ON A DEGREE OF PRIMALITY

JOSEPH PE

Abstract. In this paper we introduce a degree of primality for natural numbers, and hence, a measure of primality for intervals of consecutive numbers. We characterize maximally prime intervals of length ≤ 3 and their primalities. Maximally prime intervals of length 2 are those that contain Mersenne or Fermat primes; maximally prime intervals of length 3 are, but for a few exceptions, those whose midpoints are Dan numbers. There are relatively few maxprimes for larger lengths. We present a heuristic argument for an asymptotic form describing the distribution of maximal primalities. Finally, we mention open problems and directions for further research.

1. How Prime Is It?

Prime numbers are, of course, 'purely' prime, but can we ask meaningfully–and interestingly–how prime a composite number is? We would like to devise a degree or measure $\rho(n)$ of a number's primality. Such a scheme should assign the same maximal primality degree to prime numbers and allow for partial primality for composite numbers. (By 'number' in this paper, we mean natural number, unless otherwise specified.)

As a first approach, let's look at the number of divisors $d(n)$ of $n > 1$. Prime numbers p have the least number of divisors, with $d(p) = 2$, and numbers with few divisors would be considered 'nearly prime'. We could define a primality measure or degree $\kappa(n)$ for $n > 1$ by: $\kappa(n) = 2/d(n)$. Then $\kappa(p) = 1$ for prime p and $\kappa(n) \leq 1$ for all $n > 1$. The trouble with this approach is that it is biased against larger numbers, which tend to have more divisors than smaller ones, yet primality should not depend on a number's size. For example, $d(4) = 3$, $d(77) = 4$, hence, $\kappa(4) = 2/3 > 1/2 = \kappa(77)$, and 4 would be 'more prime' than 77 although more than half of the numbers less than 4 are divisors of 4 yet only about 4% of the numbers less than 77 divide 77.

A more sensible measure of primality uses the Euler phi-function $\varphi(n)$, defined as the number of i with $1 \leq i \leq n$ that are relatively prime to n, that is, $\gcd(i, n) = 1$. Note that n is prime if and only if $\varphi(n) = n - 1$, which is the maximum value attainable by $\varphi(n)$ relative to n. Furthermore,

more divisors mean fewer numbers relatively prime to n and vice versa. Thus, $\varphi(n)$ indirectly measures primality.

Definition 1. *The primality degree (or simply, primality) $\rho(n)$ of n is defined by: $\rho(n) = \varphi(n)/(n-1)$, where the purpose of the denominator $n-1$ is to normalize the scale of n. For the 'degenerate' case $n = 1$, we define $\rho(1) = 0$ (since 1 is not considered prime).*

Hence for all n, we have $0 \leq \rho(n) \leq 1$ and $\rho(p) = 1$ for all primes p. Going back to the example of 4 and 77, we get $\varphi(4) = 2$, $\varphi(77) = \varphi(7 \times 11) = 60$, so that $\rho(4) = 2/3 < 60/76 = \rho(77)$. This agrees with our intuition that 4 is less prime than 77.

Later on it will be helpful to consider in proofs instead an alternative but related measure of primality.

Definition 2. $\rho^*(n) = \varphi(n)/n$.

$\rho^*(n)$ is easier to manipulate than $\rho(n)$ (since it is multiplicative) and $\rho(n)/\rho^*(n) \to 1$. Note that $\rho(n) = n/(n-1)\rho^*(n)$ for $n > 1$.

We can now talk about the primality of a finite set of numbers. This can be defined as the sum of the primality of its elements. We call an interval I_n of n numbers an n-interval. (Note that an interval is just a set of consecutive natural numbers.) Given n, how prime can an n-interval be?

Definition 3. *The primality $\rho(I)$ of an n-interval $I = \{a, a+1, \ldots, a+n-1\}$ is defined by $\rho(I) = \rho(a) + \rho(a+1) + \cdots + \rho(a+n-1)$. For a fixed n, the sequence $E_n = \{[1,n], [2,n+1], [3,n+2], \ldots\}$ enumerates all the n-intervals. The corresponding sequence of primalities P_n is obtained by applying ρ to each term of E_n.*

P_n is bounded (from below by 0 and from above by n), and so the least upper bound or supremum of P_n exists. However, as some initial terms of P_n are often larger than the rest, it is more interesting to use a different kind of upper bound. For example, in the sequence $P_3 = \{2, 2.6667, 2.0667, 2.4, 1.9714, 2.3214, 1.7659, 2.1944, \ldots\}$ (to four decimal places), we would like to consider the largest limiting value of a subsequence of P_3 rather than the upper bound 2.6667. In other words, we do not want an 'isolated' upper bound. With this aim in mind, we define the following measure of maximal primality of an n-interval.

Definition 4. *The n-maxdegree, denoted by S_n, is defined by $S_n = \limsup P_n$, i.e., the limit superior or largest cluster point of P_n. (Recall that a cluster point of a sequence s is the limit of a convergent subsequence of s.)*

In the previous example, it turns out that $\limsup P_3 = 2.3333\ldots = 7/3$ (see Theorem 1 below).

2. Maxprimes

We now focus on intervals that are the 'most prime' among intervals of the same length. Intuition would suggest that these can be obtained by including as many prime numbers and as few divisors of composite numbers as possible. When $n \leq 3$, it is possible to give an exhaustive list of maximally prime intervals. For larger n the analysis is more difficult and we are able to find only few such intervals.

Definition 5. *Call an n-interval I n-maximally prime or an n-maxprime (maxprime for short) if $\rho(I) \geq S_n$. (Maxprimes generalize the usual primes in the sense that the 1-maxprimes are precisely the 1-intervals $I = \{p\}$, where p is prime.)*

Lemma 1. *If n is a multiple of $M > 1$, then*

$$\frac{\varphi(n)}{n} \leq \prod_{p|M} \left(1 - \frac{1}{p}\right)$$

where p runs through the prime factors of M. This implies, for example, that if n is a multiple of 2, then $\varphi(n) \leq n/2$; if n is a multiple of 3, then $\varphi(n) \leq 2n/3$; and if n is a multiple of 6, then $\varphi(n) \leq n/3$.

Proof. We use the following well-known theorem [1, Theorem 62].

$$\frac{\varphi(n)}{n} = \prod_{p|n} \left(1 - \frac{1}{p}\right). \tag{1}$$

The lemma follows immediately by noting that

$$\prod_{p|n} \left(1 - \frac{1}{p}\right) \leq \prod_{p|M} \left(1 - \frac{1}{p}\right) \tag{2}$$

because each factor on the right product of (2) occurs on the left product (since n is a multiple of M) and all the factors are between 0 and 1. □

Lemma 2.

(i) *If $\varphi(n) = n/2$, then n is a (positive) power of 2, and conversely.*
(ii) *If $\varphi(n) = n/3$, then $n = 2^a 3^b$ for some (positive) a and b, and conversely.*

Proof of (i). If $n = 2^a$, $a \geq 1$, then $\varphi(n)/n = (1-1/2) = 1/2$ by (1) above, hence, $\varphi(n) = n/2$. Conversely, let $\varphi(n) = n/2$. Then n is divisible by 2 and $\varphi(n)/n = 1/2 = (1 - 1/2)\prod(1 - 1/q)$, where the q's are any prime factors of n other than 2. Hence, $\prod(1 - 1/q) = 1$, so that there are zero q's and n contains only 2 as a prime factor, that is, $n = 2^a$ for some $a \geq 1$. □

Proof of (ii). If $n = 2^a 3^b$, with $a, b \geq 1$, then $\varphi(n)/n = (1-1/2)(1-1/3) = 1/3$ by (1), implying $\varphi(n) = n/3$. Conversely, let $\varphi(n) = n/3$. Then n is divisible by 3. First, we assert that n is also divisible by 2, for otherwise, $\varphi(n)/n = (1-1/3)\prod(1-1/q) = 1/3$, where the q's are any prime factors of n larger than 3, implying either the contradiction $2/3 = 1/3$ in the case of an empty set of q's, or otherwise $\prod(1-1/q) = 1/2$, which is impossible since $\prod(1-1/q) = 1/2$ implies $2\prod(q-1) = \prod q$, with the left side of the equation equal to an even number and the right side equal to an odd number since the q's are > 3. Hence, n is divisible by 2 and 3, so that $\varphi(n)/n = (1-1/2)(1-1/3)\prod(1-1/q) = 1/3$, implying $\prod(1-1/q) = 1$, which is possible only when the set of q's is empty. Therefore, $n = 2^a 3^b$. This concludes the proof of Lemma 2. □

Lemma 3. *If n is an odd composite number, then $\varphi(n) \leq n - 3$.*

Proof. If d is a divisor of n with $1 < d < n$ then $d \geq 3$, and hence n/d and $2n/d$ are integers less than n and greater than 1 and not relatively prime to n. Hence, there are at least three numbers greater than or equal to 1 and less than or equal to n that are not relatively prime to n, namely d, $2n/d$, and n. Therefore, $\varphi(n) \leq n - 3$, as required. □

In the first part of Theorem 1 below, we obtain upper bounds for the first few maxdegrees and describe an algorithm for getting similar bounds for higher-order maxdegrees. In the second part, we give a method to obtain lower bounds for maxdegrees and show that the upper bounds computed in the first part are also lower bounds, hence are optimal.

Theorem 1. *A. (i) $S_2 \leq 3/2$; and (ii) $S_3 \leq 7/3$. In general, if $n > 1$ and $\triangle$ is the product of the primes $\leq n$, then*

$$S_n \leq \max\left\{\sum_{k=0}^{n-1} \frac{\varphi(\gcd(j+k,\triangle))}{\gcd(j+k,\triangle)} : j = 1, \ldots, \triangle\right\}. \tag{3}$$

(Using (3), for example, the next few bounds for S_n can be determined: $S_4 \leq 17/6$, $S_5 \leq 107/30 = 3.566\ldots$, $S_6 \leq 59/15 = 3.933\ldots$, and $S_7 \leq 1019/210 = 4.852\ldots$.)

B. The bounds in (i) and (ii) of part A are also lower bounds, and therefore $S_2 = 3/2$ and $S_3 = 7/3$. (The same method for obtaining lower bounds of S_2, S_3 can also be used for higher-order maxdegrees, and in particular, to demonstrate the optimality of the above upper bounds for S_4, S_5, S_6, S_7.)

Proof of Part A. The number of cases is small for us to give intuitive proofs for (i) and (ii), which we shall do. Of course, they are subsumed in the general case which is handled by (3) above.

Proof of (i). Let I_2 be a 2-interval and a, b be its elements in increasing order. Then one of a and b, say b, must be a multiple of 2 (the only divisibility constraint required for a 2-interval), while a can be prime. (The case where a is a multiple of 2 is analogous.) Using Lemma 1, the largest upper bound required for a cluster point of the sequence $P_2 = \{\rho(I_2)\}$ is 3/2 since

$$\rho(I_2) = \rho(a) + \rho(b) \leq 1 + \frac{\varphi(b)}{b-1} \leq 1 + \frac{1}{2}\left(\frac{b}{b-1}\right) \to 1 + \frac{1}{2} = \frac{3}{2} \text{ as } b \to \infty.$$

Proof of (ii). Let I_3 be a 3-interval and a, b, c be its elements in increasing order. Then one of the numbers in I_3 must be a multiple of 2 and one must be a multiple of 3 (the only divisibility constraints required for a 3-interval). It is easily checked that the largest upper bound required for a cluster point of the sequence $P_3 = \{\rho(I_3)\}$ is 7/3, which occurs when b is both a multiple of 2 and of 3. For example, in the case that b is a multiple of 2 and 3,

$$\begin{aligned}\rho(I_3) = \rho(a) + \rho(b) + \rho(c) &\leq 1 + \frac{\varphi(b)}{b-1} + 1 \leq 2 + \frac{1}{3}\left(\frac{b}{b-1}\right) \\ &\to 2 + \frac{1}{3} = \frac{7}{3} \text{ as } b \to \infty.\end{aligned}$$

In the case that a is a multiple of 2 and 3, then c must also be a multiple of 2, and the corresponding upper bound (obtained using Lemma 1 as above) is $1/3 + 1 + 1/2 = 11/6 < 7/3$; the case where c is a multiple of 2 and 3 is handled similarly. If the multiples of 2 and 3 are distinct, say when a is a multiple of 3 and b is a multiple of 2, then the corresponding upper bound is $2/3 + 1/2 + 1 = 13/6 < 7/3$.

We now describe a general "brute force" but easily-automated approach to find S_n. Consider for example S_5. First we find S_5^*, the 5-maxdegree relative to the ρ^* primality measure (in Definition 2). The idea is to consider only prime divisors ≤ 5, that is, prime factors of $30 = 2 \times 3 \times 5$; this works since

$$\rho^*(n) = \frac{\varphi(n)}{n} = \prod_{p|n}\left(1 - \frac{1}{p}\right) \leq \prod_{p|n,\ p|30}\left(1 - \frac{1}{p}\right) = \frac{\varphi(\gcd(n,30))}{\gcd(n,30)}.$$

Now we observe that, since $\gcd(n, 30)$ is periodic with period 30, the problem of finding the bound for S_5^* is reduced to finding the maximum of the 30 quantities

$$\sum_{k=0}^{4} \frac{\varphi(\gcd(n+k,30))}{\gcd(n+k,30)} \quad (n = 1, \ldots, 30),$$

which can easily be done using software such as *Mathematica*. It turns out that the maximum is 107/30 for $n = 7$ and $n = 19$. We have now found S_5^*, the 5-maxdegree relative to the ρ^* primality measure, and therefore also S_5, since $\rho/\rho^* \to 1$. This completes the proof of Part A. $\square$

Proof of Part B. We use the following result from sieve theory [3, p. 58 and formula 6.107 on p. 78]. Let F be a product of k distinct irreducible polynomials with integral coefficients and positive leading coefficients, and let F have no fixed prime divisor. Then there are infinitely many positive integers m such that $F(m)$ has no prime divisors less than $m^{1/(4k+1)}$.

The main idea is that $\rho(n)$ is close to 1 if n has only large prime divisors, and the sieve theory result guarantees that there will be intervals that have only large divisors except for the few small divisors that we account for separately. Take S_3, for example. The upper bound 7/3 comes from the triple $\{6n-1, 6n, 6n+1\}$ (obtained from the maximizer in (3) above: maximum $= 7/3$ attained at $j = 5$). Factoring out the constant prime divisors and forming the polynomial $F(m) = (6m-1)m(6m+1)$, the sieve theory result implies that there are infinitely many m such that the numbers $6m-1$, m, $6m+1$ have no prime divisor less than $m^{1/13}$. Therefore, each of these numbers has at most 13 prime factors (counting multiplicity), and so, denoting by M any one of these numbers, we have $\varphi(M)/M \geq (1-1/m^{1/13})^{13}$. The last expression goes to 1 as m goes to infinity, hence $\varphi(M)/M \to 1$ for each of the three numbers M. Therefore, $\rho(6m-1) + \rho(6m) + \rho(6m+1) \to 1 + 1/3 + 1 = 7/3$, as desired.

The proof for S_2 is similar. The upper bound 3/2 comes from the pair $\{2n, 2n+1\}$ (obtained from the maximizer in (3) above: maximum $= 3/2$ attained at $j = 2$), and $F(m) = m(2m+1)$. Then $\rho(2m) + \rho(2m+1) \to 1/2 + 1 = 3/2$.

As an example of handling higher-order maxdegrees, consider S_5. The upper bound 107/30 comes from the quintuple

$$\{30n+7, 30n+8, 30n+9, 30n+10, 30n+11\}$$

(maximum $= 107/30$ attained at $j = 7$). The product of the elements of the quintuple is

$$G(n) = (30n+7)[2(15n+4)][3(10n+3)][10(3n+1)](30n+11)$$

where we have factored out the constant prime divisors. We get

$$F(m) = (30m+7)(15m+4)(10m+3)(3m+1)(30m+11).$$

However, $F(m)$ still has constant prime factors, namely 2. This can be seen by reducing $F(m)$ modulo 2 to get $F^*(m) = m(m+1)$, which is always even, or empirically by writing out the first few values of $F(m)$ (e.g. $F(1)$, …, $F(10)$).

Yet it turns out that this difficulty can be fixed by considering the subsequence

$$\begin{aligned} L(n) = G(2n) &= (60n+7)(60n+8)(60n+9)(60n+10)(60n+11) \\ &= (60n+7)[4(15n+2)][3(20n+3)][10(6n+1)](60n+11) \end{aligned}$$

and

$$Q(m) = (60m+7)(15m+2)(20m+3)(6m+1)(60m+11).$$

We see that $Q(m)$ does not have any constant prime divisors by inspecting initial terms of $Q(m)$ and observing that any constant prime divisor p of $Q(m)$ divides $\triangle$. (For, as in the derivation of F^* above, p does not exceed $N = 5$, the number of linear factors of $L(n)$, and hence p is one of the primes not exceeding N). Then as in previous cases, $\rho(60m+7)+\rho(60m+8)+\cdots+\rho(60m+11) \to 1+1/2+2/3+2/5+1 = 107/30$. (Here we have used Lemma 1 with $M = 10$.)

S_4, S_6, S_7 are handled similarly as S_5. The corresponding maxima for S_4, S_6, S_7 are attained at $j = 4,\ 26,\ 37$, respectively. This completes the proof of Part B and Theorem 1. □

Definition 6. *If p, $p+2$ is a pair of twin primes and $p+1 = 2^a3^b$, for some $a, b \geq 0$, we call $p+1$ a Dan number.*

See A027856 of [2], which lists the currently known Dan numbers. Note that, except for the first Dan number 4, we have $p+1 = 2^a3^b$, for some $a, b \geq 1$, for all Dan numbers.

In Theorem 2, we enumerate the 2- and 3-maxprimes. From this theorem and the conjectured infinity of Mersenne primes, we can conjecture that there are infinitely many 2- and 3-maxprimes.

Theorem 2.

(i) *The 2-maxprimes are the 2-intervals $J_n = \{2^n, 2^n+1\}$ where 2^n+1 is prime and $K_n = \{2^n-1, 2^n\}$ where 2^n-1 is prime. The sequence of left endpoints of the 2-maxprimes starts: 2, 3, 4, 7, 16, 31, 127, 256, 8191, ... (Note that the Fermat primes are the right endpoints of J_n, and the Mersenne primes are the left endpoints of K_n.)*

(ii) *The 3-maxprimes are the 3-intervals $\{2, 3, 4\}$, and $L_p = \{p, p+1, p+2\}$ where p and $p+2$ are twin primes and $p+1$ is a Dan number. The sequence of left endpoints of the 3-maxprimes starts: 2, 3, 5, 11, 17, 71, 107, 191, 431, 1151, ...*

Proof of (i). Since

$$\rho(J_n) = \rho(K_n) = 1 + \frac{2^{n-1}}{2^n-1} \geq 1 + \frac{2^{n-1}}{2^n} = 1 + \frac{1}{2} = \frac{3}{2}$$

then J_n and K_n are 2-maxprimes.

To see that any 2-maxprime I is among the J_n and K_n, note that some element s of I must be a multiple of 2. Indeed, $s = 2^a$ for some $a \geq 1$, for otherwise s has prime factors $q \geq 3$, and so

$$\rho(s) = \frac{\varphi(s)}{s}\frac{s}{s-1} = \frac{s}{s-1}\left(1-\frac{1}{2}\right)\prod\left(1-\frac{1}{q}\right) = \frac{1}{2}\frac{\prod\frac{q-1}{q}}{\frac{s-1}{s}} < \frac{1}{2}$$

since

$$\frac{\prod\frac{q-1}{q}}{\frac{s-1}{s}} < 1$$

(because all factors $(q-1)/q$ are less than $(s-1)/s$ and 1). But, denoting the other element of I by t, we have $\rho(s)+\rho(t) \geq 3/2$, and $\rho(s) < 1/2$ implies that $\rho(t) \geq 3/2 - \rho(s) > 3/2 - 1/2 = 1$, contradicting $\rho(t) \leq 1$. Hence, $s = 2^a$ for some $a \geq 1$.

Now, we show that t is prime, which will complete the proof of part (i).
Case (a). $t = s+1$. Since $s = 2^a$, $a \geq 1$,

$$\rho(s)+\rho(t) \geq \frac{3}{2}$$
$$\rho(t) \geq \frac{3}{2} - \rho(s) = \frac{3}{2} - \frac{2^{a-1}}{2^a-1} = \frac{2^{a+1}-3}{2^{a+1}-2}.$$

Now, $s = 2^a$ and $t = s+1$ imply $2s = 2(t-1) = 2t-2 = 2^{a+1}$. Then

$$\rho(t) \geq \frac{(2t-2)-3}{(2t-2)-2} = \frac{2t-5}{2t-4}$$
$$\frac{\varphi(t)}{t-1} \geq \frac{t-\frac{5}{2}}{t-2}$$

which, noting that $t > 2$, can be rewritten as

$$(t-2)\varphi(t) \geq \left(t-\frac{5}{2}\right)(t-1) \geq \left(t-\frac{5}{2}\right)(t-2).$$

Cancelling $(t-2)$ from both sides yields

$$\varphi(t) \geq t - 2.5. \tag{4}$$

If t were not prime, then t would be an odd composite number (because s is even), and we obtain by Lemma 3, $\varphi(t) \leq t-3$, contradicting (4). Therefore, t must be prime.
Case (b). $t = s-1$. As in Case (a) above, $s = t+1$ and $2s = 2t+2 = 2^{a+1}$, and so

$$\rho(t) \geq \frac{(2t+2)-3}{(2t+2)-2} = \frac{2t-1}{2t}$$

which can be rewritten as

$$\varphi(t) \geq t - \frac{3}{2} + \frac{1}{2t}$$

implying

$$\varphi(t) \geq t - \frac{3}{2}. \tag{5}$$

If t were not prime, then t would be an odd composite number (because s is even), and we obtain by Lemma 3, $\varphi(t) \leq t - 3$, contradicting (5). Therefore, t must be prime. $\square$

Proof of (ii). Since $\rho(\{2,3,4\}) = \rho(\{3,4,5\}) = 8/3 > 7/3$ and

$$\rho(L_p) = 2 + \frac{2^a 3^{b-1}}{2^a 3^b - 1} > 2 + \frac{2^a 3^{b-1}}{2^a 3^b} = 7/3$$

where $p + 1 = 2^a 3^b$, with $a, b \geq 1$, then $\{2,3,4\}$ and all the L_p's are 3-maxprimes.

In the rest of the proof we shall use the fact that $x/(x-1) \to 1$ from above, and so for example, if $x \geq 11$, then $x/(x-1) \leq 11/(11-1) = 11/10$.

To see that any 3-maxprime $I = \{p, p+1, p+2\}$ is among $\{2,3,4\}$ and L_p, first note that at least one element s of I is a multiple of 2 and at least one element t of I is a multiple of 3. It is easily checked that the first three 3-maxprimes are $\{2,3,4\}$, $\{3,4,5\}$, and $\{5,6,7\}$; for succeeding 3-maxprimes, we have $s, t \geq 11$ (since the next 3-maxprime is $\{11,12,13\}$), which we now assume without loss of generality. Now, $s = t$ since otherwise, using Lemma 1,

$$\begin{aligned}\rho(I) &\leq 1 + \rho(s) + \rho(t) = 1 + \frac{\varphi(s)}{s-1} + \frac{\varphi(t)}{t-1} \leq 1 + \frac{s/2}{s-1} + \frac{2t/3}{t-1} \\ &\leq 1 + \frac{11/2}{11-1} + \frac{\frac{2}{3}(11)}{11-1} = \frac{137}{60} < \frac{7}{3}\end{aligned}$$

contradicting $\rho(I) \geq 7/3$. Hence $s = t$ is a multiple of 2 and of 3.

Indeed, $s = t$ has no prime factors other than 2 and 3, that is, $s = t = 2^a 3^b$ for some $a, b \geq 1$. Otherwise, s has prime factors $q \geq 5$, and thus

$$\rho(s) = \frac{\varphi(s)}{s}\frac{s}{s-1} = \frac{s}{s-1}\left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{3}\right)\prod\left(1 - \frac{1}{q}\right) = \frac{1}{3}\frac{\prod\frac{q-1}{q}}{\frac{s-1}{s}} < \frac{1}{3}$$

since

$$\frac{\prod\frac{q-1}{q}}{\frac{s-1}{s}} < 1$$

(as in the proof of Part (i) above). But, denoting the other elements of I by x and y, we have $\rho(s) + \rho(x) + \rho(y) \geq 7/3$, and $\rho(s) < 1/3$ implies that $\rho(x) + \rho(y) \geq 7/3 - \rho(s) > 7/3 - 1/3 = 2$, contradicting $\rho(x) + \rho(y) \leq 2$. Hence, $s = 2^a 3^b$ for some $a, b \geq 1$.

Next, recalling that $I = \{p,\ p+1,\ p+2\}$ we must have $s = p + 1$, that is, s must be the midpoint of I. Otherwise, I will contain two distinct multiples of 2, namely p and $p+2$, and then, writing $p_1 = p$, $p_2 = p+1$, $p_3 = p+2$,

$$\begin{aligned}\rho(I) = \rho(p_1) + \rho(p_2) + \rho(p_3) &\le \frac{\varphi(p_1)}{p_1 - 1} + 1 + \frac{\varphi(p_3)}{p_3 - 1} \\ &\le \frac{1}{2}\frac{p_1}{p_1 - 1} + 1 + \frac{1}{2}\frac{p_3}{p_3 - 1} \\ &\le \frac{1}{2}\frac{11}{11-1} + 1 + \frac{1}{2}\frac{11}{11-1} = \frac{21}{10} < \frac{7}{3}\end{aligned}$$

(since we have assumed $p_1,\ p_3 \ge 11$), contradicting $\rho(I) \ge 7/3$. Hence, $s = p+1 = 2^a 3^b$.

It remains to show that p and $p+2$ are prime. First, we show that p is prime. Now,

$$\rho(p) \ge \frac{7}{3} - \rho(p+1) - \rho(p+2) \ge \frac{7}{3} - \rho(p+1) - 1 = \frac{4}{3} - \frac{2^a 3^{b-1}}{2^a 3^b - 1} = \frac{2^a 3^{b+1} - 4}{2^a 3^{b+1} - 3}.$$

But $p+1 = 2^a 3^b$, so $3(p+1) = 3p+3 = 2^a 3^{b+1}$, and the previous inequality becomes

$$\begin{aligned}\rho(p) &\ge \frac{(3p+3)-4}{(3p+3)-3} = \frac{3p-1}{3p} \\ \frac{\varphi(p)}{p-1} &\ge \frac{3p-1}{3p}.\end{aligned}$$

The last inequaltiy can be rewritten as

$$\varphi(p) \ge p - \frac{4}{3} + \frac{1}{3p}$$

implying

$$\varphi(p) \ge p - 2. \tag{6}$$

If p were not prime, then p would be an odd composite number (because $p+1 = s$ is even), and Lemma 3 would give $\varphi(p) \le p - 3$, contradicting (6). Therefore, p must be prime.

Finally, we show that $p+2$ is prime. As in the preceding argument,

$$\rho(p+2) = \frac{\varphi(p+2)}{(p+2)-1} = \frac{\varphi(p+2)}{p+1} \ge \frac{3p-1}{3p}$$

which can be written as

$$\varphi(p+2) \ge p + \frac{2}{3} - \frac{1}{3p}$$

and thus

$$\varphi(p+2) \ge p. \tag{7}$$

But by Lemma 3, $\varphi(p+2) \le (p+2) - 3 = p-1$, contradicting (7). Therefore, $p+2$ is prime. The proof of Theorem 2 is now complete. $\square$

3. Searching for Maxprimes

The n-maxprimes are more difficult to find for larger n. Hitherto, we have completely characterized the 2- and 3-maxprimes. In this section, we search for higher-order maxprimes.

The following *Mathematica* program can be used to search for n-intervals whose primalities are greater than or equal to some threshold value `thresh`. (To search for n-maxprimes, set `thresh` to the value of S_n.) In the program, the variable `lenint` is set to n, and `hi` is set to the largest value of an interval's left endpoint we would like to examine.

```
prm[n_]:=EulerPhi[n]/(n-1)    /;   n>1
prm[n_]:= 0                   /;   n=1
l = {}; lenint = 7; li2 = lenint - 1;
hi = 10^8; thresh = 4.85;
cnt = Apply[Plus, Map[prm, Table[i, {i, 1, lenint}]]];
For[i = 2, i <= hi, i++,
  If[cnt >= thresh,
    l = Append[l, {i - 1, N[cnt]}]];
  If[i < hi, cnt = cnt + prm[i + li2] - prm[i - 1]]];
MatrixForm[l]
```

The program output is in matrix form with two entries for each row: the first entry is the left endpoint of the n-interval and the second entry is the interval's primality (to several decimal places) $\geq$ `thresh`. The following is (partial) output for the program with inputs `lenint` =7, `thresh` =4.85, `hi` =10^8.

$$\begin{pmatrix} 1 & 5.06667 \\ 2 & 5.6381 \\ 3 & 5.3881 \\ 5 & 5.16587 \\ 7 & 5.12951 \\ 11 & 4.92994 \\ 13 & 4.91924 \\ 17 & 4.85018 \\ 21377 & 4.85139 \\ 39227 & 4.85044 \end{pmatrix} \tag{8}$$

After running the program consecutively for n = `lenint` = 4, 5, 6, 7, and `hi` = 10^8, we find very few n-maxprimes. For example, we discover only five 4-maxprimes, no 5-maxprimes with left endpoint greater than 79, only nine 6-maxprimes, and no 7-maxprimes with endpoint greater than 13.

4. A Conjecture on the Distribution of Maxprimes

What happens to S_n for larger n? It appears that $S_n/n \to C = 6/\pi^2 \sim 0.61$. We conjecture that $S_n = Cn$ asymptotically. This claim is made very plausible by the following heuristic argument, which might well be a simple matter to formalize. First,

$$\frac{S_n}{n} = \frac{\limsup P_n}{n} \sim \left(\frac{1}{n}\right)\sum_{i=1}^{n}\frac{\varphi(i)}{i}.$$

The '$\sim$' is the conjectured, unproved, part. Now it is known that

$$\sum_{i=1}^{n}\frac{\varphi(i)}{i} = \frac{n}{\zeta(2)} + O(\log n) = \frac{6}{\pi^2}n + O(\log n)$$

since $\zeta(2) = \pi^2/6$ [5, Exercise 5, p. 70]. Therefore,

$$\left(\frac{1}{n}\right)\sum_{i=1}^{n}\frac{\varphi(i)}{i} = \frac{6}{\pi^2} + O\left(\frac{\log n}{n}\right)$$

$$\frac{S_n}{n} \sim \frac{6}{\pi^2} + O\left(\frac{\log n}{n}\right)$$

which concludes the heuristic argument.

In connection with the conjecture, it is not hard to show the following.

Lemma 4.

$$S_n/n \geq 1/\zeta(2) = 6/\pi^2.$$

Proof. First, we assert that

$$\limsup \frac{1}{n}\sum_{k=1}^{n} a_k \leq \limsup a_n. \tag{9}$$

This is a special case of the Stolz-Cesàro Theorem [6]: If $\{b_n\}$ is a sequence of positive real numbers whose sum diverges, then for any sequences $\{a_n\}$ of real numbers we have

$$\limsup \frac{a_1 + a_2 + \cdots + a_n}{b_1 + b_2 + \cdots + b_n} \leq \limsup \frac{a_n}{b_n}. \tag{10}$$

Equation (9) follows immediately from (10) by taking $b_n = 1$ for all n.

Next, we apply (9) to the sequence $a_n = \rho(n) + \cdots + \rho(n+m-1)$. The same result on the average value of $\varphi(n)/n$ cited above (i.e. [5, Exercise 5, p. 70]) shows that the average value of $\rho(n)$ is also $1/\zeta(2)$, and so the average value of a_n is $m/\zeta(2)$. Using (9), $S_m = \limsup P_m = \limsup a_n \geq \limsup(\text{average value of } \{a_1, \ldots, a_m\}) = m/\zeta(2)$. Therefore, $S_m/m \geq 1/\zeta(2)$, as desired. □

5. Problems

In the following, we gather in one place a few of the many unanswered questions and problems on this topic.

(a) Are there only finitely many 4-, 5-, 6-, and 7-maxprimes? If not, exhibit larger maxprimes than the few already mentioned above. How about n-maxprimes for $n > 7$?

(b) *Near Maxprimes.* It may well be the case that there are only finitely many n-maxprimes for some n. In such a case, "near maxprimes", whose primalities are only approximately equal to their corresponding maxdegrees, would be the best we can do in the way of exhibiting 'large' n-intervals with maximal primality. Near maxprimes have a looser structure than maxprimes; for example, the maxprimes we have found can be obtained by including in an interval as many primes and as few divisors of composite numbers as divisibility constraints allow, but some near maxprimes contain no prime numbers at all.

Consider the case $n = 7$. It is impossible to have three consecutive primes each differing from the previous one by 2. (This is easily shown by considering their residues modulo 3.) Hence, the smallest possible length of an interval that contains three primes is 7; for example, with primes of the form p, $p+2$, $p+6$. Recall that in Section 3, we ran the program for `thresh` $= 1019/210 = 4.852\ldots$ (the value of S_7 obtained from Theorem 1) and found very few 7-maxprimes, indeed none with left endpoint greater than 13. Let us then select any convenient `thresh` value very slightly less than this, say `thresh` $= 4.85$. (So 7-near maxprimes are more properly called $(7, \varepsilon)$-near maxprimes, where `thresh` $= S_7 - \varepsilon$, due to the dependence on $\varepsilon > 0$.) Running the program with `thresh` $= 4.85$ and `hi` $= 10^8$ yields the near 7-maxprimes of the above sample program output in (8). There are 554 rows in this output, but only the first few rows are shown. There are 15 near 7-maxprimes that do not contain prime numbers at all, the first one being $\{26171707, 26171708, \ldots, 26171713\}$. Remarkably, after the first seven rows, all the left endpoint values end in the digit 7, that is, are congruent to 7 modulo 10.

Do all, except for finitely many, 7-maxprimes have left endpoints that are congruent to 7 modulo 10? For a fixed $n \geq 2$, given $\varepsilon > 0$, can we always find an n-interval with primality at least $S_n - \varepsilon$ and containing no prime numbers? (Other constraints on the interval can be considered, e.g. that it contain at most one prime, etc.)

(c) It seems likely that (3) of Theorem 1, Part A, gives not just upper bounds, but the actual values of maxdegrees. Obtain a formal proof of this. However, even assuming that (3) holds, we have more of an algorithm than a 'simple' formula (say, expressible in terms of elementary functions) for computing the n-maxdegrees. Is there such a simple formula?

(d) Does $\lim S_n/n = 6/\pi^2$? If this is true, then $S_n = (6/\pi^2)n$ is an asymptotic answer to the 'simple' formula required in Problem (c) above.

(e) Investigate the primalities of other finite sets of numbers, for example, sets whose elements are not necessarily consecutive. (We remark that the primality of sets of numbers in arithmetical progression has been studied; see [4] for example.)

(f) Can maxprimes–even near maxprimes–be used in lieu of prime numbers in some applications–perhaps in probabilistic scenarios where a specific prime number value can be replaced by a range of 'highly prime', but not necessarily prime, values?

6. Acknowledgments

The author would like to thank Prof. Mohammad R. Khadivi for reviewing this paper and for his valuable suggestions on its improvement. The author is grateful to the number theory referees of the *Missouri Journal of Mathematical Sciences* for their role in polishing the paper and strengthening its results.

References

[1] G. H. Hardy, E. M. Wright, et al., *An Introduction to the Theory of Numbers*, 6th edition, Oxford University Press, Oxford, 2008.

[2] N. J. A. Sloane (Ed.), The On-Line Encyclopedia of Integer Sequences, 2010. `http://www.research.att.com/~njas/sequences/`.

[3] J. Friedlander and H. Iwaniec, *Opera de Cribro*, American Mathematical Society, Providence, RI, 2010.

[4] B. Green and T. Tao, *The primes contain arbitraily long arithmetic progressions*, Annals of Mathematics, **167** (2008), 481–547.

[5] T. Apostol, *Introduction to Analytic Number Theory*, Springer-Verlag, New York, 1976.

[6] G. Nagy, *The Stolz-Cesro theorem*, `http://www.math.ksu.edu/~nagy/snippets/stolz-cesaro.pdf`.

MSC2010: 11A25, 11A41

Director, Data Mining, Catalina Marketing Corporation, 20 N. Martingale Road, Suite 300, Schaumburg, IL 60173

E-mail address: joseph.pe@catalinamarketing.com