Covering numbers, dyadic chaining and discrepancy

Christoph Aistleitner*

Abstract

In 2001 Heinrich, Novak, Wasilkowski and Woźniakowski proved that for every $s \geq 1$ and $N \geq 1$ there exists a sequence $(z_1, \ldots, z_N)$ of elements of the s-dimensional unit cube such that the star-discrepancy D_N^* of this sequence satisfies

$$D_N^*(z_1, \ldots, z_N) \leq c \frac{\sqrt{s}}{\sqrt{N}}$$

for some constant c independent of s and N. Their proof uses deep results from probability theory and combinatorics, and does not provide a concrete value for the constant c.

In this paper we give a new simple proof of this result, and show that we can choose $c = 10$. Our proof combines Gnewuch's upper bound for covering numbers, Bernstein's inequality and a dyadic partitioning technique.

1 Introduction and statement of results

Let $(z_1, \ldots, z_N)$ be a sequence of elements of the s-dimensional unit cube. The number $D_N^*(z_1, \ldots, z_N)$, which is defined as

$$D_N^*(z_1, \ldots, z_N) = \sup_{x \in [0,1]^s} \left| \lambda([0,x]) - \frac{\sum_{n=1}^N \mathbb{1}_{[0,x]}(z_n)}{N} \right|,$$

is called the star-discrepancy of $(z_1, \ldots, z_N)$. Here and in the sequel λ denotes the s-dimensional Lebesgue measure. The Koksma-Hlawka inequality states that the difference between the integral of a function f over the s-dimensional unit cube and the arithmetic mean of the function values $f(z_1), \ldots, f(z_N)$ is bounded by the product of the total variation of f (in the sense of Hardy and Krause) and the star-discrepancy $D_N^*(z_1, \ldots, z_N)$ of the sequence of sampling points $(z_1, \ldots, z_N)$. This means that sequences having small star discrepancy are useful for numerical integration. For general information on this topic we refer to [3], [6] and [14].

*Graz University of Technology, Institute of Mathematics A, Steyrergasse 30, 8010 Graz, Austria. e-mail: aistleitner@math.tugraz.at. Research supported by the Austrian Research Foundation (FWF), Project S9603-N23. This paper was written during a stay of the author at the Hausdorff Institute of Mathematics in Bonn, Germany, in the framework of the "HIM Junior Trimester Program on Stochastics". The author wants to thank the HIM for its hospitality.

Mathematics Subject Classification: 11K38, 65C05, 65D32

Sequences having small discrepancy are particularly important for the evaluation of high-dimensional integrals, which appear e.g. in financial mathematics. There exist several constructions of so-called low discrepancy sequence, i.e. sequences satisfying

$$D_N^*(z_1,\dots,z_N) \le c(\log N)^{s-1}N^{-1},$$

but these constructions are only useful if N is large compared to s. For the construction of such low-discrepancy sequences we refer to [4] and [15].

Therefore it is desirable to have sequences which have small star-discrepancy for small values of N (in comparison to s). This can be formulated in terms of the inverse of the star-discrepancy: let $N^*(s,\varepsilon)$ denote the minimal number of points with star-discrepancy at most ε. Heinrich, Novak, Wasilkowski and Woźniakowski [12] showed that

$$N^*(s,\varepsilon) = \mathcal{O}(s\varepsilon^{-2}), \tag{1}$$

where the value of the implicit constant is unknown. The dependence on the dimension is best possible in (1). In [12] the lower bound $N^*(s,\varepsilon) \gg s\log\varepsilon^{-1}$ was shown, which was improved by Hinrichs [13] to $N^*(s,\varepsilon) \gg s\varepsilon^{-1}$.

In particular (1) implies for any $s,N\ge 1$ the existence of a sequence $(z_1,\dots,z_N)$ of elements of the s-dimensional unit cube such that

$$D_N^*(z_1,\dots,z_N) \le \frac{c\sqrt{s}}{\sqrt{N}}, \tag{2}$$

where c is an absolute (but unknown) constant. The proof of (1) uses deep results from probability theory and combinatorics, namely Talagrand's maximal inequality for empirical processes [17] and Haussler's upper bound for covering numbers of Vapnik-Červonenkis classes [11], and does not seem to allow a direct calculation of c in a reasonable way. The best known results with explicit constant are typically of the form

$$D_N^*(z_1,\dots,z_N) \le \frac{c\sqrt{s\log N}}{\sqrt{N}}, \tag{3}$$

where the additional log-term essentially comes from the fact that the discrepancy D_N^* has to be discretized with respect to $\approx N^{s/2}$ sampling points (cf. [5, Theorem 3.2], [8, Theorem 2.1] and [12, Theorem 1]).[1]

In this paper we want to present a new proof of (2), which allows a simple calculation of the constant c. Our proof combines a result of Gnewuch on covering numbers (see Lemma 1 below), a standard inequality from probability theory and a dyadic partitioning technique, which is inspired by a somewhat similar technique ("dyadic chaining") which is commonly used in probabilistic discrepancy theory (cf. e.g. [1], [2] or [16]).

Basically, our proof is based on the following observations: in the known proofs, the star-discrepancy D_N^* was discretized using $\approx N^{s/2}$ sampling points, and this number of sampling

[1] In a footnote in [8] Gnewuch mentions a proof of Hinrichs for $c = 10$ in (2), presented in a seminar talk. Apparently this proof is unpublished.

points is (roughly speaking) really necessary. Thus we have a set $\gamma_1, \dots, \gamma_M$ of sampling points, and calculate, for $X_1, \dots, X_N$ being i.i.d. uniformly distributed on $[0,1]^s$, the probabilities

$$\mathbb{P}\left(\left|N\lambda([0,\gamma_i]) - \sum_{n=1}^{N} \mathbb{1}_{[0,\gamma_i]}(X_n)\right| > t\right). \tag{4}$$

If t is chosen in such a way that the sum of the probabilities in (4) for $i = 1, \dots, M$ is less than 1, this implies the existence of a realization $X_1(\omega), \dots, X_N(\omega)$ for which

$$\left|N\lambda([0,\gamma_i]) - \sum_{n=1}^{N} \mathbb{1}_{[0,\gamma_i]}(X_n(\omega))\right| \leq t \tag{5}$$

for all $1 \leq i \leq M$. Together with a (small) discretization error this gives an upper bound for the star-discrepancy.

To estimate the probabilities in (5), all proofs of results of the form (3) use Hoeffding's inequality, which states that for centered random variables with $a \leq |X_n| \leq b$ a.s., where $b - a \leq 1$, the upper bound

$$\mathbb{P}\left(\left|\sum_{n=1}^{N} X_n\right| > t\right) \leq 2e^{-2t^2/N} \tag{6}$$

holds. In our proof we use Bernstein's inequality instead of (6), which yields (under the same assumptions)

$$\mathbb{P}\left(\left|\sum_{n=1}^{N} X_n\right| > t\right) \leq 2\exp\left(-\frac{t^2/2}{\sum_{n=1}^{N} \mathbb{E}X_n^2 + t/3}\right). \tag{7}$$

For random variables with small variance the bound in (7) is in many cases much stronger than the bound in (6). Next we observe that we can write any arbitrary indicator function $\mathbb{1}_{[0,x]}$ as a sum of one indicator function of a set of Lebesgue measure $\leq 1/2$, one indicator of a set of measure $\leq 1/4$, one indicator of a set of measure $\leq 1/8$, etc. To be able to represent any interval $[0,x]$ we only need a small number of indicators of sets of measure $\approx 1/2$, some more of measure $\approx 1/4$, etc., but since the variance of

$$\sum_{n=1}^{N} \mathbb{1}_I(X_n) \tag{8}$$

is in direct accordance with the Lebesgue measure of the set I, we only need few random variables of the form (8) with large variance, and many with small variance to be able to approximate the sum

$$\sum_{n=1}^{N} \mathbb{1}_{[0,x]}(X_n)$$

for arbitrary x. Using Bernstein's inequality to estimate the probabilities of the form (4), we can get rid of the log-factor in (3).

Our main result is the following

Theorem 1 *For any $s \geq 1$ and $N \geq 1$ there exists a sequence $(z_1, \ldots, z_N)$ of elements of the s-dimensional unit cube such that*

$$D_N^*(z_1, \ldots, z_N) \leq \frac{10\sqrt{s}}{\sqrt{N}}.$$

Throughout this paper s will be a positive integer denoting the dimension. It is an easy exercise to prove the theorem for $s = 1$ and $s = 2$, so we will assume throughout the rest of this paper that $s \geq 3$. For $x, y \in [0,1]^s$, where $x = (x_1, \ldots, x_s)$ and $y = (y_1, \ldots, y_s)$, we write $x \leq y$ if $x_j \leq y_j,\ 1 \leq j \leq s$. We write 0 for the s-dimensional vector $(0, \ldots, 0)$ and 1 for $(1, \ldots, 1)$.

We will use the following Lemma 1, which is a result of Gnewuch [8, Theorem 1.15]. For the formulation we use the notation from [8] and [10]: For any $\delta \in (0,1]$ a finite set Γ of points in $[0,1]^s$ is called a δ-cover of $[0,1]^s$ if for every $y \in [0,1]^s$ there exist $x, z \in \Gamma \cup \{0\}$ such that $x \leq y \leq z$ and $\lambda([0,z)) - \lambda([0,x)) \leq \delta$. The number $\mathcal{N}(s, \delta)$ denotes the smallest cardinality of a δ-cover of $[0,1]^s$, i.e.

$$\mathcal{N}(s, \delta) = \min \left\{ |\Gamma| :\ \Gamma \text{ is a } \delta\text{-cover of } \ [0,1]^s \right\}.$$

Similarly, for any $\delta \in (0,1]$ a finite set Δ of pairs of points from $[0,1]^s$ is called a δ-bracketing cover of $[0,1]^s$, if for every pair $(x,z) \in \Delta$ the estimate $\lambda([0,z)) - \lambda([0,x)) \leq \delta$ holds, and if for every $y \in [0,1]^s$ there exists a pair (x,z) from Δ such that $x \leq y \leq z$. The number $\mathcal{N}_{[\ ]}(s, \delta)$ denotes the smallest cardinality of a δ-bracketing cover of $[0,1]^s$.

Lemma 1 *For any $s \geq 1$ and $\delta \in [0,1)$*

$$\mathcal{N}(s, \delta) \leq (2e)^s (\delta^{-1} + 1)^s$$

and

$$\mathcal{N}_{[\ ]}(s, \delta) \leq 2^{s-1} e^s (\delta^{-1} + 1)^s$$

There are several possibilities for an improvement of the constant in Theorem 1. We mention the following:

- The estimates in our proof are not everywhere best possible. It is very involved to adjust the diverse constants and probabilities to each other, and a more pedantic approach should result in an improvement of the final constant.
- In [9] Gnewuch conjectured that the upper bounds for $\mathcal{N}(s, \delta)$ and $\mathcal{N}_{[\ ]}(s, \delta)$ in Lemma 1 can be replaced by $2\delta^{-s} + o_s(\delta^{-s})$ and $\delta^{-s} + o_s(\delta^{-s})$, resp. (where o_d means that the implied constant may depend on d). An improvement of Lemma 1 could result in a significantly smaller value of the constant in Theorem 1.
- The theorem states that, with a certain positive probability, a randomly generated sequence, i.e. a so-called Monte Carlo sequence, is an appropriate choice for a sequence with small discrepancy (and may therefore be used in Quasi-Monte Carlo integration). This is a somewhat odd result, and one might expect that an appropriately designed "real" Quasi-Monte Carlo sequence should have a smaller star-discrepancy than a completely randomly generated sequence. For example, it might be reasonable to choose the

random variables $X_1, \dots, X_N$ not i.i.d. uniformly from $[0,1]^s$, but, e.g., independently and uniformly from disjoint sets $A_1, \dots, A_N$, whose union is the unit cube (this strategy is called "stratified sampling", cf. [7, Section 4.3]). This results in smaller variances for the sums

$$\sum_{n=1}^{N} \mathbb{1}_I(X_n),$$

and maybe a thought-out partition $A_1, \dots, A_N$ of $[0,1]^s$ yields a significant improvement of Theorem 1.

2 Proof of Theorem 1

For $N \leq 100s$ our theorem is trivial, so we will assume in the sequel that $N > 100s$. Set $K = \lceil (\log_2 N - \log_2 s)/2 \rceil$. Then $K \geq 4$, and

$$2^{-K} \in \left[\frac{\sqrt{s}}{2\sqrt{N}}, \frac{\sqrt{s}}{\sqrt{N}} \right]. \tag{9}$$

For $1 \leq k \leq K-1$ let Γ_k denote a 2^{-k}-cover of $[0,1]^s$, for which

$$|\Gamma_k| \leq (2e)^s (2^k + 1)^s. \tag{10}$$

Such a Γ_k exists for every $k \geq 1$ by Lemma 1.
Similarly, let Δ_K denote a 2^{-K}-bracketing cover for which

$$|\Delta_K| \leq 2^{s-1} e^s (2^K + 1)^s,$$

which also exists by Lemma 1. For notational convenience we also define

$$\Gamma_K = \{ v \in [0,1]^s : \ (v, w) \in \Delta_K \text{ for some } w \}.$$

For every $x \in [0,1]^s$ there exists a pair $(v_K, w_K) = (v_K(x), w_K(x))$ for which $(v_K, w_K) \in \Delta_K$ such that $v_K \leq x \leq w_K$ and

$$\lambda([0, w_K]) - \lambda([0, v_K]) \leq \frac{1}{2^K}.$$

For every $k,\ 2 \leq k \leq K$ and $\gamma \in \Gamma_k$ there exist $v_{k-1} = v_{k-1}(\gamma), w_{k-1} = w_{k-1}(\gamma),\ v_{k-1}, w_{k-1} \in \Gamma_{k-1} \cup \{0\}$, such that $v_{k-1} \leq \gamma \leq w_{k-1}$ and

$$\lambda([0, w_{k-1}]) - \lambda([0, v_{k-1}]) \leq \frac{1}{2^{k-1}}$$

Recursively we define

$$\begin{aligned}
p_K(x) &= v_K(x) \\
p_{K-1}(x) &= v_{K-1}(p_K(x)) = v_{K-1}(v_K(x)) \\
p_{K-2}(x) &= v_{K-2}(p_{K-1}(x)) = v_{K-2}(v_{K-1}(v_K(x))) \\
&\vdots \\
p_1(x) &= v_1(p_2(x)),
\end{aligned}$$

and, for notational convenience,

$$p_0 = 0.$$

We define for $x, y \in [0,1]^s$

$$\overline{[x,y]} := \begin{cases} [0,y]\backslash[0,x] & \text{if } x \neq 0 \\ [0,y] & \text{if } x = 0,\ y \neq 0. \\ \emptyset & \text{if } x = y = 0. \end{cases}$$

Then the sets

$$\overline{[p_k(x), p_{k+1}(x)]}, \quad 1 \leq k \leq K,$$

are disjoint, we have

$$\bigcup_{k=0}^{K-1} \overline{[p_k(x), p_{k+1}(x)]} \subset [0,x] \subset \bigcup_{k=0}^{K-1} \overline{[p_k(x), p_{k+1}(x)]} \cup \overline{[p_K(x), w_K(x)]},$$

and, accordingly, for every $x, y \in [0,1]^s$

$$\sum_{k=0}^{K-1} \mathbb{1}_{\overline{[p_k(x),p_{k+1}(x)]}}(y) \leq \mathbb{1}_{[0,x]}(y) \leq \sum_{k=0}^{K-1} \mathbb{1}_{\overline{[p_k(x),p_{k+1}(x)]}}(y) \ + \ \mathbb{1}_{\overline{[p_K(x),w_K(x)]}}(y). \tag{11}$$

Independent of x we have for $0 \leq k \leq K-1$

$$\lambda\left(\overline{[p_k(x), p_{k+1}(x)]}\right) \ \leq \ \frac{1}{2^k},$$

and

$$\lambda\left(\overline{[p_K(x), w_K(x)]}\right) \leq \frac{1}{2^K}.$$

For $0 \leq k \leq K-1$ we write A_k for the set of all sets of the form

$$\overline{[p_k(x), p_{k+1}(x)]},$$

where x can take any possible value from $[0,1]^s$. Then for $0 \leq k \leq K-1$, by (10), A_k contains at most

$$|\Gamma_{k+1}| \leq (2e)^s \left(2^{k+1}+1\right)^s.$$

elements. We write A_K for the set of all the sets of the form

$$\overline{[p_K(x), w_K(x)]},$$

where $x \in [0,1]^s$. Then A_K contains at most

$$|\Delta_K| \leq 2^{s-1} e^s \left(2^K + 1\right)^s$$

elements. We repeat that all elements of A_k, where $0 \leq k \leq K$, have Lebesgue measure bounded by 2^{-k}.

Let $X_1, \ldots, X_N$ be i.i.d. random variables defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ having uniform distribution on $[0,1]^s$.

Let $I \in A_k$ for some $k \geq 0$. Then the random variables $\mathbb{1}_I(X_1), \dots, \mathbb{1}_I(X_N)$ are i.i.d. random variables, having expected value

$$\lambda(I)$$

and variance

$$\lambda(I) - \lambda(I)^2 \leq \begin{cases} 2^{-k}(1-2^{-k}) & \text{for } k \geq 1 \\ 1/4 & \text{for } k = 0 \end{cases} \tag{12}$$

Thus, since the X_n are independent,

$$\sum_{n=1}^{N} \mathbb{1}_I(X_n)$$

has expected value $N\lambda(I)$ and variance $N(\lambda(I) - \lambda(I)^2)$.

Bernstein's inequality states that for $Z_1, \dots, Z_N$ being i.i.d. random variables, satisfying $\mathbb{E}Z_n = 0$ and $|Z_n| \leq C$ a.s. for some $C > 0$,

$$\mathbb{P}\left(\left| \sum_{n=1}^{N} Z_n \right| > t \right) \leq 2 \exp\left(- \frac{t^2}{2\left(\sum_{n=1}^{N} \mathbb{E}Z_n^2 \right) + 2Ct/3} \right).$$

Applying this inequality to the random variables $\mathbb{1}_I(X_n) - \lambda(I)$, we obtain

$$\mathbb{P}\left(\left| \sum_{n=1}^{N} \mathbb{1}_I(X_n) - N\lambda(I) \right| > t \right) \leq 2 \exp\left(- \frac{t^2}{2\left(N\lambda(I)\left(1-\lambda(I)\right)\right) + 2t/3} \right)$$

for $t > 0$. Using (12) we conclude

$$\mathbb{P}\left(\left| \sum_{n=1}^{N} \mathbb{1}_I(X_n) - N\lambda(I) \right| > t \right) \leq 2 \exp\left(- \frac{t^2}{2N2^{-k}(1-2^{-k}) + 2t/3} \right) \qquad \text{for } k \geq 2.$$

For $k = 0$ and $k = 1$ it is better to use Hoeffding's inequality, which yields

$$\mathbb{P}\left(\left| \sum_{n=1}^{N} \mathbb{1}_I(X_n) - N\lambda(I) \right| > t \right) \leq \exp\left(- \frac{2t^2}{N} \right) \qquad \text{for } k \geq 0, 1.$$

By (9) we have

$$\sqrt{sN} = N \frac{\sqrt{s}}{\sqrt{N}} \leq 2^{-K+1} N.$$

Therefore, choosing $t = c\sqrt{sN}$ for some $c > 0$, we obtain

$$\mathbb{P}\left(\left| \sum_{n=1}^{N} \mathbb{1}_I(X_n) - N\lambda(I) \right| > t \right) \leq \begin{cases} 2 \exp\left(- \frac{c^2 s}{2 \cdot 2^{-k}(1-2^{-k}) + 4c2^{-K}/3} \right) & \text{for } 2 \leq k \leq K \\ 2e^{-2c^2 s} & \text{for } k = 0, 1. \end{cases} \tag{13}$$

The set A_0 contains at most

$$(6e)^s$$

elements. Choosing

$$c_0 = \sqrt{\frac{1+\log 12}{2}} \leq 1.33$$

and using (13) with $t = c_0\sqrt{sN}$ we get

$$\mathbb{P}\left(\left|\sum_{n=1}^{N} \mathbb{1}_I(X_n) - N\lambda(I)\right| > c_0\sqrt{sN}\right) \leq 2e^{-(1+\log 12)s}$$

for every $I \in A_0$. Thus, writing

$$B_0 = \bigcup_{I \in A_0} \left(\left|\sum_{n=1}^{N} \mathbb{1}_I(X_n) - N\lambda(I)\right| > c_0\sqrt{sN}\right)$$

we have

$$\mathbb{P}(B_0) \leq 2(6e)^s e^{-(1+\log 12)s} \leq (2^{1/3} \cdot 6e)^s e^{-(1+\log 12)s} \leq 2^{-2s/3} \leq \frac{1}{4}$$

(remember that we have assumed $s \geq 3$).

A_1 contains at most

$$(10e)^s$$

elements. For

$$c_1 = \sqrt{\frac{1+\log 20}{2}} \leq 1.42,$$

choosing $t = c_1\sqrt{sN}$ we get

$$\mathbb{P}\left(\left|\sum_{n=1}^{N} \mathbb{1}_I(X_n) - N\lambda(I)\right| > c_1\sqrt{sN}\right) \leq 2e^{-(1+\log 20)s}$$

for any $I \in A_1$. Thus, writing

$$B_1 = \bigcup_{I \in A_1} \left(\left|\sum_{n=1}^{N} \mathbb{1}_I(X_n) - N\lambda(I)\right| > c_1\sqrt{sN}\right)$$

we obtain

$$\mathbb{P}(B_1) \leq 2(10e)^s e^{-(1+\log 20)s} \leq (2^{1/3} \cdot 10e)^s e^{-(1+\log 20)s} \leq 2^{-2s/3} \leq \frac{1}{4}.$$

For $2 \leq k \leq K-1$ we have

$$\begin{aligned} 2 \cdot (2e)^s (2^{k+1}+1)^s 2^k &\leq (2^{1/3} \cdot 2e)^s (2^{k+1}+1)^s 2^k \\ &\leq e^{s\left(1+\log 2+(\log 2)/3+k\log 2+\log 2+2^{-k-1}\right)+k\log 2} \\ &\leq e^{ks(0.93+2.75/k)}. \end{aligned} \tag{14}$$

For

$$c_k = \sqrt{k}\sqrt{0.93+2.75/k}\sqrt{2 \cdot 2^{-k}(1-2^{-k}) + 1.53 \cdot 4 \cdot 2^{-K}/3}$$

we have $|c_k| < 1.53$, and

$$\begin{aligned}
\frac{c_k^2}{2 \cdot 2^{-k}(1-2^{-k}) + 4c_k 2^{-K}/3} &\geq \frac{c_k^2}{2 \cdot 2^{-k}(1-2^{-k}) + 1.53 \cdot 4 \cdot 2^{-K}/3} \\
&\geq k(0.93 + 2.75/k). \qquad (15)
\end{aligned}$$

Thus we choose

$$t = c_k \sqrt{sN}$$

and get from (13) and (15)

$$\mathbb{P}\left(\left| \sum_{n=1}^{N} \mathbb{1}_I(X_n) - N\lambda(I) \right| > c_k \sqrt{sN} \right) \leq 2e^{-ks(0.93+2.75/k)}.$$

for any $I \in A_k$. Thus, writing

$$B_k = \bigcup_{I \in A_k} \left(\left| \sum_{n=1}^{N} \mathbb{1}_I(X_n) - N\lambda(I) \right| > c_k \sqrt{sN} \right),$$

we have by (14)

$$\begin{aligned}
\mathbb{P}(B_k) &\leq (2e)^s (2^{k+1}+1)^s \cdot 2e^{-ks(0.93+2.75/k)} \\
&\leq 2(2e)^s (2^{k+1}+1)^s e^{-ks(0.93+2.75/k)} \\
&\leq 2^{-k}.
\end{aligned}$$

Finally, A_K contains at most

$$2^{s-1} e^s (2^K + 1)^s$$

elements, and

$$2 \cdot 2^{s-1} e^s (2^K+1)^s 2^K \leq e^{Ks(0.93+1.76/K)} \qquad (16)$$

(where we used $K \geq 4$ and $s \geq 3$). For

$$c_K = \sqrt{K}\sqrt{0.93+1.76/K}\sqrt{2 \cdot 2^{-K}(1-2^{-K}) + 1.07 \cdot 4 \cdot 2^{-K}/3}$$

we have $|c_K| \leq 1.07$, and

$$\begin{aligned}
\frac{c_K^2}{2 \cdot 2^{-K}(1-2^{-K}) + 4c_K 2^{-K}/3} &\geq \frac{c_K^2}{2 \cdot 2^{-K}(1-2^{-K}) + 1.07 \cdot 4 \cdot 2^{-K}/3} \\
&\geq K(0.93 + 1.76/K). \qquad (17)
\end{aligned}$$

Combining (13), (16) and (17) we obtain

$$\mathbb{P}(B_K) := \mathbb{P}\left(\bigcup_{I \in A_K} \left(\left| \sum_{n=1}^{N} \mathbb{1}_I(X_n) - N\lambda(I) \right| > c_K \sqrt{sN} \right) \right) \leq 2^{-K}.$$

Combining our estimates we have

$$\sum_{k=0}^{K} \mathbb{P}(B_k) \leq \frac{3}{4} + \sum_{k=3}^{K} 2^{-k} < 1.$$

Therefore there exists at least one realization $X_1(\omega), \dots, X_N(\omega)$, such that

$$\omega \notin \bigcup_{k=0}^{K} B_k.$$

We define

$$z_n = X_n(\omega), \quad 1 \le n \le N.$$

Some calculations show that

$$\sum_{k=0}^{K} c_k \le 8.65.$$

Then, by (11), for all $x \in [0,1]$

$$\begin{aligned}
\sum_{n=1}^{N} \mathbb{1}_{[0,x]}(z_n) &\le \sum_{k=0}^{K-1} \sum_{n=1}^{N} \mathbb{1}_{\overline{[p_k(x),p_{k+1}(x)]}}(z_n) + \sum_{n=1}^{N} \mathbb{1}_{\overline{[p_K(x),w_K(x)]}}(z_n) \\
&\le N\lambda([0,w_K(x)]) + \sqrt{sN} \sum_{k=0}^{K} c_k \\
&\le N\lambda([0,x]) + N\lambda(\overline{[x,w_K(x)]}) + 8.65\sqrt{sN} \\
&\le N\lambda([0,x]) + N\frac{\sqrt{s}}{\sqrt{N}} + 8.65\sqrt{sN} \\
&\le N\lambda([0,x]) + 9.65\sqrt{sN}.
\end{aligned}$$

Similarly

$$\begin{aligned}
\sum_{n=1}^{N} \mathbb{1}_{[0,x]}(z_n) &\ge \sum_{k=0}^{K-1} \sum_{n=1}^{N} \mathbb{1}_{\overline{[p_k(x),p_{k+1}(x)]}}(z_n) \\
&\ge N\lambda([0,p_K(x)]) - \sqrt{sN} \sum_{k=0}^{K-1} c_k \\
&\ge N\lambda([0,x]) - N\lambda\left(\overline{[p_K(x),x]}\right) - 8.65\sqrt{sN} \\
&\ge N\lambda([0,x]) - 9.65\sqrt{sN}.
\end{aligned}$$

Since this applies for arbitrary $x \in [0,1]^s$, we conclude

$$D_N^*(z_1, \dots, z_N) \le 10\frac{\sqrt{s}}{\sqrt{N}},$$

which proves Theorem 1.

Acknowledgements: The author wants to thank Michael Gnewuch for several helpful suggestions. He also wants to thank two anonymous referees, whose comments led to an improvement of the presentation of this paper.

References

[1] C. Aistleitner. On the law of the iterated logarithm for the discrepancy of sequences $\langle n_{\mathbf{k}} x \rangle$ with multidimensional indices. *Unif. Distrib. Theory*, 2(2):89–104, 2007.

[2] C. Aistleitner and I. Berkes. On the law of the iterated logarithm for the discrepancy of $\langle n_k x \rangle$. *Monatsh. Math.*, 156(2):103–121, 2009.

[3] B. Chazelle. *The discrepancy method.* Cambridge University Press, Cambridge, 2000. Randomness and complexity.

[4] J. Dick and F. Pillichshammer. *Digital nets and sequences. Discrepancy theory and quasi-Monte Carlo integration.* Cambridge: Cambridge University Press. xvii, 600 p., 2010.

[5] B. Doerr, M. Gnewuch, and A. Srivastav. Bounds and constructions for the star-discrepancy via δ-covers. *J. Complexity*, 21(5):691–709, 2005.

[6] M. Drmota and R. F. Tichy. *Sequences, discrepancies and applications*, volume 1651 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin, 1997.

[7] P. Glasserman. *Monte Carlo methods in financial engineering*, volume 53 of *Applications of Mathematics (New York).* Springer-Verlag, New York, 2004. Stochastic Modelling and Applied Probability.

[8] M. Gnewuch. Bracketing numbers for axis-parallel boxes and applications to geometric discrepancy. *J. Complexity*, 24(2):154–172, 2008.

[9] M. Gnewuch. Construction of minimal bracketing covers for rectangles. *Electron. J. Combin.*, 15(1):Research Paper 95, 20, 2008.

[10] M. Gnewuch. On probabilistic results for the discrepancy of a hybrid-Monte Carlo sequence. *J. Complexity*, 25(4):312–317, 2009.

[11] D. Haussler. Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik-Chervonenkis dimension. *J. Combin. Theory Ser. A*, 69(2):217–232, 1995.

[12] S. Heinrich, E. Novak, G. W. Wasilkowski, and H. Woźniakowski. The inverse of the star-discrepancy depends linearly on the dimension. *Acta Arith.*, 96(3):279–302, 2001.

[13] A. Hinrichs. Covering numbers, Vapnik-Červonenkis classes and bounds for the star-discrepancy. *J. Complexity*, 20(4):477–483, 2004.

[14] L. Kuipers and H. Niederreiter. *Uniform distribution of sequences.* Wiley-Interscience [John Wiley & Sons], New York, 1974. Pure and Applied Mathematics.

[15] H. Niederreiter. *Random number generation and quasi-Monte Carlo methods*, volume 63 of *CBMS-NSF Regional Conference Series in Applied Mathematics.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[16] W. Philipp. Limit theorems for lacunary series and uniform distribution mod 1. *Acta Arith.*, 26(3):241–251, 1974/75.

[17] M. Talagrand. Sharper bounds for Gaussian and empirical processes. *Ann. Probab.*, 22(1):28–76, 1994.