Discrepancy Bounds for Mixed Sequences

Michael Gnewuch

Department of Computer Science, Columbia University,
1214 Amsterdam Avenue, MC 0401,
New York, 10027 NY, USA
email: gnewuch@cs.columbia.edu

November 15, 2009

A common measure for the uniformity of point distributions is the star discrepancy. Let λ_s denote the s-dimensional Lebesgue measure. Then the *star discrepancy* of a multiset $P = \{p_1, \dots, p_N\} \subset [0,1]^s$ is given by

$$D_N^*(P) := \sup_{\alpha \in [0,1]^s} \left| \lambda_s([0,\alpha)) - \frac{1}{N} \sum_{k=1}^{N} 1_{[0,\alpha)}(p_k) \right| ;$$

here $[0,\alpha)$ denotes the s-dimensional axis-parallel box $[0,\alpha_1) \times \cdots \times [0,\alpha_s)$ and $1_{[0,\alpha)}$ its characteristic function. For an infinite sequence p in $[0,1]^s$ we denote by $D_N^*(p)$ the discrepancy of its first N points. The smallest star discrepancy of any N-point set is

$$D^*(N,s) := \inf_{P \subset [0,1]^s,\, |P| = N} D_N^*(P),$$

and the *inverse of the star discrepancy* is given by

$$N^*(\varepsilon, s) := \inf\{N \in \mathbb{N} \,|\, D^*(N,s) \le \varepsilon\}.$$

There are bounds known describing the behavior of the star discrepancy in the number of points N and in the dimension s, see, e.g., [5, 2]. In [5] Heinrich, Novak, Wasilkowski, and Woźniakowski proved

$$D^*(N,s) \le C\sqrt{s/N} \quad \text{and} \quad N^*(\varepsilon,s) \le \lceil C^2 s \varepsilon^{-2} \rceil, \tag{1}$$

where the constant C does not depend on N, s or ε. The dependence of the inverse of the star discrepancy on s is optimal here; this was proved by a lower bound in [5], which was improved by Hinrichs in [6]: There exist constants $c, \varepsilon_0 > 0$ such that

$$D^*(N,s) \ge \min\{\varepsilon_0, cs/N\} \quad \text{and} \quad N^*(\varepsilon,s) \ge cs\varepsilon^{-1} \quad \text{for all } \varepsilon \in (0,\varepsilon_0). \tag{2}$$

Dagstuhl Seminar Proceedings 09391
Algorithms and Complexity for Continuous Problems
http://drops.dagstuhl.de/opus/volltexte/2009/2297

It is well known that the star discrepancy is closely related to the problem of numerical integration of certain function classes. The Koksma-Hlawka inequality tells us that the smaller the discrepancy of a multiset P, the better the worst-case error guarantee of the corresponding quasi-Monte Carlo cubature $\frac{1}{|P|}\sum_{p\in P} f(p)$, see, e.g., [7].

In many problems low-discrepancy point sets outperform random sets in moderate dimensions, but lose their effectiveness in high dimensions. For this reason researchers studied hybrid methods which try to use advantages of both Monte Carlo and quasi-Monte Carlo methods. One example are so-called mixed sequences used by Spanier [11] and studied further by Ökten and his co-authors [8, 10]. Mixed sequences showed a good performance in many numerical tests. To obtain a more objective measure for the quality of mixed sequences, probabilistic bounds on their star discrepancy have been derived.

Let $q = (q_k)$ be a sequence in $[0, 1)^d$ (ideally with a low discrepancy), and let $X = (X_k)$ be a sequence of independent and uniformly distributed random variables in $[0, 1)^{s-d}$. The resulting s-dimensional sequence $m = (m_k) = (q_k, X_k)$ is then called a *mixed sequence*.

The following probabilistic bound is due to Ökten, Tuffin, and Burago [10]:

Theorem 1. *For $s \in \mathbb{N}$, $d \in \{1, \dots, s\}$, and $\varepsilon \in (0, 1]$ we have*

$$\mathbb{P}\big(D^*_N(m) - D^*_N(q) < \varepsilon\big) \geq 1 - 2\exp\left(-\frac{\varepsilon^2 N}{2}\right) \quad \textit{for } N \textit{ sufficiently large.} \tag{3}$$

In [10] the authors did not investigate how large N actually has to be or on which parameters the required size of N really depends. For fixed q let $N(q; s, \varepsilon)$ be the smallest number such that (3) holds for all $N \geq N(q; s, \varepsilon)$. With the help of the bound (2) it was shown in [3] that for $\varepsilon < 1/64\mathrm{e}^2$ we have

$$N(q; s, \varepsilon) > \frac{1}{64\mathrm{e}^2}\frac{s}{\varepsilon} \quad \text{for all but finitely many } s. \tag{4}$$

This shows that $N(q; s, \varepsilon)$ depends significantly on s and ε.

In the literature one can find bounds of the form

$$\mathbb{P}\big(D^*_N(m) - D^*_N(q) < \varepsilon\big) \geq 1 - \frac{1}{\varepsilon^2}\frac{1}{N}(D^*_N(q) + 1)$$

for the star discrepancy and and analogous bounds for the so-called G-discrepancy. As explained in [4, 9] those results are unfortunately incorrect.

In [3] the following result was shown, which holds indeed for all values of N, without any restrictions:

Theorem 2. *For $s \in \mathbb{N}$, $d \in \{1, \dots, s\}$, and $\varepsilon \in (0, 1]$ we have*

$$\mathbb{P}\big(D^*_N(m) - D^*_N(q) < \varepsilon\big) > 1 - 2\,(2\mathrm{e})^s(2\varepsilon^{-1} + 1)^s \exp\left(-\frac{\varepsilon^2 N}{2}\right). \tag{5}$$

Let $\theta \in [0,1)$. Then we have with probability strictly larger than θ

$$D_N^*(m) < D_N^*(q) + \sqrt{\frac{2}{N}} \left(s \ln(\rho) + \ln\left(\frac{2}{1-\theta}\right)\right)^{1/2}, \tag{6}$$

where $\rho = \rho(N,s) := 6\mathrm{e}(\max\{1\,,\, N/(2\ln(6\mathrm{e})s)\})^{1/2}$.

The factor $2\,(2\mathrm{e})^s(2\varepsilon^{-1}+1)^s$ on the right hand side of of the estimate (5), which grows exponentially in s, does not mean that the estimate is weak. In fact it was shown in [3] that in any bound of the form

$$\mathbb{P}\big(D_N^*(m) - D_N^*(q) \le \varepsilon\big) \ge 1 - f(q;s,\varepsilon)\exp\left(-\frac{\varepsilon^2 N}{2}\right) \tag{7}$$

which holds for all N the function $f(q;s,\varepsilon)$ has to increase at least exponentially in s for ε sufficiently small if $\lim_{N\to\infty} D_N^*(q) = 0$.

The technique to prove Theorem 2 is similar to the one used in [1] to prove bounds on the discrepancy of the output set of a derandomized algorithm to generate samples with relatively small star discrepancy. The essential tools are bracketing covers from [2] to discretize the discrepancy, union bounds, and large deviation bounds of Chernoff-Hoefding type.

References

[1] Doerr, B., Gnewuch, M., Kritzer, P., Pillichshammer, F.: Component-by-component construction of low-discrepancy point sets of small size. Monte Carlo Methods Appl. 14 (2008), 129–149.

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

[3] Gnewuch, M.: On Probabilistic Results for the Discrepancy of a Hybrid-Monte Carlo Sequence. J. Complexity 25 (2009), 312-317.

[4] Gnewuch, M., Roşca, A. V.: On G-discrepancy and mixed Monte Carlo and quasi-Monte Carlo sequences. Acta Uni. Apul. 18 (2009), 97-110.

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

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

[7] Niederreiter, H.: Random Number Generation and Quasi-Monte Carlo Methods. CBMS-NSF Series in Applied Mathematics 63, SIAM, Philadelphia, 1992.

[8] Ökten, G.: Contributions to the Theory of Monte Carlo and Quasi-Monte Carlo Methods. PhD thesis, Claremont Graduate School, Claremont, California, 1997.

[9] Ökten, G., Gnewuch, M.: Correction of a proof in “A probabilistic result on the discrepancy of a hybrid-Monte Carlo sequence and applications”. Monte Carlo Methods Appl. 15 (2009), 169-172.

[10] Ökten, G., Tuffin, B., Burago, V.: A central limit theorem and improved error bounds for a hybrid-Monte Carlo sequence with applications in computational finance. J. Complexity 22 (2006), 435–458.

[11] Spanier, J.: Quasi-Monte Carlo methods for particle transport problems. In: Niederreiter, H., Shiue, P. J.-S. (Eds.), Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, 121–148, Lecture Notes in Statistics 106, Springer, Berlin, 1995.