1 - \exp\left(-\frac{\sqrt{\ln(\lfloor \frac{m}{d} \rfloor)}}{1.5e^{169/6}\sqrt{\pi}}\right) \geq 1 - \exp\left(-\frac{\sqrt{\ln(e^6)}}{1.5e^{169/6}\sqrt{\pi}}\right)$$

is at least some positive constant.

□

We now discuss the case that  $m$  is of similar order of magnitude as  $d$ . The following result, in particular, extends the  $\Omega(dm^{\frac{d-1}{2}})$  lower bound of Lemma 3 to arbitrary  $m = \Theta(d)$  and shows that in this case, apart from constant factors, jittered sampling point sets have an expected discrepancy not smaller than uniformly distributed random point sets. In other words, only for  $N = \omega(d)^d$  jittered sampling has a super-constant discrepancy advantage over fully random points.

**Lemma 6.** *Let  $m, d \in \mathbb{N}_{\geq 2}$ . Let  $N = m^d$  and  $P$  be a random  $N$ -point set obtained from jittered sampling. Then*

$$\mathbb{E}D^*(P) \geq \frac{4}{5e^{8+1/6}\sqrt{2\pi}} \exp\left(-\frac{32}{(m-1)^{\frac{d-1}{2}}}\right) d(m-1)^{\frac{d-1}{2}}.$$

*In particular, if  $m \geq \gamma d$  for some constant  $\gamma > 0$ , then  $\mathbb{E}D^*(P) \geq Cdm^{\frac{d-1}{2}}$  for some constant  $C > 0$  which depends only on  $\gamma$ .*

*If  $\Gamma d \geq m \geq \gamma d$  for some constants  $\Gamma \geq \gamma > 0$ , then  $\mathbb{E}D^*(P) \geq C'\sqrt{dN}$  for a constant  $C' > 0$  which depends only on  $\gamma$  and  $\Gamma$ .*

*Proof.* We use a simplified version of the proof of Lemma 3. Let  $r = \frac{m-1}{m}$ . For each  $i \in [d]$ , choose  $x_i \in [r, 1)$  such that  $R_i := [0, r)^{i-1} \times [r, x_i) \times [0, r)^{d-i}$  has maximum signed discrepancy  $\text{disc}(R_i) = |P \cap R_i| - N\lambda(R_i)$ . Let  $x = (x_1, \dots, x_d)$  and  $B :=$

$\prod_{i \in [d]} [0, x_i)$ . As in the proof of Lemma 3, we have  $\mathbb{E}D^*(P) \geq d \mathbb{E} \text{disc}(R_1)$ , so again it only remains to estimate  $\mathbb{E} \text{disc}(R_1)$ .

Clearly, we have  $\mathbb{E} \text{disc}(R_1) \geq \mathbb{E} \max\{0, \text{disc}(R'_1)\}$  with  $R'_1 := [r, r + \frac{1}{2m}) \times [0, r)^{d-1}$ . Now  $\text{disc}(R'_1)$  follows a binomial distribution with parameters  $N' = (m-1)^{d-1}$  and  $p = \frac{1}{2}$ . If  $N' \geq 16$ , then for  $\alpha = \sqrt{N'}$  we have  $\alpha + N'/\alpha \leq N'/2$  and thus equation (4) shows that with probability at least  $\frac{4}{5e^{8+1/6}\sqrt{2\pi}}e^{-32/\sqrt{N'}}$  we have a discrepancy of  $\sqrt{N'}$  or more. Consequently,  $\mathbb{E} \text{disc}(R_1) \geq \mathbb{E} \max\{0, \text{disc}(R'_1)\} \geq \frac{4}{5e^{8+1/6}\sqrt{2\pi}}e^{-32/\sqrt{N'}}\sqrt{N'}$ .

If  $N' < 16$ , then we note that  $R'_1 := [r, r + \frac{1}{2N'm}) \times [0, r)^{d-1}$  with probability  $1 - (1 - \frac{1}{2N'})^{N'} \geq 1 - \exp(-\frac{1}{2}) \geq 0.39$  contains at least one point. Hence  $\mathbb{E} \text{disc}(R_1) \geq \mathbb{E} \max\{0, \text{disc}(R'_1)\} \geq 0.39(1-0.5) \geq 0.19 \geq \frac{4}{5e^{8+1/6}\sqrt{2\pi}} \cdot 4 \geq \frac{4}{5e^{8+1/6}\sqrt{2\pi}}e^{-32/\sqrt{N'}}\sqrt{N'}$ .

Hence in either case,

$$\begin{aligned} \mathbb{E}D^*(P) &\geq d \mathbb{E} \text{disc}(R_1) \geq d \frac{4}{5e^{8+1/6}\sqrt{2\pi}}e^{-32/\sqrt{N'}}\sqrt{N'} \\ &= \frac{4}{5e^{8+1/6}\sqrt{2\pi}} \exp\left(-\frac{32}{(m-1)^{\frac{d-1}{2}}}\right) d(m-1)^{\frac{d-1}{2}}. \end{aligned}$$

If  $m \geq \gamma d$  for some constant  $\gamma > 0$ , then using  $m \geq 2$  we further estimate the above bound to  $\mathbb{E}D^*(P) \geq \frac{4e^{-32}}{5e^{8+1/6}\sqrt{2\pi}} dm^{\frac{d-1}{2}} (1 - \frac{1}{m})^{\frac{d-1}{2}}$ . Noting that  $(1 - \frac{1}{m})^m$  is increasing in  $m$  and, again, that  $m \geq 2$ , we estimate  $(1 - \frac{1}{m})^{\frac{d-1}{2}} \geq (1 - \frac{1}{m})^{m/2\gamma} \geq 2^{-1/\gamma}$ . Hence there is a constant  $C > 0$  depending on  $\gamma$  only such that  $\mathbb{E}D^*(P) \geq Cdm^{\frac{d-1}{2}}$ . When also  $m \leq \Gamma d$  for some constant  $\Gamma$ , then estimating  $dm^{\frac{d-1}{2}} \geq d^{1/2}(\frac{m}{\Gamma})^{1/2}m^{\frac{d-1}{2}} = \Gamma^{-1/2}\sqrt{dN}$  shows that we also have  $\mathbb{E}D^*(P) \geq C'\sqrt{dN}$  for some constant  $C'$  which only depends on  $\gamma$  and  $\Gamma$ .  $\square$

## 4 Proof of the Upper Bound

We now show that the lower bound proven above is tight (apart from constant factors independent of  $m$  and  $d$ ).

**Theorem 7.** *Let  $m, d \in \mathbb{N}$  with  $m \geq d \geq 2$ . Let  $N := m^d$ . Let  $P$  be a random set of  $N$  points in  $[0, 1)^d$  obtained from jittered sampling. Then*

$$\mathbb{E}D^*(P) \leq 60.9984\sqrt{dm^d} \frac{\sqrt{\ln(4em/d)} + 2.9599}{\sqrt{m/d}}.$$

The main reason why the upper bound proof of Pausinger and Steinerberger [PS16] does not give the right order of magnitude is the following. Using a similar reduction to one-dimensional discrepancies as in their lower bound proof,



Pausinger and Steinerberger again are able to give a strong bound (including an exponentially decreasing tail) for the maximum discrepancy among all boxes with upper right corner lying in the same  $m$ -cube. To obtain an upper bound valid for all boxes, a union bound is employed. Such a union bound, naturally, ignores any positive correlation between the discrepancies of boxes with corner point in different, but close-by  $m$ -cubes. So from a broader perspective, the proof again does a very precise analysis inside the  $m$ -cubes, but ignores the overall combinatorial structure of the problem.

Two arguments have been used in the past to better exploit the positive correlation of similar boxes. Heinrich et al. [HNWW01] used deep results of Talagrand and Haussler from the theory of empirical processes to give the first proof of the  $O(\sqrt{dN})$  discrepancy bound for  $N$  independent uniformly distributed random points in  $[0, 1]^d$ . With a non-trivial, purely combinatorial decomposition argument called *dyadic chaining*, Aistleitner [Ais11] and later Gnewuch and Hebbinghaus [GH21] reproved this bound and gave explicit (and small) values for the leading constant

In the following, we show that Aistleitner’s dyadic chaining technique can also be used for the non-uniformly distributed point sets stemming from jittered sampling. The main difference, and reason for the stronger discrepancy bound, comes from noting that the grid points  $\Gamma_0 = \{0, \frac{1}{m}, \frac{2}{m}, \dots, 1\}^d$  form a  $\frac{d}{m}$ -cover (see below for a definition) such that each grid point  $x \in \Gamma_0$  defines a rectangle  $[0, x)$  with discrepancy 0. Consequently, we can start the dyadic chaining construction with these grid points as coarsest cover. Note that this is not a very efficient (that is, small) cover, but this has no influence on the overall efficiency of the construction as one can verify from the proof below.

The stochastic dependencies in the jittered sampling random experiment, interestingly, impose no additional difficulties. Still the number of points in an arbitrary measurable set can be written as sum of independent 0, 1 random variables. Unlike in the proof of [Ais11], these are not identically distributed, but this has no influence on the applicability of most Chernoff-type large deviation bounds.

Of course, in addition to these observations, it remains to estimate the expected star discrepancy in an analogous way as done in [Ais11], which requires some care. As for the lower bound in the previous section, in this article we do not take great care for obtaining a small leading constant, thus simplifying some calculations as compared to [Ais11] and [GH21]. The reader familiar with Aistleitner’s proof will also note that we prefer to work with half-open rectangles, but clearly this makes no difference.

We recall the definitions of  $\delta$ -covers and  $\delta$ -bracketing covers. We use the notation  $\overline{[x, y)} := [0, y) \setminus [0, x)$  for all  $x, y \in [0, 1]^d$  with  $x \leq y$ . Let  $\delta > 0$ . A set  $\Gamma \subseteq [0, 1]^d$  is called  $\delta$ -cover if for each  $y \in [0, 1]^d$  there are  $x, z \in \Gamma \cup \{0\}$  such

that  $x \leq y \leq z$  and  $\lambda(\overline{[x, z]}) \leq \delta$ . In particular, and this is what we will need only, we have  $\lambda(\overline{[x, y]}) \leq \delta$ . A set  $\Delta \subseteq ([0, 1]^d)^2$  is called  $\delta$ -bracketing cover if for each  $x \in [0, 1]^d$  there is a pair  $(v_K^x, w_K^x) \in \Delta$  such that  $v_K^x \leq x \leq w_K^x$  and  $\lambda(\overline{[v_K^x, w_K^x]}) \leq \delta$ .

*Proof of Theorem 7.* We start by defining a sequence of  $\delta$ -covers of increasing precision. Let  $\delta_0 := \frac{d}{m}$  and  $\Gamma_0 := \{0, \frac{1}{m}, \frac{2}{m}, \dots, 1\}^d$ . Then  $\Gamma_0$  is a  $\delta_0$ -cover. Let  $K := \lfloor \frac{d-1}{2} \log_2 m \rfloor$ . For  $i = 1, \dots, K-1$ , let  $\delta_i := 2^{-i} \frac{d}{m}$  and let  $\Gamma_i$  be a  $\delta_i$ -cover with  $|\Gamma_i| \leq (4e/\delta_i)^d = (4e2^i \frac{m}{d})^d =: \gamma_i$ . Finally, let  $\delta_K := 2^{-K} \frac{d}{m}$  and let  $\Delta_K$  be a  $\delta_K$ -bracketing cover with  $|\Delta_K| \leq (4e/\delta_K)^d =: \gamma_K$ . Such covers exist by Theorem 1.15 of [Gne08], see also Lemma 1 of [Ais11].

By definition of bracketing covers, for each  $x \in [0, 1]^d$  there is a pair  $(v_K^x, w_K^x) \in \Delta_K$  such that  $v_K^x \leq x \leq w_K^x$  and  $\lambda(\overline{[v_K^x, w_K^x]}) \leq \delta_K$ . By elementary properties of the discrepancy function, we have

$$|\text{disc}([0, x])| \leq \max \{ |\text{disc}([0, v_K^x])|, |\text{disc}([0, w_K^x])| \} + N\delta_K.$$

Consequently,

$$D^*(P) \leq N\delta_K + \max \{ |\text{disc}([0, v_K^x])|, |\text{disc}([0, w_K^x])| \mid x \in [0, 1]^d \}. \quad (6)$$

Note that  $N\delta_K \leq \frac{\sqrt{dN}}{\sqrt{m/d}}$  is of asymptotic order not larger than the bound we aim at. Consequently, it suffices in the following to analyze

$$\max \{ |\text{disc}([0, v_K^x])|, |\text{disc}([0, w_K^x])| \mid x \in [0, 1]^d \}.$$

To this aim, note that for each  $i \in [0..K-1]$  and each  $x \in [0, 1]^d$  there is a  $v_i(x) \in \Gamma_i \cup \{0\}$  such that  $v_i(x) \leq x$  and  $\lambda(\overline{[v_i(x), x]}) \leq \delta_i$ .

For each  $x \in [0, 1]^d$  we define  $p_{K+1}^x := w_K^x$ ,  $p_K^x := v_K^x$ , and recursively for  $i = K-1, \dots, 0$ , we define  $p_i^x := v_i(p_{i+1}^x)$ . By construction, the sets  $B_i^x := \overline{[p_{i-1}^x, p_i^x]}$ ,  $i = 1, \dots, K+1$ , are disjoint. By the additivity of  $\text{disc}(\cdot)$  and thus the subadditivity of  $|\text{disc}(\cdot)|$ , we obtain

$$|\text{disc}([0, v_K^x])| \leq \sum_{i=1}^K |\text{disc}(B_i^x)|, \quad (7)$$

$$|\text{disc}([0, w_K^x])| \leq \sum_{i=1}^{K+1} |\text{disc}(B_i^x)|. \quad (8)$$

For this reason, we now proceed by analyzing the discrepancies of the sets  $B_i^x$ . To this aim, keep in mind that (i) we have  $\lambda(B_i^x) \leq \delta_{i-1}$  for all  $i$  and  $x$ , and that

(ii) for fixed  $i$ , the number of different  $B_i^x \neq \emptyset$  is at most  $\gamma_i = (4e2^i \frac{m}{d})^d$ , for  $i \in [K]$ , and  $\gamma_{K+1} := \gamma_K$  when  $i = K + 1$ .

Let us first regard an arbitrary measurable set  $S \subseteq [0, 1)^d$ . Let  $\mathcal{Q}$  be the set of all elementary cubes  $\prod_{j=1}^d [\frac{q_j-1}{m}, \frac{q_j}{m})$ ,  $q_1, \dots, q_d \in [m]$ . For each  $Q \in \mathcal{Q}$ , we have

$$\mathbb{E}[|P \cap S \cap Q|] = \Pr[P \cap S \cap Q \neq \emptyset] = \frac{\lambda(S \cap Q)}{\lambda(Q)} = N\lambda(S \cap Q).$$

Consequently, we may write

$$\begin{aligned} \text{disc}(S) &= |P \cap S| - N\lambda(S) \\ &= \sum_{Q \in \mathcal{Q}} (|P \cap S \cap Q| - N\lambda(S \cap Q)) \\ &= \sum_{Q \in \mathcal{Q}} (|P \cap S \cap Q| - \mathbb{E}[|P \cap S \cap Q|]) \end{aligned}$$

as a sum of  $N$  independent random variables  $Z_Q := |P \cap S \cap Q| - \mathbb{E}[|P \cap S \cap Q|]$ , each taking values in the interval  $[-1, 1]$ , each having expectation  $\mathbb{E}[Z_Q] = 0$ , and each having variance  $\text{Var}[Z_Q] = \text{Var}[|P \cap S \cap Q|] \leq \mathbb{E}[|P \cap S \cap Q|] = N\lambda(S \cap Q)$ . Consequently, Bernstein's inequality gives

$$\Pr[|\text{disc}(S)| \geq t] \leq 2 \exp\left(-\frac{t^2}{2 \sum_Q \text{Var}[Z_Q] + 2t/3}\right) \leq 2 \exp\left(-\frac{t^2}{2N\lambda(S) + 2t/3}\right).$$

For all  $i \in [K + 1]$  and  $\ell \in \mathbb{R}_{\geq 1}$  let  $t_{i,\ell} := 2C\ell d \sqrt{\frac{N \ln(2^{2i} 4e(m/d))}{m 2^{i-1}}}$  for some  $C \geq 1$ . Let  $x \in [0, 1)^d$  and  $i \in [K + 1]$ . By the above, we have

$$\begin{aligned} &\Pr[|\text{disc}(B_i^x)| \geq t_{i,\ell}] \\ &\leq 2 \exp\left(-\frac{t_{i,\ell}^2}{2N\delta_{i-1} + 2t_{i,\ell}/3}\right) \\ &\leq 2 \exp\left(-\frac{t_{i,\ell}^2}{2 \max\{2N\delta_{i-1}, 2t_{i,\ell}/3\}}\right) \\ &\leq 2 \max\left\{\exp\left(-\frac{t_{i,\ell}^2}{4N\delta_{i-1}}\right), \exp\left(-\frac{t_{i,\ell}^2}{4t_{i,\ell}/3}\right)\right\} \\ &\leq 2\left(\exp\left(-\frac{t_{i,\ell}^2}{4N\delta_{i-1}}\right) + \exp\left(-\frac{t_{i,\ell}^2}{4t_{i,\ell}/3}\right)\right) \\ &= 2 \exp\left(-\frac{4 \frac{C^2 d^2 \ell^2 N}{m} 2^{-i+1} \ln(2^{2i} 4e(m/d))}{4N 2^{-i+1} (d/m)}\right) \\ &\quad + 2 \exp\left(-\frac{3}{2} C \ell d \sqrt{\frac{N \ln(2^{2i} 4e(m/d))}{m 2^{i-1}}}\right) \end{aligned}$$

$$\begin{aligned}
&\leq 2 \exp(-d\ell^2 \ln(2^{2i} 4e(m/d))) + 2 \exp\left(-\frac{3}{2}C \sqrt{\frac{N \ln(2^{2i})}{m m^{(d-1)/2}}}\right)^{d\ell} \\
&= 2(2^{2i} 4e(m/d))^{-d\ell^2} + 2 \exp\left(\frac{3}{12}C m^{(d-1)/4} \sqrt{2i \ln(2)}\right)^{-6d\ell} \\
&\leq 2(2^{2i} 4e(m/d))^{-d\ell^2} + 2\left(\frac{1}{4}C m^{(d-1)/4} \sqrt{2i \ln(2)}\right)^{-6d\ell} =: q_{i,\ell}.
\end{aligned}$$

With a simple union bound, choosing  $C = 8^{1/12}(e/4)^{1/6}4/\sqrt{\ln(2)} = 2^{23/12}e^{1/6}/\sqrt{\ln(2)} \leq 5.3573$ , and recalling that  $d \geq 2$  and  $\ell \geq 1$ , we estimate

$$\begin{aligned}
\Pr[\exists x \in [0, 1]^d : |\text{disc}(B_i^x)| \geq t_{i,\ell}] &\leq \gamma_i q_{i,\ell} \\
&\leq (4e2^i \frac{m}{d})^d \cdot 2(2^{2i} 4e \frac{m}{d})^{-d\ell^2} \\
&\quad + (4em^{(d-1)/2} \frac{m}{d})^d \cdot 2\left(\frac{1}{4}C m^{(d-1)/4} \sqrt{2i \ln(2)}\right)^{-6d\ell} \\
&\leq 2 \cdot 2^{-id\ell^2} + 2\left(m^{-d+2} i^{-3} \frac{4e4^6}{2C^6 2^3 \ln(2)^3}\right)^{d\ell} \\
&\leq 2 \cdot 2^{-id\ell^2} + 2(\sqrt{1/8} m^{-d+2} i^{-3})^{d\ell} \\
&\leq 2 \cdot 4^{-i\ell^2} + 2 \cdot 8^{-\ell} m^{-d+2} i^{-6} \leq 2(4^{-i\ell} + 8^{-\ell} i^{-6}).
\end{aligned}$$

Another union bound together with  $\ell \geq 1$  and  $\zeta(6) = \pi^6/945 \leq 1.0174$  gives

$$\begin{aligned}
\Pr[\exists i \in [K+1] \exists x \in [0, 1]^d : |\text{disc}(B_i^x)| \geq t_{i,\ell}] \\
&\leq \sum_{i=1}^{K+1} 2(4^{-i\ell} + 8^{-\ell} i^{-6}) \\
&\leq 2 \cdot 4^{-\ell} \sum_{i=0}^{\infty} 4^{-i} + 2 \cdot 8^{-\ell} \sum_{i=1}^{\infty} i^{-6} \\
&= 4^{-\ell} \frac{8}{3} + 2 \cdot 8^{-\ell} \zeta(6) \leq 4^{-\ell} \left(\frac{8}{3} + \zeta(6)\right) < 3.7 \cdot 4^{-\ell}.
\end{aligned}$$

We compute

$$\begin{aligned}
\sum_{i=1}^{K+1} t_{i,\ell} &\leq \sum_{i=1}^{\infty} 2C\ell d \sqrt{\frac{N \ln(2^{2i} 4e(m/d))}{m 2^{i-1}}} \\
&\leq 2C\ell d \sqrt{\frac{N}{m}} \left( \sum_{i=1}^{\infty} \sqrt{\frac{2i \ln(2)}{2^{i-1}}} + \sum_{i=1}^{\infty} \sqrt{\frac{\ln(4em/d)}{2^{i-1}}} \right) \\
&\leq 2C\ell d \sqrt{\frac{N}{m}} \left( 5(1 + \sqrt{2}/2) \sqrt{2 \ln 2} + \sqrt{\ln(4em/d)} \frac{\sqrt{2}}{\sqrt{2}-1} \right) =: \ell D
\end{aligned}$$

using the estimate  $\sum_{i=1}^{\infty} \sqrt{i/2^{i-1}} \leq 1 + \sum_{i=2}^{\infty} i/2^{i/2} = 1 + \sum_{i=1}^{\infty} (2i)/2^i + \sum_{i=1}^{\infty} (2i+1)/2^{i+1/2} = 1 + 2 \sum_{i=1}^{\infty} i/2^i + \sqrt{2} \sum_{i=1}^{\infty} i/2^i + 2^{-1/2} \sum_{i=1}^{\infty} 2^{-i} = 1 + 4 + 2\sqrt{2} + 2^{-1/2} = 5(1 + \sqrt{2}/2)$ .

Consequently, by (6), (7), and (8), we have  $\Pr[D^*(P) \geq \ell D + N\delta_K] \leq 3.7 \cdot 4^{-\ell}$  for all  $\ell \geq 1$ . From this we easily derive the following bound on the expectation of  $D^*(P)$ .

$$\begin{aligned}
\mathbb{E}D^*(P) &= \int_0^\infty \Pr[D^*(P) \geq x] dx \\
&\leq D + N\delta_K + D \int_1^\infty \Pr[D^*(P) \geq \ell D + N\delta_K] d\ell \\
&\leq D(1 + 3.7 \cdot 4^{-1}/\ln(4)) + N\delta_K \leq 1.6673D + N\delta_K \\
&\leq 1.6673 \cdot 2Cd \sqrt{\frac{N}{m}} \left( 5(1 + \sqrt{2}/2)\sqrt{2\ln 2} + \frac{\sqrt{2\ln(4em/d)}}{\sqrt{2}-1} \right) + \sqrt{\frac{dN}{m/d}} \\
&\leq 17.8645 \sqrt{\frac{dN}{m/d}} (10.0499 + 3.4143\sqrt{\ln(4em/d)} + 1/17.8645) \\
&\leq 60.9948 \sqrt{\frac{dN}{m/d}} (\sqrt{\ln(4em/d)} + 2.9599). \quad \square
\end{aligned}$$

## 5 Small Numbers of Points

The results proven in the previous sections indicate that jittered sampling will not lead to discrepancies of asymptotic order smaller than those of independent random points when  $m = O(d)$ . We shall not prove this here, since an adaptation of our lower bound proof above to small numbers of points or an adaptation of the proof in [Doe14] to jittered sampling both appear not straight-forward, and since at the same time it is very hard to imagine that jittered sampling has an asymptotic advantage in some regime  $m = o(d)$  given that we have disproven such an advantage for  $m = d$ .

With no asymptotic advantage expected for  $m = O(d)$ , one could think that jittered sampling is interesting only for numbers  $N$  of points that are super-exponential in  $d$ , namely at least of the order  $\omega(d)^d$ . We now argue that this might be a too pessimistic point of view, simply because possible constant factor improvements or even advantages in lower-order terms can be interesting as well. We note that Pausinger and Steinerberger [PS16] conducted also a small experimental analysis and observed, e.g., that the (empirical) expected discrepancy of  $N = 3^5$  points in dimension  $d = 5$  (that is, we have  $m = 3$  in the jittered sampling case) is 0.1200 for independent random points and 0.1046 for jittered sampling. Such advantages may not appear large, but given that they come for free (generating a jittered sampling is not more costly than generating an independent sample), they are interesting. In addition, it is known that such point sets satisfy the known

discrepancy guarantees for independent points sets (we discuss this in more detail below), so also in terms of proven quality guarantees switching from independent random points to jittered sampling gives no disadvantages.

Since we need  $m \geq 2$  in our definition of jittered sampling, we necessarily have a number of points that is at least exponential in  $d$ . We argue now that also for smaller numbers of points some forms of jittering can be defined which promise to be at least as good as independent points, with lower-order advantages likely to appear. Assume that  $N$  is a power of two, but less than  $2^d$ , the smallest number for which our jittered sampling was defined. Hence  $N = 2^{d'}$  for some  $d' < d$ . We define a variant of jittered sampling of  $N$  points as follow. Consider the family  $B_x = \prod_{i=1}^{d'} [x_i, x_i + \frac{1}{2}) \times [0, 1)^{d-d'}$ ,  $x \in \{0, \frac{1}{2}\}^{d'}$  of boxes and define the point set  $P$  by choosing from each such box independently a point uniformly distributed in this box. This set of points is again easy to construct and is also known to satisfy the known discrepancy guarantees of independent random points.

## 5.1 Discrepancy Guarantees from the Result of Gnewuch and Hebbinghaus

We now give the promised details on why the strongest currently known discrepancy guarantee for a broad class of random point sets by Gnewuch and Hebbinghaus [GH21] (which imply those of Aistleitner [Ais11] and Aistleitner and Hofer [AH14]) also applies to the jittered samplings proposed above. All this can be found in [WGH20]. We start by describing the discrepancy result of [GH21].

Let  $\mathcal{D}_0 := \{[0, b) \setminus [0, a) \mid a, b \in [0, 1)^d, a \leq b\}$  and  $\gamma \geq 1$ . A collection  $X_1, \dots, X_N$  of random points is called  $\gamma$ -negatively dependent (with respect to  $\mathcal{D}_0$ ) if for all  $D \in \mathcal{D}_0$  and all  $J \subseteq [N]$  we have

$$\Pr[\forall j \in J : X_j \in D] \leq \gamma \prod_{j \in J} \Pr[X_j \in D],$$

$$\Pr[\forall j \in J : X_j \notin D] \leq \gamma \prod_{j \in J} \Pr[X_j \notin D].$$

Gnewuch and Hebbinghaus [GH21, Theorem 4.4] have shown that for all  $c > 0$  and  $\rho \geq 0$  the discrepancy of an  $e^{\rho d}$ -negatively dependent collection of  $N$  points in  $[0, 1)^d$ , each of which is uniformly distributed in  $[0, 1)^d$ , is at most  $c\sqrt{dN}$  with probability at least  $1 - e^{-(1.6741c^2 - 10.7042 - \rho)d}$ . Since independent uniformly distributed random points are 1-negatively dependent by definition, this reproves the known  $O(\sqrt{dN})$  upper bound for the expected discrepancy of uniformly distributed random point sets, however with better absolute bounds such as the result that with positive probability the discrepancy is at most  $2.5278\sqrt{dN}$ .

This result cannot be immediately applied to jittered sampling since there the individual points are not uniformly distributed in  $[0, 1]^d$ . The solution proposed in [WGH20] is to regard the symmetrization of the jittered sampling set. Let  $B_1, \dots, B_N$  be a partition of  $[0, 1]^d$  into sets of equal Lebesgue measure  $1/N$ . Let  $X_1, \dots, X_N$  be a jittered sampling point set with respect to these, that is, for each  $j \in [N]$  the point  $X_j$  is chosen uniformly from  $B_j$ . Let  $\sigma : [N] \rightarrow [N]$  be a permutation chosen uniformly at random. Let  $X'_i = X_{\sigma(i)}$  for all  $i \in [N]$ . Then each  $X'_i$  is uniformly distributed in  $[0, 1]^d$ . Unfortunately, the  $X'_i$  are not independent anymore (if  $X'_1 \in B_1$ , for example, then  $X'_2$  cannot be in  $B_1$ ; hence  $\Pr[X'_2 \in B_1 \mid X'_1 \in B_1] = 0 \neq \frac{1}{N} = \Pr[X'_2 \in B_1]$ ). They are, however, 1-negatively dependent as shown in [WGH20, Theorem 4.4]. Hence the discrepancy estimate of Gnewuch and Hebbinghaus applies to the  $X'_i$  and gives the same guarantees as for independent points. Since, viewed as multiset, the  $X_i$  and the  $X'_i$  are identically distributed, this discrepancy also holds for the jittered sampling  $(X_i)$ .

## 5.2 Extending Aistleitner’s Upper Bound to Jittered Sampling

Since the approach described in the previous subsection is slightly technical (we lose the independence of the points due to the symmetrization and then cope with this by establishing a negative dependence property and using a discrepancy guarantee for negatively dependent point sets), we now show that the earlier (and weaker in terms of constants) discrepancy guarantees of Aistleitner [Ais11] and Aistleitner and Hofer [AH14] quite directly apply also to jittered samplings. Since this observation might be useful also for extending future discrepancy estimates for independent uniformly distributed points to jittered sampling, we give some more details.

An easy inspection of Aistleitner’s [Ais11] proof reveals that the only way this upper bound analysis of the discrepancy of a uniformly distributed random point set relies on the randomness of the points is by applying Chernoff-type concentration bounds to the expected number of points in a measurable subset  $I$  of  $[0, 1]^d$ . If  $X_1, \dots, X_N$  are independently and uniformly distributed in  $[0, 1]^d$ , then the indicator random variables  $\mathbb{1}_I(X_j)$  for the events “ $X_j \in I$ ”,  $j \in [N]$ , are independent binary random variables with success probability  $p_{I,j} = \lambda(I)$ . Letting, for all  $j \in [N]$ ,  $Z_{I,j} = \mathbb{1}_I(X_j) - \mathbb{E}\mathbb{1}_I(X_j) = \mathbb{1}_I(X_j) - p_{I,j} = \mathbb{1}_I(X_j) - \lambda(I)$  denote the centralization of  $\mathbb{1}_I(X_j)$ , we have  $\text{disc}_P(I) = |P \cap I| - N\lambda(I) = \sum_{j=1}^N Z_{I,j} =: Z_I$ , that is, the discrepancy of  $I$  can be written as the sum of the independent random variables  $Z_{I,j}$  (having expectation zero and taking values in an interval of length 1)

and can thus be estimated via strong concentration bounds. Here Aistleitner (see the lower half of page 536 of [Ais11]) only uses the two estimates

$$\Pr[|Z_I| \geq t] \leq 2 \exp\left(-\frac{t^2}{2 \operatorname{Var}[Z_I] + \frac{2}{3}t}\right) = 2 \exp\left(-\frac{t^2}{2N\lambda(I)(1 - \lambda(I)) + \frac{2}{3}t}\right),$$

$$\Pr[|Z_I| \geq t] \leq 2 \exp\left(-\frac{2t^2}{N}\right).$$

We now argue that the same estimates hold for the corresponding expressions for jittered sampling, which shows that Aistleitner's proof also applies to jittered sampling. Since it takes no additional effort, we show this result for a more general type of random point sets. Let  $\mu_1, \dots, \mu_N$  be probability measures on  $[0, 1]^d$  (equipped with the Lebesgue  $\sigma$ -algebra) such that for each  $\lambda$ -measurable set  $I \subseteq [0, 1]^d$  we have  $\sum_{j=1}^N \mu_j(I) = N\lambda(I)$ . Let  $P'$  be a random point set obtained by sampling, for each  $j \in [N]$  independently, a point  $X'_j$  according to  $\mu_j$ . We observe that this construction includes both independent uniformly distributed points and jittered sampling (but not the generalized jittered samplings of [WGH20]).

We now show that such a point set satisfies the discrepancy guarantees for independent uniformly distributed points given by Aistleitner [Ais11]. Let  $I$  be a measurable subset of  $[0, 1]^d$ . By construction, the indicator random variables  $\mathbb{1}_I(X'_j)$ ,  $j \in [N]$ , are again independent, but now  $p'_{I,j} = \mu_j(I)$ . Let  $Z'_{I,j} := \mathbb{1}_I(X'_j) - \mathbb{E}\mathbb{1}_I(X'_j) = \mathbb{1}_I(X'_j) - p'_{I,j} = \mathbb{1}_I(X'_j) - \mu_j(I)$ . Then again  $\operatorname{disc}_{P'}(I) = |P' \cap I| - N\lambda(I) = \sum_{j=1}^N Z'_{I,j} =: Z'_I$ . Since  $Z'_I$  is a sum of independent random variables (each with expectation zero and taking values in an interval of length 1), we again have

$$\Pr[|Z'_I| \geq t] \leq 2 \exp\left(-\frac{t^2}{2 \operatorname{Var}[Z'_I] + \frac{2}{3}t}\right) \leq 2 \exp\left(-\frac{t^2}{2N\lambda(I)(1 - \lambda(I)) + \frac{2}{3}t}\right)$$

$$\Pr[|Z'_I| \geq t] \leq 2 \exp\left(-\frac{2t^2}{N}\right).$$

For the second inequality in the first line, note that  $\operatorname{Var}[Z'_I] = \sum_{j=1}^N p'_{I,j}(1 - p'_{I,j}) = \sum_{j=1}^N (p'_{I,j} - (p'_{I,j})^2)$ . We have  $\sum_{j=1}^N p'_{I,j} = \sum_{j=1}^N \mu_j(I) = N\lambda(I)$  by construction. Knowing that a sum of squares of  $n$  variables with fixed sum  $S$  is minimal for all variables being equal to  $S/n$ , that is,  $\min\{x_1^2 + \dots + x_n^2 \mid x_1, \dots, x_n \in \mathbb{R}, x_1 + \dots + x_n = S\} = n \cdot (S/n)^2$ , we obtain  $\operatorname{Var}[Z'_I] \leq N\lambda(I) - N\lambda(I)^2 = N\lambda(I)(1 - \lambda(I))$ . Consequently, we have the same concentration bounds for the discrepancy of the set  $I$  as when using independent sample points. Since Aistleitner's proof refers to the sample set only via these two estimates, his proof applies equally well to our generalized point set construction.



The same argument immediately also extends the upper bounds of Aistleitner and Hofer [AH14] to our point sets. We are optimistic that this argument can be used for future discrepancy estimates shown roughly along these lines as well. We note that [GH21] also uses only the two concentration bounds discussed above (except that it is argued why these also extend to  $\gamma$ -negatively dependent random variables). Hence the only reason why the result of [GH21] does not immediately extend to the random samplings defined above is the requirement that each point is uniformly distributed in  $[0, 1)^d$ .

## 6 Conclusion

In this work, we determined (apart from constant factors) the expected star discrepancy of a jittered sampling point set having  $m^d$  points,  $m \geq d$ . This improves both the previous upper and lower bound and also removes the unquantified restriction that “ $m$  is sufficiently large compared to  $d$ ”. These improvements are made possible by exhibiting suitable combinatorial structures in the problem which then allow an easier analysis via discrete probability theory. We feel that this might, also for other discrepancy problems on random point sets, be an approach more suitable than trying to describe the discrepancy directly via continuous-domain random variables.

From our main result and the known result that the known discrepancy guarantees for independent random points extend to jittered sampling (in particular for  $m < d$ , where we have given no guarantee), we see no reason to not generally recommend to use jittered sampling to obtain low-discrepancy point sets rather than independent random points

## Acknowledgments

This work was triggered by the 2nd International Workshop on Discrepancy Theory and Applications organized by Dmitriy Bilyk, Luca Brandolini, William Chen, Anand Srivastav, Giancarlo Travaglini. The author would like to thank the organizers for inviting him to this fruitful meeting. This work also profited from several fruitful discussions with Michael Gnewuch and Marcin Wnuk, which are gratefully acknowledged.

## References

- [AH14] Christoph Aistleitner and Markus Hofer. Probabilistic discrepancy bound for Monte Carlo point sets. *Mathematics of Computation*, 83:1373–1381, 2014.

- [Ais11] Christoph Aistleitner. Covering numbers, dyadic chaining and discrepancy. *Journal of Complexity*, 27:531–540, 2011.
- [ASYM16] Haim Avron, Vikas Sindhwani, Jiyan Yang, and Michael W. Mahoney. Quasi-Monte Carlo feature maps for shift-invariant kernels. *Journal of Machine Learning Research*, 17:120:1–120:38, 2016.
- [Bec87] József Beck. Irregularities of distribution. I. *Acta Mathematica*, 159:1–49, 1987.
- [Bel81] David R. Bellhouse. Area estimation by point-counting techniques. *Biometrics*, 37:303–312, 1981.
- [BL13] Dmitriy Bilyk and Michael Lacey. The supremum norm of the discrepancy function: recent results and connections. In *Monte Carlo and Quasi-Monte Carlo Methods 2012*, pages 23–38. Springer, 2013.
- [Cha00] Bernard Chazelle. *The Discrepancy Method*. Cambridge University Press, 2000.
- [CPC84] Robert L. Cook, Thomas K. Porter, and Loren C. Carpenter. Distributed ray tracing. In *Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1984*, pages 137–145. ACM, 1984.
- [DDG18] Benjamin Doerr, Carola Doerr, and Michael Gnewuch. Probabilistic lower bounds for the discrepancy of Latin hypercube samples. In Josef Dick, Frances Y. Kuo, and Henryk Woźniakowski, editors, *Contemporary Computational Mathematics - A Celebration of the 80th Birthday of Ian Sloan*, pages 339–350. Springer, 2018.
- [DGS04] Benjamin Doerr, Michael Gnewuch, and Anand Srivastav. Bounds and construction for the star discrepancy via  $\delta$ -covers. *Berichtsreihe des Mathematischen Seminars der Christian-Albrechts-Universität zu Kiel*, Report 04-13, 2004.
- [DGS05] Benjamin Doerr, Michael Gnewuch, and Anand Srivastav. Bounds and constructions for the star-discrepancy via  $\delta$ -covers. *Journal of Complexity*, 21:691–709, 2005.
- [Doe14] Benjamin Doerr. A lower bound for the discrepancy of a random point set. *Journal of Complexity*, 30:16–20, 2014.
- [DP10] Josef Dick and Friedrich Pillichshammer. *Digital Nets and Sequences*. Cambridge University Press, 2010.

- [DT97] Michael Drmota and Robert F. Tichy. *Sequences, Discrepancies and Applications*, volume 1651 of *Lecture Notes in Mathematics*. Springer, 1997.
- [Fel71] William Feller. *Introduction to Probability Theory and Its Applications*, volume 2. Wiley, 2nd edition, 1971.
- [FW93] Kai-Tai Fang and Yuan Wang. *Number-Theoretic Methods in Statistics*. CRC Press, 1993.
- [GH21] Michael Gnewuch and Nils Hebbinghaus. Discrepancy bounds for a class of negatively dependent random points including Latin hypercube samples. *Annals of Applied Probability*, 2021. To appear.
- [Gne08] Michael Gnewuch. Bracketing numbers for axis-parallel boxes and applications to geometric discrepancy. *Journal of Complexity*, 24:154–172, 2008.
- [Gne12] Michael Gnewuch. Entropy, randomization, derandomization, and discrepancy. In Leszek Plaskota and Henryk Woźniakowski, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2010*, volume 23 of *Proceedings in Mathematics and Statistics*, pages 43–78. Springer, 2012.
- [Hla61] Edmund Hlawka. Funktionen von beschränkter Variation in der Theorie der Gleichverteilung. *Annali di Matematica Pura ed Applicata*, 54:325–333, 1961.
- [HNWW01] Stefan Heinrich, Erich Novak, Grzegorz W. Wasilkowski, and Henryk Woźniakowski. The inverse of the star-discrepancy depends linearly on the dimension. *Acta Arithmetica*, 96:279–302, 2001.
- [KM05] Shuhei Kimura and Koki Matsumura. Genetic algorithms using low-discrepancy sequences. In *Genetic and Evolutionary Computation Conference, GECCO 2005*, pages 1341–1346. ACM, 2005.
- [Kok43] Jurjen F. Koksma. A general theorem from the uniform distribution modulo 1 (in Dutch). *Mathematica B (Zutphen)*, 1:7–11, 1942/43.
- [Mat99] Jiří Matoušek. *Geometric Discrepancy*. Springer-Verlag, Berlin, 1999.
- [MBC79] Michael D. McKay, Richard J. Beckman, and William J. Conover. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21:239–245, 1979.

- [Nie92] Harald Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*, volume 63 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.
- [OT05] Art B. Owen and Seth D. Tribble. A quasi-Monte Carlo Metropolis algorithm. *Proceedings of the National Academy of Sciences*, 102:8844–8849, 2005.
- [Owe03] Art B. Owen. Quasi-monte carlo sampling. In *Monte Carlo Ray Tracing: ACM SIGGRAPH 2003 Course Notes 44*, pages 69–88. ACM, 2003.
- [PS16] Florian Pausinger and Stefan Steinerberger. On the discrepancy of jittered sampling. *Journal of Complexity*, 33:199–216, 2016.
- [Rob55] Herbert Robbins. A remark on Stirling’s formula. *The American Mathematical Monthly*, 62:26–29, 1955.
- [TG07] Olivier Teytaud and Sylvain Gelly. DCMA: yet another derandomization in covariance-matrix-adaptation. In *Genetic and Evolutionary Computation Conference, GECCO 2007*, pages 955–963. ACM, 2007.
- [WGH20] Marcin Wnuk, Michael Gnewuch, and Nils Hebbinghaus. On negatively dependent sampling schemes, variance reduction, and probabilistic upper discrepancy bounds. In Dmitriy Bilyk, Josef Dick, and Friedrich Pillichshammer, editors, *Discrepancy Theory*, volume 26 of *Radon Series on Computational and Applied Mathematics*, pages 43–67. De Gruyter, 2020.