TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 306, Number 1, March 1988

A TRUNCATED GAUSS-KUZMIN LAW

DOUG HENSLEY

ABSTRACT. The transformations T_n which map $x \in [0, 1)$ onto 0 (if $x \leq 1/(n+1)$), and to $\{1/x\}$ otherwise, are truncated versions of the continued fraction transformation $T\colon x \to \{1/x\}$ (but $0 \to 0$).

An analog to the Gauss-Kuzmin result is obtained for these T_n, and is used to show that the Lebesgue measure of $T_n^{-k}\{0\}$ approaches 1 exponentially. From this fact is obtained a new proof that the ratios ν/k, where ν denotes any solution of $\nu^2 \equiv -1 \bmod k$, are uniformly distributed mod 1 in the sense of Weyl.

1. Introduction. The continued fraction algorithm is based on iteration of the transformation

$$T\colon x \to \{1/x\} \quad (x \neq 0), \qquad T\colon 0 \to 0$$

of the unit interval $[0, 1)$. The Gauss-Kuzmin result is that for a random variable X uniformly distributed on $[0, 1]$, the density of $T^k X$ tends to $g(t) = 1/(1+t)\log 2$, $0 \leq t \leq 1$.

The associated measure μ, determined by $\mu(a, b) = \int_a^b (g(t)\,dt$, is invariant with respect to T. That is, $\mu(T^{-1}E) = \mu(E)$ for all measurable $E \subseteq [0, 1]$. There is a considerable body of knowledge about this transformation and various related topics, such as the Jacobi-Perron algorithm. Here we mention a few of the salient points:

(1) $\bigcup_{k=1}^{\infty} T^{-k}\{0\} = \mathbf{Q} \cap [0, 1)$.

(2) T is ergodic with respect to μ.

(3) The convergence of the density $g^k(x)$ is uniform and rapid, in the sense that there exists a constant c, $0 < c < 1$, such that for all $k \geq 1$ and all t, $0 \leq t \leq 1$, $|g^k(t) - g(t)| \leq c^k$.

(4) The arithmetic mean of the partial quotients $a_k(x) := [1/T^{k-1}x]$ is infinite almost everywhere, but the geometric mean has a certain finite value a.e. These facts are well known among specialists, and are found in most standard references. See e.g. [**7**, **8**].

From (4) it follows that (m denoting Lebesgue measure) $m(\{x\colon a_k(x) \leq n$ for all $k\}) = 0$, so that $\lim_{k\to\infty} m(\{x\colon a_j(x) \leq n$ for $1 \leq j \leq k\}) = 0$.

One of our concerns here is to find out how rapidly this quantity approaches zero as $k \to \infty$. Another is to determine the asymptotic conditional density of $T^k X$ given that $T^j X > 1/(n+1)$ for $0 \leq j \leq k$. To this end we introduce the

Received by the editors November 14, 1986.

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 11A55; Secondary 11H41, 11K36.

transformations T_n mentioned in the abstract:

$$T_n(x) = \{1/x\} \qquad (1/(n+1) < x < 1),$$
$$T_n(x) = 0 \qquad (0 \le x \le 1/(n+1)).$$

Let $\mu_n(k) = 1 - m(T_n^{-k}\{0\}) = m(\{x\colon a_j(x) \le n \text{ for } 1 \le j \le k\})$. We obtain these results first:

THEOREM 1. *For each $n \ge 1$ there exists $\lambda(n)$, $0 < \lambda(n) < 1$, and $g_n(t)$, a continuous decreasing, positive, convex probability density function on* $[0,1]$, *such that for all $k \ge 1$, $\frac{1}{2}\mu_n(k) \le \lambda(n)^{k+1} \le 2\mu_n(k)$, $|\mu_n(k+1)/\mu_n(k) - \lambda(n)| \le 10(19/20)^k$, and $\mu_n(2k+2)/\mu_n(2k+1) < \lambda(n) < \mu_n(2k+1)/\mu_n(2k)$. Moreover, $\lambda(n+1) > \lambda(n)$ for all $n \ge 1$, and* $\lim_{n\to\infty} n(1-\lambda(n)) = 1/\log 2$.

Let L_n denote the linear functional

$$(L_n f)(t) = \sum_{k=1}^{n} (k+t)^{-2} f\left(\frac{1}{k+t}\right).$$

This is a truncation of

$$L_\infty f = \sum_{k=1}^{\infty} (k+t)^{-2} f\left(\frac{1}{k+t}\right),$$

which gives the density of TX if X has density f. Note that $L_\infty(g) = g$.

THEOREM 2. *The function $g_n(t)$ satisfies the condition $\lambda(n)g_n(t) = L(g_n(t))$. (Partial statement—the rest must await the introduction of further notation.) Basically, this $g_n(t)$ is a convex combination of functions $(1+\theta)/(1+\theta t)^2$ with $0 \le \theta \le 1$, as is the Gauss-Kuzmin density $g(t)$. The difference is that for $g_n(t)$ the set of θ's involved is restricted to those for which $a_k(\theta) \le n$ for all k, a set of measure zero, while $g(t) = \int_0^1 [(1+\theta)/(1+\theta t)^2]\,d\theta$.*

THEOREM 3. *Let $\nu_n(A) = \int_A g_n(t)\,dt$. Then for all Lebesgue-measurable A not containing zero, $\nu_n(T_n^{-1}A) = \lambda(n)\nu_n(A)$.*

The final result is a weaker version of the already known fact, due to Hooley, that (ν/k) is uniformly distributed mod 1 if we average over all $k \le x$ and all solutions $\nu \bmod k$ of $\nu^2 \equiv -1 \bmod k$ [**2**, **3**, **4**]. This is known for general nonsquare D in place of -1, but the case of $D = -1$ serves as a paradigm for all negative values of D.

Hooley's proof is based on the clever use of some deep results about Kloosterman sums. The proof we sketch here has its roots in a relatively simple lemma about the last continued fraction convergent c/d to a rational number a/b, other than a/b itself. Let $A(y)$ denote $\{(a,b)$ such that g.c.d.$(a,b) = 1$ and $a^2 + b^2 \le y^2\}$, and let $\sigma^*(a,b) = \sqrt{c^2+d^2}/\sqrt{a^2+b^2}$ where (c,d) satisfies $ad - bc = 1$, $c^2 + d^2 \le a^2 + b^2$, $ac + bd \ge 0$. There are two ways to write a/b as a continued fraction: $[a_1, a_2 \cdots a_n]$ or $[a_1, a_2 \cdots a_n - 1, 1]$.

The next-to-last convergents, $[a_1, a_2 \cdots a_{n-1}]$ and $[a_1, a_2 \cdots a_n - 1]$, are then equal to (c/d) and $(a-c)/(b-d)$ in one order or the other. These each give a solution to $\nu^2 \equiv -1 \bmod a^2 + b^2$: $\nu = ac + bd$ and $\nu = a(a-c) + b(b-d)$. Since (a,b) and (c,d) are nearly parallel, $\nu/(a^2+b^2)$ and $1 - \nu/(a^2+b^2)$ are well approximated by $\sigma^*(a,b)$ and $1 - \sigma^*(a,b)$, in one or the other order. Thus the question of the distribution of ν/k, $\nu^2 \equiv -1 \bmod k$, is related to the question of the distribution of $\sigma^*(a,b)$, averaged over all pairs of relatively prime integers (a,b) with $a^2+b^2 \le y^2$, say. The key to this approach is the fact that

$$\sum_{\{(a,b)\in A(y):\ \sigma^*(a,b)\le t\}} 1 = \frac{6}{\pi} t y^2 + O(t^2y^2) + O(y).$$

The distribution of ν/k is thus seen to be, at any rate, asymptotically uniform in a neighborhood of zero.

For $t = 1/(n+1)$ and n large, this estimate can be fed into a continued-fraction machinery to yield estimates of the proportion of ν/k between $[a_1, a_2 \cdots a_j]$ and $[a_1, a_2 \cdots a_j, n+1]$, still with an accuracy of about 1 part in n, provided a_1, $a_2 \cdots a_j$ are all less than $n+1$.

Knowing, as we do, the extent to which such intervals fill up $[0,1]$, the rest is just a matter of judicious balancing of various error terms. This approach also gives estimates for the distribution of $\nu/(a^2+b^2)$ averaged over pairs (a,b) confined to a wedge $\theta_1 \le \arg(a+ib) \le \theta_2$, and with $a^2+b^2 \le y^2$; it is still uniform. But I do not see any reason why Hooley's approach cannot be applied to this question, with a more accurate error bound, so the reader will be spared the details. For the record, the result one gets with this approach is that if χ denotes the characteristic function of a statement, then

$$\sum_{(a,b)\in A(x)} \chi(\sigma^*(a,b)\in[s,t] \text{ and } \theta_1 \le \arg(a+ib) \le \theta_2)$$
$$= 3(\theta_2-\theta_1)(t-s)x^2/\pi^2 + O(x^2(\theta_2-\theta_1)(\log\log x)^2/\log x) + O(x^{13/8}),$$

uniformly in $0 \le s < t \le 1$ and $0 \le \theta_1 < \theta_2 \le 2\pi$.

From this one can deduce that uniformly over $d \le x^{1/5}$ and over all classes H of the class group of $Q(\sqrt{-D})$, the distribution of $k(\mathfrak{a})/\mathrm{Norm}(\mathfrak{a})$ is uniform when averaged over all ideals $\mathfrak{a}$ in H and of norm $\le x$, as $x \to \infty$, where $k(\mathfrak{a})$ denotes the integer k, $0 \le k < \mathrm{Norm}\,\mathfrak{a}$ such that $k \equiv \sqrt{-D} \bmod \mathfrak{a}$.

Hooley does not work out, in [**2**], any error term for uniformity of distribution. But a straightforward application of the discrepancy theorem of Erdös and Turan gives the estimate (say for $D=1$),

$$\sum_{(a,b)\in A(x)} \chi(\sigma^*(a,b)\in[s,t]) = (6/\pi)(t-s)x^2 + O(x^{3/2}\exp(3\sqrt{\log x})).$$

Iwaniec gets a more general estimate which includes this as a special case, in Lemma 4 of [**4**].

2. Farey *n*-intervals. Fix n, and let $V(k) = \{(v_1, v_2, \ldots, v_k)\colon\ \le v_i \le n$ for $1 \le j \le k\}$. For $v \in V(k)$, let $I(v) = \{r\colon r = 1/v_1 + 1/v_2 + \ldots + 1/(v_k+\lambda)$ for some λ, $0 \le \lambda \le 1/(n+1)\}$. Here the notation $1/v + 1/w$ refers to the continued fraction $1/(v+1/w)$, and we shall abbreviate this to $[v,w]$. Thus $I(v) = \{r\colon r = [v_1, v_2, \ldots, v_k+\lambda]$ with $0 \le \lambda \le (n+1)^{-1}\}$. We include an *empty vector* for $V(0)$, with corresponding interval $I_0 := [0, 1/(n+1)]$. If $v \in V(k)$ we say that the rank of $I(v)$ is k.

If $n=2$, for instance, the intervals $I(v)$ of rank 0, 1, and 2 are $[0,1/3]$; $[3/4,1]$, $[3/7,1/2]$; and $[2/3,7/10]$, $[1/3,4/11]$, $[2/5,6/17]$ and $[1/2,4/7]$. For any fixed n, all the $I(v)$, taken over all $k \geq 0$ and all $v \in V(k)$, are disjoint. The only real numbers r, $0 \leq r \leq 1$ *not* belonging to any $I(v)$ are those with a continued fraction expansion $r = [w_1, w_2, w_3, \dots]$ of infinite length, such that $w_i \leq n$ for all $i \geq 1$. While there are uncountably many such numbers, it is intuitively obvious that they make up a set of Lebesgue measure zero. Later we shall prove that $R_n(k) :=$ the complement in $[0,1]$ of the union of all Farey n-intervals of rank $< k$, which is, for any finite k, a union of finitely many open intervals, has Lebesgue measure $\mu_n(k)$ which tends to zero more rapidly than $(1 - 1/(n+1))^k$ as $k \to \infty$. If we identify 0 with 1 in the unit interval to form a circle, topologically, then the mapping $T\colon [0,1] \to [0,1]$, $T(r) = \{1/r\}$ sends every interval $I(v)$ for $v \in V(k)$, $k \geq 1$, onto an interval $I(w)$ for $w \in V(k-1)$. For $k \geq 1$, and $v \in V(k)$, the restriction of T to $I(v)$ is one-to-one, continuous and differentiable, and $-(n+1)^2 \leq T'(r) \leq -1$. There are n^k elements in $V(k)$, and the preimage of each interval $I(w)$ for $w \in V(k-1)$ consists of the n intervals $I((1, w_1, \dots, w_{k-1}))$, $I((2, w_1, \dots, w_{k-1})), \dots, I((n, w_1, w_2, \dots, w_{k-1}))$. We shall abbreviate $(j, w_1, w_2, \dots, w_l)$ to (jw) from now on, and take (w, j) to be $(w_1, w_2, \dots, w_l, j)$. Extending this, we put $(v, w) = (v_1, v_2, \dots, v_j, w_1, w_2, \dots, w_l)$ for $v \in V(j)$ and $w \in V(l)$. Also, if $v = (v_1, \dots, v_j)$, put $v^- = (v_1, v_2, \dots, v_{j-1})$, and $v_- = (v_2, v_3, \dots, v_j)$. Every Farey interval $I(v)$ for $v \in V(k)$, $k \geq 1$, has either the form

$$\left[\frac{c}{d}, \frac{(n+1)c + c'}{(n+1)d + d'}\right] \quad \text{or} \quad \left[\frac{(n+1)c + c'}{(n+1)d + d'}, \frac{c}{d}\right],$$

where $cd' - c'd = -1$ in the former case and $+1$ in the latter. The former case occurs if and only if k is even. A truncated Farey n-interval $I_\Lambda(v)$ will be defined as $\{r\colon r = [v_1, v_2, \dots, v_k + \lambda]$ with $\Lambda_1 \leq \lambda \leq \Lambda_2\}$, where $\Lambda_i \leq 1/(n+1)$. These intervals behave just like the $I(v)$'s with respect to T.

Clearly any interval $[s,t] \subseteq [0,1]$ can be largely covered by Farey n-intervals $I(v) \subseteq [s,t]$ of rank $\leq k$, together with one or two truncated n-intervals perhaps, and leaving an uncovered remnant of measure $\leq \mu_n(k)$.

3. The remnant: numbers r not captured in the Farey n-intervals of small rank. Recall that $R_n(k) = \{r\colon 0 \leq r \leq 1$ and r does not belong to any Farey n-interval of rank $\leq k\}$. Equivalently,

$$(1) \quad R_n(k) = \{r \in [0,1]\colon r = [v_1, v_2, \dots, v_k + \lambda] \\ \text{with } 1/(n+1) < \lambda < 1, \text{ and } 1 \leq v_i \leq n \text{ for } 1 \leq i \leq k\},$$

and also equivalently,

$$(2) \quad R_n(k) = \{r \in [0,1]\colon r = [v_1, v_2, \dots, v_k, v_{k+1}, \rho], \\ \text{with } 1 \leq v_i \leq n \text{ for } 1 \leq i \leq k+1, \text{ and } 0 < \rho < 1\}.$$

(Neither definition covers $R_n(0) = (1/(n+1), 1)$.) Let $\mu_n(k)$ be the Lebesgue measure of $R_n(k)$, that is, the sum of the lengths of the intervals which comprise $R_n(k)$.

An example is in order: $n = 4$, $k = 0, 1$, and 2. $R_4(0) = (1/5, 1)$,

$$R_4(1) = (1/5, 5/21) \cup (1/4, 5/16) \cup (1/3, 5/11) \cup (1/2, 5/6),$$

and

$$\begin{aligned} R_4(2) = {} & (6/29, 2/9) \cup (11/49, 3/13) \cup (16/69, 4/17) \cup (21/89, 5/21) \cup (6/23, 2/7) \\ & \cup (11/38, 3/10) \cup (16/53, 4/13) \cup (21/68, 5/16) \cup (6/17, 2/5) \\ & \cup (11/27, 3/7) \cup (16/37, 4/9) \cup (21/47, 5/11) \cup (6/11, 2/3) \cup (11/16, 3/4) \\ & \cup (16/21, 4/5) \cup (21/26, 5/6). \end{aligned}$$

Doing the arithmetic, one finds that (to within the implicit accuracy of the displayed number of digits)

$$\mu_4(0) = .80000000, \quad \mu_4(1) = .55514069, \quad \mu_4(2) = .40742855,$$

and further similar calculations yield

$$\mu_4(3) = .29443213, \quad \mu_4(4) = .21382442, \quad \mu_4(5) = .15505299.$$

This is increasingly like a geometric sequence as k increases. The ratio of successive $\mu_n(k)$ seems to tend to about 0.725. Other examples with different choices of n give heuristic confirmation.

In this section, we develop a body of information about $R_n(k)$ and some associated functions and measures. For purposes of the application to uniform distribution, we only need the result that $\mu_n(k)$ decreases exponentially for each n, with a limiting ratio $\mu_n(k+1)/\mu_n(k) \to \lambda(n)$ as $k \to \infty$, that $\lambda(n)$ is increasing in n, and that $\lim_{n\to\infty} n(1-\lambda(n)) = 1/\log 2$.

The proofs are based on an analysis of the linear functional L_n, $L_n(f(t)) := \sum_{k=1}(k+t)^{-2} f(1/(k+t))$, and the effects of high order iterates of L_n on the initial function which is constant at 1 for $0 \le t \le 1$.

We begin by establishing some terminology. From now on, most of the time n will be fixed, and will be relegated to the background. Thus if the context establishes n, I will write $L(f(t)) = \sum_{k=1}^{n}(k+t)^{-2} f(1/(k+t))$, instead of $L_n(\cdots)$.

The (nonlinear) operator $S = S_n$ from the set p of positive, continuous functions on $[0,1]$ into the same set, is defined by

$$Sf(t) = \sum_{k=1}^{n}(k+t)^{-2} f\left(\frac{1}{(k+t)}\right) \Big/ \int_{1/(n+1)}^{1} f(t)\,dt. \tag{1}$$

Since

$$\int_{1/(n+1)}^{1} f(t)\,dt = \int_0^1 \sum_{k=1}^{n}(k+t)^{-2} f\left(\frac{1}{(k+t)}\right),$$

$$\int_0^1 Sf(t)\,dt = 1 \quad \text{for all } f \in \mathcal{P}. \tag{2}$$

S is simply a renormalized version of L. That is,

$$Sf = Lf \Big/ \int_0^1 Lf \tag{3}$$

and by iteration,

$$S^k f = L^k f \Big/ \int_0^1 L^k f.$$

The key lemma is that $S^k \chi_{[0,1]}(t)$ converges to a function $g(t) = g_n(t)$, and that $S(g(t)) = g(t)$. From this, we eventually obtain these theorems:

THEOREM 1. *For each* $n \geq 1$ *there exists* $\lambda(n)$, $0 < \lambda(n) < 1$, *and* $g_n(t)$, *continuous, decreasing, positive and convex on* $[0,1]$, *such that for all* $k \geq 1$,

$$\tfrac{1}{2}\mu_n(k) \leq \lambda(n)^{k+1} \leq 2\mu_n(k), \quad |\mu_n(k+1)/\mu_n(k) - \lambda(n)| \leq 10(19/20)^k,$$

and

$$\mu_n(2k+2)/\mu_n(2k+1) < \lambda(n) < \mu_n(2k+1)/\mu_n(2k).$$

For all $n \geq 1$, $\lambda(n+1) > \lambda(n)$, *and* $\lim_{n\to\infty} n(1-\lambda(n)) = 1/\log 2$.

THEOREM 2. *The function* $g_n(t)$ *satisfies the condition* $\lambda(n)g_n(t) = L(g_n(t))$. *For all initial functions* $\psi_\theta^0(t) := (1+\theta)/(1+\theta t)^2$, *with parameter* $0 \leq \theta \leq 1$, $\|S_n^k \psi_\theta^0(t) - g_n(t)\|_\infty \leq 10(19/20)^k$. *There is a probability measure* β_n, *concentrated on irrational numbers* $\alpha \in [0,1]$ *such that in the continued fraction expansion of* α *as* $[a_1, a_2, a_3, \ldots]$, *all* $a_i \leq n$, *so that*

$$g_n(t) = \int_{\theta=0}^{1} \frac{1+\theta}{(1+\theta t)^2}\, d\beta_n(\theta) = \int_{\theta=0}^{1} \psi_\theta^0(t)\, d\beta_n(\theta).$$

THEOREM 3. *Let* $T_n\colon [0,1] \to [0,1)$,

$$T_n(x) = \{1/x\} \quad \textit{if } x \geq 1/(n+1), \quad \textit{else } 0.$$

Let $\nu_n(A) = \int_A g_n(t)\, dt$. *Then for all Lebesgue-measurable* A *not containing zero,*

$$\nu_n(T_n^{-1}(A)) = \lambda(n)\nu_n(A).$$

REMARK. That is, T_n is a measure-decimating transformation. Since $\nu_n(A)$ differs from the Lebesgue measure of A by at most a factor of 2, larger or smaller, this result also gives $\mu_n(k-1)$ between $\frac{1}{2}\lambda(n)^k$ and $2\lambda(n)^k$.

A superficially attractive speculation is that the Bernoulli shift operator

$$T^*(a_1, a_2, \ldots) = (a_2, a_3, \ldots),$$

which is related to T in an obvious way, gives an alternative description of ν_n by way of, say, fixing ν_n^* on cylinders, with $\nu_n^*(a_i = j) = \int_{1/(j+1)}^{1/j} g_n(t)\, dt$.

As it happens, this does not work. The measure on $[0,1]$ corresponding to ν_n^* on sequences $(a_1, a_2, \ldots)$, is grainy, while ν_n has a smooth density. This other measure does satisfy much the same recursion as does ν_n, which shows that the proof of Theorem 3 will have to use some argument specific to the starting values for iteration of S.

Now that the results are stated, we relegate n to the background. We have already defined $\psi_\theta^0(t) = (1+\theta)/(1+\theta t)^2$. Now let $\phi_\theta^0(t) = 1/(1+\theta t)^2$, and for $r \geq 1$ let

$$\phi_\theta^r(t) = L^r \phi_\theta^0(t), \quad \psi_\theta^r(t) = S^r \psi_\theta^0(t) = \left(\phi_\theta^r(t) \Big/ \int_0^1 \phi_\theta^r(t)\, dt\right). \tag{4}$$

Recall that $V_n(r) = \{1, 2, \ldots, n\}^r$, $(v, \theta) = (v_1, v_2, \ldots, v_{r-1}, v_r + \theta)$ if

$$v = (v_1, v_2, \ldots, v_r),$$

and $\langle w \rangle$ is the denominator of the continued fraction $[w] = 1/w_1 + 1/w_2 + \cdots + 1/w_r$. (Alternatively, $\langle w \rangle = d_r$ if $d_0 = 1$, $d_1 = w_1$ and $d_i = w_i d_{i-1} + d_{i-1}$ for $2 \leq i \leq r$.)

LEMMA 1. *For all* $r \geq 1$,

$$\phi_\theta^r(t) = \sum_{v \in V_n(r)} \langle v, \theta \rangle^{-2} \phi^0_{[v,\theta]}(t).$$

COROLLARY.

$$\phi_0^r(t) = \sum_{v \in V_n(r)} (\langle v \rangle + t\langle v^- \rangle)^{-2}, \quad \textit{where } (v^-) = (v_1, v_2, \ldots, v_{r-1}).$$

PROOF. We verify the lemma directly for $r = 1$, where

$$\begin{aligned} \phi_\theta^1(t) &= \sum_{k=1}^n (k+t)^{-2} \phi_\theta^0 \left(\frac{1}{k+t} \right) \\ &= \sum_{k=1}^n (k+t)^{-2} \left(1 + \frac{\theta}{k+t} \right)^{-2} \sum_{k=1}^n (k+\theta)^{-2} \phi^0_{1/(k+\theta)}(t) \end{aligned} \tag{5}$$

as required.

Now assume the lemma holds for r. Then

$$\phi_\theta^r(t) = \sum_{v \in V_n(r)} \langle v, \theta \rangle^{-2} \phi^0_{[v,\theta]}(t), \quad \text{and} \tag{6}$$

$$\phi_\theta^{r+1}(t) = \sum_{v \in V_n(r)} \langle v, \theta \rangle^{-2} \sum_{k=1}^n (k+t)^{-2} \phi^0_{[v,\theta]} \left(\frac{1}{k+t} \right).$$

Now

$$\begin{aligned} \langle v, \theta \rangle^{-2} (k+t)^{-2} \phi^0_{[v,\theta]} \left(\frac{1}{k+t} \right) &= \frac{1}{(k+t)^2} \left(\frac{1}{1 + [v,\theta]/(k+t)} \right)^2 \langle v, \theta \rangle^{-2} \\ &= \frac{1}{(k+t+[v,\theta])^2 \langle v, \theta \rangle^2}. \end{aligned} \tag{7}$$

But

$$[v, \theta] = \langle v_2, v_3, \ldots, v_r + \theta \rangle / \langle v_1, v_2, \ldots, v_r + \theta \rangle$$

so that

$$(k + t + [v, \theta]) \langle v, \theta \rangle = (k \langle v, \theta \rangle + \langle v_2, \ldots, v_r + \theta \rangle + t \langle v, \theta \rangle). \tag{8}$$

Now $k\langle v_1, v_2, \ldots, v_r + \theta \rangle + \langle v_2, \ldots, v_r + \theta \rangle = \langle k, v_1, v_2, \ldots, v_r + \theta \rangle$, so that with $w = (k, v_1, v_2, \ldots, v_r)$,

$$(k + t + [v, \theta]) \langle v, \theta \rangle = \langle w, \theta \rangle (1 + [w, \theta] t). \tag{9}$$

Hence

$$\phi_\theta^{r+1}(t) = \sum_{v \in V_n(r)} \sum_{k=1}^n \langle k, v_1, \ldots, v_r + \theta \rangle^{-2} \phi^0_{[k, v_1, \ldots, v_r + \theta]}(t). \tag{10}$$

LEMMA 2. *For all* $r \geq 0$

$$\psi_\theta^{r+1}(t) = \sum_{k=1}^n \gamma_k^r(\theta) \psi^r_{1/(k+\theta)}(t),$$

where

$$\gamma_k^r(\theta) = \frac{\sum_{v\in V_n(r)} \langle v, k+\theta\rangle^{-1}\langle 1+v_1, v_2, \dots, v_r, k+\theta\rangle^{-1}}{\sum_v \in V_n(r) \sum_{l=1}^n \langle v, l+\theta\rangle^{-1}\langle 1+v_1, \dots, v_r, l+\theta\rangle^{-1}}$$

For $r = 0$, this should be interpreted as

$$\gamma_k^0(\theta) = \frac{(k+\theta)^{-1}(k+1+\theta)^{-1}}{\sum_{l=1}^n (l+\theta)^{-1}(l+1+\theta)^{-1}}.$$

PROOF OF LEMMA 2. There is a constant $C_r > 0$ such that

$$\psi_\theta^{r+1}(t) = C_r \sum_{v\in V_n(r)} \sum_{k=1}^n \langle v_1, \dots, v_r, k+\theta\rangle^{-2} \phi^0_{[v,k+\theta]}(t). \tag{11}$$

Now $\psi^0_{[v,k+\theta]}(t) = (1 + [v, k+\theta])\phi^0_{[v,k+\theta]}(t)$, so

$$\begin{aligned} (12)\quad & \langle v_1, v_2, \dots, v_r, k+\theta\rangle^{-2}\phi^0_{[v,k+\theta]}(t) \\ &= \langle v_1, \dots, v_r, k+\theta\rangle^{-1}(\langle v_1, \dots, v_r, k+\theta\rangle(1+[v, k+\theta]))^{-1}\psi^0_{[v,k+\theta]}(t) \\ &= \langle v_1, \dots, v_r, k+\theta\rangle^{-1}\langle 1+v_1, v_2, \dots, v_r, k+\theta\rangle^{-1}\psi^0_{[v,k+\theta]}(t). \end{aligned}$$

Thus

$$\psi_\theta^{r+1}(t) = C_r \sum_{k=1}^n \sum_{v\in V_n(r)} \{v_1, \dots, v_r, k+\theta\}\psi^0_{[v,k+\theta]}(t), \tag{13}$$

where $\{w\} := \langle w_1, \dots, w_s\rangle^{-1}\langle 1+w_1, \dots, w_s\rangle^{-1}$. Now

$$(k+\theta)\langle v_1, \dots, v_r + 1/(k+\theta)\rangle = \langle v_1, \dots, v_r, k+\theta\rangle,$$

so

$$\psi_\theta^{r+1}(t) = C_r \sum_{k=1}^n (k+\theta)^{-2} \sum_{v\in V_n(r)} \left\{v_1, \dots, v_r + \frac{1}{k+\theta}\right\} \psi^0_{[v_1,\dots,v_r+1/(k+\theta)]}(t) \tag{14}$$

We rewrite the inner sum with $a_k = 1/(k+\theta)$, as

$$\sum_{v\in V_n(r-1)} \sum_{l=1}^n \{v, l+a_k\}\psi^0_{[v,l+a_k]}(t). \tag{15}$$

This is a sum of the same form as in (13), but with r replaced by $r-1$. On the inductive assumption that the lemma holds for that case, the sum in (15) is equal to $(1/C_{r-1})\psi^r_{\alpha_k}(t)$, and since $\int_0^1$ (expression (15)) $dt = \sum_{v\in V_n(r-1)} \sum_{l=1}^n \{v, l+\alpha_k\}$, it follows that this sum is equal to C_{r-1}^{-1}. Therefore,

$$\psi_\theta^{r+1}(t) = C_r \sum_{k=1}^n (k+\theta)^{-2} \left(\sum_{v\in V_n(r)} \left\{v, \frac{1}{k+\theta}\right\} \right) \psi^r_{1/(k+\theta)}(t). \tag{16}$$

But

$$\begin{aligned} \langle v, 1/(k+\theta)\rangle(k+\theta) &= \langle v_1, v_2, \dots, v_r + 1/(k+\theta)\rangle(k+\theta) \\ &= \langle v_1, \dots, v_{r-1}\rangle(v_r(k+\theta)+1) + \langle v_1, \dots, v_{r-2}\rangle(k+\theta) \\ &= \langle v_r, \dots, v_r\rangle(k+\theta) + \langle v_1, \dots, v_{r-1}\rangle = \langle v_1, \dots, v_r, k+\theta\rangle. \end{aligned}$$

Hence

$$\psi_\theta^{r+1}(t) = C_r \sum_{v\in V_n(r)} \sum_{k=1}^{n} \{v, k+\theta\} \psi^r_{1/(k+\theta)}(t). \tag{17}$$

Since $\int_0^1 \psi_\theta^{r+1}(t)\,dt = 1$, C_r^{-1} must be $\sum_{v\in V_n(r)} \sum_{k=1}^{n}\{v, k+\theta\}$. This proves Lemma 2.

The point of Lemma 2 is that it displays the action of S on $\psi_\theta^r(t)$, as giving a weighted average of various $\psi_\alpha^r(t)$'s. If we view $\{\psi_\theta^r(t)\colon 0 \le \theta \le 1\}$ as a string, the points of which reside in some space, we see that the recursion yields on each iteration a new string which is contained in the convex hull of the preceding string. This sort of averaging ought eventually to squeeze the string down to a point, and that limit point will be our $g_n(t)$. The trick is to find the right norm.

The functions $\psi_\theta^0(t) = (1+\theta)/(1+\theta t)^2$ enjoy the property that $\psi^0_{\theta_1}(t) \succ \psi^0_{\theta_2}(t)$ if $\theta_1 > \theta_2$. That is, $\psi^0_{\theta_1}$ majorizes $\psi^0_{\theta_2}$ in the sense of Olkin and Marshall: For all s, $0 < s < 1$,

$$\int_0^s f(t)\,dt \ge \int_0^s g(t)\,dt, \qquad \text{and} \qquad \int_0^1 f(t)\,dt = \int_0^1 g(t)\,dt = 1. \tag{18}$$

When this holds for positive functions f and g, we say that $f \succ g$. We will also need the discrete analog. Given sequences $(a_1, a_2, \dots, a_r)$ and $(b_1, b_2, \dots, b_r)$ of nonnegative numbers, *not necessarily in decreasing order*, we say

$$a \succ b \quad \text{if} \sum_{i=1}^{k} a_i \ge \sum_{i=1}^{k} b_i \text{ for } 1 \le k < r, \text{ and } \sum_{i=1}^{r} a_i = \sum_{i=1}^{r} b_i = 1. \tag{19}$$

Clearly if $f \succ g$ and $g \succ h$ then $f \succ h$, and likewise for sequences. So if all the $\psi_\theta^r(t)$, for any fixed r, are comparable in this sense, we can put $\Psi_\theta^r(t) := \int_0^t \psi_\theta^r(s)\,ds$, and

$$\text{distance}(\psi^r_{\theta_1}, \psi^r_{\theta_2}) := \left| \int_0^1 \Psi^r_{\theta_1}(t) - \Psi^r_{\theta_2}(t)\,dt \right|, \tag{20}$$

and it will be a metric for $\{\psi_\theta^r(t)\colon 0 \le \theta \le 1\}$.

LEMMA 3. *For even r, $\psi^r_{\theta_1} \succ \psi^r_{\theta_2}$ if $\theta_1 \ge \theta_2$, while for odd r, $\psi^r_{\theta_1} \prec \psi^r_{\theta_2}$ if $\theta_1 \ge \theta_2$.*

This is not easily proved, and we must work up to it with some auxiliary lemmas.

LEMMA 4. *For fixed r, the sequences $\gamma_k^r(\theta_j)$ satisfy $\gamma_k^r(\theta_1) \succ \gamma_k^r(\theta_2)$ if and only if $\theta_1 \le \theta_2$.*

LEMMA 5. *Suppose $0 < c_1 \le c_2 \le \cdots \le c_n$ and $a_1, a_2, \dots, a_n > 0$. Then for all k, $1 \le k \le n$,*

$$\sum_{j=1}^{k} a_j \Big/ \sum_{j=1}^{n} a_j \ge \sum_{j=1}^{k} c_j a_j \Big/ \sum_{j=1}^{n} c_j a_j.$$

PROOF.

$$\begin{aligned}\sum_{j=1}^{n} c_j a_j \sum_{l=1}^{k} a_l - \sum_{j=1}^{k} c_j a_j \sum_{l=1}^{n} a_l &= \sum_{j=k+1}^{n} c_j a_j \sum_{l=1}^{k} a_l - \sum_{j=1}^{k} c_j a_j \sum_{l=k+1}^{n} a_l \\ &\geq c_{k+1} \sum_{k+1}^{n} \sum_{l=1}^{k} a_j a_l - c_k \sum_{j=1}^{k} \sum_{l=k+1}^{n} a_j a_l \\ &= (c_{k+1} - c_k) \sum_{j=k+1}^{n} \sum_{l=1}^{k} a_j a_l \geq 0. \quad \square\end{aligned}$$

Now this says that $(a_k)/\sum_1^n a_k \succ (c_k a_k)/\sum_1^n c_k a_k$.

To apply the lemma, we fix $0 \leq s < t \leq 1$, and put

$$a_k = \sum_{v \in V_n(r)} \{v, k+s\} = a_k(s), \quad \text{say, and} \tag{21}$$

$c_k = \sum_{v \in V_n(r)} \{v, k+t\} / \sum_{v \in V_n(r)} \{v, k+s\} = a_k(t)/a_k(s)$. To establish that $c_k \leq c_{k+1}$, $1 \leq k \leq n-1$, we prove that

$$\frac{d^2}{ds^s} \log a_k(s) > 0. \tag{22}$$

From this it will follow that

$$\frac{d}{ds} \log a_k(s) < \frac{d}{ds} \log a_k(s+1),$$

so that

$$\log a_k(t) - \log a_k(s) < \log a_{k+1}(t) - \log a_{k+1}(s),$$

and

$$a_k(t)/a_k(s) < a_{k+1}(t)/a_{k+1}(s).$$

But (22) is equivalent to the claim that

$$a_k(s) a_k''(s) > (a_k'(s))^2. \tag{23}$$

To prove this we start with the cases $r = 0$ and $r = 1$. For $r = 0$, $a_k(t) = (k+t)^{-1}(k+1+t)^{-1}$, so $\log a_k(t) = -(\log(k+t) + \log(k+1+t))$ which is thus concave up. For $r = 1$,

$$(24)\qquad a_k(t) = \sum_{j=1}^{n}((k+t)j+1)^{-1}((k+t)(j+1)+1)^{-1},$$

$$a_k'(t) = -\sum_{j=1}^{n} j((k+t)j+1)^{-2}((k+t)(j+1)+1)^{-1}$$
$$-\sum_{j=1}^{n}(j+1)((k+t)j+1)^{-1}((k+t)(t+1)+1)^{-2}, \quad \text{and}$$

$$a_k''(t) = \sum_{j=1}^{n} 2j^2((k+t)j+1)^{-3}((k+t)(j+1)+1)^{-1}$$
$$+\sum_{j=1}^{n} 2j(j+1)((k+t)j+1)^{-2}((k+t)(j+1)+1)^{-2}$$
$$+\sum_{j=1}^{n} 2(j+1)^2((k+t)j+1)^{-1}((k+t)(j+1)+1)^{-3}.$$

Thus with the notation $((k+t)j+1)^{-1} = e_j$, $((k+t)(j+1)+1)^{-1} = f_j$,

$$(25)\ a_k''(t)a_k(t) - (a_k'(t))^2 = \sum_{j=1}^{n}\sum_{l=1}^{n} 2j^2 e_j^3 f_j e_l f_l + 2j(j+1)e_j^2 f_j^2 e_l f_l$$
$$+ 2(j+1)^2 e_j f_j^3 e_l f_l - jl e_j^2 e_l^2 f_j f_l - j(l+1)e_j^2 e_l f_j f_l^2$$
$$- (j+1)l e_j f_j^2 e_l^2 f_l - (j+1)(l+1)e_j e_l f_j^2 f_l^2.$$

By symmetry this is equal to

$$(26)\qquad \sum_{j=1}^{n}\sum_{l=1}^{n} e_j e_l f_j f_l \cdot \left(\frac{1}{2} \text{ if } j = l, \text{ else } 1\right)$$
$$\cdot \left\{ \begin{array}{l} j^2e_j^2 + l^2e_l^2 - j(j+1)e_jf_j + l(l+1)e_lf_l \\ \quad + (j+1)^2f_j^2 + (l+1)^2f_l^2 - jle_je_l - j(l+1)e_jf_l \\ \quad - (j+1)lf_je_l - (j+1)(l+1)f_jf_l. \end{array} \right\}.$$

The complicated factor can be simplified a bit by setting $E_j = je_j$, etc., to read $E_j^2 + E_l^2 + E_jF_j + E_lF_l + F_j^2 + F_l^2 - E_jE_l - E_jF_l - E_lF_j - F_jF_l$. This is

$$\geq \tfrac{1}{2}E_j^2 + \tfrac{1}{2}E_l^2 + \tfrac{1}{2}F_j^2 + \tfrac{1}{2}F_l^2 + E_jF_j + E_lF_t - E_jF_l - E_lF_j$$
$$\geq \frac{1}{2}(E_j - F_l)^2 + \tfrac{1}{2}(E_l - F_j)^2 + E_jF_j + E_lF_l > 0.$$

For $r \geq 2$, we can take the approach that

$$\langle v_1, v_2, \ldots, v_r, k+t\rangle = (k+t+p_v)\langle v_1, v_2, \ldots, v_r\rangle,$$

where $p_v = \langle v_1, v_2, \ldots, v_{r-1}\rangle / \langle v_2, v_2, \ldots, v_r\rangle$, while

$$\langle 1+v_2, v_2, \ldots, v_r, k+t\rangle = (k+t+q_v)\langle 1+v_2, v_2, \ldots, v_r\rangle,$$

with the obvious meaning for q_v.

Now for general r,

$$a_k(t) = \sum_{v\in V_n(r)} \{v, k+t\} = \sum_{v\in V_n(r)} (k+t+p_v)^{-1}(k+t+q_v)^{-1}\{v\}, \tag{27}$$

$$a'_k = 2 \sum_{v\in V_n(r)} ((k+t+p_v)^{-2}(k+t+q_v)^{-1} + (k+t+p_v)^{-1}(k+t+q_v)^{-2})\{v\}$$

and

$$a''_k = 2 \sum_{v\in V_n(r)} ((k+t+p_v)^{-3}(k+t+q_v)^{-1} + (k+t+p_v)^{-2}(k+t+q_v)^{-2} + (k+t+p_v)^{-1}(k+t+q_v)^{-3})\{v\}.$$

Thus $a''_k(t)a_k(t) - (a'_k(t))^2 =$ (with $P_v = (k+t+q_v)^{-1}$ etc.)

$$\sum_{v\in V_n(r)} \sum_{w\in V_n(r)} \{v\}\{w\}P_vQ_vP_wQ_w \cdot (\tfrac{1}{2} \text{ if } v = w \text{ else } 1) \cdot \left\{ \begin{array}{l} P_v^2 + P_w^2 + P_vQ_v + P_wQ_w + Q_v^2 + Q_w^2 \\ - P_vP_w - P_vQ_w - P_wQ_v - Q_vQ_w \end{array} \right\}. \tag{28}$$

Again the last factor is positive, for the same reason as with the E_j and F_j.

Thus we may apply Lemma 5 in (21) and conclude that for arbitrary $r \geq 1$ and $0 \leq \theta_1 < \theta_2 \leq 1$,

$$\gamma_k^r(\theta_1) \succ \gamma_k^r(\theta_2). \tag{29}$$

This proves Lemma 4.

Now for $r = 0$, $\psi^0_{\theta_1}(t) \succ \psi^0_{\theta_2}(t)$ if $\theta_1 > \theta_2$. By Lemma 2,

$$\psi_\theta^{r+1}(t) = \sum_{k=1}^{n} \gamma_k^r(\theta)\psi^r_{1/(k+\theta)}(t). \tag{30}$$

Thus with the notation $\Psi_\theta^r(t) := \int_0^t \psi_\theta^r(s)\,ds$, we have

$$\Psi_\theta^{r+1}(t) = \sum_{k=1}^{n} \gamma_k^r(\theta)\Psi^r_{1/(k+\theta)}(t). \tag{31}$$

Now make the inductive assumption that Lemma 3 holds out to r (and is in doubt for $r+1$). If r is even, then $\Psi^r_{\theta_1}(t) \geq \Psi^r_{\theta_2}(t)$ if $\theta_1 \geq \theta_2$. Thus if $\theta_1 \geq \theta_2$, then

$$\sum_{k=1}^{n} \gamma_k^r(\theta_1)\Psi^r_{1/(k+\theta_1)}(t) \leq \sum_{k=1}^{n} \gamma_k^r(\theta_1)\Psi^r_{1/(k+\theta_2)}(t). \tag{32}$$

The sequence $\Psi^r_{1/(k+\theta_2)}(t)$, $1 \leq k \leq n$, is decreasing in k because $1/(k+\theta_2) > 1/(k+1+\theta_2)$. If we replace $\gamma_k^r(\theta_1)$ with $\gamma_k^r(\theta_2)$ this shifts mass into smaller values of k, where it will multiply larger $\Psi^r_{1/(k+\theta)}(t)$ values. More precisely, if $(a_k) \succ (b_k)$ and if $c_1 \geq c_2 \geq \cdots \geq c_n \geq 0$ then $\sum_{k=1}^n a_kc_k \geq \sum_{k=1}^n b_kc_k$. (This is easily proved by Abel summation and is left to the reader.) Hence $\Psi^{r+1}_{\theta_1}(t) \leq \Psi^{r+1}_{\theta_2}(t)$ for $\theta_2 \leq \theta_1$.

If r is odd, the inequalities are just reversed. The same sort of argument yields $\Psi^{r+1}_{\theta_1}(t) \geq \Psi^{r+1}_{\theta_2}(t)$, and this completes the proof of Lemma 2.

Now let

$$D^r(\theta_1,\theta_2) := \int_0^1 |\Psi^r_{\theta_1}(t) - \Psi^r_{\theta_2}(t)|.$$

By Lemma 2, $D_r(\theta_1,\theta_2) = |\Gamma^r(\theta_1) - \Gamma^r(\theta_2)|$, where $\Gamma^r(\theta): = \int_0^1 \Psi^r_\theta(t)\,dt$. Moreover, for all r, $\Gamma^r(\theta)$ is monotone (increasing or decreasing depending on the parity of r) as a function of θ. Now from (31),

$$\Gamma^{r+1}(\theta) = \sum_{k=1}^{n} \gamma^r_k(\theta)\Gamma^r\left(\frac{1}{k+\theta}\right) \tag{33}$$

Inspection of the definition of $\gamma^r_k(\theta)$ shows that

$$k^{-2} \ll \gamma^r_k(\theta) \ll k^{-2} \quad \text{(uniformly in } r \text{ and } \theta), \tag{34}$$

and that $\gamma^r_{k+1}(\theta) < \gamma^r_k(\theta)$, for $1 \le k \le n-1$. But now

$$\begin{aligned}
\Gamma^{r+1}(0) - \Gamma^{r+1}(1) &= \sum_{k=1}^{n} \gamma^r_k(1)\Gamma^r\left(\frac{1}{k+1}\right) - \gamma^r_k(0)\Gamma^r\left(\frac{1}{k}\right) \\
&= \left(\sum_{k=1}^{n-1} (\gamma^r_k(1) - \gamma^r_{k+1}(0))\Gamma^r\left(\frac{1}{k+1}\right)\right) \\
&\quad + \gamma^r_n(1)\Gamma^r\left(\frac{1}{n+1}\right) - \gamma^r_1(0)\Gamma^r(1), \\
&= \sum_{k=1}^{n+1} \delta(k,r)\Gamma^r\left(\frac{1}{k}\right), \quad \text{say.}
\end{aligned}$$

Now $\sum_{k=1}^{n+1} \delta(k,r) = 0$. Because of the cancellation among terms $\delta(k,r) = \gamma^r_k(1) - \gamma^r_{k+1}(0)$ for $1 \le k < n$, the sum of the positive $\delta(k,r)$ is less than 1, for each r. In fact, it is less than 6/7, since both $\gamma^r_1(1)$ and $\gamma^r_2(0)$ are greater than 1/7. To see this, we consider first

$$\gamma^r_2(0) = \sum_{V_n(r)} \langle v,2\rangle^{-1}\langle 1+v_1,\dots,v_r,2\rangle^{-1} \Big/ \sum_{V_n(r)}\sum_{l=1}^{n} \langle v,l\rangle^{-1}\langle 1+v_1,\dots,v_r,l\rangle^{-1}. \tag{35}$$

Now for $1 \le l \le n$, and any $v \in V_\infty(r)$, $\langle v,l\rangle/\langle v,2\rangle \le l/2$. Thus

$$\langle v,2\rangle^{-1}\langle 1+v_1,\dots,v_r,2\rangle^{-1} \le (l^2/4)\langle v,l\rangle^{-1}\langle 1+v_1,\dots,v_r,l\rangle^{-1},$$

so that $\gamma^r_2(0) \ge 1/\sum_{l=1}^{\infty} 4/l^2 = 6/4\pi^2 > 1/7$. The fraction defining $\gamma^r_1(1)$ as in (35) has the same numerator but a smaller denominator, since $l+1$ replaces l in $\langle v,l\rangle$. Thus also $\gamma^r_1(1) > 1/7$. Let $E(r) = \sum_{k:\delta(k,r)>0} \delta(k,r) = \frac{1}{2}\sum_{k=1}^{n} |\delta(k,r)|$. Then since $\Gamma^r(\theta)$ is monotone in θ, $0 \le \theta \le 1$,

$$\left|\sum_{k=1}^{n} \delta(k,r)\Gamma^r\left(\frac{1}{k}\right)\right| \le |E(r)\Gamma^r(0) - E(r)\Gamma^r(1)| \le \frac{6}{7}|\Gamma^r(0) - \Gamma^r(1)|, \tag{36}$$

that is,

$$|\Gamma^{r+1}(0) - \Gamma^{r+1}(1)| \le (\tfrac{6}{7})|\Gamma^r(0) - \Gamma^r(1)|. \tag{37}$$

Now for $r = 0$, $\Gamma^0(0) = \int_0^1 \Psi_0^0(t)\,dt = \int_0^1 t\,dt = 1/2$, while

$$\Gamma^0(1) = \int_0^1 \Psi_1^0(t)\,dt = \int_0^1 \left(2 - \frac{2}{1+t}\right) dt = 2(1 - \log 2) = .61371 \quad \text{(approx.)}.$$

Thus $|\Gamma^0(0) - \Gamma^0(1)| < 1/7$, so

$$|\Gamma^r(0) - \Gamma^r(1)| \le (1/7)(6/7)^r, \tag{38}$$

or equivalently,

$$\|\Psi_{\theta_1}^r - \Psi_{\theta_2}^r\|_1 \le (1/7)(6/7)^r \quad \text{for } \theta_i,\ 0 \le \theta_i \le 1 \text{ and } r \ge 1. \tag{39}$$

Now some elementary calculus from (39), gives for all $r \ge 1$ and $0 \le \theta_i \le 1$, $0 \le t \le 1$,

$$|\psi_{\theta_1}^r(t) - \psi_{\theta_2}^r(t)| \le 2(19/20)^r. \tag{40}$$

We begin the proof with some notation. Let $F(t) = \Psi_{\theta_1}^r(t) - \Psi_{\theta_2}^r(t)$, and assume $|F(t)|$ takes a maximum value of ε at $t = t_0$.

Since $F(t)$ is a linear combination of functions of the form $(1+\theta)t/(1+\theta t)$, with coefficients in $[-1, 1]$, and since both the sum of the positive, and the negative, coefficients is 1, we see that for $0 \le t \le 1$,

$$|F'(t)| \le 2, \quad |F''(t)| \le 4. \tag{41}$$

Now $F(0) = F(1) = 0$, so $|F(t)| \ge \varepsilon - 2(t - t_0)^2$ for $|t - t_0| \le \sqrt{\varepsilon/2}$, and the interval about t_0 of radius $\sqrt{\varepsilon/2}$ lies within $[0, 1]$. Thus

$$\int_0^1 |F(t)|\,dt \ge \left(\frac{2\sqrt{2}}{3}\right) \varepsilon^{3/2}. \tag{42}$$

In view of (39), this forces $\varepsilon \le (1/3)/(6/7)^{2r/3}$.

Now put $f(t) = F'(t) = \psi_{\theta_1}^r(t) - \psi_{\theta_2}^r(t)$, and suppose $1 \ge \delta = \sup_{0 \le t \le 1} |f(t)| = |f(t_1)|$. Then one of $t_1 \pm \frac{1}{2}\delta$ is in $[0, 1]$ (say $t_1 + \frac{1}{2}\delta$), so that as before,

$$\left| \int_{t_1}^{t_1 + \frac{1}{2}\delta} f(t)\,dt \right| \ge \frac{1}{4}\delta^2. \tag{43}$$

Since $|F(t)| \le (1/3)(6/7)^{2r/3}$, the change from t_1 to $t_1 + \frac{1}{2}\delta$ is $\le (2/3)(6/7)^{2r/3}$ in F, so $\delta \le 2(6/7)^{r/3}$. Thus $\delta < 2(19/20)^r$, which is equivalent to (40).

Now the sets $\overline{\text{Convex Hull}}\{S^r((1+\theta)/(1+\theta t)^2), 0 \le \theta \le 1\}$ are a nested sequence of compact sets, with diameter tending to zero. Therefore there is a function $g(t) = g_n(t)$ such that

$$\lim_{r \to \infty} S^r((1 + \theta)/(1 + \theta t)^2) = g(t) \tag{44}$$

uniformly for $0 \le t \le 1$. We now derive some of the properties of $g(t)$. Consider the linear operator $K\colon M \to B$, where M is the set of all functions $m(x)$ on $[0, 1]$ of bounded variation, with $m(0) = 0$. B is the set of all continuous differentiable functions on $[0, 1]$, and

$$K(m) = \int_0^1 \frac{(1 + x)\,dm(x)}{(1 + xt)^2}. \tag{45}$$

We identify m which agree almost everywhere (so that really M is a set of equivalence classes). Every m can be represented as $c_1 m_1 - c_2 m_2$, where $c_1 \geq 0$, $c_2 \geq 0$, and m_1 and m_2 are probability distribution functions ($m_i(0) \geq 0$, $m_i(1) = 1$, continuous from the right and nondecreasing).

LEMMA 6. *If m_1 and m_2 are probability distribution functions and $K(m_1) = K(m_2)$ then $m_1 = m_2$.*

PROOF. Otherwise we should have a function

$$b(t) = \int_0^1 \frac{(1+x)\, dm_1(x)}{(1+xt)^2} = \int_0^1 \frac{(1+x)\, dm_2(x)}{(1+xt)^2}, \quad \text{for } 0 \leq t \leq 1. \tag{45}$$

Thus

$$\int_0^1 \frac{(1+x)d(m_1 - m_2)(x)}{(1+xt)^2} \equiv 0 \quad \text{on } 0 \leq t \leq 1.$$

Repeated differentiation yields

$$0 \equiv \int_0^1 \frac{x^r(1+x)}{(1+xt)^{2+r}} d(m_1 = m_2)(x) \tag{47}$$

and in particular, with $t = 0$,

$$0 = \int_0^1 x^r(1+x)\; d(m_1 - m_2)(x) \tag{48}$$

for all $r \geq 0$. Now $\int_0^1 d(m_1 = m_2)(x) = 0$, so integrating (48) by parts repeatedly yields

$$\int_0^1 x^r\; d(m_1 - m_2)(x) = \sum_{k=1}^{r-1} x^k(1+x)(-1)^{r-k+1} d(m_1 - m_2)(x) = 0. \tag{49}$$

Taking linear combinations of appropriate instances of (49) and integrating once more by parts gives

$$\int_0^1 p(x)(m_1 - m_2)(x)\, dx = 0 \tag{50}$$

for all polynomials p. But this forces $m_1 = m_2$, for otherwise a sufficiently good polynomial approximation to $(m_1 - m_2)(x)$ would falsify (50).

The probability distribution functions on [0,1] form a compact metric space under the Levy metric. To each function $\psi_0^r(t) \in B$, associate

$$m_r(x) := \sum_{\theta \leq x} c_\theta, \quad \text{where } \psi_0^r(t) = \sum_\theta c_\theta \frac{1+\theta}{(1+\theta t)^2}. \tag{51}$$

With respect to this metric on M, and the L^∞ norm on B, say, K is continuous. The inverse images $m_r(x)$ of the $\psi_0^r(t)$ must have at least one accumulation point in M, since M is compact. But there cannot be two, since K would send each to $g(t)$. Thus there is a unique measure $\beta = \beta_n$ and associated probability distribution function $m(x)$, such that $K(m) = g(t)$. That is,

$$g_n(t) = \int_0^1 (1+\theta)/(1+\theta t)^2\, d\beta_n(\theta). \tag{52}$$

REMARK. Each of the functions $(1+\theta)/(1+\theta t)^2$ is analytic in the half-plane $\mathrm{Re}(t) > -1$, so $g_n(t)$ is also analytic in that domain. Thus

$$((g_n)(t))^r = \int_0^1 (1+\theta)\theta^r(-1)^r(r+1)!(1+\theta t)^{-2-r}\,d\beta_n(\theta)$$

for all $r \geq 1$.

From (44), $S(g(t)) = g(t)$. That is,

$$\left(\int_0^1 \sum_{k=1}^n (k+s)^{-2} g_n\left(\frac{1}{k+s}\,ds\right)\right) g_n(t) = \sum_{k=1}^n (k+t)^{-2} g_n\left(\frac{1}{k+t}\right). \tag{53}$$

Let

$$\lambda(n) := \int_0^1 \sum_{k=1}^n (k+s)^{-2} g_n\left(\frac{1}{k+s}\right) ds. \tag{54}$$

Then also

$$\lambda(n) = \sum_{k=1}^n \int_{1/(k+1)}^{1/k} g_n(u)\,du = \int_{1/(n+1)}^1 g_n(u)\,du,$$

so $0 < \lambda(n) < 1$.

We now have all of Theorem 2 except that β_n is concentrated on the set of irrationals in $[0,1]$ with continued fraction expansion of the form $[a_1, a_2, \dots]$, with all $a_i \leq n$. For the proof of this, consider $A_r = \{\theta\colon 0 \leq \theta \leq 1$ and $S^r(\chi_{[0,1]}(t))$ includes $(1+\theta)/(1+\theta t)^2$ with a positive coefficient$\}$. Thus $A_0 = \{0\}$, $a_1 = \{1, 1/2, \dots, 1/n\}$, and from (6), $A_r = \{\theta = [v_1, v_2, \dots, v_r]\colon v_i \leq n$ for $1 \leq i \leq r.\}$ Let $A = \{[v_1, v_2, \dots]\colon v_i \leq n$ for all $i \geq 1\}$. A is closed. For any x, $0 < x < 1$, *not* of the form $[a_1, a_2, \dots]$, all $a_i \leq n$, write $x = [a_1, a_2, \dots, a_j, b_1, b_2, \dots]$, where $b_1 > n$. We admit the possibility that $a_j > 1$ and $b_1 = \infty$, in which case there are no further b's. If $b_1 = \infty$ then consider the interval bounded by $[a_1, a_2, \dots, a_j - 1, 1, 2n+1]$ and $[a_1, a_2, \dots, a_j, 2n+1]$. For all $r \geq 1$, no element of A_r belongs to this interval. Were β to assign positive measure to the inner half of this interval, the Levy distance between the $m_r(x)$ and $m(x)$ could not converge to zero. Thus the support of β is at any rate a subset of A. Now consider $x \in A$. We must show that

$$\beta(x-\varepsilon, x+\varepsilon) > 0$$

for every $\varepsilon > 0$. So fix $\varepsilon > 0$.

There exist $a_1, a_2, \dots, a_j \leq n$ such that $|[a_1, a_2, \dots, a_j + s] - x| < \varepsilon$ for all s, $0 \leq s \leq 1$. Now for all $r > j$,

$$m_r(x-\varepsilon, x+\varepsilon) \geq \frac{1}{10}(\lambda(n))^{-r} \sum_{\substack{v=[a_1,a_2,\dots,a_j,b_1,b_2,\dots,b_{r-j}] \\ v \in V_n(r)}} \langle v\rangle^{-2}. \tag{55}$$

But that is

$$\geq \lambda(n)^{-j}\frac{1}{100}\langle a_1, a_2, \dots, a_j\rangle^{-2}\left(\sum_{w \in V_n(r-j)} \langle w\rangle^{-2}\right)(\lambda(n)^{j-r}). \tag{56}$$

Now by Lemma 1, $\sum_{v \in V_n(r)} \langle v \rangle^{-2} \geq \frac{1}{100}(\lambda(n))^r$, since for all θ, $\phi_\theta^r(t) \geq \frac{1}{10} L^r g(t)$ because $\phi_\theta^0(t) \geq \frac{1}{10} g(t)$ on $[0,1]$. Thus $m_r(x-\varepsilon, x+\varepsilon) \gg \lambda(n)^{-j} \langle a_1, \dots, a_j \rangle^{-2}$, and remains bounded away from zero as $r \to \infty$. This proves the last outstanding claim of Theorem 2.

To prove Theorem 3, it is sufficient to work with sets of the form $(0,a)$. Then

$$\begin{aligned} \nu_n(T_n^{-1}(0,a)) &= \sum_{k=1}^{n} \int_{1/(k+a)}^{1/k} g_n(t)\,dt \\ &= \lambda(n) \int_0^a g_n(t)\,dt \quad \text{by Theorem 2,} \\ &= \lambda(n) \nu_n(0,a). \end{aligned}$$

This leaves Theorem 1.

To see that $\frac{1}{2}\lambda(n)^k \leq \mu_n(k-1) \leq 2\lambda(n)^k$, we note that $\frac{1}{2} g_n(t) \leq 1 \leq 2 g_n(t)$ for $0 \leq t \leq 1$. Thus

$$\frac{1}{2} \int_0^1 L^k g_n(t)\,dt \leq \int_0^1 L^k \chi_{[0,1]}(t)\,dt \leq \int_0^1 L^k g_n(t)\,dt, \tag{57}$$

so that

$$\frac{1}{2}\lambda(n)^k \leq \int_0^1 \phi_0^k(t)\,dt \leq 2\lambda(n)^k.$$

But

$$\int_0^1 \phi_0^k(t)\,dt = \mu_n(k-1), \quad \text{for all } k \geq 1. \tag{58}$$

PROOF. For each $w \in V_n(k)$, consider the open intervals $J(w,i)$, $1 \leq i \leq n$, bounded by $[w,i]$ and $[w,i+1]$ in one order or the other. Each open interval of $R_n(k)$ has the form $\bigcup_{i=1}^n J(w,i)$, together with the connecting points. A simple calculation establishes

$$|[w,i] - [w,i+1]| = \langle w,i \rangle^{-1} (\langle w,i \rangle + \langle w \rangle)^{-1}. \tag{59}$$

Summing (59) over all choices of w, and $1 \leq i \leq n$ gives (58) with $k+1$ in place of k, in view of the corollary to Lemma 1.

REMARK. Because $S^k \phi_0^r(t)$ converges exponentially to $g(t)$, more is true: there exists a constant $c(n)$, $\frac{1}{2} < c(n) < 2$, such that

$$\mu_n(k-1) = c(n)(1 + O(19/20)^n)\lambda(n)^k.$$

We now show that $\mu_n(k+1)/\mu_n(k)$ alternates about $\lambda(n)$. First recall that we have already seen that $\psi_{\theta_2}^r(t) \prec \psi_{\theta_1}^r(t)$ if $\theta_1 \leq \theta_2$ and r is odd, while $\psi_{\theta_1}^r(t) \prec \psi_{\theta_2}^r(t)$ if $\theta_1 \leq \theta_2$ and r is even (Lemma 3). Now for even r, $\psi_0^{r+1}(t) = \sum_{k=1}^n \gamma_k^r(0) \psi_{1/k}^r(t)$. Each component of this sum majorizes $\psi_0^r(t)$, and the coefficients are positive with a sum of 1. So $\psi_0^{r+1}(t) \succ \psi_0^r(t)$. If r is odd a similar argument shows that $\psi_0^{r+1} \prec \psi_0^r$. Now

$$\begin{aligned} \frac{\mu_n(k)}{\mu_n(k-1)} &= \int_0^1 \phi_0^{k+1}(t)\,dt \Big/ \int_0^1 \phi_0^k(t)\,dt \\ &= \int_{1/(n+1)}^1 \phi_0^k(t)\,dt \Big/ \int_0^1 \phi_0^k(t)\,dt = \int_{1/(n+1)}^1 \psi_0^k(t)\,dt. \end{aligned} \tag{60}$$

For k even,

$$\int_0^{1/(n+1)} \psi_0^k(t)\,dt \geq \int_0^{1/(n+1)} \psi_0^{(k+1)}(t)\,dt,$$

and this is reversed for k odd. Consequently, $\mu_n(k)/\mu_n(k-1)$ alternates. Furthermore, two applications of Lemma 2 give

$$\psi_0^{r+2}(t) = \sum_{k=1}^{n}\sum_{j=1}^{n} \gamma_k^{r+1}(0)\gamma_j^r\left(\frac{1}{k}\right)\psi_{1/(j+1/k)}^r(t), \tag{61}$$

from which it follows that $\psi_0^0 \prec \psi_0^2 \prec \psi_0^4 \prec \cdots \prec g \prec \cdots \prec \psi_0^5 \prec \psi_0^3 \prec \psi_0^1$. Together with (60), this shows that $\mu_n(k)/\mu_n(k-1)$ alternates about $\lambda(n)$.

Next, we prove that $\lambda(n+1) > \lambda(n)$ for all $n \geq 1$. We have

$$\phi_0^k(t) = \sum_{v \in V_n(k)} (\langle v\rangle + t\langle v^-\rangle)^{-2}, \tag{62}$$

and we put

$$\tilde{\phi}_0^k(t) = \sum_{v \in V_{n+1}(k)} (\langle v\rangle + t\langle v^-\rangle)^{-2}.$$

If $\lambda(n)$ were equal to $\lambda(n+1)$ then $\tilde{\phi}_0^k(t)/\phi_0^k(t)$ would be bounded above. We prove it is not, with $t=0$.

For every $v \in V_n(k)$ consider

$$W_v := \{w \in V_{n+1}(k) : w = v \text{ in } k-1 \text{ places, and } (n+1) \text{ in one place.}\}$$

Each w with one entry equal to $n+1$ belongs to n sets W_v.

For $w \in W_v$, $\langle w\rangle \leq 4n\langle v\rangle$. Thus $\langle w\rangle^{-2} \geq (4n)^{-2}\langle v\rangle^{-2}$, so that

$$\sum_{w \in V_{n+1}(k)} \langle w\rangle^{-2} \geq \sum_{v \in V_n(k)} kn^{-1}(4n)^{-2}\langle v\rangle^{-2}.$$

For fixed n and $k \to \infty$, this gives $\tilde{\phi}_0^k(0)/\phi_0^k(0) \to \infty$.

Finally, we prove that $\lim_{n\to\infty} n(1-\lambda(n)) = 1/\log 2$. We know that

$$g_n(0) = \lim_{k\to\infty} \sum_{v\in V_n(k)} \langle v\rangle^{-2} \Big/ \sum_{v \in V_n(k)} \langle v\rangle + \langle v^-\rangle)^{-1}.$$

From Schweiger [8], if $n = \infty$, then

$$\phi_0^{(k)}(t) = \sum_{v \in V_\infty(k)} \langle v\rangle^{-2}\phi_{[v]}^0(t) = \frac{1}{(1+t)\log 2}\left(1 + O\left(\frac{3}{4}\right)^k\right),$$

and $\sum_{v\in V_\infty(k)} \langle v\rangle^{-1}(\langle v\rangle + \langle v^-\rangle)^{-1} = 1$ since for $n=\infty$, $\int_0^1 \phi_0^{(k)}(t)\,dt = 1$ for all k. Thus for fixed k as $n \to \infty$

$$g_n(0) = (1+o_k(1)) \sum_{v \in V_n(k)} \langle v\rangle^{-2} = \frac{(1+o_k(1))(1+O(3/4^k))}{\log 2}.$$

Hence $\lim_{n\to\infty} g_n(0) = 1/\log 2$. Since $1-\lambda(n) = \int_0^{1/(n+1)} g_n(t)\,dt$, and since $|g_n'(t)| \leq 2$, this integral is asymptotic to $g_n(0)/n$ as $n \to \infty$, and so $n(1-\lambda(n))$ tends toward $1/\log 2$ as claimed.

LAST REMARK. Clearly $T_n(x) \to \{1/x\}$ if $x \geq 1/(n+1)$, else 0, tends to scramble things before kicking them out of bounds to zero. Is there some analog to the ergodic theorem for measure-decimating transformation?

4. Uniform distribution of solutions to $v^2 \equiv -1 \bmod k$. Since the result obtained does not match Hooley's in accuracy, we confine ourselves to a mention of the salient steps.

The problem is converted to one of equidistribution of $\sigma^*(\alpha)$, over $\alpha \in A(x)$ as $x \to \infty$. It is noted that α^* is essentially parallel to α, and that if $\alpha = (a, b)$, $\alpha^* = (c, d)$ then $\sigma^*(\alpha)$ is close to $(ac + bd)/(a^2 + b^2) = \nu/k$, where $\nu = ac + bd \equiv \sqrt{-1} \bmod a^2 + b^2$. The important steps begin with several lemmas.

LEMMA 1.

$$\sum_{\substack{|\alpha| \leq y \\ \alpha \in \mathcal{A}}} \chi(\sigma^*(\alpha) \leq t) = \frac{6}{\pi} t y^2 + O(t^2 y^2) + O(y).$$

PROOF. We start with the identity

$$\sum_{\substack{|\alpha| \leq y \\ \alpha \in \mathcal{A}}} \chi(\sigma^*(\alpha) \leq t) = \sum_{\substack{|\beta| \leq ty \\ \beta \in \mathcal{A}}} \sum_{k=1}^{\infty} \chi(|\beta|/t \leq |\beta_* + k\beta| \leq y). \tag{1}$$

Now $k|\beta| < |\beta_* + k\beta| < (k+1)|\beta|$, so

$$\begin{aligned}\sum_{\substack{|\alpha| \leq y \\ \alpha \in \mathcal{A}}} \chi(\sigma^*(\alpha) \leq t) &\leq \sum_{\substack{|\beta| \leq ty \\ \beta \in \mathcal{A}}} \sum_{k=1}^{\infty} \chi(1/t - 1 \leq k \leq y/|\beta|) \\ &\leq \sum_{\substack{|\beta| \leq ty \\ \beta \in \mathcal{A}}} \left(\left[\frac{y}{|\beta|} \right] - \left[\frac{1}{t} \right] + 1 \right).\end{aligned} \tag{2}$$

Similarly, the rightmost term of (2), with -1 in place of $+1$, provides a lower bound for (1). Both of these bounds lie within $\sum_{|\beta| \leq ty;\ \beta \in \mathcal{A}} 2$ of $\sum_{|\beta| \leq ty;\ \beta \in \mathcal{A}} (y/|\beta| - 1/t)$, and we settle for the trivial estimate

$$\sum_{\substack{|\beta| \leq ty \\ \beta \in \mathcal{A}}} 2 \leq 2((2ty + 1)^2 - 1) = 8t^2 y^+ 8ty \leq 8t^2 y^2 + 8y. \tag{3}$$

Now we must estimate $\sum_{|\beta| \leq z;\ \beta \in \mathcal{A}} 1$ and $\sum_{|\beta| \leq z;\ \beta \in \mathcal{A}} 1/|\beta|$. This requires two subsidiary lemmas.

LEMMA 2. $\sum_{|\beta| \leq z;\ \beta \in \mathcal{A}} 1 = 6z^2/\pi + O(z)$.

LEMMA 3. $\sum_{|\beta| \leq z;\ \beta \in \mathcal{A}} 1/|\beta| = 12z/\pi + O(1)$.

COROLLARY. *Lemma* 1 *still holds if in the statement,* $\chi(\sigma^*(\alpha) \leq t)$ *is replaced by any expression of the form* $\chi(\sigma^*(\alpha) + \rho(\alpha) \leq t)$, *provided* $\rho(\alpha) = O(|\alpha|^{-3})$.

PROOF. If $\rho(\alpha) = O(|\alpha|^{-3})$ then $|\rho(\alpha)| \leq C|\alpha|^{-3}$ for some C, and all $\alpha \in \mathcal{A}$. Now for any $\delta > 0$,

$$\begin{aligned}\sum_{\substack{|\alpha| \leq y \\ \alpha \in \mathcal{A}}} \chi(\sigma^*(\alpha) + \delta \leq t) - \sum_{\substack{|\alpha| < (C/\delta)^{1/3} \\ \alpha \in \mathcal{A}}} 1 &\leq \sum_{\substack{|\alpha| \leq y \\ \alpha \in \mathcal{A}}} \chi(\sigma^*(\alpha) + \rho(\alpha) \leq t) \\ &\leq \sum_{\substack{|\alpha| \leq y \\ \alpha \in \mathcal{A}}} \chi(\sigma^*(\alpha) - \delta \leq t) + \sum_{\substack{|\alpha| \leq (C/\delta)^{1/3} \\ \alpha \in \mathcal{A}}} 1.\end{aligned} \tag{4}$$

The corollary follows from Lemmas 1 and 2 on taking $\delta = y^{-6/5}$.

LEMMA 4. $\sum_{|\beta|\le y;\ \beta\in\mathcal{A}} \chi(\sigma^*(\beta) \ge 1-t) \ll ty^2 + y$.

LEMMA 5. $\sum_{|\alpha|\le x;\ \alpha\in\mathcal{A}} \chi(s \le \sigma^*(\alpha) \le t) \ll (9/8)^n((t-s)x^2 + x)$, *uniformly in* s, t *satisfying* $0 \le s < t \le 1$ *and* $t - s \ge 3^{-n}$.

REMARK. The proof of this lemma was the central difficulty in this argument.

COROLLARY. *For any interval* $I = [s,t]$, *with* $0 \le s < t \le 1$,

$$\sum_{|\alpha|\le x;\ \alpha\in\mathcal{A}} \chi(\sigma^*(\alpha)\in I) \ll (t-s)^{7/8}x^2 + (t-s)^{-1/8}x.$$

If $t - s \le 1/x$, *the sum here is simply* $\ll x^{9/8}$.

LEMMA 6. *For every* $v \in V(k)$, *and every* $\Lambda_1 < \Lambda_2 \le 1/(n+1)$,

$$\sum_{|\alpha|\le x} \chi(\sigma^*(\alpha)\in I_\Lambda(v)) = \frac{6}{\pi}|I_\Lambda(v)|x^2 + O\left(\frac{x^2}{n^2}\right) + O(x^{9/8}).$$

COROLLARY.

$$\sum_{|\alpha|\le x} \chi(\sigma^*(\alpha)\in I(v)) = \frac{6}{\pi}x^2\left(1 + O\left(\frac{1}{n}\right)\right)|I(v)| + O(x^{9/8}).$$

The number of Farey n-intervals $I(v)$ corresponding to $v \in V(k)$, $k \le K$, is

$$\sum_{k=1}^{K} n^k = n^K\left(1 + O\left(\frac{1}{n}\right)\right).$$

Given an arbitrary interval $[s,t] \subseteq [0,1]$, we apply Lemma 6 to any Farey n-interval $I(v)$, with $v \in V(k)$ and $k \le K$, which contains as an interior point either s or t. There will be $0, 1$, or 2 such intervals. We apply the corollary to Lemma 6 to all the Farey n-intervals with $k \le K$ contained in $[s,t]$. Let $\mathcal{I}$ be the collection of all these intervals. Then

$$\begin{aligned}(5)\qquad \sum_{\substack{\alpha\in\mathcal{A}\\|\alpha|\le x}} \chi(\sigma^*(\alpha)\in[s,t]) &\ge \sum_{I\in\mathcal{I}}\sum_{\substack{\alpha\in\mathcal{A}\\|\alpha|\le x}} \chi(\sigma^*(\alpha)\in I)\\ &= \sum_{I\in\mathcal{I}} \frac{6}{\pi}x^2\left(|I|\left(1+O\left(\frac{1}{n}\right)\right) + O(x^{9/8})\#\mathcal{I}\right) + O\left(\frac{x^2}{n^2}\right)\\ &= \frac{6}{\pi}x^2\sum_{I\in\mathcal{I}}\left(|I| + O\left(\frac{x^2}{n}\right) + O(x^{9/8})\#\mathcal{I}\right).\end{aligned}$$

Now we take $n = [\frac{1}{2}\log x/(\log\log x)^2]$ and $k = [\log n]$. Then $\#\mathcal{I} \ll \exp(n\log^2 n) \ll \exp(\frac{1}{2}\log x) = \sqrt{x}$, so that $(x^{9/8}\#\mathcal{I}) \ll x^{13/8}$. The difference between $\sum_{I\in\mathcal{I}}$ and $(t-s)$ is $\ll (1-1/n)^{n\log n} \le 1/n$. It follows that

$$(6)\qquad \sum_{\substack{\alpha\in\mathcal{A}\\|\alpha|\le x}} \chi(\sigma^*(\alpha)\in[s,t]) \ge \frac{6}{\pi}x^2(t-s) + O\left(\frac{x^2(\log\log x)^2}{\log x}\right).$$

Finally, we apply (6) to $[0, s]$ and $[t, 1]$. Since $\sum_{\alpha\in\mathcal{A};\ |\alpha|\leq x} 1 = 6x^2/\pi + O(x)$ by Lemma 2, the inequality in (6) also goes the other way. Therefore

$$\sum_{\substack{\alpha\in\mathcal{A}\\ |\alpha|\leq x}} \chi(\sigma^*(\alpha) \in [s, t]) = \frac{6}{\pi}(t-s)x^2 + O\left(\frac{x^2(\log\log x)^2}{\log x}\right). \tag{7}$$

But $\sigma^*(\alpha) = s^*(\alpha) + O(|\alpha|^{-3})$, and if $\alpha = (a, b)$ then $q(a, b) = (a^2 + b^2)s^*(\alpha)$. Thus considering $[s \pm x^{6/5}, t \pm x^{-6/5}]$,

$$\sum_{\substack{|\alpha|\leq x\\ \alpha\in\mathcal{A}}} \chi(s^*(\alpha) \in [s, t]) = \frac{6}{\pi}(t-s)x^2 + O\left(\frac{x^2(\log\log x)^2}{\log x}\right),$$

as desired.

BIBLIOGRAPHY

1. E. Hecke, *Eine neu Art von Zetafunktionen und ihre Beziehungen zur Verteilung der Primzahlen*, Math. Z. **6** (1920), 11–51.
2. C. Hooley, *On the number of divisors of a quadratic polynomial*, Acta Math. **110** (1963), 97–114.
3. ——, *On the distribution of the roots of polynomial congruences*, Mathematika **11** (1964), 39–49.
4. H. Iwaniec, *Almost-primes represented by quadratic polynomials*, Invent. Math. **47** (1978), 171–188.
5. L. Kuipers and H. Niederreiter, *Uniform distribution of sequences*, Wiley, 1974.
6. E. Landau, *Vorlesungen über Zahlentheorie*, Chelsea, 1969.
7. O. Perron, *Die Lehre Von den Kettenbrüchen*, Chelsea, 1950.
8. F. Schweiger, *The metrical theory of Jacobi-Perron algorithm*, Lecture Notes in Math., vol. 334, Springer-Verlag, 1973.

DEPARTMENT OF MATHEMATICS, TEXAS A & M UNIVERSITY, COLLEGE STATION, TEXAS 77843