Notes on Number Theory and Discrete Mathematics
Print ISSN 1310–5132, Online ISSN 2367–8275
Vol. 22, 2016, No. 4, 29–40

Asymptotic formulae for the number of repeating prime sequences less than N

Christopher L. Garvie

Texas Natural Science Center, University of Texas at Austin
10100 Burnet Road, Austin, Texas 78758, USA
e-mail: clgarvie@gmail.com

Received: 21 April 2016 **Accepted:** 30 October 2016

Abstract: It is shown that prime sequences of arbitrary length, of which the prime pairs, $(p, p+2)$, the prime triplet conjecture, $(p, p+2, p+6)$ are simple examples, are true and that prime sequences of arbitrary length can be found and shown to repeat indefinitely. Asymptotic formulae comparable to the prime number theorem are derived for arbitrary length sequences. An elementary proof is also derived for the prime number theorem and Dirichlet's Theorem on the arithmetic progression of primes.
Keywords: Primes, Sequences, Distribution.
AMS Classification: 11A41.

1 Introduction

We build up the sequences of primes using a variant of the sieve process devised by Eratosthenes. In the sieve process the integers are written down sequentially up to the largest number we wish to check for primality. We first remove from consideration every number of the form 2n (n≥2), find the smallest number >2 that was not removed, that is 3 (the next prime), and similarly now remove all numbers of the form 3n (n≥2). As before find again the next number not removed, that is the next prime 5. The process is continued up to a desired prime P_r. We now write down N zeros, the position of the r^{th} zero being a placemarker for the number r. A zero at position r denotes r is either known to be a prime, or is a candidate to be prime, a one at position r denotes that r is known to be composite. Changing a zero to a one then denotes the fact that the prime candidate is found to be composite. We denote by σ_r the sequence of zeros and ones after the r^{th} prime has been used in the sieve process and denote by τ_r the repeating sequence part within σ_r. The first four are as follows:

σ_0: 000000000000000000000000000000000000000000000000000000000000 .....
σ_1: 000 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 ...
σ_2: 00010101 110101 110101 110101 110101 110101 110101 110101 110101 110101 ....
σ_3: 000101011101011101011101 111101011111011101011101011101 111101011111 .....

The spaces between the digits indicate the limits of the repeating subsequences, where τ_1=10 in σ_1, τ_2=110101 in σ_2, and τ_3=111101011111011101011101011101 in σ_3, at starting positions 2^2, 3^2, and 5^2 respectively. Each new prime P_r when used to change a zero to a one first affects the sequence in position P_r^2 as changes at positions $P_q.P_r$ $(q < r)$, have already been taken care of with prime P_q. Each new prime will therefore generate a new repeating sequence, which begins at the P_r^2 position. In sequence σ_r, all zeros at positions k, $k < P_r^2$ represent primes, at positions $k > P_r^2$, the zeros are candidates to be prime. Position 1 is not used as a prime even though it is 0.

2 Preliminary expressions

Lemma 1. *The repeating sequence length using the first n primes is given by:* $\prod_{r=1}^{n} P_r$.

Proof. Define S_n as the repeating sequence length. As every 1 within a τ_r sequence will repeat at the same relative position, we must have:

$$k_r P_r + S_n = l_r P_r, k - r > 0, l_r > 0, > k_r, r = 1, 2, \dots n,$$

and thus $P_r | S_n$ for each r, and:

$$S_n = C \prod_{r=1}^{n} P_r,$$

for some C. Choosing $l_r = k_r + 1$ we see C must be 1. □

Note that unlike the repeating sequences for primes 2 and 3, the repeating sequence is not complete before the sequence starts for the next prime 7; the τ_3 sequence for 5 starts at 25 and ends at 54, while the next prime 7 first affects the sequence at position 49. After prime P_2 =3, the repeating sequence of length S_n (starting at position P_n^2) in never completed before the next prime is added because for P_n >3 one sees that:

$$S_n > P_{n+1}^2 - P_N^2 > (P_n + 2)^2 - P_n^2 = 2P_n + 1,$$

which is clearly true for P_n >3.

When generating σ_{r+1} from σ_r the only new zeros that are changed to ones are those at positions whose prime factors include primes not already used, i.e. P_r^2, $P_r P_{r+1}$, $P_r P_{r+2}$, ... Without further primes added, the new sequence τ_r will repeat indefinitely with length $S_r = P_1 P_2 P_3 \ldots P_r$. What we now aim to show is that by adding new primes to the process results in some zeros (prime candidates) being changed to ones (composites) within the repeating τ_r sequences, but that an unchanged and increasing set of τ_r always remains after the application of ever larger primes.

3 Identifying the prime sequences

To illustrate the process we first take the simplest sequence, the prime pairs $(p, p{+}2)$. We consider σ_2 with $S_2 = 6$ and $N = 195$ (the largest number considered), so there are 31 repeating sequences of τ_2 =110101. All the 'ones' in the repeating sequences represent composite numbers, with factors $2^\alpha 3^\beta$, for some $\alpha + \beta > 1$, $\alpha, \beta \geq 0$. Applying the sieve process for all numbers $\leq$195 that have factors composed of primes $> P_2 (= 3)$, we generate 31 composite numbers. Each of those composite numbers occupying a zero position within one of the τ_2 sequences within σ_2. Within 18 of the 31 repeating τ_2 sequences a single zero is changed to a one (shown to be composite), while for two sequences both zeros are changed. The number of prime pairs is thus equal to the total number of τ_2 sequences minus the number in which one or both of the zeros is changed to ones. The number of changed sequences thus equals the total number of zeros changed to ones minus the number where both zeros are changed, which in this example is 20, leaving 11 as the correct number of prime pairs in the range (9,195).

For identifying prime triplets a similar process is be performed, but now using σ_3 with τ_3 = 111101011111011101011101011101, and $S_3 = 30$. Each τ_3 sequence includes possible prime triplets which by observation includes $(p, p+2, p+6), (p, p+2, p+8), (p, p+8, p+12)$, but also prime quadruplets, and more complicated sequences up to a maximum prime octuplet (i.e. using all the zeros within τ_3). As before with τ_2, the 'ones' in the repeating τ_3 sequences represent composite numbers, but now with factors $2^\alpha 3^\beta 5^\gamma$, for some $\alpha + \beta + \gamma > 1$, $\alpha, \beta, \gamma \geq 0$. The sieve process is applied but now with primes $> P_3 (= 5)$, and as before the number of triplets will be equal to the number of unchanged subsequences (e.g. a $(p, p+2, p+6)$ subsequence).

4 Principle theorem used and shorthand notation

The basis of the following derivations use a general theorem enunciated by Hardy & Wright [2] (see pp. 233–234), which we repeat here: If there are N objects of which N_α, have the property α, N_β have β, ..., $N_{\alpha\beta}$ have both α and β, $N_{\alpha\beta\gamma}$, have α, β, and γ, and so on, the number of objects which have none of the α, β, γ, ... is:

$$N - N_\alpha - N_\beta - \ldots + N_{\alpha\beta} + \ldots - N_{\alpha\beta\gamma} - \ldots \quad (1)$$

Following those authors, the number of integers less than or equal to N and not divisible by any one of a coprime set of integers $a, b, \ldots$, is:

$$[N] - \sum \left[\frac{N}{a}\right] + \sum \left[\frac{N}{ab}\right] - \ldots \quad (2)$$

For clarity in the ensuing equations, we use the following shorthand notation for representing sums of products of inverse primes. We define:

$$\sum_{k=1}^{<N>} \frac{1}{P_{\{k\}}}, \tag{3}$$

as the sum of all products of powers of inverse distinct primes, taken k at a time. We use the $< N >$ to indicate the series is truncated so that no inividual term is $< \frac{1}{N}$, and an $< N >$ over a product indicates the terms in the expansion are similarly truncated. We omit the lower bound if $r = 1$. Two examples:

$$\sum^{<N>} \frac{1}{P_{\{1\}}} = \sum_{r=1}^{m} \frac{1}{P_r} = \frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \ldots + \frac{1}{P_m}, \tag{4}$$

where P_m is the largest prime $\leq N$.

$$\sum^{<N>} \frac{1}{P_{\{2\}}} = \sum_{r=1}^{m} \sum_{s=r+1}^{n} \frac{1}{P_r P_s} = \frac{1}{2 \cdot 3} + \frac{1}{2 \cdot 5} + \ldots + \frac{1}{3 \cdot 5} + \frac{1}{3 \cdot 7} + \ldots, \tag{5}$$

where as before no terms are $< \frac{1}{N}$. When using integer parts, $\sum^{<N>} \left[\frac{1}{P_{\{1\}}}\right]$ indicates $\sum_{r=1}^{m} \left[\frac{1}{P_{\{r\}}}\right]$, i.e. a square bracket around each individual term, and similarly for $P_{\{2\}}$, etc.

5 Derivation of the prime number theorem

We now apply Hardy & Wrights' theorem (2) to the simplest sequence, the primes, and show how the prime number theorem can be derived from a different perspective.

Theorem 1. *If P_r is the r^{th} prime, and $e(N)$ the error term, the number of primes less than or equal to N can be written as:* $N \prod_{r=1}^{<N>} \left(1 - \frac{1}{P_r}\right) + e(N)$.

Proof. In (2) we now choose the set of integers $a, b, ...$ to be all the primes $\leq N$, $(P_1, P_2, ..., P_K, K = \pi(N))$, and in the new notation the number of integers not divisible by any one of those primes must be the first positive integer, i.e. 1, the expression is 0 if $N < P_1 (= 2)$, and we have:

$$U(N - P_1) = [N] - \sum^{<N>} \left(\left[\frac{N}{P_{\{1\}}}\right] - \left[\frac{N}{P_{\{2\}}}\right] + \ldots \right) = 1 \text{ if } N \geq P_1, 0 \text{ if } N < P_1, \tag{6}$$

and $U(N - P_1)$ is seen to be the unit step function with argument $N - P_1$. The sums within the brackets all have have a finite number of non-zero terms. Changing the step function argument, we now write:

$$U(N - q) = \left[\frac{N}{q}\right] - \sum^{<N>} \left(\left[\frac{N}{qP_{\{1\}}}\right] - \left[\frac{N}{qP_{\{2\}}}\right] + \ldots \right) = 1 \text{ if } N \geq q, 0 \text{ if } N < q, \tag{7}$$

Summing now for a set of q sucessively taking the value of all primes $\leq N$ we get:

$$\pi(N) = \sum_{k=1}^{K} U(N - P_k) = \sum_{k=1}^{K} \left(\left[\frac{N}{P_k} \right] - \sum^{<N>} \left(\left[\frac{N}{P_k P_{\{1\}}} \right] - \left[\frac{N}{P_k P_{\{2\}}} \right] + \dots \right) \right) \tag{8}$$

where $K = \pi(N)$. This is an exact representation for $\pi(N)$, the number of primes $\leq N$. To derive the standard distribution function we remove the integer parts [ ] and obtain:

$$\pi(N) = \left(\sum_{k=1}^{K} \frac{1}{P_{\{k\}}} \right) N \left(1 - \sum^{<N>} \left(\frac{1}{P_{\{1\}}} - \frac{1}{P_{\{2\}}} + \dots \right) \right) + E(N), \tag{9}$$

where $E(N)$ is the error term between the exact formulation given in (8) and the similar expression (9) without the [ ] brackets.

Clearly in (8) each P_k only contributes 1 to the final result, and using any set of K numbers within the range (P_1, N) would result in the same value of $\pi(N)$. This is not the case in the continuous formulation (9), where the error will be dependent on the choice of the P_k, due to the factor $\left(\sum_{k=1}^{K} \frac{1}{P_k} \right)$. Setting:

$$\begin{aligned} &\sum_{k=1}^{K} \left(\left[\frac{N}{P_k} \right] - \sum^{<N>} \left(\left[\frac{N}{P_k P_{\{1\}}} \right] - \left[\frac{N}{P_k P_{\{2\}}} \right] + \dots \right) \right) \\ &= \left(\sum_{k=1}^{K} \frac{1}{s_k} \right) N \left(1 - \sum^{<N>} \left(\frac{1}{P_{\{1\}}} - \frac{1}{P_{\{2\}}} + \frac{1}{P_{\{1\}}} - \dots \right) \right) + e(N), \end{aligned} \tag{10}$$

where $e(N)$ is the error term when the s_k set is chosen instead of the P_k set $(e(N \neq E(N))$. Choosing $s_k = \pi(N)$ for all k we get:

$$\begin{aligned} \pi(N) &= N \left(1 - \sum^{<N>} \left(\frac{1}{P_{\{1\}}} - \frac{1}{P_{\{2\}}} + \frac{1}{P_{\{3\}}} - \dots \right) \right) + e(N) \\ &= N \prod_{r=1}^{<N>} \left(1 - \frac{1}{P_r} \right) + e(N), \end{aligned} \tag{11}$$

and the theorem is proved. □

We now determine the magnitude of the error term in (11)

Theorem 2. *The error term* $e(N)$ *is given by:* $e(N) = O(Log(Log(N)))$, *as* $N \to \infty$.

Proof. With reference to (10) and (11) we repeat the equations, but using K instead of $\pi(N)$, for simplicity:

$$\begin{aligned} \sum_{k=1}^{K} 1 &= \sum_{k=1}^{K} \left(\left[\frac{N}{K} \right] - \sum^{<N>} \left(\left[\frac{N}{K P_{\{1\}}} \right] - \left[\frac{N}{K P_{\{2\}}} \right] + \dots \right) \right) \\ &= \sum_{k=1}^{K} \left(\frac{N}{K} - \sum^{<N>} \left(\frac{N}{K P_{\{1\}}} - \frac{N}{K P_2} + \dots \right) \right) + e(N), \end{aligned} \tag{12}$$

therefore:

$$e(N) = K\sum^{<N>}\left(\left(\frac{N}{KP_{\{1\}}} - \left[\frac{N}{KP_{\{1\}}}\right]\right) - \sum_{k=1}^{K}\left(\frac{N}{KP_{\{2\}}} - \left[\frac{N}{KP_{\{2\}}}\right]\right) + \dots\right) \tag{13}$$

For each bracketed individual term within the $\sum$, the difference between each continuous term and the corresponding integer one is: $\left(\frac{N}{a}\right) - \left[\frac{N}{a}\right]$; using the maximum difference of 1 for each even term and the minimum of 0 for each odd term we can bound the upper limit, and then using the 0 for each even term and 1 for each odd term we bound the lower limit:

$$\sum^{<N>}\left(-\frac{1}{P_{\{2\}}} - \frac{1}{P_{\{4\}}} - \dots\right) \le e(N) \le \sum^{<N>}\left(\frac{1}{P_{\{1\}}} + \frac{1}{P_{\{3\}}} + \dots\right) \tag{14}$$

Each $\sum^{<N>} \frac{1}{P_{\{r\}}}$ term is divergent as $N \to \infty$ because each includes a term $\sum_{r=k}^{N} \frac{1}{P_r}$. Using Everset & Ward [1] (p. 13), we find:

$$\sum^{<N>}\frac{1}{P_{\{1\}}} = \sum_{r=1}^{N}\frac{1}{P_r} = Log(Log(N)) + .261 + O\left(\frac{1}{Log(N)}\right), \tag{15}$$

and noting that: $\frac{1}{P_{\{k+1\}}} < \frac{1}{2}\frac{1}{P_{\{k\}}}$ for all $k \ge 1$, we get:

$$e(N) = cLog(Log(N)) + d + O\left(\frac{1}{Log(N)}\right), \tag{16}$$

with $-0.666 \le c \le 1.666$ and $-0.174 \le d \le 0.435$ resulting in the desired bound. □

Using the product definition of the Riemann Zeta function

$$\zeta(s) = \prod_{r=1}^{\infty}\left(1 - \frac{1}{P_r^s}\right)^{-1} = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \dots, s > 1, \tag{17}$$

and also the well-known expansion of the logarithm:

$$\sum_{r=1}^{N}\frac{1}{r} = Log(N) + \gamma_N, \tag{18}$$

where γ_N is the N^{th} approximant to the Euler-Mascheroni constant, we use (11) and write:

$$\begin{aligned}\pi(N) = N\prod_{r=1}^{<N>}\left(1 - \frac{1}{P_r}\right) + e(N) &= \frac{N}{\sum_{r=1}^{N}\frac{1}{r} + \Delta(N)} + e(N) \\ &= \frac{N}{Log(N) + \gamma_N + \Delta(N)} + e(N)\end{aligned} \tag{19}$$

where $\Delta(N)$ is the error term in the conversion of the product to the sum. There appears no easy way to determine $\Delta(N)$ directly from the following expression:

$$\prod_{r=1}^{<N>}\left(1 - \frac{1}{P_r}\right)\left(\sum_{r=1}^{<N>}\frac{1}{r} + \Delta(N)\right) = 1 \tag{20}$$

but it can be derived with the following limiting process:

Theorem 3. $\Delta(N) = -1$

Proof. We take equation (20) but add an extra parameter s, writing:

$$\begin{aligned}
&\prod_{r=1}^{<N>}\left(1-\frac{1}{P_r^s}\right)\left(\prod_{r=1}^{<N>}\left(1-\frac{1}{P_r^s}\right)^{-1}+\Delta_s(N)\right)\\
&=\left(\sum_{r=1}^{\infty}\frac{\mu(r)}{r^s}-\sum_{r=<N>+1}^{\infty}\frac{\mu(r)}{r^s}\right)\left(\sum_{r=1}^{\infty}\frac{1}{r^s}-\sum_{r=<N>+1}^{\infty}\frac{1}{r^s}+\Delta_s(N)\right)\\
&=\left(\zeta(s)^{-1}-\sum_{r=<N>+1}^{\infty}\frac{\mu(R)}{r^s}\right)\left(\zeta(s)-\sum_{r=<N>+1}^{\infty}\frac{1}{r^s}+\Delta_s(N)\right)=1,
\end{aligned} \tag{21}$$

where $\mu(r)$ is the Möbius function, and $\zeta(s)$ the Riemann Zeta function. For all $s > 1$ the terms in (21) are convergent. Letting $s \to 1$, results in (20). Solving for $\Delta_s(N)$ we get:

$$\Delta_s(N) = \frac{\sum_{r=<N>+1}^{\infty}\frac{\mu(r)}{r^s}+\zeta(s)^{-1}\left(\sum_{r=<N>+1}^{\infty}\frac{1}{r^s}\right)}{\zeta(s)^{-1}-\sum_{r=<N>+1}^{\infty}\frac{\mu(r)}{r^s}} \tag{22}$$

Letting $s \to 1$ now results in $\zeta(s)^{-1} \to 0$, and we find $\Delta(N) = \Delta_1(N) = -1$. □

Equation (19) thus represents the standard asymptotic prime number formula:

$$\begin{aligned}
\pi(N) &\sim \frac{N}{Log(N)+\gamma_N-1}+cLog(Log(N))+d\\
&\sim \frac{N}{Log(N)},\ \text{as } N\to\infty.
\end{aligned} \tag{23}$$

6 Derivation of Dirichlet's Theorem

Using the above results we can now deduce an alternate derivation of Dirichlet's Theorem on the distribution of primes in a linear equation. We consider the sequence:

$$d_k = \alpha k + \beta, k = 1, 2, \dots M, \tag{24}$$

with $(\alpha, \beta) = 1$, and $M = \left[\frac{N-\beta}{\alpha}\right]$. Let the prime factors of α be: $\{Q\} = Q_1, Q_2, \dots, Q_k$. $P^*_{\{r\}}$ denotes the set of primes but with factors of α omitted. We also define the sum function of the sequence as $\pi_1(\alpha, N)$ to differentiate from the prime numbers sum function $\pi(N)$. The subscript 1 refers to the fact that vwe are only considering one sequence, and as we shall see later β is not needed.

Theorem 4. *The number of times d_k is prime in range $(0, N)$ is given by:*

$$\pi_1(\alpha, N) \sim \frac{M\prod^{<M>}\left(1-\frac{1}{P_r}\right)}{\prod^{<M>}\left(1-\frac{1}{Q_r}\right)},$$

Proof. In any subset of P_i terms, P_i prime:

$$d_n = \alpha n + \beta,\ n = k, k+1, \ldots, k+P_i-1,\ (\alpha, P_i) = 1, \tag{25}$$

one of the d_n terms will be divisible by P_i, and therefore $\left[\frac{M}{P_i}\right]$ terms of the linear sequence that are divisible by P_i. With reference to (8), the number of terms not divisible by any of the primes P_r ($P_r \le N$), omitting the primes $\{Q\}$ is:

$$\pi_1(\alpha, N) = \sum_{k=1}^{L} U(M-P_k) = \sum_{k=1}^{L} \left(\left[\frac{M}{P_k^*}\right] - \sum^{<M>} \left(\left[\frac{M}{P_k^* P_{\{1\}}}\right] - \left[\frac{M}{P_k^* P_{\{2\}}}\right] + \ldots \right) \right) \tag{26}$$

and following the same procedure given in (9) and (10), we remove the [] brackets and obtain:

$$\begin{aligned} \pi_1(\alpha, N) &= \sum_{k=1}^{L} \left(\left[\frac{M}{P_k^*}\right] - \sum^{<M>} \left(\left[\frac{M}{P_k P_{\{1\}}^*}\right] - \left[\frac{M}{P_k P_{\{2\}}^*}\right] + \ldots \right) \right) \\ &= \left(\sum_{k=1}^{L} \frac{1}{s_k} \right) M \left(1 - \sum^{<N>} \left(\frac{1}{P_{\{1\}}^*} - \frac{1}{P_{\{2\}}^*} + \frac{1}{P_{\{1\}}^*} - \ldots \right) \right) + e'(M), \end{aligned} \tag{27}$$

and with reference to (11), we now choose all $s_k = \pi_1(\alpha, N)$ and obtain:

$$\pi_1(\alpha, N) = \frac{M \prod^{<M>} \left(1 - \frac{1}{P_r}\right)}{\prod^{<M>} \left(1 - \frac{1}{Q_r}\right)} + e'(M) \tag{28}$$

and the theorem is proved. □

So with reference to (19) we can now substitute for the two product terms and write:

$$\begin{aligned} \pi_1(\alpha, N) &= \left(\frac{N-\beta)}{\alpha\left(Log(\frac{N-\beta}{\alpha}) + \gamma_{N-\beta}\right)} \right) \left(\frac{1}{\prod_{r=1}^{<M>} \left(1 - \frac{1}{Q_r}\right)} \right) \\ &\sim \frac{N-\beta}{\Phi(\alpha)(Log(N-\beta) + \gamma_{N-\beta} - 1 - Log(\alpha))}, \end{aligned} \tag{29}$$

where $\Phi(\alpha)$ is the Euler Totient function, and $e'(M)$ the error term, cf. (19). Therefore as $N \to \infty$, this results in the standard Dirichlet formula:

$$\pi_1(\alpha, N) \sim \frac{N}{\Phi(\alpha) Log(N)}. \tag{30}$$

7 Main results

We assume at this point that r primes have been used in the generation of sequence σ_r in range $(0, N)$, with repeating subsequence τ_r of length S_r, starting at position P_r^2. As before, any zeros (prime candidates) in the range (P_r^2, N) can only have prime factors $P > P_r$, as prime factors

$P \leq P_r$, have already been used in the generation of positions with a one. We number the i zeros of the first occurrence of the repeating sequence τ_r in range $(P_r^2, P_r^2+S_r-1)$ as: $Z_{r_1}, Z_{r_2}, \ldots, Z_{r_k}$. With no loss of generality, and for simplicity we require $N - P_r^2 + 1$ to be divisible by S_r so that N corresponds with the last position of the final τ_r sequence. Each zero will repeat at positions:

$$nS_r + Z_{r_i}, \text{ for all } i = 1, 2, \ldots, k, \; nS_r + Z_{r_i} \leq N \tag{31}$$

We note that Z_{r_1} can only equal P_{r+1}, when $P_{r+1} \geq P_r^2$. The number of repeating τ_r intervals of length S_r, S_r, in range $(P_r^2 + S_r, N)$ is given by:

$$m_r = \left[\frac{N}{S_r}\right] - \left[\frac{P_r^2}{S_r}\right] = \left[\frac{N}{S_r}\right] \tag{32}$$

The $\left[\frac{P_r^2}{S_r}\right]$ term is zero as $P_r < 2P_{r-1}$ for all r (Hardy & Wright) [2] (p. 343). Using now $\pi_k(S_r, N)$ for the number of unchanged sequences $\leq N$, where the start of the τ_r sequences (at P_r^2), and k is the number of primes within the desired sequence, 2 for prime pairs, 3 for prime triplets, etc. It is understood that each replicating prime is defined by a different linear sequence with all $\alpha_k = S_r$ and $\beta_k = Z_{r_k}$.

Theorem 5. *The number of unchanged sequences $\leq N$ is given by:*

$$\pi_k(S_r, N) \sim \frac{S_r^{k-1}}{\Phi(S_r)^k} \frac{N}{(Log(N))^k}$$

Proof. Considering the first prime P_{r+1}, it will be a divisor of the terms generated by: $nS_r + Z_{r_i}, n = 1, 2, \ldots, m_r$, $\left[\frac{N}{P_{r+1}S_r}\right]$ times. The total number of repeating τ_r sequences for which one repeating zero is divisible by P_{r+1} is given by $\left[\frac{N}{S_rP_{r+1}}\right]$, and thus for k sets of repeating zeros there are:

$$k\left[\frac{N}{S_rP_{r+1}}\right], \tag{33}$$

of them, and thus with reference to (10), the number of σ_r sequences where no zero within a τ_r sequence is divisible by P_{r+1} is given by:

$$\left[\frac{N}{S_r}\right] - \binom{k}{1}\left[\frac{N}{S_rP_{r+1}}\right] + \binom{k}{2}\left[\frac{N}{S_rP_{r+1}^2}\right] - \binom{k}{3}\left[\frac{N}{S_rP_{r+1}^3}\right] + \ldots + (-1)^k\binom{k}{k}\left[\frac{N}{S_rP_{r+1}^k}\right] \tag{34}$$

We now remove the [ ] and can write, cf. (11), the number of generated values not divisible by P_{r+1} to be:

$$\frac{N}{S_r}\left(1 - \frac{1}{P_{r+1}}\right)^k + e_k^{"}(N) \tag{35}$$

The maximum difference between $\left(\frac{N}{S_rP_{r+1}^2}\right) - \left[\frac{N}{S_rP_{r+1}^2}\right]$ is 1, therefore the error term can be bounded as follows, cf. (14):

$$\begin{gathered} -2^{k-1} = -\binom{k}{1} - \binom{k}{3} - \ldots \leq e_k^{"}(N) \leq 1 + \binom{k}{2} + \binom{k}{4} + \ldots = 2^{k-1}, \text{ or} \\ e_k^{"}(N) \leq |2^{k-1}| \end{gathered} \tag{36}$$

For any k and N, $e''_k(N)$ is bounded and so may be ignored in the asymptotic expression and thus, cf. (28) for all primes $P > P_r$, the number of sequences not divisible by any $P > P_r$ is asymptotic to:

$$\sim \frac{N}{S_r} \prod_{i=r+1}^{<N>} \left(1 - \frac{1}{P_i}\right)^k \tag{37}$$

and thus with reference to (28) and (30) we have the following asymptotic expression for the number of unchanged σ_r sequences $\leq N$:

$$\pi_k(S_r, N) \sim \frac{S_r^{k-1}}{\Phi(S_r)^k} \frac{N}{(Log(N))^k} \tag{38}$$

□

To derive the error term for (37) we proceed as follows. With reference to (8) we have for a given S_r, N and k:

$$\pi_k(S_r, N) = \sum_{l=1}^{L} \left(\left[\frac{N}{S_r P_l}\right] - \sum^{<N>} \left(\left[\frac{N}{S_r P_l P_{\{1\}}^{\{k\}}}\right] - \left[\frac{N}{S_r P_l P_{\{2\}}^{\{k\}}}\right] + \dots \right) \right), \tag{39}$$

where $P_{\{1\}}^{\{k\}}$ runs over k equal sets of primes $> P_r$, $P_{\{2\}}^{\{k\}}$ runs over the same k equal sets two at a time, and so on. Following the same reasoning we used to derive (11) we write:

$$\pi_k(S_r, N) = \frac{N}{S_r} \prod_{l=r+1}^{<N>} \left(1 - \frac{1}{P_l}\right)^k + e_{r,k}(N) \tag{40}$$

and again using Everest & Ward [1], cf. (15), (36), we can bound the error term:

$$e_{k,r} \leq 2^{k-1} \left(Log(Log(N)) - Log(Log(P_r)) + O\left(\frac{1}{Log(N)}\right)\right) \tag{41}$$

and as $N \to \infty$, we find:$|e_{r,k}(N)| \to |2^{k-1} e(N)|$.

We list three examples below, the first are the prime pairs $(p, p+2)$ generated by σ_2, with $S_2 = 6$ and $k = 2$; the second a prime triplet $(p, p+2, p+6)$ generated by σ_3, with $S_3 = 30$, and $k = 3$, and the third also from σ_3, a prime quadruplet $(p, p+8, p+20, p+24)$ with $k = 4$.

Table 1.

Range	Prime pair actual	Prime pair eqn.	Prime pair % error	Prime triplet actual	Prime triplet eqn.	Prime triplet % error	Prime quadruplet actual	Prime quadruplet eqn.	Prime quadruplet % error
10^2	8	7	12.5	3	1	75.0	2	1	50.0
10^3	34	31	8.8	9	4	55.5	5	2	40.0
10^4	204	177	13.2	27	19	54.5	10	7	30.0
10^5	1223	1131	7.5	130	99	23.8	41	31	24.4
10^6	8168	7858	3.8	679	590	13.1	179	154	14.0
10^7	58979	57738	2.1	4319	3777	12.5	855	848	8.2
10^8	440311	442059	−3.9	27934	25636	8.2	4733	5060	−6.9
10^9	3423505	3492809	−2.0	189837	181888	4.2	28507	32021	−12.3

Authors Tomas Oliveira e Silva [3], state there are 808,675,888,577,4536 twin prime pairs below 10^{18}, and the asymptotic formula (38) gives 8.08175×10^{14} which results in a percentage error of 5.1%. The error in the formula results from three sources, the first two from ignoring the error terms $e(N)$ and $\Delta(N)$. cf. (19), ignoring $e(N)$ lowers the resulting value while ignoring $\Delta(N)$ increases it. The third derives from the fact that this is essentially an averaging algorithm as the positions of the individual terms within the $P_{\{1\}}, P_{\{2\}}, \ldots$, are not known, only their sums.

8 Refinement of the results

We can derive an alternate but similar formula for $\pi_k(S_r, N)$ which for large N has a smaller error than (38). Each repeating zero (prime candidate) is in the same position of each τ sequence, with the number of unchanged sequences defined by an expression of the form (29). For the first prime we write:

$$N_1 \sim \frac{N}{\Phi(S_r)(Log(n) + \gamma)} \tag{42}$$

N_1 being the number of unchanged τ_r sequences, with repeating sequence length S_r. Let the starting position of these unchanged τ_r sequences be:

$$P_r^2 + q_i S_r,\ i = 1, 2, \ldots, N_1 \tag{43}$$

The number of primes remaining after the second prime is used is given by:

$$\begin{aligned} N_2 &= \sum_{r=1}^{N_1} (\pi_1(S_r, S_r(q_i + 1) + P_k) - \pi_1(S_r, S_r q_i + P_k)) \\ &\sim \sum_{i=1}^{N_1} \left(\frac{S_r(q_i + 1)}{\Phi(S_r)(Log(S_r(q_i + 1)) + \gamma - Log(S_r))} - \frac{S_r q_i}{\Phi(S_r)(Log(S_r q_i) + \gamma} \right), \end{aligned} \tag{44}$$

Noting that:

$$\begin{aligned} &\sum_{i=1}^{N_1} \left(\frac{S_r(q_i + 1)}{\Phi(S_r)(Log(S_r(q_i + 1)) + \gamma - Log(S_r))} - \frac{S_r q_i}{\Phi(S_r)(Log(S_r q_i) + \gamma} \right) \\ &\to \frac{N_1 S_r}{\Phi(S_r) Log(N_1)},\ \text{as } N_1 \to \infty, \end{aligned} \tag{45}$$

we find in the limit, the position of the individual intervals within the range $(0, N)$ is irrelevant to the final result, so we position the N_1 all together in the middle of the $(0, N)$ range and use the following asymptotic equivalent recurrence relation:

$$N_{k+1} = \frac{N_k}{\Phi(S_r)} \left(\frac{\frac{N + N_k S_r}{2}}{Log\left(\frac{N + N_k S_r}{2}\right) + \gamma} - \frac{\frac{N - N_k S_r}{2}}{Log\left(\frac{N - N_k S_r}{2}\right) + \gamma} \right), k = 1, 2, \ldots, M - 1. \tag{46}$$

For finite N it is clear the value of N_{k+1} is dependent on where the range is placed within $(0, N)$, with a maximum value at the beginning and a minimum at the end. Using the known result that $\pi\left[\frac{N}{2}\right] \sim \pi[N] - \pi\left[\frac{N}{2}\right]$, as $N \to \infty$, we would expect the minimum error to be near the mid point. Both (38), and the recurrence relation (46) with initial value (42), tend to the same asymptotic value as $N \to \infty$.

Evaluating (46) with $N = 10^{18}$, we obtain the result of 808,545,347,142,264 prime pairs, giving a percentage error of .016%. Evaluating the prime quadruplet in Table 1, using a four step recurrence relation defined by (46) reduces the percentage error from -12.3% to -6.8%.

9 Conclusions and remarks

The primary result is that all primes sequences within a $(P_r^2, P_r^2 + S_r - 1)$ range will replicate indefinitely as $N \to \infty$. It is also interesting is that in the limit, all sequences of the same order (same number of primes), have the same asymptotic sum function. The value of the reccurrence relation solution in the previous section, is in its more rapid convergence, but no improvement appears possible in that direction.

Acknowledgements

I am very grateful to Professor Armitage of Durham University, U.K., for looking at an early version of this paper and encouraging me to continue with the research. Thanks also to Associate Professor Rajesh Pereira of the University of Guelph, Canada, for reviewing the paper and indicating where improvements could be made.

References

[1] Everest, G., & Ward, T. (2005) *An Introduction to Number Theory.* Springer-Verlag, London.

[2] Hardy, G.H., & Wright, E.M. (1959) *An Introduction to the Theory of Numbers.* Oxford University Press, London.

[3] Oliveira e Silva, T. (2008). "*Tables of values of pi(x) and of pi2(x)*", Aveiro University. `http://sweet.ua.pt/tos/primes.html` (retrieved 7 January 2011).