

THE STATISTICS OF THE CONTINUED FRACTION DIGIT SUM

DOUG HENSLEY

The statistics of the digits of a continued fraction, also known as partial quotients, have been studied at least since the time of Gauss. The usual measure m on the open interval $(0, 1)$ gives a probability space \mathcal{U} . Let a_k , $k \geq 1$ be integer-valued random variables which take $\alpha \in (0, 1)$ to the k^{th} partial quotient or digit in the continued fraction expansion $\alpha = 1/(a_1 + 1/(a_2 + \dots))$. Let $S_r = S_r(\alpha) = \sum_{k=1}^r a_k$. It is well known that although there is an average value for $\log a_k$, each a_k , let alone each S_r , has infinite expected value or first moment.

The main result of this work is that there exists a stable probability density function ϕ on \mathbf{R} so that

$$\lim_{r \rightarrow \infty} \sup_{z \in \mathbf{R}} \left| m(\{x \in (0, 1) : S_r(x) \leq z\}) - \int_{-\infty}^{z \log 2/r + \gamma - \log(r/\log 2)} \phi(x) dx \right| = 0.$$

Explicit error bounds, and some interesting properties of ϕ , are given. This ϕ occupies the boundary zone between distributions with support R and those with support of the form $[a, \infty)$, and was enough of an anomaly that Lévy classified it as only ‘quasi-stable’. Such distributions arise also in connection with the behavior of the sum of independent, identically distributed random variables of infinite expected value, and ϕ in particular is associated with the sum $X_1 + X_2 + \dots + X_n$ where X_j is the reciprocal of a random variable uniformly distributed on $[0, 1]$. Lévy considered this sum, and conjectured that $X_1 + X_2 + \dots + X_n < n(\log n + \log c - \log \log \log n)$ infinitely often if and only if $c < 1$. From the results obtained here, the triple logarithm should be a double, and the cutoff is at $c = e^{-\gamma}$. We give a quick proof that if $c < e^{-\gamma}$ then almost surely $X_1 + X_2 + \dots + X_n < n(\log n + \log c - \log \log n)$ occurs but finitely often.

A stable probability distribution function F is infinitely divisible and has the additional property that for all $a_1 > 0, a_2 > 0$ and for all real b_1, b_2 there exist $a > 0$ and b so that $F(a_1x + b_1) * F(a_2x + b_2) = F(ax + b)$. The history of our understanding of such F is convoluted. In the 1950’s, Lapin [10] claimed to have proved that the convolution of unimodal functions is

unimodal, but there were counterexamples, as noted by Chung in the course of translating the book ‘Limit distributions for sums of independent random variables’ [4]. In 1971, Wolfe published a paper [14, 1] on the unimodality of distributions in class L , a class which includes the stable distributions. His result was that the class of ‘one-sided’ distributions in L is unimodal. Later, Yamazato [15, 1] showed that from Wolfe’s theorem it follows that all distributions in class L are unimodal. Stable F , then, have unimodal density when the dust settles. Their characteristic functions (Fourier transforms) have the form

$$f(t) = \exp(i\gamma t - c|t|^\alpha \{1 + i\beta(t/|t|)\omega(t, \alpha)\})$$

where in general γ is an arbitrary real number, c is an arbitrary positive number, $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, and $\omega(t, \alpha) = \tan(\pi\alpha/2)$ if $\alpha \neq 1$, while if $\alpha = 1$ (the case at hand!), $\omega(t, \alpha) = (2/\pi) \log |t|$ [4]. For the particular stable density function ϕ , we have moreover:

- (1) ϕ is positive on \mathbf{R} .
- (2) $\phi(x) \approx \frac{1}{\sqrt{2\pi}} \exp(|x| - 1 - e^{|x|-1})$ as $x \rightarrow -\infty$.
- (3) $\phi(x) = x^{-2} + (2\gamma - 3 + 2 \log x)x^{-3} + O(x^{-4} \log^2 x)$ as $x \rightarrow \infty$.
- (4) ϕ extends to an entire function as the Laplace transform of $t^{-t} \sin \pi t$.
- (6) $\hat{\phi}(t) = \exp(-(\pi/2)|t| + it \log |t|)$.

A computer-generated plot of ϕ , and a couple of computer-generated histograms (10000 and 80000 values of S_{100} respectively), for the distribution of $S_r(X)$, are presented below. Some tricks are needed, and some corners have to be cut, to get this many data points in reasonable time. See the author’s home page on the Internet for details.

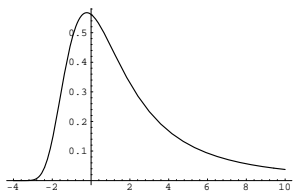


Fig. 1.

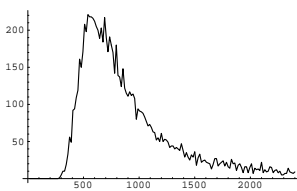


Fig. 2.

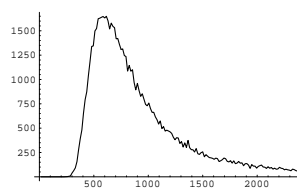


Fig. 3.

The proof of the main result breaks into two main parts. The typical digit sequence (a_i) consists mostly of small digits, punctuated by occasional larger digits. We choose the cutoff at R with $R := \lceil r^{1/2+\epsilon} \rceil$, reserving the choice of ϵ for later. (Eventually, ϵ is taken to be $1/12$.) There will almost always be on the close order of $(\log 2)^{-1} r^{1/2-\epsilon}$ of these large digits, and they are effectively statistically independent. The conditional expected value M of a small digit,

(one bounded by R), is finite. The sum of the small digits normally turns out to contribute rather more than half the grand total, and for most x is close to rM . That is, while the shape of the distribution function is determined by the large digits, the small digits move that distribution (of the sum of the large digits) to the right by a nearly-deterministic rM . Proving this involves some probabilistic approximation of the original ‘game’ with other, more suitable games that assign very nearly the same probability to ‘most’ of the possible digit sequences, followed by an appeal to known facts concerning the functional analysis of continued fractions. The idea of approximating a measure G with a similar but more tractable G^* goes back to P. Lévy, who used it in [11] to study continued fraction questions similar to ours.

The distribution of the sum of the large digits is effectively modeled by the sum of $N = [(\log 2)^{-1}r^{1/2-\epsilon}]$ independent, identically distributed random variables, each with probability density function $f_R := R/x^2$ on the interval (R, ∞) . The second main part of the proof is an analysis of this sum. For all stable laws F with *characteristic exponent* $\alpha > 1$, F extends from the real line to an entire function, [4, 10]. In our case, $\alpha = 1$ and so a particular argument is needed. From the Fourier transform of ϕ we are able to work back to the asymptotic behavior of $\phi(x)$ as $x \rightarrow \infty$ and as $x \rightarrow -\infty$, by way of classical complex analysis starting with the formula for the inverse Fourier transform. The analysis of the sum of the large digits yields, en passant, the asymptotic distribution of the sum of n independent identically distributed random variables, each with the same probability density function ($1/x^2$ for $x > 1$ else 0) as $1/X$: It too is a scaled and shifted copy of $\phi(x)$: As $n \rightarrow \infty$,

$$\text{prob} \left[\sum_1^n 1/X_k \leq n(\log n + 1 - \gamma + Z) \right] \rightarrow \int_{-\infty}^Z \phi(x) dx$$

uniformly over $Z \in R$.

2. Games with essentially equal probability measures. Here we introduce a variety of countable probability spaces, described as games of chance. Apart from the first ‘game’, in which a single number is chosen from the open interval $(0,1)$ with the usual measure, our games are Markov chains. The states of such a chain can involve a positive integer, a real number in $(0, 1)$, or both, but there is no dependence on the past, as there is in the evolution of the digits of the continued fraction expansion of a random real number. Two probability measures μ_j and μ_k say, both on the same countable event set, will be termed equivalent to within a given margin of error δ , if the set of elements a for which $\mu_j(a) \notin (1 \pm \delta) \mu_k(a)$, or vice-versa, has measure less than δ in both spaces.

The original game is to pick a random number $X \in (0,1)$, extract the first r partial quotients or digits in its continued fraction expansion, and form their sum. The issue is to characterize the asymptotic behavior of

the distribution function $p_r(z) := \text{prob}(S_r(X)) \leq z$. This game, call it G_1 , is in some respects an unsatisfactory representation of random continued fraction digits. There is some short-range correlation between one digit and the previous ones in the continued fraction expansion of X ; this game does not amount to a Markov chain on the digits. The game G_2 described below provides a Markov chain (game) for which the probability of any digits string is the same as in game 1.

Play proceeds through r stages, with one digit being chosen at each stage. At the outset, a state parameter θ is set to zero. At stage $k+1$, the probability of choosing $a_{k+1} = n$ is $(1 + \theta)/((n + \theta)(n + \theta + 1))$. Once a_{k+1} is chosen, set $\theta \leftarrow 1/(\theta + a_{k+1})$. The probability of a given sequence (a_1, a_2, \dots, a_r) of continued fraction partial quotients is exactly the same in both games: $\langle a_1, a_2, \dots, a_r \rangle^{-2} (1 + \{a_1, a_2, \dots, a_r\})^{-1}$, where $\{a_1, a_2, \dots, a_r\}$ denotes the continued fraction $1/(a_r + 1/(a_{r-1} + \dots + 1/a_1) \dots)$, and $\langle a_1, a_2, \dots, a_r \rangle$, its denominator. This fraction is the value of θ after those r digits have been chosen, incidentally. An equivalent characterisation of this game is that after digits $a_1 \cdots a_k$ have been chosen, a real number Y in $[0, 1]$ is chosen according to the probability density function $(1 + \{a\})/(1 + \{a\}t)^2$ where $\{a\}$ denotes the continued fraction $[a_k, a_{k-1}, \dots, a_1]$. The next digit is then chosen to be the integer part of $1/Y$. The other variants introduced below, of the ‘game’ of choosing digits, can likewise be given this kind of formulation. In every case, the conditional probability density function for Y is a positive linear combination of functions of the form $(1 + \theta)/(1 + \theta t)^2$, with $0 \leq \theta \leq 1$.

Equivalent $_{r-\epsilon}$ variants of this game change the probabilities, but rarely by much. That is, the set of sequences a of r positive integers falls into two subsets, ‘typical’ and ‘atypical’ say. The probabilities the two games assign to elements of ‘typical’ differ by a factor of $1 \pm O(r^{-\epsilon})$. The probabilities assigned to elements of ‘atypical’ may be quite different, but the whole of ‘atypical’ has probability $O(r^{-\epsilon})$ in both games. One such game is G_3 . In this game, we round θ down to zero following any occurrence of a large digit: $a_k \geq R \Rightarrow \theta \leftarrow 0$. Again, the ‘state’ of this Markov chain is the current value of θ . This game has the property that for all $1 \leq j \leq r$ and all sequences $a = (a_1, a_2, \dots, a_r)$ with $a_j \geq R$,

$$\text{prob}_{G_3}(a) = \text{prob}_{G_3}(a_1 \cdots a_j) \text{prob}_{G_3}(a_{j+1} \cdots a_r).$$

To see that games 2 and 3 are nearly equivalent, and find the extent δ to which they are not identical (in the probabilities assigned to digit sequences), we begin with the observation that in both games, the conditional probability, given an arbitrary initial sequence (a_1, a_2, \dots, a_k) of digits, that the next digit a_{k+1} will be large, is never more than $2/R$. Therefore, in both games the probability of having several large digits is small: If N denotes the number of large digits, then both $\text{prob}_{G_1}(N \geq z) = \text{prob}_{G_2}(N \geq z)$

and $\text{prob}_{G_3}(N \geq z)$, are less than the probability of z or more heads in r independent tosses of a coin that shows heads with probability $2/R$ and tails with probability $1 - 2/R$. Taking $z := \lceil 3r^{1/2+\epsilon} \rceil$ makes even this latter probability $\ll r^{\epsilon-1/2}$ on grounds of Cauchy's inequality. Apart from sequences (a_1, \dots, a_r) with $N(a) \geq 3r^{1/2-\epsilon}$, though, all digit sequences have essentially the same probability in either game. First of all, the two games assign exactly the same probability to any sequence a with no large digits. Consider next a sequence with exactly one large digit, say a_k .

The probability of a digit sequence $(a_1 \cdots a_k, a_{k+1} = b_1 \cdots a_r = b_{r-k}) = ab$, say, is given in games 1 and 2 by

$$\langle a \rangle^{-2} \langle b \rangle^{-2} (1 + \{a\} \{b\})^{-2} (1 + \{ab\})^{-1}$$

while in game 3, and assuming $a_k \geq R$, it is

$$\langle a \rangle^{-2} \langle b \rangle^{-2} (1 + \{a\})^{-1} (1 + \{b\})^{-1}.$$

Now $\{a\} \leq 1/R$, and $\{b\} = (1 + O(1/R))\{ab\}$. The ratio here is $(1 + O(1/R))$. For a sequence a with N large digits, the same calculation shows that the ratio of probabilities assigned to a by games 2 and 3 is $(1 + O(1/R))^N$, and with $N \ll r^{1/2-\epsilon}$ that's $1 + O(r^{-2\epsilon})$. So we have equivalence δ with $\delta = O(r^{-2\epsilon})$.

For some purposes we need yet more statistical independence in our model of digit-choosing. In game 4, large and small digits are chosen independently. This game is played exactly like game 3 except on those occasions when a large digit happens to be chosen. In this case, instead of simply using it and resetting the auxiliary θ to zero, we discard the large digit chosen, and draw again from the "urn of large digits", this time with $\text{prob}_4(n) := R/(n(n+1))$ for $n \geq R$, and zero otherwise. We then set θ to zero as before. The new probability differs by a factor of $(1 + O(1/R))$ from the conditional probability in game 3 that digit $a_k = n$ given the history $a_1 \cdots a_{k-1}$ (i.e. given θ), and given that $a_k \geq R$. Since the probabilities in games 3 and 4 are changed only by reshuffling the mass allotted to the individual large digits, the probabilities for N are unchanged and this game, too, is equivalent $_\delta$ to the others, still with $\delta = O(r^{-2\epsilon})$. Consequently, insofar as Theorem 1 is concerned, all these games are equivalent.

The point of game 4 is that we can partition the event space according to the subset $T \subset \{1, 2, \dots, r\}$ of places at which a_k is large ($\geq R$). For fixed T , we consider the conditional games $G_4(T)$ described below. Given a set $T \subset \{1, 2, \dots, R\}$, and a sequence $a = (b_0 n_1 b_1 \cdots n_N b_N)$ in which the b_j represent strings or sequences of $k_j := (t_{j+1} - t_j - 1)$ integers all between 1 and $R - 1$, and the n_j are integers $\geq R$, (that is, a sequence of r positive integers for which T is the set of positions at which 'large' integers are found), game $G_4(T)$ assigns probability $\text{prob}_4(a)/\text{prob}_4(T)$ to a . Equivalently, with $N = N(T) =$ the cardinality of T , and with b 's, k 's

and n 's defined in terms of a as above, we have

$$\begin{aligned} & \text{prob}_{4,T}(a) \\ &= \frac{\prod_{j=0}^{N-1} \langle b_j \rangle^{-2} \left(\frac{R}{R + \{b_j\}} \right) \prod_{j=1}^N n_j^{-1} (n_j + 1)^{-1} \cdot \langle b_N \rangle^{-2} (1 + \{b_N\})^{-1}}{\prod_{j=0}^{N-1} \sum_{b \in V_R(k_j)} \langle b \rangle^{-2} (R + \{b\})^{-1} \cdot \sum_{b \in V_R(k_N)} \langle b \rangle^{-2} (1 + \{b\})^{-1}}. \end{aligned}$$

For most T , (including all T with $N(T) \leq 4r^{1/2-\epsilon}$), this conditional probability agrees, to within a factor of $(1 \pm O(r^{-2\epsilon}))$ with

$$\frac{\prod_{j=0}^{N-1} \langle b_j \rangle^{-2} \prod_{j=1}^N R n_j^{-1} (n_j + 1)^{-1} \cdot \langle b_N \rangle^{-2} (1 + \{b_N\})^{-1}}{\prod_{j=0}^{N-1} \sum_{b \in V_R(k_j)} \langle b \rangle^{-2} \cdot \sum_{b \in V_R(k_N)} \langle b \rangle^{-2} (1 + \{b\})^{-1}}.$$

As mentioned earlier, the sum of the small digits is a random variable with relatively little dispersion. Specifically, we shall see that $N(T) \geq 3r^{1/2-\epsilon}$ with probability $O(r^{-2\epsilon})$ (actually, far less), and that for all other T , the conditional probability in game 4 (T) that the sum of the small digits falls outside $\frac{r}{\log 2} \left(\left(\frac{1}{2} + \epsilon \right) \log r - 1 \right) + O(r^{(3/4)+2\epsilon})$ is $O(r^{-\epsilon})$. From the perspective of the theorem, this may as well say that the sum of the small digits is always $\frac{r}{\log 2} \left(\left(\frac{1}{2} + \epsilon \right) \log r - 1 \right)$. There is a deep and fruitful connection [7, 8, 12] between the metric theory of continued fractions and the linear operator

$$G : f(t) \rightarrow \sum_{k=1}^{\infty} (k+t)^{-2} f(1/(k+t)).$$

This operator, and a variant L in which the upper limit of summation is $R-1$, will be at the center of the calculations ahead. At this point it will be convenient to introduce some more terminology.

Recall that $\langle b \rangle = \langle (b_1, b_2, \dots, b_m) \rangle$ represents the denominator of $1/(b_1 + 1/(b_2 + 1/(\dots + 1/b_m) \dots))$, that $[b]$ represents the fraction itself, and $\{b\}$, the fraction corresponding to the sequence $(b_m, b_{m-1}, \dots, b_1)$. Let $V(k)$ denote the Cartesian product of k copies of Z^+ , and $V_m(k)$ the Cartesian product of k copies of $\{1, 2, \dots, m\}$. Let $\|b\|$ denote $\sum_{j=1}^k b_j$ for $b \in V(k)$. Let L_m be the operator carrying f (a function from C to C) to the function $t \rightarrow \sum_{j=1}^m (j+t)^{-2} f(1/(j+t))$, also defined on C . For our purposes the important thing will be the action of powers of L_{R-1} on constant functions, and on functions of the form $t \rightarrow (1 + \theta t)^{-2}$.

The basis of the connection between this operator and questions of continued fractions is the fact that

$$L_m^k(1) = \sum_{b \in V_m(k)} \langle b \rangle^{-2} (1 + \{b\}t)^{-2}$$

and that $g = g_\infty = 1/(\log 2(1+t))$ is the leading eigenvalue of L_∞ . More generally,

$$L_m^k(1+\theta t)^{-2} = \sum_{b \in V_m(k)} \langle \theta + b \rangle^{-2} (1 + \{\theta + b\}t)^{-2}$$

where $\theta + b$ denotes the sequence $(\theta + b_1, b_2, \dots, b_k)$ and where $0 \leq \theta \leq 1$. We quote a result [7, Lemma 2.1 etc.], adapted and specialized to the case at hand:

Lemma 1. *There is a sequence of probability density functions $g_m : [0, 1] \rightarrow \mathbb{R}^+$, and numbers $\lambda_m \in (0, 1)$, so that*

$$\begin{aligned} \text{(a)} \quad & L_m g_m = \lambda_m g_m \\ \text{(b)} \quad & g_m = \frac{1}{\log 2} \frac{1}{1+t} \left(1 + O(m^{-1/2})\right) \\ \text{(c)} \quad & \lambda_m = 1 - \frac{6}{\pi^2 m} + O(m^{-2} \log m) \\ \text{(d)} \quad & L_m^k(1+\theta t)^{-2} = \frac{\lambda_m^k g(t)}{1+\theta} \left(1 + O(m^{-1/2}) + O(3^{-k})\right) \end{aligned}$$

The functions g_m have the form $\int_0^1 (1+\theta)(1+\theta t)^{-2} d\mu_m(\theta)$ where μ_m is a probability measure on $[0, 1]$.

With this machinery we now set about getting estimates for the quantities

$$\begin{aligned} & \sum_{b \in V_m(k)} \langle b \rangle^{-2}, \quad \sum_{b \in V_m(k)} \langle b \rangle^{-2} (1 + \{b\})^{-1} \\ & \sum_{b \in V_m(k)} \langle b \rangle^{-2} \|b\|, \quad \text{and} \quad \sum_{b \in V_m(k)} \langle b \rangle^{-2} (1 + \{b\})^{-1} \|b\| \end{aligned}$$

with particular emphasis on the case $m = R - 1$, $k \in [r^\epsilon, r^{1/2+2\epsilon}]$. Apart from a negligible fraction (no more than $O(r^{-\epsilon})$) of the mass of a 's, weighted according to the probabilities in any of the games $G_1 \dots G_4$, every interval k_j between consecutive large digits is of length at least r^ϵ . Indeed, $\text{prob}[\exists \text{ gap} < r^\epsilon] \leq \sum_{j=1}^r \sum_{k=1}^{r^\epsilon} \text{prob}[\text{both } a_j \text{ and } a_{j+k} \geq R-1]$. A long run of small digits is also unlikely: As we have seen, the conditional probability, in all games and for all antecedent digit strings a , that the next digit is large is always between $1/R$ and $2/R$. Thus in games 3 and 4, the probability of a string of k or more consecutive small digits in positions $j \dots j+k-1$ say, is $\leq (1-1/R)^k$ and so the probability of any such string is $\leq r(1-1/R)^k$. Setting $k = [R^{1/2+2\epsilon}]$ gives the claimed inequality.

The estimates from Lemma 1, applied in this case, lead to

$$\text{(i)} \quad \sum_{b \in V_m(k)} \langle b \rangle^{-2} (1 + \{b\}t)^{-2} = (1 + O(\exp(-r^\epsilon))) \lambda_m^k g_m(t),$$

$$(ii) \quad \sum_{b \in V_m(k)} \langle b \rangle^{-2} = (1 + O(\exp(-r^\epsilon))) \lambda_m^k g_m(0),$$

$$(iii) \quad \sum_{b \in V_m(k)} \langle b \rangle^{-2} (1 + \{b\})^{-1} = (1 + O(\exp(-r^\epsilon))) \lambda_m^k.$$

For the next sum

$$\sum_{b \in V_m(k)} \langle b \rangle^{-2} (1 + \{b\})^{-1} \|b\|$$

we start with the observation that this is

$$\sum_{j=1}^k \sum_{c \in V_m(j-1)} \sum_{d \in V_m(k-j)} \sum_{n=1}^m n \langle cnd \rangle^{-2} (1 + \{cnd\})^{-1}$$

where the juxtaposition cnd denotes the concatenation of the sequences c and d with the single entry n between them. Now $\{cnd\} = (1 + O(\rho^{k-j}))\{d\}$ where $\rho = 2/(1 + \sqrt{5})$. The sum here thus simplifies to

$$\sum_{j=1}^k (1 + O(\rho^{k-j})) \sum_{c \in V_m(j-1)} \sum_{n=1}^m \langle c \rangle^{-2} \frac{n}{(n + \{c\})^2}$$

times an inner sum of

$$\int_0^1 \sum_{d \in V_m(k-j)} \langle d \rangle^{-2} (1 + \theta[d])^{-2} (1 + \{\theta + d\}t)^{-2} dt$$

where $\theta = 1/(n + \{c\})$. From the identities above, this inner sum is $\int_0^1 L_m^{k-j}(1 + \theta t)^{-2} dt$. But by Lemma 1, this inner sum is $\lambda_m^{k-j} (1 + 1/(n + \{c\}))^{-1} (1 + O(m^{-1/2}) + O(3^{j-k}))$ so that

$$\begin{aligned} & \sum_{b \in V_m(k)} \langle b \rangle^{-2} (1 + \{b\})^{-1} \|b\| \\ &= \sum_{j=1}^k (1 + O(m^{-1/2}) + O(\rho^{k-j})) \lambda_m^{k-j} \\ & \quad \cdot \sum_{c \in V_m(j-1)} \sum_{n=1}^m \langle c \rangle^{-2} \frac{n}{(n + \{c\})(n + 1 + \{c\})}. \end{aligned}$$

The inner double sum can be simplified; with $h = j - 1$ it is

$$\begin{aligned} & \sum_{c \in V_m(h)} \langle c \rangle^{-2} \sum_{n=1}^m n \int_{1/(n+1)}^{1/n} L_m^h(1)(t) dt \\ &= \lambda_m^h (\log 2)^{-1} \sum_{n=1}^m n \log \left(\frac{(n+1)^2}{n(n+2)} \right) \end{aligned}$$

$$\begin{aligned}
& + O(3^{-h} \lambda_m^h \log m) + O\left(\lambda_m^h \log m / \sqrt{m}\right) \\
& = \lambda_m^h \left(\frac{\log m - 1}{\log 2} + O(\log m / \sqrt{m}) + O(3^{-h} \log m) \right).
\end{aligned}$$

and so

$$\begin{aligned}
& \sum_{b \in V_m(k)} \langle b \rangle^{-2} (1 + \{b\})^{-1} \|b\| \\
& = k \lambda_m^{k-1} \left(\frac{\log m - 1}{\log 2} \right) (1 + O(m^{-1/2}) + O(k^{-1})).
\end{aligned}$$

In the same way we have

$$\sum_{b \in V_m(k)} \langle b \rangle^{-2} \|b\| = k \lambda_m^{k-1} \left(\frac{\log m - 1}{\log^2 2} \right) (1 + O(m^{-1/2}) + O(k^{-1})).$$

Now given a fixed set T of positions at which large digits are to occur, with $N(T) \ll r^{1/2-\epsilon}$ and all gaps between consecutive elements of T at least r^ϵ , the expected value in game 4(T) of the sum of the j^{th} sequence of k_j consecutive small digits (bordered by large digits on both ends or, in the last case, just on the left), is $(1 + O(R^{-1/2}) + O(k_j^{-1}))k_j(\log m - 1)/\log 2$ and summing this over all N gaps gives an expected value for the sum of the small digits of $(1 + O(R^{-1/2}))r(\log R - 1)/\log 2$. Since this is the same, to within that margin of error, for all relevant values of T , the conditional expected value of the sum of the small digits, given that T satisfies the constraints $N(T) \ll r^{1/2-\epsilon}$, with all gaps $\geq r^\epsilon$, is $(1 + O(r^{-1/4}))r(\log R - 1)/\log 2$. We turn now to the question of the dispersion of the value of $\|b\|$ about this mean, and as one might expect, by way of an estimate of the second moment of $\|b\|$. For fixed T the events $\|b_0\|, \|b_1\| \dots \|b_N\|$ are independent. Thus

$$\text{Var} \sum_{j=0}^N \|b_j\| = \sum_{j=0}^N \text{Var} \|b_j\| \leq \sum_{j=0}^N E[\|b_j\|^2].$$

For a string of k consecutive small digits, we can estimate $E[\|b\|^2]$. The idea of the estimate is this: We break the dependence of the digits of b by assuming at each new digit the worst: $\theta = 1$, or what is roughly equivalent, the previous two digits were a large one followed by 1. This will skew to the right the distribution of $\|b\|$ and increase its second moment, which latter second moment is easily calculated because we finally have independent random digits. The technicalities of the next few paragraphs are devoted to justifying the claim that changing the odds in this fashion really does shift the distribution of $\|b\|$ right. In this calculation we are concerned with the conditional probability, (in any of games 2 through 4) that $\|b\| = t$, given an initial value for θ , and given that for (at least) the next k digits, no digit as large as R occurs. With these conditions, the probability of a sequence

b is the product of the conditional probabilities for the consecutive digits of b , when those are given by the formulae

$$p(\theta, n) := \frac{\int_{s=1/(n+1)}^{1/n} (1 + \theta s)^{-2} ds}{\int_{s=1/R}^1 (1 + \theta s)^{-2} ds}, \quad p(b) := \prod_{j=1}^k p(\{b_1, b_2, \dots, b_{j-1}\}, b_j).$$

If we use instead the worst-case odds, we have a different probability:

$$q(n) := \frac{\int_{s=1/(n+1)}^{1/n} (1 + s)^{-2} ds}{\int_{s=1/R}^1 (1 + s)^{-2} ds}, \quad q(b) := \prod_{j=1}^k q(b_j)$$

in which the digits of b are independent. Now let $p_k(t) := \text{prob}_p(\|b\| = t) = \sum_{b \in V_{R^{-1}(k)}} p(b)$, and $q_k(t) := \text{prob}_q(\|b\| = t) = \sum_{b \in V_{R^{-1}(k)}} q(b)$. The sequences $p_k(t), (t \geq 1)$ and $q_k(t), (t \geq 1)$ are sequences of finitely many positive numbers followed by zeros, and with sum 1. The claim above about skewed distributions comes to this: $(p_k(t))$ is more peaked than $(q_k(t))$. Equivalently, $\sum_{j=1}^t p_k(j) \geq \sum_{j=1}^t q_k(j)$ for all $t \geq 1$ and all $k \geq 1$. For $k = 1$, this claim boils down to the readily verified proposition that for $t \geq 1$ and $0 \leq \theta \leq 1$,

$$\frac{\int_{1/(t+1)}^1 (1 + \theta s)^{-2} ds}{\int_{1/R}^1 (1 + \theta s)^{-2} ds} \geq \frac{\int_{1/(t+1)}^1 (1 + s)^{-2} ds}{\int_{1/R}^1 (1 + s)^{-2} ds}.$$

Assuming this peakedness inequality for $k = K$, we now consider the case $k = K + 1$. We have

$$\begin{aligned} & \sum_{j=1}^t p_{K+1}(j) \\ &= \sum_{n=1}^t \sum_{\substack{a \in V_{R^{-1}(K)} \\ \|a\|=t-n}} \langle a \rangle^{-2} (1 + \{a\})^{-1} \frac{\int_{\max[1/(n+1), 1/R]}^1 (1 + s\{a\})^{-2} ds}{\int_{1/R}^1 (1 + s\{a\})^{-2} ds} \\ &\geq \sum_{n=1}^t \sum_{\substack{a \in V_{R^{-1}(K)} \\ \|a\|=t-n}} \langle a \rangle^{-2} (1 + \{a\})^{-1} \frac{\int_{\max[1/(n+1), 1/R]}^1 (1 + s)^{-2} ds}{\int_{1/R}^1 (1 + s)^{-2} ds} \\ &= \sum_{n=1}^t \sum_{\substack{a \in V_{R^{-1}(K)} \\ \|a\|=t-n}} \langle a \rangle^{-2} (1 + \{a\})^{-1} \sum_{j=1}^n q_1(j) = \sum_{n=1}^t Q_1(n) p_K(t-n) \end{aligned}$$

say, where $Q_1(n) := \sum_{j=1}^n q_1(j)$. On the other hand,

$$Q_{K+1}(t) = \sum_{j=1}^t q_{K+1}(j) = \sum_{n=1}^t Q_1(n) q_K(t-n).$$

We are now in a position to use a simple peakedness lemma. The proof is an exercise in summation by parts and is left to the reader.

Lemma 2. *If (c_n) is a positive, non-decreasing sequence, and if (d_n) and (e_n) are sequences of positive numbers with (d_n) more peaked than (e_n) , then for all $t \geq 1$, $\sum_{n=1}^{t-1} c_n d_{t-n} \geq \sum_{n=1}^{t-1} c_n e_{t-n}$. From this lemma, then, for all $t \geq 1$ we have $\sum_{j=1}^t p_K(j) \geq \sum_{j=1}^t q_K(j)$ as claimed, which completes the induction.*

Choosing k small digits for b according to the probabilities q is a game of k independent trials. The mean or expected value of a single digit's contribution to $\|b\|$ is $E_q(k) = \sum_{n=1}^{R-1} \frac{2n(R+1)}{(R-1)(n+1)(n+2)} \approx \log R$, and the variance is $\text{Var}_q(k) = \sum_{n=1}^{R-1} \frac{2(R+1)(n-E_q(k))^2}{(R-1)(n+1)(n+2)} \ll R$. Hence for $\|b\|$ itself, the variance is dominated by kR . Therefore the second moment, with q probabilities, of $\|b\|$ is dominated by $kR + (k \log R)^2$, so that also for $\|b\|$ with the original conditional probabilities p , conditional probabilities the second moment is dominated by $kR + (k \log R)^2$. That is, we have shown that

$$\lambda_{R-1}^{-k} \sum_{b \in V_{R-1}} \|b\|^2 \langle b \rangle^{-2} (1 + \{b\})^{-1} \ll kR + (k \log R)^2.$$

Now in game 4, the various sequences of small digits, punctuated by digits $\geq R$, are themselves independent. Thus for all $T \subset \{1, 2, \dots, r\}$ with maximum gap $k_j \leq 2r^{1/2+\epsilon}$, the conditional expected second moment, given that large digits are found at all positions corresponding to T and only there, is

$$\begin{aligned} & \sum_{j=0}^{N(T)} \sum_{b_0 \in V_{R-1}(k_0), \dots, b_{N(T)} \in V_{R-1}(k_{N(T)})} \|b_j\|^2 \text{prob}(b_0 \wedge b_1 \cdots \wedge b_{N(T)}) \\ & \ll \left[r^{1/2+\epsilon} R + r^{1+2\epsilon} \log^2 r \right] \cdot r^{1/2+\epsilon} \ll r^{3/2+\epsilon} \log^2 r. \end{aligned}$$

Therefore, in all but intervals of aggregate length $O(r^{-\epsilon})$ in $[0, 1]$, the sum of the small digits amongst the first r digits is $(r/\log 2)((1/2 + \epsilon) \log r - 1) + O(r^{3/4+2\epsilon})$.

This brings us to the question of the statistics of the large digits. While the small digits do contribute to the determination of the set-point of the distribution of $\|a\|$, they have essentially nothing to do with its shape. The sum of the large digits is the chancy part; the big plays decide the game.

3. The number of large digits. For fixed T , and assuming that $N(T) \leq r^{1/2-\epsilon}$, the sequence of large digits in game 4 (T) is a sequence of $N(T)$ independent, identically distributed random variables. The probability of a given digit taking value $n \geq R$ is $R/(n(n+1))$; these probabilities sum to 1. We begin the analysis of the distribution of the sum of these digits by observing that the distribution will be essentially the same if we use the

continuous probability density function Rt^{-2} on the interval $[R, \infty)$ in place of the original discrete measure: the effect is the same as if, after choosing a digit n , we add a random fractional quantity with probability density $n(n+1)(n+t)^{-2}$ on $[n, n+1]$. There is enough leeway in the statement of the main result that this change cannot affect the conclusion.

Let $Y_j, 1 \leq j \leq N$ be a sequence of N independent random variables with probability density function $f_R(t) := Rt^{-2}$ on $[R, \infty)$ (and 0 otherwise). The random variable $\text{LDSum}_N := \sum_{j=1}^N Y_j$ has a probability density function $\rho_{R,N}(x)$ which is the N -fold convolution of f_R . Before getting deeply involved in the estimation of $\hat{\rho}$ and eventually ρ itself, we need to pin down N more tightly.

In game 4, the probability of an initial run of exactly m consecutive small digits is

$$\begin{aligned} \sum_{b \in V_R(m)} \langle b \rangle^{-2} \int_0^{1/R} (1 + \{b\}t)^{-2} dt &= \int_0^{1/R} (L^m 1)(t) dt \\ &= (1 + O(R^{-1/2}) + O(3^{-m})) \frac{\lambda_R^m}{(R \log 2)}. \end{aligned}$$

This is also the conditional probability of such a run immediately following a large digit. Thus the conditional probability of a run ending with the very next digit, given that there have been exactly k small digits since the most recent large digit or since the beginning, is $(1 + O(3^{-k}) + O(R^{-1/2})) / (R \log 2)$.

In game 4, $\text{prob}[N(T) \geq 4r^{1/2-\epsilon}] \ll r^{-\epsilon}$, and $\text{prob}[\min[k_j] \leq r^\epsilon] \ll r^{-\epsilon}$. For any other T , we have

$$\begin{aligned} &\text{prob}_4[\text{large digits occur exactly for indices in } T] \\ &= \prod_{j=0}^{N(T)} \prod_{i=1}^{k_j} (1 - (1 + O(3^{-i})) / R \log 2) \cdot \prod_{j=1}^{N(T)} \frac{1}{R \log 2} (1 + O(3^{-k_j})) \\ &= (1 + O(r^{-\epsilon})) (R \log 2)^{-N(T)} (1 - 1/R \log 2)^{r-N(T)}. \end{aligned}$$

Apart from the error factor in front, this last is the probability of getting heads at exactly the tosses corresponding to elements in T , on r Bernoulli trials of a coin with probability of heads equal to $1/R \log 2$ and tails, $1 - 1/R \log 2$. Elementary calculations give $\text{prob}_{\text{Bernoulli}}[\#\text{heads} \notin \frac{r}{R \log 2} \pm O(r^{1/4})] \ll r^{-\epsilon}$ and in view of the previous estimates, the same holds for $\text{prob}_4[N(T) \notin \frac{r}{R \log 2} \pm O(r^{1/4})]$. This brings us to the main issue of the next section: what do we get when we convolve N copies of the function $f_R(x) = Rx^{-2}$ for $x \geq R$, and zero otherwise? The precise value of N , between $(r^{1/2-\epsilon} / \log 2 \pm Cr^{1/4})$, is immaterial to our purposes since the convolution of $Cr^{1/4}$ copies of f_R is, as we shall see, a probability density function with

all but $O(r^{-\epsilon})$ of its mass concentrated into the (sufficiently narrow) interval $[0, r^{(3/4)+2\epsilon}]$.

4. The sum of the reciprocals of random numbers in $(0, 1)$. To study the convolution powers of f_R , it will suffice to investigate the convolution powers of $u(x) := 1/x^2 (x \geq 1), 0 (x < 1)$, as $f_R^{*n} = (1/R)u^{*n}(x/R)$. Note that $u^{*n}(x)$ is the probability density function of the sum of n independent, identically distributed random variables $1/X_j$ when X_j has uniform distribution on $(0, 1)$. (Hence the title of this section.)

The distribution function corresponding to the density u is of course $U(x) := 1 - 1/x (x \geq 1)$. According to a theorem of Gnedenko [4, Theorem 5, p. 182], U belongs to the ‘domain of attraction’ of a stable law with $\alpha = 1$. We need to know how fast this attraction works. With $\hat{v} := f \rightarrow (t \rightarrow \int_{-\infty}^{\infty} f(z)e^{-izt} dz)$, we have

$$\begin{aligned} v(t) &:= \hat{u}(t) = \int_1^{\infty} z^{-2} e^{-izt} dz \\ &= 1 + it \log |t| - \frac{1}{2} \pi |t| + it(\gamma - 1) - \sum_{k=2}^{\infty} \frac{(-it)^k}{(k-1)k!} \end{aligned}$$

which is the known function $Ei(2, t)$. The series part of the expansion extends to an entire function, but the logarithmic term requires a branch cut, which we take along the negative real axis. This gives an extension of $v(t)$ to the rest of C . For completeness we extend v to all of C by taking limits from above the branch cut. Conversely, $u(z) := \check{v}(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} v(t)e^{itz} dt$.

Let $u_1(z) := u(z)$, and for $n \geq 1$, let $u_n := u * u_{n-1}$. Then $\widehat{u}_n(z) = (v(t))^n$. From the series expansion of $v(t)$, for $n > 1$ we have

$$v^n(t) = \exp\left(-\frac{\pi}{2}n|t| + int \log |t| + int(\gamma - 1) + O(nt^2(1 + \log^2 |t|))\right)$$

uniformly in $|t| \leq 1/(\sqrt{n} \log n)$. Now let

$$\begin{aligned} \phi(z) &:= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{\pi}{2}|s| + is \log |s| + isz\right) ds \\ &\text{so that } \hat{\phi}(t) = \exp\left(-\frac{\pi}{2}|t| + it \log |t|\right). \end{aligned}$$

Note that for $z \in R$, $|\phi(z)| \leq 1$ and $\int_{-\infty}^{\infty} \phi(z) dz = 1$, while for $t \in R$, $|\hat{\phi}(t)| \leq 1$, with $\hat{\phi}(0) = 1$. Another observation is that the Fourier transforms of suitably shifted and scaled versions of ϕ are quite close to $v^n(t)$:

$$\left[\frac{1}{n} \phi \left(\frac{z}{n} - \widehat{\log n + \gamma - 1} \right) \right] (t) = e^{-int(\log n + \gamma - 1)} \hat{\phi}(nt)$$

$$= \exp\left(-\frac{\pi}{2}n|t| + int \log |t| + int(\gamma - 1)\right).$$

Thus for $n > 1$ and uniformly over $t \in R$ with $|t| \leq 1/(\sqrt{n} \log^2 n)$,

$$v^n(t) = \left[\frac{1}{n} \phi\left(\frac{z}{n} - \log n + \gamma - 1\right) \right](t) \cdot (1 + O(nt^2(1 + \log^2 |t|))).$$

We now claim that uniformly for $z \in R$,

$$\left| \int_{-\infty}^z u_n(x) - \frac{1}{n} \phi\left(\frac{x}{n} - \log n + \gamma - 1\right) dx \right| \ll \frac{\log n}{\sqrt{n}} \quad \text{and}$$

$$\left| \int_{-\infty}^z (f_R(x))^{*n} - \frac{1}{nR} \phi\left(\frac{x}{nR} - \log n + \gamma - 1\right) dx \right| \ll \frac{\log n}{\sqrt{n}}.$$

Proof. According to the Berry-Esseen inequality [3], ‘If $F(x)$ and $G(x)$ are probability distribution functions with Fourier transforms $f(t)$ and $g(t)$ respectively, then for all $T > 0$,

$$\sup_{x \in R} |F(x) - G(x)|$$

$$\leq c_1 \left\{ \sup_{y \in R} \frac{1}{2} T \int_0^{1/T} (G(y+u) - G(y-u)) du + \int_{-T}^T \left| \frac{f(t) - g(t)}{t} \right| dt \right\}$$

where $c_1 > 0$ is an effectively computable constant.’ (Remark: The constant may be computed directly from the proof given in [3]; I get $c_1 = 24.69\dots$)

We take $G(x) := \frac{1}{n} \int_{-\infty}^x \phi\left(\frac{z}{n} - \log n + \gamma - 1\right) dz$ and $F(x) := u_n(x)$, so that $f(t) = v^n(t)$ and $g(t) = \exp\left(-\frac{\pi}{2}n|t| + int \log |t| + int(\gamma - 1)\right)$. We then take $T = 1/(\sqrt{n} \log n)$. Since $|\phi| \leq 1$ on R , the first term in the Berry-Esseen bound is no more than $1/2nT = \log n/2\sqrt{n}$. For the second term, when $|t| \leq T$ we have $|(f(t) - g(t))/t| \ll n|t|e^{-\pi n|t|/2}$. Thus $\int_{-T}^T |(f(t) - g(t))/t| dt \ll n \int_0^\infty t \exp(-\pi nt/2) dt = 4/\pi^2 n$. Both claims now follow. From the first claim our (much) earlier assertion that $\text{prob}[\sum_1^n 1/X_k \leq n(\log n + 1 - \gamma + Z)] \rightarrow \int_{-\infty}^Z \phi(x) dx$ is immediate, and with an error bound of $O(\log n/\sqrt{n})$ which is uniform in Z .

In view of our estimate for the sum of the small digits, it now follows that

$$\text{prob}_{G_1} \left[\sum_{k=1}^r a_k \leq z \right] = O(r^{-\epsilon} \log^2 r) + \int_{-\infty}^{z-h(r)} R^{-1} u_n(x/R) dx$$

where $h(r) := (r/\log 2)((1/2 + \epsilon) \log r - 1) + O(r^{3/4+2\epsilon})$, and where $n = \lceil r^{1/2-\epsilon}/\log 2 \rceil$. Routine calculus with $\rho := r/\log 2$, $\epsilon := 1/12$, and taking into account the fact that $\phi(x) \leq 1$ on R , reduces this to

$$\text{prob}_{G_1} \left[\sum_{k=1}^r a_k \leq z \right] = \int_{-\infty}^{z-\rho(\log \rho + \gamma)} \rho^{-1} \phi(s/\rho) d\rho + O(E)$$

$$= \int_{-\infty}^{z/\rho - \log \rho + \gamma} \phi(s) ds + O(E)$$

where $E = r^{-1/12} \log^2 r$. This, apart from the characterisation of ϕ , is our main theorem.

Remark on Lévy's conjecture about this sum: For arbitrary positive λ , and $Y \in R$, $\text{prob}[\sum_{k=1}^n X_k < Y] < \int_{-\infty}^Y u^{*n}(z) \exp(\lambda(Y - z)) dz$. The Fourier transform of $u(z) \exp(-\lambda z)$ is

$$1 + (\lambda + it)(\log(\lambda + it) + \gamma - 1) - \sum_{k=2}^{\infty} \frac{(-1)^k (\lambda + it)^k}{(k-1)k!}.$$

Setting $t = 0$ and using the inequality $(1 - a)^n < \exp(-na)$ for $0 < a < 1$ we have

$$\text{prob} \left[\sum_{k=1}^n X_k < Y \right] < \exp(\lambda Y + n\lambda \log \lambda + n(\gamma - 1)).$$

Setting $Y = n \log n + n\theta - n \log \log n$ and $\lambda = n^{-1} e^{-\gamma - \theta} \log n$ gives $\text{prob} < \exp(-e^{-\gamma - \theta} \log n)$. The sum over n of this upper bound is finite if and only if $\theta < -\gamma$, or equivalently, $c < e^{-\gamma}$. The bounds obtained with this technique are usually quite good. The conjecture should thus be that the sum of the reciprocals of n $[0, 1]$ uniform random variables is infinitely often less than $n \log n + \log c - \log \log n$ if and only if $c < e^{-\gamma}$.

5. Properties of $\phi(x)$. Recall that ϕ is defined as the inverse Fourier transform of $\exp(-\frac{\pi}{2}|t| + it \log |t|)$. We now put $\phi_\theta(x) := \theta^{-1} \phi(x/\theta - \log \theta)$ for $\theta > 0$. Then $\widehat{\phi_\theta}(t) = (\widehat{\phi}(t))^\theta$. Thus $(\phi * \phi_\theta)(t) = \widehat{\phi}(t) \widehat{\phi_\theta}(t) = (\widehat{\phi}(t))^{1+\theta}$, so that $\phi * \phi_\theta = \phi_{1+\theta}$. In particular, and in contrast to the way the best-known stable function $\exp(-x^2)$ behaves, $(\phi * \phi)(x) = \frac{1}{2} \phi(x/2 - \log 2)$.

The factor $\exp(-(\pi/2)|t|)$ in the definition of ϕ ensures that ϕ extends to a function analytic at least in a strip $|\Im(z)| < \pi/2$. Thus

$$\phi^{(n)}(z) = \frac{1}{2\pi} i^n \int_{-\infty}^{\infty} s^n \exp\left(-\frac{\pi}{2}|s| + is \log |s| + isx\right) ds$$

$$\text{so that } |\phi^{(n)}(x)| \leq n! 2^n \pi^{-n-2}.$$

Because ϕ is the limit of shifted and scaled versions of u_n 's, it is non-negative on R . But since it is analytic on a strip including R , it is positive a.e. on R . Finally, since $\phi * \phi(x) = \phi(x/2 - \log 2)$, ϕ is strictly positive on all of R .

Remark. Not all stable probability density functions are positive on the entire real line. Those with $\alpha < 1$ and $|\beta| = 1$ have support an interval of the form $[c, \infty)$ or $(-\infty, c]$. From this one might say that ϕ barely escapes this fate. The rapid decay of ϕ to zero on the left fits in with this observation.

The asymptotic analysis of ϕ begins with the observation that for x real,

$$\phi(x) = \frac{1}{\pi} \Re \left[\int_0^\infty e^{-\pi s/2} \exp(i(s \log s + xs)) ds \right].$$

We take an arbitrarily large M and deform the path of integration so that it runs in straight segments from 0 to Mi to $M(1+i)$ to M , and then to ∞ along R . The contributions from all but the first leg of this path tend to zero as $M \rightarrow \infty$, so (with $y = is$),

$$\phi(x) = \frac{1}{\pi} \int_0^\infty \sin(\pi y) e^{-y \log y - xy} dy.$$

From this formula it is clear that ϕ actually extends to an entire function.

Our immediate interest, though, is in the behavior as $x \rightarrow +\infty$. We have

$$\begin{aligned} \phi(x) &= \frac{1}{\pi} \int_0^{1/\sqrt{x}} (\pi y + O(y^3)) e^{-y \log y - xy} dy + O \left(\int_{1/\sqrt{x}}^\infty e^{-xy + (1/2)y \log x} dy \right) \\ &= \int_0^{1/\sqrt{x}} y e^{-xy} (1 - y \log y + O(y^2 \log^2 x)) dy + O(x^{-4}) \\ &= x^{-2} + O(x^{-3} \log x). \end{aligned}$$

Remark. The analysis may be carried to more terms, most conveniently with the aid of a computer algebra system. Mathematica gives $x^{-2} - 2x^{-3} \log x + (3 - 2\gamma)x^{-3} + O(x^{-4} \log^2 x)$ and higher-order expansions. The series does not appear to be convergent and x must be rather large before even the third term improves matters.

For $x \rightarrow -\infty$, we take $\alpha := e^{|x|^{-1}}$ and deform the path $[0, \infty)$ to run instead from 0 to $-\alpha i$ and thence to $-\alpha i + M$ to M and on to ∞ . Now if $s = \sigma - it$ with $0 \leq t \leq \alpha$ and $\sigma > 0$, then the log of the integrand is

$$\begin{aligned} &-\frac{1}{2}\pi\sigma + \frac{1}{2}\sigma \log(\sigma^2 + t^2) + tx + \sigma \tan^{-1}(t/\sigma) \\ &- it \tan^{-1}(t/\sigma) + \frac{1}{2}i\sigma \log(\sigma^2 + t^2) + ix\sigma + \frac{1}{2}i\pi t. \end{aligned}$$

The real part of this tends to $-\infty$ uniformly in $0 \leq t \leq \alpha$ which justifies omitting the return to the real axis, and instead proceeding to $-i\alpha + \infty$ after the initial down-leg. The contribution of that down-leg, though, to the real part of the integral, is zero. Now the series expansion of $-\pi s/2 + is \log s + isx$ about $-i\alpha$, is

$$-\alpha + \sum_{n=2}^{\infty} \frac{(-i)^n}{n(n-1)} e^{-\alpha(n-1)} (s + i\alpha)^n.$$

Thus for $x < 0$,

$$\begin{aligned}\phi(x) &= \frac{1}{\pi} e^{-\alpha} \Re \left[\int_{\sigma=0}^{\infty} \exp \left(\alpha \sum_{n=2}^{\infty} \frac{(-i)^n \sigma^n}{n(n-1)} \right) d\sigma \right] \\ &= \frac{\alpha}{\pi} e^{-\alpha} \Re \left[\int_{\sigma=0}^{\infty} \exp \left(-\alpha \int_0^{\sigma} \tan^{-1} \nu d\nu + \frac{i\alpha}{2} \int_0^{\sigma} \log(1 + \nu^2) d\nu \right) d\sigma \right].\end{aligned}$$

The contribution to the double integral from $\sigma \geq 1$ is $\ll \exp(-\alpha \int_0^1 \tan^{-1} \nu d\nu) \ll e^{-\alpha\pi/8}$. For the interval $\alpha^{-2/5} \leq \sigma \leq 1$, the series expansion above shows that the contribution is $O(\exp(-\frac{1}{3}\alpha^{1/5}))$. For $0 < \sigma \leq \alpha^{-2/5}$, though, the series for the integrand reduces to $(1 + O(\alpha^{-1/5})) \exp(-\frac{1}{2}\alpha\sigma^2)$. Thus $\phi(x) = \sqrt{\frac{\alpha}{2\pi}} e^{-\alpha} (1 + O(\alpha^{-1/5}))$. This proves our second decay-rate claim for $\phi(x)$.

Directly from the integrals that define ϕ it is a relatively simple matter to compute and plot ϕ and its first two derivatives. The resulting plots are shown below, and they fit in nicely with Wolfe's theorem. (Our ϕ is one of the functions covered by his special case of Yamazato's theorem to the effect that all distributions in class L are unimodal.) From the graphs, it looks as though more might be true. Does the n^{th} derivative of ϕ have exactly n zeros?

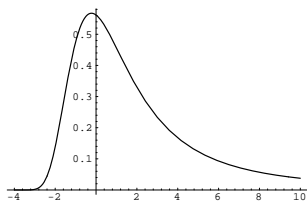


Fig. 4.

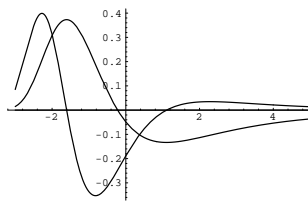


Fig. 5.

References

- [1] S. Dharmadhikara and K. Joag-dev, *Unimodality, convexity, and applications*, Academic Press, San Diego, CA, 1988, [Chapter 5, Theorems 5.5 and 5.6].
- [2] H. Diamond and J. Vaaler, *Estimates for partial sums of continued fraction partial quotients*, Pacific J. of Math., **122**(1) (1986), 73-82.
- [3] P.D.T.A. Elliott, *Probabilistic number theory I: Mean-value theorems*, Grundlehren der Mathematischen Wissenschaften, **239**, Springer-Verlag, New York-Berlin, 1979.
- [4] B.V. Gnedenko and A.N. Kolmogorov, *Limit distributions for sums of independent random variables*, Addison-Wesley, Cambridge Mass., 1954, Chapter 7 and Appendix 2.
- [5] Y. Guivarc'h and Y. Le Jan, *Asymptotic winding of the geodesic flow on modular surfaces and continued fractions*, Ann. Scient. Éc. Norm. Sup., **26** (1993), 23-50.

- [6] S. Hartman, *Quelques propriétés ergodiques des fractions continues*, *Studia Math.*, **12** (1951), 271-278.
- [7] D. Hensley, *The number of steps in the Euclidean algorithm*, *J. of Number Theory*, **49** (1994), 142-182.
- [8] ———, *Continued fraction Cantor sets, Hausdorff dimension, and functional analysis*, *J. of Number Theory*, **40** (1992), 336-358.
- [9] A.Ya. Khinchin, *Continued Fractions*, U. of Chicago Press, Chicago, 1964.
- [10] A.I. Lapin, *On some properties of stable laws*, Dissertation, 1947.
- [11] P. Lévy, *Fractions continues aléatoires*, *Rend. Ann.*, Palermo, 170-208.
- [12] D. Mayer, *On the thermodynamic formalism for the Gauss map*, *Commun. Math. Phys.*, **130** (1990), 311-333.
- [13] A. Wintner, *Asymptotic Distribution and Infinite Convolution*, Edwards Brothers, Ann Arbor, MI, (1938), 30-32.
- [14] S.J. Wolfe, *On the unimodality of L functions*, *Annals of Math. Statistics*, **42** (1971), 912-918.
- [15] M. Yamazato, *Unimodality of infinitely divisible distribution functions of class L* , *Ann. Probability*, **6** (1978), 523-531.

Received January 16, 1996 and revised June 3, 1996

TEXAS A&M UNIVERSITY
COLLEGE STATION, TX 77843-3368
E-mail address: doug.hensley@math.tamu.edu

This paper is available via <http://nyjm.albany.edu:8000/PacJ/2000/192-1-5.html>.