Normal numbers and the normality measure

Christoph Aistleitner*

Abstract

In a paper published in this journal, Alon, Kohayakawa, Mauduit, Moreira, and Rödl proved that the minimal possible value of the normality measure of an N-element binary sequence satisfies

$$\left(\frac{1}{2}+o(1)\right)\log_2 N \leq \min_{E_N\in\{0,1\}^N} \mathcal{N}(E_N) \leq 3N^{1/3}(\log N)^{2/3}$$

for sufficiently large N, and conjectured that the lower bound can be improved to some power of N. In this note it is observed that a construction of Levin of a normal number having small discrepancy gives a construction of a binary sequence E_N with $\mathcal{N}(E_N) = \mathcal{O}\left((\log N)^2\right)$, thus disproving the conjecture above.

Let a finite binary sequence $E_N = (e_1, \ldots, e_N) \in \{0,1\}^N$ be given. For $k \geq 1,\ M \geq 1$ and $X \in \{0,1\}^k$, we set

$$T(E_N, M, X) = \#\left\{n:\ 0 \leq n < M,\ \text{and } (e_{n+1}, \ldots, e_{n+k}) = X\right\},$$

which means that $T(E_N, M, X)$ counts the number of occurrences of the pattern X among the first $M+k-1$ elements of E_N. The normality measure $\mathcal{N}(E_N)$ is defined as

$$\mathcal{N}(E_N) = \max_{1\leq k\leq \log_2 N}\ \max_{X\in\{0,1\}^k}\ \max_{1\leq M\leq N+1-k} \left|T(E_N, M, X) - \frac{M}{2^k}\right|. \tag{1}$$

The normality measure was introduced in 1997 by Mauduit and Sárközy [8], together with several other measures of pseudorandomness for finite binary sequences[1]. In two papers, Alon, Kohayakawa, Mauduit, Moreira and Rödl [2, 3] studied the *minimal* and the *typical* values of the normality measure (and other measures of pseudorandomness). Concerning the typical value of $\mathcal{N}$, they proved that for any $\varepsilon > 0$ there exist $\delta_1, \delta_2 > 0$ such that for E_N uniformly distributed in $\{0,1\}^N$

$$\delta_1\sqrt{N} \leq \mathcal{N}(E_N) \leq \delta_2\sqrt{N}$$

*Department of Applied Mathematics, School of Mathematics and Statistics, University of New South Wales, Sydney NSW 2052, Australia. e-mail: aistleitner@math.tugraz.at. Research supported by a Schrödinger scholarship of the Austrian Research Foundation (FWF).

MSC 2010: Primary Classification: 68R15, Secondary Classification: 11K45, 11K16

keywords: pseudorandom sequence, normality measure, discrepancy, normal numbers

[1] Strictly speaking, Mauduit and Sárközy defined their pseudorandomness measures for sequences over the alphabet $\{-1, 1\}$ (instead of $\{0, 1\}$, as in the present paper). It is more convenient for our purpose to study sequences defined over the alphabet $\{0, 1\}$ (since they can be related to the binary representation of real numbers), and the definitions have been modified accordingly.

holds with probability at least $1-\varepsilon$ for sufficiently large N, and conjectured that a limit distribution of $\mathcal{N}(E_N)/\sqrt{N}$ exists; the latter was recently confirmed [1]. Concerning the minimal value of $\mathcal{N}$, Alon *et al.* proved that

$$\left(\frac{1}{2}+o(1)\right)\log_2 N \le \min_{E_N\in\{0,1\}^N}\mathcal{N}(E_N) \le 3N^{1/3}(\log N)^{2/3} \tag{2}$$

for sufficiently large N. The lower bound in (2) is based on a relatively simple combinatorial argument. The proof of the upper bound in (2) is rather elaborate; however, it is entirely constructive, using an explicit algebraic construction based on finite fields. Concerning a possible improvement of (2), Alon *et al.* write in [2]

> "We suspect that the logarithmic lower bound in [equation (2)] is far from the truth."

and formulate the open problem

> "Is there an absolute constant $\alpha > 0$ for which we have $\min_{E_N}\mathcal{N}(E_N) > N^\alpha$ for all large enough N?"

In [3] they write

> "The authors believe that the answer to [the open problem above] is positive."

The purpose of this note is to draw attention to a result of Levin [7] concerning the existence of a normal number with small discrepancy. This result implies the following upper bound on the minimal value of $\mathcal{N}$, closing the gap between upper and lower bounds up to a logarithmic factor, and disproving the conjecture of Alon *et al.* stated above.

Theorem 1. *There exists a constant c such that*

$$\min_{E_N\in\{0,1\}^N}\mathcal{N}(E_N) \le c(\log N)^2$$

for sufficiently large N.

Normal numbers were introduced by Borel in 1909. Let $z \in [0,1)$ be a real number, and denote its binary expansion by $z = 0.z_1z_2z_3\ldots$. Then z is called a *normal number* (in base 2, which is the only base that we are interested in in the present paper) if for any $k \ge 1$ and any block of digits $X \in \{0,1\}^k$ the relative asymptotic frequency of the number of appearances of X in the binary expansion of z is 2^{-k}. Using the terminology from the beginning of this note and writing $Z_N = (z_1,\ldots,z_N)$ for the sequence of the first N digits of z, this can be expressed as

$$\lim_{N\to\infty}\frac{T(Z_N, N+1-k, X)}{N} = 2^{-k},$$

where k is the length of X. Borel proved that almost all numbers (in the sense of Lebesgue measure) are normal. There exist many constructions of normal numbers, the first of them being obtained by concatenating the digital representations of the positive integers (Champernowne, 1933), primes (Copeland and Erdős, 1946) and values of polynomials (Davenport and Erdős, 1952). Deciding whether a given real number is normal or not is a very difficult problem, and it is unknown whether constants such as $\sqrt{2}$, e and π are normal or not.

In an informal way, normal numbers are often considered as numbers showing "random" behavior (which is justified by the aforementioned theorem of Borel). In fact, different variants of the normality property were considered as tests for pseudorandomness of (infinite) sequences of digits, for example in the monograph of Knuth [5] on *The Art of Computer Programming*, and the normality measure of Mauduit and Sárközy is a quantitative version of such a pseudorandomness test for the case of a *finite* sequence of digits. For a discussion of the connection between normal numbers, pseudorandomness of (finite) sequences, and pseudorandom number generators, see the book of Knuth and the papers of Mauduit and Sárközy *On finite pseudorandom binary sequences I-VII*, as well as [4, 9].

To proceed further, we need some notation. A sequence of real numbers $(y_n)_{n\geq 1}$ from the unit interval is called *uniformly distributed modulo one* (u.d. mod 1) if for all intervals $[a,b) \subset [0,1)$ the limit relation

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{[a,b)}(y_n) = b - a$$

holds. The quality of the uniform distribution of a sequence is measured in terms of the *discrepancy* D_N, which for $N \geq 1$ is defined as

$$D_N(y_1, \ldots, y_N) = \sup_{0\leq a<b\leq 1} \left| \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{[a,b)}(y_n) - (b-a) \right|.$$

A sequence is u.d. mod 1 if and only if its discrepancy tends to zero as $N \to \infty$.

By an observation of Wall [11], a number z is normal (in base 2) if and only if the sequence $\left(\langle 2^{n-1}z\rangle\right)_{n\geq 1}$, where $\langle\cdot\rangle$ denotes the fractional part, is u.d. mod 1. Korobov [6] posed the problem of finding a function $\psi(N)$ with maximal decay for which there exists a number z such that

$$D_N\left(z, \langle 2z\rangle, \ldots, \langle 2^{N-1}z\rangle\right) \leq \psi(N), \qquad N \geq 1.$$

The best result concerning this question is currently due to Levin [7], who proved (constructively, by giving an explicit example) the existence of a z for which

$$D_N\left(z, \langle 2z\rangle, \ldots, \langle 2^{N-1}z\rangle\right) = \mathcal{O}\left(\frac{(\log N)^2}{N}\right) \qquad \text{as} \qquad N \to \infty. \tag{3}$$

This result should be compared with a lower bound of Schmidt [10], stating that for *any* sequence $(y_n)_{n\geq 1}$

$$D_N(y_1, \ldots, y_N) \geq c_{\text{abs}} \frac{\log N}{N}$$

for infinitely many N. Thus Korobov's problem is solved, up to a logarithmic factor.

In view of Levin's result (3), Theorem 1 is a direct consequence of the following lemma.

Lemma 1. *Let $z \in [0,1)$ be a real number, whose binary expansion is given by $z = 0.z_1z_2z_3\ldots$, and assume that there exists a nondecreasing function $\Phi(N)$ such that*

$$D_N\left(z, \langle 2z\rangle, \ldots, \langle 2^{N-1}z\rangle\right) \leq \frac{\Phi(N)}{N}, \qquad N \geq 1. \tag{4}$$

Then for each $N \geq 1$ the binary sequence $Z_N = (z_1, \ldots, z_N)$ satisfies

$$\mathcal{N}(Z_N) \leq \Phi(N).$$

Proof: We may suppose that z is not a dyadic rational. For any k, any binary sequence X of length k, and any $M \leq N + 1 - k$, it is immediately seen that $T(Z_N, M, X)$ is the number of indices $0 \leq n \leq M - 1$ for which $\langle 2^n z \rangle$ falls in a certain interval of length 2^{-k}. Hence

$$\left| T(Z_N, M, X) - \frac{M}{2^k} \right| \leq M D_M \left(z, \langle 2z \rangle, \ldots, \langle 2^{M-1} z \rangle \right) \leq \Phi(M) \leq \Phi(N),$$

so $\mathcal{N}(Z_N) \leq \Phi(N)$, as claimed. By the remark before the statement of the lemma, this also proves Theorem 1.

References

[1] C. Aistleitner. On the limit distribution of the normality measure of random binary sequences. Preprint. Available at `http://arxiv.org/abs/1301.6454`.

[2] N. Alon, Y. Kohayakawa, C. Mauduit, C. G. Moreira, and V. Rödl. Measures of pseudorandomness for finite sequences: minimal values. *Combin. Probab. Comput.*, 15(1-2):1–29, 2006.

[3] N. Alon, Y. Kohayakawa, C. Mauduit, C. G. Moreira, and V. Rödl. Measures of pseudorandomness for finite sequences: typical values. *Proc. Lond. Math. Soc. (3)*, 95(3):778–812, 2007.

[4] D. H. Bailey and R. E. Crandall. Random generators and normal numbers. *Experiment. Math.*, 11(4):527–546 (2003), 2002.

[5] D. E. Knuth. *The art of computer programming. Vol. 2.* Addison-Wesley Publishing Co., Reading, Mass., second edition, 1981.

[6] N. M. Korobov. Numbers with bounded quotient and their applications to questions of Diophantine approximation. *Izv. Akad. Nauk SSSR. Ser. Mat.*, 19:361–380, 1955.

[7] M. B. Levin. On the discrepancy estimate of normal numbers. *Acta Arith.*, 88(2):99–111, 1999.

[8] C. Mauduit and A. Sárközy. On finite pseudorandom binary sequences. I. Measure of pseudorandomness, the Legendre symbol. *Acta Arith.*, 82(4):365–377, 1997.

[9] H. Niederreiter. *Random number generation and quasi-Monte Carlo methods*, volume 63 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[10] W. M. Schmidt. Irregularities of distribution. VII. *Acta Arith.*, 21:45–50, 1972.

[11] D. D. Wall. *Normal numbers.* Ph.D. thesis, University of California, Berkeley, 1949.