On the Spectra of Primes in the Sequence of Least Prime Factors

Tim Samshuijzen
TimSamshuijzen@gmail.com
Wageningen, Netherlands
September 2024

Abstract

In the sequence of least prime factors (OEIS A020639), denoted here as **LPF**, for each prime there is a periodic sequence of occurrences, or **spectrum** as we call it in this paper. In this paper, we explore the symmetries in the spectra of LPF and how they relate to the distribution of primes. From LPF we derive a novel prime-counting function that estimates the number of primes between a prime p and its value squared p^2.

Introduction

The sequence of least prime factors (OEIS A020639), denoted here as **LPF**, is a mapping from the natural numbers to its least prime factor, or 1 in the case of 1. For each prime there is a sequence of occurrences in LPF. A prime's sequence of occurrences, or **spectrum** as we call it, is known to be periodic and symmetric. The period of the spectrum of the n-th prime is equal to the n-th primorial, i.e. $p_n\#$, the product of all primes up to and including the n-th prime. For proof of these periodicities and symmetries, see *Proofs Regarding Primorial Patterns* by Dennis R. Martin, 2006.

In this paper we explore the symmetries in the spectra of LPF, and how these symmetries relate to the distribution of primes.

The Sequence of Least Prime Factors

The sequence of least prime factors, sequence OEIS A020639, denoted here as **LPF**, maps each natural number to the smallest prime that divides that number without leaving a remainder, or 1 when $n = 1$. LPF is the sequence that starts as follows:

1, 2, 3, 2, 5, 2, 7, 2, 3, 2, 11, 2, 13, 2, 3, 2, 17, 2, 19, 2, 3, 2, 23, 2, 5, 2, 3, 2, 29, 2, 31, 2, 3, 2, 5, 2, 37, 2, 3, 2, 41, 2, 43, 2, 3, 2, 47, 2, 7, 2, 3, 2, 53, 2, 5, 2, 3, 2, 59, 2, 61, 2, 3, 2, 5, 2, 67, 2, 3, ...

For each prime there is a set of occurrences in LPF. Such a set of occurrences, or **spectrum**, is periodic. For each prime there is a unique repeating pattern of occurrences. Figure 1 shows conceptually how LPF consists of a repeating pattern of occurrences per prime. Only a small section of LPF is shown, where only the 2s and 3s are visibly repeating. A larger section would show how the following primes also has its own repeating pattern of occurrences.

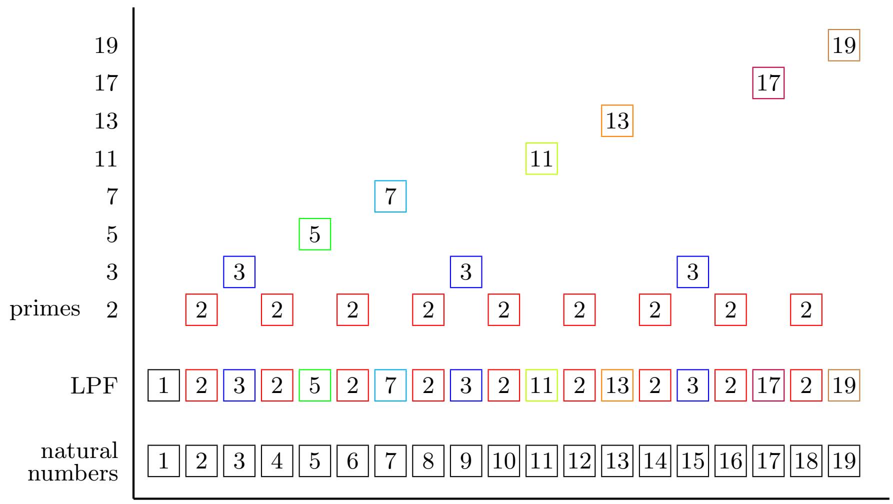

Figure 1: LPF and its spectrum per prime

The period of the n-th prime's spectrum, its **spectral period**, is the product of the prime and the spectral period of the previous prime, which is equivalent to the primorial function $p_n\#$. This primorial rate of growth makes it impossible to study the complete spectra of larger primes. For example, the spectrum of prime 53, the 16th prime, already has a spectral period of 32589158477190046000. The spectrum of prime 211, the 47th prime, has a period that spans in the order of 10^{84}. That number is a thousand times more than the number of electrons in the observable universe. Nevertheless, we can study the spectra of lower primes, and generalize the results to the higher primes.

A way to compactly visualize a prime's spectrum is to show its pattern of "jumps", i.e. the distance between consecutive occurrences. Figure 2 shows the jump pattern for the first spectral period of prime 11. Prime 11 has a spectral period that spans $(2 \cdot 3 \cdot 5 \cdot 7 \cdot 11) = 2310$ integers, and contains 48 occurrences. Each blue vertical bar corresponds with one of its 48 occurrences, and its height corresponds with the distance to its next occurrence. The first bar on the left indicates that, after prime 11's first occurrence, it took 110 integers until its second occurrence, at $11^2 = 121$.

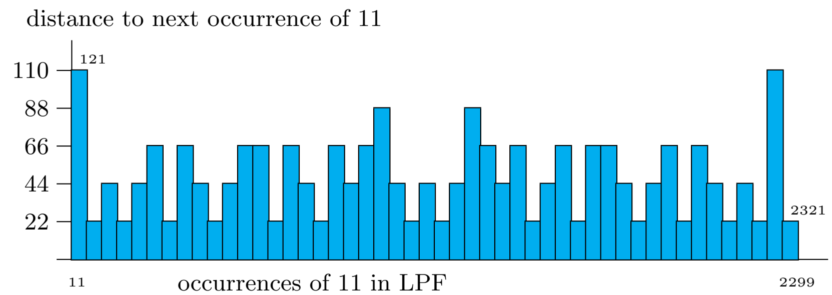

Figure 2: Jump pattern of prime 11

Notice that the pattern in Figure 2 is mirror symmetric, with the axis of symmetry halfway along the spectral period. This symmetry is present in the spectra of all primes. In the next chapter we construct a sieve that generates LPF.

A Sieve for Constructing the Spectra in LPF

Let us construct LPF using a sieving method, where we add the prime spectra one-by-one as rows to a table. Whenever a prime spectrum is added, the fundamental period of the combined spectra, i.e. the width of the table, increases by a factor equal to the added prime. This primorial rate of growth ensures that, after each addition of a prime, LPF remains periodic and symmetric as a whole.

The initial condition is the 2x2 table shown in Figure 3. The top row represents the natural numbers, which in the initial condition is just 1 and 2. The second row represents prime 2 and its occurrences. The occurrences are colored according to their prime, for easy recognition. A letter g is placed above each column that has no occurrences. The letter g stands for "gap", indicating a gap in the combined spectra. The column of natural number 1 always has a gap because 1 divided by a prime is never a natural number.

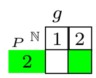

Figure 3: LPF up to prime 2

From here we construct LPF using a recurrence relation, whereby we add the spectrum of each prime one-by-one at each iteration. The recurrence relation for adding the next prime is as follows:

1. Find the first natural number greater than 1 that has a gap, i.e. no occurrence in that column. If there is no gap to be found within the table (which only happens when iterating from the initial condition), look at the progression beyond the table. Let this number p be the next prime.

2. Extend the table such that it has p times more columns, revealing that many more natural numbers.

3. make $p - 1$ copies of the pattern and append them to the right, all the way up to the end of the table.

4. Add a new row to the table. This new row represents prime p. Choose some random available color to represent p.

5. In this row, from left to right, if the cell is a gap and divisible by p, mark this cell with the chosen color for p. These markings represent the occurrences of p in LPF.

   - This procedure is equivalent to: Assign occurrences to all gaps of the previous period, including the gap at 1, and then stretch the positions to the right by a factor p.

6. Add symbols g and S above the columns to indicate gaps and axis of symmetry, respectively.

Let us apply this recurrence relation to the initial condition. As shown in Figure 3, the first gap after 1 is at 3. 3 is just outside the table, which only happens when iterating from prime 2. For all subsequent iterations the first gap is always inside the table. Next, the number of columns is increased by a factor of 3, and the block of all prior occurrences is copied 2 times and appended to the right. Executing the remaining steps of the procedure results in the table shown in Figure 4.

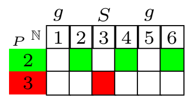

Figure 4: LPF up to prime 3

Figure 4 shows the combined spectra of primes 2 and 3. In this state the total spectral period is 6 integers in length. There are now 2 gaps, one at column 1 and the other at column 5. The symbol S, placed halfway the table, marks the axis of reflection symmetry, where the pattern of occurrences is the same on either side of this column (in the modular sense, such that the pattern is copied indefinitely on either side of the table). The least prime factor beneath S is always 3. This is because S is halfway the spectral period. The spectral period of prime p is the product of all the primes up to and including p, such that when dividing away the 2 the next lowest prime is 3. At any iteration, all occurrences in LPF, besides the occurrence of 3 at S, come in pairs, mirror symmetric on either side of S. The same reflective symmetry also holds for the gaps. The middle column is an axis of symmetry for occurrences and gaps.

The next iteration is shown in Figure 5.

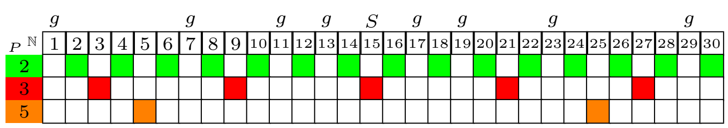

Figure 5: LPF up to prime 5

In Figure 5, notice how all the gap symbols g, except at 1, are above primes. This is not general, prime 5 is the last iteration where each gap corresponds with a prime.

In the next iteration, adding the spectrum of prime 7, the spectral period becomes $7\# = 210$, and the column of symmetry is at 105. This table is too wide to show in this paper, but the situation is well known. For example , see *An upper bound in Goldbach's conjecture*, J.M. Deshouillers, A. Granville, W. Narkiewicz, and C. Pomerance, Math. Comp. 61 (1993), 209–213, a paper that was covered by the YouTube channel *Numberphile* in the episode *210 is VERY Goldbachy*. In the table for prime 7, in the right half of the table there are gaps at composites $121, 143, 169, 187, 209$. If we ignore the outer 121 and 209, the corresponding gap pairs are at $89, 67, 41, 23$. These are the primes that do not pair with other primes to sum to 210, exactly as shown in the video of *Numberphile*.

In general, at any iteration, when adding the spectrum of prime p_n, the number of occurrences of p_n is equal to the number of gaps in the spectra of the previous primes. Furthermore, the new number of gaps is equal to the previous number of gaps, multiplied by $p_n - 1$.

Properties of the Spectra in LPF

The spectral period of prime p_n is the product of p_n and the spectral period of its previous prime. The spectral period of the n-th prime p_n is equal to the primorial function $p_n\#$.

Example
The spectral period of prime 13 is:

$$\begin{aligned} 13\# &= 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \\ &= 30030 \end{aligned}$$

The spectral periods for the primes from 2 to 31 are listed in Table 1. For the rest of the sequence, see the primorial numbers OEIS A002110.

Prime p_n	**$p_n\#$ = Spectral period**
2	2
3	6
5	30
7	210
11	2310
13	30030
17	510510
19	9699690
23	223092870
29	6469693230
31	200560490130

Table 1: The spectral periods of the first 11 primes

The number of occurrences of prime p_n in its spectral period, denoted here as $N_{occ}(p_n)$, is equal to the number of gaps in the combined spectra of its previous primes, denoted here as $N_{gaps}(p_{n-1})$.

$$N_{occ}(p_n) = N_{gaps}(p_{n-1}) \tag{1}$$

$N_{gaps}(p)$ is a function that takes as input prime p, and outputs the number of gaps in the combined spectra of primes up to and including prime p.
$N_{gaps}(p_n)$ is equal to the number of gaps in the combined spectra of the primes less than p_n, multiplied by p_n, minus the number of occurrences of p_n.

$$N_{gaps}(p_n) = \Big(N_{gaps}(p_{n-1}) \cdot p_n\Big) - N_{occ}(p_n) \tag{2}$$

Substituting (4) into (5) yields:

$$N_{gaps}(p_n) = \Big(N_{gaps}(p_{n-1}) \cdot p_n\Big) - N_{gaps}(p_{n-1}) \tag{3}$$

And:

$$N_{gaps}(p_n) = N_{gaps}(p_{n-1}) \cdot (p_n - 1) \tag{4}$$

Which can be written as:

$$N_{gaps}(p_n) = \prod_{q=1}^{q\leq n} (p_q - 1) \tag{5}$$

The ratio between the number of gaps in the combined spectra of primes up to and including prime p_n, and the number of occurrences of prime p_n, is:

$$\frac{N_{gaps}(p_n)}{N_{occ}(p_n)} = p_n - 1 \tag{6}$$

Therefore:

$$N_{gaps}(p_n) = N_{occ}(p_n) \cdot (p_n - 1) \tag{7}$$

Substituting (10) into (4) yields:

$$N_{occ}(p_n) = N_{occ}(p_{n-1}) \cdot (p_{n-1} - 1) \tag{8}$$

The number of occurrences per its spectral period for the primes from 2 to 31 are listed in Table 2. The sequence in the $N_{occ}(p_n)$ column corresponds with OEIS A005867.

Prime p_n	$N_{occ}(p_n)$ **= Occurrences / period**
2	1
3	1
5	2
7	8
11	48
13	480
17	5760
19	92160
23	1658880
29	36495360
31	1021870080

Table 2: The number of occurrences per spectral period for the first 11 primes

Alternatively, somewhat redundant, when $n > 1$, $N_{occ}(p_n)$ can be written as follows.

$$N_{occ}(p) = p_{n-1}\# \cdot \left(1 - \sum_{q=prime}^{q<p} \frac{N_{occ}(q)}{P_!(q)}\right) \tag{9}$$

Having functions for spectral period and number of occurrences per spectral period, we can calculate the average number of occurrences of prime p_n in LPF, which we call the **spectral density** of prime p_n.

Let $D_o(p_n)$ be a function that takes as input a prime p_n, and outputs the spectral density of prime p_n in LPF, defined as:

$$D_o(p_n) = \frac{N_{occ}(p_n)}{p_{n-1}} \tag{10}$$

Examples

$$D_o(2) = \frac{1}{2} = 0.5$$

$$D_o(3) = \frac{1}{6} = 0.166666.$$

$$D_o(5) = \frac{2}{30} = 0.066666.$$

$$D_o(7) = \frac{8}{210} = 0.038095.$$

$$D_o(11) = \frac{48}{2310} = 0.020779.$$

$$D_o(13) = \frac{480}{30030} = 0.015984.$$

$$D_o(17) = \frac{5760}{510510} = 0.011282.$$

$N_{gaps}(p_n)$ corresponds with sequence OEIS A005867, but shifted by 1 when compared with $N_{occ}(p_n)$. Table 3 shows how $N_{occ}(p_n)$ and $N_{gaps}(p_n)$ are related.

Prime p_n	$N_{occ}(p_n) = N_{gaps}(p_{n-1})$	$N_{gaps}(p_n) = N_{occ}(p_n) \cdot (p_n - 1)$
2	1	1
3	1	2
5	2	8
7	8	48
11	48	480
13	480	5760
17	5760	92160
19	92160	1658880
23	1658880	36495360
29	36495360	1021870080
31	1021870080	30656102400

Table 3: The values for N_{occ} and N_{gaps} for the first 11 primes

Let $D_g(p_n)$ be a function that accepts a prime p_n, and outputs the number of gaps in the combined spectra up to prime p_n, per spectral period of p_n, defined as:

$$D_g(p_n) = \frac{N_{gaps}(p_n)}{P_!(p_n)} \tag{11}$$

Table 4 shows the density of gaps when constructing LPF by adding the spectra of primes one-by-one.

Prime p_n	$p_n\#$	$N_{gaps}(p_n)$	$D_g(p_n) = \frac{N_{gaps}(p_n)}{p_{n-1}}$
2	2	1	0.5
3	6	2	0.333.
5	30	8	0.266.
7	210	48	0.228.
11	2310	480	0.207.
13	30030	5760	0.191.
17	510510	92160	0.180.
19	9699690	1658880	0.171.
23	223092870	36495360	0.163.
29	6469693230	1021870080	0.157.
31	200560490130	30656102400	0.152.

Table 4: Values for D_g for the first 11 primes

A gap corresponds with either a prime, a composite, or number 1. Table 5 shows the number of gaps per prime, how many of those gaps correspond with primes, and how many of those gaps correspond with non-primes (composites and 1). The sequence in the Prime gaps column corresponds with OEIS A048862. The sequence in the Non-prime gaps column corresponds with OEIS A048863.

Prime p_n	$N_{gaps}(p_n)$	**Prime gaps**	**Non-prime gaps**
2	1	0	1
3	2	1	1
5	8	7	1
7	48	42	6
11	480	338	142
13	5760	3242	2518
17	92160	42324	49836
19	1658880	646021	1012859
23	36495360	12283522	24211838
29	1021870080	300369786	721500294
31	30656102400	8028642999	22627459401

Table 5: The number of gaps, prime gaps, and non-prime gaps, per spectral period, per prime.

Estimating the Number of Primes Between Prime p_n and p_n^2

In the combined spectra of primes up to prime p_n, in the region between p_n and p_n^2, where the spectral density of p_n is zero, the gaps correspond exactly with all the primes greater than p_n and less than p_n^2. We do not have an exact formula for the number of gaps in this region, but we know the overall gap density. We can expect the gap density in this region to be near greater or equal to the gap density over the period whole $p_n\#$, and near less than or equal to the gap density in the combined spectra of the previous primes over the period $p_{n-1}\#$. We take the average these densities for our estimate.

Let $\pi_{pp^2}^{low}(p_n)$ be a function, which takes as input prime p_n, and outputs the expected number of primes in a region that spans from p_n and less than p_n^2, based on the gap density of the current prime, defined as:

$$\begin{aligned} \pi_{pp^2}^{low}(p_n) &= p_n^2 \cdot \frac{N_{gaps}(p_n)}{p_n\#} \\ &= p_n \cdot \frac{N_{gaps}(p_n)}{p_{n-1}\#} \end{aligned} \tag{12}$$

Let $\pi_{pp^2}^{up}(p_n)$ be a function, which takes as input prime p_n, and outputs the expected number of primes in a region that spans from p_n and less than p_n^2, based on the gap density of the previous prime, defined as:

$$\pi_{pp^2}^{up}(p_n) = p_n^2 \cdot \frac{N_{gaps}(p_{n-1})}{p_{n-1}\#} \tag{13}$$

Our estimation is then the average of $\pi_{pp^2}^{low}(p_n)$ and $\pi_{pp^2}^{up}(p_n)$. Let $\pi_{pp^2}(p_n)$ be a function, which takes as input prime p_n, and outputs an estimation for the number of primes greater than p_n and less than p_n^2, defined as:

$$\begin{aligned} \pi_{pp^2}(p_n) &= \frac{\pi_{pp^2}^{low}(p_n) + \pi_{pp^2}^{up}(p_n)}{2} \\ &= \frac{p_n \cdot \frac{N_{gaps}(p_n)}{p_{n-1}\#} + p_n^2 \cdot \frac{N_{gaps}(p_{n-1})}{p_{n-1}\#}}{2} \\ &= \frac{p_n \cdot N_{gaps}(p_n) + p_n^2 \cdot N_{gaps}(p_{n-1})}{2 \cdot p_{n-1}\#} \end{aligned} \tag{14}$$

Which, when expanded, can be written as:

$$\pi_{pp^2}(p_n) = \frac{p_n \cdot \left(\prod_{q=1}^{q\leq n}(p_q - 1)\right) + p_n^2 \cdot \left(\prod_{q=1}^{q<n}(p_q - 1)\right)}{2 \cdot \prod_{q=1}^{q<n} p_q}$$
$$= \frac{(2p_n^2 - p_n) \cdot \prod_{q=1}^{q<n}(p_q - 1)}{2 \cdot \prod_{q=1}^{q<n} p_q} \tag{15}$$

Expanding this estimation to the whole region starting from 1, as opposed to starting from p_n, we simply use the actual prime counting function $\pi(p_n)$, for calculating the number of primes up to p_n. Let $\pi_{p^2}(p_n)$ be a function that takes as input a prime p_n, and outputs an estimate for the number of primes less than p_n^2, given in:

$$\pi_{p^2}(p_n) = \pi(p_n) + \pi_{pp^2}(p_n)$$
$$= \pi(p_n) + \frac{p_n \cdot N_{gaps}(p_n) + p_n^2 \cdot N_{gaps}(p_{n-1})}{2 \cdot p_{n-1}\#}$$
$$= \pi(p_n) + \frac{(2p_n^2 - p_n) \cdot \prod_{q=1}^{q<n}(p_q - 1)}{2 \cdot \prod_{q=1}^{q<n} p_q} \tag{16}$$

Where π is the actual prime-counting function. In order to use $\pi_{p^2}(p_n)$, one is required to know all the primes up to p_n. Furthermore, $\pi_{p^2}(p_n)$ is only defined for p_n being prime. π_{p^2} relates to the prime counting function π as follows:

$$\pi(p_n^2) = \pi_{p^2}(p_n) + \epsilon \tag{17}$$

Where p_n is any prime, and ϵ is the error in estimation. The distribution of gaps is smooth at the macro scale, such that we can expect ϵ to be proportionally low for all p_n.

Example
How many primes approximately exist less than 7^2?

$$\pi_{p^2}(7) = \pi(7) + \frac{7 \cdot N_{gaps}(7) + 7^2 \cdot N_{gaps}(5)}{2 \cdot 5\#}$$
$$= 4 + \frac{7 \cdot 48 + 7^2 \cdot 8}{2 \cdot 30}$$
$$\approx \mathbf{16.13}$$

Actual value:

$$\pi(49) = \mathbf{15}$$

Example
How many primes approximately exist less than 11^2?

$$\begin{aligned}\pi_{p^2}(11) &= \pi(11) + \frac{11 \cdot N_{gaps}(11) + 11^2 \cdot N_{gaps}(7)}{2 \cdot 7\#} \\ &= 5 + \frac{11 \cdot 480 + 11^2 \cdot 48}{2 \cdot 210} \\ &\approx \mathbf{31.4}\end{aligned}$$

Actual value:

$$\pi(121) = \mathbf{30}$$

Example
How many primes approximately exist less than 13^2?

$$\begin{aligned}\pi_{p^2}(13) &= \pi(13) + \frac{13 \cdot N_{gaps}(13) + 13^2 \cdot N_{gaps}(11)}{2 \cdot 11\#} \\ &= 6 + \frac{13 \cdot 5760 + 13^2 \cdot 480}{2 \cdot 2310} \\ &\approx \mathbf{39.76}\end{aligned}$$

Actual value:

$$\pi(169) = \mathbf{39}$$

Example
How many primes approximately exist less than 17^2?

$$\begin{aligned}\pi_{p^2}(17) &= \pi(17) + \frac{17 \cdot N_{gaps}(17) + 17^2 \cdot N_{gaps}(13)}{2 \cdot 13\#} \\ &= 7 + \frac{17 \cdot 92160 + 17^2 \cdot 5760}{2 \cdot 30030} \\ &\approx \mathbf{60.80}\end{aligned}$$

Actual value:

$$\pi(289) = \mathbf{61}$$

Example
How many primes approximately exist less than 19^2?

$$\begin{aligned}\pi_{p^2}(19) &= \pi(19) + \frac{19 \cdot N_{gaps}(19) + 19^2 \cdot N_{gaps}(17)}{2 \cdot 17\#} \\ &= 8 + \frac{19 \cdot 1658880 + 19^2 \cdot 92160}{2 \cdot 510510} \\ &\approx \mathbf{71.45}\end{aligned}$$

Actual value:

$$\pi(361) = \mathbf{72}$$

Table 6 and Figure 6 shows the estimates calculated in this study. The values for $\frac{p_n}{ln(p_n)}$ (PNT) are included for comparison. With only a limited number of estimations to show for, it is too early for a thorough evaluation, but first impressions show that $\pi_{p^2}(p_n)$ follows $\pi(p_n^2)$ remarkable well. These results should proportionally scale with p_n, because $\pi_{p^2}(p_n)$ builds upon knowledge of p_n's previous primes.

prime p_n	p_n^2	$\pi(p_n^2)$	$\pi_{p^2}(p_n)$	$\frac{p_n^2}{ln(p_n^2)}$
7	49	15	16.1	12.6
11	121	30	31.4	25.2
13	169	39	39.8	32.9
17	289	61	60.8	51.0
19	361	72	71.4	61.3
23	529	99	97.5	84.6
29	841	146	145.2	124.9
31	961	162	160.3	139.9

Table 6: Number of primes less than prime squared vs. estimates

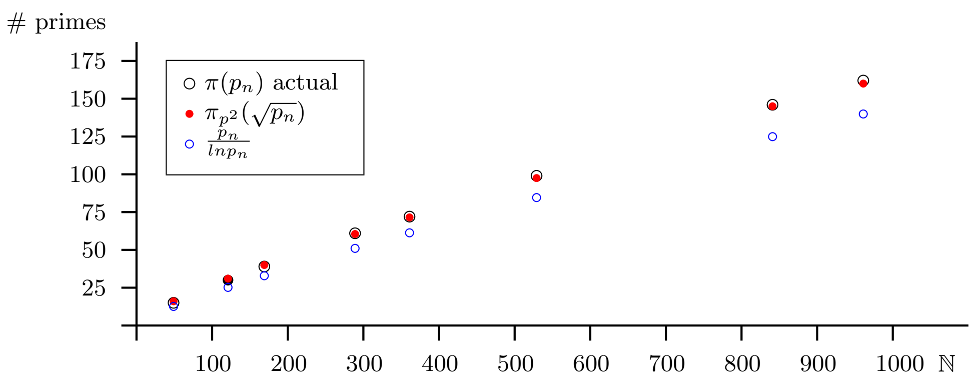

Figure 6: Comparison of prime-counting functions

Conclusion

We conclude LPF to be a useful tool for studying the distribution of primes. Much has already been written about these primorial patterns before, there is nothing new in this paper. This is just an exploratory study. The result for us is that it sets the scene for further study of primorial patterns, to more deeply understand the underlying processes that define and generate the primes.