The Expected Value of the Length of a Random Divisor

Saurabh Pandey

1. Introduction and Preliminaries

Having an understanding of, and knowing how to manipulate, sequences is essential in order to obtain a practical and thorough knowledge of mathematics. In this paper, we shall explore a particular probabilistic sequence called the *random divisor sequence*.

Definition 1. A *random divisor sequence*, or an *RDS*, is a sequence of positive integers $\{a_1, a_2 \cdots, a_{k-1}, a_k = 1\}$ with first term a_1 and last term $a_k = 1$ that satisfies the following conditions:

1. a_1 is a positive integer.
2. a_{m+1} is a randomly chosen positive divisor of a_m for every positive integer m, given that $a_m \neq 1$.
3. If for some positive integer k, some term $a_k = 1$, then the RDS terminates at a_k (There will be no other terms after a_k).

This sequence, never previously formally defined, is a generalization of a sequence introduced in USAMTS 2/3/17 [1] by Matthew Crawford.

Example 1. Possible RDSs include $\{1\}, \{125, 5, 5, 1\}$, and $\{24, 24, 6, 2, 1\}$.

It is clear from the definition of an RDS that, given a fixed first term, some RDSs are more likely to occur than others. For example, consider the RDS that begins with 2. The next term following a 2 has a $\frac{1}{2}$ chance of being 2 and a $\frac{1}{2}$ chance of being 1. Since any RDS stops when a term becomes 1, given the fixed first term 2, the RDS $\{2, 1\}$ has a $\frac{1}{2}$ probability of occurring, whereas $\{2, 2, 2, 2, 2, 2, 2, 1\}$ has a $\frac{1}{2^7} = \frac{1}{128}$ probability of occurrence. So, among RDSs starting with 2, $\{2, 1\}$ is more likely to occur than $\{2, 2, 2, 2, 2, 2, 2, 1\}$. This understanding of probability is important in understanding the expected value of the length of a sequence given a fixed first term. We will thus define three terms that rigorously develop such concepts.

Definition 2. The *probability of occurrence* of an RDS $\{a_1, a_2 \cdots, a_{k-1}, a_k = 1\}$ is defined as the probability that given a fixed first term a_1, the RDS $\{a_1, a_2 \cdots, a_{k-1}, a_k = 1\}$ will appear.

Definition 3. The *length* of an RDS $\{a_1, a_2 \cdots, a_{k-1}, a_k = 1\}$ is k, the number of terms in the sequence.

Definition 4. Let the set of possible probabilities of occurrences of an RDS with fixed first term n be $p_1, p_2, p_3 \cdots$ with $L_1, L_2, L_3 \cdots$ being the corresponding lengths of those sequences. $E(n)$ is then defined as the *expected value of the length of the sequence starting with* n, that is: $E(n) = \sum_i L_i p_i$.

We now present a recursive formula for finding $E(n)$ for any positive integer n greater than 1. This recursive formula will then be used to derive a more computationally efficient explicit formula for $E(p^\alpha)$, where p is a prime and α is a nonnegative integer.

2. A General Recursive Algorithm for finding $E(n)$

Let us examine a random divisor sequence starting with first term n. We now present a formula in Theorem 1 below that avoids the cumbersome definition of $E(n)$ above and forms a clever way of writing $E(n)$ recursively. This formula was previously briefly discussed in [2], but we will now give a more general form of it. Our proof of this formula below is also significantly more rigorous than the proof found in [2].

Theorem 1. Let n be a positive integer greater than 1. Then

$$E(n) = 1 + \frac{1}{\tau(n)} \sum_{d|n} E(d),$$

where $\tau(n)$ represents the familiar number theoretic function of the number of positive divisors of n, and we assume d is positive.

Proof: Suppose $d_1 = 1, d_2 \cdots d_{\tau(n)-1}, d_{\tau(n)} = n$ are the $\tau(n)$ positive divisors of n in increasing order ($d_k < d_{k+1}$ for every nonnegative integer k) with corresponding expected values $E(d_1), E(d_2), \ldots, E(d_{\tau(n)-1}), E(d_{\tau(n)})$. Now after the first term n, there will be a second term m, since the first term is not 1 and the RDS will thus not terminate (thus giving rise to the condition $n > 1$). Let the RDSs beginning with the fixed first term m have probabilities of occurrence $q_1, q_2, q_3 \ldots$ with $J_1, J_2, J_3 \ldots$ being the corresponding lengths of these RDSs. The corresponding RDSs beginning with the fixed first term n and fixed second term m thus have probabilities of occurrence $q_1, q_2, q_3, \ldots$ with $J_1 + 1, J_2 + 1, J_3 + 1 \ldots$ being the corresponding lengths of those sequences, because the addition of the extra term (the first term: n) only changes the length of the RDS, not the probability of occurrence of the RDS, because in our case, the probability of occurrence is solely dependent on m. So, by Definition 4,

$$E(m) = \sum_i J_i q_i.$$

We will now write $E(m)$ differently. We know that there is a $\frac{1}{\tau(n)}$ probability that $m = d_1$; a $\frac{1}{\tau(n)}$ probability that $m = d_2, \ldots$; a $\frac{1}{\tau(n)}$ probability that $m = d_{\tau(n)-1}$; and a $\frac{1}{\tau(n)}$ probability that $m = d_{\tau(n)}$, because the second term is chosen randomly among the positive divisors of n. Since m will be one of $d_1, d_2, \ldots d_{\tau(n)-1}, d_{\tau(n)}$ with probability $\frac{1}{\tau(n)}$, $E(m)$ will be the average of $E(d_1), E(d_2) \cdots E(d_{\tau(n)-1}), E(d_{\tau(n)})$, or

$$E(m) = \frac{1}{\tau(n)} \sum_{i=1}^{\tau(n)} E(d_i) = \frac{1}{\tau(n)} \sum_{d|n} E(d).$$

The second equality holds since $d_1, d_2 \cdots d_{\tau(n)-1}, d_{\tau(n)}$ are the positive divisors of n.

Again, by Definition 4,

$$E(n) = \sum_i (J_i + 1)(q_i) = \sum_i (J_i q_i + q_i) = \sum_i J_i q_i + \sum_i q_i = E(m) + \sum_i q_i.$$

But

$$E(m) = \frac{1}{\tau(n)} \sum_{d|n} E(d) \text{ and } \sum_i q_i = 1,$$

because the sum of all probabilities of occurrence of a RDS sequence with first term m is equal to 1. Thus, substituting back into the expression for $E(n)$, we have

$$E(n) = 1 + \frac{1}{\tau(n)} \sum_{d|n} E(d).$$

This is what we wanted to prove. ■

Using this formula, we can easily obtain $E(n)$ recursively. To show this, we let the prime factorization of $n = \prod_{i=1}^{k} {p_i}^{\alpha_i}$, with $p_1, p_2 \ldots p_k$ primes and $\alpha_1, \alpha_2, \ldots \alpha_k$ nonnegative integers. Note that $E(1) = 1$, as an RDS terminates when it has a term equal to 1. We can thus solve for $E(p_i), 1 \le i \le k$, since $E(p_i) = 1 + \frac{1}{2}(E(p_i) + E(1)) = 1 + \frac{1}{2}(E(p_i) + 1)$, which can be solved for $E(p_i)$. If we can solve for $E(p_i)$, we can also solve recursively for a $E(d)$, where d is a positive divisor of n that is only a product of primes. Continuing on in the same manner, we can find $E(n)$ recursively. We present the following example:

Example 2. Find $E(75)$.

Solution: $E(75) = 1 + \frac{1}{6}(E(1) + E(3) + E(5) + E(15) + E(25) + E(75))$.
$E(1) = 1$.
$E(3) = 1 + \frac{1}{2}(E(1) + E(3)) = 1 + \frac{1}{2}(E(3) + 1) \Rightarrow E(3) = 3$.
$E(5) = 1 + \frac{1}{2}(E(1) + E(5)) = 1 + \frac{1}{2}(E(5) + 1) \Rightarrow E(5) = 3$.

$E(15) = 1 + \frac{1}{4}(E(1) + E(3) + E(5) + E(15)) \Rightarrow E(15) = \frac{11}{3}$.
$E(25) = 1 + \frac{1}{3}(E(1) + E(5) + E(25)) = 1 + \frac{1}{3}(1 + 3 + E(25)) \Rightarrow E(25) = \frac{7}{2}$.

Thus $E(75) = 1 + \frac{1}{6}(E(1) + E(3) + E(5) + E(15) + E(25) + E(75)) =$
$= 1 + \frac{1}{6}(1 + 3 + 3 + \frac{11}{3} + \frac{7}{2} + E(75)) \Rightarrow E(75) = \frac{121}{30}$.

However, using the formula in Theorem 1 to find $E(n)$ is not very computationally efficient. As it is recursive, we would have to find

$$E(d_2), E(d_3) \cdots E(d_{\tau(n)-1}), E(d_{\tau(n)})$$

in order to compute $E(n)$ (the value of $E(d_1) = E(1) = 1$ is known already). Thus, we must perform $\tau(n) - 1$ computations to find $E(n)$. To help begin addressing the shortcoming, we develop a computationally efficient explicit formula for the fundamental case $E(p^\alpha)$, where p is a prime and α is a nonzero integer.

3. An Explicit Formula for $E(p^\alpha)$

We now concern ourselves with finding $E(n)$ for a most fundamental case of numbers: powers of primes. The following formerly unproven result is useful for its computational efficiency.

Theorem 2. Let p be a prime and α be an integer greater than 0. Then

$$E(p^\alpha) = \begin{cases} 1 & \text{if } \alpha = 0 \\ 2 + \displaystyle\sum_{i=1}^{\alpha} \frac{1}{i} & \text{if } \alpha > 0 \end{cases}$$

Proof:

Case 1: $\alpha = 0$. If $\alpha = 0$, then $E(p^\alpha) = E(p^0) = E(1) = 1$ by the work above, which proves the first case.

Case 2: $\alpha = 1$. For $\alpha = 1$, we will apply our recursive formula for $E(n)$. Note that $\tau(p) = 2$, since p has only the positive divisors 1 and p.

Thus, $E(p) = 1 + \dfrac{1}{2}(E(1) + E(p)), \;\Rightarrow \dfrac{E(p)}{2} = 1 + \dfrac{E(1)}{2}, \;\Rightarrow$
$E(p) = 2 + E(1) = 3 = 2 + \displaystyle\sum_{i=1}^{1} \frac{1}{i}$, as desired.

Case 3: $\alpha > 1$. By the previous section, we know that

$$E(n) = 1 + \frac{1}{\tau(n)} \sum_{d|n} E(d).$$

Consider the number p^α, with p a prime and α a positive integer. It has the positive divisors $1, p, p^2, ...p^\alpha$, so $\tau(p^\alpha) = \alpha + 1$. Using our formula for $E(n)$ developed in Theorem 1, we see that, for a prime p and $\alpha > 0$,

$$E(p^\alpha) = 1 + \frac{1}{\alpha+1}(1 + E(p) + E(p^2) + ...E(p^{\alpha-2}) + E(p^{\alpha-1}) + E(p^\alpha)).$$

Isolating the $E(p^\alpha)$ term on the LHS, it becomes quickly clear that

$$E(p^\alpha) = 1 + \frac{E(p^\alpha)}{\alpha+1} + \frac{1}{\alpha+1}(1 + E(p) + E(p^2) + ... + E(p^{\alpha-2}) + E(p^{\alpha-1})).$$

$$\Rightarrow \frac{(\alpha)(E(p^\alpha))}{\alpha+1} = 1 + \frac{1}{\alpha+1}(1 + E(p) + E(p^2) + ... + E(p^{\alpha-2}) + E(p^{\alpha-1}))$$

$$\Rightarrow E(p^\alpha) = \frac{1}{\alpha} + 1 + \frac{1}{\alpha}(1 + E(p) + E(p^2) + ... + E(p^{\alpha-2}) + E(p^{\alpha-1}))$$

But

$$E(p^{\alpha-1}) = 1 + \frac{1}{\alpha}(1 + E(p) + E(p^2) + ... + E(p^{\alpha-2}) + E(p^{\alpha-1})),$$

because the positive divisors of $p^{\alpha-1}$ are $1, p, p^2, ..., p^{\alpha-1}$ and $\tau(p) = \alpha$. However, this expression for $E(p^{\alpha-1})$ is defined only for $\alpha - 1 > 0$ or $\alpha > 1$, since if $\alpha < 1$, $p^{\alpha-1}$ would be nonintegral, and so $E(p^{\alpha-1})$ would be a nonsensical expression.

If $\alpha = 1$, then

$$E(p^{\alpha-1}) = E(1) = 1 \neq 1 + \frac{1}{\alpha}(1 + E(p) + E(p^2) + ... + E(p^{\alpha-2}) + E(p^{\alpha-1})).$$

Thus, $\alpha > 1$. Substituting our value of $E(p^{\alpha-1})$ into our expression for $E(p^\alpha)$, we see that for $\alpha > 1$, $E(p^\alpha) = E(p^{\alpha-1}) + \frac{1}{\alpha}$.

This is clearly a cleaner recursion than the one previously given. However, we still have not obtained an explicit formula. But if we actually apply our recursive formula, we see the following pattern for $\alpha > 1$:

$E(p^2) = E(p) + \frac{1}{2} = 3 + \frac{1}{2}$

$E(p^3) = E(p^2) + \frac{1}{3} = 3 + \frac{1}{2} + \frac{1}{3}$

$E(p^4) = E(p^3) + \frac{1}{4} = 3 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}$

...

$E(p^\alpha) = 3 + \sum_{i=2}^{\alpha} \frac{1}{i} = 2 + \sum_{i=1}^{\alpha} \frac{1}{i}$, by shifting indices.

To rigorously prove this assertion of the value of $E(p^\alpha)$, we turn to induction. The following proof of this lemma is essentially a formalization of the pattern identified above.

Lemma. If $E(p) = 3$ and $E(p^\alpha) = E(p^{\alpha-1}) + \frac{1}{\alpha}$, then

$$E(p^\alpha) = 2 + \sum_{i=1}^{\alpha} \frac{1}{i},$$

for $\alpha > 1$.

Proof: For $\alpha = 2$, $E(p^2) = E(p) + \frac{1}{2} = 3 + \frac{1}{2} = 2 + \sum_{i=1}^{2} \frac{1}{i}$, as desired.

Now suppose that for $\alpha - 1 > 1, E(p^{\alpha-1}) = 2 + \sum_{i=1}^{\alpha-1} \frac{1}{i}$.

Then, $E(p^\alpha) = E(p^{\alpha-1}) + \frac{1}{\alpha}$

$= 2 + \sum_{i=1}^{\alpha-1} \frac{1}{i} + \frac{1}{\alpha}$ (by our inductive hypothesis that $E(p^{\alpha-1}) = 2 + \sum_{i=1}^{\alpha-1} \frac{1}{i}$.)

$= 2 + \sum_{i=1}^{\alpha} \frac{1}{i}$, as desired.

Thus, by the Principle of Mathematical Induction, the proof of this Lemma is complete.

By transitivity, we have also proven Case 3. The completion of the proof of this final case also completes the proof of Theorem 2. ■

It is interesting to note that as α monotonically increases without bound, the difference between successive terms of the sequence $E(p^\alpha)$ monotonically decreases, yet $E(p^\alpha)$ will never converge. This, of course, is a simple consequence of the application of the integral test on the harmonic series. Whether $E(\prod_{i=1}^{k} {p_i}^{\alpha_i})$, with p_i prime and α_i a nonnegative integer for each i such that $1 \le i \le k$, converges or diverges as $a_i \to \infty$ is an open problem, however.

4. Conclusion

In this paper, we first defined the previously undefined *random divisor sequence*. We found a general formula that can be used to solve recursively for $E(n)$ for integral $n > 1$. Then we found an explicit formula for $E(p^\alpha)$, where p is a prime and α is a nonnegative integer, in terms of α, greatly improving the computational efficiency in finding $E(p^\alpha)$. The techniques used in finding an explicit formula for $E(p^\alpha)$ might also help in finding an explicit formula for $E(n) = E(\prod_{i=1}^{k} {p_i}^{\alpha_i})$, where $\prod_{i=1}^{k} {p_i}^{\alpha_i}$ is the prime factorization of n, in terms of $\alpha_1, \alpha_2, \cdots, \alpha_k$. Such explicit formulas currently appear difficult to generate because of the complexity of the recursion used in Theorem 1.

5. References

1. Crawford, Matthew. (2005, December 11). United States of America Mathematical Talent Search (USAMTS) Round 3 Problems, Year 17

   http://www.usamts.org/Tests/USAMTSProblems_17_3.pdf

2. Sato, Naoki. (2006, February 19). United States of America Mathematical Talent Search Solutions to Problem 2/3/17

   http://www.usamts.org/Solutions/Solution2_3_17.pdf

Saurabh Pandey, Stoney Creek High School, Rochester, M.I.