



HAL
open science

On Negative Dependence Properties of Latin Hypercube Samples and Scrambled Nets

Benjamin Doerr, Michael Gnewuch

► **To cite this version:**

Benjamin Doerr, Michael Gnewuch. On Negative Dependence Properties of Latin Hypercube Samples and Scrambled Nets. 2021. hal-03205337v1

HAL Id: hal-03205337

<https://hal.science/hal-03205337v1>

Preprint submitted on 22 Apr 2021 (v1), last revised 19 Sep 2021 (v2)

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.

On Negative Dependence Properties of Latin Hypercube Samples and Scrambled Nets

Benjamin Doerr* Michael Gnewuch†

April 22, 2021

Abstract

We study the notion of γ -negative dependence of random variables. This notion is a relaxation of the notion of negative orthant dependence (which corresponds to 1-negative dependence), but nevertheless it still ensures concentration of measure and allows to use large deviation bounds of Chernoff-Hoeffding- or Bernstein-type. We study random variables based on random points P . These random variables appear naturally in the analysis of the discrepancy of P or, equivalently, of a suitable worst-case integration error of the quasi-Monte Carlo cubature that uses the points in P as integration nodes. We introduce the correlation number, which is the smallest possible value of γ that ensures γ -negative dependence. We prove that the random variables of interest based on Latin hypercube sampling or on (t, m, d) -nets do, in general, not have a correlation number of 1, i.e., they are not negative orthant dependent. But it is known that the random variables based on Latin hypercube sampling in dimension d are actually γ -negatively dependent with $\gamma \leq e^d$, and the resulting probabilistic discrepancy bounds do only mildly depend on the γ -value.

1 Introduction

For the solution of discrete or continuous problems, algorithms based on independent random variables are a popular choice. This is mostly due to the following two advantages.

- Stochastic independence usually leads to an easy analysis of the resulting error. In particular, typically the concentration of measure phenomenon allows to use helpful large deviation bounds.
- Additionally, the actual error can be estimated statistically by repeating the computation several times.

*Laboratoire d'Informatique (LIX), CNRS, École Polytechnique, Institut Polytechnique de Paris, Palaiseau, France (lastname@lix.polytechnique.fr).

†Institut für Mathematik, Universität Osnabrück, Germany (michael.gnewuch@uni-osnabrueck.de).

But, depending on the specific setting, using independent random variables may also have serious disadvantages. Let us mention some examples.

- Most likely, the corresponding algorithms do not achieve the best possible asymptotic convergence rate.
- The corresponding algorithms may not provide admissible solutions. Specific examples from discrete mathematics and computer science are integer linear programs with certain hard constraints – to satisfy these hard constraints the random variables involved necessarily have to be dependent.
- If the independent random variables are random sample points, they exhibit a lack of (combinatorial, algebraic or geometric) structure. For instance, independent random samples typically contain (undesired) point clusters or gaps. Furthermore, they do not possess a group structure like, e.g., lattice rules or digital (t, m, s) -nets, which allows to use Fourier analytical methods.

Therefore it is an important and interesting problem to find algorithms based on dependent random variables that, on the one hand side, share the advantages of algorithms based on independent random variables (which are most notably the easy error analysis and the benefits of the concentration of measure phenomenon) and, on the other hand side, avoid for the specific problem at hand their unwanted disadvantages. This problem is further motivated in much more detail in the introductions of [4, 6, 9] for the specific setting where one wants to approximate expectations of random variables f or integrals $\int_{[0,1]^d} f(\mathbf{x}) d\mathbf{x}$ of integrands $f : [0, 1]^d \rightarrow \mathbb{R}$ by estimators

$$\frac{1}{N} \sum_{n=1}^N f(X_n),$$

where X_1, \dots, X_N are random points.

To tackle this problem and to analyze promising randomized algorithms, we need an appropriate probabilistic framework. A suitable framework is the one of negatively dependent (or negatively correlated) random variables. We focus here on the following notion for Bernoulli or binary random variables, i.e., random variables that only take values in $\{0, 1\}$.

Definition 1.1. Let $\gamma \geq 1$. We call binary random variables T_1, T_2, \dots, T_N *upper γ -negatively dependent* if

$$\mathbb{P} \left(\bigcap_{j \in J} \{T_j = 1\} \right) \leq \gamma \prod_{j \in J} \mathbb{P}(T_j = 1) \quad \text{for all } \emptyset \neq J \subseteq [1..N], \quad (1)$$

and *lower γ -negatively dependent* if

$$\mathbb{P} \left(\bigcap_{j \in J} \{T_j = 0\} \right) \leq \gamma \prod_{j \in J} \mathbb{P}(T_j = 0) \quad \text{for all } \emptyset \neq J \subseteq [1..N]. \quad (2)$$

We call T_1, T_2, \dots, T_N *γ -negatively dependent* if both conditions (1) and (2) are satisfied.

Essentially the same notion, called λ -correlation, was introduced earlier, but in a formally different way in [18]. 1-negative dependence is usually called *negative orthant dependence*, cf. [1]. To allow for a parameter $\gamma \geq 1$ (or $\lambda \geq 1$, respectively) instead of insisting on γ to be one relaxes the condition of negative orthant dependence, but still ensures that large deviation bounds of Chernoff-Hoeffding type hold for sums of γ -negatively dependent random variables, see [18, 6, 3].

Now one may ask the question if allowing the parameter γ to be different from 1 is a fruitful generalization. To the best of our knowledge, parameters γ (or λ for the notion of λ -correlation) different from 1 have been employed successfully for the first time in [6] to establish favorable pre-asymptotic upper bounds for the star discrepancy of Latin hypercube samples [19].

Let us explain this application in more detail: Here we deal with binary random variables $T_i, i = 1, \dots, N$, of the form $T_i = \mathbf{1}_A(X_i)$, where A is a Lebesgue-measurable subset of $[0, 1]^d$ whose characteristic function is denoted by $\mathbf{1}_A$, and X_1, \dots, X_N are random points in $[0, 1]^d$.

Definition 1.2. For $d \in \mathbb{N}$ we put

$$\mathcal{C}_0^d := \{[0, a] \mid a \in [0, 1]^d\} \quad \text{and} \quad \mathcal{D}_0^d := \{B \setminus A \mid A, B \in \mathcal{C}_0^d\}.$$

Let $\mathcal{S} \in \{\mathcal{C}_0^d, \mathcal{D}_0^d\}$. We say that a random point set $\mathbf{X} = (X_n)_{n=1}^N$ in $[0, 1]^d$ is \mathcal{S} - γ -negatively dependent if for each $S \in \mathcal{S}$ the random variables

$$\mathbf{1}_S(X_1), \dots, \mathbf{1}_S(X_N) \quad \text{are } \gamma\text{-negatively dependent.}$$

It was shown in [6, Theorem 3.6] that Latin hypercube samples \mathbf{X} are \mathcal{D}_0^d - e^d -negatively dependent. Similar results for other random point sets can be found in [23]. Furthermore, [6, Theorem 4.4] established that for each $c > 0$ a \mathcal{D}_0^d - γ -negatively dependent random point set \mathbf{X} satisfies the star discrepancy bound

$$D_N^*(\mathbf{X}) \leq c \sqrt{\frac{d}{N}} \tag{3}$$

with success probability at least $1 - \gamma \exp(-(1.6741 \cdot c^2 - 10.7042) \cdot d)$. (The constants in the bound of the success probability have been improved slightly in [7].) Note that the success probability depends only mildly on the parameter γ , hence larger values of γ can easily be leveraged by slightly increasing the parameter c . This is, e.g., visible in [6, Corollary 4.5]: Let $d, N \in \mathbb{N}$. Estimate (3) implies that there exists a realization P_{MC} of Monte Carlo points (i.e., \mathcal{D}_0^d -1-negatively dependent random points) having star discrepancy

$$D_N^*(P_{\text{MC}}) \leq 2.5287 \cdot \sqrt{\frac{d}{N}}$$

and a realization P_{LHS} of Latin hypercube sampling (i.e., of \mathcal{D}_0^d - e^d -negatively dependent random points) satisfying

$$D_N^*(P_{\text{LHS}}) \leq 2.6442 \cdot \sqrt{\frac{d}{N}}.$$

Note that the γ -values 1 and e^d differ significantly (especially in high dimension d), but the resulting constants 2.5287 and 2.6442 differ only marginally. (Both constants have recently been improved in [7] to 2.4968 and 2.6135, respectively.)

Nevertheless, the question remains if Latin hypercube samples are even \mathcal{D}_0^d -negatively orthant dependent and therefore introducing a parameter $\gamma > 1$ is unnecessary if we just use a different proof idea than the one used in the proof of [6, Theorem 3.6]?

Since Latin hypercube samples are actually \mathcal{C}_0^d -1-negatively dependent, see [6, Theorem 3.6], several colleagues expressed their belief that Latin hypercube samples are also \mathcal{D}_0^d -1-negatively dependent. In Proposition 2.3 we show that this is actually not the case. This gives strong reason to believe that there are other important families of random variables that are *not* negatively orthant dependent, but *still* γ -negatively dependent for some $\gamma > 1$ and therefore, nevertheless, fall under the concentration of measure phenomenon and satisfy large deviation bounds.

For the quantitative analysis of negatively dependent random variables we introduce the notion of the \mathcal{S} -correlation number.

Definition 1.3. Let $\mathcal{S} \in \{\mathcal{C}_0^d, \mathcal{D}_0^d\}$, and $\mathbf{X} = (X_n)_{n=1}^N$ be a random point set in $[0, 1]^d$. The \mathcal{S} -correlation number $\gamma_{\mathcal{S}}(\mathbf{X}) \in [1, \infty]$ of \mathbf{X} is defined as the smallest γ such that \mathbf{X} is \mathcal{S} - γ -negatively dependent, i.e.,

$$\gamma_{\mathcal{S}}(\mathbf{X}) := \sup \left\{ \frac{\mathbb{P}(\bigcap_{j \in J} \{\mathbf{1}_{\mathcal{S}}(X_j) = 0\})}{\prod_{j \in J} \mathbb{P}(\mathbf{1}_{\mathcal{S}}(X_j) = 0)}, \frac{\mathbb{P}(\bigcap_{j \in J} \{\mathbf{1}_{\mathcal{S}}(X_j) = 1\})}{\prod_{j \in J} \mathbb{P}(\mathbf{1}_{\mathcal{S}}(X_j) = 1)} \mid S \in \mathcal{S}, \emptyset \neq J \subseteq [1..N] \right\}.$$

In Section 2 we recall the definition of Latin hypercube samples and show that their \mathcal{D}_0^d -correlation number grows with the dimension d at least as $\Omega(\sqrt{d})$, see Proposition 2.3. In Section 3 we study scrambled (t, m, d) -nets. The random points of scrambled nets exhibit even stronger dependencies than Latin hypercube samples. Together with lattice rules, (t, m, d) -nets are arguably the most popular choice of sample points used in quasi-Monte Carlo methods. Since there are many different ways known how to scramble (t, m, d) -nets, we first introduce the notion of an abstract scrambled (t, m, d) -net, see Definition 3.3. This notion captures the crucial properties that random scrambling should exhibit. Indeed, Proposition 3.6 shows that the most important variants of scramblings directly satisfy this notion or can easily be modified to do so. In Remark 3.8 we prove that in order to obtain small correlation numbers it is indispensable that the considered scrambled (t, m, d) -nets are symmetrized, i.e., that their random points are exchangeable. In Proposition 3.9 we demonstrate that also abstract scrambled (t, m, d) -nets do, in general, not have a \mathcal{D}_0^d -correlation number of one.

2 Latin Hypercube Sampling

In this section, we show that Latin hypercube samples are not \mathcal{D}_0^d -1-negatively dependent. In fact, there is no constant γ (independent of d) such that we have \mathcal{D}_0^d - γ -negative dependence. This shows, in particular, that the use of a \mathcal{D}_0^d -correlation number larger

than one in the discrepancy analysis of Latin hypercube samples in [6] cannot be avoided (though the gap between our lower bound of $\Omega(\sqrt{d})$ and the upper bound of $\exp(d)$ used in [6] remains large).

In simple words, a Latin hypercube sample of N points in $[0, 1]^d$ is constructed by partitioning $[0, 1]^d$ into N^d axis-aligned cubes isomorphic to $[0, 1/N]^d$, selecting randomly N such cubes in a way that no two overlap in any one-dimensional projection, and then placing one point into each selected cube independently and uniformly at random. While each single point is uniformly distributed in $[0, 1]^d$, the dependent choice of the small cubes leads to dependencies among the points. The precise definition of a Latin Hypercube sampling is as follows.

Definition 2.1. A *Latin hypercube sample* (LHS) $(X_n)_{n=1}^N$ in $[0, 1]^d$ is of the form

$$X_{n,j} = \frac{\pi_j(n-1) + U_{n,j}}{N},$$

where $X_{n,j}$ denotes the j th coordinate of X_n , π_j is a permutation of $0, 1, \dots, N-1$, uniformly chosen at random, and $U_{n,j}$ is uniformly distributed in $[0, 1]$. The d permutations π_j and the dN random variables $U_{n,j}$ are mutually independent.

The definition of Latin hypercube sampling presented above was introduced by McKay, Beckman, and Conover [12] for the design of computer experiments. As shown by Owen [16], an estimator \hat{I} of an integral I based on Latin hypercube samples with N points never leads to a variance greater than that of the corresponding estimator based on $N-1$ Monte Carlo points. As an extreme example, we note that the one-dimensional projections of Latin hypercube samples are much more evenly distributed than the one-dimensional projections of Monte Carlo points. Indeed, for a one-dimensional Latin hypercube sample of size N the star discrepancy is at most $1/N$, while for a one-dimensional Monte Carlo sample of the same size the star discrepancy is of order $1/\sqrt{N}$, cf., e.g., [6, Remark 3.2].

The following theorem is a result from [6].

Theorem 2.2. Let $d \in \mathbb{N}$. Let $\mathbf{X} := (X_n)_{n=1}^N$ be a LHS in $[0, 1]^d$. For $a, b \in [0, 1]^d$ let $A := [0, a)$, $B := [0, b)$, and $D := B \setminus A$. Let

$$\gamma_{A,B} := \prod_{i=1}^d \delta_i \quad \text{with} \quad \delta_i := \begin{cases} 1 & \text{if } a_i, b_i \in \{1/N, 2/N, \dots, 1\} \quad \text{or} \quad a_i = 0, \\ e & \text{else.} \end{cases}$$

Then the random variables

$$\mathbf{1}_D(X_1), \dots, \mathbf{1}_D(X_N) \quad \text{are } \gamma_{A,B}\text{-negatively dependent.} \quad (4)$$

In particular, the random points $(X_n)_{n=1}^N$ are \mathcal{C}_0^d -1-negatively dependent as well as \mathcal{D}_0^d - e^d -negatively dependent.

We now show that if Latin hypercube samples \mathbf{X} consisting of $N = d$ points are \mathcal{D}_0^d - γ -negatively dependent for some $\gamma = \gamma_d$, then γ necessarily satisfies $\gamma_2 \geq 2$ and $\liminf_{d \rightarrow \infty} \gamma_d = \infty$ for d tending to ∞ . In particular, we cannot simply replace the $\gamma_{A,B}$ in Theorem 2.2 by 1 or any other constant independent of d .

Proposition 2.3. *Let $d \geq 2$ be a natural number, and let $\mathbf{X} := (X_n)_{n=1}^d$ be a LHS in $[0, 1]^d$. Then*

$$\gamma_{\mathcal{D}_0^2}(\mathbf{X}) \geq 2 \quad \text{and} \quad \liminf_{d \rightarrow \infty} \frac{\gamma_{\mathcal{D}_0^d}(\mathbf{X})}{\sqrt{d}} \geq \sqrt{\frac{2\pi}{e}}. \quad (5)$$

Proof. Let $\varepsilon \in (0, 1/d)$. We put

$$A = A(\varepsilon, d) := [0, \frac{d-1}{d}]^d \quad \text{and} \quad B = B(\varepsilon, d) := [0, \frac{d-1}{d} + \varepsilon]^d$$

as well as $D = D(\varepsilon, d) := B \setminus A$. We want to verify (5) by showing that

$$\frac{\mathbb{P}(X_1, \dots, X_d \in D)}{\prod_{i=1}^d \mathbb{P}(X_i \in D)} \geq (1 - g(d, \varepsilon)) \sqrt{\frac{2\pi}{e}} \cdot \sqrt{d}, \quad (6)$$

where $g(d, \varepsilon)$ is such that $\lim_{\varepsilon \rightarrow 0} g(d, \varepsilon) = 0$, and that for $d = 2$ the left hand side of (6) tends to 2 if ε tends to 0.

Let $\pi_1, \dots, \pi_d : [0..d-1] \rightarrow [0..d-1]$ be the permutations used in the construction of the Latin hypercube sample $\mathbf{X} = (X_i)_{i=1}^d$. Consider the event

$$\mathcal{E} := \{\forall i \in [0..d-1] \exists j(i) \in [1..d] : \pi_{j(i)}(i) = d-1\}.$$

To have $\{X_1, \dots, X_d\} \subset D$, it is necessary that the event \mathcal{E} occurs, hence

$$\begin{aligned} \mathbb{P}(X_1, \dots, X_d \in D) &= \mathbb{P}(\{X_1, \dots, X_d \in D\} \cap \mathcal{E}) \\ &= \mathbb{P}(\mathcal{E}) \cdot \mathbb{P}(X_1, \dots, X_d \in D \mid \mathcal{E}). \end{aligned}$$

Since we use only d permutations to construct \mathbf{X} , the event \mathcal{E} is equivalent to saying that the $\pi_j^{-1}(d-1)$, $j = 1, \dots, d$, are pairwise different. This observations allows to easily compute the probability of \mathcal{E} by

$$\mathbb{P}(\mathcal{E}) = \frac{[d \cdot (d-1)!] \cdot [(d-1) \cdot (d-1)!] \cdots [1 \cdot (d-1)!]}{[d!]^d} = \frac{d!}{d^d}.$$

The event \mathcal{E} , again by elementary counting, also implies that for each i there is exactly one j such that $\pi_j(i) = d-1$, that is, each point of \mathbf{X} has exactly one coordinate in $[\frac{d-1}{d}, 1)$. Therefore,

$$\mathbb{P}(X_i \in D \mid \mathcal{E}) = \frac{\varepsilon}{1/d} = d\varepsilon \quad \text{for all } i \in [1..d].$$

Conditioned under \mathcal{E} all events $\{X_i \in D\}$ are independent, implying

$$\mathbb{P}(X_1, \dots, X_d \in D \mid \mathcal{E}) = \prod_{i=1}^d \mathbb{P}(X_i \in D \mid \mathcal{E}) = (d\varepsilon)^d.$$

This finally gives

$$\mathbb{P}(X_1, \dots, X_d \in D) = \mathbb{P}(\mathcal{E}) \cdot \mathbb{P}(X_1, \dots, X_d \in D \mid \mathcal{E}) = \frac{d!}{d^d} (d\varepsilon)^d = d! \cdot \varepsilon^d.$$

On the other hand we have

$$\lambda^d(D) = \lambda^d(B) - \lambda^d(A) = \left(\frac{d-1}{d} + \varepsilon\right)^d - \left(\frac{d-1}{d}\right)^d = \varepsilon \left(\sum_{k=0}^{d-1} \binom{d}{k} \left(1 - \frac{1}{d}\right)^k \varepsilon^{d-1-k}\right)$$

and thus, for all $i \in [1..d]$,

$$\mathbb{P}(X_i \in D) = \lambda^d(D) = \varepsilon \cdot d \left(1 - \frac{1}{d}\right)^{d-1} + R(\varepsilon, d),$$

where

$$R(\varepsilon, d) = \varepsilon^2 \left(\sum_{k=0}^{d-2} \binom{d}{k} \left(1 - \frac{1}{d}\right)^k \varepsilon^{d-2-k}\right).$$

This gives us

$$\frac{\mathbb{P}(X_1, \dots, X_d \in D)}{\prod_{i=1}^d \mathbb{P}(X_i \in D)} = \frac{d!}{d^d} \left[\left(1 - \frac{1}{d}\right)^d\right]^{1-d} \cdot T(\varepsilon, d), \quad (7)$$

where

$$T(\varepsilon, d) := \left(1 - \frac{R(\varepsilon, d)}{\lambda^d(D)}\right)^d.$$

We now want to determine the strong asymptotic behavior of the quotient on the left hand side in equation (7) if d tends to infinity and ε tends to zero (sufficiently fast). To this purpose let us first remark that by choosing $\varepsilon = \varepsilon(d)$ small enough we get $\liminf_{d \rightarrow \infty} T(\varepsilon, d) = 1$. Furthermore, it is well-known [20] that

$$\sqrt{2\pi d} e^{-d} < \frac{d!}{d^d} < \sqrt{2\pi d} e^{-d} \exp\left(\frac{1}{12d}\right).$$

Finally, we get $(1 - 1/d)^d \rightarrow e^{-1}$ as $d \rightarrow \infty$ and

$$\begin{aligned} e^{-d} \left[\left(1 - \frac{1}{d}\right)^d\right]^{-d} &= \exp\left(-d \left(1 + d \ln\left(1 - \frac{1}{d}\right)\right)\right) \\ &= \exp\left(-d \left(1 + d \left(-\frac{1}{d} - \frac{1}{2d^2} - \frac{1}{3d^3} - \dots\right)\right)\right) \\ &= \exp\left(\frac{1}{2} + \frac{1}{3d} + \frac{1}{4d^2} + \dots\right) \\ &= e^{1/2} \cdot \exp\left(\frac{1}{d} \left(\frac{1}{3} + \frac{1}{4d} + \frac{1}{5d^2} + \dots\right)\right), \end{aligned}$$

and the last term converges to $e^{1/2}$ as d tends to infinity. Thus we get altogether

$$\frac{\mathbb{P}(X_1, \dots, X_d \in D(\varepsilon, d))}{\prod_{i=1}^d \mathbb{P}(X_i \in D(\varepsilon, d))} \approx \sqrt{\frac{2\pi}{e}} \cdot \sqrt{d} \quad (8)$$

for $d \rightarrow \infty$ in the strong asymptotic sense as long as $\varepsilon = \varepsilon(d) \rightarrow 0$ sufficiently fast.

In the case $d = 2$ our calculations above lead to $\mathbb{P}(X_1, X_2 \in D) = 2\varepsilon^2$ and

$$\mathbb{P}(X_i \in D) = \lambda^2(D) = \left(\frac{1}{2} + \varepsilon\right)^2 - \left(\frac{1}{2}\right)^2 = \varepsilon(1 + \varepsilon) \quad \text{for } i = 1, 2.$$

Hence

$$\lim_{\varepsilon \rightarrow 0} \frac{\mathbb{P}(X_1, X_2 \in D(\varepsilon, 2))}{\mathbb{P}(X_1 \in D(\varepsilon, 2))\mathbb{P}(X_2 \in D(\varepsilon, 2))} = \lim_{\varepsilon \rightarrow 0} \frac{2}{(1 + \varepsilon)^2} = 2.$$

□

3 Scrambled Nets

In this section, we show that also the dependencies stemming from scrambled net constructions can lead to a \mathcal{D}_0^d -correlation number larger than one. We give an example that is not just a Latin hypercube sample, and further, where the dependencies of Latin hypercube samples lead to a truly negative correlation. This indicates that the stronger dependencies of scrambled nets can be more harmful than those of Latin hypercubes.

Definition 3.1. Let $d, N \in \mathbb{N}$. We call an N -point multi-set (or family) $P = (p^{(n)})_{n=1}^N$ in $[0, 1]^d$ fair with respect to a Borel measurable subset F of $[0, 1]^d$ if $N\lambda^d(F) = |\{n \mid p^{(n)} \in F\}|$. In this situation we also say that F is fair with respect to P .

Definition 3.2. Let us fix integers $d, b \geq 2$, and $m \geq t \geq 0$.

- (i) An elementary interval or canonical box E in base b of order $m - t$ is a subset of $[0, 1]^d$ of the form

$$E = [k_1 b^{-\ell_1}, (k_1 + 1)b^{-\ell_1}) \times \dots \times [k_d b^{-\ell_d}, (k_d + 1)b^{-\ell_d})$$

for non-negative integers ℓ_i and $k_i < b^{\ell_i}$, $i = 1, \dots, d$, that has volume $\lambda^d(E) = b^{-(m-t)}$.

- (ii) Let $N = b^m$. An N -point multi-set $P = (p^{(n)})_{n=1}^N$ in $[0, 1]^d$ is called a (t, m, d) -net in base b if every elementary interval in base b of order $m - t$ is fair with respect to P , i.e., if it contains exactly b^t points of P .

The construction of nets goes back to Sobol' [21] and Faure [5]. The concept of (t, m, d) -nets was introduced by Niederreiter in [13] and studied in depths by him and many other researchers, see, e.g., [2, 14] and the literature mentioned therein. As these works show, the property of being a (t, m, d) -net is sufficient to lead to a small discrepancy. In practical applications, one often also wants to have random low-discrepancy point sets, where random means that the points to a certain degree resemble uniformly distributed points. To capture the important properties of existing randomized variants of (t, m, d) -nets, we formulate the following definition.

Definition 3.3. Let b, d, m, t , and N as in Definition 3.2. For all $k_1, \dots, k_d \in [0..b^{m-t} - 1]$, the set $C = \prod_{j=1}^d [k_j b^{-(m-t)}, (k_j + 1)b^{-(m-t)})$ is called a *basic cube*. Let $\mathbf{Y} = (Y_i)_{i=1}^N$ be a tuple of N random points in $[0, 1]^d$. We say that \mathbf{Y} is an *abstract scrambled (t, m, d) -net in base b* if it satisfies the following three conditions.

- (i) With probability one, $(Y_n)_{n=1}^N$ is a (t, m, d) -net in base b .
- (ii) For each $i \in [1..N]$ and each basic cube C , we have $\mathbb{P}(Y_i \in C) = b^{-d(m-t)}$.
- (iii) For all measurable $M \subseteq [0, 1]^d$, all $\emptyset \neq J \subseteq [1..N]$, and all families of basic cubes $(C_j)_{j \in J}$ that satisfy $\mathbb{P}(\forall j \in J : Y_j \in C_j) > 0$, we have

$$\mathbb{P}(\forall j \in J : Y_j \in M \mid \forall j \in J : Y_j \in C_j) = \prod_{j \in J} \frac{\lambda^d(M \cap C_j)}{\lambda^d(C_j)}.$$

We note that condition (ii) and condition (iii) immediately imply the condition

- (iv) For each $i \in [1..N]$, the point Y_i is uniformly distributed in $[0, 1]^d$.

A number of concrete realizations of randomized nets that satisfy our definition of an abstract scrambled net have been proposed in the literature, see, e.g., [15, 10, 8, 17] and the references mentioned there. Now we describe the notion of *scrambling of depth ℓ* as discussed in [10, 11], comment afterwards in Remark 3.7 on other versions of scrambled nets, and demonstrate in Remark 3.8 the importance of symmetrization. In the following, we essentially blend the definitions and notions used in [11, Sect. 2.4] and [17].

For given $b, \ell \in \mathbb{N}$, a mapping $\sigma : [0, 1] \rightarrow [0, 1]$ is called a *b -ary scrambling of depth ℓ* if it maps for all $k \in [1..\ell]$ each elementary interval E in base b of order k into an elementary interval E' in the same base having the same order, and distinct E are mapped into distinct E' .

The next statement is easily verified, cf. [15, Proposition 1] or [11, Observation 2.11].

Lemma 3.4. *If $P \subset [0, 1]^d$ is a (t, m, d) -net in base b and $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_d)$ is a d -tuple of b -ary scramblings of depth $m - t$, then $\boldsymbol{\sigma}(P) = \{\boldsymbol{\sigma}(\mathbf{p}) \mid \mathbf{p} \in P\}$ with $\boldsymbol{\sigma}(\mathbf{p}) := (\sigma_1(p_1), \dots, \sigma_d(p_d))$ is again a (t, m, d) -net in base b .*

The general form of a b -ary scrambling σ of depth ℓ is the following: Let $x \in [0, 1]$ have b -ary digit expansion

$$x = \xi_1 b^{-1} + \xi_2 b^{-2} + \xi_3 b^{-3} + \dots; \tag{9}$$

to ensure that the expansion is unique we assume that infinitely many $\xi_j \in \{0, 1, \dots, b-1\}$ are different from $b-1$. Then the first b -ary digit of $\sigma(x)$ is $\pi(\xi_1)$, where π is a permutation of $\{0, 1, \dots, b-1\}$. For $j = 2, \dots, \ell$ the j th digit of $\sigma(x)$ is given by $\pi_{\xi_1, \dots, \xi_{j-1}}(\xi_j)$, where $\pi_{\xi_1, \dots, \xi_{j-1}}$ is a permutation of $\{0, 1, \dots, b-1\}$ that may depend on the first $j-1$ digits

of x . Furthermore, let $\psi : [0, 1) \rightarrow [0, 1)$ be a mapping. Then the general form of a b -ary scrambling σ of depth ℓ is

$$\sigma(x) = \sum_{j=1}^{\ell} \pi_{\xi_1, \dots, \xi_{j-1}}(\xi_j) b^{-j} + \psi(x) b^{-\ell}.$$

A *nested uniform b -ary random scrambling of depth ℓ* is defined in the following way: We pick permutations $\pi_{\xi_1, \dots, \xi_j}$, $j = 0, 1, \dots, \ell-1$, $\xi_1, \dots, \xi_{\ell-1} \in \{0, 1, \dots, b-1\}$, independently and uniformly at random. To map a finite multiset $F = (x^{(n)})_{n=1}^N$ of $[0, 1)$, we define $\sigma(x)$ for each $x = x^{(n)}$ in F with b -ary representation as in (9) by

$$\sigma(x) = \sum_{j=1}^{\ell} \pi_{\xi_1, \dots, \xi_{j-1}}(\xi_j) b^{-j} + y_x b^{-\ell}, \quad (10)$$

where $y_x = y_{x^{(n)}} \in [0, 1)$ is chosen uniformly at random and independently for $n = 1, \dots, N$. We emphasise that $y_{x^{(i)}}$ and $y_{x^{(j)}}$ are independent random variables for $i \neq j$, even if $x^{(i)} = x^{(j)}$.

Nested uniform b -ary random scrambling is essentially what was proposed by Owen in [15], see also Remark 3.7.

A simplified version of nested uniform b -ary random scrambling of depth ℓ is *positional uniform b -ary random scrambling of depth ℓ* proposed by Matoušek in [10]: Here one only picks for each “digit position” $j \in [\ell]$ a permutation π_j of $\{0, 1, \dots, b-1\}$ independently and uniformly at random and defines for $x = x^{(n)}$ in F as in (9)

$$\sigma(x) = \sum_{j=1}^{\ell} \pi_j(\xi_j) b^{-j} + y_x b^{-\ell}, \quad (11)$$

where, as in (10), $y_x = y_{x^{(n)}}$ is chosen uniformly at random and independently for distinct $n \in [1..N]$.

So in the definitions given above “random scrambling of depth ℓ ” refers to the treatment of the different digits: for each $x \in F$ the first ℓ digits of $\sigma(x)$ are obtained by applying random permutations to the first ℓ digits of x , but the remaining digits of $\sigma(x)$ are determined by independent random variable $y_x b^{-\ell}$ in $[0, b^{-\ell})$, cf. (10) and (11). The attributes “nested” and “positional” refer to the *scrambling framework of the first ℓ digits*: “nested” refers to the case where the permutation used to scramble the j th digit of $x \in [0, 1)$ as in (9) depends (in a nested fashion) on all the previous digits of x and “positional” refers to the case where it only depends on the digit position j itself (independently of the previous digits of x). In both cases it is understood that all permutations are chosen independently at random in some subset of the set of all permutations of $\{0, 1, \dots, b-1\}$. The attribute “uniform” now refers the specific choice of the subset of permutations and the chosen probability distribution on it: we choose them uniformly at random from the set of all permutations of $\{0, 1, \dots, b-1\}$. Other choices are possible, as, e.g., the sets of all *linear permutations*, as suggested in [10], or the set of all

digital shifts, as mentioned in [8]. These sets, endowed with the corresponding uniform probability distribution, yield *nested* and *positional linear* or *nested* and *positional digital shift b -ary random scrambling of depth ℓ* , respectively. Surely, there exist more forms of random scrambling. One example is *affine matrix b -ary random scrambling of depth ℓ* , i.e., applying *affine Matrix random scrambling* as proposed in [10] to the first ℓ digits of $x \in F$ as in (9): for $k = 1, \dots, \ell$ the k th digit of $\sigma(x)$ is given by

$$\sigma(x)_k = h_{k,k}\xi_k + \left(\sum_{j=0}^{k-1} h_{k,j}\xi_j + g_k \right) =: \pi_{\xi_1, \dots, \xi_{k-1}}^{\text{mat}}(\xi_k), \quad (12)$$

where all $h_{k,k} \in \{1, \dots, b-1\}$, $h_{k,j} \in \{0, 1, \dots, b-1\}$, $1 \leq j < k$, $g_k \in \{0, 1, \dots, b-1\}$ are uniformly distributed and mutually independent. This is in some sense a hybrid of nested and positional random scrambling: the resulting random permutations $\pi_{\xi_1, \dots, \xi_{k-1}}^{\text{mat}}$ are clearly nested, but the randomization is positional, i.e., the random variables $h_{1,k}, \dots, h_{k,k}$ and g_k in (12) that generate the random permutations for the k th digits are only depending on the digit position k .

We mention random scramblings that are not “of depth ℓ ” in Remark 3.7.

Observation 3.5. If one wants to construct abstract scramblings by applying a d -tuple $\sigma = (\sigma_1, \dots, \sigma_d)$ of independent b -ary random scramblings of depth $m-t$ of the form (10) to a (t, m, d) -net P in base b , one has to fulfill conditions (i), (ii), and (iii) of Definition 3.3. We make the following observations:

1. Condition (i) is satisfied due to Lemma 3.4.
2. To ensure condition (ii) it is sufficient that for each $j \in [1..d]$ all random permutations π employed by σ_j have *single-fold uniformity* [17, Definition 2.5]

$$\mathbb{P}(\pi(i) = j) = 1/b \quad \text{for all } i, j \in \{0, 1, \dots, b-1\},$$

and all families of these random permutations that scramble different digits are mutually independent. This is obviously satisfied for nested or positional random scrambling if we choose the permutations uniformly at random from the sets of all permutations, of all linear permutations, or of all digital shifts, respectively (cf. also [17]). Furthermore, it is satisfied for affine matrix random scrambling (cf. also [10]).

3. Condition (iii) is satisfied, since the position of each $\sigma(\mathbf{p})$ inside the basic cube it ends up in is solely determined by the additive vectors $\mathbf{y}_{\mathbf{p}} b^{t-m}$ that are independent and uniformly distributed in $[0, b^{t-m})^d$, cf. (10) and (11).

We summarize the findings from Observation 3.5 in the following Proposition.

Proposition 3.6. *Let $P = \{\mathbf{p}_1, \dots, \mathbf{p}_N\} \subset [0, 1)^d$ be a (t, m, d) -net in base b . Let $\sigma = (\sigma_1, \dots, \sigma_d)$ be a d -tuple of mutually independent random scramblings of depth $m-t$ of the form (10), where for each $j \in [1..d]$ all random permutations employed by σ_j have single-fold uniformity and all families of random permutations that scramble different digits are mutually independent. Then $(\sigma(\mathbf{p}_i))_{i=1}^N$ is an abstract scrambled (t, m, d) -net.*

The statement holds, in particular, if each σ_j , $j = 1, \dots, d$, is a nested or positional (i) uniform or (ii) linear or (iii) digit shift b -ary random scrambling of depth $m - t$, or an affine matrix b -ary random scrambling of depth $m - t$.

Remark 3.7. Crucial for the realization of abstract scramblings as b -ary random scramblings of (t, m, d) -nets in base b is the concept of scrambling of depth $m - t$. When introducing b -ary random scrambling in [15], Owen first proposed infinite-digit scrambling

$$\sigma(x) = \sum_{j=1}^{\infty} \pi_{\xi_1, \dots, \xi_{j-1}}(\xi_j) b^{-j} \quad (13)$$

(but for computational issues he also proposed scrambling of depth ℓ for $\ell \geq m$, cf. [15, Section 3.3]). Assume that all random permutations considered have single-fold uniformity and are mutually independent. Then it is easy to see that for $(0, m, d)$ -nets using d -tuples of scramblings of the form (13) and (10) with $\ell \geq m$ lead to the same random scrambling. But this is not true any more if we consider (t, m, d) -nets with parameter $t > 0$; then (10) with $\ell = m - t$ yields an abstract scrambling, but (13) may violate condition (iii) of Definition 3.3.

Remark 3.8. Consider for $m \in \mathbb{N}$ and $N = b^m$ an arbitrary point set $P = (p^{(n)})_{n=1}^N$ in $[0, 1]^d$, and let $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_d)$ be a d -tuple of b -ary scramblings of depth $m - t \geq 1$ or Owen's infinite digit scrambling discussed in Remark 3.7. Then the random point set $\boldsymbol{\sigma}(P) = (\boldsymbol{\sigma}(p^{(n)}))_{n=1}^N$ cannot be \mathcal{C}_0^d - γ -negatively dependent for any γ independent of $N = b^m$. Indeed, due to the pigeon hole principle we find an elementary box E of volume b^{-1} that contains at least N/b points of P , without loss of generality the points $p^{(1)}, \dots, p^{(N/b)}$. Let c be the lower left corner of E . Put $E_0 := E - c$, i.e., E_0 is the elementary box that has the same shape as E and its lower left corner in 0. Then we have for all $j \in [1..N]$ that

$$\mathbb{P}(\boldsymbol{\sigma}(p^{(j)}) \in E_0) = b^{-1},$$

and, since $m - t \geq 1$,

$$\begin{aligned} & \mathbb{P}(\boldsymbol{\sigma}(p^{(1)}), \dots, \boldsymbol{\sigma}(p^{(N/b)}) \in E_0) \\ &= \mathbb{P}(\boldsymbol{\sigma}(p^{(1)}) \in E_0) \cdot \mathbb{P}(\boldsymbol{\sigma}(p^{(2)}), \dots, \boldsymbol{\sigma}(p^{(N/b)}) \in E_0 \mid \boldsymbol{\sigma}(p^{(1)}) \in E_0) = b^{-1} \cdot 1 = b^{-1}. \end{aligned}$$

This immediately yields

$$\gamma_{\mathcal{D}_0^d}(\boldsymbol{\sigma}(P)) \geq \gamma_{\mathcal{C}_0^d}(\boldsymbol{\sigma}(P)) \geq \frac{1}{b} \cdot b^{N/b},$$

showing that those correlation numbers grow at least exponentially in the number of points N . The unfavorable correlation numbers are due to the dependencies of the first $m - t$ digits of all components of the original points that are preserved by the specific randomizations we consider. The remedy to this problem is simple – we just have to symmetrize the random points, i.e., we have to ensure that they are exchangeable random variables. This can be done by simply giving each point of P a random label at the

beginning: Choose a permutation π uniformly at random and put $X_j := \sigma(p^{(\pi(j))})$ for all $j = 1, \dots, N$. Now the random point set $\mathbf{X} = (X_j)_{j=1}^N$ contains exactly the same random points as $\sigma(P)$, but the random points $\sigma(p^{(\pi(1))}), \dots, \sigma(p^{(\pi(N))})$ are now exchangeable. In [9, 22] the authors do random scrambling without choosing random labels; instead the symmetrization procedure is implicitly contained in the definition of the negative dependence properties studied by those authors, cf. [9, Sect. 2.1] and [22, Sect. 2.2]. We emphasize that our lower bound for the correlation number in Proposition 3.9 holds for abstract scrambled nets, regardless if they are symmetrized or not.

Since every Latin hypercube sample with N points is, in particular, an abstract scrambled $(0, 1, d)$ -net in base $b := N$, we see from Proposition 2.3 that there does not exist a constant $\gamma \in [1, \infty)$ such that all $(0, m, d)$ -nets in base b are \mathcal{D}_0^d - γ -negatively dependent. In particular, not all scrambled $(0, m, d)$ -nets are \mathcal{D}_0^d -1-negatively dependent.

The next example shows that there exist scrambled $(0, m, d)$ -nets in base b different from Latin hypercube samples that are not \mathcal{D}_0^d -1-negatively dependent.

Proposition 3.9. *Consider an abstract scrambled $(0, 2, 3)$ -net $\mathbf{Y} = (Y_i)_{i=1}^N$ in base 2. Then for all $\varepsilon > 0$ there is a $D \in \mathcal{D}_0^d$ such that*

$$\frac{\mathbb{P}(Y_1, Y_2, Y_3 \in D)}{\prod_{i=1}^3 \mathbb{P}(Y_i \in D)} \geq \left(\frac{4}{3}\right)^3 - \varepsilon = 2.\overline{370} - \varepsilon.$$

In particular, $\gamma_{\mathcal{D}_0^3}(\mathbf{Y}) \geq 2.\overline{370}$.

Proof. We note first that it follows right from the definitions that each $(0, m, d)$ net P in base b inherits the property from Latin hypercube samples that for every $j \in [1..d]$ and $k \in [1..N]$ there is exactly one point of P in each box $E_{jk} = \prod_{i=1}^d E_i$ with $E_j = [\frac{k-1}{N}, \frac{k}{N})$ and $E_i = [0, 1)$ for $i \neq j$. In particular, P cannot contain two distinct points Y and Y' such that there are $j \in [1..d]$ and $k \in [1..N]$ with $Y_j, Y'_j \in [\frac{k-1}{N}, \frac{k}{N})$. We shall use this elementary *combinatorial property* frequently in the following.

Let now (Y_1, \dots, Y_4) be an abstract scrambled $(0, 2, 3)$ net in base 2. That is, we have $N = 4$ points in $[0, 1)^d$, where $d = 3$, such that each elementary interval in base 2 of order 2 contains exactly one point. Let $\varepsilon \in (0, \frac{1}{4})$.

$$A := [0, \frac{1}{2})^3, B := [0, \frac{1}{2} + \varepsilon)^3, \text{ and } D := B \setminus A.$$

Denote by \mathcal{E} the event that each of Y_1, Y_2 , and Y_3 has exactly one coordinate entry in $[\frac{1}{2}, \frac{3}{4})$ and the other two coordinate entries in $[0, \frac{1}{2})$. Furthermore, put $U := [\frac{3}{4}, 1)^3$. We first show that the combinatorial property implies the following inclusions of events:

$$\{Y_1, Y_2, Y_3 \in D\} \subseteq \mathcal{E} \subseteq \{Y_4 \in U\}. \quad (14)$$

Indeed, assume that $Y_1, Y_2, Y_3 \in D$. Then each of these three points has at least one coordinate entry in $[\frac{1}{2}, \frac{3}{4})$ and all coordinate entries not in $[\frac{1}{2}, \frac{3}{4})$ are in $[0, \frac{1}{2})$. By the combinatorial property established in the beginning of this proof, each point can have at

most one coordinate in $[\frac{1}{2}, \frac{3}{4})$. Hence each point Y_1, Y_2 , and Y_3 has exactly one coordinate entry in $[\frac{1}{2}, \frac{3}{4})$, establishing the first inclusion.

To prove the second one, assume that \mathcal{E} holds. By the combinatorial property and elementary counting, we see that for each $j \in [1..3]$ and $k \in [1..3]$, there is a point Y_i , $i \in [1..3]$, such that $(Y_i)_j \in [\frac{k-1}{4}, \frac{k}{4})$. Hence Y_4 cannot have any coordinate in $[0, \frac{3}{4})$, showing that $Y_4 \in U$.

The crucial observation, following from the stronger net property and leading to an easy computation of $\mathbb{P}(Y_1, Y_2, Y_3 \in D \mid Y_4 \in U)$, is that the second inclusion in (14) is in fact an equality, that is, we have

$$\mathcal{E} = \{Y_4 \in U\}. \quad (15)$$

To show (15), assume that $Y_4 \in U$. Since the canonical box $B = [\frac{1}{2}, 1) \times [\frac{1}{2}, 1) \times [0, 1)$ contains Y_4 , none of the other three points can be contained in B . In particular, none of these three points can have both the first two coordinate entries in $[\frac{1}{2}, \frac{3}{4})$. The same argument applied to the canonical boxes $[\frac{1}{2}, 1) \times [0, 1) \times [\frac{1}{2}, 1)$ and $[0, 1) \times [\frac{1}{2}, 1) \times [\frac{1}{2}, 1)$ shows that each of Y_1, Y_2, Y_3 can have at most one coordinate entry in $[\frac{1}{2}, \frac{3}{4})$, and by the combinatorial property this becomes exactly one coordinate entry in $[\frac{1}{2}, \frac{3}{4})$. Again by the combinatorial property and $Y_4 \in U$, the other two coordinate entries of each of these points must lie in $[0, \frac{1}{2})$. This establishes \mathcal{E} and thus (15).

Let \mathcal{C} be the set of all $C = (C_1, \dots, C_4)$ such that (i) C_1, \dots, C_4 are basic cubes, (ii) $C_4 = U$, and (iii) the event $\mathcal{E}_C = \{\forall i \in [1..4] : Y_i \in C_i\}$ has positive probability. By (15), the event $\{Y_4 \in U\}$ implies the event \mathcal{E} . Hence also any event \mathcal{E}_C , $C \in \mathcal{C}$ implies \mathcal{E} , and thus

$$\frac{\lambda^d(D \cap C_i)}{\lambda^d(C_i)} = 4\varepsilon. \quad (16)$$

By (14), the definition of abstract scrambled nets, and (16), we have

$$\begin{aligned} \mathbb{P}(Y_1, Y_2, Y_3 \in D) &= \mathbb{P}(\{Y_1, Y_2, Y_3 \in D\} \cap \{Y_4 \in U\}) \\ &= \sum_{(C_1, \dots, C_4) \in \mathcal{C}} \mathbb{P}(\{Y_1, Y_2, Y_3 \in D\} \cap \mathcal{E}_C) \\ &= \sum_{(C_1, \dots, C_4) \in \mathcal{C}} \mathbb{P}(Y_1, Y_2, Y_3 \in D \mid \mathcal{E}_C) \mathbb{P}(\mathcal{E}_C) \\ &= \sum_{(C_1, \dots, C_4) \in \mathcal{C}} \prod_{i=1}^3 \frac{\lambda^d(D \cap C_i)}{\lambda^d(C_i)} \cdot \mathbb{P}(\mathcal{E}_C) \\ &= \sum_{(C_1, \dots, C_4) \in \mathcal{C}} (4\varepsilon)^3 \mathbb{P}(\mathcal{E}_C) \\ &= (4\varepsilon)^3 \mathbb{P}(Y_4 \in U) = (4\varepsilon)^3 \frac{1}{64} = \varepsilon^3, \end{aligned}$$

where in the last line we obtained $\mathbb{P}(Y_4 \in U) = \frac{1}{64}$ from the fact that every point in an abstract scrambled net is uniformly distributed in $[0, 1)^d$.

We easily compute

$$\lambda^3(D) = \left(\frac{1}{2} + \varepsilon\right)^3 - \left(\frac{1}{2}\right)^3 = \varepsilon \left(\frac{3}{4} + \frac{3}{2}\varepsilon + \varepsilon^2\right) = \frac{3}{4}\varepsilon + \Theta(\varepsilon^2).$$

We thus have $\mathbb{P}(Y_i \in D) = \lambda^3(D) = \varepsilon \left(\frac{3}{4} + \Theta(\varepsilon)\right)$ for all $i \in [1..3]$. Consequently, for $\varepsilon \searrow 0$ we have

$$\frac{\mathbb{P}(Y_1, Y_2, Y_3 \in D)}{\prod_{i=1}^3 \mathbb{P}(Y_i \in D)} = \left(\frac{4}{3}\right)^3 \cdot \left(\frac{1}{1 + \Theta(\varepsilon)}\right)^3 \nearrow \left(\frac{4}{3}\right)^3 = 2.\overline{370}.$$

□

Remark 3.10. As a comparison, we want to study the setting that leads to the unfavorable lower bound of the correlation constant for scrambled nets in Proposition 3.9 now for Latin hypercube samples $(X_i)_{i=1}^4$ in $[0, 1]^3$. We will observe that in this setting indeed the stricter dependencies of the scrambled net lead to a larger lower bound on the \mathcal{D}_0^3 -correlation number. We use the same notation as in Proposition 3.9 and the corresponding proof.

As in (14), we have

$$\{X_1, X_2, X_3 \in D\} \subseteq \mathcal{E} \subseteq \{X_4 \in U\}.$$

This gives us, in particular,

$$\mathbb{P}(X_1, X_2, X_3 \in D) = \mathbb{P}(X_1, X_2, X_3 \in D \mid X_4 \in U) \cdot \mathbb{P}(X_4 \in U).$$

We also have again $\mathbb{P}(X_4 \in U) = \frac{1}{64}$. What is different is that $X_4 \in U$ does not imply \mathcal{E} . In fact, if we have $X_4 \in U$, then, apart from scaling, X_1, X_2, X_3 form a Latin hypercube sample in $[0, \frac{3}{4}]^3$, and as computed in the proof of Proposition 2.3, the event \mathcal{E} occurs with probability $\frac{3!}{3^3} = \frac{2}{9}$. Consequently,

$$\begin{aligned} \mathbb{P}(X_1, X_2, X_3 \in D \mid X_4 \in U) &= \mathbb{P}(X_1, X_2, X_3 \in D \mid \mathcal{E}) \cdot \mathbb{P}(\mathcal{E} \mid X_4 \in U) \\ &= \frac{2}{9} \mathbb{P}(X_1, X_2, X_3 \in D \mid \mathcal{E}) = \frac{2}{9}(4\varepsilon)^3. \end{aligned}$$

We thus have $\mathbb{P}(X_1, X_2, X_3 \in D) = \frac{2}{9}\varepsilon^3$. If we let ε tend to zero, we obtain

$$\frac{\mathbb{P}(X_1, X_2, X_3 \in D)}{\prod_{i=1}^3 \mathbb{P}(X_i \in D)} = \frac{2}{9} \cdot \left(\frac{4}{3}\right)^3 \cdot \left(\frac{1}{1 + \Theta(\varepsilon)}\right)^3 \nearrow \frac{2}{9} \cdot \left(\frac{4}{3}\right)^3 = 0.5267\dots$$

Acknowledgment

This work was initiated at the Dagstuhl Seminar 19341, “Algorithms and Complexity for Continuous Problems”. Part of this work was done while Michael Gnewuch was a “Chercheur Invité” of École Polytechnique. He thanks his colleagues at the LIX for their hospitality.

References

- [1] H. W. BLOCK, T. H. SAVITS, AND M. SHAKED, *Some concepts of negative dependence*, Ann. Probab., 10 (1982).
- [2] J. DICK AND F. PILLICHSHAMMER, *Digital Nets and Sequences*, Cambridge University Press, Cambridge, 2010.
- [3] B. DOERR, *Probabilistic tools for the analysis of randomized optimization heuristics*, in Theory of Evolutionary Computation: Recent Developments in Discrete Optimization, B. Doerr and F. Neumann, eds., Springer, 2020, pp. 1–87. Also available at <https://arxiv.org/abs/1801.06733>.
- [4] B. DOERR, C. DOERR, AND M. GNEWUCH, *Probabilistic lower discrepancy bounds for Latin hypercube samples*, in Contemporary Computational Mathematics – a Celebration of the 80th Birthday of Ian Sloan, J. Dick, F. Y. Kuo, and H. Woźniakowski, eds., Springer-Verlag, 2018, pp. 339–350.
- [5] H. FAURE, *Discrépance de suites associées à un système de numération (en dimension s)*, Acta Arith., 41 (1982), pp. 338–351.
- [6] M. GNEWUCH AND N. HEBBINGHAUS, *Discrepancy bounds for a class of negatively dependent random points including Latin hypercube samples*, 2019. Preprint, arXiv:2102.04451 (to appear in Annals of Applied Probability).
- [7] M. GNEWUCH, H. PASING, AND C. WEISS, *A generalized Faulhaber inequality, improved bracketing covers, and applications to discrepancy*, 2020. Preprint, arXiv:2010.11479 (to appear in Mathematics of Computation).
- [8] P. L’ECUYER AND C. LEMIEUX, *Recent advances in randomized quasi-Monte Carlo methods*, in Modeling uncertainty, vol. 46 of Internat. Ser. Oper. Res. Management Sci., Kluwer Acad. Publ., Boston, MA, 2002, pp. 419–474.
- [9] C. LEMIEUX, *Negative dependence, scrambled nets, and variance bounds*, Math. Oper. Res., 43 (2018), pp. 228–251.
- [10] J. MATOUŠEK, *On the L_2 -discrepancy for anchored boxes*, J. Complexity, 14 (1998), pp. 527–556.
- [11] ———, *Geometric Discrepancy*, Springer-Verlag Berlin Heidelberg, 2nd ed., 2010.
- [12] M. D. MCKAY, R. J. BECKMAN, AND W. J. CONOVER, *A comparison of three methods for selecting values of input variables in the analysis of output from a computer code*, Technometrics, 21 (1979), pp. 239–245.
- [13] H. NIEDERREITER, *Point sets and sequences with small discrepancy*, Monatsh. Math., 104 (1987), pp. 273–337.
- [14] ———, *Random Number Generation and Quasi-Monte Carlo Methods*, vol. 63 of CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1992.
- [15] A. B. OWEN, *Randomly permuted (t, m, s) -nets and (t, s) -sequences*, in Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, H. Niederreiter and P. J.-S. Shiue, eds., New York, 1995, Springer, pp. 299–317.
- [16] ———, *Monte Carlo variance of scrambled net quadrature*, SIAM J. Numer. Anal., 34 (1997).
- [17] ———, *Variance with alternative scramblings of digital nets*, ACM Trans. Model. Comput. Simul., 13 (2003), pp. 363–378.
- [18] A. PANCONESI AND A. SRINIVASAN, *Randomized distributed edge coloring via an extension of the Chernoff-Hoeffding bounds*, SIAM J. Comput., 26 (1997), pp. 350–368.
- [19] ———, 2020. Personal communication.

- [20] H. ROBBINS, *A remark on Stirling's formula*, The American Mathematical Monthly, 62 (1955), pp. 26–29.
- [21] I. M. SOBOL', *The distribution of points in a cube and the approximate evaluation of integrals*, Zh. Vychisl. Mat. i. Mat. Fiz., 7 (1967), pp. 784–802.
- [22] J. WIART, C. LEMIEUX, AND G. Y. DONG, *On the dependence structure of scrambled (t, m, s) -nets*, 2020. Preprint, arXiv:1903.09877v5.
- [23] M. WNUK, M. GNEWUCH, AND N. HEBBINGHAUS, *On negatively dependent sampling schemes, variance reduction, and probabilistic upper discrepancy bounds*, in Discrepancy Theory, D. Bylik, J. Dick, and F. Pillichshammer, eds., Radon Series on Computational and Applied Mathematics 26, DeGruyter, 2020, pp. 43–68. arXiv:1904.10796.