Mathematika 63 (2017) 602–621 © 2017 University College London
doi:10.1112/S0025579317000109

RANDOM MULTIPLICATIVE WALKS ON THE RESIDUES MODULO n

NATHAN McNEW

Abstract. We introduce a new arithmetic function $a(n)$ defined to be the number of random multiplications by residues modulo n before the running product is congruent to zero modulo n. We give several formulas for computing the values of this function and analyze its asymptotic behavior. We find that it is closely related to $P_1(n)$, the largest prime divisor of n. In particular, $a(n)$ and $P_1(n)$ have the same average order asymptotically. Furthermore, the difference between the functions $a(n)$ and $P_1(n)$ is $o(1)$ as n tends to infinity on a set with density approximately 0.623. On the other hand, however, we see that (except on a set of density zero) the difference between $a(n)$ and $P_1(n)$ tends to infinity on the integers outside this set. Finally, we consider the asymptotic behavior of the difference between these two functions and find that $\sum_{n\leqslant x}(a(n)-P_1(n)) \sim (1-\pi/4)\sum_{n\leqslant x} P_2(n)$, where $P_2(n)$ is the second largest divisor of n.

§1. *Background.* Consider a "multiplicative" random walk on the set $\mathbb{Z}/n\mathbb{Z}$ of residues modulo n. In each step of this walk a residue modulo n is chosen, uniformly at random, and the current state is multiplied by it. Because $\mathbb{Z}/n\mathbb{Z}$ does not form a group, but rather a monoid under multiplication, this is not a transitive walk. In fact, it is an absorbing random walk with a single absorbing state, $0 \pmod n$. So if the walk proceeds long enough then, with probability one, the walk will eventually arrive (and stay) at the residue zero.

For example, when $n = 6$ this random walk is equivalent to a random walk on the directed graph depicted in Figure 1. In this walk each of the edges leaving a vertex are equally likely to be chosen. Note that it is possible to step from the residues one and five (the units modulo six) to any of the residues modulo six, but this is not true for the other residues, and that the only edge leaving zero is a loop.

In this paper, we are interested in the expected time to absorption, the number of steps it takes on average to reach $0 \pmod n$, if the initial state of the walk is $1 \pmod n$, or equivalently any unit modulo n. For any positive integer n we denote by $a(n)$ the expected time to absorption in the monoid $\mathbb{Z}/n\mathbb{Z}$, and study the behavior of this arithmetic function. Analyzing the walk depicted for $n = 6$ above, one finds, for example, that $a(6) = 3.5$.

The problem of studying random walks on groups has been studied extensively, see for example [**1**, **9**]. Many common random walks can in fact

Received 17 September 2016, published online 12 May 2017.
MSC (2010): 11N37, 37A45 (primary).

https://doi.org/10.1112/S0025579317000109 Published online by Cambridge University Press

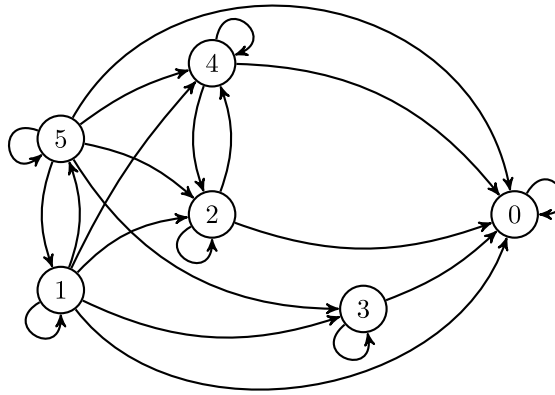


Figure 1: The directed graph depicting the random walk on the residues modulo six.

be viewed as walks on groups, for example the walk on the infinite line ($\mathbb{Z}$) or on a cycle ($\mathbb{Z}/n\mathbb{Z}$ under addition). More recently, several authors have considered random walks on monoids [8, 14]. Note that, in the literature, a random walk on a group or monoid is composed of steps which are taken from some generating set of the group (or monoid). In our case we take the generating set to be the entire monoid $\mathbb{Z}/n\mathbb{Z}$. In the case of a random walk on a group, one would not generally take the entire group as a generating set as it would then be possible to transition between any two states in a single step.

We could alternatively view our problem as a random walk on the positive integers, $\mathbb{N}$, in which we repeatedly multiply the current state by an integer chosen uniformly at random from the range $[1, n]$ (or $[1, mn]$ where $m \geqslant 1$ is an integer). Our function $a(n)$ is then the expected number of multiplications before the product is divisible by n.

The behavior of this function turns out to depend sensitively on the largest prime divisors of n. If $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$ with $p_1 > p_2 > \cdots > p_k$, we denote by

$$P_i(n) = \begin{cases} p_i & 1 \leqslant i \leqslant k, \\ 1 & i > k, \end{cases}$$

and by $B(n)$ the sum of the prime divisors of n taken with multiplicity,

$$B(n) = \sum_{i=1}^{k} \alpha_i p_i.$$

As we will see in §4 our function $a(n)$ is bounded between the functions $P_1(n)$ and $B(n)$. The behavior of these functions (along with a handful of similarly defined functions) has been studied extensively (see [2–6, 12, 13, 16]), much of it emphasizing how remarkably similar these two functions are. Given that our function is closely related to these two functions, we analyze it similarly.

https://doi.org/10.1112/S0025579317000109 Published online by Cambridge University Press

First, however, in §§2 and 3 we obtain several expressions for $a(n)$. Theorem 2.2 shows how to calculate $a(n)$ for any integer n, namely

$$a(n) = \frac{1}{n-\varphi(n)}\left(n + \sum_{\substack{d|n \\ d\neq n}} \varphi(d)a(d)\right).$$

Unfortunately, this expression is recursive, but we are able to obtain expressions that are not recursive in special cases. For prime powers (Theorem 2.1) we show that $a(p^k) = k(p-1)+1$, and for squarefree values of n we show (Theorem 3.1) that

$$a(n) = \sum_{\substack{d|n \\ d\neq 1}} (-1)^{\omega(d)+1} \frac{d}{d-\varphi(d)}.$$

In §4 we begin our study of the asymptotic behavior of this function by obtaining the average order of $a(n)$, (Corollary 4.4)

$$\sum_{n<x} a(n) \sim \frac{\pi^2 x^2}{12 \log x}$$

as $x \to \infty$, as a corollary of a result of Alladi and Erdős. This average order is asymptotically the same as both the functions $P_1(n)$ and $B(n)$.

Recall that for functions $f(x), g(x) > 0$ the notation $f(x) \sim g(x)$ is used to mean $\lim_{x\to\infty} (f(x)/g(x)) = 1$. We will also say $f(x) = O(g(x))$ or $f(x) \ll g(x)$ to mean $f(x) < Cg(x)$ for some absolute constant C and we write $f(x) = o(g(x))$ as $x \to \infty$ if $\lim_{x\to\infty} (f(x)/g(x)) = 0$.

Not only does $a(n)$ have the same average order as $P_1(n)$, but in fact we show in Theorem 5.1 that on a set of integers with density about 0.623 the function $P_1(n)$ is a very good approximation to $a(n)$ in the sense that $a(n) = P_1(n)+o(1)$ as n goes to infinity over the integers in this set. Surprisingly, however, we find that (with the exception of a set of measure zero) on the compliment of this set, where the approximation above is not true, we have that the difference between $a(n)$ and $P_1(n)$ tends to infinity.

As mentioned above, $a(n)$ is bounded between $P_1(n)$ and $B(n)$. In addition to having the same average order asymptotically, Alladi and Erdős show (see Theorem 4.2) that

$$\sum_{n<x} (B(n) - P_1(n)) \sim \sum_{n<x} P_2(n).$$

In other words, the sum of the prime factors of n, omitting the largest prime factor is completely dominated by the second largest prime factor. In light of this result, we consider the behavior of $a(n) - P_1(n)$. In Theorem 6.1 we show that there exist constants D_ℓ (whose values are given explicitly in the proof) such that for any fixed integer $N \geqslant 1$

$$\sum_{n\leqslant x} a(n) - P_1(n) = \frac{x^{3/2}}{\log^2 x} \sum_{\ell=0}^{N-1} \frac{D_\ell}{\log^\ell x} + O_N\left(\frac{x^{3/2}}{\log^{N+2} x}\right).$$

https://doi.org/10.1112/S0025579317000109 Published online by Cambridge University Press

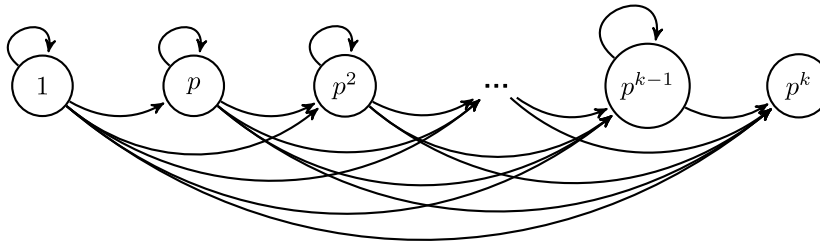


Figure 2: The directed graph depicting the random walk when $n = p^k$ is a prime power. Note that each node in this graph represents all residues that have the same greatest common divisor (gcd) with n rather than individual residues, and that not all edges have the same weight.

As in Corollary 6.2 we obtain that

$$\sum_{n<x}(a(n) - P_1(n)) \sim \left(1 - \frac{\pi}{4}\right)\sum_{n<x} P_2(n),$$

which we can interpret as saying that, in an average sense, the value of $a(n)$ is approximately the sum of the largest prime factor of n plus an extra contribution of a factor of $(1 - \pi/4)$ times its second largest prime factor.

Throughout the paper we will use X and Y to denote random variables, $\mathbb{E}[X]$ to denote the expected value of X, and $\mathbb{P}$ to denote the probability of an event occurring.

§2. *Computing the expected time to reach zero.* It does not appear to be possible to write down a general, closed formula for $a(n)$. We are, however, able to give a recursive formula and estimates that are generally quite accurate. The value of $a(n)$ is rarely an integer, and may never be unless n is a prime power. We begin with a few cases that are easily dealt with.

In the case when $n = p$ is prime, every element of $\mathbb{Z}/p\mathbb{Z}$ except $0 \pmod p$ is a unit. As a result there is only one way of reaching zero, namely by randomly choosing the residue zero as a multiple when stepping. This is then a Bernoulli trial with probability of success $1/p$, and thus the expected time to walk to zero is

$$a(p) = p.$$

The case $n = p^k$, when n is a prime power, is somewhat more enlightening. As in the prime case, the likelihood, at each step, of randomly selecting a unit (a residue coprime to p) is $(p-1)/p$. Since the only way to progress toward the absorbing state is to select a non-unit, we expect to wait for p iterations between "meaningful" steps that bring us closer to zero. Unlike the prime case, however, it is unlikely (though possible) that the first step by a non-unit will take us directly to zero. It is much more likely that we only pick up one or a couple of the necessary copies of p. In essence, we are taking a walk on the graph in Figure 2, in which each loop is weighted with probability $(p-1)/p$, each edge from p^i to p^j, $0 \leqslant i < j < k$ is weighted with probability $(p-1)/p^{i-j+1}$ and each edge from p^i to p^k by $1/p^{k-i}$.

https://doi.org/10.1112/S0025579317000109 Published online by Cambridge University Press

THEOREM 2.1. *When $n = p^k$ is a power of a prime the expected length of a multiplicative random walk on $\mathbb{Z}/p^k\mathbb{Z}$ is*

$$a(p^k) = k(p-1) + 1. \tag{1}$$

Proof. When $k = 1$ the result is the prime case discussed above. We now proceed by induction, and suppose that equation (1) holds for all exponents less than k. By linearity of expectation we can break the expected time up as

$$a(p^k) = \mathbb{E}[\text{time until first step by non-unit}] + \mathbb{E}[\text{time to step from there to zero}].$$

We know, as above, that the expected number of steps until our first non-unit is p, so we need only consider the second. Once we have taken our first step by a non-unit, say to a residue whose greatest common divisor (gcd) with p^k is p^i, we see that we need to pick up an additional $k-i$ copies of p. Doing so amounts to a walk on $\mathbb{Z}/p^{k-i}\mathbb{Z}$, which takes time $a(p^{k-i})$.

Since the probability that our first non-unit step contains exactly i copies of p is $(p-1)/p^i$ for $i < k$ and $1/p^k$ when $i = k$, we get that

$$\begin{aligned} a(p^k) &= p + \sum_{i=1}^{k-1} \frac{p-1}{p^i} a(p^{k-i}) + \frac{a(p^0)}{p^k} \\ &= p + \sum_{i=1}^{k-1} \frac{p-1}{p^i}((k-i)(p-1)+1). \end{aligned}$$

Here we have used the induction hypothesis and that $a(1) = 0$. Rewriting this sum, we get that

$$\begin{aligned} a(p^k) &= p + \frac{p-1}{p^k} \sum_{i=1}^{k-1} p^{k-i}((k-i)(p-1)+1) \\ &= p + \frac{p-1}{p^k} \sum_{i=1}^{k-1} p^i(i(p-1)+1) \\ &= p + \frac{(p-1)^2}{p^k} \sum_{i=1}^{k-1} ip^i + \frac{p-1}{p^k} \sum_{i=1}^{k-1} p^i \\ &= p + \frac{(p-1)^2}{p^k}\left(\frac{p-p^k}{(1-p)^2} - \frac{(k-1)p^k}{1-p}\right) + \frac{p-1}{p^k}\left(\frac{p-p^k}{1-p}\right) \\ &= p + \frac{p-p^k}{p^k} - (k-1)(1-p) - \frac{p-p^k}{p^k} \\ &= k(p-1)+1. \end{aligned}$$

This completes the inductive step and the proof. □

https://doi.org/10.1112/S0025579317000109 Published online by Cambridge University Press

The residues modulo an arbitrary n can be decomposed into the sets

$$R_d = \{r \in \mathbb{Z}/n\mathbb{Z} : (r, n) = d\} = d \cdot \left(\mathbb{Z}\Big/\frac{n}{d}\mathbb{Z}\right)^{\times},$$

for each of the divisors d of n. Each such set has size $\varphi(n/d)$. Our random walk on the residues modulo n can also be interpreted as a walk on the classes R_d, starting in $R_1 = (\mathbb{Z}/n\mathbb{Z})^{\times}$.

The key observation we made at the beginning of this proof to create a recursive formula for $a(p^k)$ works just as well for arbitrary composite n, although we are not able to solve the recurrence. For arbitrary n, the probability of randomly choosing a non-unit is $(n - \varphi(n))/n$, where $\varphi(n)$ is the Euler totient function, hence we expect it will take $n/(n - \varphi(n))$ steps before our first such choice. Having done so, the probability that our chosen residue has gcd d with n is $|R_d|/(n - |R_1|) = \varphi(n/d)/(n - \varphi(n))$ and thus we have the following recursive formula, which allows us to compute $a(n)$ for all n.

THEOREM 2.2. *For all $n > 1$ the expected length of a random walk on $\mathbb{Z}/n\mathbb{Z}$ is*

$$\begin{aligned} a(n) &= \frac{n}{n - \varphi(n)} + \sum_{\substack{d|n \\ d \neq 1}} \frac{\varphi(n/d)a(n/d)}{n - \varphi(n)} \\ &= \frac{1}{n - \varphi(n)}\left(n + \sum_{\substack{d|n \\ d \neq n}} \varphi(d)a(d)\right). \end{aligned}$$

§3. *Squarefree integers.* In the case that the integer n is squarefree we can solve the recursion in Theorem 2.2. It is convenient to use different methods, however, which will be useful later. In the squarefree case, reaching the residue $0 \pmod{n}$ amounts to having randomly chosen a residue divisible by each prime dividing n at least once in our walk. For any prime p dividing n, we can let X_p be a random variable which counts the number of steps taken before randomly selecting a residue divisible by p. Then all of the random variables X_p are independent, geometric random variables with $\mathbb{E}[X_p] = p$ and

$$a(n) = \mathbb{E}\Big[\max_{p|n}\{X_p\}\Big].$$

While the expected value of a maximum of random variables is not, in general, well behaved, it is straightforward to obtain a formula in the case of independent, geometric random variables. Note that this is possible despite the fact that our variables are not identically distributed. (Jeske and Blessinger observe [11] that this formula is surprisingly rare in the literature.)

THEOREM 3.1. *For $n \geqslant 2$ squarefree*

$$a(n) = \sum_{\substack{d|n \\ d \neq 1}} (-1)^{\omega(d)+1} \frac{d}{d - \varphi(d)}.$$

https://doi.org/10.1112/S0025579317000109 Published online by Cambridge University Press

Proof. Let $X = \max_{p|n}\{X_p\}$. Then

$$\begin{aligned} a(n) = \mathbb{E}\Big[\max_{p|n}\{X_p\}\Big] &= \sum_{i=0}^{\infty}\mathbb{P}[X > i] = \sum_{i=0}^{\infty}(1-\mathbb{P}[X \leqslant i]) \\ &= \sum_{i=0}^{\infty}\Big(1-\prod_{p|n}\mathbb{P}[X_p \leqslant i]\Big) \\ &= \sum_{i=0}^{\infty}\Big(1-\prod_{p|n}\Big(1-\Big(\frac{p-1}{p}\Big)^i\Big)\Big) \\ &= \sum_{i=0}^{\infty}\sum_{\substack{d|n\\ d\neq 1}}(-1)^{\omega(d)+1}\frac{\varphi(d)^i}{d^i} \\ &= \sum_{\substack{d|n\\ d\neq 1}}\frac{(-1)^{\omega(d)+1}}{1-\varphi(d)/d} = \sum_{\substack{d|n\\ d\neq 1}}(-1)^{\omega(d)+1}\frac{d}{d-\varphi(d)}. \end{aligned}$$

□

Note that for $n = p$ this reduces to the equation above, and for $n = pq$, (where $p \neq q$) gives

$$a(pq) = p + q - \frac{pq}{p+q-1} = p + \frac{q(q-1)}{p+q-1}. \tag{2}$$

This can also be obtained from Theorem 2.2.

§4. *The average order of the expected time.* Because it is necessary to have randomly chosen at least one residue divisible by each prime dividing n and, in particular, we expect the largest prime divisor, $P_1(n)$, to require $P_1(n)$ steps, we know that this function serves as a lower bound for $a(n)$. We can also see relatively easily that the function $B(n)$, the sum of all the prime divisors of n with multiplicity, gives an upper bound for $a(n)$.

LEMMA 4.1. *For any integer n, the function $a(n)$ satisfies the inequality*

$$P_1(n) \leqslant a(n) \leqslant B(n)$$

with equality occurring in each case if and only if n is prime.

Proof. Fix an integer $n > 1$. In the multiplicative walk on the residues modulo n the expected number of steps before choosing a residue divisible by $P_1(n)$ is exactly equal to $P_1(n)$. Choosing at least one residue divisible by $P_1(n)$ is necessary (but not necessarily sufficient) to have reached the absorbing state. Thus, $P_1(n) \leqslant a(n)$.

It follows, by linearity of expectation, that $a(n)$ is equal to $P_1(n)$ plus the expected number of steps required to pick up any remaining factors of n after the running product is first divisible by $P_1(n)$. If n is not prime, there is a non-zero probability there will still be some remaining factors of n to be picked up at this point (it could be, for example, that the very first step of the walk is to multiply

https://doi.org/10.1112/S0025579317000109 Published online by Cambridge University Press

by the residue $P_1(n)$), so we see that $P_1(n)$ is strictly less than $a(n)$ whenever n is composite.

We obtain the upper bound for $a(n)$ by considering a slightly different random walk. Write all of the prime factors of n (with multiplicity) as $p_1, p_2, \ldots, p_k$. This new walk begins by choosing residues modulo n uniformly at random, waiting until the first time we choose a residue divisible by p_1. Having obtained this prime factor we then repeat the process, choosing residues at random until one is chosen that is divisible by p_2, and so on, until all k prime factors of n are thus obtained. Since we expect each prime p_i dividing n to require p_i steps, the linearity of expectation implies that the expected number of steps for this walk is equal to $B(n)$.

Since this walk will take at least as long as the original random walk (any sequence of randomly chosen residues which completes this walk will also complete the original multiplicative walk) we see that $a(n) \leqslant B(n)$. Furthermore, in the case that n is composite there is a non-zero probability of obtaining multiple, or even all of the prime factors of n in a single step, and so $a(n) < B(n)$ in this case. □

Now, these functions, $P_1(n)$ and $B(n)$, despite arising as relatively crude lower and upper bounds for $a(n)$ are not so different. As a matter of fact, Alladi and Erdős studied both of them [2], and showed that both have the same average order.

THEOREM 4.2 (Alladi and Erdős [2]). *As $x \to \infty$,*

$$\sum_{n<x} P_1(n) \sim \sum_{n<x} B(n) \sim \frac{\pi^2 x^2}{12 \log x},$$

and, furthermore, the difference between these two sums is dominated by the contribution of the second largest prime factor,

$$\left(\sum_{n<x} B(n) - \sum_{n<x} P_1(n)\right) = \sum_{n<x} (B(n) - P_1(n)) = \sum_{n<x} P_2(n) + O\left(\frac{x^{4/3}}{\log^3 x}\right)$$

and that

$$\sum_{n<x} P_2(n) \sim C \frac{x^{3/2}}{(\log x)^2}$$

for some positive constant C.

Ivić [10] noted that the constant C in their result is equal to $8\zeta(3/2)/3$, attributing the result to Balasubramanian. He also showed that the sum of $P_2(n)$ admits an asymptotic expansion.

THEOREM 4.3 (Ivić [10]). *There exist constants A_l such that for any integer $N \geqslant 1$*

$$\sum_{n \leqslant x} P_2(n) = \frac{x^{3/2}}{\log^2 x} \sum_{l=0}^{N-1} \frac{A_l}{\log^l x} + O_N\left(\frac{x^{3/2}}{\log^{N+2} x}\right).$$

https://doi.org/10.1112/S0025579317000109 Published online by Cambridge University Press

We will return to Theorem 4.3 in §6. For now, we note that we can easily obtain the average order of $a(n)$ from Theorem 4.2 and Lemma 4.1.

COROLLARY 4.4. *As* $x \to \infty$,

$$\sum_{n<x} a(n) \sim \frac{\pi^2 x^2}{12 \log x}.$$

§5. *The typical behavior of* $a(n)$. In addition to having the same average order, Erdős and Pomerance [7] show that, on a set of asymptotic density one, $P_1(n) \sim B(n)$. So, in a sense these bounds answer the question of how large $a(n)$ is. But which function is a better estimate of $a(n)$? For the majority of integers n, $P_1(n)$ turns out to be a remarkably good approximation for $a(n)$, and in fact the difference between these two functions tends to zero on this set. However, for almost all of the remaining integers, a set which still has positive density, the difference $a(n) - P_1(n)$ tends to infinity.

THEOREM 5.1. *There exists a set* $\mathcal{G}$ *of asymptotic density one such that as* $n \to \infty$ *over the elements of* $\mathcal{G}$ *we have the following.*

- *If* $P_2(n) < P_1(n)^{1/2}$, *then* $a(n) = P_1(n) + o(1)$.
- *If* $P_2(n) > P_1(n)^{1/2}$, *then* $a(n) - P_1(n) \to \infty$.

The asymptotic density of the set of n *with* $P_2(n) < P_1(n)^{1/2}$ *has asymptotic density* $0.623\ldots$, *the Golomb–Dickman constant.*

Proof. The set of integers satisfying the inequality $P_2(n) < P_1(n)^{1/2}$ was studied by Wheeler [17], who showed that the density exists and is equal to the constant above. In what follows we disregard those n where any of the following hold:

(1) $P_1(n) < \log n$;
(2) $P_1(n)^{1/2}/\log n < P_2(n) < P_1(n)^{1/2} \log n$;
(3) n is divisible by a composite prime power greater than $P_2(n)$.

However, each of these conditions occur only on a set of integers with density zero. (The fact that these sets have density zero follows from standard smooth number estimates (1), from [17, Theorem 9] (2), and from the fact that $\sum_{p>x} (1/p^2)$ tends to zero as $x \to \infty$ (3), respectively.)

Let $p = P_1(n)$. By linearity of the expectation, we can write

$$a(n) = \mathbb{E}[X_p] + \mathbb{E}[Y],$$

where X_p is the random variable counting the number of steps before we randomly choose a residue divisible by p, and Y is the time required to pick up any remaining factors after choosing a residue divisible by p. Now $\mathbb{E}[X_p] = p$, so it suffices to study $\mathbb{E}[Y]$.

Let $q_1, q_2, \ldots$ denote the distinct, maximal prime powers dividing n/p and let the random variables Y_{q_i} denote the number of steps in the walk after acquiring

https://doi.org/10.1112/S0025579317000109 Published online by Cambridge University Press

a factor of p and before acquiring a factor of q_i. So

$$Y = \max_i \{Y_{q_i}\} \leqslant \sum_i Y_{q_i}.$$

In what follows we will assume that q_i is prime, as the prime power case will only be smaller. Owing to the memoryless property of the geometric distribution, we have that

$$\begin{aligned}
\mathbb{E}[Y_{q_i}] &= \mathbb{E}[X_{q_i}]P(X_p < X_{q_i}) \\
&= q_i \sum_{k=1}^{\infty} P(X_p = k)P(X_{q_i} > k) \\
&= q_i \sum_{k=1}^{\infty} \left(1 - \frac{1}{p}\right)^{k-1} \frac{1}{p} \left(1 - \frac{1}{q_i}\right)^k \\
&= \frac{q_i}{p}\left(1 - \frac{1}{q_i}\right) \sum_{k=0}^{\infty} \left(1 - \frac{1}{p}\right)^k \left(1 - \frac{1}{q_i}\right)^k \\
&= \frac{q_i - 1}{p}\left(\frac{1}{1 - (1 - 1/p)(1 - 1/q_i)}\right) \\
&= \frac{q_i - 1}{1 + p/q_i - 1/q_i} < \frac{q_i^2}{p}.
\end{aligned}$$

Now, in the first case, where $P_2(n) < P_1(n)^{1/2}$, we have that $q_i \leqslant P_2(n) \leqslant \sqrt{p}/\log n$ (since we disregarded the set of integers whose second largest prime factor was smaller than $\sqrt{p}$ by less than a factor of $\log n$) and so the expression above is less than $(\log n)^{-2}$. Since n has at most $\log n$ distinct prime factors, we have that

$$\mathbb{E}[Y] \leqslant \sum_i \mathbb{E}[Y_{q_i}] < \frac{\log n}{(\log n)^2} < \frac{1}{\log n} = o(1)$$

as $n \to \infty$.

In the second case where $P_2(n)^2 > P_1(n)$, we can assume that $q = P_2(n) > p^{1/2} \log n$, as we also disregarded above the set of integers where this was not the case, and so

$$\begin{aligned}
\mathbb{E}[Y_q] &= \frac{q - 1}{1 + p/q - 1/q} > \frac{q - 1}{1 + p/q} \\
&\gg \frac{q}{1 + p/q} \gg \frac{q}{p/q} = \frac{q^2}{p} > \log^2 n.
\end{aligned}$$

Since this goes to infinity as n does, we have that $a(n) - P_1(n) \geqslant \mathbb{E}[Y_q]$ does as well. □

The same method of proof used here can also be extended to show, for example, that (but for a set of density zero) on the integers where $P_3(n)^3 < P_1(n) \cdot P_2(n)$ we have $a(n) = P_1(n) + P_2(n)^2/(P_1(n) + P_2(n)) + o(1)$ as $n \to \infty$. Note that the set of integers on which this condition holds both contains and has density strictly greater than the set where $P_2(n)^2 < P_1(n)$.

https://doi.org/10.1112/S0025579317000109 Published online by Cambridge University Press

It similarly follows that for each fixed k there exists a set of integers on which the value of $a(n)$ is approximated, up to an error term that is $o(1)$, by a function involving only by the largest k prime factors of n, where the density of these sets approaches one as k goes to infinity. However, the expression for $a(n)$ in terms of these prime factors becomes increasingly complicated.

§6. *The average time spent waiting after the largest prime factor.* We know that $P_1(n) \leqslant a(n) \leqslant B(n)$, with equality holding in each case if and only if n is prime. In light of Theorems 4.2 and 4.3 it makes sense to similarly consider the difference $a(n) - P_1(n)$, which also corresponds, by linearity of expectation, to the expected number of steps spent waiting after the random walk has picked up a factor of the largest prime divisor of n.

THEOREM 6.1. *There exist constants D_ℓ such that, for any fixed integer $N \geqslant 1$,*

$$\sum_{n\leqslant x}(a(n) - P_1(n)) = \frac{x^{3/2}}{\log^2 x}\sum_{\ell=0}^{N-1}\frac{D_\ell}{\log^\ell x} + O_N\left(\frac{x^{3/2}}{\log^{N+2} x}\right).$$

In particular, $D_0 = (8\zeta(3/2)/3)(1 - \pi/4)$, and an expression that can be used to compute D_ℓ for any value of ℓ is given in the proof (12) below.

Comparing this with Theorem 4.2 and the constant found by Balasubramanian (which can also be obtained by methods similar to the proof of Theorem 6.1) we can get a result about how $a(n)$ depends, in an average sense, on its largest two prime factors.

COROLLARY 6.2. *The difference between $a(n)$ and $P_1(n)$ satisfies*

$$\sum_{n\leqslant x}(a(n) - P_1(n)) \sim \left(1 - \frac{\pi}{4}\right)\sum_{n\leqslant x} P_2(n).$$

Proof of Theorem 6.1. Note first that if $n = p$ is prime, then $a(p) - P_1(p) = 0$, so the primes contribute nothing to this sum. Using p and q to denote prime numbers we can then write

$$\begin{aligned}\sum_{n\leqslant x}(a(n) - P_1(n)) &= \sum_{p<x}\sum_{q\leqslant p}\sum_{\substack{m\leqslant x/pq\\ P_1(m)\leqslant q}}(a(pqm) - p)\\ &= \sum_{p<x}\sum_{q\leqslant p}\sum_{\substack{m\leqslant x/pq\\ P_1(m)\leqslant q}}(a(pq) - p + O(a(m)))\end{aligned}$$

since $a(pq) \leqslant a(pqm) \leqslant a(pq) + a(m)$. We can bound the error term above as

$$\begin{aligned}\sum_{p<x}\sum_{q\leqslant p}\sum_{\substack{m\leqslant x/pq\\ P_1(m)\leqslant q}} a(m) &\leqslant \sum_{p<x}\sum_{q\leqslant p}\sum_{\substack{m\leqslant x/pq\\ P_1(m)\leqslant q}} B(m)\\ &= \sum_{n\leqslant x}{}'(B(n) - P_1(n) - P_2(n)) = O\left(\frac{x^{4/3}}{\log^3 x}\right)\end{aligned}$$

https://doi.org/10.1112/S0025579317000109 Published online by Cambridge University Press

using Theorem 4.2. Here $\sum'$ denotes that the sum is over composite integers. Dividing the main sum into two sums based on whether the largest prime divisor divides the integer twice we get that

$$\begin{aligned}\sum_{n\leqslant x}(a(n)-P_1(n)) &= \sum_{p\leqslant\sqrt{x}}\sum_{\substack{m\leqslant x/p^2\\ P_1(m)\leqslant p}} a(p^2)-p\\ &\quad+\sum_{p<x}\sum_{q<p}\sum_{\substack{m\leqslant x/pq\\ P_1(m)\leqslant q}} a(pq)-p+O(x^{4/3})\\ &= \sum_{p\leqslant\sqrt{x}}\sum_{\substack{m\leqslant x/p^2\\ P_1(m)\leqslant p}}(p-1)\\ &\quad+\sum_{p<x}\sum_{q<p}\sum_{\substack{m\leqslant x/pq\\ P_1(m)\leqslant q}}\frac{q(q-1)}{p+q-1}+O(x^{4/3}).\end{aligned}$$

The first sum above is

$$\sum_{p\leqslant\sqrt{x}}\sum_{\substack{m\leqslant x/p^2\\ P_1(m)\leqslant p}}(p-1)\leqslant\sum_{p\leqslant\sqrt{x}}p\left\lfloor\frac{x}{p^2}\right\rfloor\leqslant\sum_{p\leqslant\sqrt{x}}\frac{x}{p}=O(x\log\log x),$$

so it can be absorbed into the error term. This leaves the second sum, which we simplify and reorder as

$$\begin{aligned}&\sum_{p<x}\sum_{q<p}\sum_{\substack{m\leqslant x/pq\\ P_1(m)\leqslant q}}\frac{q(q-1)}{p+q-1}\\ &=\sum_{p<x}\sum_{q<p}\sum_{\substack{m\leqslant x/pq\\ P_1(m)\leqslant q}}\left(\frac{q^2}{p+q}-\frac{q}{p+q-1}+\frac{q^2}{(p+q)(p+q-1)}\right)\\ &=\sum_{p<x}\sum_{q<p}\sum_{\substack{m\leqslant x/pq\\ P_1(m)\leqslant q}}\frac{q^2}{p+q}+O(x)\\ &=\sum_{m\leqslant x/6}\sum_{\substack{p<x/m\\ p>P_1(m)}}\sum_{\substack{q<p\\ q\leqslant x/mp\\ q\geqslant P_1(m)}}\frac{q^2}{p+q}+O(x)\\ &=\sum_{m\leqslant x/6}\sum_{\substack{p\leqslant\sqrt{x/m}\\ p>P_1(m)}}\sum_{\substack{q<p\\ q\geqslant P_1(m)}}\frac{q^2}{p+q}+\sum_{m\leqslant x/6}\sum_{\substack{\sqrt{x/m}<p<x/m\\ p>P_1(m)}}\sum_{\substack{q\leqslant x/mp\\ q\geqslant P_1(m)}}\frac{q^2}{p+q}+O(x)\\ &=S_1+S_2+O(x).\end{aligned}\tag{3}$$

https://doi.org/10.1112/S0025579317000109 Published online by Cambridge University Press

We evaluate the two sums S_1 and S_2 separately. Taking first S_1, we find that the sum is dominated by small values of m. The contribution to S_1 by those $m \geqslant x^\epsilon$ for any fixed small $0 < \epsilon < 1/4$ is

$$\begin{aligned}\sum_{x^\epsilon \leqslant m \leqslant x/6} \sum_{\substack{p \leqslant \sqrt{x/m} \\ p > P_1(m)}} \sum_{\substack{q < p \\ q \geqslant P_1(m)}} \frac{q^2}{p+q} &\leqslant \sum_{x^\epsilon < m \leqslant x/6} \sum_{p \leqslant \sqrt{x/m}} \sum_{q<p} q \\ &\ll \sum_{x^\epsilon \leqslant m \leqslant x/6} \sum_{p \leqslant \sqrt{x/m}} \frac{p^2}{\log p} \\ &\ll \sum_{x^\epsilon \leqslant m \leqslant x/6} \left(\frac{x}{m}\right)^{3/2} = O(x^{(3-\epsilon)/2}),\end{aligned}$$

and likewise the contribution from either p or $q \leqslant x^\epsilon$ is insignificant, so we can write

$$S_1 = \sum_{m < x^\epsilon} \sum_{x^\epsilon < p \leqslant \sqrt{x/m}} \sum_{x^\epsilon < q < p} \frac{q^2}{p+q} + O(x^{(3-\epsilon)/2}). \tag{4}$$

We treat the inner sum using partial summation

$$\sum_{x^\epsilon < q < p} \frac{q^2}{p+q} = \int_{x^\epsilon}^{p} \frac{s^2}{(p+s)\log s}\, ds + O(p^2 \mathrm{e}^{-\sqrt{\log p}}).$$

We tackle the integral by expanding the fraction as a geometric series

$$\frac{s^2}{(p+s)\log s} = \frac{s^2}{p \log s} \sum_{i=0}^{\infty} \left(-\frac{s}{p}\right)^i$$

(which converges for all $s < p$) and integrate term by term,

$$\begin{aligned}\int_{x^\epsilon}^{p} \frac{s^2}{(p+s)\log s} ds &= \sum_{i=0}^{\infty} \left(\frac{(-1)^i}{p^{i+1}} \int_{x^\epsilon}^{p} \frac{s^{i+2}}{\log s}\, ds \right) \\ &= \sum_{i=0}^{\infty} \left(\frac{(-1)^i}{p^{i+1}} (\mathrm{li}(p^{i+3}) - \mathrm{li}(x^{\epsilon(i+3)})) \right) \\ &= \sum_{i=0}^{\infty} \left(\frac{(-1)^i}{p^{i+1}} \left(\frac{p^{i+3}}{\log p^{i+3}} \left(\sum_{k=0}^{N-1} \frac{k!}{\log^k p^{i+3}} \right.\right.\right. \\ &\quad \left.\left.\left. + O_N\left(\frac{1}{\log^N p^{i+3}}\right)\right) + O\left(\frac{x^{\epsilon(i+3)}}{\epsilon(i+3)\log x}\right)\right)\right) \\ &= \frac{p^2}{\log p} \sum_{i=0}^{\infty} \sum_{k=0}^{N-1} \left(\frac{(-1)^i k!}{(i+3)^{k+1} \log^k p} \right) + O_N\left(\frac{p^2}{\log^{N+1} p}\right).\end{aligned}$$

https://doi.org/10.1112/S0025579317000109 Published online by Cambridge University Press

Here we have used the asymptotic expansion for $\mathrm{li}(x)$, where $N \geqslant 0$ is a fixed integer. Plugging this into (4) and applying partial summation to the variable p, we obtain

$$\begin{aligned} S_1 &= \sum_{m<x^\epsilon}\Bigg(\sum_{x^\epsilon<p<\sqrt{x/m}}\Bigg(\frac{p^2}{\log p}\sum_{i=0}^{\infty}\sum_{k=0}^{N-1}\Bigg(\frac{(-1)^i k!}{(i+3)^{k+1}\log^k p}\Bigg) \\ &\quad + O_N\Bigg(\frac{p^2}{\log^{N+1} p}\Bigg)\Bigg)\Bigg) \\ &= \sum_{m<x^\epsilon}\Bigg(\int_{x^\epsilon}^{\sqrt{x/m}}\frac{t^2}{\log^2 t}\sum_{i=0}^{\infty}\sum_{k=0}^{N-1}\Bigg(\frac{(-1)^i k!}{(i+3)^{k+1}\log^k t}\Bigg)dt \\ &\quad + O_N\Bigg(\frac{(x/m)^{3/2}}{\log^{N+2}(x/m)}\Bigg)\Bigg). \end{aligned} \tag{5}$$

Integrating by parts one finds that

$$\begin{aligned} \int\frac{t^2}{\log^{k+2} t}\,dt &= \frac{3^{k+1}}{(k+1)!}\mathrm{li}(t^3) - t^3\sum_{h=0}^{k}\frac{3^h(k-h)!}{(k+1)!\log^{h+1} t} + C \\ &= t^3\sum_{h=k+1}^{N}\frac{3^{k-h}h!}{(k+1)!\log^{h+1} t} + O_N\Bigg(\frac{t^3}{\log^{N+2} t}\Bigg) \end{aligned}$$

since the sum in the first line above simply subtracts off the first $k+1$ terms in the asymptotic expansion of $(3^{k+1}/(k+1)!)\mathrm{li}(t^3)$. Using this to evaluate the integral in (5) we obtain

$$\sum_{i=0}^{\infty}\sum_{k=0}^{N-1}\Bigg(\frac{(-1)^i k!}{(i+3)^{k+1}}\Bigg(t^3\sum_{h=k+1}^{N}\frac{3^{k-h}h!}{(k+1)!\log^{h+1} t} + O_N\Bigg(\frac{t^3}{\log^{N+2} t}\Bigg)\Bigg)\Bigg)\Bigg|_{t=x^\epsilon}^{t=\sqrt{x/m}}.$$

The lower limit of integration can easily be absorbed into the error term, so we need only concern the upper limit. Evaluating this expression there, it becomes

$$\begin{aligned} &\sum_{i=0}^{\infty}\sum_{k=0}^{N-1}\Bigg(\frac{(-1)^i k!}{(i+3)^{k+1}}\Bigg(\frac{x}{m}\Bigg)^{3/2}\sum_{h=k+1}^{N}\frac{3^{k-h}h!2^{h+1}}{(k+1)!\log^{h+1}(x/m)}\Bigg) \\ &= 2\Bigg(\frac{x}{m}\Bigg)^{3/2}\sum_{i=0}^{\infty}(-1)^i\sum_{h=1}^{N}\frac{2^h h!}{\log^{h+1}(x/m)}\sum_{k=0}^{h-1}\frac{3^{k-h}}{(k+1)(i+3)^{k+1}} \\ &= 2\Bigg(\frac{x}{m}\Bigg)^{3/2}\sum_{h=0}^{N-1}\frac{2^{h+1}h!}{\log^{h+2}(x/m)}\sum_{k=0}^{h}\frac{3^{k-h-1}}{k+1}\sum_{i=3}^{\infty}\frac{(-1)^{i+1}}{i^{k+1}}. \end{aligned} \tag{6}$$

https://doi.org/10.1112/S0025579317000109 Published online by Cambridge University Press

We therefore get the estimate

$$S_1 = 2\sum_{m<x^\epsilon}\left(\left(\frac{x}{m}\right)^{3/2}\left(\sum_{h=0}^{N-1}\frac{2^{h+1}h!}{\log^{h+2}(x/m)}\sum_{k=0}^{h}\frac{3^{k-h-1}}{k+1}\sum_{i=3}^{\infty}\frac{(-1)^{i+1}}{i^{k+1}} + O_N\left(\frac{1}{\log^{N+2}(x/m)}\right)\right)\right).$$

We now treat S_2. As was the case with S_1, we can ignore the contribution from those $m > x^\epsilon$, $q \geqslant x^\epsilon$ or $p > x^{1-\epsilon}$, so

$$S_2 = \sum_{m<x^\epsilon}\sum_{\sqrt{x/m}<p<x^{1-\epsilon}}\sum_{x^\epsilon<q<x/mp}\frac{q^2}{p+q} + O(x^{(3-\epsilon)/2}). \tag{7}$$

We also treat the innermost sum similarly to S_1, resulting in an integration similar to (5) but with different limits of integration:

$$\begin{aligned}\sum_{x^\epsilon<q<x/mp}\frac{q^2}{p+q} &= \int_{x^\epsilon}^{x/mp}\frac{s^2}{(p+s)\log s}\,ds = \sum_{i=0}^{\infty}\left(\frac{(-1)^i}{p^{i+1}}\int_{x^\epsilon}^{x/mp}\frac{s^{i+2}}{\log s}\,ds\right)\\ &= \sum_{i=0}^{\infty}\left(\frac{(-1)^i}{p^{i+1}}\left(\mathrm{li}\left(\left(\frac{x}{mp}\right)^{i+3}\right) - \mathrm{li}(x^{\epsilon(i+3)})\right)\right)\\ &= \sum_{i=0}^{\infty}\left(\frac{(-1)^i}{p^{i+1}}\left(\mathrm{li}\left(\left(\frac{x}{mp}\right)^{i+3}\right)\right)\right) + O\left(\frac{x^{3\epsilon}}{p}\right).\end{aligned}$$

Expanding the main term above using the asymptotic expansion for the logarithmic integral we obtain

$$\begin{aligned}&\sum_{i=0}^{\infty}\left(\frac{(-1)^i}{p^{i+1}}\left(\sum_{k=0}^{N-1}\frac{x^{i+3}k!}{(mp)^{i+3}\log((x/mp)^{i+3})^{k+1}}\right)\right)\\ &\quad + O_N\left(\sum_{i=0}^{\infty}\frac{(-1)^i x^{i+3}}{p^{i+1}(mp)^{i+3}\log^{N+1}(x/mp)^{i+3}}\right)\\ &= p^2\sum_{i=0}^{\infty}\left(\frac{x}{mp^2}\right)^{i+3}\\ &\quad\times\sum_{k=0}^{N-1}\left(\frac{(-1)^i k!}{(i+3)^{k+1}\log^{k+1}(x/m)(1-\log p/\log(x/m))^{k+1}}\right)\\ &\quad + O_N\left(\frac{x^3}{m^3p^4\log^{N+1}(x/mp)}\right)\\ &= p^2\sum_{i=3}^{\infty}\left(\frac{x}{mp^2}\right)^{i}\sum_{k=0}^{N-1}\left(\frac{(-1)^{i+1}k!}{i^{k+1}\log^{k+1}(x/m)}\sum_{j=0}^{\infty}\left(\binom{k+j}{j}\left(\frac{\log p}{\log(x/m)}\right)^{j}\right)\right)\\ &\quad + O_N\left(\frac{(x/m)^3}{p^4\log^{N+1}(x/mp)}\right).\end{aligned} \tag{8}$$

https://doi.org/10.1112/S0025579317000109 Published online by Cambridge University Press

Here we have used the series expansion $(1/(1-x))^{k+1} = \sum_{j=0}^{\infty}\binom{k+j}{j}x^j$. Inserting this into (7), and applying partial summation to p we get that

$$\begin{aligned}S_2 &= \sum_{m<x^\epsilon}\left(\int_{\sqrt{x/m}}^{x^{1-\epsilon}}\frac{t^2}{\log t}\sum_{i=3}^{\infty}\left(\frac{x}{mt^2}\right)^i\sum_{k=0}^{N-1}\frac{(-1)^{i+1}k!}{i^{k+1}\log^{k+1}(x/m)}\right.\\ &\quad\left.\times\sum_{j=0}^{\infty}\left(\binom{k+j}{j}\left(\frac{\log t}{\log(x/m)}\right)^j\right)dt + O_N\left(\frac{(x/m)^{3/2}}{\log^{N+2}(x/m)}\right)\right)\\ &= \sum_{m<x^\epsilon}\left(\sum_{i=3}^{\infty}\left((-1)^{i+1}\left(\frac{x}{m}\right)^i\sum_{k=0}^{N-1}\left(\frac{(-1)^{i+1}k!}{i^{k+1}\log^{k+1}(x/m)}\int_{\sqrt{x/m}}^{x^{1-\epsilon}}\frac{1}{t^{2i-2}\log t}\right.\right.\right.\\ &\quad\left.\left.\left.\times\sum_{j=0}^{\infty}\left(\binom{k+j}{j}\left(\frac{\log t}{\log(x/m)}\right)^j\right)dt\right)\right)\right.\\ &\quad\left.+ O_N\left(\frac{(x/m)^{3/2}}{\log^{N+2}(x/m)}\right)\right).\end{aligned} \tag{9}$$

Using the fact that

$$\int\frac{\log^{j-1}p}{t^{2i-2}}\,dt = \begin{cases}\operatorname{li}(t^{-2i+3}) + C & \text{if } j=0,\\ -\dfrac{\log^{j-1}t}{(2i-3)t^{2i-3}}\displaystyle\sum_{l=0}^{j-1}\frac{j!}{(j-l)!(2i-3)^l\log^l t} + C & \text{if } j\geqslant 1,\end{cases} \tag{10}$$

the integral in the expression above becomes

$$\left(\operatorname{li}\left(\frac{1}{t^{2i-3}}\right) - \frac{1}{t^{2i-3}}\sum_{j=1}^{\infty}\left(\binom{k+j}{j}\left(\frac{1}{\log(x/m)}\right)^j\right.\right.$$
$$\left.\left.\times\sum_{l=0}^{j-1}\frac{j!\log^{j-1}t}{(j-l)!(2i-3)^{l+1}\log^l t}\right)\right)\Bigg|_{t=\sqrt{x/m}}^{t=x^{1-\epsilon}}.$$

When evaluated at $t = x^{1-\epsilon}$, the above expression can be absorbed into the existing error term, so we need only evaluate this expression when $t = \sqrt{x/m}$, in which case it becomes

$$\begin{aligned}&-\left(\operatorname{li}\left(\left(\frac{x}{m}\right)^{-(2i-3)/2}\right) - \left(\frac{x}{m}\right)^{-(2i-3)/2}\sum_{j=1}^{\infty}\left(\binom{k+j}{j}\frac{1}{2^{j-1}\log(x/m)}\right.\right.\\ &\quad\left.\left.\times\sum_{l=0}^{j-1}\frac{j!2^l}{(j-l)!(2i-3)^{l+1}\log^l(x/m)}\right)\right)\\ &= -\left(\operatorname{li}\left(\left(\frac{x}{m}\right)^{-(2i-3)/2}\right) - \left(\frac{x}{m}\right)^{-(2i-3)/2}\sum_{l=0}^{\infty}\frac{2^{l+1}}{(2i-3)^{l+1}\log^{l+1}(x/m)}\right.\\ &\quad\left.\times\sum_{j=l+1}^{\infty}\left(\binom{k+j}{j}\frac{j!}{2^j(j-l)!}\right)\right).\end{aligned} \tag{11}$$

https://doi.org/10.1112/S0025579317000109 Published online by Cambridge University Press

The sum over j can be evaluated as $(2^{k+1}-2^{-l})(l+k)!/k!$, and the sum over l can be truncated at $N-k-1$, with the remainder of the sum being absorbed into the existing error term, $O_N((x/m)^{3/2}/\log^{N+2}(x/m))$. Use the asymptotic expansion of the logarithmic integral one more time, again truncating at $N-k-1$, turns the expression in (11) into

$$\left(\frac{x}{m}\right)^{-(2i-3)/2}\sum_{l=0}^{N-k-1}\left(\frac{2}{(2i-3)^{l+1}\log^{l+1}(x/m)}\right.$$
$$\left.\times\left(l!(-1)^l 2^l+\frac{(2^{l+k+1}-1)(l+k)!}{k!}\right)\right).$$

Inserting this back into (9) in place of the integral, and reindexing so that the summation runs over the exponent on the logarithm we get that the main term is

$$\sum_{m<x^\epsilon}\left(\frac{x}{m}\right)^{3/2}\sum_{i=3}^{\infty}\sum_{h=0}^{N-1}\left(\frac{(-1)^{i+1}}{\log^{h+2}(x/m)}\sum_{k=0}^{h}\left(\frac{2}{i^{k+1}(2i-3)^{h-k+1}}\right.\right.$$
$$\left.\left.\times(k!(h-k)!(-1)^{h-k}2^{h-k}+(2^{h+1}-1)h!)\right)\right).$$

Combining that with the corresponding estimate (6), we have that the entire main term of the sum that we are interested in (3) is equal to

$$\sum_{m\leqslant x/6}\left(\frac{x}{m}\right)^{3/2}\sum_{h=0}^{N-1}\frac{1}{\log^{h+2}(x/m)}\sum_{k=0}^{h}\sum_{i=3}^{\infty}\frac{2(-1)^{i+1}}{i^{k+1}}$$
$$\times\left(\frac{2^{h+1}3^{k-h-1}h!}{(k+1)}+\frac{(k!(h-k)!(-2)^{h-k}+(2^{h+1}-1)h!)}{(2i-3)^{h+1-k}}\right)$$

with an error term of $O_N((x/m)^{3/2}/\log^{N+2}(x/m))$. Note that the quantity given by the double summation over k and i is a constant depending only on h, which we call C_h. Thus,

$$\sum_{n\leqslant x}(a(n)-P_1(n))$$
$$=x^{3/2}\sum_{m\leqslant x^\epsilon}\left(\frac{1}{m^{3/2}}\sum_{h=0}^{N-1}\frac{C_h}{\log^{h+2}(x/m)}+O_N\left(\frac{1}{m^{3/2}\log^{N+2}(x/m)}\right)\right)$$
$$=x^{3/2}\sum_{h=0}^{N-1}C_h\sum_{m\leqslant x^\epsilon}\frac{1}{m^{3/2}}\frac{1}{\log^{h+2}(x/m)}+O_N\left(\frac{x^{3/2}}{\log^{N+2}x}\right).$$

Since

$$\sum_{m\leqslant x^\epsilon}\frac{1}{m^{3/2}\log^{h+2}(x/m)}=\sum_{m\leqslant x^\epsilon}\frac{1}{m^{3/2}\log^{h+2}x}\left(\frac{1}{1-\log m/\log x}\right)^{h+2}$$
$$=\sum_{m\leqslant x^\epsilon}\frac{1}{m^{3/2}\log^{h+2}x}\sum_{j=0}^{\infty}\binom{h+1+j}{j}\left(\frac{\log m}{\log x}\right)^j$$

https://doi.org/10.1112/S0025579317000109 Published online by Cambridge University Press

$$= \sum_{j=0}^{\infty} \frac{\binom{h+j+1}{j}}{\log^{h+j+2} x} \sum_{m \leqslant x^{\epsilon}} \frac{\log^j m}{m^{3/2}}$$
$$= \sum_{j=0}^{N-h-1} \frac{\binom{h+j+1}{j}}{\log^{h+j+2} x} (-1)^j \zeta^{(j)}\left(\frac{3}{2}\right) + O_N\left(\frac{1}{\log^{N+2} x}\right),$$

where $\zeta^{(j)}(3/2)$ denotes the jth derivative of the zeta function at $3/2$, we can conclude that

$$\sum_{n \leqslant x} (a(n) - P_1(n)) = x^{3/2} \sum_{h=0}^{N-1} C_h \sum_{j=0}^{N-h-1} \frac{\binom{h+j+1}{j}}{\log^{h+j+2} x} (-1)^j \zeta^{(j)}\left(\frac{3}{2}\right) + O_N\left(\frac{x^{3/2}}{\log^{N+2} x}\right)$$
$$= x^{3/2} \sum_{\ell=0}^{N-1} \frac{1}{\log^{\ell+2} x} \sum_{h=0}^{\ell} \binom{\ell+1}{\ell-h} C_h (-1)^j \zeta^{(\ell-h)}\left(\frac{3}{2}\right) + O_N\left(\frac{x^{3/2}}{\log^{N+2} x}\right)$$
$$= x^{3/2} \sum_{\ell=0}^{N-1} \frac{D_\ell}{\log^{\ell+2} x} + O_N\left(\frac{x^{3/2}}{\log^{N+2} x}\right)$$

where

$$D_\ell = 2 \sum_{h=0}^{\ell} \binom{\ell+1}{\ell-h} \zeta^{(\ell-h)}\left(\frac{3}{2}\right) \sum_{k=0}^{h} \sum_{i=3}^{\infty} \frac{(-1)^{\ell-h+i+1}}{i^{k+1}} \times \left(\frac{2^{h+1} h!}{3^{h-k+1}(k+1)} + \frac{(k!(h-k)!(-2)^{h-k} + (2^{h+1}-1)h!)}{(2i-3)^{h+1-k}} \right). \quad (12)$$

In the case $\ell = 0$ this reduces to

$$D_0 = 2\zeta\left(\frac{3}{2}\right) \sum_{i=3}^{\infty} \frac{(-1)^{i+1}}{i} \left(\frac{2}{3} + \frac{2}{2i-3}\right)$$
$$= 2\zeta\left(\frac{3}{2}\right) \sum_{i=3}^{\infty} (-1)^{i+1} \left(\frac{2}{3i} + \left(-\frac{2}{3i} + \frac{4}{3(2i-3)}\right)\right)$$
$$= \frac{8\zeta(3/2)}{3} \sum_{i=3}^{\infty} \frac{(-1)^{i+1}}{2i-3} = \frac{8\zeta(3/2)}{3}\left(1 - \frac{\pi}{4}\right). \qquad \square$$

§7. *Open questions and acknowledgements.* There are many remaining questions and directions that this research could be further pursued. A few examples of potentially interesting questions are given below.

https://doi.org/10.1112/S0025579317000109 Published online by Cambridge University Press

Question 7.1. Is $a(n)$ ever an integer when n is not a prime power?

Computations so far have failed to find any examples of such an integer.

Question 7.2. How many distinct residues modulo n is this random walk expected to visit?

Question 7.3. What can be said about the variance of the time required for this random walk?

Question 7.4. How do these results change if the residues chosen randomly to multiply the current state of the walk are chosen without replacement?

Acknowledgements. A portion of this work appeared in my PhD Thesis [**15**], written under the direction of Carl Pomerance, whose helpful suggestions were invaluable in the development of this paper. I would also like to thank Sarah Wolff, whose Graduate Student Seminar talk on random walks on monoids inspired questions which led to the results in this paper, and both Peter Winkler and the anonymous referee for useful suggestions.

References

1. D. Aldous, Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on Probability, XVII* (*Lecture Notes in Mathematics* **986**), Springer (Berlin, 1983), 243–297; MR 770418.
2. K. Alladi and P. Erdős, On an additive arithmetic function. *Pacific J. Math.* **71**(2) (1977), 275–294; MR 0447086.
3. K. Alladi and P. Erdős, On the asymptotic behavior of large prime factors of integers. *Pacific J. Math.* **82**(2) (1979), 295–315; MR 551689.
4. J. M. De Koninck and A. Ivić, The distribution of the average prime divisor of an integer. *Arch. Math. (Basel)* **43**(1) (1984), 37–43; MR 758338.
5. P. Erdős and A. Ivić, Estimates for sums involving the largest prime factor of an integer and certain related additive functions. *Studia Sci. Math. Hungar.* **15**(1–3) (1980), 183–199; MR 681439.
6. P. Erdős, A. Ivić and C. Pomerance, On sums involving reciprocals of the largest prime factor of an integer. *Glas. Mat. Ser. III* **21(41)**(2) (1986), 283–300; MR 896810.
7. P. Erdős and C. Pomerance, On the largest prime factors of n and $n + 1$. *Aequationes Math.* **17**(2–3) (1978), 311–321; MR 0480303.
8. D. Gretete, Random walk on a finitely generated monoid. *Int. J. Math. Combinatorics* **4** (2011), 54–58; (electronic).
9. M. Hildebrand, A survey of results on random walks on finite groups. *Probab. Surv.* **2** (2005), 33–63; MR 2121795.
10. A. Ivić, On the kth prime factor of an integer. *Zb. Rad. Prirod.-Mat. Fak. Ser. Mat.* **20**(1) (1990), 63–73; MR 1158406.
11. D. R. Jeske and T. Blessinger, Tunable approximations for the mean and variance of the maximum of heterogeneous geometrically distributed random variables. *Amer. Statist.* **58**(4) (2004), 322–327; MR 2109423.
12. J. Kemeny, Largest prime factor. *J. Pure Appl. Algebra* **89**(1–2) (1993), 181–186.
13. D. E. Knuth and L. Trabb Pardo, Analysis of a simple factorization algorithm. *Theoret. Comput. Sci.* **3**(3) (1976/77), 321–348; MR 0498355.
14. J. Mairesse, Random walks on groups and monoids with a Markovian harmonic measure. *Electron. J. Probab.* **10** (2005), 1417–1441 (electronic); MR 2191634.
15. N. McNew, Multiplicative problems in combinatorial number theory. *PhD Thesis*, Dartmouth College, 2015.

https://doi.org/10.1112/S0025579317000109 Published online by Cambridge University Press

16. E. Naslund, The average largest prime factor. *Integers* **13** (2013); Paper No. A81, 5.
17. F. S. Wheeler, Two differential-difference equations arising in number theory. *Trans. Amer. Math. Soc.* **318**(2) (1990), 491–523; MR 963247.

Nathan McNew,
Department of Mathematics,
Towson University,
8000 York Road,
Towson, MD 21252,
U.S.A.
E-mail: nmcnew@towson.edu

https://doi.org/10.1112/S0025579317000109 Published online by Cambridge University Press