New York Journal of Mathematics

New York J. Math. **4** (1998) 249–257.

Metric Diophantine Approximation and Probability

Doug Hensley

Abstract. Let $p_n/q_n = (p_n/q_n)(x)$ denote the n^{th} simple continued fraction convergent to an arbitrary irrational number $x \in (0,1)$. Define the sequence of approximation constants $\theta_n(x) := q_n^2|x - p_n/q_n|$. It was conjectured by Lenstra that for almost all $x \in (0,1)$,

$$\lim_{n\to\infty} \frac{1}{n}|\{j : 1 \le j \le n \text{ and } \theta_j(x) \le z\}| = F(z)$$

where $F(z) := z/\log 2$ if $0 \le z \le 1/2$, and $\frac{1}{\log 2}(1 - z + \log(2z))$ if $1/2 \le z \le 1$. This was proved in [BJW83] and extended in [Nai98] to the same conclusion for $\theta_{k_j}(x)$ where k_j is a sequence of positive integers satisfying a certain technical condition related to ergodic theory. Our main result is that this condition can be dispensed with; we only need that k_j be strictly increasing.

Contents

1. Introduction

Metric questions about Diophantine approximation can be approached by means of ergodic theory, dynamical systems, weak mixing and so on. At the heart of this approach lies the observation that not only is the map $T : [0,1]\backslash Q \to [0,1]\backslash Q$ given by $T : x \to \langle 1/x\rangle =: 1/x - [1/x]$ ergodic, but that $\mathcal{T} : \Omega \to \Omega := ([0,1]\backslash Q) \times [0,1]$ given by $\mathcal{T} : (x,y) \to (\langle 1/x\rangle, 1/([1/x]+y))$ is ergodic with better mixing properties. The associated measure, invariant under $\mathcal{T}$ assigns to measurable $A \subset \Omega$ mass $\int_A \frac{1}{\log 2} \int_A \frac{dx\,dy}{(1+xy)^2}$. Here we take a different tack. It goes back to the pioneering work that led to the Gauss-Kuzmin theorem, and the thread continues with the work of Wirsing and Babenko on the convergence rate in that theorem. Most recently, Vallée et al have had signal success with this circle of ideas, analyzing for

Received July 14, 1998.
Mathematics Subject Classification. 11K50 primary, 11A55, 60G50 secondary.
Key words and phrases. continued fractions, distribution, random variable.

ISSN 1076-9803/98

instance the lattice reduction algorithm in two dimensions [Val95]. This approach uses functional analysis and classical probability. Discussion of linear operators, eigenvalues, and eigenvectors requires that a linear space be specified. We shall get around to this, but for now it suffices to note that the definition below of L would make sense for any reasonable space of functions. At the heart of our approach lies the fact that if X is a random variable on $U := [0,1]\backslash Q$ with density f, then $T^n X$ has density $L^n f$ where $Lf(t) = \sum_{k=1}^{\infty}(k+t)^{-2}f(1/(k+t))$ (for $t \in [0,1]$, else zero) and that L has dominant eigenvalue 1 with corresponding eigenfunction $g(t) := 1/(\log 2(1+t))$. From this it follows that well-separated values of $T^n X$ are nearly independent random variables, so that the usual tools of classical probability can be brought into play.

These statements about random variables and density can be rephrased so as to avoid explicit mention of probability: X is a measurable function from U to U, and $\phi : [0,1] \to [0,1]$ is defined by $\phi(y) = m(\{x \in U : X(x) \le y\})$ where m denotes Lebesgue measure. If ϕ is differentiable on $[0,1]$, then $f := \phi'$ is the density of the random variable X. Similarly, the density of $T^n X$ is

$$(d/dy)\, m(\{x \in U : T^n X(x) \le y\}) = L^n f,$$

a fact which is used in the pioneering work mentioned above and in all subsequent developments along this line.

Recall (from the abstract) that $p_n/q_n = (p_n/q_n)(x)$ denotes the n^{th} simple continued fraction convergent to an arbitrary irrational number $x \in (0,1)$, while $\theta_n(x) := q_n^2|x - p_n/q_n|$. Also,

$$F(z) := \begin{cases} z/\log 2 & \text{if } 0 \le z \le 1/2 \\ \frac{1}{\log 2}(1 - z + \log(2z)) & \text{if } 1/2 \le z \le 1. \end{cases}$$

Our main result is

Theorem 1.1. *If (k_j) is a strictly increasing sequence of positive integers, and $0 < z < 1$, then (with respect to Lebesgue measure)*

$$\lim_{n\to\infty} \frac{1}{n}|\{j : 1 \le j \le n \text{ and } \theta_{k_j}(x) \le z\}| = F(z)$$

for almost all $x \in (0,1)$.

2. Probabilities and operators

The probability $Q_{r,f}(z)$ that $\theta_r(X) \le z$, when the initial density for X is f, is essentially $F(z)$, as we shall see. In the case $f \equiv 1$, this is due to Knuth. [BJW83]. What is new here is that there is a kind of near-independence of these events for widely separated values of r, and uniformly over a certain class of initial probability distributions f.

Let V_r denote the r-fold Cartesian product of the positive integers. For an arbitrary sequence $v \in V_r$ of r positive integers, let $[v] := [0; v_1, v_2, \ldots v_r] = p_r/q_r$, let $\{v\} := [0; v_r, v_{r-1}, \ldots v_1] = q_{r-1}/q_r$, and let $|v| := q_r$. Let $v_- := (v_2, \ldots v_r)$ and let $v^- := (v_1, \dots v_{r-1})$, so that $p_r = |v_-|$ and $q_{r-1} = |v^-|$. Then for $x \in [0,1]\backslash Q$,

$$
\begin{aligned}
(1) \qquad\qquad x &= \frac{p_r + (T^r x)p_{r-1}}{q_r + (T^r x)q_{r-1}} \\
\theta_r(x) &= \frac{T^r x}{1 + \{v\}T^r x} \\
L^r[(1+ut)^{-2}] &= \sum_{v \in V_r} |v|^{-2}(1 + u[v])^{-2}(1 + (\{v\} + u\{v_-\})t)^{-2} \\
(L^r f)(t) &= \sum_{v \in V_r} |v|^{-2}(1 + \{v\}t)^{-2} f([v+t])
\end{aligned}
$$

where $[v+t] := [0; v_1, v_2, \dots v_{r-1}, v_r + t]$. (Thus if f is a convex combination of probability density functions of the form $(1+u)(1+ut)^{-2}$, then so is Lf.) The claim above about the dominant eigenvalue of L can now be given specific content. For an arbitrary function f of bounded variation and zero except in $[0,1]$, (which we will call *good*) let $\|f\|$ be the total variation of f on the real line. (Thus any probability density function has norm at least 2). We have (specializing from Lemma 6 [Hen92, p 346])

Lemma 2.1. *Uniformly over $0 \le t \le 1$ and good probability density functions f of bounded variation, $L^r f = \frac{1}{\log 2(1+t)} + O(((\sqrt{5}-1)/2)^r \|f\|)$.*

Again let f be a *good* probability density function. Let $I(S) := 1$ if S is a true statement, else zero. Let X be a random variable on $[0,1]$ with density f. Then the density of $T^r X$ is $L^r f$. Let $v = v(x,r)$ denote the sequence of the first r partial quotients in the continued fraction expansion of x, so that $x = [v+t]$ where $t = T^r x$.

We now consider the probability $P_{r,z,f}$ that $\{v(X,r)\} \le z$. In the case $r = 1$, this is

$$P_{1,z,f} = \int_{x=0}^{1} f(x) I(\lfloor 1/x \rfloor \le z)\, dx = \sum_{\{k : 1/k \le z\}} \int_{t=0}^{1} (k+t)^{-2} f(1/(k+t))\, dt$$

on substituting $x = 1/(k+t)$. The similar substitution

$$x = [v+t] = [v_1, v_2, \dots v_{r-1}, v_r + t]$$

has $dx/dt = |v|^{-2}(1 + \{v\}t)^{-2}$, so the probability $P_{r,z,f}$ that $\{v\} \le z$ is given by

$$(2) \qquad P_{r,z,f} = \int_{t=0}^{1} \sum_{v \in V_r, \{v\} \le z} |v|^{-2}(1 + \{v\}t)^{-2} f([v+t])\, dt.$$

Much of what lies ahead has to do with conditional probabilities, conditional random variables, and their associated conditional densities. If E_1 is an event (measurable subset of U) with positive probability (that is, $m(E_1) > 0$), then the conditional probability of another event E_2 given E_1 is by definition

$$\mathrm{prob}(E_1 \text{ and } E_2)/\mathrm{prob}(E_1) = m(E_1 \cap E_2)/m(E_1).$$

Most of our events have to do with some random variable X on U with density f, and some function $\tau : U \to U$. Suppose $D_1 \subset R$ and let $E_1 = X^{-1}D_1 = \{x \in U : X(x) \in D_1\}$. The conditional probability that $\tau(X) \in D_2$ given $X \in D_1$ is then

$m((\tau \circ X)^{-1}D_1 \cap X^{-1}D_2)/m(X^{-1}D_1)$, and the conditional density for $\tau(X)$ given that $X \in D_1$ is the function

$$t \to (d/dt)\, m(\{x : \tau(X(x)) \le t \text{ and } X(x) \in D_1\})/m(X^{-1}(D_1)).$$

The conditional density given that $\{v(X,r)\} \le z$, for $T^r X$ is

$$g_{r,z,f} := \sum_{v \in V_r, \{v\} \le z} |v|^{-2}(1+\{v\}t)^{-2} f([v+t])/P_{r,z,f}. \tag{3}$$

Let $h_{r,z,f}(t) := g_{f,r,z}(t)P_{r,z,f}$. By convention both f and g take the value 0 off $[0,1]$.

Because $g_{r,z,f}$ is the conditional density of $T^r X$ given $v(x,r) \le z$, where X is a random variable on $(0,1)$ with density f, it follows that if Y is a random variable on $(0,1)$ with density $g := g_{r,z,f}$ and B is a measurable subset of $[0,1)$ then

$$\text{prob}[\{v(X,r)\} \le z \text{ and } T^r X \in B]/\text{prob}[\{v(X,r)\} \le z] = \text{prob}[Y \in B] \tag{4}$$

We are now in a position to state and prove

Theorem 2.1. *Uniformly over good probability density functions f, over $0 \le z \le 1$, and over $0 \le t \le 1$,*

$$h_{r,z,f} := \sum_{v \in V_r, \{v\} \le z} |v|^{-2}(1+\{v\}t)^{-2} f([v+t])$$

is good and satisfies

$$h_{r,z,f}(t) = \frac{z}{\log 2(1+tz)} + O(rz((\sqrt{5}-1)/2)^r \|f\|)$$

Proof. It is clear that the resulting function again has bounded variation. The real issue is whether it satisfies the estimate. We begin by proving a weaker version of the theorem, in which the z in the error term is replaced with 1. For this weaker version, the ground floor of induction is trivial. Let $\lambda = (\sqrt{5}-1)/2$. Let $V_r(z)$ denote the set of all $v \in V_r$ so that $\{v\} = q_{r-1}/q_r \le z$. Now assume that for some r,

$$\left| \sum_{v \in V_r(z)} |v|^{-2}(1+t\{v\})^{-2} f([v+t]) - \frac{z}{\log 2(1+tz)} \right| \le C_r \lambda^r$$

Then

$$\sum_{v \in V_{r+1}(z)} |v|^{-2}(1+t\{v\})^{-2} =$$

$$\begin{aligned} &= \sum_{n=[1/z]+1}^{\infty} n^{-2} \sum_{v \in V_r} |v|^{-2}(1+\{v\}/n)^{-2}(1+t/(n+\{v\}))^{-2} f([v,n+t]) \\ &\quad + \sum_{v \in V_r \setminus V_r(\langle 1/z \rangle)} |v|^{-2}([1/z]+t+\{v\})^{-2} \\ &= \sum_{n=[1/z]}^{\infty} (n+t)^{-2} \sum_{v \in V_r} |v|^{-2}(1+\{v\}/(n+t))^{-2} f([v+t]) \\ &\quad - \sum_{v \in V_r(\langle 1/z \rangle)} |v|^{-2}([1/z]+t)^{-2}(1+\{v\}/([1/z]+t))^{-2} f([v,[1/z]+t) \end{aligned}$$

The first term here is

$$\sum_{n=[1/z]}^{\infty} (n+t)^{-2} \cdot g(1/(n+t)) + O(\lambda^r z\|f\|)$$

from Lemma 2.1, while the second term is

$$-\left([1/z]+t)^{-2} \cdot (\frac{1}{\log 2}\frac{\langle 1/z\rangle}{1+\langle 1/z\rangle/([1/z]+t)} + \Theta C_r(\lambda^r\|f\|)\right)$$

where $|\Theta| \leq 1$, on the induction hypothesis. The total thus simplifies to

$$z/(\log 2(1+tz)) + O(\lambda^r\|f\|) + \Theta C_r[1/z]^{-2}\|f\|.$$

This leads to a recurrence $C_{r+1} = C_r + O(1)$ from which it follows that $(r^{-1}C_r)$ is bounded and the weak version of the theorem follows. For the strong version, we just note that given the weak version, the strong version follows immediately since the two error terms in passing from r to $r+1$ were $O(\lambda^r z + C_r[1/z]^{-2})$. $\square$

Corollary 2.2. *Uniformly over good probability density functions f, over $0 \leq z \leq 1$, and over $0 \leq t \leq 1$,*

$$\sum_{v\in V_r, \{v\}\geq z} |v|^{-2}(1+\{v\}t)^{-2} f([v+t]) = \frac{1}{\log 2}\frac{1-z}{(1+t)(1+tz)} + O(r(1-z)\lambda^r\|f\|).$$

In view of Corollary 2.2, the conditional density for $T^r X$ given initial density f and given that $q_{r-1}/q_r(X) \leq z$ is, on the one hand, a good density, and on the other hand, $z/((1+tz)\log(1+z)) + O(r\lambda^r\|f\|)$, while the conditional density, given that $q_{r-1}/q_r(X) \geq z$, is again on the one hand a good density, and on the other, $(1-z)/((\log 2 - \log(1+z))(1+t)(1+tz)) + O(r\lambda^r\|f\|)$. Now consider a sequence k_j of positive integers such that $k_1 \geq r$ and $k_{j+1} - k_j \geq r$ for all j.

The probability $Q_{r,f}(z)$ that $\theta_r(X) \leq z$, when the initial density for X is f, is

$$\begin{aligned} Q_{r,f}(z) &= \sum_{v\in V_r} |v|^{-2}\int_{t=0}^{1} (1+\{v\}t)^{-2} f([v+t]) I(t/(1+\{v\}t) \leq z)\, dt \\ &= \int_{t=0}^{1} \sum_{v\in V_r} |v|^{-2}(1+\{v\}t)^{-2} f([v+t])(1 - I(\{v\} \leq 1/z - 1/t))\, dt \end{aligned}$$

Let $u^* := 0$ if $u < 0$, u if $0 \leq u \leq 1$, and 1 if $u > 1$. Taking $(1/z - 1/t)^*$ in place of t in Theorem 2.1 breaks into two cases. If $z \leq 1/2$ we get

$$\begin{aligned} Q_{r,f}(z) &= 1 - \frac{1}{\log 2}\left(\int_{z}^{z/(1-z)} (1/t - z/t^2)\, dt + \int_{z/(1-z)}^{1} \frac{1}{1+t}\, dt\right) + O(\lambda^r\|f\|) \\ &= \frac{z}{\log 2} + O(\lambda^r\|f\|). \end{aligned}$$

For $z > 1/2$ we get instead

$$Q_{r,f}(z) = 1 - \frac{1}{\log 2}\int_{z}^{1} (1/t - z/t^2)\, dt + O(\lambda^r\|f\|)$$

In either case, this says

$$Q_{r,f}(z) = F(z) + O(\lambda^r)\|f\| \tag{5}$$

as claimed at the outset of this section.

Next we consider the conditional density $\rho_{r,z,f,\epsilon}(t)$ of T^rX when the initial density for X is f, given that $\theta_r(X)\gtrless z$ according to whether $\epsilon=0$ or 1. This is again a good density, with norm not more than a constant multiple of $\|f\|$, (the multiple may and does depend on z) we claim.

We first note that the probability, on initial density f, that $\theta_r(X)\le z$ is at least $Cz+O(\lambda^r)$. So for fixed $z>0$, and r sufficiently large, it is at least Kz. Similarly, the probability that $\theta_r(X)\ge z$ is at least $K(1-z)$ for r sufficiently large. Next, we need some estimate of the possible growth of $\|\rho_{r,z,f,\epsilon}\|/\|f\|$ with increasing r. The scaling factor that results from dividing

$$\sum_{v\in V_r}|v|^{-2}(1+\{v\}t)^{-2}f([v+t])(1-I(\{v\}\le 1/z-1/t))$$

by $Q_{r,f}(z)$ or by $(1-Q_{r,f}(z))$, according to whether $\epsilon=0$ or 1, is at most $O(1/z)$ or $O(1/(1-z))$. Apart from this effect, we have a norm of at most

$$\sum_{v\in V_r}|v|^{-2}\|(1+\{v\}t)^{-2}f([v+t])I(\{v\}\gtrless z^{-1}-t^{-1})\|.$$

A lemma on total variation is now needed.

Lemma 2.2. *If g is a positive function of bounded variation on R and is zero outside $[0,1]$ and h is a positive, increasing function on R with $h(1)=1$, or positive and decreasing with $h(0)=1$, then the total variation of gh is no more than that of g.*

Proof. Write g as g_1-g_2 where both are zero on $(-\infty,0)$, increasing, and constant on $(1,\infty)$, so that the total variation of g is equal to $2g_1(1)=2g_2(1)$. Then $gh=g_1h-g_2h$. This gives a representation of gh as the difference of two increasing functions, both zero on $(-\infty,0)$ and both equal and constant on $(1,\infty)$. By reflecting g we see that the same holds if h is positive and decreasing with $h(0)=1$. $\square$

Now by repeated application of the lemma, and taking into account that the total variation of $f([v+t])$ is no more than that of f, we calculate that

$$\begin{aligned}(6)\qquad &\sum_{v\in V_r}|v|^{-2}\|(1+\{v\}t)^{-2}f([v+t])I(\{v\}\gtrless z^{-1}-t^{-1})\|\\ &\le\sum_{v\in V_r}|v|^{-2}\|(1+\{v\}t)^{-2}f([v+t])\|\\ &\le\sum_{v\in V_r}|v|^{-2}\|f([v+t])\|\le(\sum_{v\in V_r}|v|^{-2})\|f\|\le 2\|f\|\end{aligned}$$

From this it follows that for all $\delta\in(0,1/2)$, and whether we require $\theta_r(X)<z$ or $\theta_r(X)>z$, there exists $R(\delta)>0$ and $K(\delta)>0$ so that $\|f_{r,z}\|\le K(\delta)\|f\|$, uniformly in $\delta\le z\le 1-\delta$ for $r\ge R(\delta)$.

Now the probability, on initial density f, that $\theta_{k_1}(X)\le z$, is $F(z)+O(\lambda^r\|f\|)$. (Recall, all $k_j>r$). The conditional density of $T^{k_1}X=Y$ (say) given this event is a normalized version of the sum of all terms (with $v\in V(k_1)$) of the form $|v|^{-2}(1+\{v\}t)^{-2}I[t\le z/(1-z\{v\})]f([v+t])$, or given instead the complementary event, the same expression but with $I[t\ge z/(1-z\{v\})]$ in place of $I[t\le z/(1-z\{v\})]$.

In either case, this is a function of bounded variation. In either case, provided $z \in [\delta, 1-\delta]$, that variation is bounded by $K(\delta)\|f\|$.

Fix n large, and consider the sequence $T^{k_j}X, 1 \le j \le n$. Fix $\gamma > 0$. To satisfy a technical condition at the end of the paper, we also require that $\gamma/(F(z)+\gamma) < 1/4$. Consider the probability that more than $(F(z)+2\gamma)n$ of the j have $\theta_{k_j} \le z$. For large n, we can break this finite sequence into $O(n^{3/4})$ subsequences, each of which has $k_{j+1} > k_j + n^{3/4}$, and each of which has at most, but on the order of, $(n^{1/4})$ entries. This way, $r = n^{3/4}$ is comfortably larger than the number of trials $O(n^{1/4})$ in a subsequence. The initial density f_0 is the density of $T^k 1$, where k is the least of the k_j in our subsequence. For each such subsequence of length N say (N depends on the subsequence), the *event* $E(\epsilon_1, \dots \epsilon_N)$, where $(\epsilon_1, \epsilon_2, \dots \epsilon_N) \in \{0,1\}^N$, is the set of all $x \in [0,1)\backslash Q$ for which $\theta_{k_j} < z$ if and only if $\epsilon_j = 0$, $1 \le j \le N$.

Let $r_j := k_{j+1} - k_j$, with $r_0 := k_1$. A *step* consists of replacing f_j with $f_{j+1} := \rho_{r_j,z,f_j,\epsilon_{j+1}}$, the conditional density, given that $\theta_{k_j}(Y) \gtrless z$, of $T^{r_j}Y$ where Y is a random variable on U with density f_j. Thus the input to the next step is again a good density, with a certain norm. The norm of the 'working' f_j may increase, by at most a factor of $K(\delta)$ each time. If the number N of steps is small compared to the minimum r_j, this is not much of a problem, because at each stage, the working 'initial' probability density function f_j has norm no greater than K^j. The probability, at each trial within the subsequence, and for any prior history of trials within that subsequence, that $\theta_{k_j} > z$, is less than $F(z)+\gamma$. That is, the conditional probability that $\theta_{k_{j_m}} > z$, given that $\theta_{k_{j_l}} < z$ exactly when $\epsilon_l = 0$ for $1 \le l \le m$, is less than $F(z)+\gamma$.

(We take n large enough that $K(\delta)^{n^{1/4}}\lambda^{n^{3/4}} < \gamma/2$). The probability that a particular subsequence has more than its own length, multiplied by $F(z)+2\gamma$, cases of $\theta_{k_j} > z$, can be shown (see below) to be bounded above by $O(\exp(-C\gamma^2 n^{1/4}))$. Keeping in mind that this claim has yet to be established, we continue with the main line of the proof.

The probability that any one of the subsequences has such an atypical success ratio, is $O(n^{3/4}\exp(-C\gamma^2 n^{1/4}))$ which tends to zero. This shows that the probability of an atypically high success ratio is asymptotically zero. The same arguments apply to the case of an atypically low success ratio, simply by redefining success to mean $\theta_{k_j} < z$. This proves that Nair's theorem[1] holds for any strictly increasing sequence (k_j) of positive integers, as claimed.

3. Non-independent trials are good enough.

'The probability that a particular subsequence has more than its own length, multiplied by $F(z)+2\gamma$, cases of $\theta_{k_j} > z$, can be shown (see below) to be bounded above by $O(\exp(-C\gamma^2 n^{1/4}))$.' We now make good on this claim. *For the remainder of this section, we shall use n to mean the number of trials. This new value of n will be on the order of the old value of $n^{1/4}$.*

We have a kind of near-independence: If

$$E := E(k_1, k_2, \dots k_n, \epsilon_1, \epsilon_2, \dots \epsilon_n) = \{x \in [0,1]\backslash Q : \theta_{k_j}(x) < z \text{ iff } \epsilon_j = 0\} \tag{7}$$

then the conditional probability that $\theta_{k_{n+1}} < z$ given that $x \in E$ is $F(z) + O(\lambda^r)$ and so less than $F(z)+\gamma$. Thus if (a_0, a_1) is the sequence

$$(a_0, a_1) := (\text{prob}(\theta_{k_1}(x) < z), \text{prob}[\theta_{k_1}(x) > z])$$

[1] See http://nyjm.albany.edu:8000/j/1998/3A-9.html.

and (b_0, b_1) the sequence $(1-\gamma-F(z), F(z)+\gamma)$, then $a_0 > b_0$ and $a_0 + a_1 = 1 = b_0 + b_1$.

Given two sequences $(a_0, a_1, \dots a_n)$ and $(b_0, b_1, \dots b_n)$ of non-negative numbers summing to 1, we say that $a :> b$ if for all $k < n$, $\sum_0^k a_j > \sum_0^k b_j$.

Lemma 3.1. *If $a :> b$, if (u_k) and (v_k) are sequences of numbers in $[0,1]$ with $u_k < v_k$ for all k, and if a' is given by $a'_j = (1-u_j)a_j + u_{j-1}a_{j-1}$ and b' by $b'_j = (1-v_j)b_j + v_{j-1}b_{j-1}$, (setting $a_{-1} = b_{-1} := 0$), then $a' :> b'$.*

Proof. We have

$$\begin{aligned}\sum_{j=0}^{k} a'_j - b'_j &= \sum_{j=0}^{k} (1-u_j)a_j + u_{j-1}a_{j-1} - (1-v_j)b_j - v_{j-1}b_{j-1} \\ &= (1-v_k)\sum_0^k (a_j - b_j) + v_k \sum_0^{k-1}(a_j - b_j) + a_k(v_k - u_k) > 0.\end{aligned}$$

□

We apply this lemma with $a = (a_0, a_1, \dots a_n)[k_1, k_2, \dots k_n, \epsilon_1, \epsilon_2, \dots \epsilon_n, z]$, defined by

$$\begin{aligned} a_m &:= \text{prob}[\#\{j : 1 \le j \le n \text{ and } \theta_{k_j} < z\} = m] \text{ for } 0 \le m \le n, \text{ and} \\ b &:= (b_0, b_1, \dots b_n), \text{ where} \\ b_m &:= \binom{n}{m}(F(z)+\gamma)^m (1-F(z)-\gamma)^{n-m}, \ 0 \le m \le n.\end{aligned}$$

The claim is that with this a and b, $a :> b$. The proof is inductive, using Lemma 3.1 n times.

We shall be using a succession of a's and b's, which will be denoted by superscripts.

$$\begin{aligned} a^1 &:= (a_0^1, a_1^1) = (\text{prob}[\theta_{k_1} \ge z], \text{prob}[\theta_{k_1} < z]), \text{ while} \\ b^1 &:= (1-F(z)-\gamma, F(z)+\gamma).\end{aligned}$$

We have $a^1 :> b^1$ because $\text{prob}[\theta_{k_1} \ge z] > 1 - F(z) - \gamma$. This so far uses only the definition of :> and earlier material but not Lemma 3.1

Now let $a_0^2 := (\text{prob}[\theta_{k_1} \ge z \text{ and } \theta_{k_2} \ge z]$, $a_1^2 := \text{prob}[\text{one small theta}]$, and $a_2^2 := \text{prob}[\theta_{k_1} \le z \text{ and } \theta_{k_2} \le z])$,

$$\begin{aligned} a^2 &:= (a_0^2, a_1^2, a_2^2), \text{ and} \\ b^2 &:= ((1-F(z)-\gamma)^2, 2(1-F(z)-\gamma)(F(z)+\gamma), (F(z)+\gamma)^2).\end{aligned}$$

We take $a = a^1$, $b = b^1$, $a' = a^2$, and $b' = b^2$ in Lemma 3.1. Then we have

$$a'_0 = (1-u_0)a_0, \ a'_1 = (1-u_1)a_1 + u_0 a_0, \ a'_2 = (1-u_2)\cdot 0 + u_1 a_1$$

where $u_0 = \text{prob}[\theta_{k_2} < z \text{ given } \theta_{k_1} \ge z]$ and $u_1 = \text{prob}[\theta_{k_2} < z \text{ given } \theta_{k_1} < z]$, while $v_0 = v_1 = F(z) + \gamma$. Applying Lemma 3.1 gives $a_2 :> b_2$. Inductively, it gives $a^n :> b^n$ which says that $a :> b$.

Thus the probability that more than $n(F(z)+2\gamma)$ cases of $\theta_{k_j} < z$ out of the first n values of θ_{k_j} is less than the probability, with a Bernoulli process which gives 'heads' with probability $F(z)+\gamma$, that more than $n(F(z)+2\gamma)$ of the first n trials come up heads.

By standard exponential centering, this probability is, we claim, less than $\exp(-(3/8)n\gamma^2)$. Let Y_j be independent Bernoulli trials with a coin taking heads with probability $\alpha = F(z)+\gamma$. (If $F(z)+\gamma > 1$ the probability in question is zero.) Now for $\zeta > 0$,

$$\begin{aligned}\mathrm{Prob}[\sum_{j=1}^{n} Y_j \geq n(\alpha+\gamma)]& \\ &\leq e^{-\zeta n(\alpha+\gamma)} \sum_{k\geq n(\alpha+\gamma)} e^{k\zeta}\mathrm{Prob}[\sum_{j=1}^{n} Y_j = k] \\ &\leq e^{-\zeta n(\alpha+\gamma)} \sum_{k=0}^{n} e^{k\zeta}\mathrm{Prob}[\sum_{j=1}^{n} Y_j = k] = e^{-\zeta n(\alpha+\gamma)}(1-\alpha+\alpha e^{\zeta})^n\end{aligned}$$

We take $\zeta = \log(1+\gamma/\alpha)$ and recall that $\gamma/\alpha < 1/4$. Thus

$$\mathrm{Prob}[\sum_{j=1}^{n} Y_j \geq n(\alpha+\gamma)] \leq (1+\gamma/\alpha)^{-n(\alpha+\gamma)}e^{n\gamma}.$$

Using the first two terms in the series expansion of $\log(1+\gamma/\alpha)$, this is less than $\exp(-(3/8)n\gamma^2/\alpha) < \exp(-n\gamma^2/4)$. This completes the proof of the theorem.

References

[BJW83] W. Bosma, H. Jager, and F. Wiedijk, *Some metrical observations on the approximation by continued fractions*, Indag. Math. **45** (1983), 281–299, MR 85f:11059.

[Hen92] D. Hensley, *Continued fraction Cantor sets, Hausdorff dimension, and functional analysis*, Journal of Number Theory **40** (1992), no. 3, 336–358, MR 93c:11058.

[Nai98] R. Nair, *On metric diophantine approximation theory and subsequence ergodic theory*, New York Journal of Mathematics **3A** (1998), 117–124.

[Val95] B. Vallée, *Méthodes d'analyse fonctionelle dans l'analyse en moyenne des algorithmes d'Euclide et de Gauss*, C.R. Acad. Sci. Paris **321** (1995), 1133–1138, MR 97c:58088.

Department of Mathematics, Texas A&M University, College Station, TX 77843
Doug.Hensley@math.tamu.edu http://www.math.tamu.edu/~doug.hensley/

This paper is available via http://nyjm.albany.edu:8000/j/1998/4-16.html.