# POWER LAWS FOR MONKEYS TYPINGS RANDOMLY: THE CASE OF UNEQUAL PROBABILITIES

BRIAN CONRAD AND MICHAEL MITZENMACHER

Abstract. An early result in the history of power laws, due to Miller, concerned the following experiment. A monkey types randomly on a keyboard with $N$ letters ($N > 1$) and a space bar, where a space separates words. A space is hit with probability $p$; all other letters are hit with equal probability $(1-p)/N$. Miller proved that in this experiment, the rank-frequency distribution of words follows a power law.

The case where letters are hit with unequal probability has been the subject of recent confusion, with some suggesting that in this case the rank-frequency distribution follows a lognormal distribution. We prove that the rank-frequency distribution follows a power law for assignments of probabilities that have rational log-ratios for any pair of keys, and we present an argument of H. Montgomery that settles the remaining cases, also yielding a power law. The key to both arguments is the use of complex analysis.

The method of proof produces simple explicit formulas for the coefficient in the power law in cases with rational log-ratios for the assigned probabilities of keys. Our formula in these cases suggests an exact asymptotic formula in the cases with an irrational log-ratio, and this formula is exactly what was proved by Montgomery.

## 1. Introduction

One of the earliest developments in the theory of power laws was the demonstration that the rank-frequency distribution of natural languages, which empirically follows a power law, could be explained by an entropy-optimization formulation developed by Mandelbrot [7]. (We provide this argument for completeness below.) This optimization framework underlies more recent work by other authors who show how it can be used to explain other power law behaviors, such as the degree distribution of the Internet graph [2, 3].

Soon after Mandelbrot's argument appeared, however, another argument by the psychologist Miller demonstrated that the power law behavior of rank-frequency distribution could be explained without the underlying optimization argument [8]. Miller describes the following experiment. A monkey types randomly on a keyboard with $N$ letters and a space bar. A space is hit with probability $p$ with $0 < p < 1$; all other letters are hit with equal probability $(1-p)/N$. A space is used to separate words. Miller demonstrates that in this experiment, the rank-frequency distribution follow a power law. (Again, we present more detail below.) Miller's result serves as warning: just because one finds a compelling mechanism to explain a power law does not mean that there are not other, perhaps simpler, explanations.

Miller only gave the proof for the case where all letters other than the space are equally likely to be hit. Interestingly, the case where letters are struck with unequal probability has recently become a point of confusion. Perline recently argued that if the letter frequencies are not equal, a lognormal distribution occurs. (This claim is repeated in [5].) Troll and bein Graben correctly argue that Perline's result simply shows that the distributions of the words of length up to $n$, for each fixed $n$, are approximately lognormal [15]. They argue that, in general, the true distribution (without truncating words up to some fixed length) is a power law, although they only give an argument for the case of two letters.

In this paper, we begin by reviewing the fascinating history of this fundamental problem. Then we use methods from complex analysis to prove that Miller's random monkey experiment yields power laws for

*Date*: October 28, 2003.

Brian Conrad was partially supported by the NSF and the Alfred P. Sloan Foundation, and would like to thank K. Conrad and C. Doering for helpful discussions. Michael Mitzenmacher was supported in part by an Alfred P. Sloan Research Fellowship and NSF grant CCR-9983832, and would like to thank N. Elkies for helpful discussions. Both authors thank W. Szpankowski for useful advice.

rank-frequency distribution with probability assignments to keys satisfying a rationality assumption on log-ratios of pairs of probabilities. We use analytic methods to establish a simple explicit power law in cases with rational log-ratios for pairs of probabilities; more specifically, we use generalized Dirichlet series and an elementary identity established by means of Fourier series. Passing to a limit on these formulas predicts an analogous result in the remaining "irrational" cases, and this prediction agrees with an unpublished theorem proved contemporaneously by H. Montgomery; in §6 we provide Montgomery's argument, which uses methods that are standard in analytic number theory.

The use of analytic techniques to study problems of this type is not in itself novel (see [4], [6], and [14], for example), and it is also a well-known phenomenon that rationality issues can lead to cases that behave in a manner somewhat different from generic cases. The novelty of this paper is therefore not in the consideration of analytic techniques but rather in the application of these techniques to an interesting non-trivial problem that has not before been studied in detail from the analytic point of view.

NOTATION AND TERMINOLOGY. Throughout this paper, the phrase *log-ratio* for a pair of positive real numbers refers to the ratio of their logarithms (to a common base, the choice of which cancels out), not the logarithm of ratios (to some base, the choice of which does make a difference). All logarithms without an indicated base are understood to be taken to the base $e = 2.71828\ldots$. We write $\lfloor \cdot \rfloor$ to denote the greatest-integer function, and $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{C}$ to denote the ring of integers and the fields of rational numbers, real numbers, and complex numbers respectively.

Finally, to permit the use of the letter $i$ as an indexing variable, we choose to write $\sqrt{-1}$ rather than the customary $i$ to denote a fixed choice of solution to $z^2 + 1 = 0$ in $\mathbf{C}$. This choice determines our sense of direction for path integrals in the complex plane.

## 2. Review of Definitions and History

Our treatment here is based on a recent survey by Mitzenmacher [9], to which we refer the reader for more information.[1] In what follows, we let $f_j$ be the (asymptotic) fraction of the time the $j$th most frequently used word appears. In many of our models several words can have the same probability of occurrence, so there may be ties. In our context, we will say that $f_j$ follows a *power law in $j$* if there exist positive constants $c_1, c_2, \alpha$ such that $c_1 j^{-\alpha} \leq f_j \leq c_2 j^{-\alpha}$ for sufficiently large $j$.

We sketch Mandelbrot's argument that leads a power law in the rank-frequency distribution of words [7]. Consider some language consisting of $W$ words. The cost of transmitting the $j$th most frequent word of the language is denoted $C_j$. For example, if we think of English text, the cost of a word might be thought of as the number of letters plus the additional cost of a space. We therefore naturally expect the most frequent words to have the smallest number of letters. Let us take the cost of a space to be 0. Then if the alphabet size is $N > 1$, there are $N^k$ possible words of length $k$ (including $k = 0$; we allow the empty word for convenience). In particular, the words with $k$ letter have frequency ranks from $1 + (N^k - 1)/(N - 1)$ to $(N^{k+1} - 1)/(N - 1)$. It follows that $\log_N j \leq C_j \leq \log_N j + 1$. Suppose that we wish to design the language to optimize the average amount of information per unit transmission cost. Here, we take the average amount of information to be the entropy. We think of each word in our transmission as being selected randomly, and the probability that a word in the transmission is the $j$th word of the language is $f_j$. Then the average information per word is the entropy $H = -\sum_{j=1}^{W} f_j \log_2 f_j$, and the average cost per word is $C = \sum_{j=1}^{W} f_j C_j$.

If we were designing the language, how would we choose the $f_j$ in order to minimize $A = C/H$? Taking derivatives, we find

$$\frac{\partial A}{\partial f_j} = \frac{C_j H + C \log_2(e f_j)}{H^2}.$$

Hence all the derivatives are 0 (and $A$ is in fact minimized) when $f_j = e^{-1} \cdot 2^{-HC_j/C}$. Since $\log_N j \leq C_j \leq \log_N j + 1$, we obtain a power law for the $f_j$; specifically,

$$(2^{-H/C} e^{-1}) j^{-H(\log_N 2)/C} \leq f_j \leq e^{-1} \cdot j^{-H(\log_N 2)/C}.$$

[1] For instance, this survey describes another argument that leads to a power law of word frequency based on preferential attachment, originally due to Simon [13]. We do not present this argument here.

Mandelbrot argues that a variation of this model matches empirical results for English quite well.

We now consider Miller's experiment [8]. Again, in his setup, a monkey types randomly on a keyboard with $N$ letters and a space bar. We again assume $N > 1$, and a space is used to separate words. A space is hit with probability $p$ (with $0 < p < 1$); all other letters are hit with equal probability $(1-p)/N$. As the monkey types, each word with $k$ (non-space) letters occurs with probability

$$p_k = \left(\frac{1-p}{N}\right)^k \cdot p,$$

and there are $N^k$ words of length $k$. (Again we allow the empty word of length 0 for convenience.) The words of longer length are less likely and hence occur lower in the rank order of word frequency. Thus, again the words with $k$ letters have frequency ranks from $1 + (N^k - 1)/(N-1)$ to $(N^{k+1} - 1)/(N-1)$. Hence, the word with rank-frequency $j$ occurs with probability $f_j$, where

$$\left(\frac{1-p}{N}\right)^{\log_N j+1} \cdot p \leq f_j \leq \left(\frac{1-p}{N}\right)^{\log_N j} \cdot p.$$

Rewriting yields

$$\left(p \cdot \frac{1-p}{N}\right) j^{\log_N(1-p)-1} \leq f_j \leq p \cdot j^{\log_N(1-p)-1},$$

and the power law behavior is apparent. (Note that this argument (and the conclusion) fails if $N = 1$ since the relevant finite geometric series behave differently.)

The above analysis of Miller's argument clearly makes use of the simplification that all letters are struck with equal probability. As previously mentioned, the case of unequal letter probabilities has met with some confusion; see, e.g., [9, 5, 10, 15]. The following example from [9] clarifies the power law behavior and lays the groundwork for our more general argument.

Consider an alphabet with two letters: "a" occurs with probability $q$, "b" occurs with probability $q^2$, and a space occurs with probability $1 - q - q^2$. The value $q$ must be chosen so that $1 - q - q^2 > 0$ (i.e., $0 < q < 1/\Phi$, where $\Phi = (1+\sqrt{5})/2$). In this case, every valid word the monkey can type occurs with probability $q^k(1-q-q^2)$ for some integer $k$. Let us say a word has pseudo-rank $k$ if it occurs with probability $q^k(1-q-q^2)$. There is 1 word with pseudo-rank 0 (the empty word), 1 with pseudo-rank 1 ("a"), 2 with pseudo-rank 2 ("aa" and "b"), and so on. A simple induction yields that the number of words with pseudo-rank $k$ is in fact the $(k+1)$th Fibonacci number $F_{k+1}$ (where $F_0 = 0$ and $F_1 = 1$). This follows easily from the fact that to obtain the words with pseudo-rank $k$ we append an "a" to a word with pseudo-rank $k-1$, or a "b" to a word with pseudo-rank $k-2$.

Recall that $F_k = \Phi^k/\sqrt{5} + o(1)$ for large $k$, where $\Phi = (1+\sqrt{5})/2$ . Also $\sum_{i=1}^k F_i = F_{k+2} - 1$. Now the argument is entirely similar to the case where all items have the same probability. When $F_{k+2} - 1 < j \leq F_{k+3} - 1$, the $j$th most frequent word has pseudo-rank $k$. For $k$ sufficiently large, we therefore have

$$k + 2 \leq \log_\Phi(\sqrt{5}(j+1)) < k + 3.$$

The frequency $f_j$ therefore satisfies

$$q^{\log_\Phi(\sqrt{5}(j+1))-2}(1-q-q^2) < f_j \leq q^{\log_\Phi(\sqrt{5}(j+1))-3}(1-q-q^2) \tag{2.1}$$

for large $j$. Again we have power law behavior.

Note that the key here was that the number of words with pseudo-rank at most $k$ grew roughly exponentially in $k$, where the base $\Phi$ was the reciprocal of the unique positive solution to the polynomial equation $x^2 + x = 1$. This is the statement we intend to generalize to more general sets of probabilities in what follows.

## 3. A General Problem

In general, we allow $N$ letters with arbitrary probabilities (adding up to something less than 1, the rest being the probability of a space), and we assume $N > 1$. If there are $n$ distinct probability values, it is convenient to label these in strictly increasing order, $0 < p_1 < p_2 < \cdots$, and write $p_i = p_1^{a_i}$. Note that in this case we have $a_1 = 1$ and $a_i < 1$ for $i > 1$. We let $w_i$ be the number of letters that are struck with the probability $p_i$. The probability of a space is therefore $1 - \sum w_i p_i$, which we require to be positive. Asking about the probability $f_j$ of the $j$th most frequent word (as a function of large $j$) can then be turned into the question of how many words have probability of occurrence greater than or equal to $p_1^\nu(1 - \sum w_i p_i)$ (as a function of large real $\nu$), as we shall see later.

More generally we have the following problem. Let $a_1 > \cdots > a_n$ be distinct positive real numbers and let $w_1, \ldots, w_n$ be positive integers. Consider the multi-set $\mathscr{A}$ that contains $a_i$ with multiplicity $w_i$ for all $i$. Say $\mathscr{A}$ has size $N = \sum w_i$, with elements enumerated as $\sigma_1, \ldots, \sigma_N$. We require $\sum w_i > 1$. Note that the case $n = 1$ (with $w_1 > 1$ counting the number of letters) corresponds to the situation considered by Miller (equal probabilities).

For each real $\nu \geq 0$, let $c_\nu$ be the number of distinct $N$-tuples of non-negative integers $(m_1, \ldots, m_N)$ such that $\nu = \sum m_i \sigma_i$. Concretely, $c_\nu$ counts how many ways $\nu$ can be expressed as a sum of elements of $\mathscr{A}$ (keep in mind that $\mathscr{A}$ is a multi-set, and our description of $c_\nu$ is consistent with the condition $c_0 = 1$). In the motivating situation of the variant on Miller's problem with unequal probabilities, we can consider the problem of counting the number of words (including the empty word) whose probability of occurrence is exactly $p_1^\nu(1 - \sum w_i p_i)$. The quantity $c_\nu$ is exactly the answer to this problem.

Algebraically, we have a formal expansion (for $x > 0$)

$$\frac{1}{1 - \sum_{i=1}^n w_i x^{a_i}} = \sum c_\nu x^\nu, \tag{3.1}$$

where $c_0 = 1$ and $c_\nu = 0$ for all $\nu$ outside of a discrete set of non-negative real numbers. Of course, when the $a_i$'s are not all integers then the right side of (3.1) is not an ordinary power series expansion around 0, so the series usually does not make sense when $x < 0$. We wish to give asymptotic bounds, in the spirit of (2.1), on $\sum_{\nu<t} c_\nu$ (respectively $\sum_{\nu\leq t} c_\nu$) as $t \to \infty$. In the application to word probabilities, these sums count the number of words whose probability of occurrence (in the sense of the discussion above) is greater than $p_1^t(1 - \sum w_i p_i)$ (respectively greater than or equal to $p_1^t(1 - \sum w_i p_i)$).

Consider the function $f_{\underline{a}}(x) = \sum_{i=1}^n w_i x^{a_i}$ on $[0, \infty)$. It would be more accurate to write $f_{\underline{a},\underline{w}}(x)$, but whereas it will be convenient to sometimes consider behavior when the $a_i$'s are varying, we shall never change the $w_i$'s. The function $f_{\underline{a}}$ is a strictly increasing continuous function with $f_{\underline{a}}(0) = 0$ and $f_{\underline{a}}(1) = \sum w_i > 1$, so there is a unique $x_0 \in (0, 1)$ with $f_{\underline{a}}(x_0) = 1$, and this is the unique solution to $f_{\underline{a}}(x) = 1$ on $[0, \infty)$. The example at the end of §2 corresponds to probabilities $p_1 = q^2$ and $p_2 = q = p_1^{1/2}$, so $a_1 = 1$ and $a_2 = 1/2$. Thus, $f_{\underline{a}}(x) = x + \sqrt{x}$ and $x_0 = \phi^2$ in this case, where $\phi = (-1 + \sqrt{5})/2$.

In general, we wish to study the behavior of $(\sum_{\nu<t} c_\nu)/(1/x_0)^t$ and $(\sum_{\nu\leq t} c_\nu)/(1/x_0)^t$ as $t \to \infty$. We will show that for explicit constants $0 < A \leq A'$ (depending on the $a_i$'s and $w_i$'s),

$$\liminf_{t\to\infty} \frac{\sum_{\nu<t} c_\nu}{(1/x_0)^t} = \liminf_{t\to\infty} \frac{\sum_{\nu\leq t} c_\nu}{(1/x_0)^t} = A, \ \limsup_{t\to\infty} \frac{\sum_{\nu<t} c_\nu}{(1/x_0)^t} = \limsup_{t\to\infty} \frac{\sum_{\nu\leq t} c_\nu}{(1/x_0)^t} = A', \tag{3.2}$$

where $A = A'$ if some $a_i/a_{i'}$ is irrational (the generic case) and otherwise $A = x_0^{1/r} A'$ with $r = D/a_1$ for the least common multiple $D$ of the denominators of the ratios $a_i/a_1 \in (0, 1]$ when these ratios all lie in $\mathbf{Q}$; we will in fact establish an exact asymptotic formula for $\sum_{\nu\leq t} c_\nu$ that is more precise than (3.2) when all ratios $a_i/a_{i'}$ are rational (and so $A < A'$), but this precise statement in a bit involved. Granting the asymptotics in (3.2), let us see how we obtain a power law for rank-frequencies of words by using Miller's argument.

Pick constants $0 < L' \leq L$ such that

$$L \cdot (1/x_0)^t \leq \sum_{\nu<t} c_\nu \leq \sum_{\nu\leq t} c_\nu \leq L' \cdot (1/x_0)^t \tag{3.3}$$

for sufficiently large $t$. The larger we take $t$, the nearer we can make $L$ and $L'$ to $A$ and $A'$ respectively. For large $j$, we wish to estimate the probability of occurrence $f_j = p_1^{t(j)}(1 - \sum_{i=1}^n w_i p_i)$ assigned to the word with rank-frequency $j$ (this identity defines the $t(j)$'s, so $0 = t(1) \leq t(2) \leq \ldots$); note that $t(j)$ is typically irrational in practice. As $j \to \infty$ we have $f_j \to 0$, so $t(j) \to \infty$. Thus, we may suppose $j$ is big enough so that (3.3) holds for $t = t(j)$. Recalling that $c_\nu$ counts the number of words whose probability of occurrence is exactly $p_1^\nu(1 - \sum w_i p_i)$, we get

$$\sum_{\nu < t(j)} c_\nu < j \leq \sum_{\nu \leq t(j)} c_\nu,$$

so $L \cdot (1/x_0)^{t(j)} < j \leq L' \cdot (1/x_0)^{t(j)}$, and finally

$$\frac{\log j - \log L'}{\log(1/x_0)} \leq t(j) < \frac{\log j - \log L}{\log(1/x_0)}.$$

Hence, for large $j$ we have

$$p_1^{(\log j - \log L)/\log(1/x_0)}(1 - \sum w_i p_i) \leq f_j \leq p_1^{(\log j - \log L')/\log(1/x_0)}(1 - \sum w_i p_i), \tag{3.4}$$

so the power law behavior of $f_j$ with respect to $j$ is obtained: we can rewrite (3.4) as

$$C(L) \cdot j^{-\log_{1/x_0}(1/p_1)} \leq f_j \leq C(L') \cdot j^{-\log_{1/x_0}(1/p_1)} \tag{3.5}$$

for large $j$, with constant $C(u) \stackrel{\text{def}}{=} (1/p_1)^{\log_{1/x_0}(u)}(1 - \sum w_i p_i)$.

By taking $j$ larger, we can make $C(L)$ and $C(L')$ as close as we please to $C(A)$ and $C(A')$ respectively. Thus, as an important consequence, if at least one ratio $a_i/a_{i'}$ is irrational then since $A' = A$ we get $f_j \sim C(A)j^{-\alpha}$ where $\alpha = \log_{1/x_0}(1/p_1)$. Even though (as we noted above) we will give an exact asymptotic formula for $\sum_{\nu \leq t} c_\nu$ in the cases with rational ratios, this formula involves the intervention of greatest-integers and the lack of control over the distribution of the fractional parts of the $t(j)$'s as $j \to \infty$ provides the obstruction to the existence of a power-like exact asymptotic formula for $f_j$ in the rational-ratio cases. In our mathematical analysis, it will be much simpler to focus on a study of the sums $\sum_{\nu < t} c_\nu$ and $\sum_{\nu \leq t} c_\nu$ as functions of $t$ rather than the word probabilities $f_j$ as a function of the rank-frequency parameter $j$, and it is for this reason that we have explained above how to extract the power law (3.5) from asymptotics on $\sum_{\nu < t} c_\nu$ and $\sum_{\nu \leq t} c_\nu$.

A useful observation is that the liminf and limsup of $(\sum_{\nu < t} c_\nu)/(1/x_0)^t$ and $(\sum_{\nu \leq t} c_\nu)/(1/x_0)^t$ (as functions of $t \to \infty$) are invariant under positive scaling of the $a_i$'s. Indeed, if $a_i' = r a_i$ for all $i$ and some $r > 0$ (with $w_i' = w_i$ for all $i$) then $f_{\underline{a}'}(x) = f_{\underline{a}}(x^r)$, so $x_0' = x_0^{1/r}$. We likewise have that $c_\nu' = c_{\nu/r}$, so

$$\frac{\sum_{\nu < t} c_\nu'}{(1/x_0')^t} = \frac{\sum_{\nu < t/r} c_\nu}{(1/x_0)^{t/r}}.$$

The scaling invariance follows, and the sums over $\nu \leq t$ go the same way. Thus, for an analysis of such liminf's and limsup's we may scale the $a_i$'s by any common positive scaling factor. While this may suggest we should exclusively consider sequences with $a_1 = 1$, it is useful to avoid such a restriction. For example, we will need to use asymptotic comparisons with $(1/x_0)^{\lfloor t \rfloor}$, and for this purpose the necessary identities are

$$\frac{\sum_{\nu < t} c_\nu'}{(1/x_0')^{\lfloor t \rfloor}} = \frac{\sum_{\nu < t/r} c_\nu}{(1/x_0)^{\lfloor r(t/r) \rfloor / r}}, \quad \frac{\sum_{\nu \leq t} c_\nu'}{(1/x_0')^{\lfloor t \rfloor}} = \frac{\sum_{\nu \leq t/r} c_\nu}{(1/x_0)^{\lfloor r(t/r) \rfloor / r}}; \tag{3.6}$$

the failure of $\lfloor r(t/r) \rfloor / r = \lfloor t \rfloor / r$ to equal $\lfloor t/r \rfloor$ implies that comparisons against $(1/x_0)^{\lfloor t \rfloor}$ obey a transformation law with respect to $a_i \mapsto r a_i$ that is more complicated than comparisons against $(1/x_0)^t$.

Here is our main result; see §6 for a discussion of the cases that violate the hypothesis of rational ratios.

**Theorem 3.1.** *Let $w_1, \ldots, w_n > 0$ be real numbers such that $\sum w_i > 1$. Let $f_{\underline{a}}(x) = \sum w_i x^{a_i}$ on $(0, \infty)$ with $a_1 > \cdots > a_n > 0$ such that $a_i/a_{i'} \in \mathbf{Q}$ for all $i$ and $i'$ (this condition is satisfied if $n = 1$), and let $D$ be the least common multiple of the denominators of the ratios $a_i/a_1 \in \mathbf{Q} \cap (0, 1]$. Let $x_0 \in (0, 1)$ be the unique positive zero of $f_{\underline{a}} - 1$, and let $r = D/a_1$ (so $r = 1$ if the $a_i$'s are integers and* $\gcd(a_1, \ldots, a_n) = 1$*). Define*

$$A' = \frac{1}{r(1 - x_0^{1/r}) x_0 f_{\underline{a}}'(x_0)} > 0. \tag{3.7}$$

*We have*

$$\lim_{t\to\infty} \frac{\sum_{\nu\le t} c_\nu}{(1/x_0)^{\lfloor rt\rfloor/r}} = A'. \tag{3.8}$$

*That is,* $\sum_{\nu\le t} c_\nu \sim A'\cdot(1/x_0)^{\lfloor rt\rfloor/r}$ *as* $t\to\infty$.

It is not evident *a priori* that the limit in (3.8) exists, and the corresponding limit using $\sum_{\nu<t}$ generally does not exist ($t$'s with $c_t\neq 0$ present obstructions in even the simplest case of geometric series; see the end of Example 3.4). Since $t-1/r<\lfloor rt\rfloor/r\le t$ for all $t$, and as $t$ varies over $[n,n+1)$ for $n\in\mathbf{Z}$ the difference $t-\lfloor rt\rfloor/r$ sweeps across the interval $[0,1/r)$, Theorem 3.1 and the *continuity* of the function $t\mapsto(1/x_0)^t$ immediately yield:

**Corollary 3.2.** *With notation and hypotheses as in Theorem* 3.1*,*

$$\liminf_{t\to\infty}\frac{\sum_{\nu<t}c_\nu}{(1/x_0)^t}=\liminf_{t\to\infty}\frac{\sum_{\nu\le t}c_\nu}{(1/x_0)^t}=x_0^{1/r}A',\ \ \limsup_{t\to\infty}\frac{\sum_{\nu<t}c_\nu}{(1/x_0)^t}=\limsup_{t\to\infty}\frac{\sum_{\nu\le t}c_\nu}{(1/x_0)^t}=A'. \tag{3.9}$$

The positive gap $A'-x_0^{1/r}A'$ between limsup and liminf in Corollary 3.2 is artificial, being entirely due to the use of division by the continuous function $(1/x_0)^t$ rather than by the step function $(1/x_0)^{\lfloor rt\rfloor/r}$. In §6 we will see how our limit formula (3.8) in the rational-ratios case predicts the correct asymptotic behavior in the remaining cases; see Theorem 6.1.

*Example* 3.3. When $n=1$ (Miller's case), we have $w=w_1>1$ and $r=1/a_1$. Thus, $f_{\underline{a}}(x)=wx^{a_1}$, so $x_0=1/w^r$. Hence, $A'=w/(w-1)$. Writing (3.1) as $1/(1-wx^{a_1})=\sum_{n\ge0}w^nx^{a_1n}$, it is trivial to directly establish (3.8) in this case.

*Example* 3.4. When $n=2$, $w_1=w_2=1$, $a_1=2$, and $a_2=1$, we have the Fibonacci-based example at the end of §2. Since $f_{\underline{a}}(x)=x^2+x$, clearly $x_0=\phi$ where $\phi=(-1+\sqrt5)/2$ satisfies $\phi^2+\phi=1$. Calculation yields $r=1/2$ and $A'=1/(1-\phi^2)(1-\phi/2)=1/(\phi+1/2)=2/\sqrt5$. Using (3.4), and recalling that $p_1=q^2$, we find that for sufficiently large $j$

$$q^{\log_\Phi j-\log_\Phi((3+\sqrt5)/\sqrt5)}(1-q-q^2)\le f_j\le q^{\log_\Phi j-\log_\Phi(2/\sqrt5)}(1-q-q^2)$$

where $\Phi=1/\phi=(1+\sqrt5)/2$.

This can be made to look more like (2.1) by rewriting it as

$$q^{\log_\Phi(\sqrt5 j)-\log_\Phi(3+\sqrt5)}(1-q-q^2)\le f_j\le q^{\log_\Phi(\sqrt5 j)-\log_\Phi 2}(1-q-q^2).$$

Compared to (2.1), this gives slightly weaker bounds in the bounding constant factors, although the power law exponent is the same. To see why it is not surprising that we obtain weaker coefficient-bounds, note that the direct analysis of rank-frequencies in this example corresponds to considering bounds as $t$ runs through the discrete set of values $t(j)$ such that the summation function $\sum_{\nu\le t}c_\nu$ jumps (i.e., the $t(j)$'s are the discretely-spread values $t$ such that $c_t\neq0$). Thus, whereas (3.3) with $L$ and $L'$ very close to $A$ and $A'$ respectively concerns optimizing bounding constants across a continuum of values of $t\to\infty$, in (2.1) we are only optimizing over a discrete set of values $t(j)$. This smaller sampling locus allows for the possibility of tighter bounding constants.

For example, in the situation analogous to that considered by Miller one sees this phenomenon: for $w>1$, consider $f(t)=\sum_{\nu\le t}w^\nu=(w^{\lfloor t\rfloor}-1)/(w-1)$, where the sum is taken over integral $\nu\in[0,t]$. Clearly $f(t)/w^t\sim1/(w-1)$ as $t$ runs through the discrete locus (of integer values) where $f$ jumps. Hence while $f(t)/w^t$ varies between $1/(w-1)$ and $1/(w(w-1))$ as $t$ grows through all positive real values, $f(w)/w^{\lfloor t\rfloor}\sim 1/(w-1)$ as $t$ grows through all positive real values. Note in particular that a gap that appears over a continuum may disappear when sampling over a discrete locus or when comparing with a well-chosen step function (such as $w^{\lfloor t\rfloor}$ replacing $w^t$).

In Example 3.4, we saw that by sampling only in a discrete locus, we can get estimates that may be much tighter than what holds over a continuum of $t$-values. At the opposite extreme, omitting a discrete set of sampling values has no impact on asymptotic power law bounding constants, as we record in the following easy lemma (that will provide a useful simplification in the subsequent analysis):

**Lemma 3.5.** *Let $0 < \tau_1 < \tau_2 < \dots$ be a discrete unbounded sequence. To prove Theorem 3.1, it suffices to prove*

$$\lim_{t\to\infty} \frac{\sum_{\nu<t} c_\nu}{(1/x_0)^{\lfloor rt\rfloor/r}} = A' \tag{3.10}$$

*as $t$ grows without bound through values distinct from the $\tau_j$'s, where $A'$ is as in (3.7). Moreover, it is enough to consider the case when the $a_i$'s are integers and $\gcd(a_1, \dots, a_n) = 1$.*

*Proof.* Since $\sum_{\nu<t} c_\nu$ is monotonically increasing and $\sum_{\nu\le t} c_\nu = \lim_{\varepsilon\to 0^+} \sum_{\nu<t+\varepsilon} c_\nu$, and the function $t \mapsto (1/x_0)^{\lfloor rt\rfloor/r}$ enjoys the same monotonicity and one-sided continuity properties, the first assertion follows. The sufficiency of considering the case of a primitive $n$-tuple of integers $a_i$ comes down to the easy verification that the limit formula (3.10) is compatible with the transformation formulas (3.6), the identity $x_0' = x_0^{1/r}$, and the identity

$$x_0' f_{\underline{a}'}'(x_0') = r x_0 f_{\underline{a}}'(x_0) \tag{3.11}$$

when $a_i' = r a_i$ for all $i$ (to prove (3.11), differentiate the identity $f_{\underline{a}'}(x) = f_{\underline{a}}(x^r)$). □

Thanks to this lemma, we shall now suppose (for the purpose of proving Theorem 3.1) that the $a_i$'s are integers without a non-trivial common factor. The advantage of this case is that $\sum w_i x^{a_i}$ is a polynomial and $r = D/a_1$ is equal to 1. The proof of Theorem 3.1 will require some techniques from complex analysis, and before giving the proof it will be convenient to make some preliminary remarks. This will also give us an opportunity to introduce some notation to be used in the proof. Making the change of variable $s = -\log x$ (that converts the positive real line into the whole real line) on (3.1), consider the meromorphic function

$$F(s) = \frac{1}{1 - \sum w_i e^{-a_i s}} = \sum_\nu c_\nu e^{-\nu s} \tag{3.12}$$

for $s \in \mathbf{C}$. The denominator is near 1 (hence non-vanishing) for $\mathrm{Re}(s) \gg 0$ since $|e^{-a_i s}| = e^{-a_i \mathrm{Re}(s)} \to 0$ as $\mathrm{Re}(s) \to \infty$. Thus, standard convergence arguments with truncated sums ensure that the right side of (3.12) is absolutely convergent and equal to the middle term for $\mathrm{Re}(s) \gg 0$. The poles of $F$ are concentrated in the vanishing locus of the denominator $1 - f_{\underline{a}}(e^{-s})$. We are going to use the behavior of $F$ as $\mathrm{Re}(s) \to \pm\infty$ to get our desired asymptotics (the reader may wish to compare our argument with the "closing-the-box" discussion in [14, pp.252ff]).

Note that the poles of $F$ (i.e., the zeros of the denominator $1 - f_{\underline{a}}(e^{-s})$) are concentrated in a vertical strip $V$ of bounded width. Indeed, as $\mathrm{Re}(s) \to -\infty$, we see that $|e^{-a_i s}| = e^{-a_i \mathrm{Re}(s)}$ has exponential growth, but the term for $i = 1$ dominates the rest since $a_1 > a_2 > \dots$, so $F(s) \cdot w_1 e^{-a_1 s} \sim 1$ (and hence $F(s)$ has exponential decay) as $\mathrm{Re}(s) \to -\infty$. We also saw above that the denominator term is near 1 as $\mathrm{Re}(s) \to \infty$. This leaves a closed vertical strip $V$ of bounded width that contains all poles of $F$. We will make this explicit in (4.1).

Due to the rationality hypothesis in Theorem 3.1, from which we brought ourselves to the case of integral $a_i$'s, $f_{\underline{a}}$ is a *polynomial.* Thus, the equation $f_{\underline{a}}(e^{-s}) = 1$ is a polynomial equation in $e^{-s}$ and so its solutions are exactly $s = -\log \rho_\alpha + 2\pi\sqrt{-1}\cdot n$ for $n \in \mathbf{Z}$, where $\rho_\alpha$ runs over the finitely many roots of $f_{\underline{a}}(z) - 1 \in \mathbf{C}[z]$. In particular, the set of solutions to $f_{\underline{a}}(e^{-s}) = 1$ is *discretely* spread out in $V$ (in the sense that there is a positive lower bound on distances between solutions), and $F$ is a vertically periodic function ($F(s + 2\pi\sqrt{-1}) = F(s)$) since $a_i \in \mathbf{Z}$ for all $i$. When at least one ratio among the $a_i$'s is irrational, the periodicity is lost and the geometry of the location of these denominator zeros (that is going to control our analysis) becomes more difficult to handle directly. The irrational case will be discussed in §6.

## 4. Preparations for the proof of Theorem 3.1

Since the $a_i$'s (and $w_i$'s) will not be changing, we write $f$ to denote $f_{\underline{a}}(x) = \sum w_i x^{a_i}$. Let $\{\rho_\alpha\}$ denote the set of roots of the polynomial $f-1$, with $\rho_{\alpha_0} = x_0$ the unique positive real root. Later on, when we need to extract a dominant term, it will be important to use:

**Lemma 4.1.** *The root $x_0$ is a* simple *root of $f-1$ and $|\rho_\alpha| > x_0$ for all $\alpha \neq \alpha_0$.*

This lemma ensures that when there exist roots $\rho \in \mathbf{C}$ to $f(z) = 1$ with $\rho \neq x_0$ (i.e., $n > 1$, or $n = 1$ and $a_1 > 1$), such $\rho$ lie strictly outside of the circle $|z| = x_0$. Though the simplicity of the root at $x_0$ is true without requiring $\gcd(a_1, \dots, a_n) = 1$, the property $|\rho_\alpha| > x_0$ for $\alpha \neq \alpha_0$ makes essential use of the condition $\gcd(a_1, \dots, a_n) = 1$ and is false without it.

*Proof.* Since $f-1$ has derivative $\sum a_i w_i x^{a_i - 1}$ that is positive on $(0, \infty)$, this derivative is nowhere vanishing on $(0, \infty)$. A double root of a polynomial must be a root of the derivative, so $x_0$ has to be a simple root.

For any $z \in \mathbf{C}$, the triangle inequality gives $|f(z)| \leq \sum w_i |z|^{a_i} = f(|z|)$, so if $|z| < x_0$ then $|f(z)| < f(x_0) = 1$. Hence, if $f(z) = 1$ then $|z| \geq x_0$. It remains to show that $f(z) = 1$ with $|z| = x_0$ forces $z = x_0$. If $|z| = x_0$ then $f(z) = 1$ forces the inequality

$$1 = |f(z)| \leq \sum w_i |z|^{a_i} = f(|z|) = f(x_0) = 1$$

to be an equality. For a set of nonzero complex numbers $\{\beta_i\}$, the triangle inequality $|\sum \beta_i| \leq \sum |\beta_i|$ is an equality if and only if there is no "angle cancellation", so the $w_i z^{a_i}$'s all lie on the same ray emanating from the origin. The sum of these $w_i z^{a_i}$'s must also lie on this ray, but this sum is $f(z) = 1$, so the ray must be $(0, \infty)$. Hence, if $f(z) = 1$ and $|z| = x_0$ then $w_i z^{a_i}$ is a positive real number for all $i$, so $z^{a_i}$ is a positive real number for all $i$. But $\gcd(a_1, \dots, a_n) = 1$ implies that $\sum a_i v_i = 1$ for some integers $v_1, \dots, v_n$, so $z = z^{\sum a_i v_i} = \prod (z^{a_i})^{v_i}$ is a positive real number. Since the equation $f(x) = 1$ has $x_0$ as its unique solution on $(0, \infty)$, we get $z = x_0$, as desired. $\square$

If $s \in \mathbf{C}$ satisfies $1 - f(e^{-s}) = 0$, then Lemma 4.1 yields $e^{-\operatorname{Re}(s)} = |e^{-s}| \geq x_0$ with equality if and only if $e^{-s} = x_0$. Defining $s_0 = -\log(x_0)$, we have $\operatorname{Re}(s) \leq s_0$ with equality if and only if $s \in s_0 + 2\pi\sqrt{-1} \cdot \mathbf{Z}$. If $\rho_0$ is a root of $f-1$ with the greatest distance from the origin (so $|\rho_0| > x_0$ if $\rho_0 \neq x_0$), then $\operatorname{Re}(s) = -\log|e^{-s}| \geq -\log|\rho_0|$ also holds. Thus, the vertical strip

$$V = \{s \in \mathbf{C} \mid -\log|\rho_0| \leq \operatorname{Re}(s) \leq s_0\} \tag{4.1}$$

enjoys the property that the denominator function $1 - f(e^{-s})$ in (3.12) is non-vanishing outside of $V$ and its zeros within $V$ consist of finitely many vertically periodic sequences $s_\alpha + 2\pi\sqrt{-1} \cdot \mathbf{Z}$, where $e^{-s_\alpha}$ ranges over the finitely many roots $\rho_\alpha$ of the polynomial $f-1$ (recall $f(0) \neq 1$, so $\rho_\alpha \neq 0$ for all $\alpha$). To remove ambiguity about the choice of $s_\alpha$'s satisfying $e^{-s_\alpha} = \rho_\alpha$, we require $-\pi \leq \operatorname{Im}(s_\alpha) < \pi$. This normalization is not significant.

In the next section, it will be convenient to make an exponential change of variable, as follows. Let $0 = \nu_1 < \nu_2 < \dots$ be the discrete set of integers $\nu$ such that $c_\nu \neq 0$. Define $n_k = e^{\nu_k}$, so $n_1 = 1$. Let $c'_k = c_{\nu_k}$. For $\operatorname{Re}(s) \gg 0$, (3.12) yields the absolutely convergent series expansion

$$F(s) = \sum_{k \geq 1} \frac{c'_k}{n_k^s}, \tag{4.2}$$

where $\{n_k\}$ is a strictly increasing unbounded sequence in $[1, \infty)$; this type of sum is called a *generalized Dirichlet series* (the usual Dirichlet series are those such that $n_k = k$). Note that $\nu_k < t$ if and only if $n_k < e^t$. Since $F$ is analytic on the open half-plane $\{\operatorname{Re}(s) > s_0\}$ to the right of $V$ and the coefficients $c'_k$ are non-negative real numbers, basic complex analysis ensures that the generalized Dirichlet series expansion for $F$ in (4.2) is absolutely convergent for $\operatorname{Re}(s) > s_0$. Using Lemma 3.5 and the identity $(1/x_0)^t = (e^t)^{s_0}$, we can reformulate Theorem 3.1 as the assertion

$$\lim_{t \to \infty} \frac{\sum_{n_k < y} c'_k}{y^{s_0} x_0^{\{t\}}} = A'$$

as $y = e^t \to \infty$ with $y$ avoiding the $n_k$'s (or avoiding a discrete locus in $(0, \infty)$ containing the $n_k$'s, such as $e^{\mathbf{Z}}$); here, $\{t\} = t - \lfloor t \rfloor$ denotes the fractional part of $t$, so $y^{s_0} x^{\{t\}} = (1/x_0)^{\lfloor t \rfloor}$.

In our later calculation of residues, we shall need two elementary formulas that we record here for convenience of reference. For a meromorphic function $g$ on $\mathbf{C}$ with a pole of order $\mu_b \geq 1$ at $b \in \mathbf{C}$, the residue $\mathrm{Res}(g; b)$ of $g$ at $b \in \mathbf{C}$ is given by

$$\mathrm{Res}(g; b) = \frac{1}{(\mu_b - 1)!} \left(\frac{\mathrm{d}}{\mathrm{d}z}\right)^{\mu_b - 1} ((z-b)^{\mu_b} g(z))|_{z=b}. \tag{4.3}$$

In particular, if $g$ has a simple pole at $b$ then $\mathrm{Res}(g; b) = \lim_{z \to b}(z - b)g(z)$. Thus, if $g$ has a simple pole at $b$ (or is analytic there) and $h$ is analytic near $b$ then this simple limit expression shows

$$\mathrm{Res}(gh; b) = h(b)\,\mathrm{Res}(g; b). \tag{4.4}$$

For poles of order greater than 1, there is no simple formula for $\mathrm{Res}(gh; b)$ in terms of $\mathrm{Res}(g; b)$ and $h(b)$ alone; one needs to use more information about the series expansions of $g$ and $h$ around $b$. Fortunately for us, Lemma 4.1 will ensure that the dominant term in the analysis involves simple poles and the unpleasant (4.3) will suffice for an estimate on the rest as an error term.

Here is the main example of interest to us: for a fixed $y > 0$, consider the function

$$H(s) = y^s \cdot \frac{F(s)}{s} = \frac{y^s}{s \cdot (1 - f(e^{-s}))} = \frac{y^s}{s(1 - \sum w_i e^{-a_i s})}.$$

This is meromorphic in $s$, and its only possible poles are where $s = 0$ and where $f(e^{-s}) = 1$. Since $f(1) = \sum w_i > 1$, we see that $H$ has a simple pole at $s = 0$ (with residue $y^0 F(0) = 1/(1 - \sum w_i)$ independent of $y$). Since $y^s$ is analytic in $s$ and nowhere vanishing, it follows that $H$ has poles at the points in $s_\alpha + 2\pi\sqrt{-1} \cdot \mathbf{Z}$ (0 is not in this latter locus, since $f(0) \neq 1$).

## 5. Proof of Theorem 3.1

For $\mathrm{Re}(s) > s_0$ and $y > 0$, (4.2) yields

$$\frac{y^s F(s)}{s} = \frac{1}{s} \cdot \sum_{k \geq 1} c'_k \left(\frac{y}{n_k}\right)^s. \tag{5.1}$$

Now comes the essential step where complex analysis provides non-trivial information: as in [14, §7.5.3] (which treats ordinary Dirichlet series), we will compute partial sums of the generalized Dirichlet series (5.1) in the form of line integrals. More precisely, for $\sigma > s_0$ we wish to study the line integral

$$I_\sigma(y) = \frac{1}{2\pi\sqrt{-1}} \int_{\sigma - \infty\sqrt{-1}}^{\sigma + \infty\sqrt{-1}} y^s F(s) \frac{\mathrm{d}s}{s} \stackrel{\mathrm{def}}{=} \lim_{R \to \infty} \frac{1}{2\pi\sqrt{-1}} \int_{\sigma - R\sqrt{-1}}^{\sigma + R\sqrt{-1}} y^s F(s) \frac{\mathrm{d}s}{s}, \tag{5.2}$$

an integral over a vertical segment of length $2R$ centered at $\sigma \in \mathbf{R}$ with endpoints $\sigma \pm R\sqrt{-1}$ going off to $\sigma \pm \infty\sqrt{-1}$. We will briefly address the convergence of $I_\sigma(y)$ after some preliminary remarks.

Such a line integral is just an ordinary calculus integral in disguise: we consider the integrand as a $\mathbf{C}$-valued function on a parameterized segment of points $\sigma + t\sqrt{-1}$ where $t \in [-R, R]$ (so $\mathrm{d}s = \mathrm{d}t$), and the real and imaginary parts are integrated separately. Because the integrand $y^s F(s)/s$ is analytic in the half-plane $\{\mathrm{Re}(s) > s_0\}$ and the vertically periodic $F(s)$ is bounded in $\{\mathrm{Re}(s) \geq s_0 + \varepsilon\}$ for any $\varepsilon > 0$, the integrand $y^s F(s)/s$ dies off like $1/|s|$ as we move horizontally up and down in $\{\sigma \leq \mathrm{Re}(s) \leq \sigma'\}$. Thus, the absence of poles for $y^s F(s)/s$ in $\{\mathrm{Re}(s) > s_0\}$ and elementary estimates on integrals of $y^s F(s)/s$ over short horizontal segments imply that the convergence and value of $I_\sigma(y)$ is independent of the choice of $\sigma > s_0$.

The key facts about $I_\sigma(y)$ are the following (which we'll justify shortly):

- $I_\sigma(y)$ is convergent, and equals $\sum_{n_k < y} c'_k$ as long as $y$ is not equal to any of the $n_k$'s;
- if we move the vertical line of integration $\sigma + \mathbf{R}\sqrt{-1}$ to the left of the vertical strip $V$ in (4.1), past all of the poles of the integrand $y^s F(s)/s$, we can compute $I_\sigma(y)$ in terms of *residues* of $y^s F(s)/s$ at all of its poles.

These facts allow us to compute the sum of interest, $\sum_{n_k<y} c'_k$, as a sum of residues of $y^sF(s)/s$, provided $y$ avoids the $n_k$'s. Estimating such sums as $y \to \infty$ will be a tractable problem. In a sense that will become clear later, the simple root $x_0$ of $f-1$ contributes the dominant term to the asymptotic and the other roots in $\mathbf{C}$ (all lying outside of the circle $|z| = x_0$, by Lemma 4.1) contribute terms of smaller order as $y$ gets large.

In order to relate $I_\sigma(y)$ to the partial sums $\sum_{n_k<y} c'_k$ when $y$ avoids the $n_k$'s, we make the following formal calculation (that can be justified rigorously): insert the generalized Dirichlet series expansion $y^sF(s) = \sum c'_k(y/n_k)^s$ into (5.2) for $\sigma > s_0$ and move the infinite sum through the integral to get

$$I_\sigma(y) = \frac{1}{2\pi\sqrt{-1}}\int_{\sigma-\infty\sqrt{-1}}^{\sigma+\infty\sqrt{-1}} \sum_{k\geq 1} c'_k\left(\frac{y}{n_k}\right)^s \frac{\mathrm{d}s}{s} = \sum_{k\geq 1}\frac{c'_k}{2\pi\sqrt{-1}}\int_{\sigma-\infty\sqrt{-1}}^{\sigma+\infty\sqrt{-1}}\left(\frac{y}{n_k}\right)^s\frac{\mathrm{d}s}{s}.$$

A direct calculation shows that for any $v > 0$ and *any* $\sigma \in \mathbf{R}$,

$$\frac{1}{2\pi\sqrt{-1}}\int_{\sigma-\infty\sqrt{-1}}^{\sigma+\infty\sqrt{-1}} v^s\frac{\mathrm{d}s}{s} = \begin{cases} 1 & v > 1, \\ 1/2 & v = 1, \\ 0 & v < 1. \end{cases}$$

Since $y$ is not equal to any of the $n_k$'s (so $y/n_k \neq 1$ for all $k$), we deduce that $I_\sigma(y)$ is convergent and in fact we obtain the well-known *Perron–Mellin formula*

$$I_\sigma(y) = \sum_{n_k<y} c'_k.$$

Our aim is therefore to prove that $I_\sigma(y)/y^{s_0}x_0^{\{t\}}$ tends to $A'$ as $y \to \infty$ with $y = e^t$ avoiding the $n_k$'s.

Our study of the vertical line integral $I_\sigma(y)$ will proceed in the usual manner by identifying it with a limit of integrals around rectangles, with the latter integrals computed in terms of residues of $y^sF(s)/s$. Recall that for a meromorphic function $h$ on $\mathbf{C}$, we can use residues to compute the path integral $(1/2\pi\sqrt{-1})\int_C h(s)\mathrm{d}s$ around a rectangular path $C$ that is disjoint from the discrete pole set of $h$ and is given the counterclockwise orientation:

$$\frac{1}{2\pi\sqrt{-1}}\int_C h(s)\mathrm{d}s = \sum_b \mathrm{Res}(h; b),$$

where $b$ runs over the finitely many poles of $h$ on the interior of the rectangle with boundary $C$. We apply this to $h(s) = y^sF(s)/s$, taking $C$ to be the rectangle $C_{R,\sigma'}$ whose right side is the vertical segment $[\sigma - R\sqrt{-1}, \sigma + R\sqrt{-1}]$ with $\sigma > s_0$ and whose left side is the vertical segment $[\sigma' - R\sqrt{-1}, \sigma' + R\sqrt{-1}]$ where $\sigma' \ll 0$. We also choose the height $R$ to ensure that the top and bottom edges avoid poles of the integrand, and the periodic nature of the pole set of $F$ provides an $\varepsilon > 0$ so that the nonzero vertical gaps between poles of the integrand $y^sF(s)/s$ are always $> \varepsilon$, regardless of how far up and down we go. Thus, we may choose the height $R \gg 0$ as if threading a needle through $V$ so that the top and bottom edges are always at least $\varepsilon/2$ away from all poles of the integrand.

Because our integrand has no poles far to the left, we can move the left side of the rectangle $C_{R,\sigma'}$ as far to the left as we please (i.e., take $\sigma' \ll 0$) without affecting the *value* of the path integral under consideration. Since $F(s)$ is vertically periodic ($F(s + 2\pi\sqrt{-1}) = F(s)$) with exponential decay far off to the left, the contribution of the left edge to the path integral tends to 0 as the left edge is moved off to $-\infty$ while keeping the top and bottom edges at a fixed height. Since $|y^s/s| = y^{\mathrm{Re}(s)}/|s|$ dies off exponentially to the left and dies off (like $1/|s|$) in bounded vertical strips as we go far up and far down, the vertical *periodicity* of $F$ implies that that there are no convergence problems along the top and bottom edges as we make them become infinitely long off to the left, so

$$\frac{1}{2\pi\sqrt{-1}}\int_{C_{R,R'}} \frac{y^sF(s)}{s}\mathrm{d}s = \frac{1}{2\pi\sqrt{-1}}\int_{C_{R,-\infty}} \frac{y^sF(s)}{s}\mathrm{d}s$$

(with the right side an absolutely convergent path integral). When estimating the decay of the integral of $y^sF(s)\mathrm{d}s/s$ along the top and bottom edges of the infinitely long "rectangle" $C_{R,-\infty}$, as long as we choose $R$ to maintain these edges at a fixed positive distance $\geq \varepsilon/2$ away from all poles of the integrand we see that the contribution of these top and bottom edge integrals vanishes in the limit as $R \to \infty$.

The upshot is that if we apply the residue theorem to the integral of $y^sF(s)/s$ around a box $C_{R,\sigma'}$, the integral is the sum of residues of the integrand at poles interior to the box, but when we first send the left side off to $-\infty$ (introducing no new poles inside of the region of integration) and then send the top and bottom off to $\pm\infty\sqrt{-1}$ (acquiring more pole terms from the residue theorem), all that survives is the integral along the right edge $\sigma + \mathbf{R}\sqrt{-1}$. This remaining line integral is exactly $I_\sigma(y)$, so we conclude from the residue theorem that

$$I_\sigma(y) = \mathrm{Res}(y^sF(s)/s;0) + \sum_\alpha \sum_{n\in\mathbf{Z}} \mathrm{Res}(y^sF(s)/s; s_\alpha + 2\pi\sqrt{-1}\cdot n), \tag{5.3}$$

where $\sum_{n\in\mathbf{Z}}$ means $\lim_{N\to\infty}\sum_{|n|\le N}$. In (5.3), $s_\alpha$ runs over the *finitely many* solutions to $f(e^{-s}) = 1$ in the horizontal region $-\pi \le \mathrm{Im}(s) < \pi$. To be precise about (5.3), we pair off $s_\alpha \pm 2\pi\sqrt{-1}\cdot n$ with $n > 0$ and fixed $\alpha$, and we handle the terms $n = 0$ separately. The residue term at $s = 0$ is $y^0F(0) = 1/(1-\sum w_i)$, a constant that is independent of $y$, so it may (and will) be ignored for our study of behavior as $y \to \infty$ (recall that $s_\alpha + 2\pi\sqrt{-1}\cdot n$ is always nonzero, since $f(0) \ne 1$).

We shall analyze (5.3) by treating the contribution of $\{s_\alpha + 2\pi\sqrt{-1}\cdot n\}_{n\in\mathbf{Z}}$ separately for each $\alpha$, first considering the case of $s_0 = s_{\alpha_0}$, the unique rightmost solution (modulo $2\pi\sqrt{-1}\cdot\mathbf{Z}$) to $f(e^{-s}) = 1$. The terms for $\alpha = \alpha_0$ in (5.3) enjoy the crucial property that the pole at each $s_0 + 2\pi\sqrt{-1}\cdot n$ is a *simple* pole, essentially due to Lemma 4.1 (and the fact that pole order is unaffected by the change of variables $x = e^{-s}$).

For clarity, consider *any* $\alpha$ such that $F$ has a simple pole at the points $s_\alpha + 2\pi\sqrt{-1}\cdot n$ (so $s_\alpha$ may not be real). By (4.4) and the definition of $F$ (as periodic mod $2\pi\sqrt{-1}\cdot\mathbf{Z}$) we get

$$\begin{aligned}\mathrm{Res}(y^sF(s)/s; s_\alpha + 2\pi\sqrt{-1}\cdot n) &= \frac{y^{s_\alpha+2\pi\sqrt{-1}\cdot n}}{s_\alpha + 2\pi\sqrt{-1}\cdot n}\mathrm{Res}(F; s_\alpha + 2\pi\sqrt{-1}\cdot n)\\ &= \frac{y^{s_\alpha+2\pi\sqrt{-1}\cdot n}}{s_\alpha + 2\pi\sqrt{-1}\cdot n}\lim_{s\to s_\alpha}\frac{s-s_\alpha}{f(e^{-s_\alpha}) - f(e^{-s})}\end{aligned} \tag{5.4}$$

(since $f(e^{-s_\alpha}) = 1$), and the final limit is the reciprocal of $(\mathrm{d}/\mathrm{d}s)(-f(e^{-s}))|_{s=s_\alpha} = \rho_\alpha f'(\rho_\alpha) \ne 0$ (recall $\rho_\alpha = e^{-s_\alpha} \ne 0$). When $\alpha = \alpha_0$, this is $x_0f'(x_0) > 0$. Adding up (5.4) over $n \in \mathbf{Z}$, for any $\alpha$ such that $\mathrm{ord}_{s_\alpha}(F) = -1$ the contribution of the $\alpha$-term in (5.3) is

$$\frac{y^{s_\alpha}}{\rho_\alpha f'(\rho_\alpha)}\sum_{n\in\mathbf{Z}}\frac{y^{2\pi\sqrt{-1}\cdot n}}{s_\alpha + 2\pi\sqrt{-1}\cdot n}. \tag{5.5}$$

The sum in (5.5) only depends on $y$ through the point $y^{2\pi\sqrt{-1}}$ on the unit circle, so we may apply:

**Lemma 5.1.** *For fixed $s \in \mathbf{C} - 2\pi\sqrt{-1}\cdot\mathbf{Z}$ and all $z \in \mathbf{C}$ with $|z| = 1$, the series*

$$H_s(z) = \sum_{n\in\mathbf{Z}}\frac{z^n}{s + 2\pi\sqrt{-1}\cdot n} = -H_{-s}(1/z)$$

*is convergent with $H_s(1) = 1/s + \sum_{n\ge1} 2s/(s^2 + 4\pi^2n^2)$, and*

$$H_s(e^{2\pi\sqrt{-1}\cdot\theta}) = \frac{e^{s(1/2-\theta)}}{e^{s/2} - e^{-s/2}} = \frac{e^{-\theta s}}{1 - e^{-s}} \tag{5.6}$$

*for $0 < \theta < 1$. In particular, $H_s(z)$ is continuous in $z$ (on the circle) away from $z = 1$, and it is bounded away from $0$ and $\infty$ in absolute value.*

*For real $s > 0$, $H_s$ away from $z = 1$ is strictly decreasing and positive as we move around the circle in the $\sqrt{-1}$-direction (i.e., counterclockwise), so $H_s$ has* $\liminf$ *and* $\limsup$ *on $\{|z| = 1, z \ne 1\}$ equal to $1/(e^s - 1)$ and $1/(1 - e^{-s})$ respectively. When $s \in \mathbf{C} - 2\pi\sqrt{-1}\cdot\mathbf{Z}$, we have $|H_s(z)| \le \max(1, e^{-\mathrm{Re}(s)})/|1 - e^{-s}|$ for $z \ne 1$ with $|z| = 1$.*

*Proof.* For $z = 1$, combining $n$th and $-n$th terms makes the formula for (and convergence of) $H_s(1)$ clear. Thus, we now consider $z$ on the unit circle with $z \ne 1$, and we shall derive the proposed explicit formula, from which everything else is obvious. For $\theta \in (0,1)$ we compute:

$$H_s(e^{2\pi\sqrt{-1}\cdot\theta}) = \sum_{n\in\mathbf{Z}} \frac{e^{2\pi\sqrt{-1}\cdot n\theta}}{s+2\pi\sqrt{-1}\cdot n} = \sum_{n\in\mathbf{Z}} \frac{(-1)^n}{s-2\pi\sqrt{-1}\cdot n}\cdot e^{2\pi\sqrt{-1}\cdot n(1/2-\theta)},$$

where $-1/2 < 1/2-\theta < 1/2$.

Thus, it suffices to prove that for $\theta' \in (-1/2,1/2)$ and $s \in \mathbf{C} - 2\pi\sqrt{-1}\cdot\mathbf{Z}$,

$$\sum_{n\in\mathbf{Z}} \frac{(-1)^n}{s-2\pi\sqrt{-1}\cdot n}\cdot e^{2\pi\sqrt{-1}\cdot n\theta'} = \frac{e^{s\theta'}}{e^{s/2}-e^{-s/2}}.$$

Since $f_s(\theta') = e^{s\theta'}/(e^{s/2}-e^{-s/2})$ is continuous on $(-1/2,1/2)$ with bounded derivative (hence $f_s$ is of bounded variation), it is pointwise equal to its formal Fourier expansion $\sum_{n\in\mathbf{Z}} \widehat{f}_s(-n)e^{2\pi\sqrt{-1}\cdot n\theta'}$, where the Fourier transform $\widehat{f}_s$ on $\mathbf{Z}$ is given by the usual formula $\widehat{f}_s(n) = \int_{-1/2}^{1/2} f_s(\theta')e^{2\pi\sqrt{-1}\cdot n\theta'}\,\mathrm{d}\theta'$. Thus, it suffices to prove that $\widehat{f}_s(n) = (-1)^n/(s+2\pi\sqrt{-1}\cdot n)$ for $n \in \mathbf{Z}$. This is a simple integral computation. □

By Lemma 5.1 with $\theta_t = \{t\}$, when $y^{2\pi\sqrt{-1}} = e^{2\pi\sqrt{-1}\{t\}} \neq 1$ (a condition that causes $y$ to avoid the discrete set $e^{\mathbf{Z}}$ that contains the $n_k$'s) we see that for $\alpha = \alpha_0$ the ratio between (5.5) and $y^{s_0}x_0^{\{t\}} = y^{s_\alpha}e^{-\theta_t s_0}$ is positive and has limiting value equal to $A' > 0$ as $y \to \infty$ (recall that $e^{-s_0} = x_0 < 1$, so $s_0 > 0$). For the *other* $s_\alpha$'s at which there is a simple pole, Lemma 5.1 provides an explicit upper bound of $A'_\alpha y^{\mathrm{Re}(s_\alpha)}$ on the $\alpha$-contribution when $y^{2\pi\sqrt{-1}} \neq 1$, where $A'_\alpha = \max(1,|\rho_\alpha|^{-1})/(|1-\rho_\alpha||\rho_\alpha f'(\rho_\alpha)|)$. Since we can make $A'_\alpha y^{\mathrm{Re}(s_\alpha)-s_0}$ as small as we please by taking $y$ sufficiently large (because $\mathrm{Re}(s_\alpha) = -\log|\rho_\alpha| < s_0$, by Lemma 4.1), this contribution is $o(y^{s_0})$. Additional division by $x_0^{\{t\}}$ does not affect this estimate since $x_0^{\{t\}}$ is bounded away from 0 and $\infty$ as $t$ varies. Provided we show that the contribution in (5.3) from the other $s_\alpha$'s (where there is a higher order pole) is also $o(y^{s_0})$ as $y$ gets large, the contribution to $I_\sigma(y)$ from the $s_0$-term dominates the rest. This would verify the criterion in Lemma 3.5, taking $t = \log(y)$ not in $\mathbf{Z}$.

In order to estimate the contributions from $s_\alpha$'s such that $f$ has a zero of order $\mu_\alpha > 1$ at $\rho_\alpha = e^{-s_\alpha}$, consider the factorization $f(z) - 1 = w_1(z-\rho_\alpha)^{\mu_\alpha}\cdot\prod_{\alpha'\neq\alpha}(z-\rho_{\alpha'})^{\mu_{\alpha'}}$ in $\mathbf{C}[z]$. Replacing $z$ with $e^{-s}$, we get $F(s) = w_1^{-1}(e^{-s}-e^{-s_\alpha})^{-\mu_\alpha}\cdot\prod_{\alpha'\neq\alpha}(e^{-s}-e^{-s_{\alpha'}})^{-\mu_{\alpha'}}$. We want to insert this into the general residue recipe (4.3) for $g(s) = y^sF(s)/s$, but we will avoid trying to be explicit with the residue computation at $s_\alpha + 2\pi\sqrt{-1}\cdot n$ since all that we require is an upper bound of the form $o(y^{s_0})$ as $y \to \infty$.

Explicitly computing the $\mu_\alpha$-fold derivative as in (4.3) is not necessary. Instead, we shall focus on the structure of the formula to get an upper bound. Using Leibnitz' rule for differentiating multiple products finitely many times and considering the higher derivatives of $y^s/s$ (finite sums of terms of the type $y^{s-m}(\log y)^q/s^r$, where $m$ and $q$ are non-negative), it is clear that the contribution to (5.3) from $s_\alpha \neq s_0$ consists of two types of sums: a finite set of sums that (up to bounded multipliers and various multinomial coefficients that we suppress) are of the shape in (5.5) with $s_0$ replaced by $s_\alpha$ (all such sums being $o(y^{s_0})$ since $\mathrm{Re}(s_\alpha) < s_0$), and finitely many more sums that (again up to bounded multipliers) have the form

$$(\log y)^q y^{s_\alpha}\cdot\sum_{n\in\mathbf{Z}} \frac{y^{2\pi\sqrt{-1}\cdot n}}{(s_\alpha+2\pi\sqrt{-1}\cdot n)^r} \tag{5.7}$$

for some $r \geq 2$ and a fixed $\alpha$. The sums in (5.7) are absolutely convergent since the terms of the sum are $O(|n|^{-r})$ as $n \to \pm\infty$ with fixed $r \geq 2$; the implicit constant in the $O$-estimate does not depend on $y$ since $|y^{2\pi\sqrt{-1}\cdot n}| = 1$. Thus, we get an estimate $O(|y^{s_\alpha}|(\log y)^q)$ for (5.7) as $y \to \infty$. But $|y^{s_\alpha}|(\log y)^q = o(y^{s_0})$ as $y \to \infty$ since $|y^{s_\alpha}| = y^{\mathrm{Re}(s_\alpha)}$ with $\mathrm{Re}(s_\alpha) < s_0$ and $(\log y)^q = o(y^\delta)$ for all $\delta > 0$. This completes the proof of Theorem 3.1.

*Remark* 5.2. For $y^{2\pi\sqrt{-1}} \neq 1$, the infinite series in (5.7) can even be computed in closed form by repeated differentiation of (5.6) with respect to $s$, so in fact (5.3) can be presented in closed form as a finite sum when $y \notin e^{\mathbf{Z}}$.

## 6. The remaining cases

Since we found explicit formulas for the limit in the case of rational ratios $a_i/a_{i'}$ in Theorem 3.1, it is tempting to try to "pass to the limit" via rational approximation of irrationals to guess what to expect in the case when some ratio $a_i/a_{i'}$ is irrational. Such irrational ratios cannot occur when $n = 1$, so in this section we assume $n > 1$.

To motivate things, observe that as an $n$-tuple $(a_1, \dots, a_n)$ with rational ratios $a_i/a_{i'}$ converges to an $n$-tuple with an irrational ratio, the value of $r = D/a_1$ must explode to $\infty$ because the common denominator $D$ must grow without bound while $a_1$ stays bounded (and away from zero). Note that the unique zero $x_0$ of $f_{\underline{a}} - 1$ on $(0, \infty)$ is a continuous (and even analytic) function of the $a_j$'s (by the Implicit Function Theorem) and $\lfloor rt \rfloor / r \in (t - 1/r, t]$ uniformly converges to $t$ as $r \to \infty$, Also, as $r \to \infty$ we have uniformly

$$\frac{1}{r(1 - {x_0}^{1/r})}, \ \frac{x_0^{1/r}}{r(1 - x_0^{1/r})} \to \frac{1}{\log(1/x_0)}$$

for $x_0 \in (0, 1)$ in a fixed small region. Since $x_0$ is continuous in the $a_i$'s, when we slightly move the $a_i$'s it follows that $x_0$ remains within a fixed small region in $(0, 1)$.

One may be tempted to believe that the asymptotics should behave roughly continuously in input data $(a_i)$ for fixed $w_1, \dots, w_n$. This is actually false, since Theorem 3.1 shows it to fail when approximating a sequence $(a_i)$ with *rational* ratios $a_i/a_{i'}$ by other arbitrarily close (but *distinct*) such sequences with rational ratios. Approximating a rational number by an infinite sequence of (distinct) rationals is pathological. If we recall that the problem of interest is one of counting the number of ways to express numbers below some bound as sums of the $a_i$'s (with fixed weights $w_i$), it is not unreasonable to imagine that the count might behave more "continuously" in the $a_i$'s for fixed weights if we only use approximations of the $a_i$'s that lead to *good* rational approximations of the $a_i/a_{i'}$'s, say through continued fractions. This restriction causes rational sequences $(a_i)$ with $a_1 = 1$ to be non-approximable by infinitely many distinct rational sequences $(a'_i)$ with $a'_1 = 1$, and so it eliminates the pathology (as well as any meaningful limit process in cases with rational ratios) and leads one to predict the asymptotic in the following result that covers *precisely* the "irrational" cases not handled by Theorem 3.1:

**Theorem 6.1.** (H. Montgomery) *Fix $n > 1$, positive reals $w_1, \dots, w_n$ satisfying $\sum w_i > 1$, and a strictly decreasing sequence $a_1 > \cdots > a_n > 0$ with some $a_i/a_{i'}$ irrational. Define $f_{\underline{a}}(x) = \sum w_i x^{a_i}$ for $x \geq 0$, and let $x_0 \in (0, 1)$ be the unique non-negative solution to the equation $1 = f_{\underline{a}}(x)$. Define*

$$B = \frac{1}{\log(1/x_0) x_0 f'_{\underline{a}}(x_0)} > 0.$$

*Consider the formal expansion $1/(1 - f_{\underline{a}}(x)) = \sum_{\nu \geq 0} c_\nu x^\nu$, with $c_\nu \neq 0$ for a discrete set of $\nu$'s. As $t \to \infty$, $\sum_{\nu \leq t} c_\nu \sim B(1/x_0)^t \sim \sum_{\nu < t} c_\nu$.*

We remind the reader that in the motivating case of the variant on Miller's experiment with letter probabilities that are not all equal, the ratio $a_i/a_{i'}$ is the exponent that arises when expressing the $i$th probability $p_i$ as a power of the $i'$th probability $p_{i'}$ (when distinct probabilities are labelled $p_1 < \cdots < p_n$). Thus, in most interesting situations at least one ratio $a_i/a_{i'}$ is irrational.

*Remark* 6.2. The scaling arguments as above Theorem 3.1 show that the assertion in the theorem (including the value of $B$) is unaffected by common scaling on the $a_i$'s. To be precise, if $a'_i = ra_i$ then we have seen that $x'_0 = x_0^{1/r}$ and $x'_0 f'_{\underline{a}'}(x'_0) = r x_0 f'_{\underline{a}}(x_0)$, so clearly

$$\log(1/x'_0) x'_0 f'_{\underline{a}'}(x'_0) = \log(1/x_0) x_0 f'_{\underline{a}}(x_0).$$

Such scaling-invariance does not play a role in the proof of the theorem.

*Remark* 6.3. Although Montgomery's proof of Theorem 6.1 will require more analytic input than the proof of our complementary Theorem 3.1, as we shall explain below, it is interesting that (as we saw above) the concrete formula (3.7) in the case of rational ratios does naturally lead to a prediction of the asymptotic proved by Montgomery in all other cases. Hence, even though the case of rational ratios is of much less significance in practice, its more elementary character and predictive power provides an interesting conceptual and intuitive way for a non-mathematician to understand the difference in behavior between the case of rational ratios and all other cases.

In the remainder of this section, we present Montgomery's proof of Theorem 6.1; the proof uses more advanced analytic methods. We are grateful to Montgomery for permission to explain his argument. The irrationality of some ratio $a_i/a_{i'}$ does play an essential role in the analysis, so the following argument does not also prove Theorem 3.1 (though a refinement of the method likely provides an alternative, more mathematically sophisticated, approach to Theorem 3.1). Let us begin by isolating a property that is a variant on Lemma 4.1 and explains the mathematical significance of an irrational ratio.

In order to describe this variant, let $w_1, \ldots, w_n$ be positive reals satisfying $\sum w_i > 1$, and let $a_1 > \cdots > a_n > 0$ be an arbitrary strictly decreasing sequence of positive real numbers. Let

$$F(s) = 1/(1 - \sum w_i e^{-a_i s}),$$

and let $s_0 > 0$ be the unique positive number such that $\sum w_i e^{-a_i s_0} = 1$. As was explained in the proof of Lemma 4.1 without using a rationality condition on the $a_i$'s, all poles of $F$ satisfy $\mathrm{Re}(s) \leq s_0$ (and there is certainly a pole at $s_0$). This is really a logarithmic reformulation of a conclusion in the proof of Lemma 4.1, adapted to the case of arbitrary (possibly non-integral) positive $a_i$'s; recall that when $a_i > 0$ is not an integer, the expression $z^{a_i}$ (as in $f_{\underline{a}}(z) = \sum w_i z^{a_i}$) is not well-defined for $z \in \mathbf{C} - \{0\}$ not a positive real number. The importance of the irrationality of some ratio $a_i/a_{i'}$ is:

**Lemma 6.4.** *With notation as above, the pole of $F$ at $s = s_0$ is the unique pole of $F$ on the line* $\mathrm{Re}(s) = s_0$ *if and only if some ratio $a_i/a_{i'}$ is irrational.*

Recalling the use of the periodic sequence of simple poles for $F$ along $s_0 + 2\pi\sqrt{-1}\cdot\mathbf{Z}$ in the computation of the dominant term in the proof of Theorem 3.1, one can appreciate a priori that different asymptotic behavior may be expected when there is *uniqueness* of $s = s_0$ as the rightmost pole of $F$ under the hypothesis of Theorem 6.1. This uniqueness from Lemma 6.4 is crucial in Montgomery's proof of Theorem 6.1.

*Proof.* Assuming $F$ to have a pole at $s_0 + \sqrt{-1}\cdot t$ with $t \in \mathbf{R}$, we wish to show that $t = 0$ is forced precisely when some ratio $a_i/a_{i'}$ is irrational. The condition of a pole at $s_0 + \sqrt{-1}\cdot t$ says

$$\sum w_i e^{-a_i s_0} e^{-a_i\sqrt{-1}\cdot t} = 1 = \sum w_i e^{-a_i s_0}.$$

Since the terms on the right all lie on the same half-line (namely, $(0,\infty)$) in $\mathbf{C}$, the equality of these sums happens if and only if the phase shifts $e^{a_i\sqrt{-1}\cdot t}$ all equal 1 (argue via angle cancellation, as in the proof of Lemma 4.1). That is, the pole condition at $s_0 + \sqrt{-1}\cdot t$ says exactly that $a_i t \in 2\pi\mathbf{Z}$ for all $i$. Hence, we need to check that the implication

$$a_1 t, \ldots, a_n t \in 2\pi\mathbf{Z} \Longrightarrow t = 0$$

holds if and only if some ratio $a_i/a_{i'}$ is irrational. Clearly if the implication fails, so $a_i t_0 = 2\pi n_i$ for some nonzero $t_0$ and (necessarily nonzero) integers $n_i$, we compute that $a_i/a_{i'} = n_i/n_{i'} \in \mathbf{Q}$ for all $i, i'$. Conversely, if all ratios $a_i/a_{i'}$ are rational, then by taking $t = 2\pi N/a_1 \neq 0$ for a positive integer $N$ divisible by the denominators of all ratios $a_i/a_1 \in \mathbf{Q}$, we get $a_i t \in 2\pi\mathbf{Z}$ for all $i$. □

Lemma 6.4 puts us in the following situation. We have a meromorphic function

$$F(s) = \frac{1}{1 - \sum w_i e^{-a_i s}} = \sum_{\nu \geq 0} c_\nu e^{-\nu s}$$

that is analytic in the half-plane $\mathrm{Re}(s) > s_0$, and has a unique pole on $\mathrm{Re}(s) = s_0$ at the real point $s = s_0$. Given this, the aim is to deduce that $\sum_{\nu<t} c_\nu \sim Be^{s_0 t}$ as $t \to \infty$, with $B = 1/\log(1/x_0)x_0 f'(x_0)$, where $x_0 \in (0,1)$ is the unique positive real number such that $f(x) = \sum w_i x^{a_i}$ is equal to 1 at $x_0$. By the elementary argument proving Lemma 3.5, this result also yields the asymptotic $\sum_{\nu\leq t} c_\nu \sim Be^{s_0 t}$.

The basic problem is to make asymptotic estimates on partial sums of (generalized) Dirichlet series coefficients when the (generalized) Dirichlet series of interest is meromorphic around a closed right half-plane with a pole on the real line that its *unique* right-most pole in $\mathbf{C}$. This type of asymptotic estimation problem is ubiquitous in analytic number theory, where one studies functions such as the Riemann zeta function $\zeta(s) = \sum_{n\geq 1} n^{-s}$ (and its more sophisticated variants) that have a meromorphic continuation beyond $\mathrm{Re}(s) \geq 1$ with a unique (simple) right-most pole that is moreover located at $s = 1$. In our situation, if we consider the function

$$G(s) = F(ss_0) = \sum_{\nu\geq 0} c_\nu e^{-\nu s_0 s} = \sum_{\nu'\geq 0} c'_{\nu'} e^{-\nu' s}$$

with $c'_{\nu'} = c_{\nu'/s_0}$ (i.e., $\nu' = \nu s_0$), the pole is shifted to $s = 1$ and $\sum_{\nu<t} c_\nu = \sum_{\nu'<ts_0} c'_{\nu'}$. Thus, one can restate Theorem 6.1 as the asymptotic $\sum_{\nu'<t'} c'_{\nu'} \sim Be^{t'}$. If we let $0 = \nu'_1 < \nu'_2 < \dots$ be the values $\nu'$ such that $c'_{\nu'} \neq 0$ and define $C_k = c'_{\nu'_k}$, then for $n'_k = e^{\nu'_k}$ we have $G(s) = \sum_{k\geq 1} C_k {n'_k}^{-s}$ and $\sum_{\nu'<t'} c'_{\nu'} = \sum_{n'_k<y} C_k$ with $y = e^{t'}$. The aim is therefore to prove $\sum_{n'_k<y} C_k \sim By$ as $y \to \infty$.

But $G$ is a generalized Dirichlet series that is absolutely and uniformly convergent in right half-planes $\mathrm{Re}(s) \geq 1 + \varepsilon$ ($\varepsilon > 0$) and $G$ has a meromorphic continuation past $\mathrm{Re}(s) = 1$ with a *unique* pole on the line $\mathrm{Re}(s) = 1$ at the point $s = 1$. We compute that $(s-1)G(s) = (s-1)F(ss_0) = (s-1)/(1-\sum w_i e^{-a_i ss_0})$ is the reciprocal of the difference quotient for the derivative of $1 - \sum w_i e^{-a_i ss_0}$ at $s = 1$, and this derivative at $s = 1$ is equal to

$$\begin{aligned} -\sum w_i e^{-a_i ss_0} \cdot (-a_i s_0)|_{s=1} = \sum w_i e^{-a_i s_0} \cdot a_i s_0 &= \sum w_i x_0^{a_i} \cdot a_i \log(1/x_0) \\ &= \log(1/x_0) x_0 \cdot \sum w_i a_i x_0^{a_i - 1} \\ &= \log(1/x_0) x_0 f'(x_0), \end{aligned}$$

a nonzero quantity. Hence, $G$ has a *simple* pole at $s = 1$, so its residue at $s = 1$ is equal to

$$\lim_{s\to 1} (s-1)G(s) = 1/\log(1/x_0)x_0 f'(x_0) = B.$$

Now consider an arbitrary generalized Dirichlet series $H(s) = \sum_{k\geq 1} A_k {m_k}^{-s}$ with $\{1 = m_1 < \cdots\}$ a strictly increasing discrete sequence in $[1, \infty)$ and the $A_k$'s are non-negative real numbers. Assume that $H$ is convergent on $\mathrm{Re}(s) > 1$ and is meromorphic on $\mathrm{Re}(s) \geq 1$ with a unique right-most pole that is moreover located at $s = 1$. Assume the pole at $s = 1$ is *simple*; we have seen that $G$ is such a function, taking $A_k = C_k$ and $m_k = n'_k$. For such $H$, we wish to establish the asymptotic estimate $\sum_{m_k<y} A_k \sim Ry$ as $y \to \infty$, where $R$ is the residue of $H$ at $s = 1$; for $H = G$ this would provide exactly what we need.

Consider the left-continuous monotonically increasing summatory function $f(u) = \sum_{\log m_k<u} A_k$ that jumps at the $\log m_k$'s with a gap of $A_k$ for the jump at $\log m_k$. The infinite series $H(s)$ for $\mathrm{Re}(s) > 1$ can be expressed as an absolutely convergent Riemann-Stieltjes integral $\int_0^\infty e^{-su}\, \mathrm{d}f(u)$. Since $H$ is assumed to meromorphically extend past the line $\mathrm{Re}(s) = 1$ with a simple pole at $s = 1$ having residue $R$, the Wiener-Ikehara Tauberian theorem gives $f(u) \sim Re^u$ as $u \to \infty$ (two references for this theorem are [11, §6.1, Thm. 3] and [16, Ch. V, §17, Thm. 17]). Now letting $y = e^u$, this says $\sum_{m_k<y} A_k \sim Ry$, as desired. This completes Montgomery's proof of Theorem 6.1.

## 7. Conclusion

Through arguments based on complex analysis, the question of power laws for word frequencies in the case of unequal key-stroke probabilities in Miller's random monkey experiment has been settled affirmatively in all cases, including some interesting subtleties. In certain cases, including the case originally analyzed by Miller, there is a multiplicative gap, corresponding to liminfs and limsups for which we have given explicit formulas. (The discussion at the end of Example 3.4 shows that, in principle, the gap might disappear when working in the discretized language of individual word probabilities $f_j$ (ordered by the rank-frequency parameter $j$), but (2.1) provides an explicit example when such a gap really occurs even at the level of word probabilities.) In the generic cases when the log-ratios of the probabilities are not all rational, this gap disappears and one has an explicit coefficient for the power law on word frequencies. Complex analysis explains this behavior via the pole structure of the generating function for the problem, when this generating function is viewed as a function of a complex variable through an exponential change of parameter $x = e^{-s}$.

It would of course be pleasant to have a proof of the power law behavior in the case of unequal probabilities that avoids some of this technical machinery. One possible approach is the following. Suppose that a random variable $X_t$ has a lognormal distribution with mean $\mu t$ and variance $\sigma^2 t$. Now consider a random variable $Y$ chosen according to the distribution $X_t$, where the value of $t$ is itself an exponentially distributed random variable. It is known that the distribution of $Y$ follows a power law [1, 9, 12]. Miller's experiment is quite similar. If we let $t$ be the number of characters in a word, it follows a geometric distribution and this approximates the exponential distribution. If we let $X_t$ correspond to the probability that a word chosen uniformly at random from all words of length $t$ is generated by Miller's experiment, then $X_t$ is approximately lognormal for sufficiently large $t$ [10]. If these approximations are sufficiently good, one might expect to obtain that the rank-frequency of words in Miller's experiment approximately follows a power law. We are somewhat skeptical, however, that an approach through probability theory can yield the rich insights obtained from utilizing the methods of analytic number theory, particularly an understanding of the significance of the arithmetic condition of rationality of log-ratios of the probabilities.

## References

[1] B. A. Huberman and L. A. Adamic. Evolutionary dynamics of the World Wide Web. Technical Report, Xerox Palo Alto Research Center, 1999. Appears as a brief communication in *Nature*, 401, p. 131, 1999.

[2] J. M. Carlson and J. Doyle. Highly optimized tolerance: a mechanism for power laws in designed systems. *Physics Review E*, 60(2):1412-1427, 1999.

[3] A. Fabrikant, E. Koutsoupias, and C. H. Papadimitriou. Heuristically Optimized Tradeoffs: A new paradigm for power laws in the Internet. In *Proceedings of the 29th International Colloquium on Automata, Languages, and Programming*, 2002.

[4] P. Flajolet and R. Sedgewick, *Analytical combinatorics*, in preparation.

[5] W. Gong, Y.Liu, V. Misra, and D. Towsley. On the tails of Web filesize distributions. In *Proceedings of the Thirty-Ninth Annual Allerton Conference on Communication, Control, and Computing*, pages 192-201, 2001.

[6] H. Mahmoud, *Evolution of random search trees*, Wiley-Interscience, New York, 1992.

[7] B. Mandelbrot. An informational theory of the statistical structure of languages. In **Communication Theory**, edited by W. Jackson, Betterworth, pages 486-502, 1953.

[8] G. A Miller. Some effects of intermittent silence. *American Journal of Psychology*, 70:311-314, 1957.

[9] M. Mitzenmacher. A Brief History of Generative Models for Power Law and Lognormal Distributions. To appear in *Internet Mathematics*.

[10] R. Perline. Zipf's law, the central limit theorem, and the random division of the unit interval. *Physical Review E*, 54(1):220-223, 1996.

[11] H. Pitt, *Tauberian theorems*, Oxford University Press, Oxford, 1958.

[12] W. J. Reed. The Pareto law of incomes – an explanation and an extension. *Physica A*, 319:469-486, 2003.

[13] H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425-440, 1955.

[14] W. Szpankowski, *Average case analysis of algorithms and sequences*, Wiley-Interscience, New York, 2001.

[15] G. Troll and P. bein Graben. Zipf's law is not a consequence of the central limit theorem. *Physical Review E*, 52(2):1347-1355, 1998.

[16] D. Widder, *The Laplace transform*, Princeton University Press, Princeton, 1941.

Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA
*E-mail address*: bdconrad@umich.edu

Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA
*E-mail address*: michaelm@eecs.harvard.edu