Probabilistic star discrepancy bounds for double infinite random matrices

Christoph Aistleitner and Markus Weimar

Abstract In 2001 Heinrich, Novak, Wasilkowski and Woźniakowski proved that the inverse of the discrepancy depends linearly on the dimension, by showing that a Monte Carlo point set $\mathscr{P}$ of N points in the s-dimensional unit cube satisfies the discrepancy bound $D_N^{*s}(\mathscr{P}) \leq c_{\text{abs}} s^{1/2} N^{-1/2}$ with positive probability. Later their results were generalized by Dick to the case of double infinite random matrices. In the present paper we give asymptotically optimal bounds for the discrepancy of such random matrices, and give estimates for the corresponding probabilities. In particular we prove that the $N \times s$-dimensional projections $\mathscr{P}_{N,s}$ of a double infinite random matrix satisfy the discrepancy estimate

$$D_N^{*s}(\mathscr{P}_{N,s}) \leq \left(2130 + 308\,\frac{\ln\ln N}{s}\right)^{1/2} s^{1/2} N^{-1/2}$$

for all N and s with positive probability. This improves the bound $D_N^{*s}(\mathscr{P}_{N,s}) \leq (c_{\text{abs}} \ln N)^{1/2} s^{1/2} N^{-1/2}$ given by Dick. Additionally, we show how our approach can be used to show the existence of completely uniformly distributed sequences of small discrepancy which find applications in Markov Chain Monte Carlo.

Christoph Aistleitner
Graz University of Technology, Institute of Mathematics A, Steyrergasse 30, 8010 Graz, Austria. Research supported by the Austrian Research Foundation (FWF), Project S9603-N23. e-mail: aistleitner@math.tugraz.at,

Markus Weimar
Friedrich-Schiller-University Jena, Institute of Mathematics, Ernst-Abbe-Platz 2, 07743 Jena, Germany. e-mail: markus.weimar@uni-jena.de

1 Introduction and statement of results

1.1 Uniform distribution and discrepancy

Let x, y be two elements of the s-dimensional unit cube $[0,1]^s$. We write $x \leq y$ if this inequality holds coordinatewise, and $x < y$ if all coordinates of x are smaller than the corresponding coordinates of y. Furthermore, $[x,y)$ denotes the set $\{z \in [0,1]^s \,|\, x \leq z < y\}$. We write 0 for the s-dimensional vector $(0,\dots,0)$, and thus $[0,x)$ denotes the set $\{z \in [0,1]^s \,|\, 0 \leq z < x\}$. Throughout the paper we will use the same notation for real numbers and for real vectors; it will be clear from the context what we mean. Moreover, by c_{abs} we will denote universal constants which may change at every occurrence.

A sequence $(x_n)_{n\in\mathbb{N}}$ of points from $[0,1]^s$ is called *uniformly distributed* (modulo 1) if for any $x \in [0,1]^s$ the asymptotic equality

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{[0,x)}(x_n) = \lambda([0,x)) \tag{1}$$

holds. Here $\mathbb{N}$ denotes the set of positive integers, and λ denotes the s-dimensional Lebesgue measure. By an observation of Weyl [23] a sequence is uniformly distributed if and only if

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n) = \int_{[0,1]^s} f(x)\,\mathrm{d}x \tag{2}$$

for any continuous s-dimensional function f. This interrelation already suggests that uniformly distributed sequences can be used for numerical integration — an idea which is the origin of the so-called *Quasi-Monte Carlo* (QMC) *method* for numerical integration. The speed of convergence in (1) and (2) can be measured by means of the *star discrepancy* of the point sequence $(x_n)_{n\in\mathbb{N}} \subset [0,1]^s$, which is defined as

$$D_N^{*s}(x_1,\dots,x_N) = \sup_{x\in[0,1]^s} \left| \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{[0,x)}(x_n) - \lambda([0,x)) \right|, \quad N \in \mathbb{N}. \tag{3}$$

A sequence is uniformly distributed if and only if the discrepancy of its first N elements tends to 0 as $N \to \infty$.

The Koksma-Hlawka inequality states that the deviation between the finite average

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

and the integral of a function f can be estimated by the product of the star discrepancy of the point set $\{x_1, \dots, x_N\}$ and the variation (in the sense of Hardy and Krause) of f; see [7, 15, 17] for details, as well as for a general introduction to uniform distribution theory and discrepancy theory. Thus, as a rule of thumb it is reasonable to perform Quasi-Monte Carlo integration by using point sets having small discrepancy. There exist many constructions of so–called *low-discrepancy point sets* and *low-discrepancy sequences*, where for many decades the main focus of research was set on finding point sets and sequences satisfying strong discrepancy bounds for large N and fixed s; however, recently, the problem asking for point sets having small discrepancy for a moderate number of points in comparison with the dimension has attracted some attention.

From a probabilistic point of view, a sequence $(x_n)_{n \in \mathbb{N}}$ is uniformly distributed if the corresponding sequence of empirical distribution functions converges to the uniform distribution. In particular, by the Glivenko-Cantelli theorem a random sequence is almost surely uniformly distributed.

1.2 The inverse of the discrepancy

Let $n^*(s,\varepsilon)$ denote the smallest possible size of a set of s-dimensional points having star discrepancy not exceeding ε. This quantity is called the *inverse of the discrepancy*. By a profound result of Heinrich, Novak, Wasilkowski and Woźniakowski [13] we know that

$$n^*(s,\varepsilon) \leq c_{\text{abs}} s\, \varepsilon^{-2}. \tag{4}$$

This upper bound is complemented by a lower bound of Hinrichs [14], stating that

$$n^*(s,\varepsilon) \geq c_{\text{abs}} s\, \varepsilon^{-1}. \tag{5}$$

Together, (4) and (5) give a complete description of the dependence of the inverse of the star discrepancy on the dimension s, while the precise dependence on ε is still an important open problem. For their proof Heinrich *et al.* use deep results of Haussler [12] and Talagrand [22]. In fact, what they exactly prove is that a randomly generated sequence satisfies (4) with positive probability. The upper bound in (4) is equivalent to the fact that for any N and s there exists a set of N points in $[0,1]^s$ satisfying the discrepancy bound

$$D_N^{*s} \leq c_{\text{abs}} \frac{\sqrt{s}}{\sqrt{N}}. \tag{6}$$

For more details on the inverse of the discrepancy and on feasibility of Quasi-Monte Carlo integration we refer to [9, 18, 19].

1.3 Double infinite matrices

Dick [4] observed that the probabilities of the exceptional sets in the argument of Heinrich *et al.* to prove (6) are summable over s and N, if the factor N is replaced by $N \ln N$. More precisely, he proved that with positive probability all the $N \times s$-dimensional projections of a randomly generated double infinite matrix $(X_{n,i})_{n,i \in \mathbb{N}}$ satisfy

$$D_N^{*s} \leq c_{\text{abs}} \sqrt{\ln N} \frac{\sqrt{s}}{\sqrt{N}}. \tag{7}$$

Dick's result has been slightly improved by Doerr, Gnewuch, Kritzer and Pillichshammer [6], again for randomly generated matrices. It is clear that such a randomly generated matrix cannot achieve the discrepancy bound (6) uniformly in s and N, since by Philipp's law of the iterated logarithm [20] for any sequence $(X_n)_{n \in \mathbb{N}}$ of independent, uniformly distributed random vectors

$$\limsup_{N \to \infty} \frac{\sqrt{N} D_N^{*s}(X_1, \ldots, X_N)}{\sqrt{\ln \ln N}} = \frac{1}{\sqrt{2}} \quad \text{almost surely.} \tag{8}$$

Thus, the factor $\sqrt{\ln N}$ in (7) cannot be reduced to a function from the class $o\left(\sqrt{\ln \ln N}\right)$, since by (8) no positive probability can exist for a random matrix satisfying such an asymptotic discrepancy bound. However, there exists a double infinite matrix constructed in a hybrid way (that is, consisting of both random and deterministic entries) whose $N \times s$-dimensional projections satisfy

$$D_N^{*s} \leq c_{\text{abs}} \frac{\sqrt{s}}{\sqrt{N}}$$

uniformly in N and s, see [1]. The purpose of the present paper is to find optimal discrepancy bounds which hold for *random* double infinite matrices with positive probability, and to give estimates for the corresponding probabilities.

1.4 Complete uniform distribution and Markov Chain Monte Carlo

A sequence $(x_n)_{n \in \mathbb{N}}$ of numbers from $[0,1]$ is called *completely uniformly distributed* (c.u.d.), if for any s the sequences

$$((x_n, \ldots, x_{n+s-1}))_{n \in \mathbb{N}} \subset [0,1]^s$$

are uniformly distributed. This property was suggested by Knuth as a test for pseudorandomness of sequences in volume II of his celebrated monograph on *The Art of Computer Programming*. However, in our context it is more sensible to use an non-overlapping version of the above construction, namely to use the first Ns elements of an infinite sequence $(x_n)_{n \in \mathbb{N}}$ to construct N points $u_1^{(s)}, \ldots, u_N^{(s)} \in [0,1]^s$ in the form

$$\begin{aligned} u_1^{(s)} &= (x_1, \ldots, x_s), \\ u_2^{(s)} &= (x_{s+1}, \ldots, x_{2s}), \\ &\vdots \\ u_N^{(s)} &= (x_{(N-1)s+1}, \ldots, x_{Ns}). \end{aligned} \tag{9}$$

These two notions of complete uniform distribution are equivalent insofar as a sequence is c.u.d. in the first sense if and only if it is c.u.d. in the second sense.

In many practical applications, e.g., in financial mathematics, a general integral of the form

$$\int_\Omega f(y)\, \mathrm{d}\mu(y)$$

for some measure space Ω and some measure μ can be transferred to the form

$$\int_{[0,1]^s} \hat{f}(y)\, \mathrm{d}y. \tag{10}$$

That is, the original function f (which can be, for example, the payoff-function of some financial derivative, where the properties of the underlying problem are described by μ) has to be replaced by a new function $\hat{f}$, which contains all the information about the change of measure from μ to λ. If $\hat{f}$ can be easily calculated and is a well-behaved function, then the integral in (10) can be directly computed using classical QMC methods. However, in many cases the function $\hat{f}$ will be difficult to handle, and it is computationally easier to directly calculate the integral

$$\int_\Omega f(y)\, \mathrm{d}\mu(y) = \int_\Omega f(y)\pi(y)\, \mathrm{d}y, \tag{11}$$

where π is the density function of μ, by sampling random variables having density π. In other words, it is necessary to sample random variables having density π, which may not be directly possible by standard methods. This problem can be solved by using *Markov Chain Monte Carlo* (MCMC). Here y_0 is a (random) starting element, and the other samples y_n are constructed iteratively in the form $y_n = \Phi(y_{n-1}, u_n)$, where $u_n \in [0,1]^s$ and Φ is an appropriate function. The distribution of $(y_n | y_0, \ldots, y_{n-1})$ is the same as the distribution of $(y_n | y_{n-1})$, which means that the sequence $(y_n)_{n\in\mathbb{N}}$ has the Markov property. Then, if π is the density of the stationary distribution of $(y_n)_{n\in\mathbb{N}}$ under Φ, the integral (11) can be estimated by

$$\frac{1}{N}\sum_{n=1}^{N} f(y_n).$$

For more background information on MCMC we refer to [16, 21].

Traditionally, the points $(u_n)_{n\in\mathbb{N}} \subset [0,1]^s$ in the aforementioned construction are sampled randomly. However, Chen, Dick and Owen [3] recently showed that it is also possible to use quasi-random points instead, namely by choosing $u_n = u_n^{(s)}$, $n \in \mathbb{N}$, constructed out of a completely uniformly distributed sequence $(x_n)_{n\in\mathbb{N}}$ according to (9). Then, under some regularity assumptions, the MCMC sampler consistently samples points having density π, provided the discrepancy of the c.u.d.-sequence is sufficiently small. The results of [3] are of a merely qualitative nature, stating that certain MCMC-methods are consistent if the discrepancy of the QMC-points, constructed according to (9), tends to zero. However, it is natural to assume that the speed of convergence of these MCMC samplers can be estimated by the speed of decay of the discrepancy of $\left\{u_n^{(s)} \,|\, 1 \le n \le N\right\}$, and therefore it is desirable to find sequences $(x_n)_{n\in\mathbb{N}}$ for which this discrepancy is small. In [3] it is noted that Dick's proof from [4] can be modified to prove the existence of a sequence $(x_n)_{n\in\mathbb{N}}$ for which

$$D_N^{*s}\left(u_1^{(s)},\ldots,u_N^{(s)}\right) \le c_{\mathrm{abs}}\sqrt{\ln N}\,\frac{\sqrt{s}}{\sqrt{N}},$$

uniformly in N and s. In the present paper we will show that the factor $\sqrt{\ln N}$ can be reduced to $\sqrt{c_{\mathrm{abs}} + (\ln\ln N)/s}$, which is already very close to the upper bound of Heinrich *et al.* in (6). We tried to find a hybrid construction achieving (6), similar to the hybrid construction of a double infinite matrix mentioned at the end of the previous section, but due to the complicated dependence between the diverse coordinates of the point sets $\left\{u_n^{(s_1)} \,|\, 1 \le n \le N_1\right\}$ and $\left\{u_n^{(s_2)} \,|\, 1 \le n \le N_2\right\}$ for different s_1, s_2 and N_1, N_2 this seems to be hopeless. The discrepancy bound in our Theorem 2 below is the strongest known discrepancy bound for c.u.d.-sequences (which is valid uniformly in N and s) at present. Furthermore, Dick's result is of limited practical use as it involves unknown constants, while our results are completely explicit and even allow to calculate the probability of a random sequence satisfying the desired discrepancy bounds.

1.5 Results

Let $X = (X_{n,i})_{n,i\in\mathbb{N}}$ be a double infinite array of independent copies of some uniformly $[0,1]$-distributed random variable. For positive integers N and s set

$$\mathscr{P}_{N,s} = \left\{X^{(1)},\ldots,X^{(N)}\right\},$$

where $X^{(n)} = (X_{n,1},\ldots,X_{n,s}) \in [0,1]^s$ for $n = 1,\ldots,N$. Hence, $\mathscr{P}_{N,s}$ is the projection of X onto its first $N \times s$ entries. As in (3) let $D_N^{*s}(\mathscr{P}_{N,s})$ denote the s-dimensional star discrepancy of these N points.

The main technical tool of the present paper is the following Lemma 1, which will be used to derive our theorems.

Lemma 1. *Let $\alpha \geq 1$ and $\beta \geq 0$ be given. Moreover, for $M, s \in \mathbb{N}$ set*

$$\Omega_{M,s} = \left\{ \max_{2^M \leq N < 2^{M+1}} N \cdot D_N^{*s}(\mathscr{P}_{N,s}) > \sqrt{\alpha A + \beta B \frac{\ln M}{s}} \sqrt{s \cdot 2^M} \right\},$$

where $A = 1165$ and $B = 178$. Then we have for all natural numbers M and s

$$\mathbb{P}(\Omega_{M,s}) < \frac{1}{(1+s)^{\alpha}} \frac{1}{M^{\beta}}.$$

The proof of Lemma 1, which is given in Section 2 below, essentially follows the lines of [2]. In addition we use a Bernstein type inequality which can be found, e.g., in Einmahl and Mason [8, Lemma 2.2]:

Lemma 2 (Maximal Bernstein inequality). *For $M \in \mathbb{N}$ let Z_n, $1 \leq n \leq 2^{M+1}$, be independent random variables with zero mean and variance $\mathbb{V}(Z_n)$. Moreover, assume $|Z_n| \leq C$ for some $C > 0$ and all $n \in \{1, \dots, 2^{M+1}\}$. Then for every $t \geq 0$*

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

At the end of Section 2, we will conclude the following two theorems from Lemma 1. Here ζ denotes the Riemann Zeta function.

Theorem 1. *Let $\gamma \geq \zeta^{-1}(2) \approx 1.73$ be arbitrarily fixed. Then with probability strictly larger than $1 - (\zeta(\gamma) - 1)^2 \geq 0$ we have for all $s \in \mathbb{N}$ and every $N \geq 2$*

$$D_N^{*s}(\mathscr{P}_{N,s}) \leq \sqrt{\gamma} \cdot \sqrt{1165 + 178 \frac{\ln \log_2 N}{s}} \cdot \sqrt{\frac{s}{N}}.$$

In particular, there exists a positive probability that a random matrix X satisfies for all $s \in \mathbb{N}$ and every $N \geq 2$

$$D_N^{*s}(\mathscr{P}_{N,s}) \leq \sqrt{2130 + 308 \frac{\ln \ln N}{s}} \cdot \sqrt{\frac{s}{N}}.$$

In our second theorem, we show how our method can be applied to obtain discrepancy bounds for completely uniformly distributed sequences. To this end let $X = (X_n)_{n \in \mathbb{N}}$ be a sequence of independent, identically distributed random variables having uniform distribution on $[0,1]$. For any $s \in \mathbb{N}$ and $N \geq 2$ define a sequence

$$\begin{aligned}
U_1^{(s)} &= (X_1, \dots, X_s),\\
U_2^{(s)} &= (X_{s+1}, \dots, X_{2s}),\\
&\vdots\\
U_N^{(s)} &= (X_{(N-1)s+1}, \dots, X_{Ns}).
\end{aligned}$$

Furthermore, let

$$\mathscr{U}_{N,s} = \left\{ U_1^{(s)}, \dots, U_N^{(s)} \right\}. \tag{12}$$

Theorem 2. *Let $\gamma \geq \zeta^{-1}(2)$ be arbitrarily fixed. Then with probability strictly larger than $1-(\zeta(\gamma)-1)^2 \geq 0$ we have for all $s \in \mathbb{N}$ and every $N \geq 2$*

$$D_N^{*s}(\mathscr{U}_{N,s}) \leq \sqrt{\gamma} \cdot \sqrt{1165 + 178 \frac{\ln \log_2 N}{s}} \cdot \sqrt{\frac{s}{N}}.$$

In particular, there exists a positive probability that a random sequence X is completely uniformly distributed and satisfies for all $s \in \mathbb{N}$ and every $N \geq 2$

$$D_N^{*s}(\mathscr{U}_{N,s}) \leq \sqrt{2130 + 308 \frac{\ln \ln N}{s}} \cdot \sqrt{\frac{s}{N}}.$$

Our results are essentially optimal in two respects. On the one hand, for any N and s satisfying $N \leq \exp(\exp(c_{\text{abs}} s))$ our Theorem 1 gives (with positive probability) a discrepancy estimate of the form

$$D_N^{*s}(\mathscr{P}_{N,s}) \leq c_{\text{abs}} \frac{\sqrt{s}}{\sqrt{N}} \tag{13}$$

and by this means resembles the aforementioned result of Heinrich *et al.* (note that a discrepancy estimate of the form (13) is not of much use if $N > \exp(\exp(c_{\text{abs}} s))$, since in this case the well-known bounds for low-discrepancy sequences are much smaller). Hence, any improvement of our Theorem 1 (up to the values of the constants) would require an improvement of (4). Furthermore, recent research of Doerr [5] shows that the expected value of the star-discrepancy of a set of N points in $[0,1]^s$ is of order $s^{1/2}N^{-1/2}$. The probability estiamtes in [5] can be used to show that for an i.i.d. random matrix X there exist absolute constants K_1, K_2 such that for every s the probability, that the $N \times s$-dimensional projections of X have a discrepancy bounded by $K_1 s^{1/2} N^{-1/2}$ for all $N \leq \exp(\exp(K_2 s))$, is zero. This means that for N in this range our Theorem 1 is essentially optimal. It should also be mentioned that it is possible that (4) is already optimal and cannot be improved.

On the other hand, for fixed s and large N our discrepancy estimate is of the form

$$D_N^{*s}(\mathscr{P}_{N,s}) \leq c(s) \frac{\sqrt{\log \log N}}{\sqrt{N}}.$$

This discrepancy bound is asymptotically optimal for random matrices (up to the value of the constants c(s)), since in view of the law of the iterated logarithm (8) no random construction can achieve a significantly better rate of decay of the discrepancy with positive probability.

We note that constructing a sequence of elements of $[0,1]$ satisfying good discrepancy bounds in the sense of complete uniform distribution is much more difficult than constructing point sets in $[0,1]^s$ for a fixed number of points and fixed dimension s. There exist constructions of sequences having good c.u.d.-behavior, but usually the corresponding discrepancy bounds are only useful if N is much larger than s; it is possible that the discrepancy estimates in Theorem 2 are optimal in the sense that they give good results uniformly for *all* possible values of N and s, and that Theorem 2 cannot be significantly improved (up to the values of the constants) in this regard.

2 Proofs

Proof of Lemma 1: Since the proof is somewhat technical we split it into different steps. The main ingredients in the proof are a dyadic decomposition of the unit cube, which was introduced in [2], a maximal version of Bernstein's inequality (Lemma 2), which is also used to prove the law of the iterated logarithm in probability theory, and Dick's observation from [4] that the exceptional probabilities are exponentially decreasing in s and are therefore summable over s. Our results could not be proved using the method in [13] (which is tailor-made for fixed N and s), since there does not exist a maximal version of Talagrand's large deviations inequality for empirical processes, which is the crucial ingredient in [13].

Step 1. Let $M,s \in \mathbb{N}$ be fixed. Without loss of generality we can assume

$$\frac{1}{2}\sqrt{\alpha A + \beta B \frac{\ln M}{s}}\sqrt{\frac{s}{2^M}} < 1 \tag{14}$$

because otherwise

$$\Omega_{M,s} \subseteq \left\{ \max_{2^M \le N < 2^{M+1}} N \cdot D_N^{*s}(\mathscr{P}_{N,s}) > 2^{M+1} \right\} = \emptyset.$$

For a moment assume $L \geq 2$ to be given. Let $(a_k)_{k=-1}^{L}$ and $(b_k)_{k=-1}^{L}$ be two non-negative, non-increasing sequences such that

$$A \geq 2\left(\sum_{k=-1}^{L}\sqrt{a_k}\right)^2 \quad \text{and} \quad B \geq 2\left(\sum_{k=-1}^{L}\sqrt{b_k}\right)^2 \tag{15}$$

and set

$$y_k = \alpha a_k + \beta b_k \frac{\ln M}{s} \quad \text{and} \quad t_k = \sqrt{y_k}\sqrt{s \cdot 2^M}. \tag{16}$$

Hence, using (14), as well as (15), we have

$$1 > \frac{1}{\sqrt{2}} \sqrt{\alpha a_{-1} + \beta b_{-1} \frac{\ln M}{s}} \sqrt{\frac{s}{2^M}} = 2 \cdot \frac{1}{2\sqrt{2}} \sqrt{y_{-1}} \sqrt{\frac{s}{2^M}}.$$

If we choose $L \in \mathbb{N}$ such that

$$\frac{1}{2}\left(\frac{1}{2\sqrt{2}}\sqrt{y_{-1}}\sqrt{\frac{s}{2^M}}\right) < 2^{-L} \leq \frac{1}{2\sqrt{2}}\sqrt{y_{-1}}\sqrt{\frac{s}{2^M}} \tag{17}$$

this implies $L \geq 2$.

Since the square root function is sublinear and concave we may use Jensen's inequality to obtain

$$\begin{aligned}
\sum_{k=-1}^{L} \sqrt{y_k} &\leq \sum_{k=-1}^{L} \sqrt{\alpha a_k} + \sum_{k=-1}^{L} \sqrt{\beta b_k \frac{\ln M}{s}} \\
&= \sqrt{\alpha \left(\sum_{k=-1}^{L} \sqrt{a_k}\right)^2} + \sqrt{\beta \left(\sum_{k=-1}^{L} \sqrt{b_k}\right)^2 \frac{\ln M}{s}} \\
&\leq \sqrt{\alpha 2 \left(\sum_{k=-1}^{L} \sqrt{a_k}\right)^2 + \beta 2 \left(\sum_{k=-1}^{L} \sqrt{b_k}\right)^2 \frac{\ln M}{s}} \\
&\leq \sqrt{\alpha A + \beta B \frac{\ln M}{s}} \qquad (18)
\end{aligned}$$

out of (15) and the definition of y_k.

Step 2. In what follows we use a decomposition of the s-dimensional unit cube in terms of δ-covers and δ-bracketing covers. A detailed description of this decomposition can be found in [2]. We briefly sketch the main points.
For any given $\delta \in (0,1]$ 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 two elements $x,z \in \Gamma$ such that $x \leq y \leq z$ and $\lambda([0,z)\backslash[0,x)) \leq \delta$. Furthermore, a finite set Δ of pairs of points from $[0,1]^s$ is called a δ-*bracketing cover* if for every pair $(x,z) \in \Delta$ we have $\lambda([0,z)\backslash[0,x)) \leq \delta$, and if for every $y \in [0,1]^s$ there exists a pair $(x,z) \in \Delta$ such that $x \leq y \leq z$. The concepts of δ-covers and δ-bracketing covers were investigated in detail in [10].
For $1 \leq k < L$ let Γ_k denote a 2^{-k}-cover of $[0,1]^s$. Moreover, let Δ_L denote a 2^{-L}-bracketing cover of $[0,1]^s$. For notational convenience we set

$$\begin{aligned}
\Gamma_L &= \left\{ p_L \in [0,1]^s \,\middle|\, (p_L, p_{L+1}) \in \Delta_L \text{ for some } p_{L+1} \right\}, \\
\Gamma_{L+1} &= \left\{ p_{L+1} \in [0,1]^s \,\middle|\, (p_L, p_{L+1}) \in \Delta_L \text{ for some } p_L \right\}
\end{aligned}$$

and $p_0 = 0 \in [0,1]^s$. Furthermore, for points $a,b \in [0,1]^s$ we define

$$\overline{[a,b)} = \begin{cases} [0,b) \setminus [0,a), & \text{if } a \neq 0, \\ [0,a), & \text{if } a = 0 \text{ and } b \neq 0, \end{cases}$$

as well as $\overline{[0,b)} = \emptyset$ if $b = 0$.
By the definition of a δ-bracketing cover, to every $x \in [0,1]^s$ we can assign a set $\overline{[p_L(x), p_{L+1}(x))}$, such that $p_k \in \Gamma_k,\ k = L, L+1$, and $\lambda(\overline{[p_L(x), p_{L+1}(x))}) \leq 2^{-L}$. Now, by the definition of a δ-cover, we can assign a point $p_{L-1}(x) \in \Gamma_{L-1} \cup \{0\}$ such that $p_{L-1}(x) \leq p_L(x)$ and $\lambda(\overline{[p_{L-1}(x), p_L(x))}) \leq 2^{-L+1}$. Next we assign a point $p_{L-2}(x) \in \Gamma_{L-2} \cup \{0\}$ such that $p_{L-2}(x) \leq p_{L-1}(x)$ and $\lambda(\overline{[p_{L-2}(x), p_{L-1}(x))}) \leq 2^{-L+2}$. Proceeding inductively, also for every $k = 1, \ldots, L-3$ we find a point $p_k(x) \in \Gamma_k \cup \{0\}$ such that $p_k(x) \leq p_{k+1}(x)$ and $\lambda(\overline{[p_k(x), p_{k+1}(x))}) \leq 2^{-k}$. Finally for every $k = 1, \ldots, L+1$ we have assigned points $p_k(x),\ 1 \leq k \leq L+1$, belonging to $\Gamma_k \cup \{0\}$ for each k, such that, writing $I_k(x) = \overline{[p_k(x), p_{k+1}(x))}$, $1 \leq k \leq L$, and setting $I_0(x) = \overline{[0, p_1(x))}$, we have

$$\bigcup_{k=0}^{L-1} I_k(x) \subset [0,x) \subset \bigcup_{k=0}^{L} I_k(x). \tag{19}$$

and

$$\lambda(I_k(x)) \leq 2^{-k}, \quad k \in \{0, \ldots, L\}.$$

For every $k \in \{0, \ldots, L\}$, let $\mathscr{A}_k = \{I_k(x) \,|\, x \in [0,1]^s\}$ denote the collection of all possible sets $I_k(x)$, as x runs through the whole unit cube $[0,1]^s$. Then the cardinality of these sets is bounded by $\#\Gamma_{k+1}$. Using Theorem 1.15 from Gnewuch [10] we see that we can choose our 2^{-k}-covers Γ_k such that

$$\begin{aligned} \#\mathscr{A}_k &\leq \#\Gamma_{k+1} \leq 2^s \frac{s^s}{s!}(2^{k+1}+1)^s < \frac{1}{2}\sqrt{2/\pi}\left(2e(2^{k+1}+1)\right)^s \\ &\leq \frac{1}{2}\exp\left(\ln\sqrt{2/\pi} + \alpha s \ln\left(2e(2^{k+1}+1)\right)\right), \qquad k \in \{0, \ldots, L-2\} \end{aligned} \tag{20}$$

where we used Stirling's formula, as well as $s \geq 1$, and $\alpha \geq 1$. Similarly, we can choose the 2^{-L}-bracketing cover Δ_L in a way that

$$\begin{aligned} \#\mathscr{A}_k &\leq \#\Gamma_L = \#\Delta_L \leq 2^{s-1}\frac{s^s}{s!}(2^L+1)^s < \frac{1}{2}\sqrt{1/(2\pi)}\left(2e(2^L+1)\right)^s \\ &\leq \frac{1}{2}\exp\left(\ln\sqrt{1/(2\pi)} + \alpha s \ln\left(2e(2^L+1)\right)\right), \qquad k \in \{L-1, L\}. \end{aligned}$$

Step 3. Given the decomposition from Step 2 we define for $k = 0, \ldots, L$ and $I \in \mathscr{A}_k$

$$E_k(I) = \left\{ \max_{2^M \leq N < 2^{M+1}} \left| \sum_{n=1}^{N} \mathbb{1}_I\left(X^{(n)}\right) - N\lambda(I) \right| > t_k \right\},$$

where the numbers t_k were defined in (16). Moreover, let $E_k = \bigcup_{I \in \mathscr{A}_k} E_k(I)$ and $E = \bigcup_{k=0}^{L} E_k$. If we can show that independently of $I \in \mathscr{A}_k$ we have

$$2^{k+1}(1+s)^{\alpha}M^{\beta}\cdot\#\mathscr{A}_k\cdot\mathbb{P}(E_k(I))<1 \qquad \text{for} \qquad k\in\{0,\dots,L\}, \tag{21}$$

then this leads to

$$\mathbb{P}(E)\leq\sum_{k=0}^{L}\sum_{I\in\mathscr{A}_k}\mathbb{P}(E_k(I))\leq\sum_{k=0}^{L}2^{-(k+1)}(1+s)^{-\alpha}M^{-\beta}<\frac{1}{(1+s)^{\alpha}}\frac{1}{M^{\beta}}. \tag{22}$$

In order to show (21) for fixed $I\in\mathscr{A}_k$ with $k\in\{0,\dots,L\}$ let us define the random variables $Z_n=\mathbb{1}_I(X^{(n)})-\lambda(I)$, where $n=1,\dots,2^{M+1}$. Obviously, all the Z_n are independent and bounded by

$$|Z_n|\leq C=\max\{\lambda(I),1-\lambda(I)\}.$$

Furthermore, we have $\mathbb{E}(Z_n)=0$ and $\mathbb{V}(Z_n)=\lambda(I)(1-\lambda(I))$. From Lemma 2 applied to $\pm Z_n$ and $t=t_k$ we conclude

$$\mathbb{P}(E_k(I))\leq 2\exp\left(-\frac{t_k^2/2^M}{4\lambda(I)(1-\lambda(I))+2Ct_k/(3\cdot 2^M)}\right). \tag{23}$$

Since $t_k=\sqrt{y_k}\sqrt{s\cdot 2^M}\leq\sqrt{y_{-1}}\sqrt{s\cdot 2^M}$ estimate (17) implies

$$2\frac{t_k}{3\cdot 2^M}<2\frac{4\sqrt{2}}{3}2^{-L}<\begin{cases}2^{-k+1}, & \text{for } k=0,\dots,L-1,\\ 2^{-L+2}, & \text{for } k=L.\end{cases}$$

For $k\in\{0,1\}$ we use $\lambda(I)(1-\lambda(I))\leq 1/4$ and $C\leq 1$ to estimate the denominator in (23) by 3 and 2, respectively. If $L>2$ and $k\in\{2,\dots,L-1\}$ it is easy to see that maximizing $4\lambda(1-\lambda)+2^{-k+1}(1-\lambda)$ subject to $\lambda\in[0,2^{-k}]$ gives the bound $3\cdot 2^{-k+1}(1-2^{-k})$. For $K=L$ a similar argument shows that the denominator in (23) is less than $2^{-L+3}(1-2^{-L})$. Hence, we have

$$\mathbb{P}(E_k(I))\leq 2\exp(-y_k s/\Lambda_k), \quad \text{where} \quad \Lambda_k=\begin{cases}3, & k=0,\\ 2^{-k+3}(1-2^{-k}), & k>0.\end{cases} \tag{24}$$

Using $s\geq 1$ it is easily seen that for any $k\geq 0$

$$2^{k+1}(1+s)^{\alpha}M^{\beta}\leq\exp\left(\ln 2^{k+1}+\alpha s\ln 2+\beta\ln M\right) \tag{25}$$

such that we can conclude (21) if we choose a_k and b_k in the right way. We explain the necessary arguments for the case $k=0$ explicitly. The case $k>0$ then works in the same manner. Combining the estimates (20), (24) and (25) we have because of $\alpha s\geq 1$ that

$$
\begin{aligned}
&2^1(1+s)^\alpha M^\beta \cdot \#\mathscr{A}_0 \cdot \mathbb{P}(E_0(I)) \\
&\leq \exp\left(\ln\left(2\sqrt{2/\pi}\right) + \alpha s \ln\left(4e(2^1+1)\right) + \beta \ln(M) - \alpha s a_0/\Lambda_0 - \beta \ln(M) b_0/\Lambda_0\right) \\
&\leq \exp\left(\alpha s\left(\ln\left(24e\sqrt{2/\pi}\right) - a_0/\Lambda_0\right) + \beta \ln(M)\left(1-b_0/\Lambda_0\right)\right) \leq \exp(0) = 1,
\end{aligned}
$$

if we choose

$$
b_0 = \Lambda_0 \quad \text{and} \quad a_0 = \Lambda_0 \ln\left(24e\sqrt{2/\pi}\right). \tag{26}
$$

Similarly, we choose

$$
b_k = \Lambda_k \quad \text{and} \quad a_k = \Lambda_k \cdot \ln\left(2^{k+3} e\sqrt{2/\pi}(2^{k+1}+1)\right) \tag{27}
$$

to obtain (21) also for $k = 1, \dots, L$.

Step 4. To conclude the main statement of Lemma 1 by applying (22) it remains to show that $\Omega_{M,s} \subseteq E$. To this end let $N \in [2^M, 2^{M+1})$ be arbitrary, but fixed. Due to the definitions in Step 3 for every $\omega \in E^C = \bigcap_{k=0}^L \bigcap_{I \in \mathscr{A}_k} E_k(I)^C$ and $x_n = X^{(n)}(\omega) \in [0,1]^s$ (for $n = 1, \dots, N$) we have

$$
\left|\sum_{n=1}^N \mathbb{1}_I(x_n) - N\lambda(I)\right| \leq t_k \qquad \text{for all } k \in \{0, \dots, L\} \text{ and every } I \in \mathscr{A}_k.
$$

Thus, from (19) we conclude for every $x \in [0,1]^s$

$$
\begin{aligned}
\sum_{n=1}^N \mathbb{1}_{[0,x)}(x_n) &\leq \sum_{k=0}^L \sum_{n=1}^N \mathbb{1}_{\overline{[p_k(x), p_{k+1}(x))}}(x_n) \\
&\leq \sum_{k=0}^L \left(N\lambda\left(\overline{[p_k(x), p_{k+1}(x))}\right) + t_k\right) \\
&= N\lambda([0,x)) + N\lambda\left(\overline{[x, p_{L+1}(x))}\right) + \sum_{k=0}^L t_k.
\end{aligned}
$$

Since $p_L(x) \leq x \leq p_{L+1}(x)$ the volume of the set $\overline{[x, p_{L+1}(x))}$ can be estimated from above by 2^{-L} what is no larger than $1/2\sqrt{y_{-1}}\sqrt{s/2^M}$ due to (17). Hence, because of $N < 2 \cdot 2^M$, the second term in the above sum is less than $\sqrt{y_{-1}}\sqrt{s \cdot 2^M}$. Consequently,

$$
\begin{aligned}
\sum_{n=1}^N \mathbb{1}_{[0,x)}(x_n) &< N\lambda([0,x)) + \sqrt{s \cdot 2^M} \sum_{k=-1}^L \sqrt{y_k} \\
&\leq N\lambda([0,x)) + \sqrt{\alpha A + \beta B \frac{\ln M}{s}} \sqrt{s \cdot 2^M},
\end{aligned}
$$

where we used (18) from Step 1 for the last estimate. In a similar way we obtain the corresponding lower bound

$$\begin{aligned}\sum_{n=1}^{N} \mathbb{1}_{[0,x)}(x_n) &\geq \sum_{k=0}^{L-1}\sum_{n=1}^{N} \mathbb{1}_{\overline{[p_k(x),p_{k+1}(x))}}(x_n) \\ &\geq N\lambda([0,x)) - N\lambda\left(\overline{[x,p_{L+1}(x))}\right) - \sum_{k=0}^{L-1} t_k \\ &> N\lambda([0,x)) - \sqrt{\alpha A + \beta B \frac{\ln M}{s}}\sqrt{s \cdot 2^M}.\end{aligned}$$

Both the estimates, together with the definition of D_N^{*s}, imply

$$N \cdot D_N^{*s}(x_1,\ldots,x_N) \leq \sqrt{\alpha A + \beta B \frac{\ln M}{s}}\sqrt{s \cdot 2^M}$$

since $x \in [0,1]^s$ was arbitrary. Due to the fact that this holds for all $N \in [2^M, 2^{M+1})$ and for every $\omega \in E^C$ we have shown $\Omega_{M,s} \subseteq E$.

Step 5. Finally, we need to check that the sequences $(a_k)_{k=-1}^{L}$ and $(b_k)_{k=-1}^{L}$, which were defined in (26), (27), and (24), satisfy the assumptions made at the beginning of Section 2. We already checked that $L \geq 2$, see (17). Moreover, it is obvious that both sequences are non-negative and non-increasing for $k \geq 0$. Hence, we define $a_{-1} = a_0$ and $b_{-1} = b_0$ to guarantee that this holds for all k. It remains to show (15). To this end we calculate for $c \in \{a, b\}$

$$2\left(\sum_{k=-1}^{L}\sqrt{c_k}\right)^2 \leq 2\left(2\sqrt{c_0} + \sum_{k=1}^{\infty}\sqrt{c_k}\right)^2 \leq \begin{cases} 1164.87, & \text{if } c = a, \\ 177.41, & \text{if } c = b. \end{cases}$$

This completes the proof choosing $A = 1165$ and $B = 178$. $\square$

Proof of Theorem 1: Let ζ denote the Riemann Zeta function, and let $\gamma \geq \zeta^{-1}(2)$. Due to the choice of $A > 9/2$ in Lemma 1 we have $\Omega_{1,s} = \emptyset$ for all $s \in \mathbb{N}$. Hence, for $\alpha = \beta = \gamma$ it follows

$$\mathbb{P}\left(\bigcup_{s\geq 1}\bigcup_{M\geq 1}\Omega_{M,s}\right) \leq \sum_{s\geq 1}\sum_{M\geq 2}\mathbb{P}(\Omega_{M,s}) < \sum_{s\geq 1}\sum_{M\geq 2}\frac{1}{(1+s)^\gamma}\frac{1}{M^\gamma} = (\zeta(\gamma)-1)^2 \leq 1.$$

In particular, this implies $\mathbb{P}\left(\left(\bigcup_{s\geq 1}\bigcup_{M\geq 1}\Omega_{M,s}\right)^C\right) > 0$. Since $\zeta^{-1}(2) < 1.73$ and $B\frac{\ln\log_2 N}{s} \leq B\frac{\ln\ln N}{s} - B\ln\ln 2 \leq 66 + B\frac{\ln\ln N}{s}$, we can choose $\gamma = 1.73$ and obtain

$$\begin{aligned} D_N^{*s}(\mathscr{P}_{N,s}) &\leq \sqrt{1.73}\sqrt{1165 + 66 + 178\frac{\ln\ln N}{s}} \cdot \sqrt{\frac{s}{N}} \\ &\leq \sqrt{2130 + 308\frac{\ln\ln N}{s}} \cdot \sqrt{\frac{s}{N}}\end{aligned}$$

with positive probability. This proves Theorem 1. $\square$

Proof of Theorem 2: For any fixed N and s, the point set $\mathscr{U}_{N,s}$ is an array of $N \times s$ i.i.d. uniformly distributed random variables, just like $\mathscr{P}_{N,s}$ in the assumptions of Lemma 1. For given $\alpha \geq 1,\ \beta \geq 0,\ M,s \in \mathbb{N}$, as well as $A = 1165$ and $B = 178$, set

$$\Omega_{M,s} = \left\{ \max_{2^M \leq N < 2^{M+1}} N \cdot D_N^{*s}(\mathscr{U}_{N,s}) > \sqrt{\alpha A + \beta B \frac{\ln M}{s}} \sqrt{s \cdot 2^M} \right\}$$

where $\mathscr{U}_{N,s}$ now is defined in (12). Then Lemma 1 yields

$$\mathbb{P}(\Omega_{M,s}) < \frac{1}{(1+s)^\alpha} \frac{1}{M^\beta}.$$

With this estimate for the probabilities of the exceptional sets, the rest of the proof of Theorem 2 can be carried out in exactly the same way as the proof of Theorem 1. $\square$

Note that the choice of the constants $\alpha = \beta = \gamma$ in the above proofs is not essential. Alternatively, it would be sufficient to take any pair of parameters $1 < \alpha, \beta < \infty$ such that $(\zeta(\alpha) - 1)(\zeta(\beta) - 1) \leq 1$. Using this trade-off it is possible to fine-tune the absolute constants in our theorems in order to minimize the discrepancy bounds for given N and s. Moreover, better estimates on the size of the used δ-(bracketing) covers may lead to (minor important) improvements of these constants. For details we refer to [2] and the conjectures in Gnewuch [11].

Acknowledgements The questions considered in the present paper arose during discussions with several colleagues at the MCQMC 2012 conference. We particularly want to thank Joseph Dick for pointing out to us the connection between the discrepancy of random matrices, complete uniform distribution and Markov chain Monte Carlo.
Finally, we wish to express our gratitude to the two anonymous referees who helped to improve the presentation of the paper.

References

1. Aistleitner, C.: On the inverse of the discrepancy for infinite dimensional infinite sequences. Preprint. To appear in: J. Complexity. DOI 10.1016/j.jco.2012.06.002. URL http://dx.doi.org/10.1016/j.jco.2012.06.002
2. Aistleitner, C.: Covering numbers, dyadic chaining and discrepancy. J. Complexity **27(6)**, 531–540 (2011). DOI 10.1016/j.jco.2011.03.001. URL http://dx.doi.org/10.1016/j.jco.2011.03.001
3. Chen, S., Dick, J., Owen, A.B.: Consistency of Markov chain quasi-Monte Carlo on continuous state spaces. Ann. Statist. **39**(2), 673–701 (2011). DOI 10.1214/10-AOS831. URL http://dx.doi.org/10.1214/10-AOS831

4. Dick, J.: A note on the existence of sequences with small star discrepancy. J. Complexity **23**(4-6), 649–652 (2007). DOI 10.1016/j.jco.2007.01.004. URL http://dx.doi.org/10.1016/j.jco.2007.01.004
5. Doerr, B.: A Lower Bound for the Discrepancy of a Random Point Set. Preprint. Available at: http://arxiv.org/abs/1210.0572
6. Doerr, B., Gnewuch, M., Kritzer, P., Pillichshammer, F.: Component-by-component construction of low-discrepancy point sets of small size. Monte Carlo Methods Appl. **14**(2), 129–149 (2008). DOI 10.1515/MCMA.2008.007. URL http://dx.doi.org/10.1515/MCMA.2008.007
7. Drmota, M., Tichy, R.F.: Sequences, discrepancies and applications, *Lecture Notes in Mathematics*, vol. 1651. Springer-Verlag, Berlin (1997).
8. Einmahl, U., Mason, D.M.: Some universal results on the behavior of increments of partial sums. Ann. Probab. **24(3)**, 1388–1407 (1996). DOI 10.1214/aop/1065725186. URL http://dx.doi.org/10.1214/aop/1065725186
9. Gnewuch, M.: Entropy, randomization, derandomization, and discrepancy. To appear in: L. Plaskota, H. Wozniakowski (Eds.), Monte Carlo and Quasi-Monte Carlo Methods 2010, Springer-Verlag.
10. Gnewuch, M.: Bracketing numbers for axis-parallel boxes and applications to geometric discrepancy. J. Complexity **24(2)**, 154–172 (2008). DOI 10.1016/j.jco.2007.08.003. URL http://dx.doi.org/10.1016/j.jco.2007.08.003
11. Gnewuch, M.: Construction of minimal bracketing covers for rectangles. Electron. J. Combin. **15(1)** (2008), Research Paper 95, 20pp.
12. Haussler, D.: 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). DOI 10.1016/0097-3165(95)90052-7. URL http://dx.doi.org/10.1016/0097-3165(95)90052-7
13. Heinrich, S., Novak, E., Wasilkowski, G.W., Woźniakowski, H.: The inverse of the star-discrepancy depends linearly on the dimension. Acta Arith. **96**(3), 279–302 (2001). DOI 10.4064/aa96-3-7. URL http://dx.doi.org/10.4064/aa96-3-7
14. Hinrichs, A.: Covering numbers, Vapnik-Červonenkis classes and bounds for the star-discrepancy. J. Complexity **20**(4), 477–483 (2004). DOI 10.1016/j.jco.2004.01.001. URL http://dx.doi.org/10.1016/j.jco.2004.01.001
15. Kuipers, L., Niederreiter, H.: Uniform distribution of sequences. Wiley-Interscience [John Wiley & Sons], New York (1974). Pure and Applied Mathematics.
16. Liu, J.S.: Monte Carlo strategies in scientific computing, *Springer Series in Statistics*. Springer, New York (2008).
17. Matoušek, J.: Geometric discrepancy. An illustrated guide. Revised paperback reprint of the 1999 original. *Algorithms and Combinatorics, 18*. Springer–Verlag, Berlin (2010).
18. Novak, E., Woźniakowski, H.: Tractability of multivariate problems. Vol. I: Linear information, *EMS Tracts in Mathematics*, vol. 6. European Mathematical Society (EMS), Zürich (2008). DOI 10.4171/026. URL http://dx.doi.org/10.4171/026
19. Novak, E., Woźniakowski, H.: Tractability of multivariate problems. Vol. II: Standard information for functionals, *EMS Tracts in Mathematics*, vol. 12. European Mathematical Society (EMS), Zürich (2010). DOI 10.4171/084. URL http://dx.doi.org/10.4171/084
20. Philipp, W.: Mixing sequences of random variables and probablistic number theory. American Mathematical Society, Providence, R. I. (1971). Memoirs of the American Mathematical Society, No. 114.
21. Robert, C.P., Casella, G.: Monte Carlo statistical methods, second edn, *Springer Texts in Statistics*. Springer-Verlag, New York (2004).
22. Talagrand, M.: Sharper bounds for Gaussian and empirical processes. Ann. Probab. **22**(1), 28–76 (1994).
23. Weyl, H.: Über die Gleichverteilung von Zahlen mod. Eins. Math. Ann. **77**(3), 313–352 (1916). DOI 10.1007/BF01475864. URL http://dx.doi.org/10.1007/BF01475864