

## Low-discrepancy point sets for non-uniform measures

by

CHRISTOPH AISTLEITNER and JOSEF DICK (Sydney)

**1. Introduction.** Let  $x_1, \dots, x_N$  be a set of points in the  $d$ -dimensional unit cube  $[0, 1]^d$ . The *star-discrepancy*  $D_N^*$  of  $x_1, \dots, x_N$  is defined as

$$(1) \quad D_N^*(x_1, \dots, x_N) = \sup_{A \subset [0, 1]^d} \left| \frac{1}{N} \sum_{n=1}^N \mathbb{1}_A(x_n) - \lambda(A) \right|,$$

where  $\mathbb{1}_A$  denotes the indicator function of  $A$ ,  $\lambda$  is the  $d$ -dimensional Lebesgue measure, and the supremum is extended over all axis-parallel boxes  $A$  which have one vertex at the origin. By the Koksma–Hlawka inequality, for any  $x_1, \dots, x_N$  and any function  $f$  which has total variation  $\text{Var } f$  on  $[0, 1]^d$  (in the sense of Hardy and Krause) we have

$$(2) \quad \left| \frac{1}{N} \sum_{n=1}^N f(x_n) - \int_{[0, 1]^d} f(x) dx \right| \leq D_N^*(x_1, \dots, x_N) \text{Var } f.$$

Consequently, point sets having small discrepancy can be used for numerical integration. This method is called *quasi-Monte Carlo method* (as opposed to the *Monte Carlo method*, which uses randomly sampled points), and is heavily used in applications such as option pricing in financial mathematics. There exist several constructions of  $d$ -dimensional  $N$ -element point sets achieving a discrepancy of order

$$(3) \quad D_N^*(x_1, \dots, x_N) \ll \frac{(\log N)^{d-1}}{N}.$$

Consequently, the error of quasi-Monte Carlo integration for cleverly chosen sampling points is asymptotically significantly smaller than the (probabilistic) error of order  $N^{-1/2}$  of the Monte Carlo method. For more information on discrepancy theory and the quasi-Monte Carlo method, see for example [15, 17, 28].

2010 *Mathematics Subject Classification*: 11K38, 65D30, 65C05, 62G30.

*Key words and phrases*: discrepancy, quasi-Monte Carlo method, Beck–Fiala theorem, inverse of discrepancy.

Upper bounds of the form (3) are only useful if the number  $N$  of points is large in comparison with the dimension  $d$ ; in particular, the expression on the right-hand side of (3) is increasing for  $N \leq e^{d-1}$ . To investigate low-discrepancy point sets whose cardinality is of moderate size in comparison with the dimension  $d$ , the notion of *inverse of star-discrepancy* was introduced. Let  $n^*(d, \varepsilon)$  denote the smallest possible cardinality of a point set in  $[0, 1]^d$  achieving a star-discrepancy of at most  $\varepsilon$ . Heinrich, Novak, Wasilkowski and Woźniakowski [23] proved that

$$(4) \quad n^*(d, \varepsilon) \leq c_{\text{abs}} d \varepsilon^{-2}$$

for some absolute constant  $c_{\text{abs}}$ , and Hinrichs [24] obtained the lower bound

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

Thus the inverse of star-discrepancy depends linearly on the dimension; the precise dependence of  $n^*(d, \varepsilon)$  on  $\varepsilon$  is still an open problem.

The notion of star-discrepancy can be easily generalized to measures  $\mu$  different from the Lebesgue measure  $\lambda$  (which is the measure of uniform distribution on  $[0, 1]^d$ ). In the following we will assume that  $\mu$  is a real, non-negative, normalized Borel measure on  $[0, 1]^d$ . Similarly to (1), we define the *star-discrepancy* of a point set  $x_1, \dots, x_N \in [0, 1]^d$  with respect to  $\mu$  by

$$(6) \quad D_N^*(x_1, \dots, x_N; \mu) = \sup_{A \subset [0,1]^d} \left| \frac{1}{N} \sum_{n=1}^N \mathbb{1}_A(x_n) - \mu(A) \right|,$$

where again the supremum is extended over all axis-parallel boxes  $A$  which have one vertex at the origin. Beck [5] proved that for any  $\mu$ , any  $d \geq 1$  and any  $N \geq 1$  there exists a point set  $x_1, \dots, x_N \in [0, 1]^d$  for which

$$(7) \quad D_N^*(x_1, \dots, x_N; \mu) \ll \frac{(\log N)^{2d}}{N},$$

where the (unspecified) implied constant may only depend on  $d$ , but not on  $N$  and, somewhat surprisingly, also not on the measure  $\mu$ . As a consequence for any  $\mu$  and any  $d \geq 1$  there also exists an infinite sequence  $(x_n)_{n \geq 1}$  of points from  $[0, 1]^d$  for which

$$D_N^*(x_1, \dots, x_N; \mu) \ll \frac{(\log N)^{2d+2}}{N} \quad \text{for all } N \geq 1.$$

The purpose of the present paper is to obtain an improved and numerically explicit version of Beck’s theorems, using results on the balancing of vectors and the discrepancy of hypergraphs (these concepts are explained in Section 2). For the proof we need to approximate a general measure by a discrete measure, and we show how this problem is connected with the theory of large deviations of empirical processes and the inverse of star-discrepancy. We also prove a general, numerically explicit version of (4) (see Section 3).

Finally, combining our results with a recent version of the Koksma–Hlawka inequality due to Brandolini et al. [10], we prove the existence of cubature rules achieving fast asymptotic convergence rates for the numerical integration of functions having discontinuities of a certain form (see Section 4). Some open problems are stated in Section 5, and proofs are given in Section 6.

**THEOREM 1.** *Let  $\mu$  be any non-negative normalized Borel measure on  $[0, 1]^d$ , where  $d \geq 1$  is arbitrary. Then for any  $N \geq 1$  there exist points  $x_1, \dots, x_N \in [0, 1]^d$  such that*

$$D_N^*(x_1, \dots, x_N; \mu) \leq 63\sqrt{d} \frac{(2 + \log_2 N)^{(3d+1)/2}}{N}.$$

**THEOREM 2.** *Let  $\mu$  be any non-negative normalized Borel measure on  $[0, 1]^d$ , where  $d \geq 1$  is arbitrary. Then there exists an infinite sequence  $(x_n)_{n \geq 1}$  of points from  $[0, 1]^d$  such that*

$$D_N^*(x_1, \dots, x_N; \mu) \leq 133\sqrt{d+1} \frac{(4 + 2\log_2 N)^{(3d+4)/2}}{N} \quad \text{for } N \geq 1.$$

In the statement of these theorems and throughout the rest of this paper,  $\log$  denotes the natural logarithm, and  $\log_2$  denotes the dyadic logarithm.

**REMARK.** We want to thank Harald Niederreiter for bringing to our attention the papers of Hlawka and Mück [26, 27] and Hartinger and Kainhofer [21], which we were not aware of while originally writing this manuscript and which deal with similar issues. All three of these papers are mainly devoted to the practical problem of how to construct a point set with small discrepancy with respect to a measure  $\mu$ , and use for this purpose a suitable transformation of a small-discrepancy point set with respect to  $\lambda$  (which is also the approach used in [33]). Generally the discrepancy bounds obtained by this method are much worse than the one in our Theorem 1, as they only give a discrepancy estimate of order roughly  $N^{-1/d}$  (see e.g. [27, Satz 2]); however, in the case when  $\mu$  has a continuous density which is the product of one-dimensional densities, then a discrepancy with respect to  $\mu$  as small as (3) can be achieved (see [25] and [21, Theorem 2]). In [21] there is also mention of a variant of the Koksma–Hlawka inequality for non-uniform measures (somewhat similar to Lemma 3 in the present paper), which is implicit in the PhD thesis of Chelson [12]. Chelson’s results were never published outside his PhD thesis, and seem to be cited in occasionally incorrect form in the literature. We plan to discuss Chelson’s results, their consequences and possible applications in proper detail in another paper in the near future.

**2. Discrepancy of hypergraphs and matrices.** Let  $X$  be a finite set, and let  $\mathcal{E}$  be a family of subsets of  $X$  (which are called (*hyper*)edges).

Then the pair  $\mathcal{H} = (X, \mathcal{E})$  is called a (finite) *hypergraph*. We can partition the elements of  $X$  into two classes by a coloring function  $b: X \rightarrow \{-1, 1\}$ . The *discrepancy* of  $\mathcal{H}$  is then defined as

$$\text{disc}(\mathcal{H}) = \min_{b: X \rightarrow \{-1, 1\}} \max_{E \in \mathcal{E}} |b(E)|,$$

where  $b(E) = \sum_{x \in E} b(x)$ . The concept of discrepancy of hypergraphs can be generalized to discrepancy of matrices in a natural way. Let  $A = (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$  be an  $m \times n$ -matrix, and set

$$\text{disc}(A) = \min_{b \in \{-1, 1\}^n} \|Ab\|_\infty.$$

This notion of discrepancy of matrices indeed contains the notion of discrepancy of hypergraphs; to see this, let  $X = \{x_1, \dots, x_n\}$  and  $\mathcal{E} = \{E_1, \dots, E_m\}$  and set  $A = (a_{ij})$ , where  $a_{ij} = 1$  whenever  $x_j \in E_i$  and  $a_{ij} = 0$  otherwise.  $A$  is called the *incidence matrix* of  $\mathcal{H}$ , and  $\text{disc}(A) = \text{disc}(\mathcal{H})$ .

There are another two closely related notions of discrepancy of a matrix. One is *linear discrepancy*, which is defined as

$$\text{lindisc}(A) = \max_{\beta \in [-1, 1]^n} \min_{b \in \{-1, 1\}^n} \|A(\beta - b)\|_\infty.$$

The second is *hereditary discrepancy*, which is defined as

$$\text{herdisc}(A) = \max_{J \subset \{1, \dots, n\}} \text{disc}((a_{ij})_{1 \leq i \leq m, j \in J}).$$

Both notions can also be used to define the respective discrepancies for hypergraphs, by identifying a hypergraph with its incidence matrix. Then  $\text{herdisc}(\mathcal{H})$  is the maximum discrepancy of all induced subgraphs of  $\mathcal{H}$ , while  $\text{lindisc}(\mathcal{H})$  gives a bound for the error for approximating reals by integers (in the spirit of inequality (11) and Lemma 1 below).

An important relation which we will need is that for any  $A$  we have

$$(8) \quad \text{lindisc}(A) \leq 2 \text{herdisc}(A);$$

this inequality is due to Beck–Spencer [7] and Lovász–Spencer–Vesztergombi [29]. More information on the discrepancy of hypergraphs and matrices can be found in the books of Chazelle [11] and Matoušek [31].

Let  $\Delta(\mathcal{H})$  denote the *maximum degree* of a hypergraph  $\mathcal{H}$ , that is,

$$\Delta(\mathcal{H}) = \max_{x \in X} \#\{E \in \mathcal{E} : x \in E\}.$$

In other words, no element  $x$  of  $X$  is contained in more than  $\Delta(\mathcal{H})$  sets  $E \in \mathcal{E}$ . The Beck–Fiala theorem states that

$$(9) \quad \text{disc}(\mathcal{H}) \leq 2\Delta(\mathcal{H}) - 1$$

for any hypergraph  $\mathcal{H}$ . Clearly the maximum degree of an induced subgraph of  $\mathcal{H}$  cannot exceed the maximum degree of  $\mathcal{H}$ . Thus by a combination of (8)

and (9) we have

$$(10) \quad \text{lindisc}(\mathcal{H}) \leq 4\Delta(\mathcal{H}) - 2$$

for any hypergraph  $\mathcal{H}$ . A direct consequence of this inequality is the following fact. Let  $\beta_1, \dots, \beta_n \in [0, 1]$  be real numbers, and let  $\mathcal{E}$  be a family of subsets of  $\{1, \dots, n\}$  such that each  $i \in \{1, \dots, n\}$  belongs to at most  $\Delta$  elements of  $\mathcal{E}$ . Then there exist integers  $b_1, \dots, b_n \in \{0, 1\}$  such that

$$(11) \quad \left| \sum_{i \in E} \beta_i - \sum_{i \in E} b_i \right| \leq 2\Delta - 1 \quad \text{for all } E \in \mathcal{E}.$$

(Note that this statement corresponds to using “colors”  $\{0, 1\}$  rather than  $\{-1, 1\}$ , which saves us a factor 2 in comparison with (10)). Inequality (11) is (in a slightly stronger form) the key ingredient in Beck’s proof of (7) in [5].

The Beck–Fiala theorem is remarkable insofar as the upper bound in (9) depends only on  $\Delta(\mathcal{H})$ , but *not* on the number of vertices  $n$  or the number of edges  $m$ . However, it turns out that we can improve (7) by using a different bound for the discrepancy of hypergraphs, which gives an improved dependence on  $\Delta(\mathcal{H})$  at the cost of an additional dependence on  $m$ . The following Lemma 1 is a consequence of a result of Banaszczyk [4], which we state first.

Banaszczyk proved the following ([4, Theorem 1 and the remark on p. 353]): Let  $K$  be a convex body in  $\mathbb{R}^m$  whose  $m$ -dimensional (standard) Gaussian measure  $\gamma_m$  is at least  $\gamma_m(K) \geq 1/2$  and which contains the origin in its interior. Then for any points  $u_1, \dots, u_n \in \mathbb{R}^m$  satisfying  $\|u_i\|_2 \leq 1/5$ ,  $1 \leq i \leq n$ , there exist signs  $b_1, \dots, b_n \in \{-1, 1\}$  such that  $b_1u_1 + \dots + b_nu_n \in K$ .

LEMMA 1. *Let real numbers  $\beta_1, \dots, \beta_n \in [0, 1]$  be given, and let  $\mathcal{E}$  be a family of subsets of  $\{1, \dots, n\}$ . Assume that each  $i \in \{1, \dots, n\}$  belongs to at most  $\Delta$  elements of  $\mathcal{E}$ , and let  $m$  denote the cardinality of  $\mathcal{E}$ . Then there exist  $b_1, \dots, b_n \in \{0, 1\}$  such that*

$$\left| \sum_{i \in E} \beta_i - \sum_{i \in E} b_i \right| \leq 5\sqrt{2\Delta \log(2m)} \quad \text{for all } E \in \mathcal{E}.$$

Here we can choose  $b_i = 0$  whenever  $\beta_i = 0$ .

The fact that Banaszczyk’s theorem implies an upper bound for the discrepancy of a hypergraph of given maximum degree is known, and is for example mentioned at the end of Section 4.3 of Matoušek’s book [31, pp. 115–116], where the upper bound  $\mathcal{O}(\sqrt{\Delta \log n})$  is given. However, we have not found any place where the value of the implied constant is explicitly stated; thus a detailed deduction of Lemma 1 is given in Section 6.

Banaszczyk’s theorem can also be applied to another interesting problem. Let  $u_1, \dots, u_n \in \mathbb{R}^m$  have Euclidean norm at most 1. Then by Banaszczyk’s

theorem there exist signs  $b_1, \dots, b_n \in \{-1, 1\}$  such that  $\|b_1 u_1 + \dots + b_n u_n\|_\infty = \mathcal{O}(\sqrt{\log m})$ . A famous conjecture of Komlos states that here the factor  $\mathcal{O}(\sqrt{\log m})$  can be replaced by an absolute constant. If the Komlos conjecture is true, then the upper bound in (9) can be replaced by  $\mathcal{O}(\sqrt{\Delta(\mathcal{H})})$  (which is known as the Beck–Fiala conjecture), the upper bound in (11) and Lemma 1 can also be replaced by  $\mathcal{O}(\sqrt{\Delta})$ , and the asymptotic order of the logarithmic terms in Theorems 1 and 2 can be improved from  $(3d + 1)/2$  and  $(3d + 4)/2$  to  $3d/2$  and  $(3d + 3)/2$ , respectively.

**3. Empirical processes.** One important ingredient in the proofs of Theorems 1 and 2 is the fact that any measure  $\mu$  on  $[0, 1]^d$  can be approximated by a discrete measure. In this section it is shown how the existence of such an approximation, together with convergence rates, can be proved by considering the empirical process of  $\mu$ -distributed independent random variables, and how large deviation bounds for empirical processes indexed by sets can be used to obtain a generalization of the *inverse of discrepancy* result (4). The results which we obtain in this section are much stronger than what would be necessary for the proofs of Theorems 1 and 2, but they are of some independent interest.

Throughout this section, let  $X_1, X_2, \dots$  be a sequence of independent, identically distributed random variables taking values in a measurable space  $(X, \mathcal{A})$  and having distribution  $P$ , and let  $\mathcal{C}$  be a class of subsets of  $X$ . Let  $P_N$  denote the empirical distribution of  $X_1, \dots, X_N$ , that is,

$$P_N = \frac{1}{N} \sum_{n=1}^N \delta_{X_n},$$

where  $\delta_Y$  denotes the Dirac measure centered on  $Y$ , and let

$$\alpha_N(\mathcal{C}) = N^{1/2}(P_N(\mathcal{C}) - P(\mathcal{C})), \quad \mathcal{C} \in \mathcal{C},$$

denote the empirical process indexed by  $\mathcal{C}$ . To avoid measurability problems we will assume throughout this paper that  $\mathcal{C}$  is countable. In this section we will be concerned with probabilistic estimates for the size of

$$(12) \quad \sup_{\mathcal{C} \in \mathcal{C}} |\alpha_N(\mathcal{C})|,$$

which of course will depend on the complexity of the class  $\mathcal{C}$ .

Let  $A$  be a finite subset of  $X$ . Then  $\mathcal{C}$  is said to *shatter*  $A$  if for every subset  $B$  of  $A$  there exists a  $C \in \mathcal{C}$  such that  $B = A \cap C$ . If there exists a largest finite number  $d$  such that  $\mathcal{C}$  shatters at least one set of cardinality  $d$ , then  $\mathcal{C}$  is called a *Vapnik–Chervonenkis class* (VC class) of index  $d$ . The notion of VC classes is well-established in the theory of empirical processes indexed by sets, since there exist strong bounds for the metric entropy of such classes. Using such entropy bounds (due to Haussler [22]), Talagrand [34, Theorem

1.1(i)] proved that there exists an absolute constant  $c_{\text{abs}}$  such that for each VC class  $\mathcal{C}$  of index  $d \geq 1$  we have

$$(13) \quad \mathbb{P}\left(\sup_{C \in \mathcal{C}} |\alpha_N(C)| \geq t\right) \leq \frac{c_{\text{abs}}}{t} \left(\frac{c_{\text{abs}} t^2}{d}\right)^d e^{-2t^2} \quad \text{for all } t > 0.$$

Talagrand’s result, which is the key ingredient in the proof of (4), is a significant improvement of earlier results of a similar type (for a comparison, see Table 1 in [37]). Regrettably, Talagrand’s paper does not give an explicit value of the constant  $c_{\text{abs}}$  in (13). Yet there do exist weaker versions of (13) which do not involve any unknown constants, such as for example the original estimate of Vapnik and Chervonenkis [36], which together with Sauer’s lemma [32] gives

$$(14) \quad \mathbb{P}\left(\sup_{C \in \mathcal{C}} |\alpha_N(C)| \geq t\right) \leq 8 \left(\frac{eN}{d}\right)^d e^{-t^2/32} \quad \text{for all } t > 0$$

(for a textbook treatment, see for example [13, Theorem 12.5]). Note that (14) is weaker than (13) in its dependence on both  $N$  and  $t$ , and could not be used to prove (4). Another completely explicit large deviations bound for empirical processes, which will be perfectly suitable for our purpose, is due to Alexander. We state it as a lemma.

LEMMA 2. *Let  $N \geq 1$  and let  $\mathcal{C}$  be a countable VC class of index  $d$ . Then for any  $t$  satisfying both*

$$(15) \quad t > \frac{2^{33/2} d}{\sqrt{N}} \log\left(\max\left(\frac{N}{2d}, e\right)\right)$$

and

$$(16) \quad t > \sqrt{2^{25} d \log 4},$$

we have

$$\mathbb{P}\left(\sup_{C \in \mathcal{C}} |\alpha_N(C)| > t\right) \leq 16e^{-t^2}.$$

REMARKS. This is a special case of [3, Theorem 2.8(ii)]. In fact, Alexander’s theorem is much more general than Lemma 2, insofar as it is formulated for a VC graph class of functions with  $(d, k)$ -constructible graph region class rather than for a VC class of sets. However, as noted in [3, p. 1049], a VC class of index  $d$  is always a VC graph class with  $(d, 1)$ -constructible graph region class; moreover, every result formulated for a class of functions is clearly also applicable for a class of sets (by considering the indicator functions of the sets). Furthermore, for the parameters in Alexander’s theorem we have chosen  $\psi = 2M^2$  (which is permitted by the remark after [3, equation (1.7)]) and  $\varepsilon = 1/2$ . Since our functions  $f$  are of the form  $\mathbb{1}_C(X_n)$  for some sets  $C$  we have  $0 \leq f(X_n) \leq 1$ , and hence the variance of  $f(X_n)$  is at most  $1/4$ ; consequently, we can choose  $\alpha = 1/4$ . Finally, the estimate in

Alexander’s theorem is for the outer measure  $\mathbb{P}^*$  rather than the measure  $\mathbb{P}$ ; however, by assuming that  $\mathcal{C}$  is countable we can avoid all measurability problems, which means that the set  $\{\sup |\alpha_N| > t\}$  is measurable and we can use the measure  $\mathbb{P}$ .

As mentioned before, Talagrand’s inequality (13) is the key ingredient in the proof of (4), together with the well-known fact that the class of axis-parallel boxes in  $[0, 1]^d$  which have one vertex at the origin is a VC class of index  $d$ . However, as observed by Heinrich et al., Talagrand’s inequality (13) can be used to obtain a much more general *inverse of star-discrepancy* result for VC classes. Using the notation from above, let  $t_1, \dots, t_N$  be points in  $X$  and set

$$D_N^{(\mathcal{C}, P)}(t_1, \dots, t_N) = \sup_{C \in \mathcal{C}} \left| \frac{1}{N} \sum_{n=1}^N \mathbb{1}_C(t_n) - P(C) \right|.$$

Assume that  $\mathcal{C}$  is a VC class of index  $d$ . Then by (13) there exist points  $t_1, \dots, t_N \in X$  such that

$$(17) \quad D_N^{(\mathcal{C}, P)}(t_1, \dots, t_N) \leq c_{\text{abs}} \frac{\sqrt{d}}{\sqrt{N}},$$

which means that there always exists a point set of cardinality at most  $c_{\text{abs}} d \varepsilon^{-2}$  achieving a  $D_N^{(\mathcal{C}, P)}$ -discrepancy of at most  $\varepsilon$ . Since Talagrand’s inequality contains an unspecified constant, the same is the case for (17). However, as we will see (in the proof of Theorem 3 in Section 6), using Lemma 2 instead of (13) we can obtain a fully explicit version of (17), which we state as the following theorem.

**THEOREM 3.** *Let  $(X, \mathcal{A}, P)$  be a probability space, and let  $\mathcal{C}$  be a class of subsets of  $X$  which is a countable VC class of index  $d$ . Then for any  $N \geq 1$  there exist points  $t_1, \dots, t_N \in X$  for which*

$$D_N^{(\mathcal{C}, P)}(t_1, \dots, t_N) \leq 2^{13} \frac{\sqrt{d}}{\sqrt{N}}.$$

*Consequently, for any  $\varepsilon \in (0, 1)$  there exists a set of at most  $N(\varepsilon) = 2^{26} d \varepsilon^{-2}$  points whose  $D_N^{(\mathcal{C}, P)}$ -discrepancy does not exceed  $\varepsilon$ .*

Theorem 3 is the first fully explicit *inverse of discrepancy* type result for VC classes. However, the value of the constants in Theorem 3 is not very satisfactory, given the fact that in inequality (4) (that is, in the case of star-discrepancy on  $[0, 1]^d$ ) we may choose  $c_{\text{abs}} = 100$  (see [1]). One way for a possible improvement of the constants in Theorem 3 would of course be to try to find an explicit version of Talagrand’s inequality (13). Another possible way would be to use Massart’s [30] explicit version of some other large deviations bounds for empirical processes due to Talagrand [35]; however,



these bounds involve the quantity

$$(18) \quad \mathbb{E} \left( \sup_{C \in \mathcal{C}} |\alpha_N(C)| \right),$$

which depends on the entropy of the class  $\mathcal{C}$ . Using Haussler’s entropy bounds for VC classes and the method for estimating (18) which is indicated in [18, Section 5.2] it might be possible to obtain an improved version of Theorem 3. For the problem of estimating (18) in the context of low-discrepancy point sets on  $[0, 1]^d$  see [2, 16].

To conclude this section, we repeat an argument from [23] to show how to apply the results obtained for the discrepancy  $D_N^{(\mathcal{C}, P)}$  in Theorem 3 to the case of the discrepancy  $D_N^*(\cdot; \mu)$ , as defined in (6). Clearly, using the notation from the present section, we have to choose  $X = [0, 1]^d$  and  $P = \mu$ . However, we cannot take for  $\mathcal{C}$  the class of all axis-parallel boxes in  $[0, 1]^d$  having one vertex at the origin, as this class is obviously uncountable. Instead we take for  $\mathcal{C}$  the class of all such boxes which have only rational coordinates. Restricting the test sets  $A$  in definition (6) to boxes having only rational coordinates does not change the value of star-discrepancy. Consequently, with these settings the discrepancies  $D_N^{(\mathcal{C}, P)}$  and  $D_N^*(\cdot; \mu)$  coincide, and Theorem 3 includes an *inverse of star-discrepancy* result for the discrepancy  $D_N^*(\cdot; \mu)$ , which is stated as a corollary below.

**COROLLARY 1.** *Let  $\mu$  be any non-negative, normalized Borel measure on  $[0, 1]^d$ . Then for any  $N \geq 1$  there exist points  $x_1, \dots, x_N \in [0, 1]^d$  such that*

$$D_N^*(x_1, \dots, x_N; \mu) \leq 2^{13} \frac{\sqrt{d}}{\sqrt{N}}.$$

Furthermore, for any  $\varepsilon \in (0, 1)$  there exists a set of at most  $N(\varepsilon) = 2^{26} d \varepsilon^{-2}$  points whose star-discrepancy with respect to  $\mu$  does not exceed  $\varepsilon$ .

**4. Numerical integration of discontinuous functions.** The Koksma–Hlawka inequality (2) can be rewritten for smooth functions  $f$  on  $[0, 1]^d$  in the form

$$(19) \quad \left| \frac{1}{N} \sum_{n=1}^N f(x_n) - \int_{[0,1]^d} f(x) dx \right| \leq D_N^*(x_1, \dots, x_N) V_{1,1}(f),$$

where

$$(20) \quad V_{1,1}(f) = \sum_{\mathbf{u} \subset \{1, \dots, d\}} \int_{[0,1]^{|\mathbf{u}|}} \left| \frac{\partial^{|\mathbf{u}|} f}{\partial x_{\mathbf{u}}} (x_{\mathbf{u}}, 1) \right| dx_{\mathbf{u}}.$$

Here the sum is extended over all subsets  $\mathbf{u}$  of  $\{1, \dots, d\}$ . The symbol  $x_{\mathbf{u}}$  means that only those coordinates of  $x$  are considered whose index is con-

tained in  $\mathbf{u}$ , and  $(x_{\mathbf{u}}; 1)$  means that all those components of  $x$  whose index is not contained in  $\mathbf{u}$  are replaced by 1. The quantity  $V_{1,1}(f)$  is the norm corresponding to an inner product in a certain reproducing kernel Hilbert space of Sobolev type; for more details, see for example [14, Chapter 3] or [15, Chapter 2].

The inequality (19) allows a nice function-analytic interpretation of the error of quasi-Monte Carlo integration for smooth functions; however, it is not of any use if the function  $f$  has discontinuities. The Koksma–Hlawka inequality in the form (2) is slightly more flexible, since it also yields upper bounds for the numerical integration of discontinuous functions. In particular, this is the case if  $f$  is the indicator function of an axis-parallel box having one vertex at the origin (which is not surprising in view of the definition of star-discrepancy). However, the variation of a function  $f$  in the sense of Hardy and Krause can be infinity even if  $f$  is a rather simple function, for example the indicator function of a polytope which may have non-axis-parallel faces.

Let  $\Omega$  be a subset of  $[0, 1]^d$ , and assume that  $f(x)$  is of the form

$$(21) \quad g(x) \cdot \mathbb{1}_{\Omega}(x),$$

where  $g$  is a smooth function. Let  $x_1, \dots, x_N$  be points in  $[0, 1]^d$ . Can we then obtain a bound for the QMC integration error

$$(22) \quad \left| \frac{1}{N} \sum_{n=1}^N f(x_n) - \int_{[0,1]^d} f(x) dx \right|$$

in terms of an (appropriately defined) discrepancy of  $x_1, \dots, x_N$  which holds *uniformly* for all possible sets  $\Omega$ ? Probably not without strong a priori restrictions on the class of sets in which  $\Omega$  has to lie. For example, even if we set  $g \equiv 1$  but allow  $\Omega$  to be any subset of  $[0, 1]^d$ , there will always be a choice of  $\Omega$  such that the error in (22) equals 1 (namely if we take  $\Omega = [0, 1]^d \setminus \{x_1, \dots, x_N\}$ ). Also if we impose some restrictions on  $\Omega$  we will not necessarily get any good bounds for the QMC error (22): For example, if  $g \equiv 1$  and  $\Omega$  can be any convex set, then independently of the point set  $x_1, \dots, x_N$  the error in (22) will not be better than  $\gg N^{-2/(d+1)}$  for some choice of  $\Omega$  (a variant of the Koksma–Hlawka inequality for functions having convex super-level sets has actually recently been proved by Harman [20]). In other words, we cannot expect to find a point set which gives good error bounds for QMC integration for functions of the form (21) which hold *uniformly* for all sets  $\Omega$  from a large class of subsets of  $[0, 1]^d$ .

However, the problem of numerical integration of functions having discontinuities cannot be entirely neglected, since such functions arise in a natural way in one of the main areas of applications of QMC methods. More

precisely, QMC methods are a standard tool for pricing financial derivatives in financial mathematics, and there are classes of such derivatives, for instance digital options and barrier options, which have a discontinuous payoff function. In a standard model of financial mathematics the value of the underlying asset is assumed to follow a geometric Brownian motion, and the value of a financial derivative is given by the expected value of the payoff under the risk-neutral measure. The Brownian motion is discretized into  $d$  intervals, and points in  $[0, 1]^d$  are used, together with a suitable path-generating method, to simulate a path of the geometric Brownian motion. Accordingly, the payoff of the financial derivative can be estimated by the  $d$ -dimensional integral of a function  $f$ , where the properties of  $f$  depend both on the payoff of the derivative and on the path-generating method used. For the basic principles of this method see Glasserman [19]. The influence of the path-generating method on the effectiveness of QMC integration has been studied in detail by Wang and Tan [38, 39].

We will show that once we assume that the set  $\Omega$  (where the discontinuities occur) is *fixed*, then we can find QMC point sets with asymptotic convergence rate almost as good as (3). The key ingredient (besides Theorem 1) is a recent variant of the Koksma–Hlawka inequality for general integration domains, which is due to Brandolini, Colzani, Gigante and Travaglini [10]. They proved (amongst many other results) the following inequality, which we state as a lemma.

LEMMA 3. *Let  $f$  be a continuous function on  $\mathbb{T}^d$ , and let  $\Omega$  be a Borel subset of  $[0, 1]^d$ . Then for  $x_1, \dots, x_N$  in  $[0, 1]^d$  we have*

$$\left| \frac{1}{N} \sum_{n=1}^N (f \cdot \mathbb{1}_\Omega)(x_n) - \int_{\Omega} f(x) \, dx \right| \leq D_N^\Omega(x_1, \dots, x_N) V(f),$$

where

$$(23) \quad D_N^{(\Omega)}(x_1, \dots, x_N) = 2^d \sup_{A \subset [0, 1]^d} \left| \frac{1}{N} \sum_{n=1}^N \mathbb{1}_{\Omega \cap A}(x_n) - \lambda(\Omega \cap A) \right|,$$

$$(24) \quad V(f) = \sum_{\mathbf{u} \subset \{1, \dots, d\}} 2^{d-|\mathbf{u}|} \int_{[0, 1]^d} \left| \frac{\partial^{|\mathbf{u}|}}{\partial x_{\mathbf{u}}} f(x) \right| dx;$$

here  $\frac{\partial^{|\mathbf{u}|}}{\partial x_{\mathbf{u}}} f(x)$  is the partial derivative of  $f$  with respect to those components of  $x$  with index in  $\mathbf{u}$ , and the supremum is taken over all axis-parallel boxes  $A$ .

As a consequence of Lemma 3, together with Theorems 1 and 3, we get the following result. The proof is given at the end of Section 6.

THEOREM 4. *Let  $f$  be a continuous function on  $\mathbb{T}^d$ , and let  $\Omega$  be a Borel subset of  $[0, 1]^d$  which has positive  $d$ -dimensional Lebesgue measure  $\lambda$ . Then*

for any  $N \geq 1$  there exist points  $x_1, \dots, x_N$  in  $[0, 1]^d$  for which

$$(25) \quad \left| \frac{1}{N} \sum_{n=1}^N f(x_n) - \frac{1}{\lambda(\Omega)} \int_{\Omega} f(x) d\lambda(x) \right| \leq \frac{(63\sqrt{d} 4^d (2 + \log_2 N)^{(3d+1)/2} + 1)V(f)}{N} + \frac{\|f\|_{\infty}}{N},$$

where  $V(f)$  is defined in (24).

The purpose of Theorem 4 is to show that *in principle* there exist QMC point sets which yield fast asymptotic convergence rates for QMC integration for functions of the form (21). Note, however, that the point set  $x_1, \dots, x_N$  in Theorem 4 depends on  $\Omega$ , and consequently we would have to construct *different* point sets for functions having their discontinuities at different positions (which is not surprising, as noted at the beginning of this section). Thus for calculating the risk-neutral price of a financial derivative with discontinuous payoff the point set yielding small QMC integration error has to be chosen with respect to both the barrier (where the discontinuity occurs) and the path-generating method (but *not* depending on the value of the payoff). Additionally, equation (25) contains the term  $\lambda(\Omega)$ , which might be unknown. Furthermore the proof of Theorem 1 is of a purely probabilistic/combinatorial nature and *not* constructive, so it remains an open problem how to effectively construct such point sets (possibly in a parametric form, depending on  $\Omega$ ). Thus Theorem 4 is far away from a possible implementation, and is mostly of theoretical interest.

Note that the assumption  $\lambda(\Omega) > 0$  in Theorem 4 is no real restriction, since in the case  $\lambda(\Omega) = 0$  we clearly have

$$\int_{\Omega} f(x) d\lambda(x) = 0$$

for any function  $f(x)$ , which makes numerical integration pointless. Note also that Theorem 4 is *not* a trivial consequence of Theorem 1, combined with Lemma 3, since the two notions of discrepancy in (6) and (23) are defined in a different way. We will choose  $\mu$  in such a way that

$$(26) \quad \mu(A) = \frac{\lambda(\Omega \cap A)}{\lambda(\Omega)}.$$

Then  $\mu$  is a normalized measure. However, the discrepancy defined in (6) is then given by

$$(27) \quad \sup_{A \subset [0,1]^d} \left| \frac{1}{N} \sum_{n=1}^N \mathbb{1}_A(x_n) - \frac{\lambda(\Omega \cap A)}{\lambda(\Omega)} \right|,$$

which is not the same as (23); in particular, these discrepancies are only

comparable for our point set from Theorem 1 if we can guarantee that all points constructed there for the measure  $\mu$  given by (26) are contained in  $\Omega$  (which is actually the case, as the proof of Theorem 4 will show).

**5. Open problems.** As mentioned in (3), for any  $d \geq 1$  and  $N \geq 1$  there exists a (finite) point set  $x_1, \dots, x_N$  whose star-discrepancy (with respect to the uniform measure on  $[0, 1]^d$ ) is bounded by

$$(28) \quad D_N^*(x_1, \dots, x_N) \ll (\log N)^{d-1} N^{-1};$$

here the implied constant depends on  $d$ . Furthermore, there exists an infinite sequence  $(x_n)_{n \geq 1}$  of points from  $[0, 1]^d$  whose star-discrepancy is bounded by

$$(29) \quad D_N^*(x_1, \dots, x_N) \ll (\log N)^d N^{-1} \quad \text{for all } N \geq 1,$$

where again the implied constant depends on  $d$ . It is unknown if these upper bounds are optimal, and the problem of finding the optimal convergence rates in these inequalities is called the *Great Open Problem* of discrepancy theory (cf. [6, Conjecture, p. 6]). The best known lower bound is

$$D_N^*(x_1, \dots, x_N) \gg (\log N)^{(d-1)/2+\eta(d)}$$

for some  $\eta(d) > 0$ , and for any (finite) set of  $N$  points in  $[0, 1]^d$ , due to Bilyk, Lacey and Vagharshakyan [9] (see [8] for a survey). Obviously the upper bounds in Theorems 1 and 2 are of asymptotic order

$$(30) \quad D_N^*(x_1, \dots, x_N; \mu) \ll (\log N)^{(3d+1)/2} N^{-1}$$

for a finite point set  $x_1, \dots, x_N$ , and

$$(31) \quad D_N^*(x_1, \dots, x_N; \mu) \ll (\log N)^{(3d+4)/2} N^{-1} \quad \text{for all } N \geq 1$$

for an infinite sequence  $(x_n)_{n \geq 1}$ ; it would be very interesting to know if the asymptotic upper bounds in these inequalities can be improved to (28) and (29), respectively, or, if this is not the case for general  $\mu$ , to classify those measures  $\mu$  for which it is possible. As mentioned in Section 2, a proof of the Komlos conjecture (or the Beck–Fiala conjecture) would imply that the exponents of the logarithmic terms in (30) and (31) could both be reduced by 1/2, which would still leave a gap to (28) and (29).

As mentioned in Section 1, the problem concerning the inverse of the discrepancy is open even in the case of the “classical” star-discrepancy (that is, with respect to the uniform measure). As noted in Section 3, the upper bound (4) remains valid if the uniform measure is replaced by another measure on  $[0, 1]^d$ ; on the other hand, Hinrichs’ proof of the lower bound (5) depends on the metric entropy of the system of axis-parallel boxes *with respect to the uniform measure*, and can (as far as we understand) not easily be generalized to other measures. Thus many problems concerning the inverse

of the discrepancy for the star-discrepancy  $D_N^*$ , the discrepancy  $D_N^*(\cdot; \mu)$  and the discrepancy  $D_N^{(C,P)}$  remain open. As noted after the statement of Theorem 3, it would also be interesting to get an improved version of this theorem with a significantly smaller value for the numerical constant, in the general case of the inverse of discrepancy with respect to general measures and VC classes.

Generalizing and improving Lemma 3 of Brandolini, Colzani, Gigante and Travaglini would be very interesting. In particular, it would be helpful to get rid of the terms which are exponential in  $d$ , and which spoil a possible application of the point set from Corollary 1 in order to obtain a version of Theorem 4 which is suitable for point sets which have moderate cardinality  $N$  in comparison with  $d$ . Furthermore, it would be nice to have a version of Lemma 3 which is formulated for the discrepancy in (27) rather than (23); such a Koksma–Hlawka inequality would be more convenient in the case when  $\Omega$  is fixed and has small volume, and could for example render the laborious technical argument in the proof of Theorem 4 unnecessary.

**6. Proofs.** In this section we give the proofs of Theorems 1–4. We start by showing that Theorem 3 follows from Lemma 2. Subsequently we give a proof of Lemma 1 from Section 2, which, together with Corollary 1 (a corollary of Theorem 3) and Lemmas 4 and 5 below, will be used to prove Theorems 1 and 2. The proofs for Theorems 1 and 2 are similar to those given for (7) by Beck in [5], the only major differences being that we use Lemma 1 instead of (11), and carefully take into account the dependence of all the quantities involved on the dimension  $d$ . Finally, we use Theorems 1 and 3 and Lemmas 3 and 5 to prove Theorem 4.

*Proof of Theorem 3.* Theorem 3 follows easily from Lemma 2. In fact, using Lemma 2 for  $t = 2^{13}\sqrt{d}$ , we immediately see that condition (16) is satisfied. On the other hand, for our choice of  $t$  condition (15) is satisfied if

$$2^{13}\sqrt{d} > \frac{2^{33/2}d}{\sqrt{N}} \log\left(\max\left(\frac{N}{2d}, e\right)\right),$$

which is equivalent to

$$(32) \quad \frac{\sqrt{N}}{\sqrt{d}} > 2^{7/2} \log\left(\max\left(\frac{N}{2d}, e\right)\right).$$

Equation (32) holds whenever  $\sqrt{N} \geq 96\sqrt{d}$ . Thus, using Lemma 2 under the additional assumption that  $\sqrt{N} \geq 96\sqrt{d}$  we get

$$\mathbb{P}\left(\sup_{C \in \mathcal{C}} |\alpha_N(C)| > 2^{13}\sqrt{d}\right) \leq 16e^{-2^{26}d} < 1,$$

which proves the existence of points  $t_1, \dots, t_N \in X$  for which

$$D_N^{(C,P)}(t_1, \dots, t_N) \leq 2^{13} \frac{\sqrt{d}}{\sqrt{N}}.$$

On the other hand, in the case  $\sqrt{N} < 96\sqrt{d}$  the conclusion of Theorem 3 holds trivially, since  $D_N^{(C,P)}$  is always bounded by 1. This proves Theorem 3. ■

*Proof of Lemma 1.* The density of the  $m$ -dimensional standard Gaussian measure is given by

$$\gamma_m(y_1, \dots, y_m) = (2\pi)^{-m/2} \exp\left(-\frac{y_1^2 + \dots + y_m^2}{2}\right), \quad (y_1, \dots, y_m) \in \mathbb{R}^m.$$

Consequently, for  $K = [-\sqrt{2\log(2m)}, \sqrt{2\log(2m)}]^m$  we have

$$\begin{aligned} \gamma_m(K) &= \left(\frac{1}{\sqrt{2\pi}} \int_{-\sqrt{2\log(2m)}}^{\sqrt{2\log(2m)}} e^{-y^2/2} dy\right)^m \\ &\geq \left(1 - \frac{2}{\sqrt{2\pi}} \int_{\sqrt{2\log(2m)}}^{\infty} \frac{y}{\sqrt{2\log(2m)}} e^{-y^2/2} dy\right)^m \\ &= \left(1 - \frac{1}{\sqrt{\pi \log(2m)}} e^{-\log(2m)}\right)^m = \left(1 - \frac{1}{2m\sqrt{\pi \log(2m)}}\right)^m \geq 1/2 \end{aligned}$$

(note that  $\sqrt{\pi \log(2m)} \geq 1$  for  $m \geq 1$ ). Let  $X = \{1, \dots, n\}$  and the family  $\mathcal{E}$  of subsets of  $X$  be given, and write  $\mathcal{H} = (X, \mathcal{E})$ . Then by the assumptions of the lemma the hypergraph  $\mathcal{H}$  has maximum degree at most  $\Delta$ , and consequently the incidence matrix  $A$  of  $\mathcal{H}$  has column vectors  $u_1, \dots, u_n$  for which  $\|u_i\|_2 \leq \sqrt{\Delta}$ ,  $1 \leq i \leq n$ . Thus by Banaszczyk’s theorem there exist colors  $b_1, \dots, b_n \in \{-1, 1\}$  such that

$$\|b_1 u_1 + \dots + b_n u_n\|_\infty \leq 5\sqrt{2\Delta \log(2m)}.$$

In other words, for the discrepancy of  $\mathcal{H}$  we have

$$\text{disc}(\mathcal{H}) \leq 5\sqrt{2\Delta \log(2m)}.$$

Since the maximum degree of any induced subgraph of  $\mathcal{H}$  is also at most  $\Delta$ , together with (8) this yields

$$\text{lindisc}(\mathcal{H}) \leq 10\sqrt{2\Delta \log(2m)},$$

which implies Lemma 1 (note again that in Lemma 1 we use “colors”  $\{0, 1\}$  instead of  $\{-1, 1\}$ , which saves us a factor 2). The fact that we can choose  $b_i = 0$  whenever  $\beta_i = 0$  is trivial (we can simply ignore all vertices  $i$  for which  $\beta_i = 0$  in the application of Banaszczyk’s theorem). This proves Lemma 1. ■

As a consequence of Lemma 1, we get the following result.

LEMMA 4. For  $(i_1, \dots, i_d) \in \{1, \dots, N\}^d$ , let  $\beta_{(i_1, \dots, i_d)} \in [0, 1]$  be given. Then there exist  $b_{(i_1, \dots, i_d)} \in \{0, 1\}$  such that  $b_{(i_1, \dots, i_d)} = 0$  whenever  $\beta_{(i_1, \dots, i_d)} = 0$ , and such that for all  $(J_1, \dots, J_d) \in \{1, \dots, N\}^d$  we have

$$\left| \sum_{i_1=1}^{J_1} \cdots \sum_{i_d=1}^{J_d} (b_{(i_1, \dots, i_d)} - \beta_{(i_1, \dots, i_d)}) \right| \leq 10\sqrt{d} (2 + \log_2 N)^{(3d+1)/2}.$$

*Proof.* For simplicity, we assume that  $N = 2^m$  for some integer  $m \geq 1$  (this assumption will be dropped at the end of the proof). We use a classical dyadic decomposition method. Let  $m_1, \dots, m_d$  be non-negative integers such that  $0 \leq m_1, \dots, m_d \leq m$ , and let  $\mathcal{E}(m_1, \dots, m_d)$  denote the class of all sets of the form

$$\prod_{s=1}^d \{j_s 2^{m_s} + 1, \dots, (j_s + 1) 2^{m_s}\}$$

(where the product  $\prod$  denotes the Cartesian product of sets) for  $0 \leq j_s < 2^{m-m_s}$ ,  $1 \leq s \leq d$ . Then for any fixed  $m_1, \dots, m_d$  the class  $\mathcal{E}(m_1, \dots, m_d)$  forms a partition of  $\{1, \dots, N\}^d$ , which implies that for fixed  $m_1, \dots, m_d$  any  $(i_1, \dots, i_d)$  appears in only one set of  $\mathcal{E}(m_1, \dots, m_d)$ . On the other hand, for fixed  $m_1, \dots, m_d$  the class  $\mathcal{E}(m_1, \dots, m_d)$  consists of at most  $N^d$  elements. The number of possible values for  $(m_1, \dots, m_d)$  is  $(m+1)^d$ . Thus by Lemma 1 there exist  $b_{(i_1, \dots, i_d)} \in \{0, 1\}$  such that for any set  $E$  which is contained in  $\mathcal{E}(m_1, \dots, m_d)$  for some  $m_1, \dots, m_d$  we have

$$\begin{aligned} \left| \sum_{(i_1, \dots, i_d) \in E} (b_{(i_1, \dots, i_d)} - \beta_{(i_1, \dots, i_d)}) \right| &\leq 5\sqrt{2(m+1)^d \log(2(m+1)^d N^d)} \\ &\leq 10\sqrt{d(m+1)^d \log(2N)}, \end{aligned}$$

and such that additionally  $b_{(i_1, \dots, i_d)} = 0$  whenever  $\beta_{(i_1, \dots, i_d)} = 0$ .

For arbitrary  $J_1, \dots, J_d$  from  $\{1, \dots, N\}$  the set

$$\prod_{s=1}^d \{1, \dots, J_s\}$$

can be written as a disjoint union of sets from

$$\bigcup_{0 \leq m_s \leq m \text{ for } 1 \leq s \leq d} \mathcal{E}(m_1, \dots, m_d),$$

in such a way that we take at most one set from  $\mathcal{E}(m_1, \dots, m_d)$  for any possible  $m_1, \dots, m_d$  (such a representation is easily found from the binary



representation of the numbers  $J_1, \dots, J_d$ ). Consequently, we have

$$(33) \quad \left| \sum_{(i_1, \dots, i_d) \in \prod_{s=1}^d \{1, \dots, J_s\}} (b_{(i_1, \dots, i_d)} - \beta_{(i_1, \dots, i_d)}) \right| \leq 10\sqrt{d(m+1)^d \log(2N)} (m+1)^d$$

for any  $J_1, \dots, J_d$ . Finally, assume that  $N$  is not an integral power of 2. Then we write  $\hat{N}$  for the smallest number exceeding  $N$  which is an integral power of 2, and set

$$b_{(i_1, \dots, i_d)} = 0, \quad \beta_{(i_1, \dots, i_d)} = 0 \quad \text{for } (i_1, \dots, i_d) \in \{1, \dots, \hat{N}\}^d \setminus \{1, \dots, N\}^d.$$

Then from (33) we get

$$\begin{aligned} \max_{\substack{J_1, \dots, J_d: \\ 1 \leq J_s \leq \hat{N} \text{ for } 1 \leq s \leq d}} \left| \sum_{(i_1, \dots, i_d) \in \prod_{s=1}^d \{1, \dots, J_s\}} (b_{(i_1, \dots, i_d)} - \beta_{(i_1, \dots, i_d)}) \right| &\leq 10\sqrt{d(\lceil \log_2 N \rceil + 1)^d \log(4N)(\lceil \log_2 N \rceil + 1)^d} \\ &\leq 10\sqrt{d} (2 + \log_2 N)^{(3d+1)/2}, \end{aligned}$$

which proves Lemma 4. ■

LEMMA 5. *Given any (not necessarily distinct) points  $z_1, \dots, z_K$  in  $[0, 1]^d$  and any number  $N$  satisfying  $N \leq \sqrt{K}$  we can find an  $N$ -element subset  $\{x_1, \dots, x_N\}$  of  $\{z_1, \dots, z_K\}$  such that for any axis-parallel box  $A \subset [0, 1]^d$  which has one vertex at the origin we have*

$$\left| \sum_{n=1}^N \mathbb{1}_A(x_n) - \frac{N}{K} \sum_{n=1}^K \mathbb{1}_A(z_n) \right| \leq 60\sqrt{d} (2 + \log_2 N)^{(3d+1)/2} + 4d + 6.$$

REMARK. Since we do not assume that the points  $z_1, \dots, z_K$  are distinct, the sets in the statement and in the proof of this lemma are strictly speaking multisets. However, to keep the presentation short, we simply neglect this minor technicality.

*Proof.* We can rearrange the points  $z_1, \dots, z_K$  in  $d$  different ways in increasing order, according to their  $d$  coordinates. More precisely, for each  $s \in \{1, \dots, d\}$  we define  $z_1^{(s)}, \dots, z_K^{(s)}$  in such a way that

$$\{z_1^{(s)}, \dots, z_K^{(s)}\} = \{z_1, \dots, z_K\}$$

and so that the  $s$ th coordinate of  $z_m^{(s)}$  does not exceed the  $s$ th coordinate of  $z_n^{(s)}$  for  $m \leq n$ .

For  $1 \leq s \leq d$  and  $1 \leq i \leq N$  we set

$$F_i^{(s)} = \{z_n^{(s)} : \lfloor (i-1)K/N \rfloor < n \leq \lfloor iK/N \rfloor\},$$

and, for  $(i_1, \dots, i_d) \in \{1, \dots, N\}^d$ ,

$$\beta_{(i_1, \dots, i_d)} = \frac{N}{K + N} \#\left\{n : z_n \in \bigcap_{s=1}^d F_{i_s}^{(s)}\right\}.$$

By Lemma 4 there exist  $b_{(i_1, \dots, i_d)} \in \{0, 1\}$  such that for all  $(J_1, \dots, J_d) \in \{1, \dots, N\}^d$  we have

$$\left| \sum_{i_1=1}^{J_1} \dots \sum_{i_d=1}^{J_d} (b_{(i_1, \dots, i_d)} - \beta_{(i_1, \dots, i_d)}) \right| \leq 10\sqrt{d} (2 + \log_2 N)^{(3d+1)/2},$$

and such that  $b_{(i_1, \dots, i_d)} = 0$  whenever  $\beta_{(i_1, \dots, i_d)} = 0$ . Now we select one point  $x_{(i_1, \dots, i_d)}$  from  $\prod_{s=1}^d F_{i_s}^{(s)}$  whenever  $b_{(i_1, \dots, i_d)} = 1$ , and write  $\mathcal{P}$  for the set consisting of all such points.

Let  $J_1, \dots, J_d \in \{1, \dots, N\}$  be fixed, and set

$$G = G_{(J_1, \dots, J_d)} = \bigcup_{1 \leq i_1 \leq J_1} \dots \bigcup_{1 \leq i_d \leq J_d} \left( \bigcap_{s=1}^d F_{i_s}^{(s)} \right).$$

Then we have

$$\begin{aligned} (34) \quad & \left| \#(\mathcal{P} \cap G) - \frac{N}{K} \#(G) \right| \\ & \leq \left| \#(\mathcal{P} \cap G) - \frac{N}{K + N} \#(G) \right| + \underbrace{\left| \left( \frac{N}{K} - \frac{N}{K + N} \right) \#(G) \right|}_{\leq N^2 / (K + N) \leq 1} \\ & \leq \left| \sum_{i_1=1}^{J_1} \dots \sum_{i_d=1}^{J_d} (b_{(i_1, \dots, i_d)} - \beta_{(i_1, \dots, i_d)}) \right| + 1 \\ & \leq 10\sqrt{d} (2 + \log_2 N)^{(3d+1)/2} + 1. \end{aligned}$$

In particular, by choosing  $J_1 = \dots = J_d = N$ , we see that the cardinality of  $\mathcal{P}$  differs from  $N$  by at most  $10\sqrt{d} (2 + \log_2 N)^{(3d+1)/2} + 1$ . Consequently, there exists an  $N$ -element subset  $\mathcal{Q}$  of  $\{z_1, \dots, z_K\}$  such that

$$|\#(\mathcal{P}) - \#(\mathcal{Q})| \leq 10\sqrt{d} (2 + \log_2 N)^{(3d+1)/2} + 1,$$

and by (34) we have

$$(35) \quad \left| \#(\mathcal{Q} \cap G) - \frac{N}{K} \#(G) \right| \leq 20\sqrt{d} (2 + \log_2 N)^{(3d+1)/2} + 2$$

for all  $G = G_{(J_1, \dots, J_d)}$ . We will show that  $\mathcal{Q}$  satisfies the conclusion of Lemma 5.

Let  $(a_1, \dots, a_d) \in [0, 1]^d$  be given, and let  $A$  denote the axis-parallel box  $\prod_{s=1}^d [0, a_s]$ . Then there exist  $J_1, \dots, J_d \in \{1, \dots, N - 1\}$  such that

$$G_{(J_1, \dots, J_d)} \subset \{z_n : z_n \in A\} \subset G_{(J_1 + 1, \dots, J_d + 1)}.$$

Furthermore, since  $\#(G(J_1 + 1, \dots, J_d + 1) \setminus G(J_1, \dots, J_d)) \leq d(K/N + 1) \leq 2dK/N$  we have

$$(36) \quad \#(\{z_n : z_n \in A\} \setminus G(J_1, \dots, J_d)) \leq 2dK/N,$$

and consequently, by (35),

$$(37) \quad \begin{aligned} & |\#(\mathcal{Q} \cap A) - \#(\mathcal{Q} \cap G(J_1, \dots, J_d))| \\ & \leq |\#(\mathcal{Q} \cap G(J_1 + 1, \dots, J_d + 1)) - \#(\mathcal{Q} \cap G(J_1, \dots, J_d))| \\ & \leq \left| \frac{N}{K} \left( \#(G(J_1 + 1, \dots, J_d + 1)) - \#(G(J_1, \dots, J_d)) \right) \right| \\ & \quad + 40\sqrt{d} (2 + \log_2 N)^{(3d+1)/2} + 4 \\ & \leq 2d + 40\sqrt{d} (2 + \log_2 N)^{(3d+1)/2} + 4. \end{aligned}$$

Thus, by (35), (36) and (37) we have

$$\begin{aligned} & \left| \sum_{n=1}^N \mathbb{1}_A(x_n) - \frac{N}{K} \sum_{n=1}^K \mathbb{1}_A(z_n) \right| \\ & = \left| \#(\mathcal{Q} \cap A) - \frac{N}{K} \#\{n : z_n \in A\} \right| \\ & \leq \left| \#(\mathcal{Q} \cap A) - \#(\mathcal{Q} \cap G(J_1, \dots, J_d)) \right| \\ & \quad + \left| \#(\mathcal{Q} \cap G(J_1, \dots, J_d)) - \frac{N}{K} \#(G(J_1, \dots, J_d)) \right| \\ & \quad + \left| \frac{N}{K} (\#(G(J_1, \dots, J_d)) - \#\{n : z_n \in A\}) \right| \\ & \leq 2d + 40\sqrt{d} (2 + \log_2 N)^{(3d+1)/2} + 4 + 20\sqrt{d} (2 + \log_2 N)^{(3d+1)/2} + 2 + 2d \\ & \leq 60\sqrt{d} (2 + \log_2 N)^{(3d+1)/2} + 4d + 6. \end{aligned}$$

This proves Lemma 5. ■

*Proof of Theorem 1.* All that is left for the proof of Theorem 1 is to show that we can closely approximate the measure  $\mu$  by a discrete measure of the form  $K^{-1} \sum_{n=1}^K \delta_{z_n}$  for some  $z_1, \dots, z_K$ . Such points exist by Corollary 1: For a given  $N$  we can choose  $K = 2^{26}dN^2$  points  $z_1, \dots, z_K \in [0, 1]^d$  such that

$$(38) \quad \sup_{A \subset [0,1]^d} \left| \frac{1}{K} \sum_{n=1}^K \mathbb{1}_A(z_n) - \mu(A) \right| \leq \frac{1}{N},$$

where the supremum is extended over all axis-parallel boxes which have one vertex at the origin. Then by Lemma 5 there exist points  $x_1, \dots, x_N$  such

that

$$\begin{aligned} & \sup_{A \subset [0,1]^d} \left| \frac{1}{N} \sum_{n=1}^N \mathbb{1}_A(x_n) - \mu(A) \right| \\ & \leq \sup_{A \subset [0,1]^d} \left| \frac{1}{N} \sum_{n=1}^N \mathbb{1}_A(x_n) - \frac{1}{K} \sum_{n=1}^K \mathbb{1}_A(z_n) \right| + \frac{1}{N} \\ & \leq \frac{60\sqrt{d}(2 + \log_2 N)^{(3d+1)/2} + 4d + 7}{N} \leq \frac{63\sqrt{d}(2 + \log_2 N)^{(3d+1)/2}}{N}. \end{aligned}$$

This proves Theorem 1. ■

*Proof of Theorem 2.* Theorem 2 follows from Theorem 1. To show this, we use a classical method of constructing finite point sets in the  $d + 1$ -dimensional unit cube in order to obtain a  $d$ -dimensional infinite sequence. Let  $\mu$  be the given measure on  $[0, 1]^d$ , and set  $\nu = \mu \times \lambda$ , where  $\lambda$  is the 1-dimensional Lebesgue measure. Furthermore, for  $i \geq 1$  we set

$$N_i = 2^{(2^i-2)} \quad \text{and} \quad M_i = \begin{cases} 0 & \text{for } i = 1, \\ \sum_{l=1}^{i-1} N_l & \text{for } i \geq 2. \end{cases}$$

By Theorem 1 for any  $i \geq 1$  there exists an  $N_i$ -element point set  $\mathcal{Q}_i \subset [0, 1]^{d+1}$  whose star-discrepancy (with respect to  $\nu$ ) is bounded by

$$\frac{63\sqrt{d+1}(2 + \log_2 N_i)^{(3d+4)/2}}{N_i}.$$

We write  $x_1^{(i)}, \dots, x_{N_i}^{(i)}$  for the elements of  $\mathcal{Q}_i$  rearranged in increasing order according to their last coordinate. Since the order of points does not influence discrepancy, we have

$$(39) \quad D_{N_i}^*(x_1^{(i)}, \dots, x_{N_i}^{(i)}; \nu) \leq \frac{63\sqrt{d+1}(2 + \log_2 N_i)^{(3d+4)/2}}{N_i}.$$

We write  $y_n^{(i)}$  for the last coordinate of  $x_n^{(i)}$ . A simple approximation argument shows that we can assume that  $y_n^{(i)}$  is strictly increasing for  $1 \leq n \leq N_i$ .

Now let  $n \geq 1$  be given. There exists a number  $i$  such that  $n = M_i + j$  for some  $j \in \{1, \dots, N_i\}$ . We set

$$x_n = \bar{x}_j^{(i)},$$

where  $\bar{x}_j^{(i)}$  is the projection of  $x_j^{(i)}$  onto  $[0, 1]^d$  by omitting its last coordinate. Now let  $A$  be any axis-parallel box in  $[0, 1]^d$  which has one vertex at the origin, and let  $N \geq 1$ . There exists some  $i$  such that  $N = M_i + j$  for some  $j \in \{1, \dots, N_i\}$ . Note that in this case  $\log_2 N \geq 2^{i-1} - 2$ , and consequently

$4 + 2 \log_2 N \geq 2^i$ . Then by the triangle inequality we have

$$\begin{aligned}
 (40) \quad & |\#\{n \leq N : x_n \in A\} - N\mu(A)| \\
 & \leq \left| \sum_{l=1}^{i-1} \#\{1 \leq n \leq N_l : x_n^{(l)} \in A \times [0, 1]\} - N_l\nu(A \times [0, 1]) \right| \\
 (41) \quad & + |\#\{1 \leq n \leq j : x_n^{(i)} \in A \times [0, y_j^{(i)}]\} - N_i\nu(A \times [0, y_j^{(i)}])| \\
 (42) \quad & + |N_i\nu(A \times [0, y_j^{(i)}]) - j\mu(A)|.
 \end{aligned}$$

The terms in lines (41) and (42) are each bounded by  $N_i D_{N_i}^*(x_1^{(i)}, \dots, x_{N_i}^{(i)}; \nu)$ . Thus by (39) we conclude that (40) is bounded by

$$\begin{aligned}
 & \left( \sum_{l=1}^{i-1} 63\sqrt{d+1} (2 + \log_2 N_l)^{(3d+4)/2} \right) + 126\sqrt{d+1} (2 + \log_2 j)^{(3d+4)/2} \\
 & \leq 126\sqrt{d+1} \left( \underbrace{\frac{1}{2} \left( \sum_{l=1}^{i-1} (2^l)^{(3d+4)/2} \right)}_{\leq (2^i)^{(3d+4)/2} / 20} + (2^i)^{(3d+4)/2} \right) \leq 133\sqrt{d+1} (2^i)^{(3d+4)/2} \\
 & \leq 133\sqrt{d+1} (4 + 2 \log_2 N)^{(3d+4)/2}.
 \end{aligned}$$

Thus we have

$$D_N^*(x_1, \dots, x_N; \mu) \leq 133\sqrt{d+1} \frac{(4 + 2 \log_2 N)^{(3d+4)/2}}{N},$$

which proves Theorem 2. ■

*Proof of Theorem 4.* In the case  $\Omega = [0, 1]^d$  the two discrepancies in (6) and (23) coincide, up to an additional multiplicative factor  $2^d$  in (23), and Theorem 4 follows directly from the combination of Theorem 1 and Lemma 3.

Now assume that  $\Omega \neq [0, 1]^d$ . Thus there exists a point  $P \in [0, 1]^d$  which is *not* contained in  $\Omega$ . We define the measure  $\mu$  as in (26), that is,

$$\mu(A) = \frac{\lambda(\Omega \cap A)}{\lambda(\Omega)}$$

for every Borel subset  $A$  of  $[0, 1]^d$ . Since by assumption  $\lambda(\Omega) > 0$ , the measure  $\mu$  is a non-negative, normalized Borel measure on  $[0, 1]^d$ . Let  $\mathcal{B}$  denote the class of Borel subsets of  $[0, 1]^d$ , and define

$$\mathcal{A} = \{B \cap \Omega : B \in \mathcal{B}\}.$$

Then the pair  $(\Omega, \mathcal{A})$  forms a measurable space. Since  $\Omega$  is a measurable subset of  $[0, 1]^d$ , the restriction of  $\mu$  to  $(\Omega, \mathcal{A})$  is also a measure; with a slight abuse of notation we will denote this measure also by  $\mu$ . By definition  $\mu(\Omega) = 1$ , and thus  $(\Omega, \mathcal{A}, \mu)$  is a probability space.

As mentioned in Section 3, it is well-known that the class of axis-parallel boxes which are contained in  $[0, 1]^d$  and have a vertex at the origin is a VC class of dimension  $d$ . Let  $\mathcal{C}$  denote the class of all sets which can be obtained by intersecting such an axis-parallel box with  $\Omega$ . It is an easy exercise to check that then the class  $\mathcal{C}$  is also a VC class, of dimension at most  $d$ . Upon applying Theorem 3 to the probability space  $(\Omega, \mathcal{A}, \mu)$  and the class  $\mathcal{C}$  this implies that there exist  $z_1, \dots, z_K \in \Omega$  for which

$$D_N^*(z_1, \dots, z_K; \mu) = D_N^{(\mathcal{C}, \mu)}(z_1, \dots, z_K) \leq 2^{13} \frac{\sqrt{d}}{\sqrt{N}}$$

(the class  $\mathcal{C}$  is not countable, but this problem can be overcome in the way described before the statement of Corollary 1). It is important that in this approximation,  $z_1, \dots, z_K$  are elements of  $\Omega$ . Taking  $K = 2^{26} d N^2$  we get the points  $z_1, \dots, z_K$  for the proof of Theorem 1, which satisfy (38). The proof of Lemma 5 reveals that the points  $x_1, \dots, x_N$  in the conclusion of Theorem 1 are chosen from among  $z_1, \dots, z_K$ . Consequently, since in the present case we have  $z_1, \dots, z_K \in \Omega$ , the points from the conclusion of Theorem 1 also satisfy  $x_1, \dots, x_N \in \Omega$ .

For these points we have

$$D_N^*(x_1, \dots, x_N; \mu) \leq \frac{63\sqrt{d} (2 + \log_2 N)^{(3d+1)/2}}{N}.$$

Let

$$M = \left\lceil \frac{N}{\lambda(\Omega)} \right\rceil;$$

then

$$(43) \quad \left| \frac{1}{M} - \frac{\lambda(\Omega)}{N} \right| \leq \frac{\lambda(\Omega)^2}{N^2} \leq \frac{\lambda(\Omega)}{N^2}.$$

We define

$$x_{N+1} = \dots = x_M = P$$

(where  $P$  is the point for which  $P \notin \Omega$ ). Note that for the discrepancy defined in (23) the supremum is taken over all axis-parallel boxes contained in  $[0, 1]^d$ , without the additional assumption that one corner of each such box has to be at the origin as in the definition of discrepancy in (6). Restricting the supremum to only those axis-parallel boxes which have a vertex at the origin will add a multiplicative factor  $2^d$ . Thus, using (43), we get

$$\begin{aligned} D_M^{(\Omega)}(x_1, \dots, x_M) &\leq 4^d \sup_{A \subset [0, 1]^d} \left| \frac{1}{M} \sum_{n=1}^M \mathbb{1}_{\Omega \cap A}(x_n) - \lambda(\Omega \cap A) \right| \\ &= 4^d \sup_{A \subset [0, 1]^d} \left| \frac{1}{M} \sum_{n=1}^N \mathbb{1}_A(x_n) - \mu(A)\lambda(\Omega) \right| \end{aligned}$$

$$\begin{aligned} &\leq 4^d \left( \frac{\lambda(\Omega)}{N} + \sup_{A \subset [0,1]^d} \left| \frac{\lambda(\Omega)}{N} \sum_{n=1}^N \mathbb{1}_A(x_n) - \mu(A)\lambda(\Omega) \right| \right) \\ &= 4^d \lambda(\Omega) \left( \frac{1}{N} + D_N^*(x_1, \dots, x_N; \mu) \right) \\ &\leq 4^d \lambda(\Omega) \frac{63\sqrt{d}(2 + \log_2 N)^{(3d+1)/2} + 1}{N}, \end{aligned}$$

where the supremum is always taken over those axis-parallel boxes  $A$  which have a vertex at the origin. Now by Lemma 3 we have, using (43) again,

$$\begin{aligned} &\left| \frac{1}{N} \sum_{n=1}^N f(x_n) - \frac{1}{\lambda(\Omega)} \int_{\Omega} f(x) d\lambda(x) \right| \\ &\leq \frac{1}{\lambda(\Omega)} \left| \frac{1}{M} \sum_{n=1}^M (f \cdot \mathbb{1}_{\Omega})(x_n) - \int_{\Omega} f(x) d\lambda(x) \right| + \frac{\|f\|_{\infty}}{N} \\ &\leq \frac{1}{\lambda(\Omega)} D_M^{(\Omega)}(x_1, \dots, x_M) \cdot V(f) + \frac{\|f\|_{\infty}}{N} \\ &\leq 4^d \frac{63\sqrt{d}(2 + \log_2 N)^{(3d+1)/2} + 1}{N} V(f) + \frac{\|f\|_{\infty}}{N}, \end{aligned}$$

which proves Theorem 4. ■

**Acknowledgments.** We want to thank Richard Nickl for several helpful remarks concerning empirical processes indexed by sets and the metric entropy of VC classes, and Benjamin Doerr for valuable explanations concerning the discrepancies of hypergraphs and matrices. Thanks also to Jerome Spanier, who helped us to get hold of a copy of Chelson’s PhD thesis.

Christoph Aistleitner is supported by a Schrödinger scholarship of the Austrian Research Foundation (FWF).

Josef Dick is supported by an Australian Research Council Queen Elizabeth 2 Fellowship.

### References

- [1] C. Aistleitner, *Covering numbers, dyadic chaining and discrepancy*, J. Complexity 27 (2011), 531–540.
- [2] C. Aistleitner and M. Hofer, *Probabilistic discrepancy bounds for Monte Carlo point sets*, Math. Comp. 83 (2014), 1373–1381.
- [3] K. S. Alexander, *Probability inequalities for empirical processes and a law of the iterated logarithm*, Ann. Probab. 12 (1984), 1041–1067; Correction, ibid. 15 (1987), 428–430.
- [4] W. Banaszczyk, *Balancing vectors and Gaussian measures of  $n$ -dimensional convex bodies*, Random Structures Algorithms 12 (1998), 351–360.

- [5] J. Beck, *Some upper bounds in the theory of irregularities of distribution*, Acta Arith. 43 (1984), 115–130.
- [6] J. Beck and W. W. L. Chen, *Irregularities of Distribution*, Cambridge Univ. Press, Cambridge, 2008.
- [7] J. Beck and J. Spencer, *Integral approximation sequences*, Math. Program. 30 (1984), 88–98.
- [8] D. Bilyk, *On Roth’s orthogonal function method in discrepancy theory*, Unif. Distrib. Theory 6 (2011), 143–184.
- [9] D. Bilyk, M. T. Lacey, and A. Vagharshakyan, *On the small ball inequality in all dimensions*, J. Funct. Anal. 254 (2008), 2470–2502.
- [10] L. Brandolini, L. Colzani, G. Gigante, and G. Travaglini, *On the Koksma–Hlawka inequality*, J. Complexity 29 (2013), 158–172.
- [11] B. Chazelle, *The discrepancy method*, Cambridge Univ. Press, Cambridge, 2000.
- [12] P. O. Chelson, *Quasi-random techniques for Monte Carlo methods*, PhD thesis, The Claremont Graduate Univ., 1976.
- [13] L. Devroye, L. Györfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition*, Appl. Math. (New York) 31, Springer, New York, 1996.
- [14] J. Dick, F. Y. Kuo, and I. H. Sloan, *High-dimensional integration: The quasi-Monte Carlo way*, Acta Numer. 22 (2013), 133–288.
- [15] J. Dick and F. Pillichshammer, *Digital Nets and Sequences. Discrepancy Theory and Quasi-Monte Carlo Integration*, Cambridge Univ. Press, Cambridge, 2010.
- [16] B. Doerr, *A lower bound for the discrepancy of a random point set*, J. Complexity 30 (2014), 16–20.
- [17] M. Drmota and R. F. Tichy, *Sequences, Discrepancies and Applications*, Lecture Notes in Math. 1651, Springer, Berlin, 1997.
- [18] E. Giné and R. Nickl, *An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation*, Probab. Theory Related Fields 143 (2009), 569–596.
- [19] P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Appl. Math. (New York) 53, Springer, New York, 2004.
- [20] G. Harman, *Variations on the Koksma–Hlawka inequality*, Unif. Distrib. Theory 5 (2010), 65–78.
- [21] J. Hartinger and R. Kainhofer, *Non-uniform low-discrepancy sequence generation and integration of singular integrands*, in: Monte Carlo and Quasi-Monte Carlo Methods 2004, Springer, Berlin, 2006, 163–179.
- [22] D. Haussler, *Sphere packing numbers for subsets of the Boolean  $n$ -cube with bounded Vapnik–Chervonenkis dimension*, J. Combin. Theory Ser. A 69 (1995), 217–232.
- [23] S. Heinrich, E. Novak, G. W. Wasilkowski, and H. Woźniakowski, *The inverse of the star-discrepancy depends linearly on the dimension*, Acta Arith. 96 (2001), 279–302.
- [24] A. Hinrichs, *Covering numbers, Vapnik–Červonenkis classes and bounds for the star-discrepancy*, J. Complexity 20 (2004), 477–483.
- [25] E. Hlawka, *Gleichverteilung und Simulation*, Österreich. Akad. Wiss. Math.-Natur. Kl. Sitzungsber. II 206 (1997), 183–216.
- [26] E. Hlawka and R. Mück, *A transformation of equidistributed sequences*, in: Applications of Number Theory to Numerical Analysis (Montreal, 1971), Academic Press, New York, 1972, 371–388.
- [27] E. Hlawka und R. Mück, *Über eine Transformation von gleichverteilten Folgen. II*, Computing (Arch. Elektron. Rechnen) 9 (1972), 127–138.
- [28] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, Wiley-Interscience, New York, 1974.



- [29] L. Lovász, J. Spencer, and K. Vesztegombi, *Discrepancy of set-systems and matrices*, Eur. J. Combin. 7 (1986), 151–160.
- [30] P. Massart, *About the constants in Talagrand's concentration inequalities for empirical processes*, Ann. Probab. 28 (2000), 863–884.
- [31] J. Matoušek, *Geometric Discrepancy. An Illustrated Guide*, Algorithms Combin. 18, Springer, Berlin, 2010.
- [32] N. Sauer, *On the density of families of sets*, J. Combin. Theory Ser. A 13 (1972), 145–147.
- [33] C. Schretter and H. Niederreiter, *A direct inversion method for non-uniform quasi-random point sequences*, Monte Carlo Methods Appl. 19 (2013), 1–9.
- [34] M. Talagrand, *Sharper bounds for Gaussian and empirical processes*, Ann. Probab. 22 (1994), 28–76.
- [35] M. Talagrand, *Concentration of measure and isoperimetric inequalities in product spaces*, Inst. Hautes Études Sci. Publ. Math. 81 (1995), 73–205.
- [36] V. N. Vapnik and A. Ja. Červonenkis [A. Ya. Chervonenkis], *The uniform convergence of frequencies of the appearance of events to their probabilities*, Teor. Veroyatnost. i Primenen. 16 (1971), 264–279 (in Russian).
- [37] N. Vayatis and R. Azencott, *Distribution-dependent Vapnik–Chervonenkis bounds*, in: Computational Learning Theory (Nordkirchen, 1999), Lecture Notes in Comput. Sci. 1572, Springer, Berlin, 1999, 230–240.
- [38] X. Wang and K. S. Tan, *How do path generation methods affect the accuracy of quasi-Monte Carlo methods for problems in finance?*, J. Complexity 28 (2012), 250–277.
- [39] X. Wang and K. S. Tan, *Pricing and hedging with discontinuous functions: Quasi-Monte Carlo methods and dimension reduction*, Management Sci. 59 (2013), 376–389.

Christoph Aistleitner, Josef Dick  
School of Mathematics and Statistics  
University of New South Wales  
Sydney, NSW 2052, Australia  
E-mail: aistleitner@math.tugraz.at  
josef.dick@unsw.edu.au

Received on 23.8.2013  
and in revised form on 1.4.2014

(7564)

