Journal of Mathematical Sciences, Vol. 146, No. 2, 2007

ON GAUSS–KUZ'MIN STATISTICS FOR FINITE CONTINUED FRACTIONS

A. V. Ustinov

UDC 511.37+511.336

ABSTRACT. The article is devoted to finite continued fractions for numbers a/b when integer points (a, b) are taken from a dilative region. Properties similar to the Gauss–Kuz'min statistics are proved for these continued fractions.

1. Notation

(1) The expression $[x_0; x_1, \dots, x_s]$ denotes the continued fraction

$$x_0 + \cfrac{1}{x_1 + \cfrac{}{\ddots + \cfrac{1}{x_s}}}$$

of length s with formal variables $x_0, x_1, \dots, x_s$.

(2) For rational r, the representation $r = [t_0; t_1, \dots, t_s]$ is a canonical expansion of r in the continued fraction, where $t_0 = [r]$ (the integer part of r), $t_1, \dots, t_s$ are positive integers such that $t_s \geq 2$ for $s \geq 1$.

(3) For $x \in [0, 1]$ and rational $r = [t_0; t_1, \dots, t_s]$, $s_x(r)$ is the quantity of numbers $j \in \{1, \dots, s\}$ such that $[0; t_j, \dots, t_s] \leq x$. In particular, the length of the continued fraction $s = s(r)$ is $s_1(r)$.

(4) The asterisk in any double sum like

$$\sum_n \sum_m{}^{*} \dots$$

means that the variables also satisfy the condition $(m, n) = 1$.

(5) If A is a proposition that can be true or false, then the bracketed notation $[A]$ stands for 1 if A is true and 0 otherwise.

(6) For natural q, the symbol $\delta_q(a)$ denotes the characteristic function of divisibility by q:

$$\delta_q(a) = [a \equiv 0 \pmod p] = \begin{cases} 1 & \text{if } a \equiv 0 \pmod q, \\ 0 & \text{if } a \not\equiv 0 \pmod q. \end{cases}$$

(7) The finite differences of the function $a(u, v)$ are

$$\Delta_{1,0} a(u, v) = a(u+1, v) - a(u, v), \quad \Delta_{0,1} a(u, v) = a(u, v+1) - a(u, v),$$
$$\Delta_{1,1} a(u, v) = \Delta_{0,1}(\Delta_{1,0} a(u, v)) = \Delta_{1,0}(\Delta_{0,1} a(u, v)).$$

(8) The sum of the kth powers of the divisors is denoted as

$$\sigma_k(q) = \sum_{d|q} d^k.$$

Translated from Fundamentalnaya i Prikladnaya Matematika, Vol. 11, No. 6, pp. 195–208, 2005.

2. Introduction

A detailed analysis of the Euclidean algorithm leads to various problems concerning the statistical properties of finite continued fractions (see [10, Sec. 4.5.3]). If a pair of positive integers a and b ($a < b$) is supplied at the input of the algorithm, then the number of divisions $s(a, b)$ is of main interest. It is equal to the number of quotients in the continued fraction

$$\frac{a}{b} = [0; t_1, \ldots, t_s].$$

The first result about the average length of the Euclidean algorithm belongs to Heilbronn [7], who proved that

$$\frac{1}{\varphi(b)} \sum_{\substack{1 \le a \le b \\ (a,b)=1}} s\left(\frac{a}{b}\right) = \frac{12 \log 2}{\pi^2} \log b + O\left(\frac{b}{\varphi(b)} \sigma_{-1}^3(b)\right).$$

Later, Porter [12] obtained an asymptotic formula with two significant terms for the same sum

$$\frac{1}{\varphi(b)} \sum_{\substack{1 \le a \le b \\ (a,b)=1}} s\left(\frac{a}{b}\right) = \frac{12 \log 2}{\pi^2} \log b + C_P + O\big(b^{-1/6+\varepsilon}\big),$$

where

$$C_P = \frac{\log 2}{\zeta(2)} \left(3 \log 2 + 4\gamma - 4\frac{\zeta'(2)}{\zeta(2)} - 2\right) - \frac{1}{2}$$

is Porter's constant; its closed form was found by Wrench (see [9]). Intermediate results in this direction belong to Tonkov [13, 14].

By averaging over both parameters a and b, it is possible to obtain more detailed information. So, Dixon [6] showed that, for any positive ε, there is a constant $c_0 > 0$ such that

$$\left| s\left(\frac{a}{b}\right) - \frac{12 \log 2}{\pi^2} \log b \right| < (\log b)^{1/2+\varepsilon}$$

for all pairs (a, b) in the region $1 \le a \le b \le R$, except for at most $R^2 \exp(-c_0 (\log R)^{\varepsilon/2})$ pairs.

Hensley [8] improved Dixon's result and proved that, asymptotically, the difference between $s(a/b)$ and its average is normally distributed and the parameters of this distribution can be explicitly written.

Recently, Vallée developed an approach allowing (by averaging over a and b) to investigate the average operating time of various variants of Euclidean algorithms [17] including binary algorithm [16].

Moreover, in many cases, the length of Euclidean algorithms is Gaussian [5].

More exact information on the continued fraction of number a/b is provided by the quantity $s_x(a/b)$, which is a discrete analog of the Gauss–Kuz'min statistics $F_n(x)$. For fixed $x \in [0, 1]$, the function $F_n(x)$ is defined as the measure of the numbers

$$\alpha = [0; t_1, \ldots, t_n, t_{n+1}, \ldots] \in [0, 1]$$

such that

$$\alpha_n = [0; t_{n+1}, t_{n+2}, \ldots] \in [0, x].$$

In [11], Kuz'min proved the Gauss conjecture

$$\lim_{n \to \infty} F_n(x) = \log_2(x + 1.)$$

The final result in this direction belongs to Babenko [4] who proved that

$$F_n(x) = \log_2(x + 1) + \sum_{j=1}^{\infty} \lambda_j^n \psi_j(x),$$

where $\lambda_j \to 0$, $|\lambda_1| > |\lambda_2| \ge |\lambda_3| \ge \ldots$, and $\psi_j(x)$ are analytical functions in $\mathbb{C} \setminus (-\infty, -1)$.

Arnold (see [1, No. 1993-11]) posed the problem about the statistical properties of elements of continued fractions for numbers a/b such that the points (a, b) are inside a sector $a, b > 0$, $a^2 + b^2 \leq R^2$ or inside an extending region $\Omega(R)$ of the general form. The problem about the asymptotic behavior of the sum

$$N_x(R) = \sum_{(a,b)\in\Omega(R)} s_x\left(\frac{a}{b}\right)$$

is more general. It is similar to the problem on Gauss–Kuz'min statistics. For the sector, this problem was solved by Avdeeva and Bykovsky in [3]. In [2], Avdeeva proved the more precise asymptotic formula

$$N_x(R) = \frac{3}{\pi}\log(1+x)R^2\log R + O(R^2),$$

with a better error term than in [3]. Later, Ustinov obtained the asymptotic formula with two significant terms (see [15])

$$N_x(R) = \frac{3}{\pi}R^2\left[\log(1+x)\log R + C(x)\right] + O(R^{17/9}\log^2 R),$$

with a complicated function $C(x)$.

In the present paper, a similar problem is considered for the general region $\Omega(R)$, which is the image of a fixed region Ω_0 under a homothety with ratio $R > 1$:

$$\Omega(R) = R\cdot\Omega_0 = \left\{(x, y)\colon x, y > 0,\ \left(\frac{x}{R}, \frac{y}{R}\right) \in \Omega_0\right\}.$$

The region Ω_0 is defined in polar coordinates

$$\Omega_0 = \left\{(\rho, \varphi)\colon 0 \leq \varphi \leq \frac{\pi}{4},\ 0 \leq \rho \leq r(\varphi)\right\}$$

and its area is

$$V_0 = \frac{1}{2}\int\limits_0^{\pi/4} r^2(\varphi)\, d\varphi.$$

If the function $r(\varphi)$ is defined on the closed interval $[0, \pi/4]$ and satisfies the conditions

$$r(\varphi) \geq \varepsilon_0 > 0, \quad r'(\varphi) \leq r(\varphi)\arctan\varphi,$$

then, for the sum

$$N_x(R) = \sum_{(a,b)\in\Omega(R)} s_x\left(\frac{a}{b}\right),$$

the asymptotic formula with two significant terms

$$N_x(R) = \frac{2V_0}{\zeta(2)}R^2(\log(x+1)\log R + C(x)) + O(R^{2-1/5}\log^3 R)$$

is proved. This general result is more precise than that in [15] and it shows that the principal term in the Gauss–Kuz'min statistics for finite continued fractions depends not on the form of the region Ω_0 but only on its area.

3. Auxiliary Transformation

Let $N_x^*(R)$ be the sum

$$N_x^*(R) = \sum_{\substack{(a,b)\in\Omega(R)\\(a,b)=1}} s_x\left(\frac{a}{b}\right).$$

Since

$$N_x(R) = \sum_{d\leq R} N_x^*\left(\frac{R}{d}\right),$$

it is sufficient to obtain the asymptotic formula for $N_x^*(R)$.

Let $T_x^*(R)$ be the number of solutions to the system

$$\begin{cases} PQ' - P'Q = \pm 1, \\ mP + nP' = a, \\ mQ + nQ' = b, \\ a^2 + b^2 \le R^2 r^2 \left(\arctan \dfrac{a}{b}\right), \end{cases} \tag{1}$$

where

$$1 \le Q \le Q', \quad 1 \le P' \le Q', \quad 0 \le P \le Q, \quad 1 \le m \le xn, \quad (m,n) = 1. \tag{2}$$

Similarly to [15, Lemma 3], it can be proved that, for any $R \ge 2$ and $x \in [0;1]$, the following formula holds:

$$N_x^*(R) = T_x^*(R) + \frac{R^2}{\zeta(2)}[x < 1]V_0(x) + O(R \log R), \tag{3}$$

where

$$V_0(x) = \frac{1}{2} \int_0^{\arctan x} r^2(\varphi)\, d\varphi.$$

For the further study of $T_x^*(R)$, we introduce a parameter U lying in the interval $1 \le U \le R$. Let T_1 be the number of solutions to system (1) with constraints (2) that satisfy the additional condition $Q' \le U$. Denote by T_2 the number of solutions with $Q' > U$. Then

$$T_x^*(R) = T_1 + T_2.$$

Each of the terms T_1 and T_2 will be considered separately.

4. The Evaluation of T_1

Lemma 1. *Let $q \ge 1$ be an integer and the function $a(u,v)$ be defined at integer points (u,v) such that $1 \le u, v \le q$. We also assume that this function satisfies the inequalities*

$$a(u,v) \ge 0, \quad \Delta_{1,0}a(u,v) \le 0, \quad \Delta_{0,1}a(u,v) \le 0, \quad \Delta_{1,1}a(u,v) \ge 0 \tag{4}$$

at all points where these conditions are defined. Then, for the sum

$$W = \sum_{u,v=1}^{q} \delta_q(uv-1)a(u,v),$$

the following asymptotic formula holds:

$$W = \frac{\varphi(q)}{q^2} \sum_{u,v=1}^{q} a(u,v) + O(A\psi(q)\sqrt{q}),$$

where $\psi(q) = \sigma_0(q)\sigma_{-1/2}(q)\log^2(q+1)$ and $A = a(1,1)$ is the greatest value of the function $a(u,v)$.

For the proof see [15].

Let q be a positive integer and $x \in [0,1]$. For integers u and v $(1 \le u, v \le q)$, by $I_q(u,v)$, we denote the interval

$$\left[\arctan\left(\frac{u}{q} - \frac{x}{q(q+vx)}\right), \arctan\frac{u}{q}\right].$$

Lemma 2. *Let $r(\varphi) \in C^{(1)}([0,\pi/4])$ be a nonnegative function satisfying the condition $r'(\varphi) \le r(\varphi)\tan\varphi$ for $\varphi \in [0,\pi/4]$. Then, for the sum*

$$W_1(q) = \frac{1}{2} \sum_{u,v=1}^{q} \delta_q(uv-1) \int_{I_q(u,v)} r^2(\varphi)\, d\varphi,$$

the following asymptotic formula holds:

$$W_1(q) = V_0 \log(1+x)\frac{\varphi(q)}{q^2} + O\left(\frac{\psi(q)}{q^{3/2}}\right),$$

where

$$V_0 = \frac{1}{2}\int\limits_0^{\pi/4} r^2(\varphi)\,d\varphi.$$

Proof. From conditions $r(\varphi) \geq 0$ and $r'(\varphi) \leq r(\varphi)\tan\varphi$, it follows that the function

$$a(u,v) = \int\limits_{I_q(u,v)} r^2(\varphi)\,d\varphi$$

satisfies conditions (4). Hence, by Lemma 1, we have

$$W_1(q) = \frac{\varphi(q)}{2q^2}\sum_{u,v=1}^{q}\int\limits_{I_q(u,v)} r^2(\varphi)d\varphi + O\left(\frac{\psi(q)\sqrt{q}}{q^2}\right).$$

Lagrange's theorem implies that

$$\int\limits_{I_q(u,v)} r^2(\varphi)\,d\varphi = \frac{x}{q(q+vx)}\,\frac{1}{1+\frac{u^2}{q^2}}\left(r^2\left(\arctan\frac{u}{q}\right) + O\left(\frac{1}{q^2}\right)\right),$$

$$\frac{x}{q+vx} = \log(q+vx) - \log(q+(v-1)x) + O\left(\frac{1}{q^2}\right),$$

$$\frac{1}{q(1+\frac{u^2}{q^2})}r^2\left(\arctan\frac{u}{q}\right) = \int\limits_{u-1}^{u} r^2\left(\arctan\frac{z}{q}\right)\,d\arctan\frac{z}{q} + O\left(\frac{1}{q^2}\right).$$

Therefore,

$$W_1(q) = \frac{\varphi(q)}{2q^2}\sum_{u=1}^{q}\int\limits_{u-1}^{u} r^2\left(\arctan\frac{z}{q}\right)\,d\arctan\frac{z}{q}\sum_{v=1}^{q}[\log(q+vx) - \log(q+(v-1)x)] + O\left(\frac{\psi(q)}{q^{3/2}}\right)$$

$$= V_0\log(1+x)\frac{\varphi(q)}{q^2} + O\left(\frac{\psi(q)}{q^{3/2}}\right).$$

The following proposition easily follows from Lemma 2.

Corollary 1. *Let $N \geq 1$. Then, for the sum*

$$W_2 = \sum_{q\leq N} W_1(q),$$

we have

$$W_2 = V_0\frac{\log(1+x)}{\zeta(2)}\left(\log N + \gamma - \frac{\zeta'(2)}{\zeta(2)}\right) + f(x) + O\left(\frac{\log^5(N+1)}{\sqrt{N}}\right), \tag{5}$$

where $f(x)$ is the function defined by the series

$$f(x) = \sum_{q=1}^{\infty}\left(W_1(q) - V_0\log(1+x)\frac{\varphi(q)}{q^2}\right). \tag{6}$$

Remark. In the same way, for the sum

$$W_3(q) = \frac{1}{2} \sum_{u,v=1}^{q} \delta_q(uv-1) \int\limits_{I'_q(u,v)} r^2(\varphi)\, d\varphi,$$

where

$$I'_q(u,v) = \left[\arctan \frac{u}{q}, \arctan\left(\frac{u}{q} + \frac{x}{q(q+vx)}\right)\right]$$

and $q \geq 2$, we we obtain the asymptotic formula

$$W_3(q) = \frac{\pi}{4} \log(1+x) \frac{\varphi(q)}{q^2} + O\left(\frac{\psi(q)}{q^{3/2}}\right).$$

Accordingly, the sum

$$W_4 = \sum_{2 \leq q \leq N} W_3(q)$$

for $N \geq 1$ has the following representation:

$$W_4 = V_0 \frac{\log(1+x)}{\zeta(2)} \left(\log N + \gamma - \frac{\zeta'(2)}{\zeta(2)}\right) + g(x) + O\left(\frac{\log^5(N+1)}{\sqrt{N}}\right), \tag{7}$$

where

$$g(x) = \sum_{q=2}^{\infty} \left(W_3(q) - V_0 \log(1+x) \frac{\varphi(q)}{q^2}\right). \tag{8}$$

Summation begins with $q = 2$ because $q = 1$ leads to the numbers $Q' = Q = P' = 1$, $P = 2$, which do not satisfy conditions (2).

Formulas (5) and (7) (as in [15]) give an asymptotic formula for T_1.

Theorem 1. *Let* $1 \leq U \leq R$. *Then*

$$T_1 = \frac{2V_0}{\zeta^2(2)} R^2 (\log(x+1) \log U + C_1(x)) + O(R^2 U^{-1/2} \log^5 R) + O(RU \log R),$$

where

$$C_1(x) = \log(x+1) \left(\gamma - \frac{\zeta'(2)}{\zeta(2)}\right) + \frac{\zeta(2)}{2V_0} (f(x) + g(x)), \tag{9}$$

and the functions $f(x)$ *and* $g(x)$ *are defined by* (6) *and* (8).

5. Evaluation of T_2

Lemma 3. *Let* $q \geq 1$ *be an integer and* $f(u)$ *be a nonnegative nonincreasing function on* $[0; q]$ *and* $f(0) \leq q$. *Then*

$$\sum_{u=1}^{q} \sum_{1 \leq v \leq f(u)} \delta_q(uv \pm 1) = \frac{\varphi(q)}{q^2} V(\Omega) + O(q^{3/4} \sigma_0(q) \log(q+1)),$$

where Ω *is the region on the* uv*-plane defined by the inequalities* $0 \leq u \leq q$, $0 \leq v \leq f(u)$ *and* $V(\Omega)$ *is its area.*

For the proof see [15, Lemma 7].

Consider a function $v(u)$ defined in the region $1 \leq u \leq q-1$, $0 \leq v \leq q$ as an implicit function by the equation

$$a^2 + b^2 = R^2 r^2 \left(\arctan \frac{a}{b}\right), \tag{10}$$

where

$$a = m \frac{uv \pm 1}{q} + nu, \quad b = mv + nq.$$

In the following proposition, we assume that the function $v(u)$ is defined at least at one point.

Lemma 4. *Let $R \geq 1$ be a real number, m, n, and q be positive integers, and $1 \leq m \leq n$. Moreover, assume that the function $r(\varphi) \in C^{(1)}([0, \pi/4])$ satisfies the conditions*

$$r(\varphi) \geq \varepsilon_0 > 0, \quad r'(\varphi) \leq r(\varphi) \arctan \varphi.$$

Then, for

$$q^2 > U_0 = \frac{13}{\varepsilon_0^2} \max_{\varphi \in [0, \pi/4]} r(\varphi)|r'(\varphi)|, \tag{11}$$

the function $v(u)$ is defined on a closed interval $[q_0, q-1]$ $(1 \leq q_0 \leq q-1)$ and $v(u)$ is nonincreasing on this interval.

Proof. First, note that

$$\frac{\partial(a^2+b^2)}{\partial v} = 2a\frac{mu}{q} + 2bm \geq 2m,$$
$$\left|\frac{R^2 \partial r^2(\arctan a/b)}{\partial v}\right| = \frac{2R^2 r|r'|m}{(a^2+b^2)q^2} \leq \frac{2R^2 r|r'|m}{n^2 q^4},$$

where $r = r(\arctan a/b)$ and $r' = r'(\arctan a/b)$.

Since the function $v(u)$ must be defined at least at one point, the parameter R satisfies the inequality

$$\left(m\frac{q^2+1}{q} + nq\right)^2 + (mv+nq)^2 \geq R^2 \varepsilon_0^2.$$

Hence, $R^2 \leq 13n^2q^2/\varepsilon_0^2$, and, for $q^2 > U_0$,

$$\left|\frac{R^2 \partial r^2(\arctan a/b)}{\partial v}\right| \leq \frac{2 \cdot 13}{\varepsilon_0^2} \frac{r|r'|m}{q^2} < 2m \leq \frac{\partial(a^2+b^2)}{\partial v}.$$

Therefore, for fixed u, Eq. (10) determines at most one value of v.

We can differentiate $v(u)$ as an implicit function:

$$\frac{dv}{du} = -\frac{b^2}{m} \frac{a/b - r'/r}{au + bq \pm mr'/r}.$$

The assumption $r'/r \leq a/b$ implies that the numerator of the above expression is nonnegative. Since $m/n \leq 1$ and $q^2 > r'/r$, we have

$$au + bq \pm m\frac{r'}{r} \geq nq^2 \pm m\frac{r'}{r} = n\left(q^2 \pm \frac{m}{n}\frac{r'}{r}\right) > 0.$$

Hence, the function $v(u)$ does not increease and is defined on a closed interval segment $[u_0, q-1]$, where $1 \leq u_0 \leq q-1$.

Lemma 5. *Suppose that function $r(\varphi)$ satisfies the conditions of Lemma 4, U_0 is defined by (11), $U_0^{1/2} \leq U < R$, and $R_1 = R/U$. Then*

$$T_2 = 2\sum_{n<R_1} \sum_{m \leq nx}{}^{*} \sum_{U<q\leq R} \frac{\varphi(q)}{q^2} V(m,n,q) + O(R^2 U^{-1/4} \log^2 R),$$

where $V(m,n,q)$ is the area of the region $\Omega(m,n,q)$ on the uv-plane defined by the inequalities

$$0 \leq u, v \leq q, \quad \left(\frac{u^2}{q^2}+1\right)(mv+nq)^2 \leq R^2 r^2\left(\arctan\frac{u}{q}\right).$$

Proof. By definition of T_2, we have

$$T_2 = \sum_{1\le n<R_1} \sideset{}{^*}\sum_{m\le nx} \sum_{U<q\le R/n} \sum_{u,v=1}^{q} \delta_q(uv\pm 1)\left[a^2+b^2 \le R^2r^2\left(\arctan\frac{a}{b}\right)\right],$$

where

$$a = m\frac{uv\pm 1}{q}+nu, \quad b = mv+nq.$$

From Lemmas 4 and 3, it follows that

$$T_2 = \sum_{1\le n<R_1} \sideset{}{^*}\sum_{m\le nx} \sum_{U<q\le R/n} \left(\frac{\varphi(q)}{q^2}V_\pm(m,n,q)+O(q^{3/4}\sigma_0(q)\log q)\right),$$

where $V_\pm(m,n,q)$ is the area of the region $\Omega_\pm(m,n,q)$ on the uv-plane defined by

$$0\le u,v\le q, \tag{12}$$

$$\left(m\frac{uv\pm 1}{q}+nu\right)^2+(mv+nq)^2 \le R^2r^2\left(\arctan\frac{u}{q}\pm\frac{m}{q(mv+nq)}\right). \tag{13}$$

To complete the proof, it is sufficient to show that

$$V_\pm(m,n,q) = V(m,n,q)+O(q). \tag{14}$$

Indeed, this equality implies the asymptotic formula

$$T_2 = 2\sum_{n<R_1} \sideset{}{^*}\sum_{m\le nx} \sum_{U<q\le R/n} \frac{\varphi(q)}{q^2}V(m,n,q)+O(R^2U^{-1/4}\log^2 R),$$

and this formula is equivalent to the assertion of the lemma, because the condition $q\le R/n$ can be replaced by $q<R$ (for $nq>R$, the region $\Omega(m,n,q)$ is empty and $V(m,n,q)=0$).

To prove (14), we consider the difference between the intervals of v specified by (12) and (13) for a fixed u $(1\le u\le q-1)$. At least one of these intervals should be nonempty. Therefore, one of the following inequalities is valid:

$$\left(m\frac{uv\pm 1}{q}+nu\right)^2+(mv+nq)^2 > R^2r^2\left(\arctan\frac{u}{q}\pm\frac{m}{q(mv+nq)}\right),$$

and $R\ll nq$. Let $(u,v)\in\Omega(m,n,q)\setminus\Omega_\pm(m,n,q)$ or $(u,v)\in\Omega_\pm(m,n,q)\setminus\Omega(m,n,q)$. Using the formulas

$$\sqrt{\left(m\frac{uv\pm 1}{q}+nu\right)^2+(mv+nq)^2} = (mv+nq)\sqrt{\frac{u^2}{q^2}+1}+O\left(\frac{m}{q}\right)$$

and

$$r\left(\arctan\frac{u}{q}\pm\frac{m}{q(mv+nq)}\right) = r\left(\arctan\frac{u}{q}\right)+O\left(\frac{m}{nq^2}\right),$$

we get

$$Rr\left(\arctan\frac{u}{q}\right)-(mv+nq)\sqrt{\frac{u^2}{q^2}+1} \ll \frac{Rm}{nq^2}+\frac{m}{q}\ll\frac{m}{q}.$$

Therefore, for fixed u $(1\le u\le q-1)$, the variable v varies within an interval of length $O(1/q)$.

The difference between the areas of the regions $\Omega(m,n,q)$ and $\Omega_\pm(m,n,q)$ within the strips $0\le u\le 1$ and $q-1\le u\le q$ is less than q. This implies formula (14). Lemma 5 is proved.

Lemma 6. *Let $1 \le U \le R$ and $R_1 = R/U$. Then, for the sum*

$$W_5 = \sum_{n<R_1} \sideset{}{^*}\sum_{m \le nx} \sum_{U<q\le R} \frac{\varphi(q)}{q^2} V(m,n,q),$$

we have

$$W_5 = \frac{U^2}{\zeta(2)} \int\limits_0^1 dt \int\limits_0^{R_1(t)} \xi F^*(\xi)\, d\xi + O(R^2 U^{-1} \log R),$$

where $R_1(t) = R_1 r(\arctan t)/\sqrt{t^2+1}$ and

$$F^*(\xi) = \sum_{n<\xi} \sideset{}{^*}\sum_{m\le nx} \frac{1}{m}\left(\frac{1}{n} - \frac{1}{m+n}\right)[\xi \ge m+n] + \sum_{n<\xi} \sideset{}{^*}\sum_{m\le nx} \frac{1}{m}\left(\frac{1}{n} - \frac{1}{\xi}\right)[\xi < m+n].$$

For the proof see [15, Lemma 9].

Corollary 2. *Let $1 \le U \le R$, $R_1 = R/U$, and*

$$R_1(t) = R_1 \frac{r(\arctan t)}{\sqrt{t^2+1}}.$$

Then

$$T_2 = 2\frac{U^2}{\zeta(2)} \int\limits_0^1 dt \int\limits_0^{R_1(t)} \xi F^*(\xi)\, d\xi + O(R^2 U^{-1/4} \log^2 R).$$

The proof follows from Lemmas 5 and 6.

Lemma 7. *Let $N > 1$. Then, for the sum*

$$F^*(N) = \sum_{n<N} \sideset{}{^*}\sum_{m\le nx} \frac{1}{m}\left(\frac{1}{n} - \frac{1}{m+n}\right) - \sum_{n<N} \sideset{}{^*}\sum_{\substack{m\le nx \\ m+n>N}} \frac{1}{m}\left(\frac{1}{N} - \frac{1}{m+n}\right),$$

the following asymptotic formula holds:

$$F^*(N) = \frac{\log(x+1)}{\zeta(2)} \log N + \frac{H(x)}{\zeta(2)} + O\left(\frac{\log^2(N+1)}{N}\right),$$

where

$$H(x) = \log(x+1)\left(\log x - \frac{\zeta'(2)}{\zeta(2)} - \frac{1}{2}\log(x+1) + \gamma - 1\right) + h(x)$$

and

$$h(x) = \sum_{m=1}^{\infty}\left(\sum_{m/x\le n<m/x+m} \frac{1}{n} - \log(x+1)\right). \tag{15}$$

For the proof see [15, Lemma 10].

The following proposition can be proved by direct calculations.

Lemma 8. *Let $R_1 > 0$ and*

$$R_1(t) = R_1 \frac{r(\arctan t)}{\sqrt{t^2+1}}$$

for $t \in [0,1]$. *Then*

$$\int_0^1 dt \int_0^{R_1(t)} \xi \, d\xi = V_0 R_1^2,$$

$$\int_0^1 dt \int_0^{R_1(t)} \xi \log \xi \, d\xi = V_0 R_1^2 \left(\log R_1 - \frac{1}{2} \right) + V_1 R_1^2,$$

where

$$V_1 = \frac{1}{2} \int_0^{\pi/4} r^2(\varphi) \log(r(\varphi)\cos\varphi) \, d\varphi.$$

Theorem 2. *Let* $1 \le U \le R$ *and* $R_1 = R/U$. *Then*

$$T_2 = \frac{2V_0}{\zeta^2(2)} R^2 (\log(x+1)\log R_1 + C_2(x)) + O(R^2 U^{-1/4} \log^2 R) + O(RU \log^2 R),$$

where

$$C_2(x) = \log(x+1) \left(\log x - \frac{\zeta'(2)}{\zeta(2)} + \gamma - \frac{\log(x+1)}{2} - \frac{3}{2} + \frac{V_1}{V_0} \right) + h(x) \tag{16}$$

and the function $h(x)$ *is defined by* (15).

The proof follows from Lemmas 7 and 8 and Corollary 2.

6. Main Result

Theorem 3. *Let* $R \ge 2$. *Then*

$$N_x(R) = \frac{2V_0}{\zeta(2)} R^2 (\log(x+1)\log R + C(x)) + O(R^{2-\frac{1}{5}} \log^3 R),$$

where

$$C(x) = \log(x+1) \left(\log x - \frac{\zeta'(2)}{\zeta(2)} + 2\gamma - \frac{\log(x+1)}{2} - \frac{3}{2} + \frac{V_1}{V_0} \right) + h(x) + \frac{\zeta(2)}{2V_0}(f(x) + g(x) + V_0(x)[x<1])$$

and $f(x)$, $g(x)$ *are defined by* (6) *and* (8).

Proof. From Theorems 1 and 2, we obtain

$$T_x^*(R) = T_1 + T_2 = \frac{2V_0}{\zeta^2(2)} R^2 (\log(x+1)\log R + C_1(x) + C_2(x)) \\ + O(R^2 U^{-1/2} \log^5 R) + O(R^2 U^{-1/4} \log^2 R) + O(RU \log^2 R).$$

Choosing $U = R^{4/5}$ and substituting the result into (3), we obtain

$$N_x^*(R) = \frac{2V_0}{\zeta^2(2)} R^2 (\log(x+1)\log R + C_3(x)) + O(R^{9/5} \log^2 R),$$

where

$$C_3(x) = C_1(x) + C_2(x) + \frac{\zeta(2)}{2} \frac{V_0(x)}{V_0}[x<1].$$

Finally, applying the formula

$$N_x(R) = \sum_{d \le R} N_x^* \left(\frac{R}{d} \right),$$

we complete the proof of the theorem.

This work was supported by INTAS, grant No. 03-51-5070.

REFERENCES

1. *Arnold's Problems* [in Russian], Fazis, Moscow (2000).
2. M. O. Avdeeva, "On the statistics of partial quotients of finite continued fractions," *Funkts. Anal. Prilozhen.*, **38**, No. 2, 1–11 (2004).
3. M. O. Avdeeva and V. A. Bykovskii, *A Solution of Arnold's Problem on the Gauss–Kuzmin Statistic*, preprint FEB RAS Khabarovsk Division of the Institute for Applied Mathematics, No. 8, Dalnauka, Vladivostok (2002).
4. K. I. Babenko, "On a problem of Gauss," *Sov. Math. Dokl.*, **19**, 136–140 (1978).
5. V. Baladi and B. Valle, "Euclidean algorithms are Gaussian," *J. Number Theory*, **110**, No. 2, 331–386 (2005).
6. J. D. Dixon, "The number of steps in the Euclidean algorithm," *J. Number Theory*, **2**, 414–422 (1970).
7. H. Heilbronn, "On the average length of a class of finite continued fractions," in: *Abhandlungen aus Zahlentheorie und Analysis*, VEB Deutsher Verlag der Wissenschaften, Berlin, Plenum Press, New York (1968), pp. 89–96.
8. D. Hensley, "The number of steps in the Euclidean algorithm," *J. Number Theory*, **49**, No. 2, 142–182 (1994).
9. D. E. Knuth, "Evaluation of Porter's constant," *Comput. Math. Appl.*, **2**, 137–139 (1976).
10. D. E. Knuth, *The Art of Computer Programming*, Vol. 2, 2nd ed., Massachusetts, Addison-Wesley (1981).
11. R. O. Kuz'min, "Sur un problème de Gauss," in: *Atti del Congresso Internazionale dei Matematici*, Bologna (1928), pp. 83–89.
12. J. W. Porter, "On a theorem of Heilbronn," *Mathematika*, **22**, no. 1, 20–28 (1975).
13. T. Tonkov, "On the average length of finite continued fractions," *Acta Arith.*, **26**, 47–57 (1974).
14. T. Tonkov, "On the average length of finite continued fractions," *Math. Balkanica*, **4**, 617–629 (1974).
15. A. V. Ustinov, "On statistic properties of continued fractions," in: *Works on Number Theory* [in Russian], Zap. Nauchn. Sem. St. Peterb. Otd. Mat. Inst. im. V. A. Steklova, Ross. Akad. Nauk (POMI), Vol. 322, St. Petersburg (2005), pp. 186–211.
16. B. Vallée, "Dynamics of the binary Euclidean algorithm: Functional analysis and operators," *Algorithmica*, **22**, 660–685 (1998).
17. B. Vallée, "A unifying framework for the analysis of a class of Euclidean algorithms," in: *Proceedings of LATIN'2000*, Lect. Notes Comp. Sci., Vol. 1776, Springer, pp. 343–354.

A. V. Ustinov
Moscow State University
E-mail: `ustinov@mech.math.msu.su`