LAWS OF LARGE NUMBERS AND KHINCHIN'S CONSTANT

ALY SOLIMAN

Abstract. The main goal of this paper is to prove the existence of Khinchin's constant as an application of a variation of the strong law of large numbers. We also give our own concise proof of a weaker variation of the Gauss-Kuzmin Theorem. We assume that the reader is familiar with probability theory and measure theory.

Contents

1. Introduction

The **simple continued fraction** of a real number α is

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \dots}}},$$

where a_0 is an integer and $a_1, a_2, \dots$ are positive integers. These integers are known as the **natural elements** of α. A more concise notation for the continued fraction is $\alpha = [a_0; a_1, a_2, \dots]$. The **metric theory of continued fractions** is generally concerned with using measure theory to classify elements of $\mathbb{R}^+$ whose natural elements satisfy certain properties. Many of the results studied are about the natural elements of almost every real number, that is, all numbers except for a null set with respect to some measure. In this paper, we explore and prove several results in the metric theory of continued fractions. Metrical results can be rephrased as results in probability theory and vice versa. Thus, we can rephrase the metrical results we discuss as probabilistic results and, in fact, apply results in probability theory to prove metrical results on continued fractions. In particular, we can apply a variation of the **strong law of large numbers** to prove a fascinating metrical result: there exists a constant K_0, known as **Khinchin's constant**, such that for almost all real numbers $\alpha = [0; a_1, a_2, \ldots]$, with respect to the Lebesgue measure,

$$\lim_{n\to\infty} (a_1 a_2 \ldots a_n)^{1/n} = K_0. \tag{1.1}$$

In the first section, we prove the variation of the strong law of large numbers. In the two sections after, we explore several results in the metrical theory of continued fractions. In the final section, we tie up all of the previous sections to prove the existence of Khinchin's constant.

2. Laws of Large Numbers

True to the intuitive interpretation of the expected value being the "center" of potential outputs of a random variable, **laws of large numbers** are statements about the arithmetic mean of random variables converging to a limit dependent only on the expected values of the variables. Let $\mu(X)$ and $\sigma^2(X)$ denote the expected value and variance of a random variable X, respectively.

Theorem 2.1 (**The Strong Law of Large Numbers**). *If $(X_k)_{k\geq 1}$ is a collection of independent and identically distributed random variables with expected values $\mu(X_1) = \mu(X_2) = \ldots = \mu < \infty$, then*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \overset{a.s.}{\to} \mu \text{ as } n \to \infty. \tag{2.2}$$

We will discuss a proof for a variation of the above theorem that will eventually prove the existence of Khinchin's constant. First, we introduce the following two theorems. The next theorem is a variation of one direction of the **Borel–Cantelli Lemma**.

Theorem 2.3. *Let $(X_n)_{n\geq 1}$ be a sequence of random variables with sample space Ω and let $(x_n)_{n\geq 1}$ be a sequence of real numbers such that the series*

$$\sum_{n=1}^{\infty} \Pr(X_n \geq x_n)$$

converges. If $A = \{\omega \in \Omega \mid X_n(\omega) \geq x_n \text{ for infinitely many } n\}$ then $\mathrm{P}(A) = 0$ or, equivalently,

$$\Pr(X_n < x_n \text{ for sufficiently large } n) = 1.$$

Proof. Let $A_k = \{\omega \in \Omega \mid X_k(\omega) \geq x_k\}$ and note that, for any integer $m \geq 1$,

$$A \subset \bigcup_{k=m}^{\infty} A_k, \text{ and so } 0 \leq \mathrm{P}(A) \leq \sum_{k=m}^{\infty} \mathrm{P}(A_k) = \sum_{k=m}^{\infty} \Pr(X_k \geq x_k).$$

For any $\epsilon > 0$, there exists some integer $m \geq 1$ such that $\sum_{k=m}^{\infty} \Pr(X_k \geq x_k) < \epsilon$. This implies $0 \leq \mathrm{P}(A) < \epsilon$, and so taking $\epsilon \to \infty$ proves the theorem. □

The proofs for the following theorems can be found in Andrew M Rockett and Peter Szusz's introductory text to continued fractions [1].

Theorem 2.4 (**Chebyshev's inequality**). *Let X be a random variable with finite expected value μ and finite variance $\sigma^2 > 0$. For any real number $h > 0$,*

$$\Pr(|X - \mu| \geq h\sigma) \leq \frac{1}{h^2}. \tag{2.5}$$

For any two functions $f : \mathbb{N} \to \mathbb{R}$ and $g : \mathbb{N} \to \mathbb{R}$, we say that $f(n) = O(g(n))$ if there exists a constant $C > 0$ such that $|f(n)| \leq C|g(n)|$ for all $n \geq 1$.

Theorem 2.6. *Assume that a sequence of random variables $(X_n)_{n\geq 1}$ is such that $\sigma^2\left(\sum_{k=n+1}^{m} X_k\right) = O(m-n)$ and $\mu(X_k) < \infty$ for all $k \geq 1$, It follows that*

$$\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} \mu(X_k) \xrightarrow{a.s.} 0 \text{ as } n \to \infty. \tag{2.7}$$

Proof. Let $\xi > 0$. For any integer $m \geq 1$, let

$$\mu_m = \mu\left(\sum_{k=1}^{m} X_k\right) = \sum_{k=1}^{m} \mu(X_k), \text{ and } \sigma_m^2 = \sigma^2\left(\sum_{k=1}^{m} X_k\right).$$

From the given conditions, $\sigma_m^2 = O(m)$, and so there exists some constant $C > 0$ such that $\sigma_m^2 \leq mC$. Note that $m\xi \geq \sigma_m \xi\sqrt{m}/\sqrt{C}$, and so applying inequality (2.5) for $h = \xi\sqrt{m}/\sqrt{C}$ gives

$$\Pr\left(\left|\sum_{k=1}^{m} X_k - \mu_m\right| \geq m\xi\right) \leq \Pr\left(\left|\sum_{k=1}^{m} X_k - \mu_m\right| \geq \sigma_m\xi\sqrt{m}/\sqrt{C}\right) \leq \frac{C}{m\xi^2}.$$

Therefore, the series

$$0 \leq \sum_{m=1}^{\infty} \Pr\left(\left|\sum_{k=1}^{m^2} X_k - \mu_{m^2}\right| \geq m^2\xi\right) \leq \sum_{m=1}^{\infty} \frac{C}{m^2\xi^2} = \frac{C\pi^2}{6\xi^2}$$

converges. Hence, by Theorem 2.3, the probability that

$$\left|\sum_{k=1}^{m^2} X_k - \mu_{m^2}\right| < m^2\xi, \text{ or, equivalently, } \frac{1}{m^2}\left|\sum_{k=1}^{m^2} X_k - \mu_{m^2}\right| < \xi,$$

for sufficiently large m is 1. However, the above bound works only for square integers. We now look at the "gaps" in between. Assume that an integer $M \geq 1$ is **not** a square integer. This implies $m^2 < M < (m+1)^2$ for some integer $m \geq 1$. Let $\sigma_{M,m}^2 = \sigma^2\left(\sum_{k=m^2+1}^{M} X_k\right)$. Once again, applying (2.5) for $h = m^2\xi/\sigma_{M,m}$ gives

$$\Pr\left(\left|\sum_{k=m^2+1}^{M} X_k - (\mu_M - \mu_{m^2})\right| \geq m^2\xi\right) \leq \frac{\sigma_{M,m}^2}{m^4\xi^2} = \frac{O(M-m^2)}{m^4\xi^2} = O\left(\frac{1}{m^3}\right),$$

since $m^2 < M < (m+1)^2$. The series $\sum_{m=1}^{\infty} 1/m^3$ converges, and so the series

$$\sum_{m=1}^{\infty} \Pr\left(\left|\sum_{k=m^2+1}^{M_m} X_k - (\mu_{M_m} - \mu_{m^2})\right| \geq m^2\xi\right) \tag{2.8}$$

also converges. Note that for every $m \geq 1$ in the above series, we have to **choose** some integer $m^2 < M_m < (m+1)^2$ in order for each term in the series to be well-defined. Although (2.8) converges no matter what the choice function for M_m might be, we will see in a second that it's best if we choose M_m to give the maximum value of

$$\Pr\left(\left|\sum_{k=m^2+1}^{M} X_k - (\mu_{M_m} - \mu_{m^2})\right|\right),$$

among all $m^2 < M < (m+1)^2$. Since (2.8) converges, we can apply Theorem 2.3, and so the probability that

$$\frac{1}{m^2}\left|\sum_{k=m^2+1}^{M_m} X_k - (\mu_{M_m} - \mu_{m^2})\right| < \xi,$$

for sufficiently large m, is 1. Due to our choice of M_m, this statement holds for any $m^2 < M < (m+1)^2$, and not just M_m. Thus, the probability that

$$\frac{1}{M}\left|\sum_{k=1}^{M} X_k - \mu_M\right| \leq \frac{1}{m^2}\left|\sum_{k=1}^{m^2} X_k - \mu_{m^2}\right| + \frac{1}{m^2}\left|\sum_{k=m^2+1}^{M} X_k - (\mu_M - \mu_{m^2})\right| < 2\xi$$

for sufficiently large m or, equivalently, M is 1. Therefore, we got the bound to work for square and non-square numbers. To summarize what was done, for any $\xi > 0$, the probability that $\frac{1}{n}\left|\sum_{k=1}^{n}(X_k - \mu(X_k))\right| < \xi$ for sufficiently large n is 1. Taking $\xi \to 0$ proves the theorem. □

3. Continued Fractions

One of the earliest uses of continued fractions was in in the study of **Diophantine Approximation**, which is concerned with finding "good" rational approximations of real numbers.

Definition 3.1. A **good rational approximation** of a real number α is a rational p/q such that $|q\alpha - p| < |b\alpha - a|$ for any rational a/b such that $0 < b < q$. Intuitively, it is the "closest" rational to α among all rationals with smaller denominator.

It turns out that we can find good rational approximations for any real number by truncating its simple continued fraction.

Definition 3.2. For an integer $n \geq 0$, assume that a real number α has at least $n+1$ natural elements and denote them as $a_0, a_1, \ldots a_n$. The rational numbers

$$c_k = [a_0; a_1, \ldots, a_k],$$

for $k = 0, \ldots, n$ are called the **convergents** of α. Furthermore, let the integers p_n and $q_n > 0$ denote the co-prime integers such that $c_n = p_n/q_n$.

Remark 3.3. If α is a rational number with $n+1$ natural elements then $c_n = \alpha$. However, if α is irrational then c_n exists for all $n \geq 0$. Furthermore, a real number has only finitely many natural elements if it is rational and infinitely many if it is irrational. Therefore, from now on, we can assume that an arbitrary real number has infinitely many natural elements since then our metrical results would be excluding the rationals - a null set with respect to the Lebesgue measure. We also restrict the metrical result to numbers in $[0, 1]$ since the first element a_0 doesn't really affect the results and can be set to zero.

It can be shown that the convergents of a real number are good approximations of it, but we are only stating this to provide some motivation behind the definition of convergents and will not be proving it. As we will see, investigating the convergents can be fruitful in the metric theory of continued fractions. First, let us go through some properties of the convergents of an arbitrary real number. All of the results in this section and their proofs can be found in Khinchin's introductory text to continued fractions [2]. The following theorem shows that the values a_n, c_{n-1}, and c_{n-2} are sufficient to compute c_n.

Theorem 3.4. *The sequences $(p_n)_{n\geq 0}$ and $(q_n)_{n\geq 0}$ satisfy the recurrence relation*

$$u_n = u_{n-1}a_n + u_{n-2}, \tag{3.5}$$

for any $n \geq 2$.

Proof. Let $P(k)$ be the proposition that, for any real number with at least k natural elements, relation (3.5) holds for all p_n and q_n with $3 \leq n < k$. Assume that $P(k)$ is true for some $k > 3$. Let α be a real number with at least $k+1$ natural elements and let us consider the rational number $r = [a_1; a_2, \ldots, a_k]$. Let $c'_i = p'_i/q'_i$ denote the convergents of r for $i = 0, \ldots, k-1$. Note that $r = p'_{k-1}/q'_{k-1}$, since the first natural element of r is a_1 not a_0. Furthermore,

$$\frac{p_k}{q_k} = a_0 + \frac{1}{r} = a_0 + \frac{q'_{k-1}}{p'_{k-1}} = \frac{a_0 p'_{k-1} + q'_{k-1}}{p'_{k-1}},$$

and, similarly,

$$\frac{p_{k-1}}{q_{k-1}} = a_0 + \frac{q'_{k-2}}{p'_{k-2}} = \frac{a_0 p'_{k-2} + q'_{k-2}}{p'_{k-2}}.$$

The same patterns holds for p_{k-2}/q_{k-2}. Thus, by $P(k)$,

$$\begin{aligned} p_k &= a_0(p'_{k-2}a_k + p'_{k-3}) + (q'_{k-2}a_k + q'_{k-3}) \\ &= a_k(a_0 p'_{k-2} + q'_{k-2}) + (a_0 p'_{k-3} + q'_{k-3}) \\ &= a_k p_{k-1} + p_{k-2}, \end{aligned}$$

and $q_k = a_k p'_{k-2} + p'_{k-3} = a_k q_{k-1} + q_{k-2}$. Therefore, $P(k)$ implies $P(k+1)$. We can tediously verify that $P(2)$ and $P(3)$ are true. Hence, by strong induction, $P(k)$ is true for all $k \geq 2$. ☐

Remark 3.6. Setting $p_{-1} = 1$ and $q_{-1} = 0$ for any real number extends the above theorem for $k \geq 1$. Furthermore, note that we can rewrite a real number $\alpha = [a_0; a_1, \ldots, a_n, \ldots]$ as

$$\alpha = [a_0; a_1, \ldots, a_{n-1}, r_n]$$

where $n \geq 1$ and $r_n = [a_n; a_{n+1}, \ldots]$. Thus, by replacing a_n with r_n in relation (3.5), the above proof also gives

$$\alpha = \frac{p_{n-1}r_n + p_{n-2}}{q_{n-1}r_n + q_{n-2}}. \tag{3.7}$$

Theorem 3.8. *We have $q_n p_{n-1} - q_{n-1}p_n = (-1)^n$ for any integer $n \geq 1$.*

Proof. Let $P(n)$ be the proposition that $q_n p_{n-1} - q_{n-1}p_n = (-1)^n$ for $n \geq 1$. Note that $p_0 = a_0$ and $q_0 = 1$ since $c_0 = a_0$. Furthermore, $p_1 = a_0 a_1 + 1$ and $q_1 = a_1$ since $c_1 = a_0 + 1/a_1$. Thus, $q_1 p_0 - q_0 p_1 = a_0 a_1 - (a_0 a_1 + 1) = -1$, and so $P(1)$ holds. Multiplying (3.5) for $u_n = p_n$ by q_{n-1} gives

$$p_n q_{n-1} = p_{n-1}q_{n-1}a_n + p_{n-2}q_{n-1},$$

and multiplying it for $u_n = q_n$ by p_{n-1} gives

$$q_n p_{n-1} = q_{n-1}p_{n-1}a_n + q_{n-2}p_{n-1}.$$

Subtracting the above two equations gives $p_n q_{n-1} - q_{n-1}p_n = q_{n-2}p_{n-1} - p_{n-1}q_{n-2}$, and so $P(n-1)$ implies $P(n)$ for any $n \geq 2$. Thus, by induction, $P(n)$ is true for all $n \geq 1$. ☐

We are now ready to discuss some results in the metric theory of continued fractions. Let λ denote the Lebesgue measure.

Definition 3.9. For any integer $n \geq 1$, let $\mathscr{S}_n = (b_i)_{i=1}^n$ denote a sequence of positive integers of length n, and let Q_n denote the set of all possible $\mathscr{S}_n$.

Definition 3.10. For positive integers n and b, let $S(n, b)$ denote the set of real numbers $\alpha = [0; a_1, \ldots, a_n, \ldots]$ such that $a_n = b$. Furthermore, for positive integers $n_1 < n_2$, b_1, and b_2, let $S(n_1, b_1, n_2, b_2)$ denote the set of real numbers $\alpha = [0; a_1, \ldots, a_{n_1}, \ldots, a_{n_2}, \ldots]$ such that $a_{n_1} = b_1$ and $a_{n_2} = b_2$.

Example 3.11. The set $S(3, 7)$ is the set of real numbers in $[0, 1]$ with natural element $a_3 = 7$, and the set $S(3, 7, 6, 2)$ is the set of real numbers in $[0, 1]$ with natural elements $a_3 = 7$ and $a_6 = 2$.

Remark 3.12. Taking a probability space with sample space $\Omega = [0, 1]$ and probability measure being the Lebesgue measure, the above definitions can be redefined probabilistically. In particular, defining the discrete random variable $X_k : \Omega \to \mathbb{N}$ as $X_k(\alpha) = a_k$, where a_k is the k^{th} natural element of α, it follows that

$$S(n, b) = \Pr(X_n = b), \text{ and } S(n_1, b_1, n_2, b_2) = \Pr(X_{n_1} = b_1 \text{ and } X_{n_2} = b_2).$$

Definition 3.13. For any sequence $\mathscr{S}_n = (b_i)_{i=1}^n$ in Q_n, let $J(\mathscr{S}_n)$ denote the set of real numbers $\alpha = [0; a_1, a_2, \ldots, a_n, \ldots]$ such that $a_i = b_i$ for all $i = 1, 2, \ldots, n$.

Example 3.14. For $\mathscr{S}_4 = (1, 3, 7, 5)$, $J(\mathscr{S}_4)$ is the set of all real numbers in $[0, 1]$ with natural elements $a_1 = 1, a_2 = 3, a_3 = 7$, and $a_4 = 5$.

Proposition 3.15. *For any sequence $\mathscr{S}_n = (b_i)_{i=1}^n$, let $c'_{n-1} = p_{n-1}/q_{n-1}$ and $c'_n = p_n/q_n$ denote the $n-1$ and n convergents of the rational number*

$$[0; b_1, \ldots, b_n] = \cfrac{1}{b_1 + \cfrac{1}{b_2 + \ldots + \cfrac{1}{b_n}}} = c'_n,$$

respectively. It follows that

$$J(\mathscr{S}_n) = \left(\frac{p_n + p_{n-1}}{q_n + q_{n-1}}, \frac{p_n}{q_n} \right). \tag{3.16}$$

We say that $J(\mathscr{S}_n)$ is an interval of order n.

Proof. Let $\alpha = [0; a_1, a_2, \ldots, a_n, \ldots]$ be a number in $J(\mathscr{S}_n)$ and let $c_0, \ldots, c_n$ denote the first $n+1$ convergents of α. Since $c_{n-1} = p_{n-1}/q_{n-1}$ and $c_n = p_n/q_n$, (3.7) implies

$$\alpha = \frac{p_n r_{n+1} + p_{n-1}}{q_n r_{n+1} + q_{n-1}} = [0, a_1, \ldots, a_n + 1/r_{n+1}],$$

where $r_k = [a_k; a_{k+1}, \ldots]$. Note that $1 \leq r_k < \infty$ since $a_k \geq 1$ for any $k \geq 1$. Therefore, starting at $r_{n+1} = 1$ and taking $r_{n+1} \to \infty$ gives the end points

$$\frac{p_n + p_{n-1}}{q_n + q_{n-1}}, \text{ and } \lim_{r_{n+1} \to \infty} [0; a_1, \ldots, a_n + 1/r_{n+1}] = \frac{p_n}{q_n},$$

for any α in $J(\mathscr{S}_n)$. Hence, $J(\mathscr{S}_n)$ is the interval with such end points. $\square$

Remark 3.17. The above proposition implies

$$\lambda(J(\mathscr{S}_n)) = \left|\frac{p_n}{q_n} - \frac{p_n + p_{n-1}}{q_n + q_{n-1}}\right|. \tag{3.18}$$

Furthermore, for any $n \geq 1$, the disjoint intervals of order n partition the interval $[0, 1]$.

Theorem 3.19. *For any positive integers n and b,*

$$\frac{1}{3b^2} < \lambda\left(S(n+1, b)\right) < \frac{2}{b^2}. \tag{3.20}$$

Proof. Fix a sequence $\mathscr{S}_n = (b_i)_{i=1}^n$ in Q_n, and let $\mathscr{S}^*_{n+1} = (b^*_i)_{i=1}^{n+1}$ be such that $b^*_i = b_i$ for $1 \leq i \leq n$ and $b^*_{n+1} = b$. By roughly the same argument in proposition 3.15,

$$\lambda(J(\mathscr{S}^*_{n+1})) = \left|\frac{p_n b + p_{n-1}}{q_n b + p_{n-1}} - \frac{p_n(b+1) + p_{n-1}}{q_n(b+1) + q_{n-1}}\right|, \tag{3.21}$$

where $c'_{n-1} = p_{n-1}/q_{n-1}$ and $c'_n = p_n/q_n$ are the convergents of $[0; b_1, \ldots, b_n]$. By Theorem 3.8,

$$\lambda(J(\mathscr{S}_n)) = \left|\frac{p_n}{q_n} - \frac{p_n + p_{n-1}}{q_n + q_{n-1}}\right| = \left|\frac{(-1)^n}{q_n(q_n + q_{n-1})}\right| = \frac{1}{q_n^2\left(1 + \frac{q_{n-1}}{q_n}\right)}.$$

Similarly, we can rewrite (3.21) as

$$\begin{aligned}\lambda(J(\mathscr{S}^*_{n+1})) &= \left|\frac{(-1)^n}{(q_n b + p_{n-1})(q_n(b+1) + q_{n-1})}\right| \\ &= \frac{1}{q_n^2 b^2\left(1 + \frac{q_{n-1}}{bq_n}\right)\left(1 + \frac{1}{b} + \frac{q_{n-1}}{bq_n}\right)}.\end{aligned}$$

Therefore,

$$\frac{\lambda(J(\mathscr{S}^*_{n+1}))}{\lambda(J(\mathscr{S}_n))} = \frac{1}{b^2} \cdot \frac{\left(1 + \frac{q_{n-1}}{q_n}\right)}{\left(1 + \frac{q_{n-1}}{bq_n}\right)\left(1 + \frac{1}{b} + \frac{q_{n-1}}{bq_n}\right)}. \tag{3.22}$$

Noting by Theorem 3.4 that $q_{n-1} < q_n$, we can bound the factor on the right side in the previous equation from above:

$$\frac{\left(1 + \frac{q_{n-1}}{q_n}\right)}{\left(1 + \frac{q_{n-1}}{bq_n}\right)\left(1 + \frac{1}{b} + \frac{q_{n-1}}{bq_n}\right)} < 1 + \frac{q_{n-1}}{q_n} < 2.$$

Furthermore, since $1 + q_{n-1}/q_n \geq 1 + q_{n-1}/bq_n$ and $(1 + 1/b + q_{n-1}/bq_n) < 3$, we can also bound the factor from below:

$$\frac{1}{3} < \frac{1}{1 + \frac{1}{b} + \frac{q_{n-1}}{bq_n}} < \frac{1 + \frac{q_{n-1}}{q_n}}{(1 + q_{n-1}/bq_n)\left(1 + \frac{1}{b} + \frac{q_{n-1}}{bq_n}\right)}.$$

Thus, we have

$$\frac{1}{3b^2}\lambda(J(\mathscr{S}_n)) < \lambda(J(\mathscr{S}^*_{n+1})) < \frac{2}{b^2}\lambda(J(\mathscr{S}_n)).$$

Since the intervals $J(\mathscr{S}_n)$ partition $[0,1]$,

$$\sum_{\mathscr{S}_n \in Q_n} \lambda(J(\mathscr{S}_n)) = 1, \text{ and so } \frac{1}{3b^2} < \sum_{\mathscr{S}_n \in Q_n} \lambda(J(\mathscr{S}^*_{n+1})) < \frac{2}{b^2}.$$

The summation on the right over all $\mathscr{S}_n$in Q_n gives the measure of the set of real numbers in $[0,1]$ such that $a_{n+1} = b^*_{n+1} = b$. In other words,

$$\sum_{\mathscr{S}_n \in Q_n} \lambda(J(\mathscr{S}^*_{n+1})) = \lambda\left(S\left(n+1, b\right)\right),$$

and so the proof is complete. □

Example 3.23. The measure of the set of numbers in $[0,1]$ such that $a_7 = 2$ is

$$\frac{1}{12} < \lambda\left(S\left(7,2\right)\right) < \frac{1}{2}.$$

In the next section, we will explore more estimates on the measure of S.

4. The Gauss-Kuzmin Theorem

The **Gauss–Kuzmin Theorem** was first formulated by Gauss in 1800 and it may have been the first theorem in the metric theory of continued fractions. For a real number $\alpha = [0; a_1, a_2, \ldots]$ and integer $n \geq 0$, let

$$z_n(\alpha) = [0; a_{n+1}, a_{n+2}, \ldots]. \tag{4.1}$$

Note that $0 \leq z_n(\alpha) < 1$ for any $n \geq 0$. For any $0 \leq s \leq 1$, let $\lambda_n(s)$ denote the Lebesgue measure of the set of real numbers α in the unit interval $[0,1]$ for which $z_n(\alpha) < s$. Gauss noted in one of his diaries that "by a very simple argument",

$$\lim_{n\to\infty} \lambda_n(s) = \frac{\ln\left(1+s\right)}{\ln 2}, \text{ for all } 0 \leq s \leq 1. \tag{4.2}$$

This limit became known as the Gauss–Kuzmin theorem, although Gauss's original proof was never found. An actual proof for the limit was not discovered until 1928 by Kuzmin [3], who showed that

$$\lambda_n(s) = \frac{\ln\left(1+s\right)}{\ln 2} + O\left(e^{-c\sqrt{n}}\right), \tag{4.3}$$

for some constant $c > 0$ that only depends on s. This was an improvement upon Gauss's result since he was not able to bound the rate of convergence of limit (4.2). Let us define the function $\phi : [0,1] \to [0,1]$ as

$$\phi(s) := \lim_{n\to\infty} \lambda_n(s). \tag{4.4}$$

We will present a proof that if ϕ is in $C^1([0,1])$ (continuously differentiable) then $\phi(s) = \ln(1+s)/\ln 2$ - a weaker variation of the theorem. Let's first look at a recurrence relation satisfied by $\lambda_n(s)$. For any real number $\alpha = [0; a_1, a_2, \ldots]$, we have

$$z_n(\alpha) = \frac{1}{a_{n+1} + z_{n+1}(\alpha)}.$$

This means that, for a fixed α and $0 \leq s \leq 1$, we get $z_{n+1}(\alpha) < s$ if, and only if,

$$\frac{1}{a_{n+1} + s} < z_n(\alpha) \leq \frac{1}{a_{n+1}}.$$

Thus, the set of points for which $z_{n+1}(\alpha) < s$ is precisely the set of points such that $1/(k+s) < z(\alpha, n) \leq 1/k$ for some integer $k \geq 1$, and so we get the recurrence relation

$$\lambda_{n+1}(s) = \sum_{k=1}^{\infty} \left(\lambda_n \left(\frac{1}{k} \right) - \lambda_n \left(\frac{1}{k+s} \right) \right). \tag{4.5}$$

This relation was probably known to Gauss and it has been used in most of the proofs for the Gauss-Kuzmin theorem. We can quickly verify that the function $f(s) = c \ln(1+s)$, for some constant c, satisfies the equation

$$f(s) = \sum_{k=1}^{\infty} \left(f \left(\frac{1}{k} \right) - f \left(\frac{1}{k+s} \right) \right). \tag{4.6}$$

We now show that if $f : [0,1] \to \mathbb{R}$ is in $C^1([0,1])$ and satisfies relation (4.6), then $f(s) = c \ln(1+s)$ for some constant c.

Lemma 4.7. *Let $I \subset \mathbb{R}$ be a compact set and let $(g_k)_{k \geq 1}$ be a sequence of functions from I to I converging point-wise to a constant function. Furthermore, let $(\eta_k)_{k \geq 1}$ be a sequence of functions from I to $(0, \infty)$ such that $\sum_{k=1}^{\infty} \eta_k(s) = 1$ for all s in I. A function $R : I \to \mathbb{R}$ is a constant function if, and only if, it is a continuous function such that*

$$R(s) = \sum_{k=1}^{\infty} \eta_k(s) R(g_k(s)). \tag{4.8}$$

Proof. If R is a constant function then (4.8) follows since $\sum_{k=1}^{\infty} \eta_k(s) = 1$. Now assume that R is continuous and satisfies (4.8). There are points p and q in I such that $R(p)$ and $R(q)$ are the maximum and minimum of $R(I)$, respectively. If there are $k \geq 1$ such that $R(p) > R(g_k(p))$ then

$$R(p) = \sum_{k=1}^{\infty} \eta_k(p) R(g_k(p)) < \sum_{k=1}^{\infty} \eta_k(p) R(p) = R(p).$$

This is a contradiction, and so we must have $R(p) = R(g_k(p))$ for all $k \geq 1$. This implies $R(p) = R(w)$, where $w = \lim_{k \to \infty} g_k(p)$. The same argument can be made to show that $R(q) = R(w)$ since we also have $w = \lim_{k \to \infty} g_k(q)$. Thus, $R(s) = R(w)$ for all s in I, and so R is a constant function. □

Theorem 4.9. *If $f : [0,1] \to \mathbb{R}$ is in $C^1([0,1])$ and satisfies relation (4.6) then $f(s) = c \ln(1+s)$ for constant c.*

Proof. Differentiating (4.6) yields

$$f'(s) = \sum_{k=1}^{\infty} \frac{1}{(k+s)^2} f' \left(\frac{1}{k+s} \right).$$

Thus, if we define the function $R : [0,1] \to \mathbb{R}$ as $R(s) = (1+s) f'(s)$ then

$$\begin{aligned} R(s) &= \sum_{k=1}^{\infty} \frac{1+s}{(k+s)^2} \cdot \frac{k+s}{k+s+1} \cdot \left(1 + \frac{1}{k+s+1} \right) R \left(\frac{1}{k+s} \right) \\ &= \sum_{k=1}^{\infty} \frac{1+s}{(k+s)(k+s+1)} R \left(\frac{1}{k+s} \right). \end{aligned}$$

Let $\eta_k(s) = \dfrac{1+s}{(k+s)(k+s+1)}$ and and $g_k(s) = \dfrac{1}{k+s}$ for any s in $[0,1]$ and integer $k \geq 1$. It follows that $\eta_k(s) > 0$ and

$$\sum_{k=1}^{\infty} \eta_k(s) = \sum_{k=1}^{\infty} \frac{1+s}{(k+s)(k+s+1)} = (1+s)\sum_{k=1}^{\infty}\left(\frac{1}{k+s} - \frac{1}{k+s+1}\right) = \frac{1+s}{1+s} = 1,$$

for any $0 \leq s \leq 1$. Furthermore, the sequence of functions $(g_k(s))_{k\geq 1}$ converges point-wise to the constant function $g(s) = 0$. Hence, by Lemma 4.7, R is a constant function. In other words, we have $f'(s) = c/(1+s)$ for some constant c, and so $f(s) = c\ln(1+s) + h$ for some constant h. Since $f(0) = 0$, we have $h = 0$. □

The above theorem allows us to prove a weaker variation of the Gauss-Kuzmin theorem.

Theorem 4.10. *If $\phi(s)$ exists for every $s \in [0,1]$ and the function ϕ is in $C^1([0,1])$ then $\phi(s) = \ln(1+s)/\ln 2$.*

Proof. Fix a point s in $[0,1]$. Taking the limit on both sides in (4.5) as $n \to \infty$ yields

$$\phi(s) = \sum_{k=1}^{\infty}\left(\phi\left(\frac{1}{k}\right) - \phi\left(\frac{1}{k+s}\right)\right),$$

and so, by Theorem 4.9, we get $\phi(s) = c\ln(1+s)$ where $c = \phi(1)/\ln 2 = 1/\ln 2$. □

Thus, proving from properties of $\lambda_n(s)$ that $\phi(s)$ is a function in $C^1([0,1])$ can potentially produce a simple and concise proof for 4.2, but we were not successful at proving this condition independently. Let us now discuss a well-known proof for 4.2. The proof for the following theorem, and the rest of the proofs after, can be found in Andrew M Rockett and Peter Szusz's text.

Theorem 4.11. *Let $(f_n)_{n\geq 0}$ be a sequence of functions from $[0,1]$ to $\mathbb{R}$ such that f_0 is twice differentiable with $f_0(0) = 0$ and $f_0(1) = 1$, and*

$$f_{n+1}(s) = \sum_{k=1}^{\infty}\left(f_n\left(\frac{1}{k}\right) - f_n\left(\frac{1}{k+s}\right)\right). \tag{4.12}$$

It follows that

$$f_n(s) = \frac{\ln(1+s)}{\ln 2} + O(\theta^n),$$

*for some constant $0 < \theta < 0.76$ that **doesn't** depend on the sequence $(f_n)_{n\geq 0}$.*

Remark 4.13. From now on we will denote by θ the constant mentioned in the above theorem.

The above theorem can be applied to $\lambda_n(s)$ since it satisfies relation (4.5) and the function $\lambda_0(s) = s$ is twice differentiable. Thus,

$$\lambda_n(s) = \frac{\ln(1+s)}{\ln 2} + O(\theta^n). \tag{4.14}$$

Remark 4.15. Let b be a positive integer. For a number $\alpha = [0, a_1, a_2, \ldots, a_n, \ldots]$, we have $0 \leq z_n(\alpha) < 1$, and so $a_n = b$ if, and only if,

$$\frac{1}{b+1} < z_{n-1}(\alpha) = \frac{1}{a_n + z_n(\alpha)} \leq \frac{1}{b}. \tag{4.16}$$

Therefore, the Gauss-Kuzmin Theorem readily gives us a good estimate on the measure in 3.19. In particular, (4.16) implies

$$\lambda\left(S\left(n,b\right)\right)=\lambda_{n-1}\left(\frac{1}{b}\right)-\lambda_{n-1}\left(\frac{1}{b+1}\right),$$

and so (4.14) gives

$$\begin{aligned}\lambda\left(S\left(n,b\right)\right)&=\frac{1}{\ln 2}\ln\left(1+\frac{1}{b}\right)-\frac{1}{\ln 2}\ln\left(1+\frac{1}{b+1}\right)+O(\theta^{n})\\&=\frac{1}{\ln 2}\ln\left(1+\frac{1}{b^{2}+2b}\right)+O(\theta^{n}).\end{aligned}\tag{4.17}$$

In other words, considering remark 3.12, the probability $\Pr(X_n=b)$ becomes less dependent on n as $n\to\infty$. The generality of Theorem 4.11 also allows us to prove the next result.

Theorem 4.18. *For any positive integers $n_2>n_1$ and b_1,b_2, we have*

$$|\lambda(S(n_1,b_1,n_2,b_2))-\lambda(S(n_1,b_1))\lambda(S(n_2,b_2))|=O\left(\theta^{n_2-n_1}\right)\lambda(S(n_1,b_1))\lambda(S(n_2,b_2)).$$

Proof. Let $k=n_2-n_1$ and $\mathscr{S}_{n_1}=(s_i)_{i=1}^{n_1}$ be a sequence from Q_{n_1}. For the interval $J(\mathscr{S}_{n_1})$ of order n_1, let $\lambda_k(\mathscr{S}_{n_1},s)$ denote the Lebesgue measure of the set of numbers α in $J(\mathscr{S}_n)$ such that $z_{n_1+k}(\alpha)<s$. We can quickly verify that

$$\lambda_{k+1}(\mathscr{S}_{n_1},s)=\sum_{i=1}^{\infty}\left(\lambda_k\left(\mathscr{S}_{n_1},\frac{1}{i}\right)-\lambda_{k+1}\left(\mathscr{S}_{n_1},\frac{1}{i+s}\right)\right),\tag{4.19}$$

following the same argument used with $\lambda_k(s)$. The function

$$f_k\left(\mathscr{S}_n,s\right)=\frac{\lambda_k(\mathscr{S}_n,s)}{\lambda(J(\mathscr{S}_n))}$$

satisfies a relation like (4.19) since it is just a constant multiple of $\lambda_k(\mathscr{S}_n,s)$. Note that $f_0\left(\mathscr{S}_n,0\right)=0$ and $f_0\left(\mathscr{S}_n,1\right)=1$. Now let's take a closer look at the function $f_0\left(\mathscr{S}_n,s\right)$. It represents the fraction of numbers in $J(\mathscr{S}_{n_1})$ such that $z_{n_1}(\alpha)<s$. Any number α in $J(\mathscr{S}_{n_1})$ can be written as

$$\alpha=[0;s_1,\ldots,s_{n_1}+z_{n_1}(\alpha)]=\cfrac{1}{s_1+\cfrac{1}{\ldots+\cfrac{1}{s_{n_1}+\cfrac{1}{(1/z_{n_1}(\alpha))}}}}.$$

Thus, ranging $z_{n_1}(\alpha)$ from 0 to s gives the end points

$$\frac{p_{n_1}}{q_{n_1}},\text{ and }\frac{p_{n_1}(1/s)+p_{n_1-1}}{q_{n_1}(1/s)+q_{n_1-1}}=\frac{p_{n_1}+sp_{n_1-1}}{q_{n_1}+sq_{n_1-1}},$$

where $c_{n_1-1}=p_{n_1-1}/q_{n_1-1}$ and $c_{n_1}=p_{n_1}/q_{n_1}$ are the n_1-1 and n_1 convergents of $[0;s_1,\ldots,s_{n_1}]$, respectively. The function $\lambda_0(\mathscr{S}_{n_1},s)$ can then be written as

$$\lambda_0(\mathscr{S}_{n_1},s)=\left|\frac{p_{n_1}}{q_{n_1}}-\frac{p_{n_1}+sp_{n_1-1}}{q_{n_1}+sq_{n_1-1}}\right|=\left|\frac{(-1)^{n_1}s}{q_{n_1}(q_{n_1}+sq_{n_1})}\right|=\frac{s}{q_{n_1}(q_{n_1}+sq_{n_1})},$$

which is twice differentiable with respect to s. Since $f_0(\mathscr{S}_n, s)$ is just a constant multiple of $\lambda_0(\mathscr{S}_{n_1}, s)$, it is also twice differentiable. Therefore, we can apply Theorem 4.11 to get

$$f_k(\mathscr{S}_{n_1}, s) = \frac{\ln(1+s)}{\ln 2} + O\left(\theta^k\right),$$

and so

$$\lambda_k(\mathscr{S}_{n_1}, s) = \lambda(J(\mathscr{S}_{n_1}))\frac{\ln(1+s)}{\ln 2} + \lambda(J(\mathscr{S}_n))O\left(\theta^k\right). \tag{4.20}$$

Let $S_k(\mathscr{S}_{n_1}, b_2)$ denote the set of numbers in the interval $J(\mathscr{S}_{n_1})$ with natural element $a_{n_1+k} = b_2$. Similar to (4.16), a number α would have $a_{n_1+k} = b_2$ if, and only if,

$$\frac{1}{b_2+1} < z_{n_1+k-1}(\alpha) = \frac{1}{a_{n_1+k} + z_{n_1+k}(\alpha)} \leq \frac{1}{b_2}.$$

This implies

$$\lambda(S_k(\mathscr{S}_{n_1}, b_2)) = \lambda_{k-1}\left(\mathscr{S}_{n_1}, \frac{1}{b_2}\right) - \lambda_{k-1}\left(\mathscr{S}_{n_1}, \frac{1}{b_2+1}\right).$$

Therefore, by roughly the same argument for (4.17), we can use (4.20) to get

$$\lambda\left(S_k(\mathscr{S}_{n_1}, b_2)\right) = \frac{\lambda(J(\mathscr{S}_n))}{\ln 2}\ln\left(1 + \frac{1}{b_2^2 + 2b_2}\right) + \lambda(J(\mathscr{S}_n))O\left(\theta^k\right).$$

Let Q denote the set of sequences $\mathscr{S}_{n_1} = (s_i)_{i=1}^{n_1}$ in Q_{n_1} such that $s_{n_1} = b_1$. Since the intervals of order n_1 partition the interval $[0,1]$, it follows that

$$\begin{aligned}
\lambda\left(S\left(n_1, b_1, n_2, b_2\right)\right) &= \sum_{\mathscr{S}_{n_1} \in Q} \lambda\left(S_k(\mathscr{S}_{n_1}, b_2)\right) \\
&= \frac{1}{\ln 2}\ln\left(1 + \frac{1}{b_2^2 + 2b_2}\right)\sum_{\mathscr{S}_{n_1} \in Q}\lambda(J(\mathscr{S}_{n_1})) + O\left(\theta^k\right)\sum_{\mathscr{S}_{n_1} \in Q}\lambda(J(\mathscr{S}_{n_1})) \\
&= \frac{1}{\ln 2}\ln\left(1 + \frac{1}{b_2^2 + 2b_2}\right)\lambda\left(S\left(n_1, b_1\right)\right) + O\left(\theta^k\right). \qquad (4.21)
\end{aligned}$$

Combining (4.17) and (4.21) proves the theorem with a few steps using the triangle-inequality. □

In light of remark 3.12, the above theorem can be rephrased as

$$\Pr(X_{n_1} = b_1 \text{ and } X_{n_2} = b_2) = \Pr(X_{n_1} = b_1)\Pr(X_{n_2} = b_2)(1 + O(\theta^{n_2 - n_1})),$$

which means that the random variables X_{n_1} and X_{n_2} are "weakly" independent. Intuitively, this is the key behind meeting the conditions for Theorem 2.6 - a result that's almost like Theorem 2.1 for identically distributed and independent variables.

5. Khinchin's Constant

This section is a combination of Khinchin's text, and Andrew M Rockett and Peter Szusz's text. In particular, we present Khinchin's proofs for the results but frame the results from a probabilistic perspective as done by Andrew M Rockett and Peter Szusz. Let us define a probability space with sample space $\Omega = [0,1]$ and probability measure being the Lebesgue measure. The event space is the set of Lebesgue measurable subsets of $[0,1]$. Let $f : \mathbb{N} \to [0, \infty)$ be a function such that $f(n) = O\left(n^\delta\right)$ for some constant $0 < \delta < 1/2$. Fix an integer $k \geq 1$ and let us define the random variable $X_k : \Omega \to \mathbb{R}$ as $X_k(\alpha) := f(a_k)$, where a_k is the k^{th} natural element of α. Note that, for any $k \geq 1$, $X_k(\Omega)$ is countable since $a_k \in \mathbb{Z}$. Therefore, the random variables $X_1, X_2, \ldots$ are discrete.

Proposition 5.1. *For a random variable X_k,*

$$\mu(X_k) = \sum_{r=1}^{\infty} f(r)\lambda\left(S\left(k,r\right)\right), \tag{5.2}$$

$$\sigma^2(X_k) = \sum_{r=1}^{\infty} f(r)^2\lambda\left(S\left(k,r\right)\right) - \mu(X_k)^2. \tag{5.3}$$

Proof. Note that $X(\Omega) = \{f(1), f(2), f(3), \ldots\}$. Since X_k is discrete, its expected value is given as

$$\mu(X_k) = \sum_{r=1}^{\infty} f(r)\Pr(X_k = f(r)).$$

Therefore,

$$\mu(X_k) = \sum_{r=1}^{\infty} f(r)\lambda(\{\omega \in [0,1] \mid f(a_k) = f(r)\}) = \sum_{r=1}^{\infty} f(r)\lambda(\{\alpha \in [0,1] \mid a_k = r\}).$$

This proves (5.2) since $S(k,r) = \{\alpha \in [0,1] \mid a_k = r\}$. A similar argument can be for the variance. $\square$

Theorem 5.4. *For any $k \geq 1$, $\mu(X_k) = O(1)$ and $\sigma^2(X_k) = O(1)$.*

Proof. By (3.20) and (5.2),

$$\mu(X_k) = \sum_{r=1}^{\infty} f(r)\lambda\left(S\left(k,r\right)\right) = \sum_{r=1}^{\infty} O(r^{\delta-2}) < \infty.$$

Furthermore, since the series on the right doesn't depend on k, $\mu(X_k)$ has an upper bound independent of k. Once again, (3.20) gives

$$\sum_{r=1}^{\infty} f(r)^2\lambda\left(S\left(k,r\right)\right) = \sum_{r=1}^{\infty} O(r^{2\delta-2}) < \infty,$$

since $2\delta - 2 < -1$. Therefore, $\sigma^2(X_k)$ is finite for any $k \geq 1$. Since the series on the right in the above equation doesn't depend on k, $\sigma^2(X_k)$ has an upper bound independent of k. $\square$

Note that, by (4.17) and (5.2),

$$\lim_{k\to\infty} \mu(X_k) = \frac{1}{\ln 2}\sum_{r=1}^{\infty} f(k)\ln\left(1 + \frac{1}{r^2+2r}\right).$$

The Cesàro mean of $(\mu(X_k))_{k\geq 1}$ converges to the same limit as $(\mu(X_k))_{k\geq 1}$, and so

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mu(X_k) = \frac{1}{\ln 2}\sum_{k=1}^{\infty} f(k)\ln\left(1 + \frac{1}{k^2+2k}\right). \tag{5.5}$$

For positive integers $k > i$, let

$$\begin{aligned} g_{i,k} &:= \sum_{r_1, r_2 \geq 1} f(r_1) f(r_2) \lambda\left(S\left(i, r_1, k, r_2\right)\right) - \mu(X_i)\mu(X_k) \\ &= \int_0^1 f(a_i) f(a_k) d\lambda - \mu(X_i)\mu(X_k). \end{aligned}$$

Lemma 5.6. *We have $g_{i,k} = O(\theta^{k-i})$.*

Proof. By Theorem 4.18,

$$\begin{aligned} &\left| g_{i,k} - \sum_{r_1, r_2 \geq 1} f(r_1) f(r_2) \lambda\left(S\left(i, r_1\right)\right) \lambda\left(S\left(k, r_2\right)\right) + \mu(X_i)\mu(X_k) \right| \\ &= O\left(\theta^{k-i}\right) \sum_{r_1, r_2 \geq 1} f(r_1) f(r_2) \lambda\left(S\left(i, r_1\right)\right) \lambda\left(S\left(k, r_2\right)\right). \end{aligned} \tag{5.7}$$

Note that

$$\mu(X_i)\mu(X_k) = \sum_{r_1, r_2 \geq 1} f(r_1) f(r_2) \lambda\left(S\left(i, r_1\right)\right) \lambda\left(S\left(k, r_2\right)\right),$$

and so $|g_{i,k}| = O(\theta^{k-i})\mu(X_i)\mu(X_k) = O(\theta^{k-i})$, since $\mu(X_i)\mu(X_k) = O(1)$, by Theorem 5.4. □

Theorem 5.8. *We have $\sigma^2\left(\sum_{k=n+1}^m X_k\right) = O(m-n)$.*

Proof. Note that

$$\sigma^2\left(\sum_{k=n+1}^m X_k\right) = \int_0^1 \left(\sum_{k=n+1}^m \left(X_k - \mu(X_k)\right)\right)^2 d\lambda.$$

We can expand the square of the sum to get

$$\begin{aligned} \sum_{k=n+1}^m \int_0^1 (X_k - \mu(X_k))^2 d\lambda + 2 \sum_{i=n+1}^m \sum_{k=i+1}^m \int_0^1 (X_i - \mu(X_i))(X_k - \mu(X_k)) d\lambda \\ = \sum_{k=n+1}^m \sigma^2(X_k) + 2 \sum_{i=n+1}^m \sum_{k=i+1}^m g_{i,k} \\ = \sum_{k=n+1}^m O(1) + 2 \sum_{i=n+1}^m \sum_{k=i+1}^m O(\theta^{k-i}) \\ = O(m-n), \end{aligned}$$

which is what we want. □

By Theorem 5.8, we can apply Theorem 2.6, and so

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{a.s.} \frac{1}{n} \sum_{k=1}^n \mu(X_k). \tag{5.9}$$

In other words, considering (5.5),

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^n f(a_k) = \frac{1}{\ln 2} \sum_{k=1}^\infty f(k) \ln\left(1 + \frac{1}{k^2 + 2k}\right),$$

for almost all real numbers in $[0,1]$. Suitable choices of f give numerous corollaries on almost all simple continued fractions. In particular, setting $f(n) = \ln n$ yields

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \ln a_k = \frac{1}{\ln 2} \sum_{k=1}^{\infty} \ln(k) \ln\left(1 + \frac{1}{k^2+2k}\right).$$

Equivalently, we get the fascinating result

$$\lim_{n\to\infty} (a_1 a_2, \ldots a_n)^{1/n} = \prod_{k=1}^{\infty} \left(1 + \frac{1}{k^2+2k}\right)^{\ln k/\ln 2} = K_0 = 2.6854520\ldots$$

for almost all real numbers in $[0,1]$. This result could have been proved without ever evoking probability theory, as done in Khinchin's text. However, it is really interesting to see how the proofs Khinchin's gave can be simplified by rephrasing some of the results and seeing how they relate to statements in probability. Instead of having an isolated proof, this result can instead be viewed as part of a bigger picture, being an application of a law of large numbers.

Acknowledgments

It is a pleasure to thank my mentor Megan Roda for her great efforts helping me learn measure theory, and for giving me readings that I found really interesting. I would also like to thank Peter May for organizing the REU - it has been a great experience.

References

[1] Rockett, A.M., and P., Szusz. Continued Fractions.World Scientific Publishing Company, 1992.
[2] A. Ya. Khinchin. Continued fractions. The University of Chicago Press, Chicago, Ill.-London, 1964.
[3] Kuzmin, R. O. (1928). "On a problem of Gauss". Dokl. Akad. Nauk SSSR: 375–380