On some questions of V.I. Arnold on the stochasticity of geometric and arithmetic progressions

Christoph Aistleitner

Department of Mathematics and Statistics, Graduate School of Science, Kobe University, 1-1 Rokkodai, Nada-ku, Kobe 657-8501, Japan

E-mail: aistleitner@math.tugraz.at

Received 6 March 2015, revised 25 June 2015
Accepted for publication 14 August 2015
Published 8 September 2015

Recommended by Dr Dmitry Dolgopyat

Abstract
In some of his final papers, V I Arnold studied pseudorandomness properties of finite deterministic sequences, which he measured in terms of their ‘stochasticity parameter’. In the present paper we illustrate the background in probability theory and number theory of some of his considerations, and give answers to some of the questions raised in his papers.

Keywords: pseudorandomness, stochasticity, arithmetic progressions, geometric progressions, equidistribution
Mathematics Subject Classification numbers: 11K45, 60C05, 37A50, 37A45

1. Introduction
In some of his final papers, V I Arnold investigated pseudorandomness properties of finite deterministic sequences of integers or reals. Amongst several other types of sequences, he in particular investigated arithmetic progressions, geometric progressions, continued fraction expansions, permutations and quadratic residues; see the papers [5–15] in the bibliography below. To quantify the degree of pseudorandomness of these sequences, Arnold used a ‘stochasticity parameter’ $\lambda_n$, and several of the mentioned papers of Arnold begin with a short history of the introduction of this stochasticity parameter in Kolmogorov’s seminal ‘Italian paper’ [26].

1 An English translation of Kolmogorov’s Italian paper, together with an introduction by M A Stephens, can be found in [27].

The author is supported by a Schrödinger scholarship of the Austrian Research Foundation (FWF).
Let $x_1, \ldots, x_n$ be a finite sequence of $n$ real numbers, sorted in increasing order. Their empirical counting function $C_n(X)$ is defined as the number of elements $x_m$ which are not larger than $X$; that is, we have

$$C_n(X) = \begin{cases} 0 & \text{for } X < x_1, \\ m & \text{for } x_m \leq X < x_{m+1}, \\ n & \text{for } x_n \leq X, \end{cases}$$

for every $X \in \mathbb{R}$. In contrast, the theoretical counting function $C_0(X)$ is given by

$$C_0(X) = n \mathbb{P}(x \leq X),$$

that is by the expected number of values not exceeding $X$ of independent observations of the random variable $x$ (in other words, this is $n$ times the cumulative distribution function of $x$). Let

$$\mathcal{F}_n = \sup_X |C_n(X) - C_0(X)|.$$

Then the stochasticity parameter $\lambda_n$ is defined by

$$\lambda_n = \frac{\mathcal{F}_n}{\sqrt{n}}.$$

Kolmogorov proved, under the assumption that the cumulative distribution function of $x$ is continuous, that $\lambda_n$ has a limiting distribution $\Phi$ as $n \to \infty$, which is given by

$$\Phi(X) = \lim_{n \to \infty} \mathbb{P}(\lambda_n \leq X) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 X^2}, \quad \text{for } X > 0. \quad (2)$$

Note that the distribution $\Phi$, which is now known as the Kolmogorov distribution, is universal—it does not depend on the initial distribution of $x$. Thus a given (large) sample of observations may be accepted as a realization of a sequence of i.i.d. random variables having a specific distribution if and only if the value of its stochasticity parameter $\lambda_n$, calculated with respect to this distribution, is contained in an interval which contains the largest part of the mass of the Kolmogorov distribution. This principle is the basis of the Kolmogorov–Smirnov test in statistics. A possible choice for such an interval may be $[0.4, 1.8]$, since the Kolmogorov distribution assigns a probability of less than one per cent to its complement. However, note that by the Chung–Smirnov law of the iterated logarithm (established by Chung [19] and Smirnov [32]; see also [36, p 505]) for a sequence of independent, identically distributed (i.i.d.) random variables having a continuous distribution we have

$$\limsup_{n \to \infty} \frac{\lambda_n}{\sqrt{\log \log n}} = \frac{1}{\sqrt{2}} \quad \text{almost surely.} \quad (3)$$

In other words, even if the typical value of $\lambda_n$ for fixed $n$ should be somewhere between 0.4 and 1.8, in the long run for an infinite sample of observations we should expect values of $\lambda_n$ as large as roughly $\sqrt{\log \log n}$ to occur from time to time (there even exist precise quantitative results how often such large values should be observed; see [33]).

In the papers [5–15], Arnold collected a large number of empirical observations, rigorous mathematical theorems and open problems concerning the ‘randomness’ of deterministic sequences. The purpose of the present paper is to comment on some of the observations, illustrate the context of these investigations in probability theory and number theory, and to answer some particular problems.
It should be noted that Kolmogorov’s stochasticity parameter is just one out of many possible ways to measure the randomness of a given sequence. Later in his life, Kolmogorov himself established a complexity theory, which can be used to formalize randomness (see [29]). More recently, an effort to measure the pseudorandomness properties of finite sequences was made by Mauduit and Sárközy, who introduced and studied several new measures of pseudorandomness (first for binary sequences, starting with [30], and later for sequences of more symbols [1, 2]). The problem is also discussed in detail in volume 2 of Knuth’s ‘The art of computer programming’ [25].

The outline of the remaining part of this paper is as follows. In the subsequent section, we will introduce the notion of the star-discrepancy, which is a classical concept in number theory. We will show that in the case of \( \lambda_n \) being calculated with respect to the continuous uniform distribution on \([0, 1]\), the star-discrepancy and Kolmogorov’s stochasticity parameter coincide, and that consequently known results from discrepancy theory can be utilized to answer Arnold’s questions. In section 3 we discuss a conjecture of Arnold on the typical value of the stochasticity parameter for sequences of the form \( ax^A \mod N \), where \( A \) is fixed and \( x = 1, 2, \ldots \). Here the word ‘typical’ means that we want to obtain results which hold for almost all parameters \( a \), in the sense of Lebesgue measure. In section 4 we discuss the closely related problem asking for the typical value of the stochasticity parameter of \( a^x \mod N \) where now \( a > 1 \) is fixed, \( x = 1, 2, \ldots \), and \( A \) is taken uniformly from \([0, N]\). In section 5 we discuss the problem of arithmetic progressions with real (not necessarily rational) step size, which is closely connected with the theory of continued fractions. Finally, section 6 contains the proof of a theorem stated in section 4.

2. Uniform distribution modulo 1 and discrepancy theory

Let \( x_1, x_2, \ldots \) be an infinite sequence of real numbers. This sequence is called uniformly distributed modulo one (u.d. mod 1) if for all \( X \in [0, 1] \) the asymptotic relation

\[
\lim_{n \to \infty} \frac{C_n(X)}{n} = X
\]

holds. Here \( C_n \) is the empirical counting function of the sequence of fractional parts of \( x_1, x_2, \ldots \) (therefrom the name ‘uniform distribution modulo one’). In a vague sense a sequence which is u.d. mod 1 can be interpreted as showing ‘random’ behavior, since by the Glivenko-Cantelli theorem a sequence of i.i.d. uniformly \([0, 1]\)-distributed random variables satisfies (4) almost surely. The degree of uniformity of the distribution of a finite point set \( x_1, \ldots, x_n \) can be measured in terms of its star-discrepancy, which is defined as

\[
D^*_n(x_1, \ldots, x_n) = \sup_{X \in [0,1]} \left| \frac{C_n(X)}{n} - X \right|
\]

where again \( C_n \) is the empirical counting function of the fractional parts of \( x_1, x_2, \ldots \). It is easy to see that there is a close connection between the star-discrepancy \( D^*_n \) and the stochasticity parameter \( \lambda_n \) in the case when the sequence is contained in \([0, 1]\) and the underlying distribution in the stochasticity parameter is assumed to be the uniform distribution on \([0, 1]\). More precisely, in this case these two quantities coincide up to normalization, and we have

\[
\sqrt{n} D^*_n(x_1, \ldots, x_n) = \lambda_n.
\]
a subinterval of \([A, B]\), and comparing to the normalized Lebesgue measure). Thus in the case of real sequences from a finite interval, which are compared to the uniform distribution on this interval, results from discrepancy theory can be directly translated into results for the stochasticity parameter. We will use this fact in sections 3–6 below. It should be noted that while the uniform distribution may be the ‘natural’ choice to use for comparison with the empirical distribution of a deterministic set of real numbers, there also exist many classes of sequences of reals whose limit distribution is different from the uniform distribution; many examples can be found in the book of Strauch and Porubský [34]. More background on uniform distribution theory and discrepancy theory can be found in the monographs of Drmota–Tichy [20] and Kuipers–Niederreiter [28].

3. The stochasticity parameter of geometric progressions

In [14, p 36], Arnold formulates the following conjecture:

Conjecture. The Kolmogorov stochasticity parameter \(\lambda_n\) of residues modulo \(N\) of \(n\) terms of a geometric progression with an arbitrary ratio \(a > 1\) does not tend to zero as \(n \to \infty\) (for almost all \(a\), so that exceptional values form a set of Lebesgue measure zero on the real line).

The solution of this conjecture can be deduced from a result which has been recently obtained by the author in [3]; the answer is affirmative. Arnold’s conjecture, asserting that the Kolmogorov parameter \(\lambda_n\) of a ‘typical’ geometric progression is not too small, should be compared to the case of arithmetic progressions, where the Kolmogorov parameter of a typical sequence actually is too small (it tends to 0 as \(n \to \infty\)); see section 5 below. However, in comparison with (3) the assertion that \(\lambda_n\) does not tend to 0 as \(n \to \infty\) is too weak to capture the behavior of the Kolmogorov stochasticity parameter for a typical i.i.d. random sequence. Under the supposition that a typical geometric progression behaves similar to a typical realization of an i.i.d. random sequence, one could actually conjecture that even \(\sqrt{\log \log n} \lambda_n\) does not tend to 0 as \(n \to \infty\) for almost all \(a > 1\). As the following result from [3] shows this stronger statement is also true, and the Kolmogorov stochasticity parameter for typical geometric progressions satisfies the Chung–Smirnov law of the iterated logarithm in exactly the same way as an i.i.d. random sequence.

Theorem A. Let \(A > 0\) and \(N > 0\) be fixed real numbers. Then for the sequence of remainders \(aA, a^2A, a^3A, \ldots\) modulo \(N\) we have

\[
\limsup_{n \to \infty} \frac{\lambda_n}{\sqrt{\log \log n}} = \frac{1}{\sqrt{2}}
\]

for almost all \(a \in \mathbb{R}, a > 1\),

where the Kolmogorov stochasticity parameter \(\lambda_n\) is calculated with respect to the uniform distribution on \([0, N]\).

This theorem is stated in [3] only for the case of the fractional part of a sequence, that is for the case of \(a^iA, a^{i+1}A, \ldots\) being reduced modulo 1. However, it is easily seen that by a simple change of scale the theorem also covers the case of \(a^iA, a^{i+1}A, \ldots\) being reduced modulo \(N\), by means of replacing \(A\) by \(NA\). Thus, the answer to Arnold’s conjecture is affirmative.

4. The stochasticity parameter of lacunary sequences

In [14, p 36], following the conjecture mentioned in the previous section, Arnold formulates the following conjecture:
Moreover, one can conjecture that for almost any base \( a > 1 \) the following more general statement holds: The distribution of the values \( \lambda_n(A) \) of the Kolmogorov stochasticity parameter \( \lambda_n \) of the sequences of \( n \) remainders modulo \( N \) of geometric progressions starting at different points \( A (0 < A < N) \), tend[s], as \( n \to \infty \), to the universal Kolmogorov distribution \( \Phi \) (under the assumption that the starting point \( A \) is uniformly distributed on the interval \( 0 < A < N \)).

Note that this conjecture is much stronger than the conjecture from the previous section, where it was only required that \( \lambda_n \) does not tend to 0 as \( n \to \infty \). However, there is also a difference between the probabilistic model which is used to specify a class of parametric sequences. In the previous section, the sequence \( a^x A, x = 1, 2, \ldots \) was obtained by assuming \( A \) to be fixed and allowing different values for the parameter \( a \). In the present case, \( a \) is fixed and \( A \) is variable.

Thus to solve the problem from the previous section (and in the case of reduction modulo 1) it was, roughly speaking, necessary to show that the functions \( \{aA\}, \{a^2A\}, \{a^3A\}, \ldots \) understood as functions of \( a \), show a behavior which is similar to that of sequences of i.i.d. random variables. In the present case it has to be shown that the same functions, now understood as functions of \( A \), also behave like i.i.d. random variables. These two problems are technically quite different, and require different methods. Generally speaking, the case of lacunary sequences (that is, of assuming that \( A \) is the variable and \( a \) is fixed, as in the present section) is the case which has a longer research history, is better understood, and is easier to handle.

The asymptotic behavior of the Kolmogorov stochasticity parameter (or, in other words: the star-discrepancy) of lacunary sequences is an intensively studied subject. It turns out that precise results depend on fine number-theoretic properties of the growth factor \( a > 1 \) in a very sensitive way. Quite recently, Fukuyama [22] proved the following.

**Theorem B.** The Kolmogorov stochasticity parameter \( \lambda_n \) of the sequence \( aA, a^2A, a^3A, \ldots \) modulo 1 satisfies, for almost all \( A \in [0, 1] \), the asymptotic relation

\[
\limsup_{n \to \infty} \frac{\lambda_n}{\sqrt{\log \log n}} = \begin{cases} 
\frac{\sqrt{84}}{9} & \text{for } a = 2, \\
\frac{(a+1)a(a-2)}{2(a-1)^3} & \text{if } a \geq 4 \text{ is an even integer,} \\
\frac{a+1}{2(a-1)} & \text{if } a \geq 3 \text{ is an odd integer,} \\
\frac{1}{\sqrt{2}} & \text{if } a > 1 \text{ and } a^x \not\in \mathbb{Q} \text{ for } x = 1, 2, \ldots.
\end{cases}
\]

The last case is particularly interesting; it covers the case when \( a \) is a transcendental number. Since almost all numbers are transcendental, this is the typical case with respect to Lebesgue measure, and as in section 3 there is a perfect accordance with the Chung–Smirnov law of the iterated logarithm (3) for i.i.d. random variables. A corresponding limit theorem for the distribution of \( \lambda_n \) has not been proved so far; we state it below as a theorem.

**Theorem 1.** Let \( a > 1 \) be a fixed real number for which \( a^x \not\in \mathbb{Q} \) for \( x = 1, 2, \ldots \), and let \( N > 0 \) also be fixed. Then for the Kolmogorov stochasticity parameter \( \lambda_n \) of the sequence \( aA, a^2A, a^3A, \ldots \mod N \) we have
\[
\lim_{n \to \infty} \mathbb{P}(A \in [0, N] : \lambda_n \leq X) = \Phi(X) \quad \text{for all } X \in \mathbb{R},
\]

where \(\mathbb{P}\) denotes the normalized Lebesgue measure on \([0, N]\), where \(\Phi\) is the distribution function of the Kolmogorov distribution as defined in (2), and where \(\lambda_n\) is calculated with respect to the uniform distribution on \([0, N]\).

Note that, as in Fukuyama’s theorem above, the set of real numbers \(a\) for which \(a^s \notin \mathbb{Q}\) for \(x = 1, 2, \ldots\) has full Lebesgue measure. Thus theorem 1 proves Arnold’s conjecture. The proof of theorem 1 will be given in section 6, at the end of this paper. If the assumption \(a^s \notin \mathbb{Q}\) for \(x = 1, 2, \ldots\) in the statement of theorem 1 is replaced by \(a^s \in \mathbb{Q}\) for some positive integer \(x\), then there still exists a limit distribution of the Kolmogorov stochasticity parameter \(\lambda_n\). However, in this case the limit distribution depends on number-theoretic properties of \(a\) and \(x\) in a very complicated way, and is different from Kolmogorov’s distribution.

### 5. The stochasticity parameter of arithmetic progressions

In [9], Arnold proves several theorems on the stochasticity parameter of arithmetic progressions\(^2\); then, at the end of [9], he writes:

I do not know whether the value of the Kolmogorov stochasticity parameter \(\lambda_n\) of an arithmetic progression of fractional parts of the \(n\) numbers \(kx\) tends to zero for almost all real numbers \(k\), or whether it is just as often unbounded (it might also be ‘generically’ bounded away from 0 and \(\infty\)). The ergodicity of the Gauss–Kuzmin dynamical system \(z \mapsto \{1/z\}\) suggests that any such asymptotic behavior of the stochasticity parameter \(\lambda_n\) should have probability either 0 or 1 (in the space of values of the parameter \(k\)) (provided that it depends only on the asymptotic behavior of the partial quotients \(a_s\) of the continued fraction of \(k\) as \(s \to \infty\)). But I do not know whether the probability is 0 or 1 for the types of behavior described above for the stochasticity parameter.

Arnold’s observation that the problem of the stochasticity parameter (or, in the language of section 2: the star-discrepancy) of a sequence of fractional parts \(\{kx\}, x = 1, 2, \ldots\) is intimately connected with the continued fraction expansion of the step \(k\) is absolutely right. Roughly speaking, the smaller the continued fraction coefficients of \(k\) are, the smaller the discrepancy of \(\{kx\}, x = 1, 2, \ldots\) is. There also exist many precise quantitative results giving discrepancy bounds in terms of the continued fraction coefficients of \(k\); such results are presented in great detail in [28], chapter 2, section 3] and [20, section 1.4.1]. Together with the profound results of Khintchine [23, 24] on the metric theory of continuous fractions one obtains the following result ([20, theorem 1.72]):

Suppose that \(\psi(n)\) is a positive increasing function. Then

\[
\lim_{n \to \infty} nD_n^s((k), \ldots, (kn)) = O((\log n)\psi(\log \log n))
\]

for almost all \(k \in \mathbb{R}\) if and only if

\[
\sum_{n=1}^{\infty} \frac{1}{\psi(n)} < \infty.
\]

\(^2\) As noted in [9], by suitably choosing the scale the general case of arithmetic progressions modulo \(N\) can be reduced to the case of arithmetic progressions modulo 1, that is to the case of fractional parts of arithmetic progressions.
In particular, this implies that for arbitrary \( \varepsilon > 0 \) we have
\[
D^{x}_{\varepsilon}([k], \ldots, (kn)) = O\left(\frac{(\log n)(\log \log n)^{1+\varepsilon}}{n}\right)
\]
as \( n \to \infty \)
for almost all \( k \in \mathbb{R} \). Consequently, by (5), we also have
\[
\lambda_n \to 0 \quad \text{for almost all } k,
\]
which provides the solution of Arnold’s problem.

6. Proof of theorem 1

As noted before, for the proof of theorem 1 we may assume without loss of generality that \( N = 1 \), which means that \( A \) is taken uniformly from \([0, 1]\) and the sequence we consider is the sequence of fractional parts \( \{aA\}, \{a^2A\}, \{a^3A\}, \ldots \).

Our proof of theorem 1 follows the one given in [4] for the case of quickly increasing integer sequences \( a_n, x = 1, 2, \ldots \), and which we adopt to the sequence \( a^x, 1, 2, \ldots \) for real \( a \) instead. The key ingredient is the following result of Fukuyama [21]. It is stated in [21] in a much more general multi-dimensional form, but we only need a special case of the one-dimensional version.

**Lemma 1 ([21, theorem 1]).** Let \( f(y) \) be a measurable function which is of bounded variation on \([0, 1]\) and satisfies
\[
\int_{0}^{1} f(y) \, dy = 0, \quad \int_{0}^{1} f(y) \, dy = 1.
\]
Assume that \( a^x \not\in \mathbb{Q} \) for \( x = 1, 2, \ldots \). Then for all \( X \in \mathbb{R} \) we have
\[
\lim_{n \to \infty} P\left( A \in [0, 1] : \frac{1}{\sqrt{n}} \sum_{x=1}^{n} f(a^x A) \leq X \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{X} e^{-y^2/2} \, dy,
\]
where \( P \) denotes the Lebesgue measure on \([0, 1]\).

**Proof of theorem 1.** Let a number \( a \) satisfying the assumptions of the theorem be given. As noted above, we may assume without loss of generality that \( N = 1 \). Using the notations from the previous sections, we set
\[
G_n(X) = \frac{C_d(X) - C_d(0)}{\sqrt{D}}.
\]
This stochastic process is called the *empirical process*, and we clearly have
\[
\lambda_n = \sup_{X} |G_n(X)|.
\]
In our present setting the empirical process is given by
\[
G_n(t) = \frac{\sum_{x=1}^{n} 1_{[0,t]}(\{a^x A\}) - nt}{\sqrt{n}}, \quad t \in [0, 1].
\]
For each \( n \), the paths of the process \( G_n \) are so-called càdlàg-functions (which means that they are right-continuous and have left limits everywhere), and consequently \( G_n \) is a stochastic process on the Skorokhod space \( D[0, 1] \) (which simply is the space of càdlàg-functions). This space can be equipped with the so-called Skorokhod metric, which is given by

\[
\sigma(f, g) = \inf_{\lambda \in \Lambda} \max\{ \| \lambda - I \|, \| f - g \circ \lambda \| \}.
\]

Here \( \| \cdot \| \) is the sup-norm on \([0, 1]\), the infimum is taken over the class \( \Lambda \) of all strictly increasing continuous bijections from the unit interval to itself, and \( I \) denotes the identity function. Intuitively speaking, two functions \( f, g \) have a small distance \( \sigma(f, g) \) if the sup-norm of \( f - g \) can be made small by ‘wiggling’ the argument of one of the functions a little bit. For more details on these definitions see for example [18]. We want to show that \( (G_n)_{n \geq 1} \) converges weakly to a standard Brownian bridge process \( B(t) \), which is a Gaussian process having (almost surely) continuous paths, mean zero and covariance function

\[
\text{Cov}(B(t_1), B(t_2)) = \mathbb{E}(B(t_1)B(t_2)) = t_1(1 - t_2) \quad \text{for} \quad t_1 < t_2.
\]

Informally speaking, one may imagine a standard Brownian bridge as a standard Brownian motion (also called Wiener process) under the additional assumption that \( B(1) = 0 \). Weak convergence means by definition that we have

\[
\mathbb{E}(h(G_n)) \to \mathbb{E}(h(B))
\]

for all functionals \( h \) on \( D[0, 1] \) which are bounded and continuous with respect to the topology induced by \( \sigma \); our desired result then follows from the continuous mapping theorem and the well-known fact that the Kolmogorov distribution is the distribution of the supremum of the standard Brownian bridge.

To prove weak convergence \( G_n \Rightarrow B \), by [18, theorem 13.1] we have to show that all finite-dimensional distributions of \( G_n \) converge to the corresponding finite-dimensional distributions of \( B \), and that the sequence of distributions of the processes \( G_n(t), n = 1, 2, \ldots \) is tight. Tightness intuitively means that a sequence of measures is not allowed to ‘escape to infinity’; formally a sequence of probability measures is tight if for all \( \varepsilon > 0 \) there exists a compact set \( C \) such that each of the measures in the sequences assigns probability at least \( 1 - \varepsilon \) to \( C \). However, instead of proving tightness directly we will rather use a standard criterion to show that the sequence of distributions induced by \( (G_n)_{n \geq 1} \) is tight (see below).

By the well-known Cramér–Wold device (see for example [16, p 343]), for the convergence of all finite-dimensional distributions of \( G_n \) to those of \( B \) it is sufficient to show that

\[
b_1G_n(t_1) + \cdots + b_mG_n(t_m) \xrightarrow{D} b_1B(t_1) + \cdots + b_mB(t_m)
\]

for all \( m \geq 1 \) and all \( (b_1, \ldots, b_m) \in \mathbb{R}^m \), \( 0 \leq t_1 < \cdots < t_m \leq 1 \). Here \( \xrightarrow{D} \) denotes convergence in distribution. Thus, let \( (b_1, \ldots, b_m) \in \mathbb{R}^m \) and \( 0 \leq t_1 < \cdots < t_m \leq 1 \) be given. For \( t \in [0, 1] \), let \( I_{[0, t]}(y) \) denote the function \( \mathbb{1}_{[0, t]}(y) - t \); in other words, \( I_{[0, t]} \) is the indicator function of \([0, t]\), centered at expectation and extended with period 1. Then we have

\[
G_n(t) = \frac{\sum_{i=1}^{n} I_{[0, t]}(a^nA)}{\sqrt{n}},
\]

and consequently

\[
b_1G_n(t_1) + \cdots + b_mG_n(t_m) = \frac{1}{\sqrt{n}} \sum_{x=1}^{n} \sum_{k=1}^{m} b_k I_{[0, t]}(a^nA).
\]
The function
\[ \sum_{k=1}^{m} b_k \mathbf{1}_{(0,a]}(y) \]
has integral zero (on \([0, 1]\)) and is periodic with period 1. Furthermore, some simple calculations show that we have
\[ \int_{0}^{1} \left( \sum_{k=1}^{m} b_k \mathbf{1}_{(0,a]}(y) \right)^2 dy = \sum_{k=1}^{m} b_k^2 a(1 - t_k) + 2 \sum_{1 \leq k < \ell \leq m} b_k b_{\ell} f_k(1 - t_k). \quad (11) \]
Thus, by lemma 1, the distribution of (10) converges to a normal distribution with mean zero and variance given by the right-hand side of (11). On the other hand, using (7) we can easily show that
\[ \mathbb{E}(b_1B(t_1) + \cdots + b_mB(t_m))^2 = \sum_{k=1}^{m} b_k^2 a(1 - t_k) + 2 \sum_{1 \leq k < \ell \leq m} b_k b_{\ell} f_k(1 - t_k). \]
Thus the distribution of the expression on the right-hand side of (8) is also the normal distribution with mean zero and variance given by the right-hand side of (11). In other words, we have established (8), which proves that the finite-dimensional distributions of \(G_n\) converge to those of \(B\).

To prove that the sequence of probability measures induced by \(G_n(t), \ n = 1, 2, \ldots\) is tight we will use the following criterion, which is stated in [18, theorem 13.2]. A sequence \((P_n)_{n \geq 1}\) of probability measures is tight if and only if

(i) \( \lim_{K \to \infty} \limsup_n P_n(\{ y : \|y\| \geq K \}) = 0 \), and

(ii) for each fixed \( \varepsilon > 0 \) we have \( \lim_{\delta \to 0} \limsup_n P_n(\{ y : w'(\delta) \geq \varepsilon \}) = 0 \).

In the statement of this criterion, \( w' \) is a particular modulus of continuity which is adjusted to the structure of the space \(D[0, 1]\), and which is defined as follows. For a subinterval \(T \subset [0, 1]\) we set
\[ w_T(T) = \sup_{s,t \in T} |y(s) - y(t)|, \]
and we call a set \(0 = t_0 < t_1 < \cdots < t_v = 1\) \(\delta\)-sparse’ if \(\min(t_{i+1} - t_i) > \delta\). Then, for \(\delta \in (0, 1)\), \(w'\) is defined as
\[ w'(\delta) = w'(y, \delta) = \inf \max_{|\{t_i\}| \leq v} \{ w_T(t_i-1, t_i) \}, \]
where the infimum is taken over all \(\delta\)-sparse sets \(\{t_i\}\) (of arbitrary cardinality).

By the corollary stated after theorem 13.2 in [18], instead of (i) it is also sufficient to establish

(i') For each \( t \) in a set \( T \) that is dense in \([0, 1]\) and contains 1 we have
\[ \lim_{K \to \infty} \limsup_n P_n(\{ y(t) : |y(t)| \geq K \}) = 0. \]

3671
Thus to obtain our result it remains to establish that (i’) and (ii) hold in our setting. To establish
(i’), let \( t \in [0, 1] \) be arbitrary, and let \( K > 0 \) be given. We have to show that

\[
\lim_{K \to \infty} \limsup_{n} \mathbb{P} \left( A \in [0, 1] : \frac{\sum_{x=1}^{n} I_{0,x}(a^x)A}{\sqrt{n}} \geq K \right) = 0.
\]

Excluding the trivial cases \( t = 0 \) and \( t = 1 \) we have, by lemma 1 and since \( \int_{0}^{1} I_{0,t}^2 f(y) \, dy = t(1-t) \),

\[
\lim_{n \to \infty} \mathbb{P} \left( A \in [0, 1] : \frac{1}{\sqrt{n}} \sum_{x=1}^{n} I_{[0,t]}(a^x)A \geq K \right) = \frac{1}{\sqrt{2\pi}} \int_{K/(\sqrt{n})-1}^{\infty} e^{-y^2/2} \, dy.
\]

To show (ii), we will use the following lemma. It can be shown in exactly the same way as
the (much more general) Proposition in [31], with the only difference of omitting the factor
\((\log \log N)^{1/2}\) in the definition of \( Q \) on page 247 of [31]. (In [31] the results are stated for the
case of exponentially growing sequences of integers, but, as noted in the proof of lemma (3.4)
of [17], they remain valid in the real case). We write \( \|f\|_{L^2} \) for the \( L^2(0,1) \) norm of \( f \).

**Lemma 2.** Assume that \( \|f\|_{L^2} \geq c_n n^{-1/4} \) for some constant \( c_1 \). Assume that \( f \) satisfies the first
two properties in line (6), and has variation at most 2 on \([0,1]\). Then there exist constants
\( c_2, c_3, n_0 \), all of which are independent of \( f \), such that for \( n \geq n_0 \) we have

\[
\mathbb{P} \left( A \in [0, 1] : \sum_{x=1}^{n} f(a^x)A \geq c_2 \|f\|_{L^2}^{1/4} \sqrt{n} \right) \leq c_3 \exp(-2 \|f\|_{L^2}^{1/2}) + n^{-3/4}.
\]

To establish (ii) we will assume that \( \delta = 2^{-M} \) for some integer \( M \geq 1 \). Let \( \varepsilon > 0 \) be fixed.
Using (9) and the fact that \( I_{[0,t]} - I_{[s,t]} = I_{[s,t]} \) for \( s < t \), it is sufficient to show that we have

\[
\lim_{M \to \infty} \limsup_{n} \mathbb{P} \left( A \in [0, 1] : \max_{t} \sup_{t \in [((i-1)2^{-M}, i2^{-M})]} \left| \sum_{x=1}^{n} I_{[0,t]}(a^x)A \right| \geq \varepsilon \sqrt{n} \right) = 0.
\]

Clearly we have

\[
\mathbb{P} \left( A \in [0, 1] : \max_{t} \sup_{t \in [((i-1)2^{-M}, i2^{-M})]} \left| \sum_{x=1}^{n} I_{[0,t]}(a^x)A \right| \geq \varepsilon \sqrt{n} \right) \leq \sum_{j=1}^{2^M} \mathbb{P} \left( A \in [0, 1] : \sup_{t \in [(i-1)2^{-M}, i2^{-M})]} \left| \sum_{x=1}^{n} I_{[(i-1)2^{-M}, i2^{-M})]}(a^x)A \right| \geq \varepsilon \sqrt{n}/2 \right).
\]

We will only show how to prove an estimate for the probability in (13) in the case \( i = 0 \); the
other cases can be treated in exactly the same way. We will use a method called dyadic chaining,
which is commonly used in probability theory to estimate probabilities concerning the supremum
of a stochastic process (such as, in our case, an empirical process). We choose a number \( N \)
such that \( \varepsilon/\sqrt{4n} \leq 2^{-N} \leq \varepsilon/\sqrt{8n} \). For given \( t \in [0, 2^{-M}) \) let \( r^t \) denote the number whose first
$N$ binary digits (after the decimal point) coincide with those of $t$, but whose other binary digits are all zeros. Furthermore, set $t^* = t + 2^{-N}$. Then for the centered indicator functions we have

$$I_{[0,t^*]}(y) - 2^{-N} \leq I_{[0,t]}(y) \leq I_{[0,t^*]}(y) + 2^{-N}.$$ 

Consequently, by our choice of $N$, we have

$$
\mathbb{P} \left( A \in [0,1] : \sup_{t \in \{0,2^{-N}\}} \left| \sum_{s=1}^{n} I_{[0,t]}(a^sA) \right| \geq \varepsilon \sqrt{n} \right) \\
\leq \mathbb{P} \left( A \in [0,1] : \max_{t=m2^{-N}} n \sum_{s=1}^{n} I_{[0,t]}(a^sA) \geq \varepsilon \sqrt{n} / 4 \right). \tag{14}
$$

We can represent every interval $[0, t]$ for $t$ in the form given in (14) as the disjoint union of

- at most one interval of length $2^{-M}$,
- at most one interval of length $2^{-M-1}$,
- ... 
- at most one interval of length $2^{-N}$.

according to the binary representation of $t$. Furthermore, in order to be able to represent every interval $[0, t]$ in such a way we have to consider at total of

- one interval of length $2^{-M}$,
- two intervals of length $2^{-M-1}$,
- ... 
- $2^N-M$ intervals of length $2^{-N}$.

Let $f$ be the centered indicator function of an interval of length $2^{-k}$ for some $k$ in the range $\{2^{-M}, 2^{-M-1}, \ldots, 2^{-N}\}$. Then $\|f\|_2 = \sqrt{2^{-k}(1 - 2^{-k})} \geq c_1 n^{-1/4}$ for some constant $c_1$ (depending on $\varepsilon$) by our choice of $N$. Using lemma 2 for such a function we have

$$
\mathbb{P} \left( A \in [0,1] : \sum_{s=1}^{n} f(a^sA) \geq c_2(2^{-k}(1 - 2^{-k}))^{1/4} \sqrt{n} \right) \leq c_1 \exp(-2^{k/4} + 1) + n^{-3/4}.
$$

Summing the probabilities of the exceptional sets over all centered indicator functions of the form described above we get a total probability of all exceptional sets of at most

$$
\sum_{\ell=M}^{N} 2^\ell(c_3 \exp(-2^{\ell/4 + 1}) + n^{-3/4}). \tag{15}
$$

On the complement of the union of all these exceptional sets we have

$$
\max_{t=m2^{-N}} n \sum_{s=1}^{n} I_{[0,t]}(a^sA) \leq \sum_{\ell=M}^{N} c_3(2^{-k}(1 - 2^{-k}))^{1/4} \sqrt{n}. \tag{16}
$$

If $M$ is chosen sufficiently large, then the right-hand side of (16) is dominated by $\varepsilon \sqrt{n} / 4$, as desired. On the other hand, by choosing $M$ large the probability in (15), even when multiplied with a factor $2^M$ coming from the $2^M$ possible values of $i$ in (13), can be made arbitrarily small (provided that $n \geq n_0$). This proves (12), which, as noted above, establishes property (ii). Thus
all the properties required for the tightness criterion are given, and, also noted, consequently we have weak convergence $G_n \Rightarrow B$.

It is well known that the functional $f \mapsto \sup_{0 \leq t \leq 1} |f(t)|$ is a continuous functional on $D[0, 1]$. Thus by the continuous mapping theorem (see for example [35, theorem 1.3.6]), and since we have already established $G_n \Rightarrow B$, the distribution of $\sup_{0 \leq t \leq 1} |G_n(t)|$ converges to the distribution of $\sup_{0 \leq t \leq 1} |B(t)|$. However, since $\lambda_n = \sup_{0 \leq t \leq 1} |G_n(t)|$, and since the distribution of the maximum of the standard Brownian bridge is the Kolmogorov distribution, this proves the theorem.

Acknowledgments

I want to thank K Fukuyama for his support during my one-year stay at Kobe University, and for his remarks concerning this manuscript. Many thanks also to the administrative staff at the Department of Mathematics of Kobe University.

References

[1] Ahlswede R, Mauduit C and Sárközy A 2006 Large families of pseudorandom sequences of $k$ symbols and their complexity. I General Theory of Information Transfer and Combinatorics (Lecture Notes in Computer Science vol 4123) (Berlin: Springer) pp 293–307
[2] Ahlswede R, Mauduit C and Sárközy A 2006 Large families of pseudorandom sequences of $k$ symbols and their complexity. II General Theory of Information Transfer and Combinatorics (Lecture Notes in Computer Science vol 4123) (Berlin: Springer) pp 308–25
[3] Aistleitner C 2014 Quantitative uniform distribution results for geometric progressions Isr. J. Math. 204 155–97
[4] Aistleitner C and Berkes I 2013 Limit distributions in metric discrepancy theory Mon. hfte Math. 169 253–65
[5] Arnold V I 2007 Continued fractions of square roots of rational numbers and their statistics Usp. Mat. Nauk. 62 3–14
[6] Arnold V I 2008 Orbits’ statistics in chaotic dynamical systems Nonlinearity 21 T109–12
[7] Arnold V I 2008 Statistics of the period lengths of the continued fractions for the eigenvalues of the integer matrices of order two Funct. Anal. Math. 2 15–26
[8] Arnold V I 2008 Statistics of the periods of continued fractions for quadratic irrationals Izv. Ross. Akad. Nauk, Ser. Mat. 72 3–38
[9] Arnold V I 2008 To what extent are arithmetic progressions of fractional parts random? Usp. Mat. Nauk. 63 5–20
[10] Arnold V I 2009 Empirical study of stochasticity for deterministic chaotic dynamics of geometric progressions of residues Funct. Anal. Math. 2 139–49
[11] Arnold V I 2009 Lengths of periods of continued fractions of square roots of integers Funct. Anal. Math. 2 151–64
[12] Arnold V I 2009 Permutations Usp. Mat. Nauk 64 3–44
[13] Arnold V I 2009 Random and algebraic permutations’ statistics Funct. Anal. Math. 2 247–8
[14] Arnold V I 2009 Stochastic and deterministic characteristics of orbits in chaotically looking dynamical systems Trans. Mosc. Math. Soc. 2009 31–69
[15] Arnold V I 2010 Are quadratic residues random? Regul. Chaotic Dyn. 15 425–30
[16] Ash R B 2000 Probability and Measure Theory 2nd edn (Burlington, MA: Academic)
[17] Berkes I 1976 On the asymptotic behaviour of $S(n)x$ main theorems Z. Wahrscheinlichkeitstheor. Verwandte Geb. 34 319–45
[18] Billingsley P 1999 Convergence of probability measures Wiley Series in Probability and Statistics: Probability and Statistics 2nd edn (New York: Wiley)
[19] Chung K-L 1949 An estimate concerning the kolmogoroff limit distribution Trans. Am. Math. Soc. 67 36–50
[20] Drmota M and Tichy R F 1997 Sequences, Discrepancies and Applications (Lecture Notes in Mathematics vol 1651) (Berlin: Springer)
[21] Fukuyama K 1994 The central limit theorem for Riesz–Raikov sums Probab. Theory Relat. Fields 100 57–75
[22] Fukuyama K 2008 The law of the iterated logarithm for discrepancies of \( \theta^n \) Acta Math. Hung. 118 155–70
[23] Khintchine A 1923 Ein satz über kettenbrüche, mit arithmetischen anwendungen Math. Z. 18 289–306
[24] Khintchine A 1924 Einige sätze über kettenbrüche, mit anwendungen auf die theorie der diophantischen approximationen Math. Ann. 92 115–25
[25] Knuth D E 1998 The Art of Computer Programming vol 2 3rd edn (Reading, MA: Addison-Wesley)
[26] Kolmogorov A 1933 Sulla determinazione empirica di una legge di distribuzione Giorn. Ist. Ital. Attuari 4 83–91
[27] Kotz S and Johnson N L 1992 Breakthroughs in Statistics (Springer Series in Statistics: Perspectives in Statistics vol II) (New York: Springer)
[28] Kuipers L and Niederreiter H 1974 Uniform Distribution of Sequences (New York: Wiley)
[29] Li M and Vitányi P 2008 An introduction to kolmogorov complexity and its applications 3rd edn (Texts in Computer Science) (New York: Springer)
[30] Mauduit C and Sarkozy A 1997 On finite pseudorandom binary sequences. I. Measure of pseudorandomness, the legendre symbol Acta Arith. 82 365–77
[31] Philipp W 1974 Limit theorems for lacunary series and uniform distribution mod 1 Acta Arith. 26 241–51
[32] Smirnov N V 1944 Approximate laws of distribution of random variables from empirical data Usp. Matem. Nauk. 10 179–206
[33] Strassen V 1964 An invariance principle for the law of the iterated logarithm Z. Wahrscheinlichkeitstheor. Verwandte Geb. 3 211–26
[34] Strauch O and Porubský Š 2005 Distribution of Sequences: a Sampler (Schriftenreihe der Slovakischen Akademie der Wissenschaften [Series of the Slovak Academy of Sciences] vol 1) (New York: Peter Lang)
[35] van der Vaart A W and Wellner J A 1996 Weak Convergence and Empirical Processes (Springer Series in Statistics) (New York: Springer)
[36] Shorack G R and Wellner J A 1986 Empirical processes with applications to statistics Probability and Mathematical Statistics (New York: John Wiley & Sons)