Discrepancy and Uniformity

STEVEN FINCH

February 19, 2004

Let $X = \{x_n\}_{n=1}^{\infty}$ be an infinite sequence of points in the interval $[0,1)$ and let X_N denote the finite subsequence $\{x_n\}_{n=1}^{N}$. Define, for each N, the **discrepancy**

$$D_N(X) = \sup_{0 \le a < b < 1} \left| \frac{|X_N \cap [a,b)|}{N} - (b-a) \right|$$

and the **star discrepancy**

$$D_N^*(X) = \sup_{0 \le c < 1} \left| \frac{|X_N \cap [0,c)|}{N} - c \right|.$$

It can be proved that $1/N \le D_N \le 1$ and $1/(2N) \le D_N^* \le D_N \le 2D_N^*$. The sequence X is uniformly distributed in $[0,1)$ if and only if $\lim_{N\to\infty} D_N(X) = 0$. We are interested in low-discrepancy sequences, that is, sequences X for which $D_N(X)$ is small for all N. The efficient construction of such X is essential in quasi-Monte Carlo algorithms used, for example, to approximate a multivariate integral or to simulate certain random processes [1, 2, 3].

On the one hand, Béjian [4, 5] showed that

$$S(X) = \limsup_{N\to\infty} \frac{N}{\ln(N)} D_N(X) \ge C$$

$$S^*(X) = \limsup_{N\to\infty} \frac{N}{\ln(N)} D_N^*(X) \ge C^*$$

for all sequences X, where

$$C = \max_{r \ge 2} \frac{r-2}{4(r-1)\ln(r)} = 0.120386..., \qquad C^* = \frac{1}{24\ln(2)} = 0.060112....$$

This is a consequence of work by van Aardenne-Ehrenfest [6, 7], Roth [8] and Schmidt [9] regarding the unavoidable irregularities that occur in any point distribution. Improvement is likely. On the other hand, there are special sequences X for which [10, 11, 12]

$$S(X) \le \frac{23}{35\ln(6)} = 0.366758..., \qquad S^*(X) \le \frac{1919}{3454\ln(12)} = 0.223584...$$

---

and we shall discuss such examples shortly. The gap between lower and upper bounds is surprisingly wide: It would be good someday for these estimates to be tightened.

Students of statistics will recognize $D_N^*(X)$ as the **Kolmogorov-Smirnov one-sample statistic**, under the hypothesis that the sequence X_N is a random sample drawn from a Uniform $(0,1)$ distribution. Call this hypothesis H_0. We have the following asymptotic result [13, 14]:

$$\lim_{N\to\infty} \mathrm{P}\left(\sqrt{N}D_N^*(X) \le z \mid H_0 \text{ is true}\right) = 1 - 2\sum_{n=1}^{\infty}(-1)^{n-1}e^{-2n^2z^2}.$$

Call this expression $\theta(z)$. In an experimental data analysis setting, if $\sqrt{N}D_N^*(X)$ exceeds a sufficiently large threshold w (in the sense that the probability $1-\theta(w)$ is suitably small), then one must doubt the truth of H_0.

0.1. Scrambled van der Corput Sequences. Given an integer $n \ge 1$, write the base b representation for $n-1$ as

$$n - 1 = \sum_{k=0}^{\infty} m_k(n)b^k, \qquad m_k \in B$$

where $B = \{0, 1, \ldots b-1\}$. Then, for any permutation $\sigma : B \to B$, the scrambled van der Corput sequence X_b^σ has n^{th} term given by

$$x_n = \sum_{j=0}^{\infty} \sigma(m_j(n))b^{-j-1}.$$

For example, if $b = 2$ and σ is the identity permutation ε, then

$$X_2^\varepsilon = \{0,\ 0.1,\ 0.01,\ 0.11,\ 0.001,\ 0.101,\ 0.011, ...\},$$

that is, x_n is simply the reflection of $n-1$ across the decimal point. Haber [15, 16, 17] computed that $S^*(X_2^\varepsilon) = 1/(3\ln(2)) = 0.480898....$ Faure [10] proved the more general result

$$S(X_b^\varepsilon) = S^*(X_b^\varepsilon) = \begin{cases} \dfrac{b-1}{4\ln(b)} & \text{if } b \text{ is odd} \\ \dfrac{b^2}{4(b+1)\ln(b)} & \text{if } b \text{ is even.} \end{cases}$$

Introducing a non-identity permutation σ provides the smallest discrepancy currently known [12]:

$$S(X_{36}^\sigma) = 23/(35\ln(6)) = 0.366758...$$

where σ has cycle structure

$$\begin{array}{l}(0)\ (1\ 25\ 33\ 16\ 23\ 4\ 31\ 30\ 10\ 34\ 24\ 21\ 28\ 2\ 17)\ (3\ 7)\\ (5\ 11\ 22\ 14\ 29\ 19\ 32\ 6\ 20\ 8\ 27\ 26\ 12)\ (9\ 13\ 15)\ (18)\ (35).\end{array}$$

It is useful to generalize X_b^σ to X_b^Σ, where Σ is a sequence of permutations σ. If $\Sigma = \{\sigma_j\}_{j=0}^\infty$, then the n^{th} term of X_b^Σ is simply given by

$$x_n = \sum_{j=0}^{\infty} \sigma_j(m_j(n)) b^{-j-1}.$$

For example, if $j \geq 0$ is an integer, let h be the smallest integer $\geq \max\{1, \sqrt{j}\}$. If $h(h-1)+1 \leq j \leq h^2$, define σ_j to be the permutation

$$(0)\ (1\ 5)\ (2\ 9)\ (3)\ (4\ 7)\ (6\ 10)\ (8)\ (11);$$

otherwise define σ_j to be the permutation

$$(0\ 11)\ (1\ 6)\ (2)\ (3\ 8)\ (4)\ (5\ 10)\ (7)\ (9).$$

The alternating character of Σ plays a role in reducing the star discrepancy to [10]

$$S^*(X_{12}^\Sigma) = 1919/(3454 \ln(12)) = 0.223584...$$

Again, this is the smallest such value currently known.

0.2. $\{n\alpha\}$-Sequences. Let $\alpha > 0$ be irrational. Define a sequence Y^α to have n^{th} term [18]

$$y_n = \{n\alpha\} = n\alpha \bmod 1,$$

that is, the fractional part of $n\alpha$. Let $D_N(\alpha) = D_N(Y^\alpha)$ for convenience. It is known that $D_N(\alpha) \to 0$ as $N \to \infty$, just as for van der Corput sequences. The corresponding values of $S(\alpha)$ and $S^*(\alpha)$ are not as small as earlier, but are nevertheless interesting.

Dupain & Sós [19] proved that

$$\inf_\alpha S^*(\alpha) = S^*(\sqrt{2}) = \frac{1}{4\ln(1+\sqrt{2})} = 0.283648...$$

and Schoissengeier [20, 21, 22, 23] expressed $S^*(\alpha)$ in terms of the continued fraction expansion of α. Baxa [24, 25] demonstrated that the image of the set of all irrational α under the map $\alpha \mapsto S^*(\alpha)$ is the ray $[S^*(\sqrt{2}), \infty]$, which contrasts sharply against the Lagrange and Markov spectra [26].

Ramshaw [27] proved that

$$S(\varphi) = \frac{1}{5\ln(\varphi)} = 0.415617...$$

where $\varphi = (1+\sqrt{5})/2$ is the Golden mean. A proof that

$$\inf_{\alpha} S(\alpha) = S(\varphi)$$

has never been published [28, 29]; a gap as such in the literature deserves to be filled.

0.3. Self-Similar Sequences. (At some later time, we will introduce the work of Borel [30, 31] on a sequence defined via the recursive bit substitution $0 \mapsto 01$ and $1 \mapsto 010$.)

0.4. Erdös-Turán Inequality. Erdös & Turán [32] proved that there exist constants c_1, c_2 such that

$$D_N(X) \leq \frac{c_1}{K+1} + c_2 \sum_{k=1}^{K} \frac{1}{k} \left| \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i k x_n} \right|$$

for any positive integers N, K and any sequence $X \subseteq [0,1)$. There is considerable flexibility in the choice of the two constants, as indicated here:

$(c_1, c_2) = (6, \frac{4}{\pi})$	(Kuipers & Niederreiter [1])
$(c_1, c_2) = (1, 3)$	(Baker [33] and Montgomery [34])
$(c_1, c_2) = (1, 1)$	(Mauduit, Rivat & Sárközy [35]).

Rivat & Tenenbaum [36], building on the work of Vaaler [37], recently determined constants that are believed to be close to best for the Erdös-Turán inequality:

$$(c_1, c_2) = \left(1, \tfrac{2}{\pi}\gamma\right) = (1, 0.6527196578...)$$

$$(c_1, c_2) = \left(1+\xi, \tfrac{2}{\pi}\right) = (1.1434819845..., 0.6366197723...)$$

where

$$f(t) = \sqrt{[\pi t(1-t)\cot(\pi t) + t]^2 + [\pi t(1-t)]^2}, \qquad 0 \leq t \leq 1,$$

$$\gamma = \max_{0 \leq t \leq 1} f(t) = 1.0252896410...,$$

$$g(x,t) = \left(1 - 3x^2 + 3x^2 \left|\cos\left(\tfrac{\pi t}{3x}\right)\right|\right) f(t), \qquad 0 \leq x \leq \tfrac{\sqrt{3}}{3},$$

and $\xi = 0.1434819845...$ is the smallest value of x for which $\max_{0 \leq t \leq 1} g(x,t) \leq 1$. In fact, Rivat & Tenenbaum found a one-parameter *family* of admissible constants (c_1, c_2), but we have indicated only the endpoints of this family.

A similar set of formulas occur in the determination of close-to-best constants for the Berry-Esseen inequality. (This is a somewhat different version of the inequality than that discussed in [38].) Let F, G be two distribution functions with corresponding characteristic functions φ, ψ. Assume that G is differentiable and that $\sup_x |G'(x)| = M < \infty$. Then there exist constants c_1, c_2 such that

$$\sup_x |F(x) - G(x)| \leq c_1 \frac{M}{T} + c_2 \int\limits_{-T}^{T} \left| \frac{\varphi(t) - \psi(t)}{t} \right| dt$$

for all $T > 0$. Admissible values for these constants include

$$\begin{array}{ll} (c_1, c_2) = (\frac{24}{\pi}, \frac{1}{\pi}) & \text{(Feller [39] and Loève [40])} \\ (c_1, c_2) = (\pi, \frac{1}{2\pi}\gamma) & \text{(Vaaler [37] and Tenenbaum [41])}, \end{array}$$

where $\gamma/(2\pi) = 0.1631799144...$ and γ is exactly as before. The new approach in [36] can perhaps be applied here too.

References

[1] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, Wiley, 1974; MR0419394 (54 #7415).

[2] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, 1992; MR1172997 (93h:65008).

[3] M. Drmota and R. F. Tichy, *Sequences, Discrepancies and Applications*, Lect. Notes in Math. 1651, Springer-Verlag, 1997; MR1470456 (98j:11057).

[4] R. Béjian, Minoration de la discrépance à l'origine d'une suite quelconque, *Annales Faculté Sciences Toulouse Math.* 1 (1979) 201–213; MR0568146 (82b:10076).

[5] R. Béjian, Minoration de la discrépance d'une suite quelconque sur T, *Acta Arith.* 41 (1982) 185–202; MR0674832 (83k:10101).

[6] T. van Aardenne-Ehrenfest, Proof of the impossibility of a just distribution of an infinite sequence of points over an interval, *Proc. Konink. Nederl. Akad. Wetensch. Sci. Sect.* 48 (1945) 266–271; *Indag. Math.* 7 (1945) 71–76; MR0015143 (7,376l).

[7] T. van Aardenne-Ehrenfest, On the impossibility of a just distribution, *Proc. Konink. Nederl. Akad. Wetensch. Sci. Sect.* 52 (1949) 734–739; *Indag. Math.* 11 (1949) 264–269; MR0032717 (11,336d).

[8] K. F. Roth, On irregularities of distribution, *Mathematika* 1 (1954) 73–79; MR0066435 (16,575c).

[9] W. M. Schmidt, Irregularities of distribution. VII, *Acta Arith.* 21 (1972) 45–50; MR0319933 (47 #8474).

[10] H. Faure, Discrépances de suites associées à un système de numération (en dimension un), *Bull. Soc. Math. France* 109 (1981) 143–182; available online at http://www.numdam.org/numdam-bin/browse; MR0623787 (82i:10069).

[11] A. Thomas, Discrépance en dimension un, *Annales Faculté Sciences Toulouse Math.* 10 (1989) 369–399; MR1425493 (97g:11083).

[12] H. Faure, Good permutations for extreme discrepancy, *J. Number Theory* 42 (1992) 47–56; MR1176419 (93j:11049).

[13] M. H. DeGroot, *Probability and Statistics*, Addison-Wesley, 1975, pp. 463–471.

[14] J. D. Gibbons, *Nonparametric Statistical Inference*, McGraw-Hill, 1971, pp. 75–85; MR0286223 (44 #3437).

[15] S. Haber, On a sequence of points of interest for numerical quadrature, *J. Res. Nat. Bur. Standards B* 70 (1966) 127–136; MR0203938 (34 #3785).

[16] R. Béjian and H. Faure, Discrépance de la suite de van der Corput, *C. R. Acad. Sci. Paris Sér. A-B* 285 (1977) A313–A316; MR0444600 (56 #2950).

[17] R. Béjian and H. Faure, Discrépance de la suite de van der Corput, *Séminaire Delange-Pisot-Poitou (Théorie des nombres)*, 19^{e} année, 1977/78, n. 13, pp. 1–14; MR0520308 (80b:10078).

[18] S. R. Finch, Pisot-Vijayaraghavan-Salem constants, *Mathematical Constants*, Cambridge Univ. Press, 2003, pp. 192–199.

[19] I. Dupain and V. T. Sós, On the discrepancy of $(n\alpha)$ sequences, *Topics in Classical Number Theory*, v. I, Proc. 1981 Budapest conf., Colloq. Math. Soc. János Bolyai 34, North-Holland, 1984, pp. 355–387; MR0781147 (87c:11065).

[20] J. Schoissengeier, On the discrepancy of $(n\alpha)$, *Acta Arith.* 44 (1984) 241–279; MR0774103 (86c:11056).

[21] J. Schoissengeier, On the discrepancy of $(n\alpha)$. II, *J. Number Theory* 24 (1986) 54–64; MR0852190 (88d:11074).

[22] J. Schoissengeier, The discrepancy of $(n\alpha)_{n\geq 1}$, *Math. Annalen* 296 (1993) 529–545; MR1225989 (94e:11086).

[23] C. Baxa and J. Schoissengeier, Minimum and maximum order of magnitude of the discrepancy of $(n\alpha)$, *Acta Arith.* 68 (1994) 281–290; MR1308128 (95j:11073).

[24] C. Baxa, On the discrepancy of the sequence $(n\alpha)$, *J. Number Theory* 55 (1995) 94–107; MR1361562 (96i:11086).

[25] C. Baxa, On the distribution of the sequence $(n\alpha)$ with transcendental α, *Acta Arith.* 81 (1997) 357–363; MR1472815 (98m:11080).

[26] S. R. Finch, Freiman's constant, *Mathematical Constants*, Cambridge Univ. Press, 2003, pp. 199–203.

[27] L. Ramshaw, On the discrepancy of the sequence formed by the multiples of an irrational number, *J. Number Theory* 13 (1981) 138–175; MR0612680 (82k:10071).

[28] C. Baxa, Comparing the distribution of $(n\alpha)$-sequences, *Acta Arith.* 94 (2000) 345–363; MR1779947 (2001e:11086).

[29] J.-P. Borel, Symbolic representation of piecewise linear functions on the unit interval and application to discrepancy, *Theoret. Comput. Sci.* 123 (1994) 61–87; MR1257204 (94m:11030).

[30] J.-P. Borel, Sur une suite de fonctions liées à un procédé itératif. I, *Publ. Dépt. Math. Limoges* 12 (1990) 1–16.

[31] J.-P. Borel, Sur une suite de fonctions liées à un procédé itératif. II, *J. Théor. Nombres Bordeaux* 5 (1993) 235–261; MR1265904 (95h:11072).

[32] P. Erdös and P. Turán, On a problem in the theory of uniform distribution. II, *Proc. Konink. Nederl. Akad. Wetensch. Sci. Sect.* 51 (1948) 1262–1269; *Indag. Math.* 10 (1948) 406–413; MR0027896 (10,372d).

[33] R. C. Baker, *Diophantine Inequalities*, Oxford Univ. Press, 1986, p. 19; MR0865981 (88f:11021).

[34] H. L. Montgomery, *Ten Lectures on the Interface between Analytic Number Theory and Harmonic Analysis*, Amer. Math. Soc., 1994, p. 8; MR1297543 (96i:11002).

[35] C. Mauduit, J. Rivat and A. Sárközy, On the pseudo-random properties of n^c, *Illinois J. Math.* 46 (2002) 185–197; available online at http://www.math.uiuc.edu/~hildebr/ijmwww/46-1/; MR1936084 (2003h:11088).

[36] J. Rivat and G. Tenenbaum, Constantes d'Erdös-Turán, *Ramanujan J.*, to appear (2004); available online at http://www.iecn.u-nancy.fr/~tenenb/PUBLIC/Prepublications_et_publications/ET.pdf.

[37] J. D. Vaaler, Some extremal functions in Fourier analysis, *Bull. Amer. Math. Soc.* 12 (1985) 183–216; MR0776471 (86g:42005).

[38] S. R. Finch, Berry-Esseen constant, *Mathematical Constants*, Cambridge Univ. Press, 2003, pp. 264–266.

[39] W. Feller, *An Introduction to Probability Theory and its Applications*, v. II., 2nd ed., Wiley, 1971, pp. 536–538, 542–546; MR0270403 (42 #5292).

[40] M. Loève, *Probability Theory. I*, 4th ed., Springer-Verlag, 1977, p. 297; MR0651017 (58 #31324a).

[41] G. Tenenbaum, *Introduction to Analytic and Probabilistic Number Theory*, Cambridge Univ. Press, 1995, pp. 240–241; MR1342300 (97e:11005b).