AN ELEMENTARY CHARACTERIZATION OF THE GAUSS-KUZMIN MEASURE IN THE THEORY OF CONTINUED FRACTIONS

Shreyas Singh, Zhuo Zhang, AJ Hildebrand
Department of Mathematics, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA
{singh88,zhuoz4,ajh}@illinois.edu

Received: , Revised: , Accepted: , Published:

Abstract

By a classical result of Gauss and Kuzmin, the frequency with which a string $\mathbf{a} = (a_1, \dots, a_n)$ of positive integers appears in the continued fraction expansion of a "random" real number is given by $\mu_{GK}(I(\mathbf{a}))$, where $I(\mathbf{a})$ is the set of real numbers in $[0, 1)$ whose continued fraction expansion begins with the string $\mathbf{a}$ and μ_{GK} is the *Gauss-Kuzmin measure*, defined by $\mu_{GK}(I) = \frac{1}{\log 2}\int_I \frac{1}{1+x}dx$, for any interval $I \subseteq [0, 1]$.

It is known that the Gauss-Kuzmin measure satisfies the symmetry property $(*)$ $\mu_{GK}(I(\mathbf{a})) = \mu_{GK}(I(\overleftarrow{\mathbf{a}}))$, where $\overleftarrow{\mathbf{a}} = (a_n, \dots, a_1)$ is the reverse of the string $\mathbf{a}$. We show that this property in fact characterizes the Gauss-Kuzmin measure: If μ is any probability measure with continuous density function on $[0, 1]$ satisfying $\mu(I(\mathbf{a})) = \mu(I(\overleftarrow{\mathbf{a}}))$ for all finite strings $\mathbf{a}$, then $\mu = \mu_{GK}$.

We also consider the question whether symmetries analogous to $(*)$ hold for permutations of $\mathbf{a}$ other than the reverse $\overleftarrow{\mathbf{a}}$; we call such a symmetry *nontrivial*. We show that strings $\mathbf{a}$ of length 3 have no nontrivial symmetries, while for each $n \geq 4$ there exists an infinite family of strings $\mathbf{a}$ of length n that do have nontrivial symmetries. Finally we present numerical data supporting the conjecture that, in an appropriate asymptotic sense, "almost all" strings $\mathbf{a}$ have no nontrivial symmetries.

1. Introduction

If one picks a "random" real number x and expands it in base 10, then 1/10 of the digits will be 0, 1/10 will be 1, and so on. More generally, any finite string $(d_1, \dots, d_n)$ of digits in $\{0, 1, \dots, 9\}$ occurs in the decimal expansion of the number with frequency $1/10^n$ in the sense that

$$\lim_{N\to\infty} \frac{1}{N}\#\{0 \leq i \leq N-1 : d_{i+1}(x) = d_1, \dots, d_{i+n}(x) = d_n\} = \frac{1}{10^n}, \tag{1.1}$$

where $d_1(x), d_2(x), \dots$ is the sequence of decimal digits of the fractional part of x. A number x with this property is called *normal* with respect to base 10; normality with respect to other bases is defined analogously. By a classical result of Borel [3], almost all real numbers x are normal with respect to all integer bases $n \geq 2$.

In this paper we consider analogous questions for continued fraction expansions of numbers, that is, expansions of the form

$$x = a_0(x) + \cfrac{1}{a_1(x) + \cfrac{1}{a_2(x) + \cfrac{1}{\ddots}}} = [a_0(x); a_1(x), a_2(x), \dots], \tag{1.2}$$

where $a_0(x) = \lfloor x \rfloor$ and $a_i(x)$, $i = 1, 2, \dots$, are positive integers, which we call the *continued fraction digits*[1] of x. It is well-known (see Section 2 for further details and references) that any irrational number x has a unique *infinite* continued fraction expansion of the form (1.2), and that, conversely, for any integer a_0 and any infinite sequence $a_1, a_2, \dots$ of positive integers there exists a unique irrational number x whose continued fraction digits are given by this sequence. Thus, continued fraction expansions are analogous to decimal and base b expansions in that they provide a way to "encode" real numbers in terms of sequences of integers.

The continued fraction analogs of Borel's results on the frequencies of digits and strings of digits in base b expansions of real numbers are the following theorems, which have their origins in work of Gauss and which are key results in the metric theory of continued fractions developed by Kuzmin [7].

Gauss-Kuzmin Theorem ([1, (3.25)], [5, Proposition 4.1.1])**.** Almost all real numbers x satisfy

$$\lim_{N\to\infty} \frac{1}{N} \#\{1 \leq i \leq N : a_i(x) = a\} = \log_2 \left(1 + \frac{1}{a(a+2)}\right) \tag{1.3}$$

for every positive integer a, where $\log_2 x = (\log x)/(\log 2)$ denotes the base 2 logarithm.

Thus, for almost all x, the continued fraction expansion of x contains the digit 1 with frequency $\log_2(1 + 1/3) = 0.415037\dots$, the digit 2 with frequency $\log_2(1 + 1/8) = 0.169925\dots$, and so on. The numbers $P_{GK}(a)$ defined by

$$P_{GK}(a) = \log_2 \left(1 + \frac{1}{a(a+2)}\right) \quad (a \in \mathbb{N}) \tag{1.4}$$

form a discrete probability distribution on $\mathbb{N}$, called the *Gauss-Kuzmin distribution.*

[1] In the literature, the numbers $a_i(x)$ are usually called *partial quotients*. We use the term *digits* here to emphasize the analogy to digits in ordinary decimal and base b expansions.

More generally, given a finite string $\mathbf{a} = (a_1, \ldots, a_n)$ of positive integers, set[2]

$$I(\mathbf{a}) = \{x \in [0,1) : a_i(x) = a_i \quad (i = 1, \ldots, n)\}. \tag{1.5}$$

Thus, $I(\mathbf{a})$ is the set of real numbers in $[0,1)$ whose continued fraction expansion (ignoring the leading term $a_0(x) = 0$) *begins* with the string $\mathbf{a}$. Let μ_{GK} be the *Gauss-Kuzmin measure* on the interval $[0,1]$ defined by

$$\mu_{GK}(I) = \frac{1}{\log 2} \int_I \frac{1}{1+x}\,dx, \tag{1.6}$$

for any interval $I \subset [0,1]$. Finally, set

$$P_{GK}(\mathbf{a}) = \mu_{GK}(I(\mathbf{a})) = \frac{1}{\log 2} \int_{I(\mathbf{a})} \frac{1}{1+x}\,dx. \tag{1.7}$$

With these notations we have:

Generalized Gauss-Kuzmin Theorem ([5, Proposition 4.1.2])**.** Almost all real numbers x satisfy

$$\lim_{N\to\infty} \frac{1}{N} \#\{0 \le i \le N-1 : a_{i+1}(x) = a_1, \ldots, a_{i+n}(x) = a_n\} = P_{GK}(\mathbf{a}) \tag{1.8}$$

for every finite string $\mathbf{a} = (a_1, \ldots, a_n)$ of positive integers.

Thus, $P_{GK}(\mathbf{a})$ is the frequency with which a "random" real number contains the string $\mathbf{a}$ in its continued fraction expansion. When restricted to strings $\mathbf{a}$ of a *fixed* length n, the frequencies $P_{GK}(\mathbf{a})$ form a discrete probability measure on the set $\mathbb{N}^n$.

As an illustration of this result, consider the string $\mathbf{a} = (3,1,4)$. In this case the set $I(\mathbf{a})$ is an interval with endpoints

$$[0;3,1,4] = \cfrac{1}{3 + \cfrac{1}{1 + \cfrac{1}{4}}} = \frac{5}{19}, \quad [0;3,1,5] = \cfrac{1}{3 + \cfrac{1}{1 + \cfrac{1}{5}}} = \frac{6}{23}. \tag{1.9}$$

Thus, by the generalized Gauss-Kuzmin Theorem the string $(3,1,4)$ occurs in the continued fraction expansion of a "random" number x with frequency

$$\begin{aligned} P_{GK}((3,1,4)) &= \mu_{GK}\left(\left(\frac{6}{23}, \frac{5}{19}\right]\right) = \frac{1}{\log 2} \int_{6/23}^{5/19} \frac{1}{1+x}\,dx \\ &= \log_2\left(\frac{1+\frac{5}{19}}{1+\frac{6}{23}}\right) = -\log_2\left(1 - \frac{1}{551}\right) = 0.002620\ldots \end{aligned} \tag{1.10}$$

[2]There is a slight ambiguity in this definition due to the ambiguity (see (2.2) below) in the continued fraction representation of a *rational* number. This ambiguity does not affect the results stated here since rational numbers represent a set of Lebesgue measure 0 and we are only concerned with integrals over the sets $I(\mathbf{a})$, but it could be resolved by requiring the last digit in the continued fraction representation of a rational number to be strictly greater than 1.

(Recall that $\log_2(x) = (\log x)/\log 2)$ denotes the logarithm of x with respect to base 2.) For comparison, by (1.1), the frequency with which this string occurs in the *decimal* expansion of a random number is $1/10^3$.

For single digit strings $\mathbf{a} = (a)$, the set $I(\mathbf{a})$ reduces to the interval $(1/(a+1), 1/a]$ and the frequency $P_{GK}(\mathbf{a})$ becomes

$$\begin{aligned} P_{GK}(\mathbf{a}) &= \mu_{GK}\left(\left(\frac{1}{a+1}, \frac{1}{a}\right]\right) = \frac{1}{\log 2}\int_{1/(a+1)}^{1/a} \frac{1}{1+x}dx \\ &= \log_2\left(1 + \frac{1}{a(a+2)}\right), \end{aligned} \tag{1.11}$$

which is the Gauss-Kuzmin distribution $P_{GK}(a)$ defined in (1.4).

Although continued fraction expansions share many properties with ordinary decimal and base b expansions, there are three key differences. The most obvious difference is that, while the frequencies of base b digits are uniformly distributed on the finite set $\{0, 1, \ldots, b-1\}$, continued fraction digits can take any positive integer value, and their frequencies, given by (1.4), are non-uniform.

A second difference is that consecutive continued fraction digits are not independent; that is, in general we have $P_{GK}((a_1, \ldots, a_n)) \neq P_{GK}(a_1) \ldots P_{GK}(a_n)$. For example, by (1.10) the string $(3, 1, 4)$ occurs with frequency $P_{GK}((3, 1, 4)) = -\log_2(1 - 1/551) = 0.002620\ldots$, while from (1.4) we get

$$\begin{aligned} P_{GK}(3)P_{GK}(1)P_K(4) &= \log_2\left(1 + \frac{1}{3\cdot 5}\right)\log_2\left(1 + \frac{1}{1\cdot 3}\right)\log_2\left(1 + \frac{1}{4\cdot 6}\right) \\ &= 0.002275\ldots \end{aligned}$$

A third difference is that frequencies of strings of continued fraction digits depend on the order in which these digits occur in the string; that is, different permutations of the same string *in general* have different frequencies. For example, by (1.10) the string $(3, 1, 4)$ occurs with frequency $-\log_2(1 - 1/551)$, while an analogous calculation shows that the string $(3, 4, 1)$ occurs with the slightly smaller frequency $-\log_2(1 - 1/608)$. It is this dependency of the frequency of a string of digits in continued fraction expansions on the order of the digits in the string that we will focus on in this paper.

There is a notable exception to the dependency on the order of the digits: the reverse of a string *always* has the same frequency as the string itself. Specifically, given a finite string $\mathbf{a} = (a_1, \ldots, a_n)$, let $\overleftarrow{\mathbf{a}}$ denote the string obtained by reversing the order of the digits a_i, i,e.,

$$\overleftarrow{\mathbf{a}} = (a_n, \ldots, a_1). \tag{1.12}$$

Then the following result holds:

Symmetry Property. For all finite strings $\mathbf{a} = (a_1, \ldots, a_n)$ of positive integers we have

$$\mu_{GK}(I(\mathbf{a})) = \mu_{GK}(I(\overleftarrow{\mathbf{a}})) \tag{1.13}$$

and thus

$$P_{GK}(\mathbf{a}) = P_{GK}(\overleftarrow{\mathbf{a}}). \tag{1.14}$$

This surprising property of the Gauss-Kuzmin measure is an elementary consequence of the definitions (1.6) and (1.5) of the Gauss-Kuzmin measure and the sets $I(\mathbf{a})$. The property has been observed in the literature (see, e.g., [8, p. 189] and [11, p. 430]), though does not seem to be widely known; we will provide a proof in Section 2 (see Lemma 2.5). We remark that $I(\mathbf{a})$ and $I(\overleftarrow{\mathbf{a}})$ in general represent intervals with different endpoints and of different lengths, so there is no *obvious* reason why an identity such as (1.13) should hold.

Table 1 illustrates the dependence of the frequencies $P_{GK}(\mathbf{a})$ on the permutations of the string $\mathbf{a}$ as well as the symmetry property (1.14). The table shows the frequencies of all six permutations of the string $(3, 1, 4)$. As predicted by (1.14) the reverse of the string $(3, 1, 4)$, i.e., the permutation $(4, 1, 3)$, has the same frequency as the string itself, and the same holds for the pairs $\{(1, 3, 4), (4, 3, 1)\}$ and $\{(3, 4, 1), (1, 4, 3)\}$.

String $\mathbf{a}$	Interval $I(\mathbf{a})$	Frequency $P_{GK}(\mathbf{a})$
$(1, 3, 4)$	$\left(\frac{16}{21}, \frac{13}{17}\right]$	$-\log_2\left(1 - \frac{1}{629}\right)$
$(1, 4, 3)$	$\left(\frac{17}{21}, \frac{13}{16}\right]$	$-\log_2\left(1 - \frac{1}{608}\right)$
$(3, 1, 4)$	$\left(\frac{6}{23}, \frac{5}{19}\right]$	$-\log_2\left(1 - \frac{1}{551}\right)$
$(3, 4, 1)$	$\left(\frac{9}{29}, \frac{5}{16}\right]$	$-\log_2\left(1 - \frac{1}{608}\right)$
$(4, 1, 3)$	$\left(\frac{5}{24}, \frac{4}{19}\right]$	$-\log_2\left(1 - \frac{1}{551}\right)$
$(4, 3, 1)$	$\left(\frac{7}{30}, \frac{4}{17}\right]$	$-\log_2\left(1 - \frac{1}{629}\right)$

Table 1: The frequencies $\mu_{GK}(I(\mathbf{a}))$ of the 6 permutations $\mathbf{a}$ of the string $(3, 1, 4)$.

In our first result we show that the symmetry property (1.13) in fact characterizes the Gauss-Kuzmin measure μ_{GK}:

Theorem 1 (Characterization of the Gauss-Kuzmin measure)**.** *Let $S \subset \mathbb{N}$ be an infinite set of positive integers. Let μ be a probability measure on $[0,1]$ with continuous density function satisfying*

$$\mu(I(\mathbf{a})) = \mu(I(\overleftarrow{\mathbf{a}})) \tag{1.15}$$

for all strings $\mathbf{a} = (a_1, \ldots, a_n)$ of positive integers of length $n \in S$. Then $\mu = \mu_{GK}$, i.e., μ is the measure with density function $f(x) = (1/\log 2)(1+x)^{-1}$.

This characterization is best-possible[3] in the sense that if we impose the symmetry property (1.15) only for strings $\mathbf{a}$ of length from a *finite* set S, the conclusion $\mu = \mu_{GK}$ need not hold:

Theorem 2 (Optimality of the characterization)**.** *Let $N \in \mathbb{N}$ be given. Then there exists a probability measure μ on $[0,1]$ with continuous density function that satisfies (1.15) for all strings $\mathbf{a} = (a_1, \ldots, a_n)$ of positive integers of length $n \leq N$, but $\mu \neq \mu_{GK}$.*

In the example shown in Table 1, it is the case that two different permutations of $(3,1,4)$ have the same frequency *if and only if* one is the reverse of the other. We next explore the question whether this holds for more general strings.

Given a permutation σ of the indices $1, 2, \ldots, n$ and a string $\mathbf{a} = (a_1, \ldots, a_n)$, let $\sigma(\mathbf{a})$ denote the permutation of $\mathbf{a} = (a_1, \ldots, a_n)$ induced by σ, i.e.,

$$\sigma(\mathbf{a}) = (a_{\sigma(1)}, \ldots, a_{\sigma(n)}). \tag{1.16}$$

Definition 1.1 (Non-trivial symmetries)**.** Let $\mathbf{a}$ be a string of positive integers of length n. We say that the string $\mathbf{a}$ has a *nontrivial symmetry* if there exists a permutation σ of $1, \ldots, n$ with $\sigma(\mathbf{a}) \neq \mathbf{a}$ and $\sigma(\mathbf{a}) \neq \overleftarrow{\mathbf{a}}$ such that

$$P_{GK}(\sigma(\mathbf{a})) = P_{GK}(\mathbf{a}). \tag{1.17}$$

From Table 1 we see that the string $(3,1,4)$ has *no* nontrivial symmetries. This raises the question of whether the same holds for more general strings. For strings of length 2, this is trivially the case as the only permutations of such a string are the string itself and its reverse. The question becomes nontrivial for strings of length 3 and larger. We prove the following result:

Theorem 3 (Strings with nontrivial symmetries)**.**

(i) There exists no string $\mathbf{a}$ of length 3 with a nontrivial symmetry.

(ii) For each $n \geq 4$ there exists an infinite, $\lfloor (n-2)/2 \rfloor$-parameter, family of strings $\mathbf{a}$ of length n that have a nontrivial symmetry.

[3]But see Remark 3.2 for comments on possible refinements of the statement of the theorem.

In Section 6 we present numerical data suggesting that strings with nontrivial symmetries are quite rare in the following sense.

Conjecture 4 (Strings with nontrivial symmetries)**.** Let $n \geq 4$ be given. Then

$$\lim_{N\to\infty} \frac{1}{N^n} \#\left\{\mathbf{a} \in \{1,\ldots,N\}^n : \mathbf{a} \text{ has a nontrivial symmetry}\right\} = 0. \tag{1.18}$$

This may be interpreted as saying that almost all strings $\mathbf{a}$ have no nontrivial symmetries.

2. Background on continued fractions

2.1. Continued fraction basics

We begin by recalling some key definitions and facts from the elementary theory of continued fractions. Details and proofs can be found, for example, in [1, Chapter 2], [4, Chapter 9], [6, Chapters I–II], and [9, Chapter 5].

A *continued fraction* is a finite or infinite expression of the form

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}}} = [a_0; a_1, a_2, \ldots], \tag{2.1}$$

where a_0 is an arbitrary integer, and $a_1, a_2, \ldots$ are positive integers.

Clearly, any *finite* (i.e., terminating) continued fraction represents a rational number. Conversely, any rational number can be represented as a finite continued fraction $[a_0; a_1, \ldots, a_n]$. There is a slight ambiguity in this representation due to the identity

$$[a_0; a_1, \ldots, a_n] = [a_0; a_1, \ldots, a_n - 1, 1] \quad \text{if } a_n > 1. \tag{2.2}$$

This ambiguity can be eliminated by requiring the last digit in the representation to be strictly greater than 1.

An *infinite* continued fraction is defined as the limit, as $n \to \infty$, of the finite continued fractions obtained by truncating the given infinite continued fraction after n terms:

$$[a_0; a_1, a_2, \ldots] = \lim_{n\to\infty} [a_0; a_1, a_2, \ldots, a_n]. \tag{2.3}$$

It is known (see, e.g., [9, Theorem 5.11]) that any *irrational* real number has a unique infinite continued fraction expansion; that is, there exists a unique integer a_0 and and unique positive integers $a_1, a_2, \ldots$ such that

$$x = \lim_{n\to\infty} [a_0; a_1, a_2, \ldots, a_n]. \tag{2.4}$$

Conversely, given any integer a_0 and any sequence $a_1, a_2, \ldots$ of positive integers, the limit (2.3) exists and represents an irrational number.

The *convergents* of a (finite or infinite) continued fraction $[a_0; a_1, a_2, \ldots]$ are the rational numbers obtained by truncating the continued fraction after finitely many terms. The nth convergent is the continued fraction truncated at a_n and is denoted by p_n/q_n; that is, p_n and q_n are integers satisfying

$$[a_0; a_1, \ldots, a_n] = \frac{p_n}{q_n}, \tag{2.5}$$

with the convention that

$$\begin{cases} p_0 = a_0, \ q_0 = 1, \\ p_n \in \mathbb{Z}, \ q_n \in \mathbb{N}, \quad (p_n, q_n) = 1 \quad (n \geq 1). \end{cases} \tag{2.6}$$

The numbers p_n and q_n can be computed recursively in terms of the digits a_i, or equivalently through the matrix identity (see, for example, [2, Section 1.3.2] or—in a slightly different, but equivalent form—[10, §5])

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix} = \begin{pmatrix} p_{n-1} & p_n \\ q_{n-1} & q_n \end{pmatrix} \quad (n \geq 1). \tag{2.7}$$

Taking the determinant on both sides of (2.7) yields the identity

$$p_n q_{n-1} - q_n p_{n-1} = (-1)^{n-1} \quad (n \geq 1). \tag{2.8}$$

Notational conventions. In the remainder of this paper we will focus on continued fractions with leading term $a_0 = 0$, i.e., continued fractions representing numbers in the unit interval $[0; 1)$. In this case, we suppress the term $a_0 (= 0)$ in the continued fraction notation (2.1) and thus write

$$[a_1, a_2, \ldots] = [0; a_1, a_2, \ldots] = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}}}. \tag{2.9}$$

Note that when $a_0 = 0$, we have, by (2.6),

$$p_0 = 0, \quad q_0 = 1. \tag{2.10}$$

Given a finite string $\mathbf{a} = (a_1, \ldots, a_n)$ of positive integers, we let

$$[\mathbf{a}] = [a_1, \ldots, a_n] \, (= [0; a_1, \ldots, a_n]) \tag{2.11}$$

denote the continued fraction with digits given by this string. Given two finite strings $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_m)$ of positive integers, we denote by

$(\mathbf{a},\mathbf{b}) = (a_1,\dots,a_n,b_1,\dots,b_m)$ the concatenation of these strings and by $[\mathbf{a},\mathbf{b}] = [a_1,\dots,a_n,b_1,\dots,b_m]$ the corresponding continued fraction. In the case when the string $\mathbf{b} = (b)$ consists of a single digit b, we suppress the parentheses around b and write $(\mathbf{a}, b)$ and $[\mathbf{a}, b]$ instead of $(\mathbf{a},(b))$ and $[\mathbf{a},(b)]$, respectively; $(b,\mathbf{a})$ and $[b,\mathbf{a}]$ are to be interpreted analogously.

2.2. Convergent matrices

It will be convenient to encode the last two convergents of a finite continued fraction as a 2×2 matrix, defined as follows.

Definition 2.1 (Convergent matrix)**.** Given a finite string $\mathbf{a} = (a_1,\dots,a_n)$ of positive integers, the *convergent matrix* associated with this string (or with the continued fraction, $[\mathbf{a}]$, defined by this string) is the matrix $C(\mathbf{a})$ defined by

$$C(\mathbf{a}) = \begin{pmatrix} p_{n-1} & p_n \\ q_{n-1} & q_n \end{pmatrix}, \tag{2.12}$$

where p_{n-1}/q_{n-1} and p_n/q_n denote the $(n-1)$th and nth convergents of the continued fraction $[\mathbf{a}] = [a_1,\dots,a_n]$.

Note that the matrix $C(\mathbf{a})$ defined in (2.12) is exactly the matrix appearing on the right side of the identity (2.7). Since, by our assumption $a_0 = 0$,

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & a_0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

the first two matrices on the left of (2.7) are inverses of each other, so the identity simplifies to

$$C(\mathbf{a}) = \begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix}\cdots\begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix}. \tag{2.13}$$

We next derive explicit formulas for the convergent matrix $C(\overleftarrow{\mathbf{a}})$ associated with the reverse $\overleftarrow{\mathbf{a}}$ of a string $\mathbf{a}$, and the convergent matrices $C(\mathbf{a}, t)$ and $C(t,\mathbf{a})$ associated with the strings $(\mathbf{a}, t)$ and $(t, \mathbf{a})$, obtained by appending or prepending a single digit t to the string $\mathbf{a}$.

Lemma 2.2 (Explicit formulas for convergent matrices)**.** *Let* $\mathbf{a} = (a_1,\dots,a_n)$ *be a string of positive integers, and denote the convergent matrix of* $\mathbf{a}$ *by*

$$C(\mathbf{a}) = \begin{pmatrix} p' & p \\ q' & q \end{pmatrix} \tag{2.14}$$

(so that $(p, q) = (p_n, q_n)$ *and* $(p', q') = (p_{n-1}, q_{n-1})$*). Then we have:*

(i) $C(\overleftarrow{\mathbf{a}}) = \begin{pmatrix} p' & q' \\ p & q \end{pmatrix}$.

(ii) $C(\mathbf{a},t) = \begin{pmatrix} p & p'+tp \\ q & q'+tq \end{pmatrix} \quad (t \in \mathbb{N})$.

(iii) $C(t,\mathbf{a}) = \begin{pmatrix} q' & q \\ p'+tq' & p+tq \end{pmatrix} \quad (t \in \mathbb{N})$.

(iv) $C(t,\overleftarrow{\mathbf{a}}) = \begin{pmatrix} p & q \\ p'+tp & q'+tq \end{pmatrix} \quad (t \in \mathbb{N})$.

Proof. (i). Taking the transpose on both sides of the identity (2.13) and noting that the matrices $\begin{pmatrix} 0 & 1 \\ 1 & a_i \end{pmatrix}$ are symmetric, we obtain

$$\begin{aligned} C(\mathbf{a})^T &= \left(\begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix} \right)^T = \begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix}^T \cdots \begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix}^T \\ &= \begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix} = C(\overleftarrow{\mathbf{a}}). \end{aligned}$$

Since

$$C(\mathbf{a})^T = \begin{pmatrix} p' & p \\ q' & q \end{pmatrix}^T = \begin{pmatrix} p' & q' \\ p & q \end{pmatrix},$$

this proves (i).

To prove (ii)–(iv), let t be a positive integer and apply the identity (2.13) to the strings $(\mathbf{a},t) = (a_1,\ldots,a_n,t)$, $(t,\mathbf{a}) = (t,a_1,\ldots,a_n)$, and $(t,\overleftarrow{\mathbf{a}}) = (t,a_n,\ldots,a_1)$. We obtain

$$\begin{aligned} C(\mathbf{a},t) &= \begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & t \end{pmatrix} = C(\mathbf{a}) \begin{pmatrix} 0 & 1 \\ 1 & t \end{pmatrix} \\ &= \begin{pmatrix} p' & p \\ q' & q \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & t \end{pmatrix} = \begin{pmatrix} p & p'+tp \\ q & q'+tq \end{pmatrix}, \\ C(t,\mathbf{a}) &= \begin{pmatrix} 0 & 1 \\ 1 & t \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & t \end{pmatrix} C(\mathbf{a}) \\ &= \begin{pmatrix} 0 & 1 \\ 1 & t \end{pmatrix} \begin{pmatrix} p' & p \\ q' & q \end{pmatrix} = \begin{pmatrix} q' & q \\ p'+tq' & p+tq \end{pmatrix}, \\ C(t,\overleftarrow{\mathbf{a}}) &= \begin{pmatrix} 0 & 1 \\ 1 & t \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & t \end{pmatrix} C(\overleftarrow{\mathbf{a}}) \\ &= \begin{pmatrix} 0 & 1 \\ 1 & t \end{pmatrix} \begin{pmatrix} p' & q' \\ p & q \end{pmatrix} = \begin{pmatrix} p & q \\ p'+tp & q'+tq \end{pmatrix}, \end{aligned}$$

as claimed. □

Remark 2.3. The identity in part (i) of the lemma relates the last two convergents of a finite continued fraction corresponding to a string **a** to the last two convergents of the continued fraction corresponding to the reverse string $\overleftarrow{\mathbf{a}}$. Identities of this type have long been known in the literature (see, e.g., [10, §11], [8, p. 189], and [6, Theorem 6]). It is this identity that lies at the root of the symmetry property (1.13). By the same token, the (apparent) lack of analogous identities for permutations other than the reversal may explain why for "most" strings $\mathbf{a}$, the reversal $\overleftarrow{\mathbf{a}}$ seems to be the only permutation of $\mathbf{a}$ under which the Gauss-Kuzmin measure is invariant.

2.3. Fundamental intervals

Let $\mathbf{a} = (a_1, \ldots, a_n)$ be a string of positive integers, and let $C(\mathbf{a}) = \left(\begin{smallmatrix} p' & p \\ q' & q \end{smallmatrix}\right)$ be the associated convergent matrix (cf. (2.14)). We begin by obtaining an explicit formula for the set $I(\mathbf{a})$ defined in (1.5) in terms of the entries p', p, q', q of this matrix. By definition $I(\mathbf{a})$ is the set of real numbers $x \in [0, 1)$ whose continued fraction expansion begins with the digits of $\mathbf{a}$, i.e., the set of real numbers of the form $x = [a_1, \ldots, a_n]$ or $x = [a_1, \ldots, a_n, *]$, with the asterisk denoting one or more positive integers. It follows easily from the definitions (2.1) and (2.3) of finite and infinite continued fractions that this set is an interval with endpoints

$$[a_1, \ldots, a_n] = \frac{p_n}{q_n} = \frac{p}{q}$$

and

$$[a_1, \ldots, a_n + 1] = [a_1, \ldots, a_n, 1] = \frac{p' + p}{q' + q},$$

where the first of these two endpoints is included, and the second excluded. We call this interval a *fundamental interval*. Using the notation

$$\langle \alpha, \beta \rangle = \begin{cases} [\alpha, \beta) & \text{if } \alpha < \beta, \\ (\beta, \alpha] & \text{if } \alpha > \beta, \end{cases} \tag{2.15}$$

we thus have (cf. [1, (3.6)] or [2, Ex. 1.3.15])

$$I(\mathbf{a}) = \left\langle \frac{p}{q}, \frac{p' + p}{q' + q} \right\rangle. \tag{2.16}$$

Since

$$\frac{p' + p}{q' + q} - \frac{p}{q} = \frac{(p' + p)q - p(q' + q)}{q(q' + q)} = \frac{p'q - pq'}{q(q' + q)} = \frac{(-1)^n}{q(q' + q)},$$

where the last step follows from (2.8), (2.16) can be written as

$$I(\mathbf{a}) = \left\langle \frac{p}{q}, \frac{p}{q} + \frac{(-1)^n}{q(q' + q)} \right\rangle. \tag{2.17}$$

Next, we derive explicit formulas for the fundamental intervals associated with strings of the form $\overleftarrow{\mathbf{a}}$, $(\mathbf{a}, t)$, $(t, \mathbf{a})$, and $(t, \overleftarrow{\mathbf{a}})$.

Lemma 2.4 (Explicit formulas for fundamental intervals)**.** *Let* $\mathbf{a} = (a_1, \dots, a_n)$ *be a string of* n *positive integers with convergent matrix* (2.14)*. Then we have:*

(i) $I(\overleftarrow{\mathbf{a}}) = \left\langle \frac{q'}{q}, \frac{q'}{q} + \frac{(-1)^n}{q(p+q)} \right\rangle.$

(ii) $I(\mathbf{a}, t) = \left\langle \frac{p'+tp}{q'+tq}, \frac{p'+tp}{q'+tq} + \frac{(-1)^{n+1}}{(q'+tq)(q'+(t+1)q)} \right\rangle \quad (t \in \mathbb{N}).$

(iii) $I(t, \mathbf{a}) = \left\langle \frac{q}{p+tq}, \frac{q}{p+tq} + \frac{(-1)^{n+1}}{(p+tq)(p'+p+t(q'+q))} \right\rangle \quad (t \in \mathbb{N}).$

(iv) $I(t, \overleftarrow{\mathbf{a}}) = \left\langle \frac{q}{q'+tq}, \frac{q}{q'+tq} + \frac{(-1)^{n+1}}{(q'+tq)(p'+q'+t(p+q))} \right\rangle \quad (t \in \mathbb{N}).$

Proof. By part (i) of Lemma 2.2, the convergent matrix $C(\overleftarrow{\mathbf{a}})$ associated with the reverse string $\overleftarrow{\mathbf{a}}$ is obtained from the convergent matrix $C(\mathbf{a}) = \left(\begin{smallmatrix} p' & p \\ q' & q \end{smallmatrix}\right)$ by interchanging p and q', i.e., by making the substitutions

$$p \to q', \quad q' \to p. \tag{2.18}$$

Performing the same substitutions in the formula (2.17) for the interval $I(\mathbf{a})$ then yields the formula for $I(\overleftarrow{\mathbf{a}})$ asserted in part (i) of the lemma.

The formulas for $I(\mathbf{a}, t)$, $I(t, \mathbf{a})$, and $I(t, \overleftarrow{\mathbf{a}})$ in parts (ii)–(iv) of the lemma can be obtained similarly using the formulas for $C(\mathbf{a}, t)$, $C(t, \mathbf{a})$, and $C(t, \overleftarrow{\mathbf{a}})$ from Lemma 2.2 and making the substitutions

$$p' \to p, \quad q' \to q, \quad p \to p' + tp, \quad q \to q' + tq, \tag{2.19}$$

$$p' \to q', \quad q' \to p' + tq', \quad p \to q, \quad q \to p + tq, \tag{2.20}$$

$$p' \to p, \quad q' \to p' + tp, \quad p \to q, \quad q \to q' + tq, \tag{2.21}$$

and also replacing n by $n+1$ to account for the additional digit t in the strings $(\mathbf{a}, t)$, $(t, \mathbf{a})$, and $(t, \overleftarrow{\mathbf{a}})$. □

2.4. The Gauss-Kuzmin measure

Using the formula (2.17) for $I(\mathbf{a})$ we can compute the Gauss-Kuzmin probabilities (see (1.6) and (1.7))

$$P_{GK}(\mathbf{a}) = \mu_{GK}(I(\mathbf{a})) = \frac{1}{\log 2} \int_{I(\mathbf{a})} \frac{1}{1+x}\, dx \tag{2.22}$$

in terms of the entries p, q, p', q' of the convergent matrix associated with the string $\mathbf{a}$.

Lemma 2.5 (Explicit formula for the Gauss-Kuzmin measure)**.** *Let* $\mathbf{a} = (a_1, \ldots, a_n)$ *be a string of positive integers, with convergent matrix given by* (2.14)*. Then we have:*

$$P_{GK}(\mathbf{a}) = P_{GK}(\overleftarrow{\mathbf{a}}) = \left|\log_2\left(1 + \frac{(-1)^n}{(p+q)(q'+q)}\right)\right|. \tag{2.23}$$

In particular, the Gauss-Kuzmin measure satisfies the symmetry property (1.14)*.*

Proof. By (2.17), $I(\mathbf{a})$ is an interval with endpoints

$$\alpha = \frac{p}{q}, \quad \beta = \frac{p}{q} + \frac{(-1)^n}{q(q'+q)}. \tag{2.24}$$

Hence

$$\begin{aligned} P_{GK}(\mathbf{a}) &= \mu_{GK}(I(\mathbf{a})) = \mu_{GK}(\langle \alpha, \beta \rangle) \\ &= \frac{1}{\log 2}\left|\int_\alpha^\beta \frac{1}{1+x}\,dx\right| = \left|\log_2 \frac{1+\beta}{1+\alpha}\right|. \end{aligned} \tag{2.25}$$

Using (2.24) we get

$$\frac{1+\beta}{1+\alpha} = \frac{1 + \frac{p}{q} + \frac{(-1)^n}{q(q'+q)}}{1 + \frac{p}{q}} = \frac{(p+q)(q'+q) + (-1)^n}{(p+q)(q'+q)} = 1 + \frac{(-1)^n}{(p+q)(q'+q)}. \tag{2.26}$$

Substituting (2.26) into (2.25) yields the desired formula for $P_{GK}(\mathbf{a})$.

To obtain an analogous formula for $P_{GK}(\overleftarrow{\mathbf{a}})$, note that the denominator in the formula (2.23) for $P_{GK}(\mathbf{a})$, i.e., the expression $(p+q)(q'+q)$, is the product of the sum of the entries in the second column of $C(\mathbf{a})$ with the sum of the entries in the second row of $C(\mathbf{a})$. If $\mathbf{a}$ is replaced by $\overleftarrow{\mathbf{a}}$, then by Lemma 2.2, the entries in the second column of $C(\overleftarrow{\mathbf{a}})$ become (q', q) and thus have sum $q' + q$, while the entries in the second row become (p, q) and thus have sum $p + q$. The product of these two sums is $(q'+q)(p+q)$. The latter expression is identical to the corresponding expression, $(p+q)(q'+q)$, for the matrix $C(\mathbf{a})$. Since the strings $\mathbf{a}$ and $\overleftarrow{\mathbf{a}}$ both have the same length n, it follows that $P_{GK}(\overleftarrow{\mathbf{a}}) = P_{GK}(\mathbf{a})$. $\square$

3. Proof of Theorem 1

Throughout this section we assume μ is a probability measure on $[0, 1]$ with continuous density function $f(x)$, so that

$$\mu(I) = \int_I f(x)dx \tag{3.1}$$

for any interval $I \subseteq [0,1]$ and

$$\mu([0,1]) = \int_0^1 f(x)dx = 1. \tag{3.2}$$

The crux of the proof lies in the following result.

Lemma 3.1. *Let* $\mathbf{a} = (a_1, \ldots, a_n)$ *be a string of positive integers, and let*

$$r = [a_1, \ldots, a_n] \tag{3.3}$$

be the rational number represented by the continued fraction $[\mathbf{a}]$. *Suppose that the measure* μ *satisfies the symmetry property for all strings of the form* $(\mathbf{a}, t) = (a_1, \ldots, a_n, t)$, $t \in \mathbb{N}$*; i.e., suppose that*

$$\mu(I(\mathbf{a}, t)) = \mu(I(t, \overleftarrow{\mathbf{a}})) \quad (t \in \mathbb{N}). \tag{3.4}$$

Then we have

$$f(r) = \frac{f(0)}{1+r}. \tag{3.5}$$

Proof. By Lemma 2.4 we have, for any $t \in \mathbb{N}$,

$$I(\mathbf{a}, t) = \left\langle \alpha(t), \alpha(t) + (-1)^{n+1}\delta(t) \right\rangle, \tag{3.6}$$

$$I(t, \overleftarrow{\mathbf{a}}) = \left\langle \beta(t), \beta(t) + (-1)^{n+1}\epsilon(t) \right\rangle, \tag{3.7}$$

where

$$\alpha(t) = \frac{p' + tp}{q' + tq}, \quad \delta(t) = \frac{1}{(q' + tq)(q' + (t+1)q)}, \tag{3.8}$$

$$\beta(t) = \frac{q}{q' + tq}, \quad \epsilon(t) = \frac{1}{(q' + tq)(p' + q' + t(p+q))}. \tag{3.9}$$

(Recall that p'/q' and p/q denote the last two convergents of the continued fraction $[a_1, \ldots, a_n]$, so that $[a_1, \ldots, a_{n-1}] = p'/q'$ and $[a_1, \ldots, a_n] = p/q = r$, with the convention that, when $n = 1$, $(p', q') = (p_0, q_0) = (0, 1)$.)

Using (3.1) and the mean value theorem for integrals, it follows that

$$\mu(I(\mathbf{a}, t)) = \left| \int_{\alpha(t)}^{\alpha(t)+(-1)^{n+1}\delta(t)} f(x)dx \right| = f(\xi(t))\delta(t), \tag{3.10}$$

$$\mu(I(t, \overleftarrow{\mathbf{a}})) = \left| \int_{\beta(t)}^{\beta(t)+(-1)^{n+1}\epsilon(t)} f(x)dx \right| = f(\eta(t))\epsilon(t), \tag{3.11}$$

where $\xi(t)$ and $\eta(t)$ are real numbers in $[0,1]$ satisfying

$$|\xi(t) - \alpha(t)| \le \delta(t), \tag{3.12}$$

$$|\eta(t) - \beta(t)| \le \epsilon(t). \tag{3.13}$$

Now note that, as $t \to \infty$, we have, by (3.8) and (3.9),

$$\alpha(t) = \frac{p + p'/t}{q + q'/t} \to \frac{p}{q} = r, \quad \delta(t) \to 0,$$
$$\beta(t) = \frac{p'}{p + tp'} \to 0, \quad \epsilon(t) \to 0.$$

Using (3.12) and (3.13), it follows that

$$\lim_{t\to\infty} \xi(t) = \lim_{t\to\infty} \alpha(t) = r,$$
$$\lim_{t\to\infty} \eta(t) = \lim_{t\to\infty} \beta(t) = 0,$$

and hence, by the continuity of $f(x)$,

$$\lim_{t\to\infty} f(\xi(t)) = f(r), \tag{3.14}$$
$$\lim_{t\to\infty} f(\eta(t)) = f(0). \tag{3.15}$$

On the other hand, the symmetry assumption (3.4) along with the formulas (3.10) and (3.11) imply that

$$f(\xi(t))\delta(t) = f(\eta(t))\epsilon(t),$$

and hence

$$f(\xi(t)) = f(\eta(t))\frac{\epsilon(t)}{\delta(t)}.$$

Letting $t \to \infty$ on both sides of the latter identity, we obtain, in view of (3.14) and (3.15),

$$\begin{aligned} f(r) &= f(0) \lim_{t\to\infty} \frac{\epsilon(t)}{\delta(t)} \\ &= f(0) \lim_{t\to\infty} \frac{(q' + tq)(q' + (t+1)q)}{(q' + tq)(p' + q' + t(p+q))} = \frac{f(0)}{1 + p/q} = \frac{f(0)}{1 + r}, \end{aligned}$$

as claimed. □

Proof of Theorem 1, Completion. Assume now that μ satisfies the full strength of the hypothesis of Theorem 1, i.e., that the symmetry property $\mu(I(\mathbf{a})) = \mu(I(\overleftarrow{\mathbf{a}}))$ holds for all finite strings $\mathbf{a}$ of positive integers whose length belongs to a given infinite set S. By Lemma 3.1 we then have

$$f(r) = \frac{f(0)}{1 + r} \tag{3.16}$$

for all rational numbers r of the form

$$r = [a_1, \dots, a_n], \quad n + 1 \in S, \quad a_i \in \mathbb{N} \quad (i = 1, \dots, n). \tag{3.17}$$

To complete the proof, it remains to show that the set of numbers r of the form (3.17) is dense in the interval $[0, 1]$. Indeed, if this set is dense in $[0, 1)$, then the continuity of f implies that the relation $f(x) = f(0)/(1 + x)$ holds for *all* real $x \in (0, 1)$, while the assumption that f is a probability density function forces $f(0) = 1/\log 2$. Hence f is the density function of the Gauss-Kuzmin measure, and we conclude $\mu = \mu_{GK}$, as claimed.

To prove the above claim, it is enough to show that every *irrational* number in $(0, 1)$ can be approximated arbitrarily closely by numbers of the form (3.17). Fix an irrational number $x \in (0, 1)$, let

$$x = [a_1, a_2, \dots]$$

be the (infinite) continued fraction expansion of x, and let

$$r_n = [a_1, \dots, a_n]$$

be the nth convergent of this continued fraction. Note that r_n is of the form (3.17) whenever $n + 1 \in S$. Since $\lim_{n\to\infty} r_n = x$ and the set S is infinite, it follows that x can be approximated arbitrarily closely by numbers r_n with $n + 1 \in S$, and hence by numbers of the form (3.17). This completes the proof of Theorem 1. $\square$

Remark 3.2. The above proof shows that to obtain the conclusion of the theorem, it suffices to impose the symmetry property (1.15) on strings of the form $\mathbf{a} = (\mathbf{a}', t)$, $t \in \mathbb{N}$, where $\mathbf{a}'$ runs through a set of finite strings of positive integers with the property that the rational numbers $[\mathbf{a}']$ represented by these strings are dense in $[0, 1]$.

In fact, using set theoretic arguments one can show the following refinement of the theorem: Given any infinite set S of positive integers, there exists a sequence of "test strings" $\mathbf{a}_s$, $s \in S$, where $\mathbf{a}_s$ has length s, such that the conclusion of the theorem remains valid if the symmetry property holds for the strings $\mathbf{a}_s$, $s \in S$.

4. Proof of Theorem 2

Given $N \in \mathbb{N}$, we seek to construct a probability measure μ on $[0, 1]$ with continuous density function that satisfies the symmetry property

$$\mu(I(\mathbf{a})) = \mu(I(\overleftarrow{\mathbf{a}})) \tag{4.1}$$

for all strings $\mathbf{a}$ of length at most N, but is different from the Gauss-Kuzmin measure μ_{GK}.

The key to our argument is contained in the following lemma.

Lemma 4.1. *Let $N \in \mathbb{N}$ be given and let μ be a probability measure on $[0,1]$ satisfying*

$$\mu(I(\mathbf{a})) = \mu_{GK}(I(\mathbf{a})) \tag{4.2}$$

for all strings $\mathbf{a} = (a_1, \dots, a_N)$ of positive integers of length exactly N. Then μ satisfies the symmetry property (4.1) for all strings $\mathbf{a} = (a_1, \dots, a_n)$ of length $n \le N$.

Proof. First note that, since the Gauss-Kuzmin measure μ_{GK} satisfies the symmetry property (4.1), the assumptions of the lemma imply that for strings $\mathbf{a}$ of length exactly N,

$$\mu(I(\mathbf{a})) = \mu_{GK}(I(\mathbf{a})) = \mu_{GK}(I(\overleftarrow{\mathbf{a}})) = \mu(I(\overleftarrow{\mathbf{a}})) \tag{4.3}$$

Thus μ satisfies the symmetry property (4.1) for strings of length N.

It therefore remains to show that (4.1) also holds for strings $\mathbf{a}$ of length $n < N$. This will follow if we can show that the assumption (4.2) remains valid for such strings.

To see this, let $\mathbf{a} = (a_1, \dots, a_n)$ be a string of positive integers of length $n < N$. From the definition of $I(\mathbf{a})$ as the set of real numbers in $(0,1)$ whose continued fraction expansion begins with the digits $a_1, \dots, a_n$ it is clear that, modulo a set of measure zero[4], $I(\mathbf{a})$ is the disjoint union of sets $I(\mathbf{a}')$, where $\mathbf{a}'$ runs through strings of the form $\mathbf{a}' = (\mathbf{a}, \mathbf{b}) = (a_1, \dots, a_n, b_1, \dots, b_{N-n})$, with $b_i \in \mathbb{N}$. Since the strings $\mathbf{a}'$ have length exactly N, the assumption of the lemma applies to these strings, so we have

$$\mu(I(\mathbf{a})) = \sum_{\mathbf{b} \in \mathbb{N}^{N-n}} \mu(I(\mathbf{a}, \mathbf{b})) = \sum_{\mathbf{b} \in \mathbb{N}^{N-n}} \mu_{GK}(I(\mathbf{a}, \mathbf{b})) = \mu_{GK}(\mathbf{a}),$$

as claimed. □

Proof of Theorem 2, Completion. Let $N \in \mathbb{N}$ be given. In view of the lemma, it suffices to construct a probability measure μ on $[0,1]$ with continuous density function that takes on the same value as the Gauss-Kuzmin measure μ_{GK} on sets of the form $I(\mathbf{a})$, $\mathbf{a} \in \mathbb{N}^N$, but is different from μ_{GK}.

Choose a particular string $\mathbf{a}_0 = (a_{1,0}, \dots, a_{N,0})$ of N positive integers, let $\alpha < \beta$ be the endpoints of the interval $I(\mathbf{a}_0)$, and let μ be the measure on $[0,1]$ with density function given by

$$f(x) = \begin{cases} f_0(x) & \text{if } \alpha < x < \beta, \\ f_{GK}(x) & \text{otherwise,} \end{cases} \tag{4.4}$$

where

$$f_{GK}(x) = \frac{1}{(\log 2)(1+x)} \quad (0 \le x \le 1)$$

[4]The exceptional set consists of rational numbers whose continued fraction expansion begins with the string $\mathbf{a}$, but has fewer than N digits and thus is not counted in any set $I(\mathbf{a}')$, where $\mathbf{a}' \in \mathbb{N}^N$.

is the density function of the Gauss-Kuzmin measure μ_{GK} and $f_0(x)$ is any non-negative continuous function on $[\alpha, \beta]$ satisfying

$$f_0(\alpha) = f_{GK}(\alpha), \quad f_0(\beta) = f_{GK}(\beta), \tag{4.5}$$

$$f_0(x) \neq f_{GK}(x) \quad \text{for some } x \in (\alpha, \beta), \tag{4.6}$$

$$\int_\alpha^\beta f_0(x)dx = \int_\alpha^\beta f_{GK}(x)dx. \tag{4.7}$$

The definition (4.4) of $f(x)$ implies that $f(x)$ is identical to $f_{GK}(x)$ outside the interval $I(\mathbf{a}_0)$. Since the intervals $I(\mathbf{a})$, $\mathbf{a} \in \mathbb{N}^N$, are pairwise disjoint, it follows that $\mu(I(\mathbf{a})) = \mu_{GK}(I(\mathbf{a}))$ holds for all strings $\mathbf{a} \in \mathbb{N}^N$ with $\mathbf{a} \neq \mathbf{a}_0$, while condition (4.7) ensures that $\mu(I(\mathbf{a})) = \mu_{GK}(I(\mathbf{a}))$ also holds for $\mathbf{a} = \mathbf{a}_0$. On the other hand, by (4.6) and the continuity of f_0, the measure μ is different from the Gauss-Kuzmin measure μ_{GK}.

The conditions (4.5) and the assumption that f_0 is continuous on (α, β) ensure that the function $f(x)$ defined by (4.4) is continuous on the entire interval $[0, 1]$. Moreover, (4.7) implies that $\int_0^1 f(x)dx = \int_0^1 f_{GK}(x)dx = 1$, so that $f(x)$ is a continuous probability density function on $[0, 1]$. Thus, the measure μ has all of the required properties, and the proof is complete. □

5. Proof of Theorem 3

5.1. Characteristic numbers

By Lemma 2.5 the Gauss-Kuzmin measure $P_{GK}(\mathbf{a})$ of a string $\mathbf{a} = (a_1, \dots, a_n)$ of positive integers is given by

$$P_{GK}(\mathbf{a}) = \left| \log_2 \left(1 + \frac{(-1)^n}{(p+q)(q'+q)} \right) \right|. \tag{5.1}$$

where p', q', p, q are the entries of the convergent matrix $C(\mathbf{a}) = \left(\begin{smallmatrix} p' & p \\ q' & q \end{smallmatrix}\right)$. In particular, $P_{GK}(\mathbf{a})$ depends only on the parity of n and the quantity $(p+q)(q'+q)$ appearing in the denominator on the right side of (5.1). In view of the key role played by this quantity we introduce the following definition.

Definition 5.1 (Characteristic number)**.** Let $\mathbf{a} = (a_1, \dots, a_n)$ be a finite string of positive integers with convergent matrix $C(\mathbf{a}) = \left(\begin{smallmatrix} p' & p \\ q' & q \end{smallmatrix}\right)$. The *characteristic number*, $\chi(\mathbf{a})$, of the string $\mathbf{a}$ is defined as

$$\chi(\mathbf{a}) = (p+q)(q'+q). \tag{5.2}$$

Thus $\chi(\mathbf{a})$ is the product of the sum of the entries in the second column with the sum of the entries of the second row of the convergent matrix, $C(\mathbf{a})$, associated with this string.

In light of the above remarks we then have:

Lemma 5.2. *If* $\mathbf{a}$ *and* $\mathbf{b}$ *are strings of positive integers whose lengths have the same parity, then* $P_{GK}(\mathbf{a}) = P_{GK}(\mathbf{b})$ *holds if and only if* $\chi(\mathbf{a}) = \chi(\mathbf{b})$.

In particular, *two permutations of a string have the same Gauss-Kuzmin measure if and only if they have the same characteristic number.* The proofs of both parts of Theorem 3 as well as the experimental results presented in Section 6 are based on this crucial observation.

5.2. Proof of Theorem 3(i)

Given a string (a, b, c) of positive integers, let $\chi(a, b, c)$ be the characteristic number of this string. We seek to show that if (a', b', c') is a permutation of (a, b, c) such that $(a', b', c') \neq (a, b, c)$ and $(a', b', c') \neq (c, b, a)$, then $\chi(a', b', c') \neq \chi(a, b, c)$. We first consider the particular permutation $(a', b', c') = (b, a, c)$, i.e., the permutation that interchanges a with b.

Lemma 5.3. *Let* (a, b, c) *be a string of positive integers. If* $a \neq b$, *then*

$$\chi(a, b, c) \neq \chi(b, a, c). \tag{5.3}$$

Proof. The proof is based on an explicit calculation of the characteristic number of a string (a, b, c) of positive integers as a polynomial in the variables a, b, c.

Fix a string (a, b, c) of positive integers with $a \neq b$, and suppose, to get a contradiction, that

$$\chi(b, a, c) = \chi(a, b, c). \tag{5.4}$$

Using (2.13), we can calculate the convergent matrix of the string (a, b, c) as

$$C(a, b, c) = \begin{pmatrix} 0 & 1 \\ 1 & a \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & b \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & c \end{pmatrix} = \begin{pmatrix} b & bc+1 \\ ab+1 & abc+a+c \end{pmatrix},$$

so we have $p' = b$, $q' = ab + 1$, $p = bc + 1$, and $q = abc + a + c$. Substituting these values into (5.2), we obtain

$$\chi(a, b, c) = (p + q)(q' + q) = (abc + bc + a + c + 1)(abc + ab + a + c + 1). \tag{5.5}$$

Introducing the polynomial

$$S = S(a, b, c) = abc + a + b + c + 1, \tag{5.6}$$

we can rewrite (5.5) as

$$\chi(a, b, c) = (S + bc - b)(S + ab - b) = S^2 + S(ab + bc - 2b) + b^2(a - 1)(c - 1). \tag{5.7}$$

Now note that, since S is a symmetric polynomial in a, b, c, permuting these variables does not affect the value of S. Hence, interchanging a and b in (5.7) and subtracting the resulting expression from the expression on the right of (5.7) yields, in view of our assumption (5.4),

$$\begin{aligned} 0 &= \chi(a,b,c) - \chi(b,a,c) \\ &= (bc - 2b - ac + 2a)S + (b^2(a-1) - a^2(b-1))(c-1) \\ &= (b-a)\left((c-2)S + (ab-a-b)(c-1)\right), \end{aligned} \tag{5.8}$$

and hence, since $a \neq b$,

$$(c-2)S = (c-1)(a+b-ab) = (c-1)\big[1-(a-1)(b-1)\big]. \tag{5.9}$$

We show that (5.9) cannot hold. If $c = 1$, the right-hand side of (5.9) vanishes, while the left-hand side is negative, so we have a contradiction. If $c = 2$, the left-hand side vanishes, while the right-hand side is non-zero since $(a-1)(b-1) \neq 1$ by our assumption $a \neq b$, so this case is also impossible. Finally, if $c \geq 3$, then (5.9) implies

$$S = \frac{c-1}{c-2}\big[1-(a-1)(b-1)\big] \leq \frac{c-1}{c-2} \leq 2,$$

which is again a contradiction since, by (5.6), $S = abc + a + b + c + 1 \geq 5$. □

Proof of Theorem 3(i). Let (a, b, c) be a string of positive integers, and let (a', b', c') be a permutation of (a, b, c). Suppose that

$$\chi(a', b', c') = \chi(a, b, c). \tag{5.10}$$

We seek to show that (5.10) can only hold if the permutation (a', b', c') is either the string (a, b, c) itself, or its reverse, (c, b, a).

If $b' = b$, the desired conclusion obviously holds. Assume therefore that $b' \neq b$. Then $b' = a$ or $b' = c$, and by the symmetry property we may assume without loss of generality that $b' = a$. Thus we have either $(a', b', c') = (b, a, c)$ or $(a', b', c') = (c, a, b)$. Since, by the symmetry property, $\chi(b, a, c) = \chi(c, a, b)$, it suffices to consider the case $(a', b', c') = (b, a, c)$. But in this case Lemma 5.3 along with our assumption (5.10) implies that $a = b$. Hence $(a', b', c') = (a, a, c) = (a, b, c)$, i.e., the permutation (a', b', c') is the identity permutation. This completes the proof. □

5.3. Proof of Theorem 3(ii)

For part (ii) of Theorem 3 we seek to construct, for any given length $n \geq 4$, an $\lfloor (n-2)/2 \rfloor$-parameter family of strings $\mathbf{a}$ of length n that have a nontrivial symmetry in the sense of Definition 1.1. In view of Lemma 5.2, this amounts to constructing strings $\mathbf{a}$ for which there exists a permutation $\sigma(\mathbf{a})$ with $\sigma(\mathbf{a}) \neq \mathbf{a}$ and $\sigma(\mathbf{a}) \neq \overleftarrow{\mathbf{a}}$ that has the same characteristic number as the string $\mathbf{a}$.

The key to our construction lies in a special class of strings defined as follows.

Definition 5.4 (Stable strings)**.** Let $\mathbf{a} = (a_1, \ldots, a_n)$ be a finite string of positive integers and let $C(\mathbf{a}) = \left(\begin{smallmatrix} p' & p \\ q' & q \end{smallmatrix}\right)$ be the convergent matrix of $\mathbf{a}$. The string $\mathbf{a}$ is called *stable* if it satisfies

$$p = 2q', \tag{5.11}$$

In other words, a stable string is a string whose convergent matrix has the property that its $(1, 2)$ entry is exactly twice its $(2, 1)$ entry.

We will show in Lemma 5.7 below that a stable string of length n gives rise to a string of length $n + 2$ with a nontrivial symmetry. Thus our task is reduced to constructing, for any given length $n \geq 2$, an infinite $\lfloor n/2 \rfloor$-parameter family of stable strings of length n. This is accomplished by the following lemma.

Lemma 5.5.

(i) The following families of strings are stable:

$$(t, 2t) \quad (t \in \mathbb{N}), \tag{5.12}$$

$$(t, 1, 2t + 1) \quad (t \in \mathbb{N}). \tag{5.13}$$

(ii) If $\mathbf{a} = (a_1, \ldots, a_n)$ is a stable string, then any string of the form

$$(t, \overleftarrow{\mathbf{a}}, 2t) = (t, a_n, \ldots, a_1, 2t) \quad (t \in \mathbb{N}) \tag{5.14}$$

is also stable.

(iii) For each $n \geq 2$ there exists an $\lfloor n/2 \rfloor$-parameter family of stable strings of length n.

Proof. (i) Using (2.13), we calculate the convergent matrices associated with the strings (5.12) and (5.13):

$$C(t, 2t) = \begin{pmatrix} 0 & 1 \\ 1 & t \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 2t \end{pmatrix} = \begin{pmatrix} 1 & 2t \\ t & 1 + 2t^2 \end{pmatrix},$$

$$C(t, 1, 2t + 1) = \begin{pmatrix} 0 & 1 \\ 1 & t \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 2t + 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 + 2t \\ 1 + t & 1 + 4t + 2t^2 \end{pmatrix}.$$

Both of these matrices satisfy the stability condition (5.11) for any $t \in \mathbb{N}$, so the associated families of strings are stable as claimed.

(ii) Assume $\mathbf{a} = (a_1, \ldots, a_n)$ is a stable string with convergent matrix $C(\mathbf{a}) = \left(\begin{smallmatrix} p' & p \\ q' & q \end{smallmatrix}\right)$. Using the relation (cf. Lemma 2.2(i)) $C(\overleftarrow{\mathbf{a}}) = \left(\begin{smallmatrix} p' & q' \\ p & q \end{smallmatrix}\right)$, we obtain, for any $t \in \mathbb{N}$,

$$
\begin{aligned}
C(t, \overleftarrow{\mathbf{a}}, 2t) &= \begin{pmatrix} 0 & 1 \\ 1 & t \end{pmatrix} C(\overleftarrow{\mathbf{a}}) \begin{pmatrix} 0 & 1 \\ 1 & 2t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & t \end{pmatrix} \begin{pmatrix} p' & q' \\ p & q \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 2t \end{pmatrix} \\
&= \begin{pmatrix} q & p + 2tq \\ q' + tq & p' + 2tq' + p + 2tq \end{pmatrix} = \begin{pmatrix} r' & r \\ s' & s \end{pmatrix},
\end{aligned}
$$

say. Since the string $\mathbf{a}$ is stable, we have $p = 2q'$. It follows that $r = p + 2tq = 2q' + 2tq = 2(q' + tq) = 2s'$, so the string $(t, \overleftarrow{\mathbf{a}}, 2t)$ is stable as well.

(iii) This follows from parts (i) and (ii) by starting out with the families of strings (5.12) and (5.13) and inductively applying the procedure described in part (ii) of the lemma. At each step the length of the string is increased by 2 and an additional free parameter is introduced. Thus the total number of free parameters in the families of strings generated by this process is $\lfloor n/2 \rfloor$, where n is the length of the string. □

Remark 5.6. The families of stable strings generated by the iterative procedure of Lemma 5.5 can be described explicitly. In the case $n = 2m$ is even, the strings are of the form

$$
(\delta_1 t_m, \delta_2 t_{m-1}, \ldots, \delta_m t_1, \delta_{m-1} t_1, \ldots, \delta_1 t_{m-1}, \delta_0 t_m) \quad (t_1, \ldots, t_m \in \mathbb{N}), \tag{5.15}
$$

where $\delta_i = 1$ if i is odd and $\delta_i = 2$ if i is even. A similar, though slightly more complicated, explicit description could be given for the case when n is odd. Table 2 shows the families of strings obtained from the lemma for lengths $n \leq 7$.

n	$\mathbf{a}$
2	$(t_1, 2t_1)$
3	$(t_1, 1, 2t_1 + 1)$
4	$(t_2, 2t_1, t_1, 2t_2)$
5	$(t_2, 2t_1 + 1, 1, t_1, 2t_2)$
6	$(t_3, 2t_2, t_1, 2t_1, t_2, 2t_3)$
7	$(t_3, 2t_2, t_1, 1, 2t_1 + 1, t_2, 2t_3)$

Table 2: Families of stable strings $\mathbf{a}$ of length $n \in \{2, 3, 4, 5, 6, 7\}$ constructed by the procedure of Lemma 5.5. Here $t_1, t_2, \ldots$ are arbitrary positive integer parameters.

To complete the proof of Theorem 3(ii), we show in the following lemma that any stable string of length n yields a string of length $n + 2$ that has a nontrivial symmetry. Given a string $\mathbf{a} = (a_1, \ldots, a_n)$ of positive integers, let $\mathbf{a}^+$ be the string defined by

$$\mathbf{a}^+ = (2, 1, a_1, \ldots, a_{n-1}, a_n + 1). \tag{5.16}$$

Thus, $\mathbf{a}^+$ is the string obtained from $\mathbf{a}$ by prepending the digits 2 and 1 and incrementing the last digit in $\mathbf{a}$ by 1.

Lemma 5.7. *If $\mathbf{a} = (a_1, \ldots, a_n)$ is a stable string of positive integers, then the string $\mathbf{a}^+$ defined by* (5.16) *has a nontrivial symmetry σ given by*

$$\sigma(\mathbf{a}^+) = (2, a_n + 1, a_{n-1}, \ldots, a_1, 1). \tag{5.17}$$

Proof. Note that $\sigma(\mathbf{a}^+)$ is the permutation that reverses the last $n+1$ digits of $\mathbf{a}^+$. Since $a_n + 1 \neq 1$, this permutation cannot be the identity permutation, and since $2 \neq 1$, it cannot be the permutation that reverses the digits of $\mathbf{a}^+$. Thus, σ is a nontrivial permutation in the sense of Definition 1.1, and in view of Lemma 5.2, it therefore remains to show that the strings $\mathbf{a}^+$ and $\sigma(\mathbf{a}^+)$ have the same characteristic number, i.e., that

$$\chi(\mathbf{a}^+) = \chi(\sigma(\mathbf{a}^+)). \tag{5.18}$$

Using the identity $\left(\begin{smallmatrix} 0 & 1 \\ 1 & a+1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 & 1 \\ 1 & a \end{smallmatrix}\right)\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and (2.13) we obtain

$$\begin{aligned} C(\mathbf{a}^+) &= \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix}\cdots\begin{pmatrix} 0 & 1 \\ 1 & a_{n-1} \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & a_n+1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} C(\mathbf{a}) \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} p' & p \\ q' & q \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} p'+q' & p'+q'+p+q \\ 2p'+3q' & 2p'+3q'+2p+3q \end{pmatrix} = \begin{pmatrix} r' & r \\ s' & s \end{pmatrix}, \end{aligned} \tag{5.19}$$

say. It follows that

$$\begin{aligned} \chi(\mathbf{a}^+) &= (r+s)(s'+s) = (3p'+4q'+3p+4q)(4p'+6q'+2p+3q) \\ &= (3p'+10q'+4q)(4p'+10q'+3q), \end{aligned} \tag{5.20}$$

where the last step follows from the assumption that the string $\mathbf{a}$ is stable and thus satisfies $p = 2q'$.

Similarly, noting that $\sigma(\mathbf{a}^+)$ is the string obtained by reversing the last $n+1$ digits of $\mathbf{a}^+$ (i.e., all digits of $\mathbf{a}^+$ after the leading digit 2), we obtain

$$\begin{aligned} C(\sigma(\mathbf{a}^+)) &= \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}\left[\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} p' & p \\ q' & q \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\right]^T \\ &= \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} p' & q' \\ p & q \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} q'+q & p'+q'+p+q \\ 3q'+2q & 3p'+3q'+2p+2q \end{pmatrix} = \begin{pmatrix} u' & u \\ v' & v \end{pmatrix}, \end{aligned} \tag{5.21}$$

say, and hence

$$\begin{aligned}\chi(\sigma(\mathbf{a}^+)) &= (u+v)(v'+v) = (4p'+4q'+3p+3q)(3p'+6q'+2p+4q)\\ &= (4p'+10q'+3q)(3p'+10q'+4q). \qquad (5.22)\end{aligned}$$

Comparing (5.20) with (5.22) yields the desired symmetry relation (5.18). $\square$

Table 3 illustrates the construction of strings with nontrivial symmetries from stable strings described in Lemmas 5.5 and 5.7.

$\mathbf{a}$	$\mathbf{a}^+$	$\sigma(\mathbf{a}^+)$
$(t_1, 2t_1)$	$(2, 1, t_1, 2t_1 + 1)$	$(2, 2t_1 + 1, t_1, 1)$
$(t_1, 1, 2t_1 + 1)$	$(2, 1, t_1, 1, 2t_1 + 2)$	$(2, 2t_1 + 2, 1, t_1, 1)$
$(t_2, 2t_1, t_1, 2t_2)$	$(2, 1, t_2, 2t_1, t_1, 2t_2 + 1)$	$(2, 2t_2 + 1, t_1, 2t_1, t_2, 1)$
$(t_2, 2t_1 + 1, 1, t_1, 2t_2)$	$(2, 1, t_2, 2t_1 + 1, 1, t_1, 2t_2 + 1)$	$(2, 2t_2 + 1, t_1, 1, 2t_1 + 1, t_2, 1)$

Table 3: Families of stable strings $\mathbf{a}$ of length $n \in \{2, 3, 4, 5\}$ and the associated families of strings $\mathbf{a}^+$ of length $n + 2$ with nontrivial symmetries $\sigma(\mathbf{a}^+)$ given by (5.16) and (5.17) Here t_1 and t_2 are arbitrary positive integer parameters.

.

Remark 5.8. The families of strings $\mathbf{a}^+$ of *even* lengths constructed via Lemmas 5.5 and 5.7 involve each of the digits 2 and 1 exactly once, along with digits of the forms $(*)$ t_i, $2t_i$, $2t_i+1$ (cf. Table 3). By requiring the parameters t_i to be pairwise distinct and satisfy $t_i \equiv 1 \bmod 4$ and $t_i > 2$ one can ensure that the digits of the strings $\mathbf{a}^+$ of even length obtained from this construction are pairwise distinct.

For strings $\mathbf{a}^+$ of *odd* length ≥ 5 this is not the case as the construction given above *necessarily* involves two occurrences of the digit 1 (see the cases $n = 3$ and $n = 5$ in Table 3), although by restricting the parameters t_i as before one can ensure that all remaining digits of $\mathbf{a}^+$ are pairwise distinct. The duplication of the digit 1 in the case of strings of odd length can be avoided by a generalized version of the construction given in Lemmas 5.5 and 5.7 that involves an additional parameter s. We give a brief sketch of the argument.

Given a positive integer s, define a string $\mathbf{a}$ to be *s-stable* if it satisfies $p = (s^2 + s)q'$. The latter condition generalizes the stability condition (5.11), which corresponds to the case $s = 1$. Similar to Lemmas 5.5 and 5.7 one can verify that all strings of the form $(t, (s^2+s)t)$ and $(t, s^2+s-1, (s^2+s)t+1)$, where $t \in \mathbb{N}$, are s-stable, and then use an inductive process to construct, for each $n \geq 2$, an infinite $\lfloor n/2 \rfloor$-parameter family of s-stable strings, each of which gives rise to a string $\mathbf{a}^+$ with a nontrivial symmetry. For *odd* $n \geq 5$ the strings $\mathbf{a}^+$ of length n obtained in this manner involve $\lfloor (n-2)/n \rfloor$ additional parameters t_i and digits of the form 2, s, $2s$, t_i, $(s^2+s)t_i$, and $(s^2+s)t_i+s$. These digits will be pairwise distinct if we require the parameter s to be greater than 2 and the parameters t_i to be pairwise distinct and congruent to 1 modulo $2s^2+2s$.

6. Numerical data and conjectures

Recall (cf. Definition 1.1) that a string $\mathbf{a} = (a_1, \dots, a_n)$ of positive integers is said to have a *nontrivial symmetry* if there exists a permutation of $\mathbf{a}$ other than the identity and the reverse that preserves the Gauss-Kuzmin measure $P_{GK}(\mathbf{a})$ of the string. By Theorem 3, strings of length $n = 3$ have no nontrivial symmetries, while for each $n \geq 4$ there exists an infinite family of strings of length n that *do* have a nontrivial symmetry. Conjecture 4 states that strings of the latter type are the exception rather than the rule in the sense that their proportion among all strings of length n of digits in $\{1, \dots, N\}$ tends to 0 as $N \to \infty$. In this section, we provide numerical evidence supporting this conjecture, and we propose refined versions of this conjecture.

For simplicity, we consider only strings of *distinct* digits. This restriction does not affect the assertion of Conjecture 4 since, for each fixed n, the proportion of strings $\mathbf{a} = (a_1, \dots, a_n) \in \{1, \dots, N\}^n$ that have distinct digits is $N(N-1)\dots(N-n+1)/N^n$, which converges to 1 as $N \to \infty$.

Given an n-tuple $(a_1, \dots, a_n)$ of distinct positive integers, let

$$\nu(\mathbf{a}) = \#\{P_{GK}(\sigma(\mathbf{a})) : \sigma \in S_n\}, \tag{6.1}$$

where S_n is the set of all permutations on $\{1, \dots, n\}$. Thus, $\nu(\mathbf{a})$ is the number of distinct values of the Gauss-Kuzmin measure $P_{GK}(\sigma(\mathbf{a}))$ as $\sigma(\mathbf{a})$ runs through the $n!$ permutations of $\{a_1, \dots, a_n\}$. Note that $\nu(\mathbf{a})$ depends only on the digits $a_1, \dots, a_n$, not on the order in which these digits occur in the string $\mathbf{a}$. We may therefore assume that $a_1 < \dots < a_n$.

We have trivially $\nu(\mathbf{a}) \leq n!$. Moreover, pairing up each permutation of $\mathbf{a}$ with its reverse, we see that the symmetry property (1.14) implies $\nu(\mathbf{a}) \leq n!/2$, *with equality if and only if none of the permutations of* $\mathbf{a}$ *has a nontrivial symmetry.* Thus, a natural way to quantify the occurrence of strings with nontrivial symmetries is to compare, for large $N \in \mathbb{N}$, the number of tuples $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{N}^n$ with $a_1 < \dots < a_n \leq N$ that satisfy the condition $\nu(\mathbf{a}) < n!/2$ with the total number of such tuples, i.e., with $\binom{N}{n}$. Set

$$f(N,n) = \#\left\{\mathbf{a} = (a_1,\ldots,a_n) \in \mathbb{N}^n : a_1 < \cdots < a_n \le N,\ \nu(\mathbf{a}) < \frac{n!}{2}\right\}, \tag{6.2}$$

$$\delta(N,n) = \frac{f(N,n)}{\binom{N}{n}}. \tag{6.3}$$

The quantity $\delta(N,n)$ represents the probability that a random sample of n distinct digits in $\{1,\ldots,N\}$ has a permutation with a nontrivial symmetry.

The quantities $\nu(\mathbf{a})$ defined in (6.1), and hence the numbers $f(N,n)$ and $\delta(N,n)$, can be computed from the explicit formula (2.23) for the Gauss-Kuzmin measure $P_{GK}(\mathbf{a})$. Using the symbolic computation software *Mathematica* we carried out these computations for $n \in \{4,5,6\}$ and a range of values of N.

For $n = 4$, we computed the exact values of $f(N,4)$ for all positive integers $N \le 120$. Table 4 and Figure 1 below show the results of these computations. The table lists, for $N = 10, 20, \ldots, 120$, the total number, $\binom{N}{4}$, of unordered 4-tuples of distinct positive integers $\le N$ along with the number, $f(N,4)$, and proportion, $\delta(N,4) = f(N,4)/\binom{N}{4}$, of these tuples that have a permutation with a nontrivial symmetry. Also shown is the ratio $f(N,4)/N$, which measures the rate of growth of the function $f(N,4)$ compared to that of a linear function in N.

N	$\binom{N}{4}$	$f(N,4)$	$f(N,4)/N$	$\delta(N,4)$
10	210	10	1.0000	0.047619
20	4845	30	1.5000	0.006192
30	27405	47	1.5667	0.001715
40	91390	66	1.6500	0.000722
50	230300	87	1.7400	0.000378
60	487635	104	1.7333	0.000213
70	916895	121	1.7286	0.000132
80	1581580	142	1.7750	0.000090
90	2555190	159	1.7667	0.000062
100	3921225	178	1.7800	0.000045
110	5773185	199	1.8091	0.000034
120	8214570	216	1.8000	0.000026

Table 4: Strings of length 4 with nontrivial symmetries.

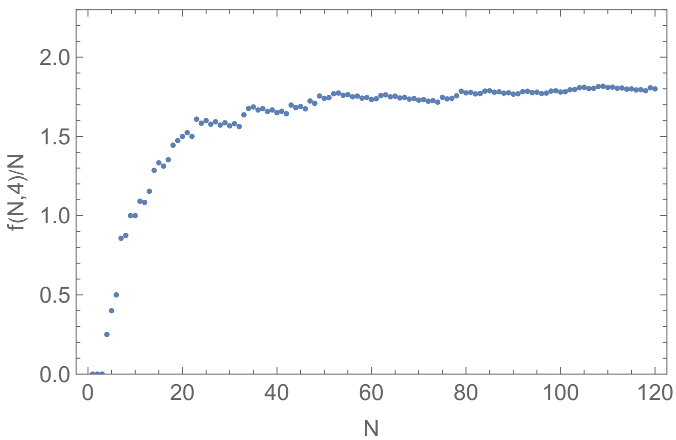

Figure 1: The ratios $f(N,4)/N$ for $N \leq 120$.

The data shown in Table 4 and Figure 1 provide rather compelling evidence that strings of length 4 with nontrivial symmetries are exceedingly rare: For example, among the $\binom{120}{4} = 8,214,570$ unordered 4-tuples of distinct integers in $\{1,\ldots,120\}$ only 216 have a permutation with a nontrivial symmetry. The probabilities $\delta(N,4)$ listed in the last column of the table seem to converge rapidly to 0 as $N \to \infty$. Moreover, the counts $f(N,4)$ of strings with nontrivial symmetries appear to grow at rate that is roughly linear in N. We are therefore led to the following conjecture.

Conjecture 5. We have:

(i) $\lim\limits_{N\to\infty} \delta(N,4) = 0.$

(ii) $\lim\limits_{N\to\infty} \dfrac{\log f(N,4)}{\log N} = 1.$

Since $\delta(N,4) = f(N,4)/\binom{N}{4} = O(f(N,4)/N^4)$, part (ii) of the conjecture is a stronger version of part (i), implying that $\delta(N,4)$ converges to 0 at a rate $O(1/N^c)$ for any constant $c < 3$. We note that the construction given in the proof of Theorem 3(ii) implies $f(N,4) \gg N$ and hence yields $\liminf_{N\to\infty} \log f(N,4)/\log N = 1$, i.e., the lower bound in part (ii) of the conjecture.

Tables 5 and 6 below show the results of analogous computations for strings of length $n \in \{5,6\}$. Because of running time limitations[5], the range for the variable N had to be restricted to $N \leq 60$ for $n = 5$ and $N \leq 35$ for $n = 6$.

[5]The number of cases to be examined grows at a rate proportional to N^n.

N	$\binom{N}{5}$	$f(N,5)$	$\delta(N,5)$
10	252	8	0.031746
20	15504	43	0.002773
30	142506	85	0.000596
40	658008	137	0.000208
50	2118760	184	0.000087
60	5461512	236	0.000043

Table 5: Strings of length 5 with nontrivial symmetries.

N	$\binom{N}{6}$	$f(N,6)$	$\delta(N,6)$
10	210	23	0.109524
15	5005	100	0.019980
20	38760	276	0.007121
25	177100	496	0.002801
30	593775	746	0.001256
35	1623160	1088	0.000670

Table 6: Strings of length 6 with nontrivial symmetries.

Again, the tables provide convincing evidence that the probabilities $\delta(N,n)$ approach 0 as $N \to \infty$. We make the following conjecture, which generalizes part (i) of the latter conjecture to strings of arbitrary length $n \geq 4$, and implies Conjecture 4.

Conjecture 6. For any integer $n \geq 4$ we have

$$\lim_{N\to\infty} \delta(N,n) = 0.$$

Regarding the rate of growth of $f(N,n)$ for general $n \geq 4$, it seems plausible that, in analogy to part (ii) of Conjecture 5, a relation of the form

$$\lim_{N\to\infty} \frac{\log f(N,n)}{\log N} = \alpha_n, \tag{6.4}$$

or possibly even

$$\lim_{N\to\infty} \frac{f(N,n)}{N^{\alpha_n}} = c_n, \tag{6.5}$$

holds with suitable constants α_n and $c_n > 0$. Unfortunately, we do not have enough data to support a specific conjecture of this type. We note that the families of strings $\mathbf{a}$ constructed in the proof of Theorem 3(ii) involve $\lfloor (n-2)/2 \rfloor$ parameters, and since the digits in $\mathbf{a}$ depend linearly on these parameters, $f(N, n)$ must grow at a rate at least $N^{\lfloor (n-2)/2 \rfloor}$ as $N \to \infty$. Thus, the constant α_n in (6.4) and (6.5) must satisfy

$$\alpha_n \geq \left\lfloor \frac{n-2}{2} \right\rfloor. \tag{6.6}$$

For $n = 4$, the bound (6.6) becomes $\alpha_4 \geq 1$, and the data in Table 4 suggests that this bound is sharp, i.e., that $\alpha_4 = 1$. Whether this remains true for $n > 4$ is an open question.

7. Concluding remarks

While the digits in decimal (and base b) expansions of real numbers behave essentially like independent identically distributed random variables, the statistical behavior of digits in continued fraction expansions of real numbers is more complicated and far less intuitive: The frequency with which a string $\mathbf{a}$ of digits occurs in the continued fraction expansion of a "random" real number is given by the Gauss-Kuzmin measure $P_{GK}(\mathbf{a})$ of this string. Classically, this measure arises either as a solution to a functional equation (see, e.g., [6]) or as the invariant measure with respect to the ergodic transformation corresponding to the continued fraction algorithm (see, e.g., [2]). In this paper we provided a new, elementary, derivation of this measure by showing in Theorem 1 that the Gauss-Kuzmin measure is the *only* (continuous) measure under which the reverse of a finite string of continued fraction digits occurs in the continued fraction expansion of a random real number with the same frequency as the original string.

Motivated by this result, we investigated more generally the extent to which the frequency of a string of digits in the continued fraction expansion depends on the order in which these digits appear in the string. Specifically, we considered the question whether the reverse of a string is the *only* permutation under which this frequency is invariant. In Theorem 3 we proved that this indeed holds for *all* strings of length 3, while for each length $n \geq 4$ there exists an infinite family of strings of length n that do have a permutation other than the reversal under which the frequency remains invariant. Supported by experimental data, we conjecture (Conjecture 4) that strings of the latter type are the exception in the sense that for a "typical" string $\mathbf{a}$ of continued fraction digits the reversal of the string is the *only* permutation that leaves this frequency invariant.

We conclude this paper by mentioning some open problems suggested by these results. The main—and arguably most interesting—set of open problems concern the frequencies of strings with nontrivial symmetries. Call such a string *exceptional*. One way to quantify the occurrence of exceptional strings is via the function $\delta(N,n)$ defined in Section 6, which can be interpreted as the probability that a randomly chosen string of n distinct digits in $\{1,\ldots,N\}$ has a *permutation* that is exceptional.[6] What can one say about the asymptotic behavior of $\delta(N,n)$ as $N\to\infty$?

Alternatively, one can consider, for a *given* string $\mathbf{a}=(a_1,\ldots,a_n)$ of n distinct positive integers, the quantity $\epsilon(\mathbf{a})=1-\nu(\mathbf{a})/(n!/2)$, where $\nu(\mathbf{a})$ is defined in (6.1). As noted in Section 6, we have $\nu(\mathbf{a})\le n!/2$, and hence $\epsilon(\mathbf{a})\ge 0$, with equality if and only if *none* of the permutations of the string $\mathbf{a}$ is exceptional. Thus, $\epsilon(\mathbf{a})$ can be viewed as a measure for the frequency of exceptional strings among all *permutations* of $\mathbf{a}$, and it is of interest to investigate the asymptotic behavior of $\epsilon(\mathbf{a})$ as the length n of the string $\mathbf{a}$ tends to infinity. In particular, if $\mathbf{a}^{(n)}=(1,2,\ldots,n)$, does $\epsilon(\mathbf{a}^{(n)})$ converge to 0 as $n\to\infty$? Is it the case that $\epsilon(\mathbf{a})\to 0$ *uniformly in* $\mathbf{a}$ as the length n of the string $\mathbf{a}$ tends to infinity?

A related problem is to characterize *all* exceptional strings. The families of strings constructed in the proof of Theorem 3(ii) are examples of such strings, but there exist exceptional strings that are not part of these families (nor the generalized families mentioned in Remark 5.8). For example, one can verify that any string of the form $(t+1,1,t+3,t+2)$, where $t\in\mathbb{N}$, has a nontrivial symmetry given by $(t+2,1,t+1,t+3)$.

Another circle of questions concerns the set of frequencies of the strings $\sigma(\mathbf{a})$ as σ runs through all permutations of a given string $\mathbf{a}$. What can one say about the maximal and minimal frequencies in this set, and the permutations under which these extremal frequencies are attained?

References

[1] J. Borwein, A. van der Poorten, J. Shallit, and W. Zudilin, *Neverending fractions*, vol. 23 of *Australian Mathematical Society Lecture Series*, Cambridge University Press, Cambridge, 2014, URL `https://doi.org/10.1017/CBO9780511902659`.

[2] K. Dajani and C. Kraaikamp, *Ergodic theory of numbers*, vol. 29 of *Carus Mathematical Monographs*, Mathematical Association of America, Washington, DC, 2002.

[3] M. Émile Borel, Les probabilités dénombrables et leurs applications arithmétiques, *Rendiconti del Circolo Matematico di Palermo (1884-1940)* **27** (1909)(1), 247–271.

[4] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, Oxford University Press, Oxford, sixth edn., 2008, revised by D. R. Heath-Brown and J. H. Silverman, With a foreword by Andrew Wiles.

[6]Note that, in contrast to the quantity $\nu(\mathbf{a})$ defined in (6.1), which depends only on the *set* (or multi-set) of digits in a string $\mathbf{a}$, whether or not $\mathbf{a}$ is exceptional (i.e., has a nontrivial symmetry) depends on the order of the digits in $\mathbf{a}$ and thus is not invariant with respect to permuting these digits.

[5] M. Iosifescu and C. Kraaikamp, *Metrical theory of continued fractions*, vol. 547 of *Mathematics and its Applications*, Kluwer Academic Publishers, Dordrecht, 2002, URL `https://doi.org/10.1007/978-94-015-9940-5`.

[6] A. Y. Khinchin, *Continued fractions*, University of Chicago Press, Chicago, Ill.-London, 1964.

[7] R. Kuzmin, Sur un problème de Gauss, in *Atti del Congresso Internazionale dei Matematici: Bologna del 3 al 10 de settembre di 1928*, 1929, pp. 83–90.

[8] P. Lévy, Sur les lois de probabilité dont dépendent les quotients complets et incomplets d'une fraction continue, *Bulletin de la Societé Mathématique de France* **57** (1929), 178–194.

[9] I. Niven, *Irrational numbers*, vol. No. 11 of *The Carus Mathematical Monographs*, Mathematical Association of America, 1956.

[10] O. Perron, *Die Lehre von den Kettenbrüchen. Bd I. Elementare Kettenbrüche*, B. G. Teubner Verlagsgesellschaft, Stuttgart, 1954, 3te Aufl.

[11] J. Vandehey, New normality constructions for continued fraction expansions, *J. Number Theory* **166** (2016), 424–451, URL `https://doi.org/10.1016/j.jnt.2016.01.030`.