The power of online thinning in reducing discrepancy

Raaz Dwivedi* Ohad N. Feldheim† Ori Gurel-Gurevich‡ Aaditya Ramdas§

September 13, 2017

Abstract

Consider an infinite sequence of independent, uniformly chosen points from $[0,1]^d$. After looking at each point in the sequence, an overseer is allowed to either keep it or reject it, and this choice may depend on the locations of all previously kept points. However, the overseer must keep at least one of every two consecutive points. We call a sequence generated in this fashion a *two-thinning* sequence. Here, the purpose of the overseer is to control the discrepancy of the empirical distribution of points, that is, after selecting n points, to reduce the maximal deviation of the number of points inside any axis-parallel hyper-rectangle of volume A from nA. Our main result is an explicit low complexity two-thinning strategy which guarantees discrepancy of $O(\log^{2d+1} n)$ for all n with high probability (compare with $\Theta(\sqrt{n \log\log n})$ without thinning). The case $d = 1$ of this result answers a question of Benjamini.

We also extend the construction to achieve the same asymptotic bound for $(1+\beta)$-thinning, a set-up in which rejecting is only allowed with probability β independently for each point. In addition, we suggest an improved and simplified strategy which we conjecture to guarantee discrepancy of $O(\log^{d+1} n)$ (compare with $\theta(\log^d n)$, the best known construction of a low discrepancy sequence). Finally, we provide theoretical and empirical evidence for our conjecture, and provide simulations supporting the viability of our construction for applications.

Keywords. Two-choices, thinning, discrepancy, subsampling, online, Haar.

1 Introduction

Let $(\Omega, \mathcal{F}, \mu)$ be a probability space and let $\mathcal{R}$ be a class of subsets of Ω. The $\mathcal{R}$-*discrepancy* of S, a subset of Ω of size n, with respect to μ is defined as

$$\mathrm{Dis}_{\mathcal{R}}(S) := \sup_{R \in \mathcal{R}} \Big| |S \cap R| - n\mu(R) \Big|.$$

Let $X = \{X_n\}_{n\in\mathbb{N}}$ be a sequence of elements in Ω, and write $X^n = \{X_i\}_{i\in[n]}$. The discrepancy of X is defined as the sequence of discrepancies $\{\mathrm{Dis}_{\mathcal{R}}(X^n)\}_{n\in\mathbb{N}}$.

Throughout we consider only with $\mathcal{R}$-discrepancy with respect to Lebesgue measure on $\Omega = [0,1]^d$, where $\mathcal{R} = \{\bigotimes_{i=1}^d [a_i, b_i) \subseteq [0,1)^d : 0 \leqslant a_i < b_i \leqslant 1\}$ are the axis aligned hyper-rectangles. For brevity we call this simply *discrepancy*, and denote $\mathrm{Dis}(X^n) = \mathrm{Dis}_{\mathcal{R}}(X^n)$.

*University of California, Berkeley, email:raaz.rsk@berkeley.edu.

†Hebrew University of Jerusalem 1613091, email:Ohad.Feldheim@mail.huji.ac.il, research conducted in Stanford University and supported in part by NSF grant DMS 1613091.

‡Hebrew University of Jerusalem, email:Ori.Gurel-Gurevich@mail.huji.ac.il, research was supported by the Israel Science Foundation (grant No. 1707/16).

§University of California, Berkeley, email:aramdas@berkeley.edu.

The best known upper bound for the discrepancy of X is $\mathrm{Dis}(X^n) = O(\log^d n)$ and several lattice related constructions are known (see, e.g. [6]). However, in many applications only restricted control over the locations of the points X_n is available so that an optimal discrepancy sequence cannot be used. The most extreme case is the *Monte-Carlo* setting, where the points are independent samples of the uniform distribution over $[0,1)^d$. In this case classical results in probability theory imply that $\mathrm{Dis}(X^n) = O(\sqrt{n \log\log n})$ and that this estimate is tight. Due to the significant gap between the optimal discrepancy obtainable by an infinite sequence and the discrepancy of a sequence of independent samples it has been desirable to look for variations on the Monte-Carlo setting which obtain lower discrepancy by allowing an overseer mild control over the sequence X. The most well known result in this line of investigation is the "power of two-choices" paper, by Azar-Broder-Karlin-Upfal [1], who show that in the setting of $\Omega = [N]$, uniform μ and $\mathcal{R} = \{\{n\} \ : \ n \in [N]\}$, by allowing the overseer to choose X_n among two i.i.d. μ-distributed samples it is possible to obtain an exponential improvement in the discrepancy.

In this work we investigate a related, weaker sense of control. Consider an infinite sequence $U_1^\infty := \{U_n\}_{n\in\mathbb{N}}$ of i.i.d. uniform random variables on $[0,1)^d$. These points are shown to an overseer one by one, who may depend on his past choices in deciding whether to keep each point or reject it. However his control is restricted by the constraint of keeping at least one of every two consecutive points. We call a strategy executed by the overseer in producing such a sequence a *two-thinning strategy*. We also consider an even weaker setting, in which, in addition to the restriction of a two-thinning, each point has independent probability β to be rejectable and otherwise it must be kept. Inspired by the work of Peres-Talwar-Wieder [23] on $(1+\beta)$-choice, we call this setting $(1+\beta)$*-thinning*. More precise definitions of the above terminology are provided in Section 3.

Our main result is an explicit $(1+\beta)$-thinning strategy on $\mathrm{Unif}[0,1)^d$, which we call *Haar* strategy which satisfies the following.

Theorem 1. *The Haar $(1+\beta)$-thinning strategy yields a sequence Z which almost surely satisfies*

$$\limsup_{n\to\infty} \frac{\mathrm{Dis}(Z^n)}{\log^{2d+1}(n)} \leqslant \frac{100(d^2+1)}{\beta} \ .$$

This result is obtained as an immediate consequence of the following more detailed theorem.

Theorem 2. *The Haar $(1+\beta)$-thinning strategy yields a sequence Z which for all $n \in \mathbb{N}$ and $\Delta > 0$ satisfies*

$$\mathbb{P}\Big(\mathrm{Dis}(Z^n) \geqslant \beta^{-1} \log^{2d}(n)(\Delta + 1000 + 100d^2 \log n)\Big) \leqslant \beta^{-2} e^{-\frac{\Delta}{50}}.$$

Moreover, in order to apply this strategy the overseer requires $\mathcal{O}(n \log^d n)$ memory and $\mathcal{O}(n \log^d n)$ computations to produce the first n samples.

In section 6 we suggest a heuristic improvement of our analysis, bringing us to make the following conjecture.

Conjecture 1. *The Haar $(1+\beta)$-thinning strategy yields a sequence Z that almost surely satisfies*

$$\limsup_{n\to\infty} \frac{\mathrm{Dis}(Z^n)}{\log^{3d/2+1}(n)} < \infty.$$

In the same section we also suggest a simplified strategy with the same complexity which we call *greedy-Haar* strategy, which we conjecture to provide an additional improvement over the result above. Namely

Conjecture 2. *The greedy-Haar $(1+\beta)$-thinning strategy yields a sequence Z that almost surely satisfies*

$$\limsup_{n\to\infty} \frac{\mathrm{Dis}(Z^n)}{\log^{d+1}(n)} < \infty.$$

To demonstrate that our constructions are also viable in practice as an alternative for Monte-Carlo i.i.d. sampling we dedicate Section 7 to simulations, comparing the performance of our strategies with classical Monte-Carlo discrepancy. Further discussion on the potential applications of our results in statistics is provided in Section 2.4.

2 Related Work

In this section we briefly survey related work on the power of two-choices and discrepancy theory and present possible applications of our work to numerical integration and statistics.

2.1 Two-choices and weaker forms of choices

The power of two choices is a phenomenon discovered and popularized by Azar, Broder, Karlin and Upfal [1], who consider a setting in which the underlying space is the discrete set $[M] = \{1, \dots, M\}$ and the discrepancy is measured with respect to $\mathcal{R} = \{\{m\} : m \in [M]\}$. Thinking of the points $Z_1, \dots Z_N$ as balls and of their values in $[M]$ as bins, the authors considered a process where at each step a ball is assigned to the least occupied among two bins chosen uniformly and independently. They show that when $N = \Theta(M)$ this yields with high probability a discrepancy of $\mathrm{Dis}_{\mathcal{R}}(Z^N) = (1+o(1))\,(\log\log N/\log 2)$ (compare with $\mathrm{Dis}_{\mathcal{R}}(U^N) = (1+o(1))\,(\log N/\log\log N)$ when U_i are i.i.d. uniform). When $N \gg M$ their results imply that $\mathbb{P}(\mathrm{Dis}_{\mathcal{R}}(Z^N) > \Delta \log M)$ decays exponentially fast in Δ, uniformly in N, so that the discrepancy does not grow with N. In addition, in this model, the load of a typical bin deviates from N/M by merely a constant (compare with a typical deviation of $\Theta(\sqrt{N/M})$ and $\mathrm{Dis}_{\mathcal{R}}(U^N) = \Theta(\sqrt{N \log M/M})$ for U_i i.i.d. uniform). It was later discovered that these results are tight up to a constant in the exponent (see e.g. [23]). Note, however, that significantly better iterated log bounds were obtained by Berenbrink, Czumaj, Steger and Vöcking [5] for the one-sided gap between the load of the most loaded bin and the average load. For a simpler proof see Talwar and Wieder [27].

While considering applications of the power of two choices to queuing theory, Mitzenmacher, in his thesis [20], suggested the following more robust setting of "two-choices with errors". Peres, Talwar and Wieder [23] later formulated this process, defining the equivalent $(1+\beta)$-choice process for $\beta \in [0,1]$. In this process, with probability β (independent of everything else) the overseer is offered two uniformly distributed independent bins and with probability $(1-\beta)$ only one such bin is offered and no choice is allowed. $(1+\beta)$-thinning processes are closely related to $(1+\beta)$-choice processes. In fact, a two-thinning set-up is equivalent to the corresponding two-choices set-up where the overseer is oblivious to the second available bin. Extending this argument, we see that every $(1+\beta)$-thinning processes is a $(1+\beta)$-choice process (i.e., every process that could be realized by a $(1+\beta)$-thinning strategy could also be realized by a $(1+\beta)$-choice strategy). On the other hand, Proposition 3.1 below guarantees that every $(1+\beta)$-choice process for $\beta \leqslant \frac{1}{2}$ is a $(1+2\beta)$-thinning process. As Theorems 1 and 2 are obtained for $(1+\beta)$-thinning processes with arbitrarily small β, they are also valid in the $(1+\beta)$-choice setting.

In the balls and bins setting, both $(1+\beta)$-choice processes and $(1+\beta)$-thinning processes achieve the same asymptotic discrepancy of $\Theta(\log M)$ when $N \gg M$, the same discrepancy that could be achieved by a two-choices process (this follows from results of [23]). On the other hand, if one

measures discrepancy by the one-sided maximal load semi-norm given by

$$\max_{i\in[M]} \#\{n \in [N] \ : \ \xi(n) = i\} - N/M,$$

then Berenbrink, Czumaj, Steger and Vöcking [5] show that two-choices process can, in fact, achieve $\Theta(\log\log M)$, while both $(1+\beta)$-choice process and $(1+\beta)$-thinning processes still achieve only $\Theta(\log M)$ (again by [23]). A similar gap between two choices, $(1+\beta)$-choice and $(1+\beta)$-thinning for $\beta < 1$ exists also in the regime $N \asymp M$, and in this regime both notions obtain no significant improvement over a no-choice setting. Curiously, when $\beta = 1$, the optimal discrepancy obtainable by two-thinning strategy is $\Theta(\sqrt{\log N/\log\log N})$ which is strictly between the discrepancy in the no-choice setting, which is $\Theta(\log N/\log\log N)$ and the optimum in the 2-choice setting which is $\Theta(\log\log N)$. This is shown in a separate recent note by the second and third author [8].

2.2 Interval subdivision processes

The case $\Omega = [0,1]$ of our result relates to a long line of investigation of so called *interval subdivision processes*.

An *interval subdivision process* is a sequence of points $(X_i)_{i=1}^{\infty}$ where $X_i \in [0,1]$. The intervals of the process at the n-th step are the gaps between adjacent points in $(X_i)_{i=1}^{n}$, while the *empirical measure* at that step is defined as $\frac{1}{n}\sum_{i=0}^{n-1}\delta_{X_i}$, where δ_j is the Dirac delta measure. When the points are chosen independently according to the uniform distribution on $[0,1]$ we call this the *uniform interval subdivision process*. By the law of large numbers, the empirical measure of this process converges to the uniform measure almost surely as n tends to infinity.

In 1975 Kakutani [10] suggested a couple of alternative models for interval subdivision which he conjectured to be more regular then the uniform process in the sense that their empirical measures should converge to the uniform distribution more rapidly. In one of these processes, which we refer to here as the *Kakutani process*, the n-th point is selected uniformly on the largest interval (observe that there are no ties almost surely). Kakutani conjectured that the empirical measure of the Kakutani process converges to the uniform measure. This fact was later proved by van Zwet in [28] and independently by Lootgiester in [18]. Once convergence was established it remained to recover in what sense the Kakutani process is more regular than the i.i.d. uniform subdivision.

One natural measure for regularity of the convergence of the empirical measure is the discrepancy of the sequence. A classical result of Kolmogorov and Smirnov (communicated by Donsker [7]), implies that the difference between t and the empirical measure of interval $[0,t]$ of the uniform interval subdivision process, normalized by a factor of $\sqrt{n}$ converges to the standard Brownian bridge. Hence, the discrepancy of the uniform interval subdivision process is of order $\Theta(\sqrt{n})$. However, the the interval variation discrepancy of the Kakutani process was not easy to handle, and in the 1980s other properties of the process have been studied (see [24]). Analysis of the interval variation discrepancy was made possible only in 2004 when Pyke and van Zwet [25] were able to compute the empirical process of the Kakutani process and showed that the difference between t and the empirical measure of the process on the interval $[0,t]$, normalized by a factor of $\sqrt{n}$, converges to a Brownian bridge with half the standard deviation. In particular, this implied that the Kakutani process achieves an improvement of merely a constant factor in the interval variation discrepancy over the uniform interval subdivision process.

Circa 2014, Benjamini (see [19] and [9]) suggested investigating how a two-choice variant of the uniform interval subdivision process behaves. One family of algorithms which Benjamini suggested are *local* algorithms, namely ones in which the player considers only the size of the intervals which

contain the new sampled points. Two natural examples being *max-2* and *furthest-2* whose respective descriptions are "pick the point located in the larger interval" and "pick the point furthest from all previously chosen points".

Following the work of Maillard and Paquette [19] who studied other properties of the *max-2* process, Junge [9] showed that the empirical measure of the *max-2* process indeed converges to the uniform measure. However, both simulations and comparison with Kakutani processes, indicate that max-2 is likely to be at most as regular as the Kakutani process, thus demonstrating discrepancy of $\Theta(\sqrt{n})$. This has been the primary instigator of our present work, where we show that that even in the weaker setting of $(1+\beta)$-thinning, by adopting a *global* strategy the player can obtain a near optimal interval variation discrepancy of $O(\log^3 n)$.

2.3 Discrepancy theory

Discrepancy theory – the study of discrete objects which imitate regularity properties of a continuous counterpart, has, in fact, originated at the study of low discrepancy sequences with respect to the uniform measure on $[0,1)^d$, the very object investigated here. Traditionally, this theory is concerned with deterministic objects, trying to obtain bounds on the lowest discrepancy possible for prefixes $X_1, \ldots, X_n$ of a sequence $X = \{X_i\}_{i \in \mathbb{N}}$ of points in $[0,1)^d$.

In $d \geqslant 2$ the exact optimal asymptotic behavior of the discrepancy is unknown. There exist explicit constructions of sequences X whose discrepancy is $\mathrm{Dis}(X^n) \leqslant C_d \log^d n$ while for every sequence X it is known (by [16]) that there exist infinitely many n-s such that $\mathrm{Dis}(X^n) \geqslant c_d \log^{(d+1+\gamma_d)/2}(n)$, where and $c_d, C_d, \gamma_d \in (0,1)$ are constants depending on dimension.

The upper bound is achieved using *lattice rules* or *digital nets*—for example, Hammersley point sets which are based on the infinite van der Corput sequence (see, e.g., [6]) achieve the upper bound. The lower bounds were obtained by Bilyk, Lacey and Vagharshakyan [16], building upon the work of Roth [15]. We remark that arguments involving Haar wavelets, which play a key role in our construction, are used to prove lower bounds in classical literature (see, e.g., Ch. 3 of the book [17]). While there seem to be no prior work involving Haar wavelets as a tool for obtaining computationally-efficient constructive upper bounds, there exist recent works [2], [3], [4] which use the related Walsh wavelets to control an L_2 notion of discrepancy (weaker than the L_∞ notion we consider).

Discrepancy theory is the main motivation for Conjecture 2, as this conjecture would establish that online thinning typically achieves discrepancy which deviates by merely $\log n$ factor from the minimal discrepancy of any infinite sequence.

2.4 Applications

In this section we list a few potential applications of our results. It is important to notice that our output sequence has the desirable property that it is unbiased. This is expressed in the following claim.

Claim 2.1. *Let Z be the output sequence of either the Haar $(1+\beta)$-thinning strategy or the greedy-Haar $(1+\beta)$-thinning strategy. Then for any integrable f we have*

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} f(Z_i)\right] = \int_{[0,1)^d} f(x)dx.$$

We postpone the proof of this claim to Section 5.4. Next, we divide the description of potential applications to one-dimensional and multi-dimensional.

One dimensional applications to statistics. Our results could be used to obtain a new method for on-line sample thinning in statistics. Typically, thinning is not an effective practice in statistics. However, there are settings in which it is actually beneficial. To make this concrete lets us illustrate the application of our method through an example from botany. Consider a setting in which a researcher wishes to assess the expectation of a parameter Y - the amount of a certain bacteria on a type of wild plants. It is well known that Y is strongly dependent in an unknown yet smooth way on the mass of the sampled plant, a well studied parameter which we denote by X. To obtain Y the researcher must harvest the plant, keep it in cold storage and run an expensive procedures, hence it is much more costly to assess Y for any particular sample than to measure X. The researcher now travels in the jungle and measures X for different plants, he can then either discard them or keep them for measuring Y. By applying our results to the percentile distribution of X (which is uniform by definition), we can thin an arbitrarily low percentage of our samples on-line and obtain an empirical percentile distribution of the samples of X which has discrepancy of $O(\log^3 n)$ rather than $O(\sqrt{n})$ discrepancy without any thinning. As a result the average of sampled Y will suffer from less variance caused by the variance of the sampled values of X. Hence the researcher will be able to obtain better precision for a given cost. Notice that by Claim 2.1 this method will not create any bias in the estimate of $\mathbb{E}(Y)$.

Other settings in which a similar application is viable include Experimental agriculture, where an organism (a plant or an animal) is raised and the parameter X could be assessed at a much earlier stage of growth in comparison with Y and Monte-carlo simulations in which Y is obtained from X by heavy computations. To read more on the benefit of thinning for Markov chain Monte-carlo (MCMC) samplers in a similar setting, see a recent work by Owen [22].

Multi-dimensional applications. While the law of large numbers guarantees that a sequence of n independent uniform random converge to the uniform distribution, the rate of this convergence is often slower than desired for practical applications. One setting where this is the case is that of Monte-Carlo numerical integration. In this setting one approximates an intractable continuous integral $\int f(x)dx$ by a discrete average $\frac{1}{n}\sum_{i=1}^{n} f(U_i)$ for uniform U_i. For any arbitrary point sequence $\mathcal{P}$, and compact subset $\mathfrak{F}$ of a Banach space, Holder's inequality implies that $|\frac{1}{n}\sum_{i=1}^{n} f(P_i) - \int f(x)dx| \leqslant \sup_{f\in\mathfrak{F}} \frac{1}{n}\|f\| \operatorname{Dis}(\mathcal{P}_1^n)$ for appropriate norms $\|\cdot\|$ that measure variation of functions. This is called the "Koksma-Hlawka" inequality when discrepancy is measured by axis-aligned rectangles and the functions have bounded "Hardy-Krause" variation (bounded mixed partial derivatives). Since $\operatorname{Dis}(U^n) = \Theta(\sqrt{n})$, the Monte-Carlo sum converges at a $1/\sqrt{n}$ rate to the integral.

One can achieve a much better rate of convergence by replacing random i.i.d. sequences by non-i.i.d. random sequence or even by a deterministic pseudo-random sequence that has lower asymptotic discrepancy than U_1^n. As a result the theory of numerical and *Quasi*-Monte-Carlo (QMC) integration have found applications to several results from discrepancy theory. For more details on bounds related to discrepancy theory and QMC, readers may refer to the books [17, 12, 11] and the surveys [13, 14] and the references therein.

Our sampling algorithm also provides an unbiased estimate for the integral (by Claim 2.1). While the discrepancy and complexity of the algorithm are not as good as low discrepancy methods such as digital nets, it has the benefit of working even in setting where one cannot choose the points at which the function is evaluated. Moreover the output sequence has less structure than lattices based constructions.

Finally, considering application where actual thinning is undesirable, we remark that if rather than allowing to discard every point with probability $1-\beta$, we instead weight each point with a weight of either 1 or $1-\beta$, then our result can be shown to persist.

3 Preliminaries

In this section we formally define $(1+\beta)$-*thinning strategies*, and related notions that are useful for our proofs. We then give a sufficient condition that describes which distributions can be realized by a single step of $(1+\beta)$-thinning, and provide a few technical lemmata required to prove Theorem 2. Throughout we follow the convention that the notation log denotes the logarithm with base 2.

3.1 Thinning functions and strategies

A *thinning function* is a measurable function $f : [0,1)^d \to [0,1]$. We think of the input of such a function as a random element in $[0,1)^d$, and of its output as the probability that we decide to keep the chosen element. Formally, given $X_1, X_2 \in [0,1)^d$ and an independent $U_1 \sim \mathrm{Unif}[0,1)$, we let Z' be equal to X_1 if $U_1 \leqslant f(X_1)$ and equal to X_2 otherwise. We call Z_1 the *two-thinned sample* produced by f.

A *two-thinning strategy* is an instrument instructing the overseer how to choose a thinning function to produce Z_n given $Z_1, \dots, Z_{n-1}$. Formally, such a strategy is a countable collection of measurable functions $f_n : ([0,1)^d)^{n-1} \times [0,1)^d \to [0,1]$, such that for every fixed value of the first $n-1$ entries, the function on the last entry is a thinning function.

A two-thinning strategy is applied to produce a random *two-thinning sequence* in the following way. Denote by X a sequence of i.i.d. uniform random variables on $[0,1)^d$. We now inductively define Z as a subsequence of X produced by the strategy. To do so, we shall employ $U = \{U_n\}_{n\in\mathbb{N}}$ a sequence of i.i.d. $\mathrm{Unif}[0,1]$ random variables, independent from everything else, serving as an external source of randomness. Given $Z_1, \dots, Z_{n-1}$, inductively define

$$\chi_n = \mathbb{1}\{U_n > f_n((Z_1, \dots, Z_{n-1}), X_{n+\sum_{i=1}^{n-1}\chi_i})\}.$$

Here, χ_n represents the decision whether to reject (1) or keep (0) in the n-th step so that $\sum_{i=1}^{n}\chi_i$ is the number of rejections made by our algorithm in the process of allocating the first n balls. Using these we set $Z_n = X_{n+\sum_{i=1}^{n}\chi_i}$. Observe that, conditioned on $Z_1, \dots, Z_{n-1}$, the variable Z_n indeed has the distribution of a two-thinning sample according to $f(\cdot) = f_n((Z_1, \dots, Z_{n-1}), \cdot)$.

3.2 $(1+\beta)$-thinning strategy

Given a fixed $\beta \leqslant 1$, a thinning function f satisfying $f \geqslant 1-\beta$ almost surely is called a $(1+\beta)$-*thinning function* and a two-thinning sample of such a function is called a $(1+\beta)$-*thinned sample*. A $(1+\beta)$-*thinning strategy* is a two-thinning strategy which, for every given $Z_1, \dots, Z_{n-1}$, satisfies that $f(x) = f_n((Z_1, \dots, Z_{n-1}), x)$ is an $(1+\beta)$-thinning function. Observe that such a strategy rejects each sample, conditioned on the past, with probability at most β and that the case $\beta = 1$ coincides with our previous definitions.

3.3 Distribution realization via $(1+\beta)$-thinning

In this section we provide a sufficient condition for a distribution on $[0,1)^d$ to be realizable as a $(1+\beta)$-thinned sample.

Proposition 3.1. *Let μ be an absolutely continuous probability measure on $[0,1)^d$ whose density g satisfies*

$$1 - \frac{\beta}{2} \leqslant g(x) \leqslant 1 + \frac{\beta}{2}.$$

Then, $f(x) = g(x) - \frac{\beta}{2}$ defines a $(1+\beta)$-thinning function whose $(1+\beta)$-thinned sample is distributed according to μ.

Proof. Let $X_1, X_2 \sim \mathrm{Unif}([0,1)^d)$ and $U_1 \sim \mathrm{Unif}([0,1])$, independent from one another and let Z' be equal to X_1 if $U_1 \leqslant f(X_1)$ and to X_2 otherwise, so that Z' is a $(1+\beta)$-thinned sample of f. We compute

$$\begin{aligned}
\mathbb{P}(Z' \in A) &= \mathbb{P}\Big(X_1 \in A, U_1 \leqslant f(X_1)\Big) + \mathbb{P}\Big(X_2 \in A, U_1 > f(X_1)\Big) \\
&= \int_A \left(g(z) - \frac{\beta}{2}\right) dz + \int_{[0,1)^d} \left(1 - g(z) + \frac{\beta}{2}\right) dz \cdot |A| \\
&= \left(\mu(A) - \frac{\beta}{2}|A|\right) + \left(1 - 1 + \frac{\beta}{2}\right)|A| = \mu(A),
\end{aligned}$$

where $|A|$ is the Lebesgue measure of A. The proposition follows. □

Proposition 3.1 is pivotal in the indirect constructions of this paper. Rather than describing thinning functions we shall describe a discrete time stochastic process on $[0,1)^d$ whose n-th entry represents the location of the n-th ball. We then show that almost surely at every step the distribution of the next ball is realizable as a $(1+\beta)$-thinned sample for some easily computable f.

3.4 Processes defined via a conditional density function

Let $(\Omega, \mathcal{F})$ be the measurable space on $([0,1)^d)^{\mathbb{N}}$ with the sigma field generated by the cylindrical Borel topology. We call an $(\Omega, \mathcal{F})$-measurable random variable *a discrete time process* on $[0,1)^d$. Each process $Z = \{Z_n\}_{n\in\mathbb{N}}$ of this sort is associated with a counting process $\nu = \{\nu_n\}_{n\in\mathbb{N}}$ defined by $\nu_n = \sum_{i=1}^n \delta_{Z_i}$ where δ_x is a dirac delta measure at x. We will only concern ourselves with processes whose counting measure ν is Markovian. That is,

$$\nu_n \mid \nu_1, \ldots, \nu_{n-1} \stackrel{d}{=} \nu_n \mid \nu_{n-1}.$$

These are processes satisfying that the distribution of Z_n depends only on the overall locations of the previous $n-1$ balls, and not on their order.

One way to construct a exchangeable discrete time process on $[0,1)^d$ is via a *conditional density function*, which we define as sequence of measurable functions $\lambda_n(\nu)$, each of which takes as input a counting measure of n elements in $[0,1)^d$ and produces a density function λ_n of a probability measure on $[0,1)^d$. Given such λ_n, we write

$$\lambda_n^A(\nu) = \int_A \lambda_n(\nu)(x)dx,$$

for every measurable $A \subset [0,1)^d$.

Given such a conditional density function λ, we define the process Z associated with it by

$$\mathbb{P}\Big(Z_n \in A \;\Big|\; \{Z_i\}_{i<n}\Big) = \lambda_n^A\left(\nu_{n-1}\right). \tag{1}$$

We call Z the process associated with counting measure ν and conditional density λ.

3.5 Balancing pairs

Let Z be a process on $[0,1)^d$ associated with counting measure ν and conditional density λ.

Given two disjoint sets $A, B \subset [0,1)^d$ satisfying $|A| = |B| = \kappa$, we say that Z is *θ-balancing* with respect to the pair $\{A, B\}$ from time $s \in \mathbb{N}$ if almost surely, $\kappa \leqslant \lambda_n^A + \lambda_n^B \leqslant 3\kappa$ and

$$\begin{aligned} \lambda_n^A &\geqslant \lambda_n^B + \theta\kappa \quad \text{if } \nu_n(A) < \nu_n(B), \\ \lambda_n^B &\geqslant \lambda_n^A + \theta\kappa \quad \text{if } \nu_n(B) < \nu_n(A), \end{aligned} \tag{2}$$

for all $n \geqslant s$.

We now turn to show a key concentration property of balancing pairs. Assume that Z is θ-balancing with respect to $\{A, B\}$ from time s and let $n \geqslant s$. By Equation (1) we have,

$$\begin{aligned} \mathbb{P}\Big(Z_n \in A \;\Big|\; \{Z_i\}_{i<n},\ Z_n \in A \cup B, \nu_n(B) < \nu_n(A)\Big) &= \frac{\lambda_n^A(\nu_{n-1})}{\lambda_n^A(\nu_{n-1}) + \lambda_n^B(\nu_{n-1})} \\ &\leqslant \frac{\frac{1}{2}\Big(\lambda_n^A(\nu_{n-1}) + \lambda_n^B(\nu_{n-1}) - \theta\kappa\Big)}{\lambda_n^A(\nu_{n-1}) + \lambda_n^B(\nu_{n-1})} \\ &= \frac{1}{2} - \frac{1}{2}\frac{\theta\kappa}{\lambda_n^A(\nu_{n-1}) + \lambda_n^B(\nu_{n-1})} \leqslant \frac{1}{2} - \frac{\theta}{6}. \end{aligned} \tag{3}$$

3.6 Concentration bounds for balancing processes

The following lemma shows that being θ-balancing with respect to a pair $\{A, B\}$ implies exponential concentration of the difference between the number of balls in A and B.

Lemma 3.2. *Let $s \in \mathbb{N}$, $0 < \theta < 1$ and $A, B \subset [0,1)^d$ be disjoint. If Z is a process on $[0,1)^d$ which is θ-balancing with respect to $\{A, B\}$ from time s and satisfies $\mathbb{E}\Big(\exp\big(\theta\frac{|\nu_s(A)-\nu_s(B)|}{2}\big)\Big) \leqslant \frac{150}{\theta^2}$, then for all $n \geqslant s$ we have*

$$\mathbb{E}\left(\exp\left(\theta\frac{|\nu_n(A) - \nu_n(B)|}{2}\right)\right) \leqslant \frac{150}{\theta^2}.$$

To show Lemma 3.2 we shall employ the following super-martingale type criterion.

Lemma 3.3. *Let $(M_k)_{k\geqslant 0}$ be random variables taking values in $\mathbb{R}_+$ which satisfy*

$$\mathbb{E}(M_k \mid F_{k-1}) \leqslant \alpha M_{k-1} + \beta$$

for some $0 < \alpha < 1$, $\beta > 0$, where $F_n = \sigma\big((M_k)_{0\leqslant k\leqslant n}\big)$. Then

$$\mathbb{E}(M_k) \leqslant \left(1 - \alpha^k\right)\frac{\beta}{1-\alpha} + \alpha^k\mathbb{E}(M_0) \qquad \textit{for all } k.$$

Proof.

$$\mathbb{E}(M_k) = \mathbb{E}(\mathbb{E}(M_k \mid F_{k-1})) \leqslant \alpha\mathbb{E}(M_{k-1}) + \beta.$$

Using induction over k the lemma follows. □

We are now ready to prove Lemma 3.2.

Proof of Lemma 3.2. Let $A, B \subset [0,1)^d$ be disjoint, $0 < \theta < 1$ and $s \in \mathbb{N}$, and assume that Z is a process on $[0,1)^d$ which is θ-balancing with respect to $\{A, B\}$ from time s, which satisfies

$$\mathbb{E}\left(\exp\left(\theta \frac{|\nu_s(A) - \nu_s(B)|}{2}\right)\right) \leqslant \frac{150}{\theta^2} \tag{4}$$

Writing $\mathcal{F}_n = \sigma\big((Z_k)_{0 \leqslant k \leqslant n}\big)$, we observe that for any $b > 0$,

$$\begin{aligned}
\mathbb{E}\left(\exp\left(\theta \frac{|\nu_n(A) - \nu_n(B)|}{2}\right) \;\middle|\; \exp\left(\theta \frac{|\nu_{n-1}(A) - \nu_{n-1}(B)|}{2}\right) = e^{\frac{b\cdot\theta}{2}}, \mathcal{F}_{n-1}\right) \\
&\leqslant e^{\frac{b\cdot\theta}{2}}\left(1 - \kappa + \kappa\left(\frac{1-\theta/3}{2}e^{\frac{\theta}{2}} + \frac{1+\theta/3}{2}e^{-\frac{\theta}{2}}\right)\right) \\
&= e^{\frac{b\cdot\theta}{2}}\left(1 - \kappa + \kappa\left(\cosh\frac{\theta}{2} - \frac{\theta}{3}\sinh\frac{\theta}{2}\right)\right) \\
&\leqslant e^{\frac{b\cdot\theta}{2}}\left(1 - \kappa(1 - e^{-\frac{\theta^2}{24}})\right)
\end{aligned}$$

Where the first inequality follows from (3) and the last inequality uses a Taylor expansion of $\cosh(x)$ and $\sinh(x)$.

For $b = 0$, we have

$$\begin{aligned}
\mathbb{E}\left(\exp\left(\theta \frac{|\nu_n(A) - \nu_n(B)|}{2}\right) \;\middle|\; \exp\left(\theta \frac{|\nu_{n-1}(A) - \nu_{n-1}(B)|}{2}\right) = 1, F_{n-1}\right) \\
&\leqslant 1 - 3\kappa + 3\kappa e^{\frac{\theta}{2}} \\
&\leqslant 1 - \kappa(1 - e^{-\theta^2/24}) + 3\kappa e^{\frac{\theta}{2}}.
\end{aligned}$$

Using the above bounds, we obtain

$$\begin{aligned}
\mathbb{E}\left(\exp\left(\theta \frac{|\nu_n(A) - \nu_n(B)|}{2}\right) \;\middle|\; \exp\left(\theta \frac{|\nu_{n-1}(A) - \nu_{n-1}(B)|}{2}\right), F_{n-1}\right) \\
&\leqslant \alpha\left(\exp\left(\theta \frac{|\nu_{n-1}(A) - \nu_{n-1}(B)|}{2}\right)\right) + \beta
\end{aligned}$$

with $\alpha = 1 - \kappa(1 - e^{-\theta^2/24})$ and $\beta = 3\kappa e^{\frac{\theta}{2}}$. Taking $M_k = \exp\left(\theta \frac{|\nu_k(A) - \nu_k(B)|}{2}\right)$ and observing that for all n we have $\sigma\big((M_k)_{0 \leqslant k \leqslant n}\big) \subseteq F_n$, we apply Lemma 3.3 together with (4) to get

$$\mathbb{E}\left(\exp\left(\theta \frac{|\nu_n(A) - \nu_n(B)|}{2}\right)\right) < (1 - \alpha^n)\frac{3\kappa e^{\frac{\theta}{2}}}{\kappa(1 - e^{-\theta^2/24})} + \alpha^n \mathbb{E}\Big(\exp\left(\theta \frac{|\nu_s(A) - \nu_s(B)|}{2}\right)\Big) < \frac{150}{\theta^2}.$$

□

We also make the following observation.

Observation 3.4. *Let $A^1, \dots, A^k$ be a collection of random variables such that for all $i \in [k]$ we have $\mathbb{E}(\exp(cA^i)) < C$ for some constants c, C. Then for any non-negative $a_1, \dots, a_k$ such that $\sum_{i=1}^k a_i \leqslant 1$, we have*

$$\mathbb{E}\left(\exp\left(c\sum_{i=1}^k a_i A^i\right)\right) < C.$$

Proof. This is an immediate consequence of Jensen's inequality and the convexity of the exponential function. □

Finally, we require the following estimate.

Observation 3.5. *Let Z be an process on $[0,1)^d$, associated with counting measure ν and conditional density λ with $\lambda_n(x) < 2$ for all n, x. Let $\mathcal{D} \subseteq [0,1)^d$ be a measurable set. Then, for any $s \in \mathbb{N}, 0 < \alpha < 1$, we have*

$$\mathbb{E}(e^{\alpha(\nu_s(\mathcal{D}) - \nu_0(\mathcal{D}))}) \leqslant e^{4s|\mathcal{D}|\alpha}$$

Proof. If $|\mathcal{D}| > \frac{1}{2}$, the inequality is straightforward. Otherwise, $\mathbb{E}(e^{\alpha(\nu_s - \nu_0)\mathcal{D}})$ is bounded from above by the moment generating function of Binomial distribution with parameters s and $2|\mathcal{D}|$, which is $(1 + 2|\mathcal{D}|(e^\alpha - 1))^s$. Using the fact that $1 + x < e^x < 1 + 2x$ for all $x \in [0,1]$, we bound this by $e^{2s|\mathcal{D}|(e^\alpha - 1)} \leqslant e^{4s|\mathcal{D}|\alpha}$. □

4 Haar Functions

A *diadic interval* is an interval of the form $I = [a2^{-\ell}, (a+1)2^{-\ell})$ for $\ell, a \in \mathbb{Z}$. We call ℓ the *order of* I and write $\mathcal{O}(I) = \ell$.

Given a diadic interval I of order ℓ, we define I_{even} and I_{odd} as its left and right halves – in particular, they are the unique diadic intervals satisfying $\mathcal{O}(I_{\text{even}}) = \mathcal{O}(I_{\text{odd}}) = \ell+1$, $\inf I_{\text{even}} = \inf I$ and $\sup I_{\text{odd}} = \sup I$. Each diadic interval I satisfying $I_{\text{even}} \subset [0,1)$ is associated with a *Haar function* $H_I : [0,1) \to \{-1, 0, 1\}$ defined by

$$H_I(x) = \begin{cases} 1 & x \in I_{\text{even}}, \\ -1 & x \in I_{\text{odd}}, \\ 0 & \text{otherwise}, \end{cases}$$

and we define the order of H_I by $\mathcal{O}(H_I) = \mathcal{O}(I_{\text{even}}) = \mathcal{O}(I_{\text{odd}}) = \mathcal{O}(I) + 1$. It is not hard to verify that Haar functions associated with different diadic intervals are orthogonal with respect to the inner product $\langle f, g\rangle := \int_{[0,1)^d} f(t)g(t)dt$, and that they forms an orthogonal basis for $L^2([0,1])$. This is known as the *Haar wavelet basis*. Note that the functions here are not normalized so that $\langle H, H\rangle = |supp(H)| \leqslant 1$. Also note that the indicator function of any diadic interval I of order ℓ is orthogonal to all Haar functions of order greater than ℓ.

These notion generalize naturally to $d > 1$. A *diadic rectangle* $R \subset \mathbb{R}^d$ is the cartesian product of diadic intervals $I_1 \times \cdots \times I_d$. The Haar function $H_R : [0,1)^d \to \{\pm 1, 0\}$ associated with this rectangle is $H_R = \prod_{i=1}^d H_{I_i}$. The orders of these are given by $\mathcal{O}(R) := \sum_{i=1}^d \mathcal{O}(I_i)$ and $\mathcal{O}(H_R) = \sum_{i=1}^d \mathcal{O}(H_{I_i})$.

Write $\mathcal{H}_{h_0}^{h_1} = \left\{H_R \ : \ h_0 \leqslant \mathcal{O}(R) \leqslant h_1\right\}$ for the set of diadic Haar functions on $[0,1)^d$ of order between h_0 and h_1. As before, Haar functions form the orthogonal *Haar wavelet basis* of $L^2([0,1]^d)$. For a Haar function we also define

$$H^+ := \{x \in [0,1)^d \ : \ H(x) = 1\} \text{ and } H^- := \{x \in [0,1)^d \ : \ H(x) = -1\},$$

so that $\langle \nu_t, H\rangle = \nu_t(H^+) - \nu_t(H^-)$.

4.1 Writing arbitrary rectangles in terms of Haar functions

As mentioned in the overview, our strategy maintains balance with respect to all Haar functions up to a certain granularity in order to control the discrepancy on arbitrary rectangles. To this end we first express every *diadic* rectangle as a linear combination of Haar functions. We then use this construction to use Haar functions to represent every rectangle whose corners are located on a lattice, and later to approximate any arbitrary rectangle.

Proposition 4.1. *For any diadic rectangle R in $[0,1)^d$ of order ℓ we have*

$$\mathbb{1}_R = \sum_{H \in \mathcal{H}_0^\ell} \frac{\langle \mathbb{1}_R, H\rangle}{\langle H, H\rangle} H.$$

Moreover, $\sum_{H\in\mathcal{H}_0^\ell} \left|\frac{\langle \mathbb{1}_R, H\rangle}{\langle H, H\rangle}\right| = 1$.

Proof. Since $\mathcal{H}_0^\infty$ is an orthogonal basis for $L^2([0,1]^d)$, it would suffice to show that $\langle \mathbb{1}_R, H'\rangle = 0$ for all $H' \in H_{\ell+1}^\infty$.

To this end let H' be a Haar function of order greater than ℓ and denote $H = \prod_{i=1}^d H_i$ and $R = \otimes_{i=1}^d I_i$. Since $\mathcal{O}(H) > \mathcal{O}(R)$ there must exist $j \in [d]$ such that $\mathcal{O}(H_j) > \mathcal{O}(I_j)$. As noted before, this implies that $\langle \mathbb{1}_{I_j}, H_j\rangle = 0$. Since $\mathbb{1}_R = \prod_{i=1}^d \mathbb{1}_{I_i}$ and $H = \prod_{i=1}^d H_i$ we obtain that $\langle \mathbb{1}_R, H\rangle = \prod_{i=1}^d \langle \mathbb{1}_{I_i}, H_i\rangle = 0$ as required.

To see the last part, observe that if $\langle \mathbb{1}_{D_i}, H\rangle \neq 0$ then either $D_i \subseteq supp(H^-)$ or $D_i \subseteq supp(H^+)$. Hence for any point $x \in D_i$ we have $\frac{\langle \mathbb{1}_{D_i}, H\rangle}{\langle H, H\rangle} H(x) \geqslant 0$ from which the last part follows. □

Define a *lattice rectangle* in $[0,1)^d$ of order $\ell \in \mathbb{N}$ to be a rectangle whose corners are on the lattice $2^{-\ell}\mathbb{Z}$. In the next proposition we provide a decomposition of lattice rectangles of order ℓ into diadic rectangles.

Proposition 4.2. *Every lattice rectangle of order $\ell \geqslant 1$ in $[0,1)^d$ can be written as the disjoint union of at most $(2\ell)^d$ disjoint diadic rectangles of order at most ℓ.*

Proof. We begin by showing that any interval of order $\ell \geqslant 1$ in $d = 1$ can be written as the disjoint union of at most 2ℓ disjoint diadic intervals. We prove using induction on ℓ. For the case $\ell = 1$ the statement is straightforward. For a diadic interval $I = [a2^{-\ell}, b2^{-\ell})$ with $0 \leqslant a < b \leqslant 2^\ell$ we write

$$I = [a2^{-\ell}, a'2^{-\ell}) \cup [a'2^{-\ell}, b'2^{-\ell}) \cup [b'2^{-\ell}, b2^{-\ell})$$

where $a' = a + \mathbb{1}_{a \text{ is odd}}$, $b' = b - \mathbb{1}_{b \text{ is odd}}$ and we interpret $[c, c) = \varnothing$. Since the middle interval is of order at most $\ell - 1$, by our induction assumption, it can be written as a disjoint union of at most $2\ell - 2$ diadic intervals.

For general d, given $R = \otimes_{j=1}^d I_j$ with $I_j = [a_j 2^{-\ell}, b_j 2^{-\ell})$ this allows us to decompose each I_j into disjoint diadic intervals $I_{j1}, \ldots, I_{jk_j}$ for $k_j \leqslant 2\ell$. Writing $R = \bigcup_{1\leqslant m_j \leqslant k_j} \otimes_{j=1}^d I_{jm_j}$. □

Finally, we bound the error when approximating any rectangle by a pair of lattice rectangles, one of which is slightly larger and one which is slightly smaller.

Proposition 4.3. *Let $\ell, d \in \mathbb{N}$. For any rectangle R contained in $[0,1)^d$ there exist lattice rectangles R^-, R^+ of order at most $d\ell$ such that $R^- \subseteq R \subseteq R^+$ and $0 \leqslant |R^1 \backslash R^2| \leqslant 2d2^{-\ell}$.*

Proof. Let $R = \otimes_{i=1}^d [x_i, y_i) \subseteq [0,1)^d$, and write $R^- = \otimes_{i=1}^d [\lceil 2^\ell x_i\rceil 2^{-\ell}, \lfloor 2^\ell y_i\rfloor 2^{-\ell})$ and $R^+ = \otimes_{i=1}^d [\lfloor 2^\ell x_i\rfloor 2^{-\ell}, \lceil 2^\ell y_i\rceil 2^{-\ell})$. Clearly $R^- \subseteq R \subseteq R^+$. Writing $r_i = \lfloor 2^\ell y_i\rfloor 2^{-\ell} - \lceil 2^\ell x_i\rceil 2^{-\ell}$ we have $|R^-| = \prod_{i=1}^d r_i$ and $|R^+| \leqslant \prod_{i=1}^d (r_i + 2^{1-\ell})$. The Propositions follows. □

5 The Haar $(1+\beta)$-thinning strategy

In this section we present the *Haar* $(1+\beta)$*-thinning strategy* which guarantees asymptotically low discrepancy, and show that it satisfies Theorem 2.

Throughout, let $h = h(n) = \lfloor \log n \rfloor$. This will serve as the largest order Haar function being considered by our strategy at time n. We denote $W(s) = \sum_{i=1}^{s} \binom{i+d-1}{d-1}$ and let Z be a process on $[0,1)^d$ associated with counting measure ν and conditional density λ, defined by

$$\lambda_n(\nu_n)(x) := 1 + \frac{\beta}{2W(h)} \sum_{H \in \mathcal{H}_1^h} \operatorname{sgn}\langle \nu_n, -H \rangle H(x). \tag{5}$$

We begin by observing that

Observation 5.1. *Z is a $(1+\beta)$-thinned sample of a $(1+\beta)$-thinning strategy.*

Proof. Observe that $\int_{[0,1)^2} H(x)dx = 0$ for all $H \in \mathcal{H}_1^h$ and therefore, $\int_{[0,1)^2} \lambda_n(\nu_n)(x)dx = 1$. We only need to verify that the condition of Proposition 3.1 is satisfied at every $n \in \mathbb{N}$, i.e. that for all $x \in [0,1]^d$ and $n \in \mathbb{N}$ we have

$$1 - \frac{\beta}{2} \leqslant \lambda(x) \leqslant 1 + \frac{\beta}{2},$$

which follows immediately from (5), and from the fact that for all $x \in [0,1)$ we have

$$\sum_{H \in \mathcal{H}_1^h} |H(x)| = |\{s \in \mathbb{N}_0^d, 0 < \sum_{i=1}^{d} s_i \leqslant h\}| = \sum_{i=1}^{h} \binom{i+d-1}{d-1} = W(h). \tag{6}$$

□

In light of the claim we call the strategy producing Z the *Haar* $(1+\beta)$-thinning strategy.

Next, in Section 5.1 we discuss the complexity of realizing this strategy. In Section 5.2 we show exponential concentration properties related to Z. Finally in section 5.3 we use these to prove Theorem 2.

5.1 Realizing the Haar thinning strategy

In this section we discuss the time and memory complexity required for the overseer to realize the Haar thinning strategy. In particular we show the following.

Proposition 5.2. *In order to apply the $(1+\beta)-$Haar thinning strategy the overseer requires $\mathcal{O}(n \log^d n)$ memory and $\mathcal{O}(n \log^d n)$ computations to produce the first n samples.*

Proof. Recall that in our set-up the overseer is given a uniformly distributed point $X_{n+\sum_{i=1}^{n-1}\chi_i}$ in $[0,1)^d$. Then, relying upon a data structure which he maintain, the overseer he must compute a threshold $\tau_n \in [0,\beta]$. Then with probability τ_n the value of χ_n is set to be 1 and otherwise it is set to be 0. In light of Proposition 3.1, in order to realize Z we must set $\tau_n = \lambda_n(\nu_n)(x) - \frac{\beta}{2}$. The rest of the section discusses the complexity of computing this function.

We remark that, as the custom goes, complexity estimates are given for integer computations and ignore the increase in storage, reference and computation costs for large numbers. If these were taken into account additional poly-$\log\log n$ factors would multiply both time and memory.

As before, let $n \in \mathbb{N}$, recall that $h = h(n) = \lfloor \log n \rfloor$ and $W(s) = \sum_{i=1}^{s} \binom{i+d-1}{d-1}$, and denote by H_I a Haar function corresponding to the diadic rectangle I. For each function H_I of order ℓ with

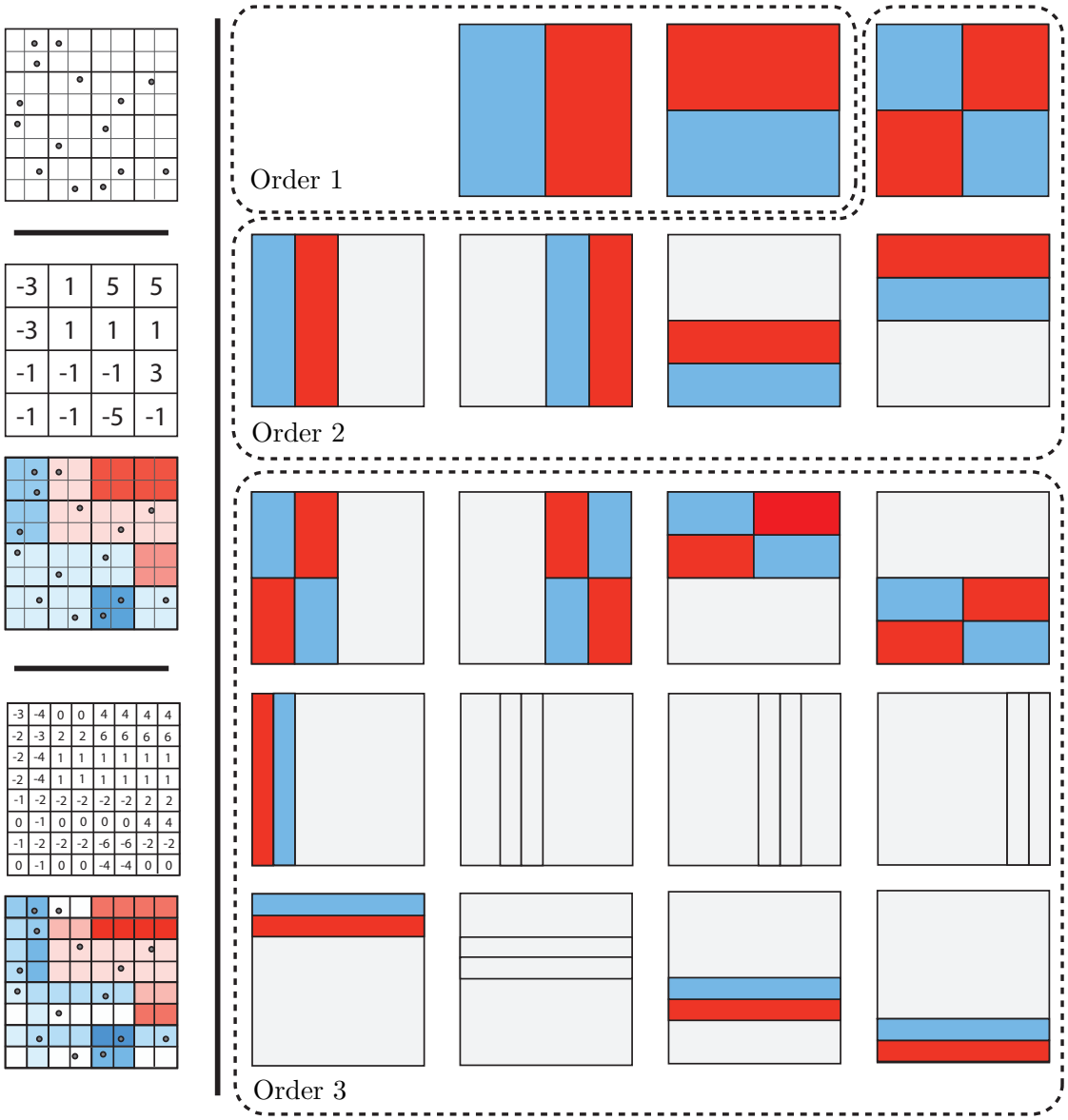


Figure 1: **Left, top.** ν_{15} the empirical measure of a sequence sampled according to our thinning strategy at $n = 15$. **Left, middle.** $\sum_{H \in \mathcal{H}_1^3} \operatorname{sgn}\langle \nu_n, -H \rangle H(x)$, averaged on diadic squares of side $1/4$ along with a visual representation of the conditional density of λ_{15} averaged on diadic squares of side $1/4$. Warmer color indicates higher density. **Left, bottom.** The conditional density of λ_{15} along with a visual representation. **Right.** Haar functions of orders one to three, multiplied by $\operatorname{sgn}\langle \nu_t, -H \rangle$. Gray indicates the value 0, blue – the value -1 and red – the value 1. Notice that the fully grayed out functions of order three are the ones which are perfectly balanced.

$I = I_1 \times \cdots \times I_d$ and $\mathcal{O}(I_i) = \ell_i$ so that $\ell = 1 + \sum \ell_i$, we call $\ell_1, \ldots, \ell_d$ the *shape* of I. At time n we maintain an array A_n of gradually increasing size. A_n consists of data cells corresponding to each Haar function in $\mathcal{H}_1^h$. Each of these cells associated with $H_I \in \mathcal{H}_1^h$ contains the present value of $\langle \nu_n, H_I \rangle$. Observe that the size of such an array is bounded by the total number of shapes which is $W(h) = O(\log^d n)$, multiplied by the maximal number of elements of each shape of order $\ell \leqslant h$ which is $2^{\ell+d} = O(n)$, giving total memory complexity of $O(n \log^d n)$, as required.

The arrangement of A is as follows. We order the data cells first by their order ℓ, then lexicographically by shape and then lexicographically by the point whose coordinate sum is minimal in I. With this arrangement we can find the cells of all Haar functions of order less than ℓ containing a particular point at the cost of a constant number of arithmetic operations per function.

Given this array, computing the value of τ_n takes $W(h) = O(\log^d n)$ operation, one for each element of the sum. Using this we can determine the value of Z_n. We then update A_n by altering the value of all entries corresponding to Haar functions associated with rectangles containing Z_n. As noted in (6), this takes $W(h) = O(\log^d n)$ operations. In addition, for each n such that $h(n) > h(n-1)$ we must allocate additional entries to A for the new $\binom{h+d-1}{d-1}$ shapes of order $h(n)$. There are less than $2n$ Haar functions for each of these shapes so that this operation takes less than $2nW(h)$ operations. We then go over all points $Z_1, \ldots, Z_n$ and update the entries of A_n corresponding to Haar functions associated with the new shapes at the cost of $O(nW(h))$ steps. Hence to produce the first n entries and the time complexity is

$$O(n \log^d n) + \sum_{s=1}^{\lfloor \log n \rfloor} \Big(O(2^{d+s} W(s)) + O(2^s W(s)) \Big) = O(n \log^d n),$$

concluding the proof of the proposition. □

5.2 Concentration properties of Z

We begin by showing that diadic projections of Z have a balancing nature. Recall that $h = \lfloor \log n \rfloor$.

Proposition 5.3. *For any Haar function H on $[0,1)^d$ we have*

$$\mathbb{E}\left(e^{\frac{\beta |\langle \nu_n, H \rangle|}{2W(h)}} \right) < \frac{600 W(h)^2}{\beta^2}. \tag{7}$$

Proof. Let H be a Haar function on $[0,1)^d$ of order $\ell \in \mathbb{N}$. We use different arguments for times before and after 2^ℓ. We begin by showing that Z is $\frac{\beta}{W(h)}$-balancing with respect to $\{H^+, H^-\}$ starting from time $s = 2^\ell$. We write $\kappa := |H^-| = |H^+|$, let $2^\ell \leqslant M \leqslant n$ and observe that

$$\begin{aligned} \lambda_M^{H+} - \lambda_M^{H-} &= \langle H, \lambda \rangle = \left\langle H, 1 + \frac{\beta}{2W(\lfloor \log M \rfloor)} \sum_{G \in \mathcal{H}_1^{\lfloor \log M \rfloor}} G(x) \operatorname{sgn}\langle \nu_M, -G \rangle \right\rangle \\ &= \frac{2\beta\kappa \operatorname{sgn}\langle \nu_M, -H \rangle}{2W(\lfloor \log M \rfloor)}. \end{aligned}$$

Hence the conditions of (2) are satisfied with $\theta = \frac{\beta}{W(h)}$. Next we show (7). Indeed, for any time $M \leqslant 2^\ell$, we have

$$\mathbb{E}(e^{\frac{\beta}{2W(h)} |\langle \nu_M, H \rangle|}) \leqslant \mathbb{E}(e^{\frac{\beta}{2W(h)} \langle \nu_M, |H| \rangle}) \overset{(i)}{\leqslant} e^{\frac{2\beta}{W(h)} 2^{\min(d,h)}} \overset{(ii)}{\leqslant} 100 < \frac{600 W(h)^2}{\beta^2}. \tag{8}$$

Here inequality (i) follows from Observation 3.5 using $M \leqslant 2^\ell, |\mathcal{D}| = |supp(H)| \leqslant \min(2^{d-\ell}, 1)$. To see inequality (ii) we claim that $W(h) \geqslant 2^{\min(h,d)-1}$. Indeed, if $h \leqslant d$, then

$$W(h) \geqslant \binom{h+d-1}{h} \geqslant \frac{d^h}{h!} \geqslant \frac{h^h}{h!} \geqslant 2^{h-1},$$

while if $h > d$, then

$$W(h) \geqslant \binom{h+d-1}{d-1} \geqslant \frac{h^{d-1}}{(d-1)!} \geqslant \frac{d^{d-1}}{(d-1)!} \geqslant 2^{d-1}.$$

From this we also deduce that (7) holds in the case $n \leqslant 2^\ell$. By applying Lemma 3.2 with $\{H^+, H^-\}$, $s = 2^\ell$ and $\theta = \frac{\beta}{W(h)}$ we get that (7) holds also in the case $n > 2^\ell$, concluding the proof of the proposition. □

Next, we use this to show concentration of ν_n on low-order lattice rectangles.

Proposition 5.4. *For any $n \in \mathbb{N}$ and any lattice rectangle $R \subset [0,1)^d$ with of order at most h we have*

$$\mathbb{E}\left(e^{\frac{\beta|\nu_n(R)-n|R||}{2^{d+1}h^d W(h)}}\right) \leqslant \frac{600W(h)^2}{\beta^2}.$$

Proof. Let R be a lattice rectangle of order at most h. By Proposition 4.2 there exist disjoint diadic rectangles $D_1, \ldots, D_k$ of order at most h such that $k \leqslant (2h)^d$ and $R = \cup_{i=1}^k D_i$. By Proposition 4.1, each D_i satisfies

$$\mathbb{1}_{D_i} = \sum_{H \in \mathcal{H}_0^h} \frac{\langle \mathbb{1}_{D_i}, H\rangle}{\langle H, H\rangle} H,$$

with $\sum_{H\in\mathcal{H}_0^\ell} \left|\frac{\langle \mathbb{1}_{D_i}, H\rangle}{\langle H,H\rangle}\right| = 1$. We observe that $\langle \mathbb{1}_{D_i}, \mathbb{1}_{[0,1)^d}\rangle = n|D_i|$ where $\mathbb{1}_{[0,1)^d}$ is the only Haar function of order 0. We conclude that there exist coefficients a_H for $H \in \mathcal{H}_1^h$ with $\sum_{H\in\mathcal{H}_1^h} |a_i| \leqslant (2h)^d$ such that

$$\mathbb{1}_R = \sum_{H\in\mathcal{H}_1^h} a_i H + n|R|.$$

By Proposition 5.3, together with Observation 3.4, this implies that

$$\mathbb{E}\left(e^{\frac{\beta|\nu_n(R)-n|R||}{2(2h)^d W(h)}}\right) \leqslant \frac{600W(h)^2}{\beta^2}, \tag{9}$$

as required. □

5.3 Proof of Theorem 2

Let $n \geqslant 4$ and observe for $n < 2^d$ the theorem is straightforward as $\mathrm{Dis}(Z^n) \leqslant n$ almost surely. As before denote $h = \lfloor \log n \rfloor$. We begin by bounding $W(h)$. We compute

$$W(h) \leqslant \frac{h(h+d-1)^{d-1}}{(d-1)!} \leqslant \frac{2^{d-1}h^d}{(d-1)!} \leqslant 25\left(\frac{h}{2}\right)^d.$$

Next, let R be a lattice rectangle of order at most h. By Proposition 5.4 we have

$$\mathbb{E}\left(e^{\frac{\beta|\nu_n(R)-n|R||}{50h^{2d}}}\right) \leqslant \mathbb{E}\left(e^{\frac{\beta|\nu_n(R)-n|R||}{2^{d+1}h^d W(h)}}\right) \leqslant \frac{600W(h)^2}{\beta^2},$$

so that by Markov's inequality for all $\Delta' > 0$ we have

$$\mathbb{P}\left(\left|\nu_n(R) - n|R|\right| \geqslant \Delta' \log^{2d} n\right) \leqslant \frac{2^{19}\log^{2d} n}{2^{2d}\beta^2} e^{-\frac{\Delta'\beta}{50}}.$$

Observe that there are at most n^{2d^2} lattice rectangle of order at most $d \log n$. Denoting $\mathcal{R}_i = \{\text{lattice rectangle } R \subset [0,1)^d \ : \ \mathcal{O}(R) \leqslant i\}$ we have

$$\mathbb{P}\left(\sup_{R \in \mathcal{R}_i} \left|\nu_n(R) - n|R|\right| \geqslant \Delta' \log^{2d} n\right) \leqslant \frac{2^{19}\log^{2d} n}{2^{2d}\beta^2} n^{2d^2} e^{-\frac{\Delta'\beta}{50}}.$$

By proposition 4.3 applied with $\ell = \lfloor \log n \rfloor$, for each rectangle R in $[0,1)^d$ there exist $R^+, R^- \in \mathcal{R}_{d \log n}$ such that $R^- \subseteq R \subseteq R^+$ and $|R^+| \leqslant |R^-| + 4d/n$. Hence

$$\nu_n(R^-) - n|R^-| - 4d \leqslant \nu_n(R^-) - n|R^+| \leqslant \nu_n(R) - n|R| \leqslant \nu_n(R^+) - n|R^-| \leqslant \nu_n(R^+) - n|R^+| + 4d.$$

Hence, for $\Delta' > 0$ we get

$$\mathbb{P}\left(\mathrm{Dis}(Z^n) \geqslant \Delta' \log^{2d} n + 4d\right) \leqslant \beta^{-2} \exp\left(\frac{2d^2 \log n}{\log e} + \log(\log n) + \log(2d) + \frac{19 - 2d}{\log e} - \frac{\beta \Delta'}{50}\right).$$

Plugging in $\Delta' = \beta^{-1}(\Delta + 1000 + 100d^2 \log n)$ we obtain

$$\mathbb{P}\Big(\mathrm{Dis}(Z^n) \geqslant \beta^{-1} \log^{2d}(N)(\Delta + 1000 + 100d^2 \log n)\Big) \leqslant \beta^{-2} e^{-\frac{\Delta}{50}},$$

as required. □

5.4 Proof of Claim 2.1

It would suffice to show the Z is unbiased for $\mathbb{1}_R$ where $R = \otimes_{i=1}^d I_i$ is a diadic rectangle ℓ. To see this we show that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \mathbb{1}_R(Z_i)\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{R'}(Z_i)\right] \tag{10}$$

for any $R' = \otimes_{i=1}^d I_i'$ with $\mathcal{O}(I_i') = \mathcal{O}(I_i)$. This is a consequence of the diadic tree symmetry. To see this, consider the binary representation of R and R' in each dimension and write D_i for the digits in which they disagree in dimension i. Let $g : [0,1)^d \to [0,1)^d$ be the measure preserving bijection which maps a point x to a point $g(x)$ whose binary representation in each coordinate i is flipped exactly on D_i. Now couple the sequence X_i and with a sequence $X_i' = g(X_i)$ and apply the same strategy to produce $\{Z_j\}_{j \in \mathbb{N}}$ and $\{Z_j'\}_{j \in \mathbb{N}}$ using the same sequence U used to determine our thinning decisions as in Section 3.1. Observe that in this case $Z_i' = g(Z_i)$ so that for all n we have

$$\frac{1}{n}\sum_{i=1}^n \mathbb{1}_R(Z_i) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{R'}(Z_i'),$$

and hence, as $Z_i' \stackrel{d}{=} Z_i$, (10) holds. □

6 The greedy-Haar strategy

In this section we describe the empirically more efficient variant of our strategy called the *greedy-Haar strategy*. We then provide heuristic justification for Conjectures 1 and 2.

Unlike the case of the Haar strategy, we describe the strategy directly by

$$f_n((Z_1,\dots,Z_{n-1}),x) = \begin{cases} 1 & \sum\limits_{H\in\mathcal{H}_1^h} \operatorname{sgn}\langle \nu_n, -H\rangle H(x) < 0 \\ \frac{1}{2} & \sum\limits_{H\in\mathcal{H}_1^h} \operatorname{sgn}\langle \nu_n, -H\rangle H(x) = 0 \\ 0 & \sum\limits_{H\in\mathcal{H}_1^h} \operatorname{sgn}\langle \nu_n, -H\rangle H(x) > 0 \end{cases}$$

The name greedy-Haar corresponds to the a point of view by which each Haar function H wishes to reduce $\langle \nu_n, H\rangle$. Hence we compute $\sum_{H\in\mathcal{H}_1^h} \operatorname{sgn}\langle \nu_n + \mathbb{1}_x, H\rangle - \sum_{H\in\mathcal{H}_1^h} \operatorname{sgn}\langle \nu_n, H\rangle$ and if this quantity is positive we keep x, if it is negative we reject it, and if it is 0, we break the tie by a fair coin-toss.

6.1 Heuristic analysis

We begin by describing the $\log^{d/2}(n)$ heuristic improvement to Theorem 1, giving rise to Conjecture 1. We then describe an additional heuristic $\log^{d/2}(n)$ improvement stemming from the greedy-Haar strategy which adds up to Conjecture 2.

Improvement of the analysis (Conjecture 1). We conjecture that the usage of Observation 3.4 to obtain (9) is not tight. In this transition we decompose each rectangle R to the sum Haar functions whose coefficients add up to at most $\log_2^d n$. We then bound the rectangle's discrepancy by a triangle inequality using the bound for each individual Haar funciton. However, for a rectangle R and a Haar function h we have $|\langle \mathbb{1}_R, h\rangle|/\langle h, h\rangle \leqslant 1$, so the coefficient of each particular Haar function is at most 1. Hence, assuming sufficient independence between the coefficients $\langle \mathbb{1}_R, h\rangle$ for different Haar functions h, we should expect the sum of $\langle \nu, H\rangle$ to produce a discrepancy of $\log_2^{d/2} n$, and not $\log_2^d n$.

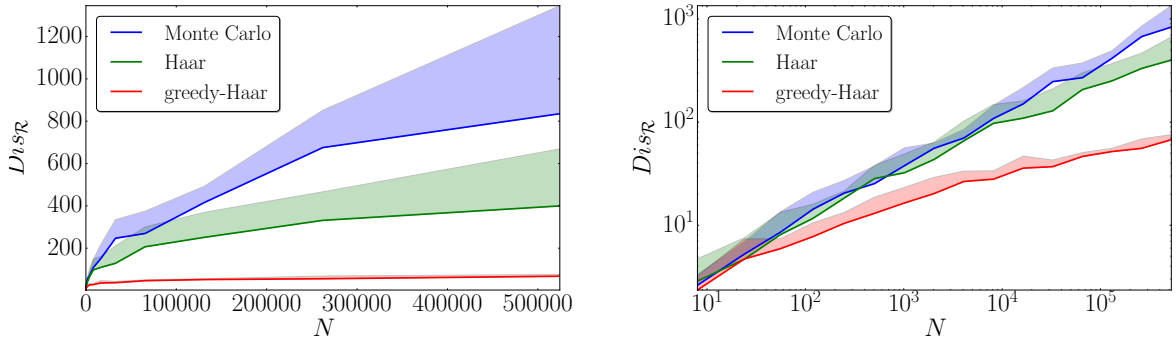
Better concentration inequalities for the greedy-Haar strategy (Conjecture 2). Let $H \in \mathcal{H}_1^h$ be a particular Haar function and assume that $\langle \nu_n, -H\rangle > 0$. Denote by k the number of elements of $\mathcal{H}_1^h$ whose support contains a given point. Also recall the notation G^- and G^+, the positive and negative domains of a Haar function G. We examine the probability of that a point falls in H^+ compared with the probability that it falls in H^-. Observe that every other Haar function $G \in \mathcal{H}_1^h \backslash \{H\}$ is orthogonal to H so that $(\operatorname{sgn}\langle \nu_n, -G\rangle)\langle H, G\rangle) = 0$. In addition, if we approximate the signs of $\langle \nu_n, -G\rangle$ for $G \in \mathcal{H}_1^h \backslash \{H\}$ by independent random variables , then their total value would have a binomial$(\frac{1}{2}, k)$ distribution, so that typically on a region of size $\frac{k^{-1/2}}{2} supp(H)$ they are tied and the sign of $\langle \nu_n, -H\rangle$ determines whether to accept or reject. Hence we expect the process $\langle \nu_n, -H\rangle$ to behave roughly like an $\Theta(1/h^{d/2})$ balancing process which would yield an improvement of $\log_2^{d/2} n$ to the bound.

7 Empirical results

In this section we provide simulation results both for the Haar and the greedy-Haar 2-thinning strategies. As evident from these simulations, the greedy-Haar strategy is significantly better than the Haar strategies, and both strategies perform somewhat better than shown by our Theorems.

Strategy	$n = 56$	$n = 2^7$	$n = 2^9$	$n = 2^{11}$	$n = 2^{13}$	$n = 2^{15}$	$n = 2^{17}$	$n = 2^{19}$
Monte Carlo	8.6 (2.5)	14.3 (3.3)	25.3 (6.1)	55.8 (3.5)	108.6 (18.4)	247.1 (43.9)	415.8 (38.4)	835.3 (255.2)
Haar	8.1 (2.6)	11.7 (2.2)	28.3 (4.9)	43.5 (10.2)	97.3 (25.4)	128.8 (40.9)	251.4 (59.1)	399.8 (134.7)
greedy-Haar	5.9 (0.7)	7.8 (1.4)	13.0 (8.4)	20.3 (5.5)	28.0 (2.9)	37.0 (2.9)	51.9 (1.9)	67.5 (4.0)

Table 1: Some Values of Discrepancy for different strategies in one-dimension. Given are the mean (and standard error) across 20 experiments.



(a) $d = 1$, Linear Scale

(b) $d = 1$, Log Scale

Figure 2: Plots of discrepancy (averaged over 20 experiments) in one dimensions for Monte Carlo (blue), Haar 2-Thinning (green) and Greedy-Thinning 2-Thinning (red) strategies. The plots are provided both in linear and log-scale.

We begin by showing discrepancy results, and then discuss the bias of particular rectangles. In all simulations we compare the three methods, i.i.d. samples which we refer to here as Monte-Carlo, Haar 2-thinning, and greedy-Haar 2-thinning. Unfortunately the simulations are not sufficient to determine the power of the log in the decay of the discrepancy with sufficient certainty to scientifically estimate the exponent of the log in Conjecture 2.

7.1 Main Simulations

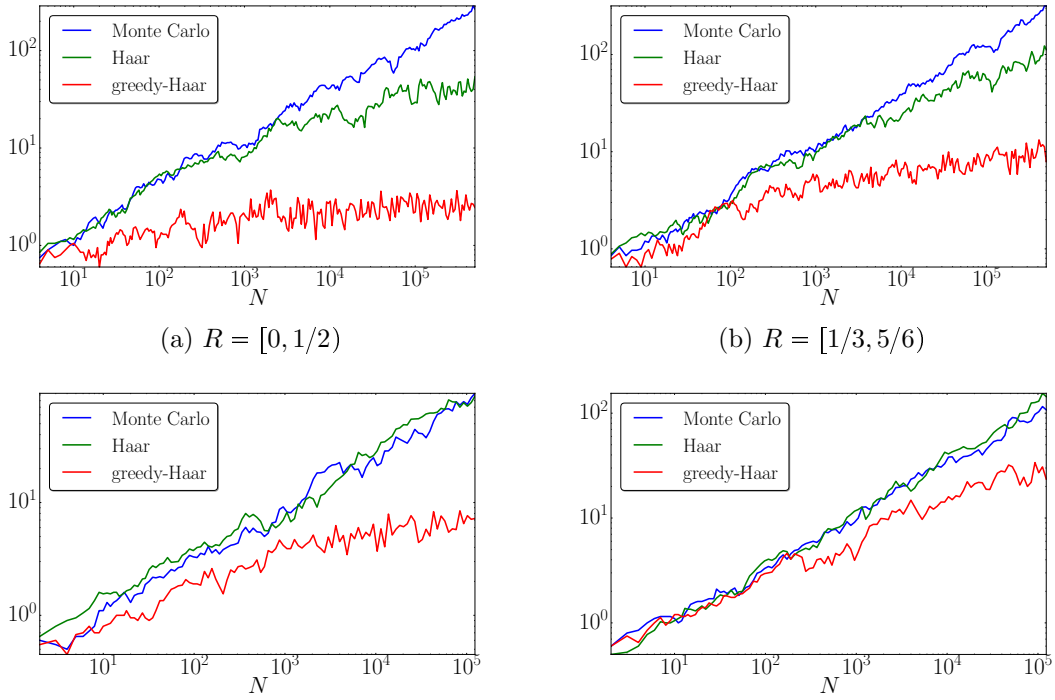
We have averaged 20 simulated outputs of 2^{19} samples for each of the three strategies in one dimension. For this case, we have computed the rectangle R which has maximal $|\nu_N(R) - |R||$ whenever $N = \lceil 2^k \rceil$. Our results are summarized in Table 1 and Figure 2.

7.2 Other Simulations

We were also interested in the performance of the strategies on a diadic rectangles and on a given rectangle whose decomposition intro Haar-functions has high coefficients. These show the validity of our estimates for such rectangles, and verify the logic of the proof. For this purpose we chose the intervals $[0, \frac{1}{2}]^d$ and $[\frac{1}{3}, \frac{5}{6}]^d$, the first of which is diadic while the other has a very complex diadic decomposition. Comparison between those rectangles in one and two dimensions are given in Table 2 and Figure 3. The results clearly indicate the the biases of these rectangles are dominated by a different power of $\log(n)$.

	$d = 1$			$d = 2$		
	Monte Carlo	Haar	greedy-Haar	Monte Carlo	Haar	greedy-Haar
$R = [0, 1/2)^d$						
$n = 10$	1.0 (1.0)	1.2 (0.9)	1.1 (0.7)	1.1 (0.6)	1.6 (1.2)	0.8 (0.5)
$n = 100$	4.9 (3.4)	5.3 (3.0)	1.1 (1.5)	3.2 (2.2)	3.8 (2.7)	1.9 (1.7)
$n = 1000$	10.1 (7.0)	7.9 (6.2)	2.0 (2.2)	8.2 (6.1)	7.5 (4.3)	4.7 (3.4)
$n = 10000$	42.5 (36.2)	22.9 (18.3)	2.5 (2.7)	22.4 (17.9)	27.1 (17.9)	5.1 (4.5)
$n = 100000$	102.6 (57.3)	28.2 (26.7)	3 (2.2)	76.2 (54.8)	73.1 (75.0)	6.7 (4.9)
$R = [1/3, 5/6)^d$						
$n = 10$	1.2 (0.8)	1.4 (0.9)	0.9 (0.6)	1.1 (0.7)	1.1 (.06)	1.2 (0.7)
$n = 100$	3.4 (2.2)	3.1 (2.6)	2.8 (1.7)	3.1 (2.2)	4.1 (3.0)	3.0 (1.7)
$n = 1000$	10.5 (8.5)	9.1 (10.1)	5.4 (4.0)	10.2 (6.7)	12.9 (7.4)	5.0 (3.1)
$n = 10000$	36.5 (25.0)	22.9 (17.8)	6.5 (5.7)	36.1 (22.3)	40.6 (24.8)	13.8 (12.5)
$n = 100000$	123.9 (118.1)	57.3 (37.7)	10.4 (6.2)	102.3 (87.4)	131.4 (88.7)	30.6 (22.1)

Table 2: Mean (standard error) biases for different rectangles in one and two dimensions.



(a) $R = [0, 1/2)$ (b) $R = [1/3, 5/6)$

(c) $R = [0, 1/2)^2$ (d) $R = [1/3, 5/6)^2$

Figure 3: Plots of the bias $|\nu(R) - |R||$ in logarithmic scale averaged over 20 experiments for different rectangles R in one and two dimensions.

7.3 Observations from the simulations

We draw the following observations from the simulations

- Both Haar and greedy-Haar seem to be always at least as good as Monte-Carlo sampling.
- Greedy-Haar strategy seem to be always at least as good the Haar strategy.
- In one dimension greedy-Haar performs significantly better than Monte-Carlo sampling for as little as 50 samples.

Acknowledgments

The authors wish to thank Itai Benjamini for suggesting the model of the power of two choices on interval partitions, to Yuval Peres for introducing to us related recent litrature, and to Art Owen and Asaf Nachmias for useful discussions.

References

[1] Y. Azar, A. Broder, A. Karlin and E. Upfal, *Balanced allocations*, SIAM Journal of Computing 29 no. 1 (1999), pp. 180-200.

[2] W. W. L. Chen, and M. M. Skriganov. *Explicit constructions in the classical mean squares problem in irregularities of point distribution*, J. Reine Angew. Math. 545, 6795 (2002).

[3] W. W. L. Chen, and M. M. Skriganov. *Davenport's theorem in the theory of irregularities of point distribution*, Journal of Mathematical Sciences 115, no. 1 (2003): 2076-2084.

[4] M. M. Skriganov. *Harmonic analysis on totally disconnected groups and irregularities of point distributions*, J. Reine Angew. 600, 2549 (2006).

[5] P. Berenbrink, A. Czumaj, A. Steger and B. Vöcking, *Balanced allocations: The heavily loaded case*, SIAM Journal on Computing 35 no. 6 (2006), pp. 1350-1385.

[6] D. Bilyk, Chapter *Roths orthogonal function method in discrepancy theory and some new connections*, in *A panorama of discrepancy theory.* Springer International Publishing (2014), pp. 71-158.

[7] M. Donsker, *Justification and extension of Doobs heuristic approach to the Kolmogorov Smirnov theorems*, The Annals of Mathematical Statistics 23 (1952), pp. 277-281.

[8] O. N. Feldheim and O. Gurel-Gurevich, *The power of thinning in ballanced allocation*, In perparation.

[9] M. Junge, *Choices, intervals and equidistribution*, Electronic Journal of Probability 20 (2015).

[10] S. Kakutani, *A problem of equidistribution on the unit interval* $[0, 1]$, Proceedings of Oberwolfach Conference on Measure Theory. Lecture Notes in Math. 541 (1975), pp. 367-376.

[11] Niederreiter, Harald, *Random number generation and quasi-Monte Carlo methods*, Society for Industrial and Applied mathematics (1992).

[12] Dick, Josef, and Friedrich Pillichshammer, *Digital nets and sequences: Discrepancy Theory and QuasiMonte Carlo Integration*, Cambridge University Press (2010).

[13] Dick, J., Kuo, F.Y. and Sloan, I.H., *High-dimensional integration: the quasi-Monte Carlo way*, Acta Numerica, 22 (2013), pp.133-288.

[14] Dick, J. and Pillichshammer, F., *Discrepancy theory and quasi-Monte Carlo integration*, A Panorama of Discrepancy Theory (2014), pp. 539-619.

[15] Roth, K.F., *On irregularities of distribution*, Mathematika, 1(02) (1954), pp.73-79.

[16] Bilyk, D., Lacey, M.T. and Vagharshakyan, A., *On the small ball inequality in all dimensions*, Journal of Functional Analysis, 254(9) (2008), pp.2470-2502.

[17] Chazelle, Bernard, *The discrepancy method: randomness and complexity*, Cambridge University Press, 2000.

[18] J. C. Lootgieter. *Sur la répartition des suites de Kakutani.* Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques 13 no. 4 (1977), pp. 385-410.

[19] P. Maillard and E. Paquette, *Choices and intervals*, Israel Journal of Mathematics (2014). pp. 1-48.

[20] M. Mitzenmacher, *The Power of Two Choices in Randomized Load Balancing*, PhD thesis, University of California, Berkeley, CA, 1996.

[21] M. Mitzenmacher and E. Upfal. *Probability and computing: Randomized algorithms and probabilistic analysis,* Cambridge University Press, 2005.

[22] A. B. Owen, *Statistically efficient thinning of a Markov chain sampler*, Journal of Computational and Graphical Statistics, 2017, accepted.

[23] Y. Peres, K. Talwar and U. Wieder, *Graphical Balanced Allocations and the $(1+\beta)$-Choices Process*, Random Structures & Algorithms 47, no. 4 (2015), pp. 760-775 157-163.

[24] R. Pyke, *The Asymptotic Behavior of Spacings Under Kakutani's Model for Interval Subdivision*, The Annals of Probability 8 no. 1 (1980), pp. 157-163.

[25] R. Pyke and W. R. van Zwet, *Weak convergence results for the Kakutani interval splitting procedure*, The Annals of Probability 32 no. 1 (2004), pp. 380-423.

[26] A. W. Richa, M. Mitzenmacher and R. Sitarman, *The power of two random choices: A survey of techniques and results*, Combinatorial Optimization 9 (2001), pp. 255-304.

[27] K. Talwar and U. Wieder, *Balanced allocations: A simple proof for the heavily loaded case.* International Colloquium on Automata, Languages, and Programming. Springer Berlin Heidelberg, 2014.

[28] W. R. van Zwet, *A Proof of Kakutanis Conjecture on Random Subdivision of Longest Intervals*, The Annals of Probability 6 no. 1 (1978), pp. 133-137.