Monte Carlo Methods Appl. Vol. ■■ No. ■ (■■■■), pp. 1–18
DOI 10.1515 / MCMA.■■■■.■■■

Improved Halton sequences and discrepancy bounds

Henri Faure and Christiane Lemieux

Abstract. For about fifteen years, the surprising success of quasi-Monte Carlo methods in finance has been raising questions that challenge our understanding of these methods. At the origin are numerical experiments performed with so-called GSobol' and GFaure sequences by J. Traub and his team at Columbia University, following the pioneering work of S. Tezuka in 1993 on generalizations of Niederreiter $(t, s)-$sequences, especially with $t = 0$ (Faure sequences). Then in the early 2000, another breakthrough was achieved by E. Atanassov, who found clever generalizations of Halton sequences by means of permutations that are even asymptotically better than Niederreiter-Xing sequences in high dimensions. Unfortunately, detailed investigations of these GHalton sequences, together with numerical experiments, show that this good asymptotic behavior is obtained at the expense of remaining terms and is not sensitive to different choices of permutations of Atanassov. As the theory fails, the reasons of the success of GHalton, as well as GFaure, must be sought elsewhere, for instance in specific selections of good scramblings by means of tailor-made permutations. In this paper, we report on our assertions above and we give some tracks to tentatively remove a part of the mystery of QMC success.

Keywords. 65C05, 65C10, 11K38.

AMS classification. Low discrepancy sequences, Quasi-Monte Carlo methods, Computational finance.

1. Introduction

This paper aims to provide some new contributions to the thorny problem of high dimensional integration. It is well known that the number of nodes N required to approximate an s-dimensional integral with accuracy ε is exponential in s [17, 27]. For instance, using a cartesian product of trapezoidal rules gives an error in $O(N^{-\varepsilon/s})$, hence needing $N = 10^s$ points if $\varepsilon = 10^{-2}$. This phenomenon is named the "curse of dimensionality" and for more than fifty years it has been the subject of countless tentatives to overcome its evil spell.

In short, these tentatives can be divided in two families, both being still unsatisfactory: (1) Monte Carlo (MC) methods seem to give a solution with an approximation in $O(1/\sqrt{N})$, but it is only a probabilistic error bound still too slow and (2) quasi-Monte Carlo (QMC) methods do not have this disadvantage, but the deterministic bounds in

The second author's research was supported by an NSERC Discovery Grant.

$O((\log N)^{s-1}/N)$ that they provide, although asymptotically good, require a number of points practically too large to be meaningful. Nevertheless special sets of nodes used in QMC methods, among the so-called *low discrepancy point sets* to be discussed in Section 2, give very good results in finance—far better than MC methods—where integration problems are often in several hundreds of dimensions, see [16], [20] and [27, Section 2] where a very interesting "little history" of investigations on that dilemma at Columbia University is given. How can we explain that? Leaving aside the functional aspect of this problem based on weighted spaces and tractability, as surveyed in [27], we have taken in [9] and [14] another direction with the deepening of number theoretical properties of two main families of low discrepancy sequences, namely generalized Halton and Faure sequences. In the present study, we report some surprising results of experiments confirming that our knowledge on irregularities of distribution of multidimensional sequences is still poor in spite of remarkable results obtained in the last three decades. Its main lack is that theoretical bounds ignore the fundamental importance of permutations in the behavior of scrambled multidimensional sequences. Even for the modified Halton sequences studied by Atanassov [1], we will see that good asymptotic bounds depending on a clever choice of permutations are in fact deceiving.

On the other hand, in one dimension there are precise studies [4, 7, 8] that yield families of permutations with a tight control on the discrepancy of the resulting scrambled sequences. Using such scrambled one-dimensional sequences as coordinates of generalized Halton or generalized Faure sequences gives very good results in experiments with high dimensional integrals [9, 14], as good as Sobol' sequences with well-chosen direction numbers. Moreover, it is easy to choose which coordinates have to be scrambled or not in relation with the presumed importance of such coordinates for the problem at hand. In this way, its effective dimension could be checked experimentally by comparing the results for different, more or less numerous, scrambled coordinates.

The paper is organized as follows. Bounds for the discrepancy of Halton sequences are recalled in Section 3, with the required background definitions given in Section 2. Similar bounds but for the class of digital (t, s)-sequences are given in Section 4, and allow us, in Section 5, to establish comparisons between these sequences and Halton sequences with respect to the asymptotic behavior of their discrepancy bounds. In Section 6, we focus instead on the behavior of the bounds in the non-asymptotic regime, thus unveiling the huge gap between the behavior in these two regimes, in particular with respect to the prominent example of *modified Halton sequences* derived from the remarkable theoretical work of Atanassov [1]. We provide in Section 7 a discussion of digit permutations and their effect on the discrepancy in these two regimes, and then conclude the paper in Section 8.

2. Definitions

2.1. Halton and Faure sequences

Halton sequences are s-dimensional sequences, with values in the hybercube $I^s = [0,1)^s$. They are obtained using one-dimensional van der Corput sequences in base b for each coordinate, but in order to have good distribution properties, it is necessary to choose pairwise coprime bases, the simplest choice being to take the first s primes. So, let us first recall the definition of a *van der Corput sequence* in base b, denoted S_b: Its nth term ($n \geq 1$) is defined as

$$S_b(n) = \sum_{r=0}^{\infty} \frac{a_r(n)}{b^{r+1}}, \tag{2.1}$$

where $a_r(n)$ is the r-th digit of the b-adic expansion of

$$n-1 = \sum_{r=0}^{\infty} a_r(n)\, b^r.$$

An s-dimensional *Halton sequence* [11] $X_1, X_2, \ldots$ in I^s is defined as

$$X_n = (S_{b_1}(n), \ldots, S_{b_s}(n)), \tag{2.2}$$

where the b_j's, for $j = 1, \ldots, s$, are pairwise coprime.

A one-dimensional *generalized van der Corput sequence* [4] is obtained first by choosing a sequence $\Sigma = (\sigma_r)_{r\geq 0}$ of permutations of $Z_b = \{0, 1, \ldots, b-1\}$; then, the nth term of the sequence is defined as

$$S_b^{\Sigma}(n) = \sum_{r=0}^{\infty} \frac{\sigma_r\big(a_r(n)\big)}{b^{r+1}}. \tag{2.3}$$

If the same permutation σ is used for all digits, (i.e., if $\sigma_r = \sigma$ for all $r \geq 0$), then we use the notation S_b^{σ} to denote S_b^{Σ}. The van der Corput sequence in base b defined in (2.1) is obtained by taking $\sigma_r = id$ for all $r \geq 0$, where id stands for the identity permutation over Z_b.

A *generalized Halton sequence* [6] is defined by choosing s sequences of permutations $\Sigma_j = (\sigma_{j,r})_{r\geq 0}$, $j = 1, \ldots, s$, and then by defining the nth point $X_n \in I^s$ of the sequence as

$$X_n = (S_{b_1}^{\Sigma_1}(n), \ldots, S_{b_s}^{\Sigma_s}(n)),\ n \geq 1, \tag{2.4}$$

where the b_j's are pairwise coprime bases. In applications, these b_j's are usually chosen as the first s prime numbers. In this case, we denote the jth base as p_j.

Another way of enriching the van der Corput sequence is by applying a linear transformation to the digits $a_0(n), a_1(n), \ldots$ before outputting a number between 0 and 1.

We call this a *linearly scrambled* van der Corput sequence. For a prime base b, it is obtained by choosing an $\infty \times \infty$ matrix $C = (C_{r,l})$ with elements in $\mathbb{Z}_b$, and then defining the nth term of this sequence as

$$S_b^C(n) = \sum_{r=0}^{\infty} \sum_{l=0}^{\infty} C_{r+1,l+1} a_l(n) b^{-(r+1)}.$$

A *generalized Faure sequence in base* b (where b is prime) has its nth point $X_n \in I^s$ defined as

$$X_n = (S_b^{C_1}, \ldots, S_b^{C_s}), \ n \geq 1, \tag{2.5}$$

where $C_j = A_j P_j$, with P_j the (upper triangular) Pascal matrix in $\mathbb{Z}_b$ raised to the power $j-1$, and A_j a nonsingular lower triangular matrix [24]. The original Faure sequences from [5] amounts to take A_j as the identity matrix. We recall these definitions of $(0,s)-$sequences because we will point out some interesting connections with generalized Halton sequences in Section 7.

2.2. Discrepancy

Irregularities of distribution of sequences are measured with the concept of *discrepancy*. Various notions exist but, for short, here we only consider the so-called *extreme discrepancy*, which corresponds to the worst case error in the domain of complexity of multivariate problems. Let be given a point set $\mathcal{P}_N = \{X_1, \ldots, X_N\} \subseteq I^s$ and an interval J of I^s of the form $J = \prod_{j=1}^s [y_j, z_j)$, where $0 \leq y_j < z_j \leq 1$. Then the *discrepancy function* of $\mathcal{P}_N$ on J is the difference

$$E(J; N) = A(J; N) - NV(J),$$

where $A(J; N) = \#\{n; 1 \leq n \leq N, X_n \in J\}$ is the number of points in $\mathcal{P}_N$ that fall in the subinterval J, and $V(J) = \prod_{j=1}^s (z_j - y_j)$ is the volume of J.

Let us denote by J^* subintervals of the form $J^* = \prod_{j=1}^s [0, z_j)$, where $0 \leq z_j \leq 1$. Then, the *star (extreme) discrepancy* D^* and the *(extreme) discrepancy* D of $\mathcal{P}_N$ are defined by

$$D^*(\mathcal{P}_N) = \sup_{J^*} |E(J^*; N)| \quad \text{and} \quad D(\mathcal{P}_N) = \sup_J |E(J; N)|.$$

It is well known that $D^*(\mathcal{P}_N) \leq D(\mathcal{P}_N) \leq 2^s D^*(\mathcal{P}_N)$. For an infinite sequence X, we denote by $D(N, X)$ and $D^*(N, X)$ the discrepancies of its first N points. Note that several authors have a $1/N$ factor when defining the above quantities.

A sequence satisfying $D^*(N, X) \in O((\log N)^s)$ is typically considered to be a *low-discrepancy sequence* (according to the present knowledge on irregularities of distribution). But the constant hidden in the O notation needs to be made explicit

to make comparisons possible across sequences. This is achieved in many papers with an inequality of the form

$$D^*(N, X) \leq c_s(\log N)^s + O((\log N)^{s-1}).$$

Implicit in that inequality is a bound of the form (for some integer N_0)

$$D^*(N, X) \leq c_s(\log N)^s + d_s(\log N)^{s-1}, \text{ for all } N \geq N_0. \tag{2.6}$$

Then a common approach to compare low-discrepancy sequences is to assume the second term above can be neglected and to study the behavior of c_s as a function of s, i.e., people consider the asymptotic behavior of the sequence as N goes to infinity.

But in applications, the number N of sample points used is limited and seldom goes beyond 10^5 or 10^6. On the other hand, for the second term to be neglected, we should have, for example,

$$d_s(\log N)^{s-1} \leq c_s(\log N)^s,$$

which holds if and only if $\log N > (d_s/c_s)$, i.e., $N \geq N_1 \equiv e^{d_s/c_s}$. It means that N might have to be huge before c_s can be a meaningful way of comparing sequences, making c_s completely useless as a quality measure for common ranges of experiments in QMC methods. Hence, it is important to take into account the comparative values of c_s and d_s if we want to give more realistic comparisons of low discrepancy sequences. Moreover, in the Koksma-Hlawka inequality, for high dimensional numerical integration, dividing (2.6) by N does not help because the term $(\log N)^s/N$ remains prominent, regardless of how small the leading constant c_s is. Such a behavior was already noted, especially by Morokoff and Caflisch [16] and Kocis and Whiten [12]. See also Section 2 in the remarkable survey of Woźniakowski [27], who gives an account of the origin of research on tractability problems in relation with the surprising performance of QMC methods in finance.

3. Discrepancy bounds for Halton sequences

This subsection describes results that have been established since the introduction of Halton sequences in 1960 and that deal with the discrepancy of generalized Halton sequences. In order to avoid confusion, we will denote the original Halton sequences as in (2.1) by H, the generalized Halton sequences as in (2.4) by GH and the so-called *modified Halton sequences* constructed by Atanassov as HA (see (3.7) below). Note that these HA sequences are a special case of GH sequences.

It was first Meijer [15] who improved a loose bound of Halton [11] to inequality

$$D^*(N, H) \leq \prod_{j=1}^{s} \frac{b_j - 1}{\ln b_j} (\log N)^s + O((\log N)^{s-1}).$$

Next, Faure [5, 3] further improved the bound to

$$D^*(N,H) \le \prod_{j=1}^{s} \frac{b_j-1}{2\ln b_j}(\log N)^s + O((\log N)^{s-1}), \tag{3.1}$$

rewritten with another proof by Niederreiter [17, p. 29] as (with the same c_s)

$$D^*(N,H) < \prod_{j=1}^{s}\left(\frac{b_j-1}{2\log b_j}\log N + \frac{b_j+1}{2}\right) + s. \tag{3.2}$$

Next, Faure [6] remarked that his proof in [3] extends easily to GH sequences. But the major theoretical improvement goes back to Atanassov [1, Theorem 2.1], with a completely different proof using an argument of diophantine geometry:

$$D^*(N,H) \le \frac{1}{s!}\prod_{j=1}^{s}\left(\frac{(b_j-1)\ln N}{2\ln b_j} + s\right) + \sum_{k=0}^{s-1}\frac{b_{k+1}}{k!}\prod_{j=1}^{k}\left(\left\lfloor\frac{b_j}{2}\right\rfloor\frac{\ln N}{\ln b_j} + k\right) + u, \tag{3.3}$$

where u is 0 when all numbers b_j are odd, and

$$u = \frac{b_j}{2(s-1)!}\prod_{1\le i\le s, i\ne j}\left(\frac{(b_i-1)\ln N}{2\ln b_i} + s - 1\right)$$

if b_j is the even number among them. Therefore estimate (2.6) holds with constant

$$c_s = \frac{1}{s!}\prod_{j=1}^{s}\frac{b_j-1}{2\ln b_j}. \tag{3.4}$$

By making the constant c_s smaller by a factor $s!$, it is now going to 0 as s goes to infinity, whereas previously it was tending to infinity super-exponentially! Moreover, as for the preceding bounds, it is easy to extend this last result to generalized Halton sequences GH. Therefore, in the comparisons to be made in Sections 5 and 6, we use the value of c_s given by (3.4) for GH sequences and denote it $c_s(GH)$.

But in the same paper, when the bases are distinct primes, Atanassov was able to make the constant c_s even smaller in two cases. In the following, we deal only with the most interesting of these cases for applications. We need a definition before we can state the result.

Definition. Let $p_1,\ldots,p_s$ be s distinct primes. The integers $k_1,\ldots,k_s$ are called *admissible* for them, if $p_j \not| k_j$ and for each set of integers $b_1,\ldots,b_s$, $p_j \not| b_j$, there exists a set of integers $\alpha_1,\ldots,\alpha_s$, satisfying the congruences

$$k_j^{\alpha_j}\prod_{1\le i\le s, i\ne j} p_i^{\alpha_i} \equiv b_j \pmod{p_j}, \qquad j = 1,\ldots,s. \tag{3.5}$$

Finding sets of admissible integers is not a difficult task and amounts to solving integer linear systems. Many choices are possible and lists have been provided, see for instance [2] and [25].

With the help of admissible integers, Atanassov defines his modified Halton sequence HA as follows: For each coordinate j $(1 \leq j \leq s)$, let $T_j = (\tau_{j,r})_{r\geq 0}$ be the sequence of permutations of Z_{p_j} defined by

$$\tau_{j,r}(x) = k_j^r x \pmod{p_j}, \quad r \geq 0, \tag{3.6}$$

where $k_1, \ldots, k_s$ are admissible integers for the prime numbers $p_1, \ldots, p_s$. Then, the *modified Halton sequence* HA associated with admissible integers $k_1, \ldots, k_s$ is the generalized Halton sequence (as defined in (2.4)):

$$HA = (S_{p_1}^{T_1}, \ldots, S_{p_s}^{T_s}). \tag{3.7}$$

This special choice of permutations T_j leads to the best bounds found until now for generalized Halton sequences (where $K = \prod_{j=1}^{s}(p_j - 1)$):

$$\begin{aligned} D^*(N, HA) \quad \leq \quad & \left(\frac{1}{s!} \prod_{j=1}^{s} \left(\frac{\log N}{K \log p_j} + s \right) \right) \cdot K^s \cdot \\ & \left(1 + \sum_{j=1}^{s} \log p_j \prod_{j=1}^{s} \frac{p_j}{p_j - 1} \left(-1 + \prod_{j=1}^{s} (1 + \log p_j) \right) \right) \\ & + \sum_{k=0}^{s-1} \frac{p_{k+1}}{k!} \prod_{j=1}^{k} \left(\left\lfloor \frac{p_j}{2} \right\rfloor \frac{\log N}{\log p_j} + k \right) \\ & + \sum_{j=1}^{s} \frac{1}{(s-1)!} \prod_{\substack{k=1 \\ k\neq j}}^{s} \left(\frac{p_k}{2} \frac{\log N}{\log p_k} + s - 1 \right), \end{aligned} \tag{3.8}$$

The proof of this result is very long and difficult to follow. In the technical report [26], we checked carefully each step of the proof and provided three alternative proofs for an important intermediary result (Prop. 4.1 in [1]) whose proof contains a little inaccuracy in [1].

After rewriting, Atanassov took out the leading constant c_s as in (2.6), everything else being rejected in the complementary term $O((\log N)^{s-1}) = d_s(\log N)^{s-1}$, thus getting:

$$c_s(HA) = \frac{1}{s!} \sum_{j=1}^{s} \log p_j \prod_{j=1}^{s} \frac{p_j(1 + \log p_j)}{(p_j - 1)\log p_j}. \tag{3.9}$$

4. Discrepancy bounds for $(t, s)-$sequences

Before proceeding with comparisons between these various bounds for generalized Halton sequences, we briefly recall the main bounds corresponding to the other family

of low discrepancy sequences, namely the $(t,s)-$sequences. By doing so, we will be able to establish further comparisons in the next section, which hopefully will be useful to QMC practitioners.

For $(0,s)-$sequences in prime base $b \geq s$, in other words Faure sequences, denoted F, we have an explicit bound deduced from [5, Section 6.4]: for all $N \geq 1$

$$D^*(N,F) \leq \left(\frac{b-1}{2}\right)^s \frac{1}{(s-1)!} \sum_{m=0}^{n} (m+s)^{s-1}, \tag{4.1}$$

where the integer n satisfies $(\log N + \log 2)/\log b - 1 \leq n \leq (\log N + \log 2)/\log b$.

From this bound, we get after some calculations that the constant c_s in (2.6) for Faure sequences is

$$c_s = \frac{1}{s!}\left(\frac{b-1}{2\log b}\right)^s. \tag{4.2}$$

The bound in (4.1) and constant c_s in (4.2) remain valid for generalized Faure sequences as defined in (2.5). In the following sections, we use that value of c_s given in (4.2) for $c_s(GF)$.

For $(t,s)-$sequences, e.g., Niederreiter sequences (and Sobol' sequences in base 2), it is proved (see for instance [17]) that

$$c_s = \frac{b^t}{s!}\frac{b-1}{2\lfloor \frac{b}{2} \rfloor}\left(\frac{\lfloor \frac{b}{2} \rfloor}{\log b}\right)^s. \tag{4.3}$$

As for Faure sequences, this constant is also valid if the coordinates are scrambled by permutations of the digits. Note that in these constructions, the base b and the dimension s can be chosen independently from each other, but the quality parameter t depends on b and s. While $t \in O(s \log s)$ for the Sobol' and Niederreiter sequences in a given base b, $t \in O(s)$ for the Niederreiter-Xing sequences studied in [18]. Note that Kritzer [13] recently improved the constants c_s in (4.2) and (4.3) by a factor $1/2$ for odd $b \geq 3$ and $s \geq 2$, and by a factor $1/3$ for $b = 2$ and $s \geq 5$ (a similar result holds for even b).

Finally, while (4.3) only gives the constant c_s, Niederreiter also obtains explicit upper bounds for $(t,s)-$sequences (see, for instance, [17, Theorem 4.12]). However, since these bounds are valid for integers $N \geq b^t$, they are useless for usual ranges of samples in high dimensional integration, therefore we do not present them and refer the interested reader to [17].

5. Asymptotic comparisons

The results in this section come from computations done in [9]. We report them for sake of completeness and to enlighten our goal. They concern the constants c_s in the leading term of the upper bound (2.6). Such comparisons are a matter of Number

Theory competition, where the most tiny improvement in a direction can have a great importance, especially if people tried for a long time to progress in that direction. But regarding our present goal of practical implementations for MC and QMC users, asymptotic results that require astronomical and even cosmological numbers of trials to become relevant seem of little interest. Nevertheless, we think it is worthwhile to present them, because it can help clarify different points of view.

There are two possibilities for comparing the quantities c_s given in Sections 3 and 4: searching for asymptotic estimates as s grows to infinity or computing c_s for different values of the dimension s.

In the first approach, it is easier to compare the logarithms, using Stirling formula and results from Number Theory. We refer to our paper [9] and Section B of its online appendix for details. The estimates are as follows, where $c_s(NX)$ is the value of the constant c_s given in (4.3), but in the special case of Niederreiter-Xing sequences, for which $t \in O(s)$ as mentioned above:

$$\begin{aligned}
-s\log(1.5) < \quad & \log(c_s(GH)) \quad < -\log s + O(1),\\
-s\log s < \quad & \log(c_s(HA)) \quad < -s\log s + O(s),\\
-s\log\log s < \quad & \log(c_s(GF)) \quad < -s\log\log s + O(s),\\
-s\log s < \quad & \log(c_s(NX)) \quad < -s\log s + O(s),
\end{aligned}$$

Therefore, from the point of view of asymptotic upper bounds, the best sequences are Halton-Atanassov and Niederreiter-Xing sequences, with order $-s\log s$, followed by Faure $(0,s)-$sequences, with order $-s\log\log s$ and Halton sequences, with order $-\log s$. Hence, this shows that modified Halton sequences HA can have bounds on their discrepancy that are asymptotically as good as the Niederreiter-Xing sequences, which as mentioned before, are designed so that the quality parameter t behaves optimally as a function of s. By constrast, the value of c_s given in (3.1)—so what was the best bound prior to the work of Atanassov—is such that $\log c_s \in O(s\log s)$.

As a last comment for this first approach, we note that for Sobol' sequences, $\log c_s$ is bounded from above by $s\log\log s + O(s)$, whatever the direction numbers are. This fact is a first *surprise* since Sobol' sequences with carefully chosen direction numbers are among the best in numerical experiments.

Next, for the second approach, we give in Table 1 the values of c_s for Halton, Faure, Niederreiter-Xing and Halton-Atanassov sequences, for values of s ranging between 10 and 50. We have kept the values computed with D (instead of D^*) in [9], hence the corrective factor 2^s for the leading constants in Table 1. The values t_s for the Niederreiter-Xing sequences were obtained using the database `MinT` [21]. That is, for each s we searched for the pair (b,t) that was yielding the smallest value for c_s in (4.3) and we used the small improvement by Kritzer for (4.3).

These results show that for smaller values of s, the Niederreiter-Xing sequences have the smallest constant, but around $s = 30$, the Halton-Atanassov sequences start to

s	Halton $2^s c_s(GH)$	Faure $2^s c_s(GF)$	Nied-Xing $2^s c_s(NX)$	Halton-Atan $2^s c_s(HA)$
10	2.18e-01	2.19e-01	4.46e-03	2.82e+00
20	2.85e-02	1.72e-02	1.75e-08	1.44e-07
30	3.53e-03	3.27e-05	1.40e-16	1.91e-17
40	5.78e-04	1.20e-07	1.07e-24	6.17e-29
50	9.50e-05	1.19e-09	7.56e-35	1.34e-41

Table 1. Values of constants in the discrepancy bounds for each sequence and different dimensions s.

have a better constant, hence suggesting an advantage in favor of the Halton-Atanassov sequences inside the "$-s \log s$" order. Here is a second *surprise* since well chosen generalized Halton sequences and generalized Faure sequences (see Faure–Lemieux digital scramblings in [9] and [14]) perform as well as Sobol' sequences in applications whereas Halton–Atanassov sequences as defined in [1] behave quite badly (see [25] and [9, Section 3 (2) and Appendix (Section H)] .

6. Comparisons in the non-asymptotic regime

While we tried to understand the different steps of the proof of Theorem 2.3 [1, Section 4], which asserts that inequality (2.6) holds with leading constant (3.9), our attention focused on a quantity introduced in the final step, the constant $K = \prod_{j=1}^{s}(p_j - 1)$. This constant, which is already big (recall that $p_1, \ldots, p_s$ are distinct primes, hence at least the first s primes), appears in the first line of (3.8) at the power s. Of course, it is compensated in the leading term because of the division by K, but a bit of rewriting shows that the term Ks, which is multiplied s times, remains in the complementary term $d_s(\log N)^{s-1}$ of (2.6). This causes the bound (3.8) to be overall extremely large. This is a third *surprise* since nobody before paid attention to this complementary term when comparing the merits of low discrepancy sequences, neither for Halton nor for $(t, s)-$sequences.

In order to have a concrete idea of the importance of the complementary term in comparison with the leading one, we have computed the bounds (3.2) and (3.3) for generalized Halton sequences GH (with the first s prime numbers, excluding prime 2 in (3.3)) and the bound (3.8) for modified Halton sequences HA, but without the two last lines (so as to lighten the computation) as in Atanassov's paper [1, p. 31]. We report these results in Tables 2, 3, 4. As a matter of fact, since the Koksma-Hlawka inequality involves the ratio $D^*(N, X)/N$ and since numbers occuring in the results are very huge, we have actually computed the bounds of $D^*(N, X)/N$ and $c_s(\log N)^s/N$. And finally, since the bounds of $D^*(N, X)/N$ were still too large, we

N	bounds	c_5	$\log\left(\text{bd}\frac{D^*(N)}{N}\right)$	$c_5(\log N)^5/N$	$\frac{c_5(\log N)^5}{\text{bd } D^*(N)}$
10^4	(3.2) GH	2.62e+00	4.03	1.74e+01	3.10e-01
	(3.3) GH	7.09e-02	1.09	4.70e-01	1.59e-01
	(3.8) HA	5.04e+00	33.07	3.34e+01	1.45e-13
10^5	(3.2) GH	2.62e+00	2.63	5.31e+00	3.84e-01
	(3.3) GH	7.09e-02	-0.45	1.43e-01	2.24e-01
	(3.8) HA	5.04e+00	30.77	1.02e+01	4.41e-13
10^6	(3.2) GH	2.62e+00	1.09	1.32e+00	4.45e-01
	(3.3) GH	7.09e-02	-2.08	3.57e-02	2.85e-01
	(3.8) HA	5.04e+00	28.47	2.54e+00	1.09e-12
10^{10}	(3.2) GH	2.62e+00	-5.88	1.70e-03	6.05e-01
	(3.3) GH	7.09e-02	-9.23	4.59e-05	4.68e-01
	(3.8) HA	5.04e+00	19.28	3.26e-03	1.38e-11

Table 2. Computations of bounds (3.2), (3.3) and (3.8) in dimension 5.

decided to show their logarithm. The results have been rounded to two decimals.

To sum up, in the first column we give the value of c_s, in the second one the value of $\log(\text{bound}\,(D^*(N,X)/N))$, in the third one the value of $c_s(\log N)^s/N$ and in the last one the ratio $c_s(\log N)^s/\text{bound}(D^*(N,X))$. Hence, a small ratio means the complementary term $d_s(\log N)^{s-1}$ in (2.6) is largely dominating the leading term $c_s(\log N)^s$.

In light of these results, it is obvious that a bound in which c_s behaves extremely well does not necessarily mean that overall the bound is better. Namely, although modified Halton sequences HA have the constant c_s with the best behavior, we can see that when s is small, $c_s(HA)$ is actually larger than $c_s(GH)$ for generalized Halton sequences; and when s is large, the good behavior of $c_s(HA)$ would require astronomical values of N in order to provide smaller discrepancy bounds. In addition, it should be pointed out that except for GH when $s = 5$, all the above bounds are still bigger than 1, while from the definitions it is clear that $D^*(N,X)/N \leq 1 \ldots$ These experiments dramatically show that the good behavior of c_s is obtained at the expense of d_s for HA sequences.

7. Primordial importance of permutations

A remarkable characteristic of bounds on the discrepancy of Halton and $(t,s)-$sequences listed in Sections 3–4 and compared in Sections 5–6 is that they are valid for their generalizations with sequences of permutations, whatever these sequences of

N	sequence	c_{10}	$\log\left(\text{bd}\dfrac{D^*(N)}{N}\right)$	$c_{10}(\log N)^{10}/N$	$\dfrac{c_{10}(\log N)^{10}}{\text{bd } D^*(N)}$
10^4	(3.2) GH	7.72e+02	22.34	3.39e+08	6.73e-02
	(3.3) GH	1.29e-03	11.09	5.66e+02	8.68e-03
	(3.8) HA	2.76e-03	222.53	1.21e+03	2.74e-94
10^5	(3.2) GH	7.72e+02	21.79	3.16e+08	1.09e-01
	(3.3) GH	1.29e-03	10.19	5.27e+02	1.98e-02
	(3.8) HA	2.76e-03	220.23	1.13e+03	2.55e-93
10^6	(3.2) GH	7.72e+02	20.97	1.96e+08	1.53e-01
	(3.3) GH	1.29e-03	9.12	3.27e+02	3.57e-02
	(3.8) HA	2.76e-03	217.93	6.98e+02	1.58e-92
10^{10}	(3.2) GH	7.72e+02	16.16	3.24e+06	3.10e-01
	(3.3) GH	1.29e-03	3.76	5.40e+00	1.26e-01
	(3.8) HA	2.76e-03	208.72	1.15e+01	2.61e-90

Table 3. Computations of bounds (3.2), (3.3) and (3.8) in dimension 10.

N	sequence	c_{50}	$\log\left(\text{bd}\dfrac{D^*(N)}{N}\right)$	$c_{50}(\log N)^{50}/N$	$\dfrac{c_{50}(\log N)^{50}}{\text{bd } D^*(N)}$
10^4	(3.2) GH	2.57e+45	225.77	4.20e+89	3.72e-09
	(3.3) GH	2.49e-18	91.34	4.07e+26	8.73e-14
	(3.8) HA	1.19e-56	10521	1.95e-12	0
10^5	(3.2) GH	2.57e+45	231.32	2.94e+93	1.02e-07
	(3.3) GH	2.49e-18	95.79	2.85e+30	7.17e-12
	(3.8) HA	1.19e-56	10519	1.36e-08	0
10^6	(3.2) GH	2.57e+45	235.79	2.68e+96	1.06e-06
	(3.3) GH	2.49e-18	99.37	2.60e+33	1.82e-10
	(3.8) HA	1.19e-56	10517	1.24e-05	0
10^{10}	(3.2) GH	2.57e+45	247.08	3.31e+103	1.64e-04
	(3.3) GH	2.49e-18	108.3	3.21e+40	3.07e-07
	(3.8) HA	1.19e-56	10507	1.54e+02	0

Table 4. Computations of bounds (3.2), (3.3) and (3.8) in dimension 50.

permutations are, either the identity or sophisticated ones.

As we saw in Section 3, there is an exception to this, which are the Halton–Atanassov sequences tailored with sequences of permutations obtained from admissible integers. But in this case, there still exists many possibilities for the choice of admissible integers (see our comment after (3.5)) and once again many sequences of permutations give the same bound. But there is more: in the definition of sequences T_j of permutations in [1], recalled in (3.6), the first permutation for each coordinate $\tau_{j,0}$ is equal to the identity id and therefore, the first digit of each coordinate is never scrambled whatever the admissible integers are.

However, there is no need to begin with $r = 0$ to define the sequences of permutations T_j. In [2] Atanassov and Durchova suggest, "because it is intuitively better", to change $\tau_{j,r}$ in (3.6) for $\tau_{j,r}(x) = k_j^{r+1}x + b_{j,r} \pmod{p_j}$ with $r \geq 0$ and with arbitrary integers $b_{j,r}$ which, moreover, does not change the bound (3.8). The proof of that assertion, outlined in [2], can be extended to permutations $\tau_{j,r}(x) = k_j^{r+a_j}x + b_{j,r} \pmod{p_j}$ with arbitrary integers a_j and even in a more general setting involving matrices, as discussed in [10].

Our implementations of Halton–Atanassov sequences, named AD in [9], using power $r + 1$ (instead of r) perform well in various experiments and are comparable to our own scrambled Halton sequences, named FL in [9], themselves comparable to GSobol' and GFaure sequences as shown in [14]. But on the contrary, Halton–Atanassov sequences are hardly distinguishable from original Halton sequences H when using power r. This fact has already been noted by Vandewoestyne and Cools [25] and is quite prominent in examples shown in Figures 1 and 3 where "HApowr" stands for HA sequences with power r in (3.6). With such examples, we hope it is now clear for the reader that using the identity for the first digit only is the cause of a dramatic damage.

The same situation prevails for implementations of GFaure sequences. That is, generalized Faure sequences whose first digit is not permuted do not perform much better than the original Faure sequences. This is illustrated on Figures 2 and 3, where GF is obtained from the first construction in [14, Section 4] and GFunperm is the version of GF where the first digit of each coordinate is not permuted, while F is the original Faure sequence.

All these observations confirm the primordial importance of permutations in the use of low discrepancy sequences for QMC methods, in spite of the theory of irregularities of distribution which, until now, provides only upper bounds that are unable to discriminate the identity among other permutations. The only exception is *dimension one* for which several theoretical and practical results exist showing the gap between the identity and well chosen permutations. To end this section, we give below a short excerpt of these results (from [4] and [8]) to further emphasize the crucial role of permutations for readers of the MCM community.

We only consider sequences $S_b^\sigma := S_b^\Sigma$ from (2.3) where $\sigma_r = \sigma$ for all $r \geq 0$ and

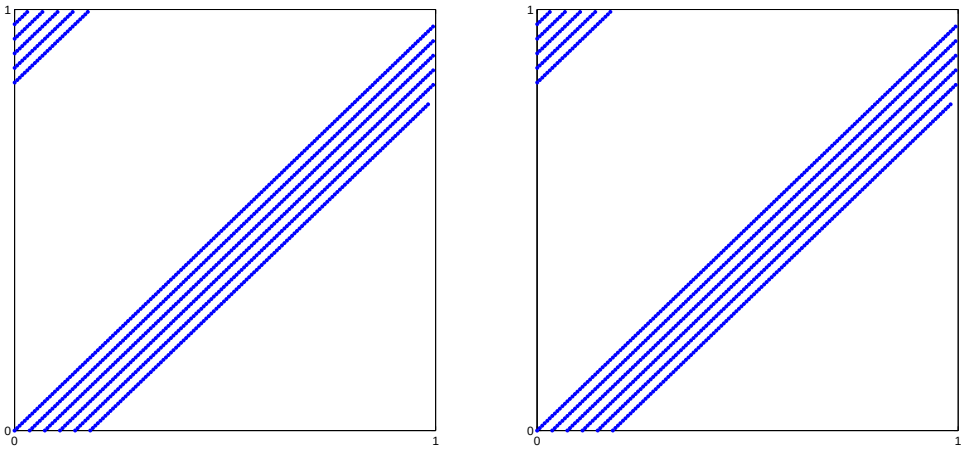

Figure 1. 1000 first points of HApowr (left) and H (right), dimensions 39 and 40

we give results for D only because they are easier to state, but analog formulas also exist for D^*. For such sequences, there exists effectively computable constants α_b^σ, which here depend on the permutation σ, such that for all $N \geq 1$

$$D(N, S_b^\sigma) \leq \frac{\alpha_b^\sigma}{\log b} \log N + \alpha_b^\sigma + 2 \quad \text{and} \quad \limsup_{N\to\infty} \frac{D(N, S_b^\sigma)}{\log N} = \frac{\alpha_b^\sigma}{\log b}. \tag{7.1}$$

It is important to point out that these formulas directly give both the asymptotic and the non-asymptotic regimes.

For large b, computing α_b^σ is a difficult task, but it has a good approximation by means of another quantity which is easy to compute: set $Z_b^\sigma = \left(\frac{\sigma(0)}{b}, \cdots, \frac{\sigma(b-1)}{b}\right)$ and define $d_b^\sigma := \max\limits_{1\leq k\leq b} \max\limits_{0\leq h'<h<b} \left| E\left(\left[\frac{h'}{b}, \frac{h}{b}\right[; k; Z_b^\sigma\right)\right|$, the so-called *discrete discrepancy* of the two-dimensional net $\left\{\left(\frac{k-1}{b}, Z_b^\sigma(k)\right) ; 1 \leq k \leq b\right\}$). Then $d_b^\sigma - 1 \leq \alpha_b^\sigma \leq d_b^\sigma$.

First of all, we recall formulas for the original van der Corput sequences S_b^{id}: for $b = 2$, $d_2^{id} = 1/2$ and $\alpha_2^{id} = 1/3$. Then for odd $b \geq 3$ (analogs exist for even b)

$$d_b^{id} = \frac{b-1}{4} + \frac{b-1}{4b} \quad \text{and} \quad \alpha_b^{id} = \frac{b-1}{4}.$$

For instance, with the least prime base larger or equal to 360, we obtain $d_{367}^{id} = 33672/367 \approx 91.75$ and $\alpha_{367}^{id} = 366/4 = 91.5$.

Let us now introduce a new definition, which we restrict to a prime base b for simplicity: a *linear digit scrambling* is a permutation π of the set Z_b of the form $\pi(k) = fk + g \pmod b$, where $f \neq 0$ and g are given in Z_b. If $g = 0$, we obtain

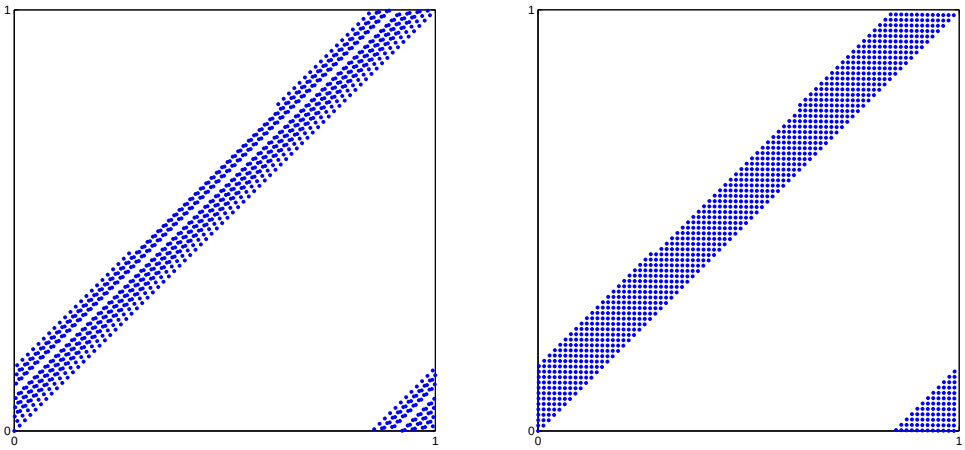

Figure 2. 1000 first points of GFunperm (left) and F (right), dimensions 39 and 40

the so-called *multipliers* f of our preceding papers [9], [14]. The additive factor g is a translation usually called *digital shift*. In base 367, with one of the best multipliers, $f = 97$, we get (π is identified to f) $d^{97}_{367} = 1088/367 \approx 2.96$ while $d^{id}_{367} \approx 91.75$, more than 30 times bigger!

Among all prime bases up to $b = 1301$, the best leading constant in (7.1), that is $\alpha^{\sigma}_b / \log b$, is obtained with base 233 with $\alpha^{89}_{233}/\log 233 \approx 0.44$. Among the 212 prime numbers we have considered, few such constants are above 0.52, mainly for small b. Recall that $\alpha^{id}_2 = 1/3$ and $\alpha^{id}_3 = 1/2$ so that $\alpha^{id}_2/\log 2 \approx 0.48$ and $\alpha^{id}_3/\log 3 \approx 0.45$. Good multipliers in any base often give bounds close to these exact values, whereas without scrambling, we would have $\alpha^{I}_b/\log b = (b-1)/4 \log b$, a quantity increasing to infinity with b.

8. Conclusion

In this paper we have reviewed discrepancy bounds for Halton sequences and their generalizations. We have provided numerical evidence demonstrating that the implied constant in these bounds—while being a useful tool to make comparisons in the asymptotic regime—could dramatically misrepresent the behavior of the bounds in the non-asymptotic regime. We also discussed the primordial importance of permutations, whose effect on the quality of a sequence is well understood theoretically in one dimension, but less so in higher dimensions.

In view of all this, we are seemlingly left with no good answer to the question: *how should one choose permutations and/or scrambling matrices to construct "good" generalized Halton or Faure sequences?* On one hand, as we have seen in the previous sections, there is a huge gap between what appears to be a good sequence depending

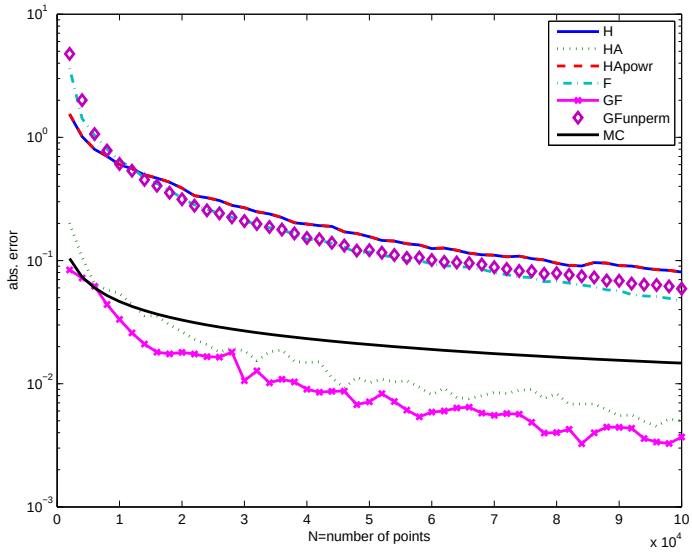

Figure 3. Absolute error for digital option problem (see [14] for more details) with $s = 75$: H and HApowr are indistinguishable

on whether we consider the asymptotic or non-asymptotic regime; on the other hand, tight theoretical results are only known for one-dimensional sequences. In an attempt to best take into account these two realities, the approach we adopted in the two recent papers [9, 14] is to make use of the better knowledge for one-dimensional sequences as a primary criterion for building sequences, while also looking at the discrepancy of low-dimensional projections in the non-asymptotic regime as a secondary criterion. This approach automatically avoids using the identity to permute the first digit, because as we just saw the identity yields a large discrepancy in one dimension. In the case of GF sequences though, not only the identity should be avoided, but also using the same permutations for different coordinates should be avoided as much as possible, as this causes the corresponding projections to be almost the same as if the identity had been used on each of them. We are working to overcome such difficulties with the use of larger bases, around $2s$, instead of the least prime base greater than s. We plan to report on this later in another paper.

To conclude the present study, we want to emphasize that like many people do, we think the superiority of QMC methods over MC for high dimensional integration, as in finance, comes from the fact that these problems involve only a reasonable set of "nevralgic" coordinates. Finding such sets (even roughly) could be helpful to remove a part of the mystery evoqued by Woźniakowski in his paper [27] at the end of Sections 3, 4 and, last but not least, 10. The remarks on effective dimension given in the introduction suggest directions for future research. We will pursue our efforts in these

directions in the near future.

Finally, recall that the great open conjecture in irregularities of distribution, i.e., the exact order for the extreme discrepancy of s-dimensional sequences is $(\log N)^s$, already starts with dimension $s = 2$. Hence it is quite hopeless to wait for theoretical results able to distinguish between permutations in dimensions greater than one and so empirical studies like [9] and [14] currently seem the only way to improve scramblings of quasi-random sequences.

References

1. E. I. Atanassov, On the discrepancy of the Halton sequences, *Math. Balkanica, New Series* **18.1-2** (2004), 15–32.
2. E. I. Atanassov and M. K. Durchova, Generating and testing the modified Halton sequences, In: Dimov and all (Eds.) Numerical Methods and Applications 2002, LNCS vol. 2542 (2003), 91–98.
3. H. Faure, Suites à faible discrépance dans $\mathbb{T}^s$, Publ. Dép. math., Université de Limoges, France (1980).
4. H. Faure, Discrépance de suites associées à un système de numération (en dimension un), *Bull. Soc. math. France* **109** (1981), 143–182.
5. H. Faure, Discrépance de suites associées à un système de numération (en dimension s), *Acta Arith.* **61** (1982), 337–351.
6. H. Faure, On the star-discrepancy of generalized Hammersley sequences in two dimensions, *Monats. Math.* **101** (1986), 291–300.
7. H. Faure, Good permutations for extreme discrepancy, *J. Number Theory* **42** (1992), 47–56.
8. H. Faure, Selection criteria for (random) generation of digital $(0, s)-$sequences, In: H. Niederreiter and D. Talay (Eds.) Monte Carlo and Quasi-Monte Carlo Methods 2004. Springer, Berlin Heidelberg New York (2006), 113–126.
9. H. Faure and C. Lemieux Generalized Halton sequences in 2008: A comparative study. *ACM Trans. Model. Comp. Sim.*, **19**, 4 (2009), Article 15.
10. H. Faure, C. Lemieux and X. Wang, Extensions of Atanassov's method for Halton sequences. Working paper (2010).
11. J. H. Halton, On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals, *Numer. Math.* **2** (1960), 184–90.
12. L. Kocis and W. J. Whiten, Computational investigations of low–discrepancy sequences, *ACM Trans. Math. Software*, **23–2** (1997), 266–294.
13. P. Kritzer, Improved upper bounds on the star discrepancy of $(t, m, s)-$nets and $(t, s)-$sequences, *J. Complexity*, **22** (2006), 336–347.
14. C. Lemieux and H. Faure, New perspectives on $(0, s)$-sequences. Monte Carlo and Quasi-Monte Carlo Methods 2008, L'Ecuyer P. and Owen A. (Eds.), Springer (2010), 113–130.
15. H. G. Meijer, The discrepancy of a $g-$adic sequence, *Indag. Math.*, **30** (1968), 54–66.
16. W. J. Morokoff and R. E. Caflisch, Quasi random sequences and their discrepancies, *SIAM J. Sci. Comput.*, **15.6** (1994), 1251–1279.

17. H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods.* CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia (1992).

18. H. Niederreiter and C. P. Xing, Quasirandom points and global functions fields. In: S. Cohen and H. Niederreiter (Eds) Finite Fields and Applications, London Math. Soc. Lectures Notes Series, **233** (1996), 269–296.

19. A. B. Owen, Randomly permuted (t, m, s)-nets and (t, s)-sequences. In: H. Niederreiter and P. Shiue (Eds.) Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing. Springer, Berlin Heidelberg New York (1995), 299–317.

20. A. Papageorgiou and J. F. Traub, Beating Monte Carlo, *Risk*, **9** (1996), 63–65.

21. R. Schürer and W. Schmid, MinT: a database for optimal net parameters. In: H. Niederreiter and D. Talay (Eds.) Monte Carlo and Quasi-Monte Carlo Methods 2004. Springer, Berlin Heidelberg New York (2006), 457–469.

22. I. M. Sobol', The distribution of points in a cube and the approximate evaluation of integrals, *U.S.S.R. Comput. Math. and Math. Phys.* **7** (1967), 86–112.

23. S. Tezuka, Polynomial arithmetic analogue of Halton sequences, *ACM Trans. Model. Comp. Sim.*, **3** (1993), 99–107.

24. S. Tezuka, *Uniform Random Numbers: Theory and Practice.* Kluwer Academic Publishers, Boston (1995).

25. B. Vandewoestyne and R. Cools, Good permutations for deterministic scrambled Halton sequences in terms of L_2-discrepancy, *J. Comput. and Appl. Math.*, **189(1-2)** (2006), 341–361.

26. X. Wang, C. Lemieux, H. Faure, A note on Atanassov's discrepancy bound for the Halton sequence. Technical report, University of Waterloo, Canada (2008).

27. H. Woźniakowski, Tractability of multivariate problems. In *Foundations of Computational Mathematics, Hong Kong, 2008*, F. Cucker, A. Pinkus and M. Todd (Eds.), London Math. Soc. Lecture Note Ser. 363, Cambridge University Press, (2009), 236–276.

Received ■■■

Author information

Henri Faure, Institut de Mathématiques de Luminy, UMR 6206 CNRS, 163 Av. de Luminy, case 907, 13288 Marseille cedex 9, France.
Email: faure@iml.univ-mrs.fr

Christiane Lemieux, Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West, Waterloo, ON, N2L 3G1, Canada.
Email: clemieux@uwaterloo.ca