

Prime numbers and the (double) primorial sieve.

1. The quest.

This research into the distribution of prime numbers started by placing a prime number on the short leg of a Pythagorean triangle. The properties of the Pythagorean triples imply that prime numbers $> p_4$ (with $p_4 = 7$) are not divisible by $p_i \in \{2, 3, 5, 7\}$. These divisors have a combined repeating pattern based on the primorial $p_4\#$, the product of the first four prime numbers. This pattern defines the fourth (double) primorial sieve.

The $P_4\#$ -sieve is the first sieve with all the properties of the infinite set of primorial sieves. It also gives an explanation for the higher occurrence of the (9, 1) last digit gap among prime numbers.

The (double) primorial sieve has two main functions.

The primorial sieve is a method to generate a consecutive list of prime numbers, since the base of the n -th primorial sieve contains all prime numbers $< p_n\#$.

The double primorial sieve offers preliminary filtering of natural numbers when they are sequential stacked above the base of the sieve. Possible prime numbers are only found in the columns supported by the $\varphi(p_n\#)$ struts.

Each strut is relative prime to $p_n\#$ and is like a support beam under a column.



Fig. 1: Examples of imaginary struts to support a specific column.

2. The principal of the (double) primorial sieve.

The width of the primorial sieve $P_n\#$ -sieve is determined by the primorial $p_n\#$, the product of the first n prime numbers. All natural numbers $g > p_n\#$ form a matrix of infinite height when stacked sequential on top of the base of the sieve (Fig. 2abc). The $\varphi(p_n\#)$ struts support the columns that contain possible prime numbers. Natural numbers $> p_n\#$ in the unsupported columns are filtered and definitively composite numbers. For the struts S_j applies that $\gcd(S_j, p_n\#) = 1$ with $1 \leq j \leq \varphi(p_n\#)$ and $\varphi(m)$ Euler's totient function. The list of all prime numbers $< p_n\#$ consist of the prime numbers $\leq p_n$ and the non-composite struts.

Each primorial sieve is constructed from the previous sieve and therefore from all previous sieves. In order to generate the $P_1\#$ -sieve the $P_0\#$ -sieve is defined, based on the "prime number" $p_0 = 1$. The $P_0\#$ -sieve (Fig. 2a) has a width of $p_0\# = 1$ and $\varphi(p_0\#) = 1$ strut. All natural numbers $g > p_0\#$ are situated above the one strut S_1 and are possible prime; there is no filtering.

The $P_1\#$ -sieve (Fig. 2b) has a width of $p_1\# = p_1 = 2$ and filters out the even numbers via $g \equiv 0 \pmod{p_1}$. Natural numbers $g > p_1\#$ above the $\varphi(p_1\#) = 1$ strut S_1 are possible primes.

The $P_2\#$ -sieve (Fig. 2c) has a width of $p_2\# = p_2 \cdot p_1\# = 6$ and filters $g \equiv 0 \pmod{p_2}$. Natural numbers $g > p_2\#$ above the $\varphi(p_2\#) = 2$ struts $S(p_2\#)_i \in \{S_1, S_2\}$ comply with $\gcd(g, p_2) = 1$ as well as $\gcd(g, p_2\#) = 1$ due to the construction via all previous sieves. Possible prime numbers $g > p_2\#$ are of the form $6a \pm 1$ with $a \in \mathbb{N}^+$.

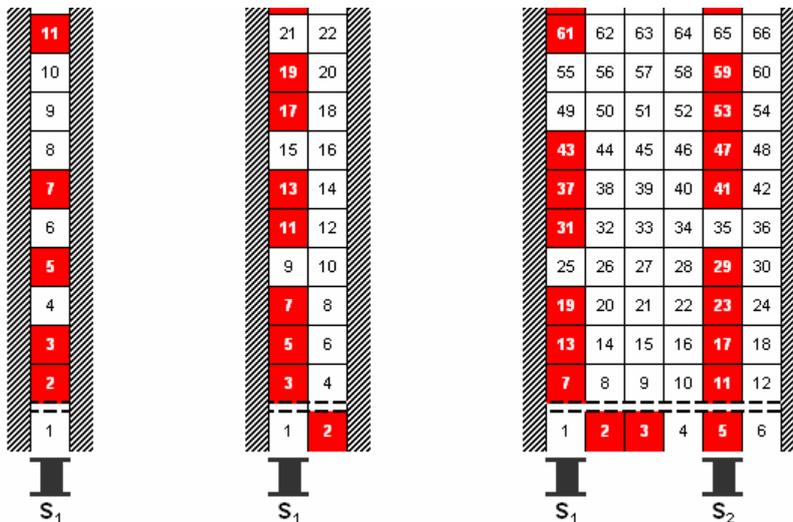


Fig. 2abc: The double primorial sieves: $P_0\#$ -sieve, $P_1\#$ -sieve and $P_2\#$ -sieve.

The third (double) primorial sieve.

From the $P_3\#$ -sieve onwards there is an unambiguous algorithm to build the struts from the previous sieve. The $S(p_2\#)_i$ struts of the $P_2\#$ -sieve are repeated p_3 times, so that $S(p_3\#)_k = S(p_2\#)_i + m \cdot p_2\#$ with $0 \leq m < p_3$. Struts with $S(p_3\#)_k \equiv 0 \pmod{p_3}$ are removed (Fig. 3) and thus (multiples of) $S(p_3\#)_2$.

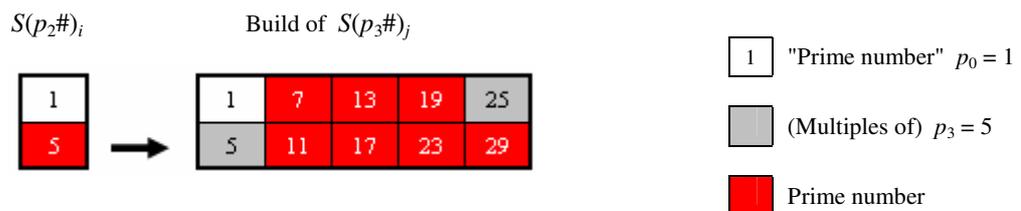


Fig. 3: $P_3\#$ -sieve: generating the struts.

The $P_3\#$ -sieve has a width of $p_3\# = p_3 \cdot p_2\# = 30$ and $\varphi(p_3\#) = 8$ struts. The $P_3\#$ -sieve provides the list of prime numbers $< p_3\#$ consisting of the prime numbers $p_i \in \{2, 3, 5\}$ and the struts S_j that satisfy $\gcd(S_j, p_3\#) = 1$ with $1 < j \leq \varphi(p_3\#)$. Note that with the third primorial sieve all struts > 1 are prime numbers. Potential prime numbers $> p_3\#$ are situated above the struts and meet both $\gcd(g, p_3) = 1$ and $\gcd(g, p_3\#) = 1$ (Fig. 4a).

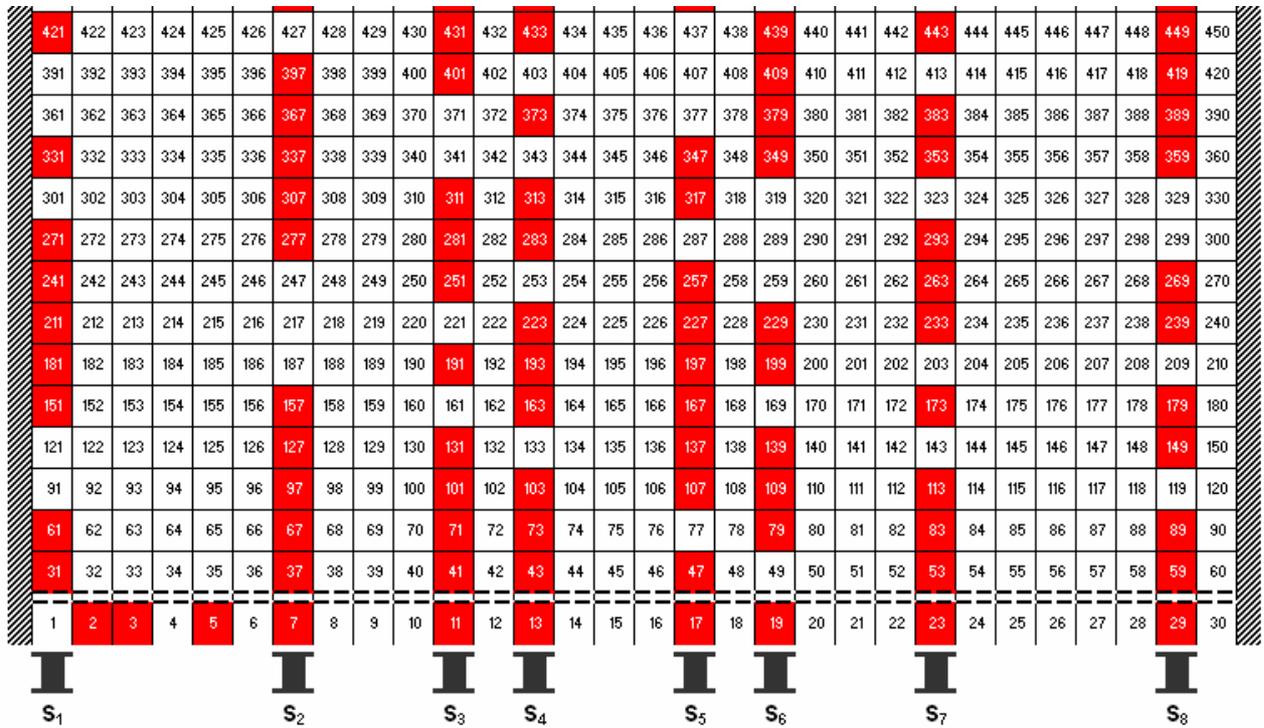


Fig. 4a: The third double primorial sieve.

The $P_3\#$ -sieve has many similarities with the Wheel Factorization method of Paul Pritchard. Fig. 4b shows a wheel with the inner circle formed by the first 30 natural numbers, and thus with a $p_3\# = 30$ base. The spokes of the wheel that contain possible prime numbers have the same functionality as the columns above the struts of the primorial sieve. The graphical representation of the wheel is in this case more concrete. Clearly visible is the symmetry of the spokes in $p_3\# / 2$.

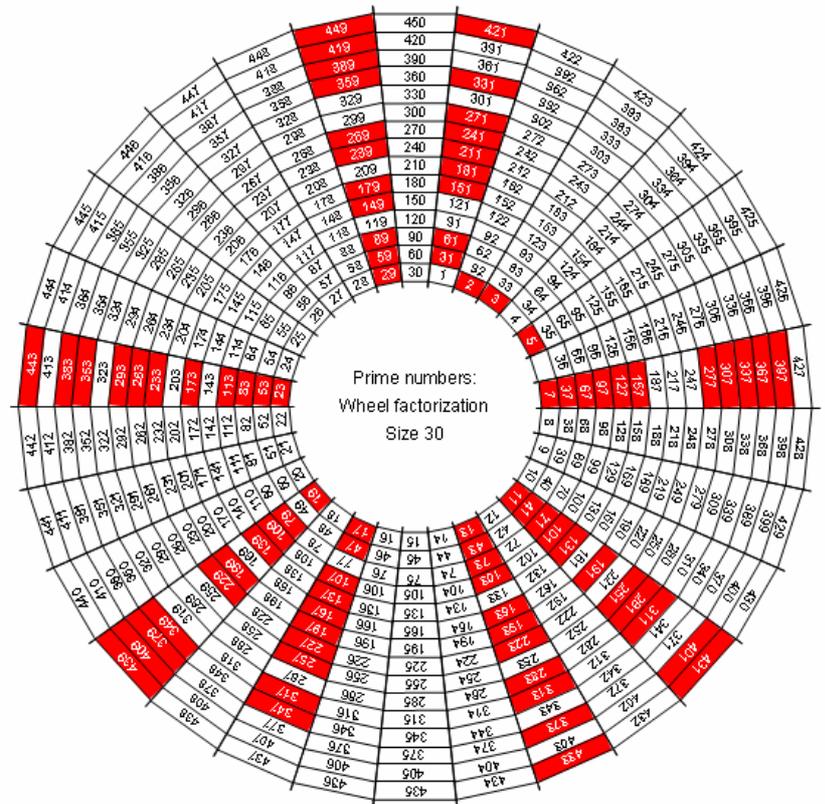


Fig. 4b: Wheel factorization with size 30.

The fourth double primorial sieve.

The struts of the $P_4\#$ -sieve are built via the $P_3\#$ -sieve by repeating the pattern of the $\varphi(p_3\#) = 8$ struts p_4 times, so that $S(p_4\#)_k = S(p_3\#)_i + m \cdot p_4$ with $0 \leq m < p_4$. The struts to be removed with $S(p_4\#)_k \equiv 0 \pmod{p_4}$, consist of $p_4 \cdot S(p_3\#)_i$ with $1 \leq i \leq \varphi(p_3\#)$ (Fig. 5). The $P_4\#$ -sieve has $j = \varphi(p_4\#) = (p_4 - 1) \cdot \varphi(p_3\#) = 48$ struts. The $P_4\#$ -sieve is the first sieve with all the characteristics of the double primorial sieve, since the struts are no longer exclusively prime numbers, for example $S(p_4\#)_{39} = 169 = 13 \cdot 13$.

To generate the list $> p_4$ with all prime numbers $< p_4\#$ from the struts of the $P_4\#$ -sieve the composite struts $S(p_4\#)_j$ with $j \in \{28, 33, 39, 43, 48\}$ are marked negative (Fig. 5).

These composite struts are found via $S(p_4\#)_j \cdot S(p_4\#)_i < p_4\#$ with $i, j > 1$ and $S(p_4\#)_j \leq S(p_4\#)_i$.

Thus: $11 \cdot 11 = 121$, $11 \cdot 13 = 143$, $11 \cdot 17 = 187$, $11 \cdot 19 = 209$ and $13 \cdot 13 = 169$.

The prime numbers $\leq p_4$ plus the non-composite struts $> p_4$ supply the list of the 46 prime numbers $< p_4\#$.



Fig. 5: $P_4\#$ -sieve: building the struts

Fig. 6 shows the equal distribution of the $\pi(10^9) = 50,847,534$ prime numbers above the struts of the $P_4\#$ -sieve, with a deviation relative to $\pi(10^9) / \varphi(p_4\#)$ of less than 0,05%. Among the struts of the $P_4\#$ -sieve the influence is still visible of the repeated pattern of the 8 struts $\{1, 7, 11, 13, 17, 19, 23, 29\}$ from the $P_3\#$ -sieve.

The distance d between S_1 and S_2 of the $P_4\#$ -sieve equals to $d = S(p_4\#)_2 - S(p_4\#)_1 = p_5 - p_0 = 11 - 1 = 10$.

This is the biggest gap between the struts (Fig. 5 and Fig. 6). Due to the symmetry in $(p_4\# / 2)$ the distance d is also found between the second to last and the last strut of the sieve.

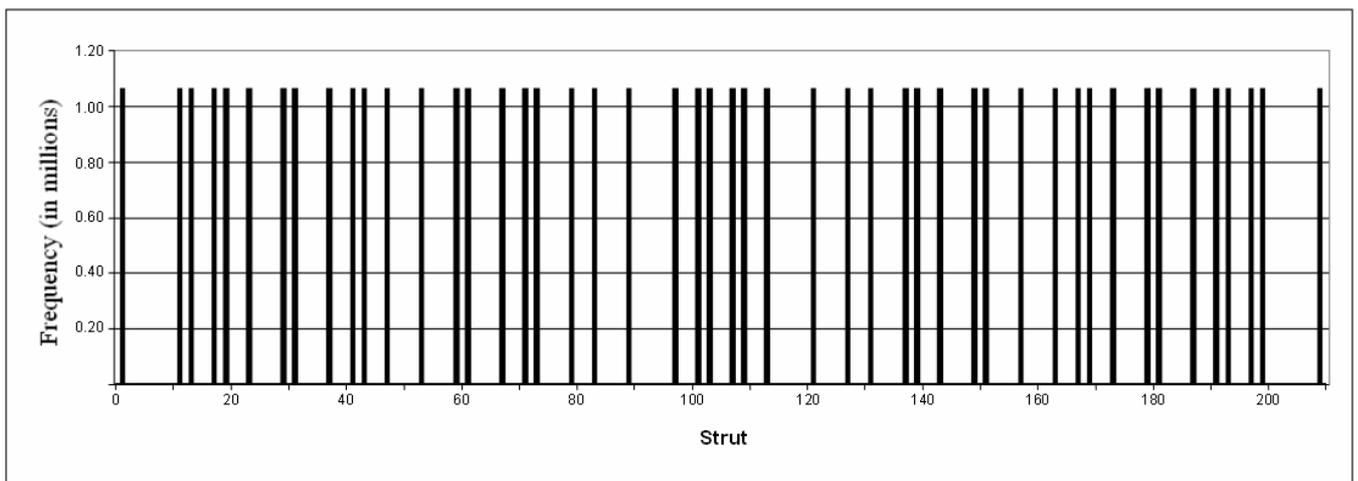


Fig. 6: $P_4\#$ -sieve: distribution of prime numbers $< 10^9$.

3. The double primorial sieve in practical applications.

3.1.1. Example of how to implement the $P_4\#$ -sieve.

Each double primorial sieve with the base of $p_n\#$ can be used to determine whether a natural number g is a prime number. The following distinction is made:

- A number $g \leq p_n$ is a prime number when $g \in \{p_1, p_2, \dots, p_n\}$
- A number $p_n < g \leq p_n\#$ is a prime number when $g \in \{S(p_n\#)_j \mid \gcd(g, p_k) = 1 \text{ for all } p_n < p_k \leq \sqrt{g}\}$.
The values of p_k correspond with the struts $p_n < S(p_n\#)_j \leq \sqrt{g}$ that are not composite.
- A number $p_n\# < g \leq (p_n\#)^2$ can be a prime number when $g \pmod{p_n\#} \in \{S(p_n\#)_j\}$, and thus $g \in \{S(p_n\#)_j + m \cdot p_n\# \mid 1 \leq j \leq \varphi(p_n\#) \wedge m \in \mathbf{N}^+\}$
The number g is a prime number as further applies that $\gcd(g, p_k) = 1$ for all $p_n < p_k \leq \sqrt{g}$.
The values of p_k correspond with the struts $p_n < S(p_n\#)_j \leq \sqrt{g}$ that are non-composite.
- A number $g > (p_n\#)^2$ can be a prime number when $g \pmod{p_n\#} \in \{S(p_n\#)_j\}$.
Checking for possible prime of $g > (p_n\#)^2$ via $p_k \mid g$ for all $p_n < p_k \leq \sqrt{g}$ can be approximated by $d \mid g$ with $d \in \{S(p_n\#)_j + m \cdot p_n\# \mid 1 \leq j \leq \varphi(p_n\#) \wedge m \in \mathbf{N}_0\}$
For $d < p_n\#$ division by the struts $> p_n$ that are non-composite suffices.

To determine whether $g \pmod{p_n\#} \in \{S(p_n\#)_j\}$ the property can be used that with bigger $P_n\#$ -sieves the $\varphi(p_n\#)$ struts $S(p_n\#)_j$ are almost proportional split over the $p_n\#$ -base.

For example: Fig. 7a is based on the $P_6\#$ -sieve with $p_6\# = 30030$ and $(p_6\#)^2 \approx 9 \cdot 10^8$.

The natural number $g_1 = 20893$ is near the strut $g_1 \cdot \varphi(p_6\#) / p_6\# = 4007$ of the $P_6\#$ -sieve and ultimately coincides with strut 4009, that is marked as composite. The prime factors are not found.

The natural numbers $g_2 = g_1 + 10^3 \cdot p_6\# < (p_6\#)^2$ and $g_3 = g_1 + 10^5 \cdot p_6\# > (p_6\#)^2$ with $g_2 \equiv g_3 \equiv g_1 \pmod{p_6\#}$ are possible prime, since both numbers coincides with the strut 4009 (even though the strut is composite).

The list of prime numbers up to $\sqrt{g_2}$ from the base of the sieve suffices to find a prime divisor of g_2 .

Examining g_3 starts with all prime numbers from the base of the sieve, e.g. the struts $p_6 < S(p_6\#)_j \leq p_6\#$ that are non-composite. Prime divisors $> p_6\#$ are approximated by $S(p_6\#)_j + m \cdot p_6\#$ with $m \geq 1$.

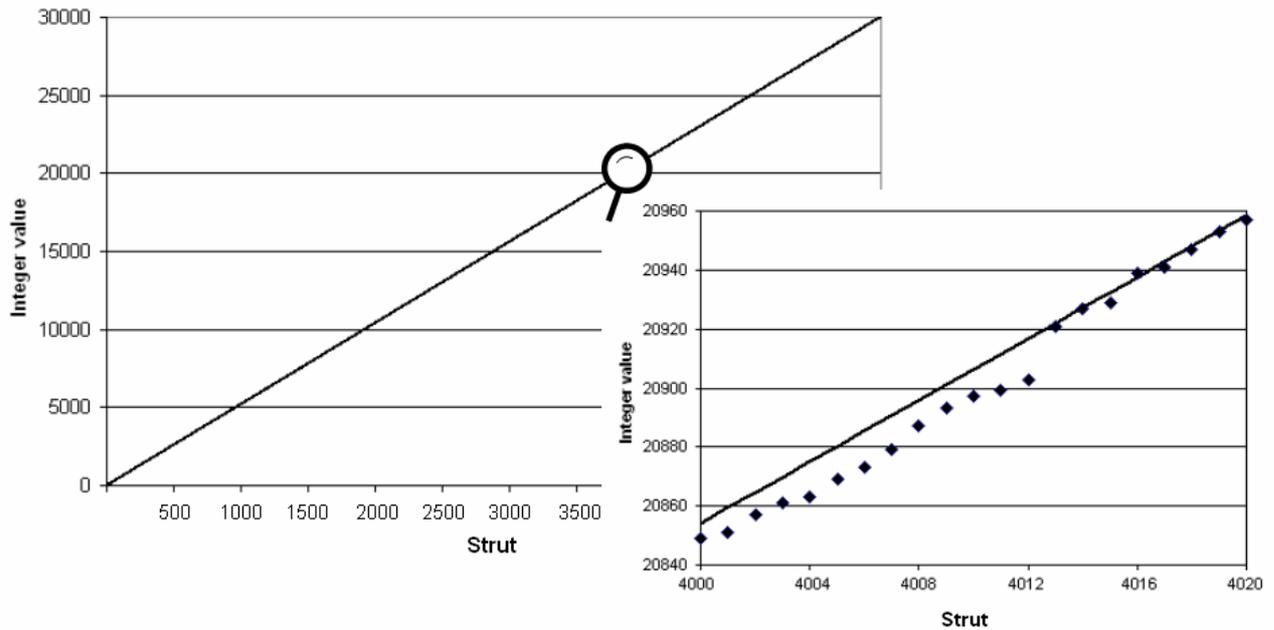


Fig. 7a: $P_6\#$ -sieve: the distribution of the struts over the base.

3.2.1. The gap between prime numbers.

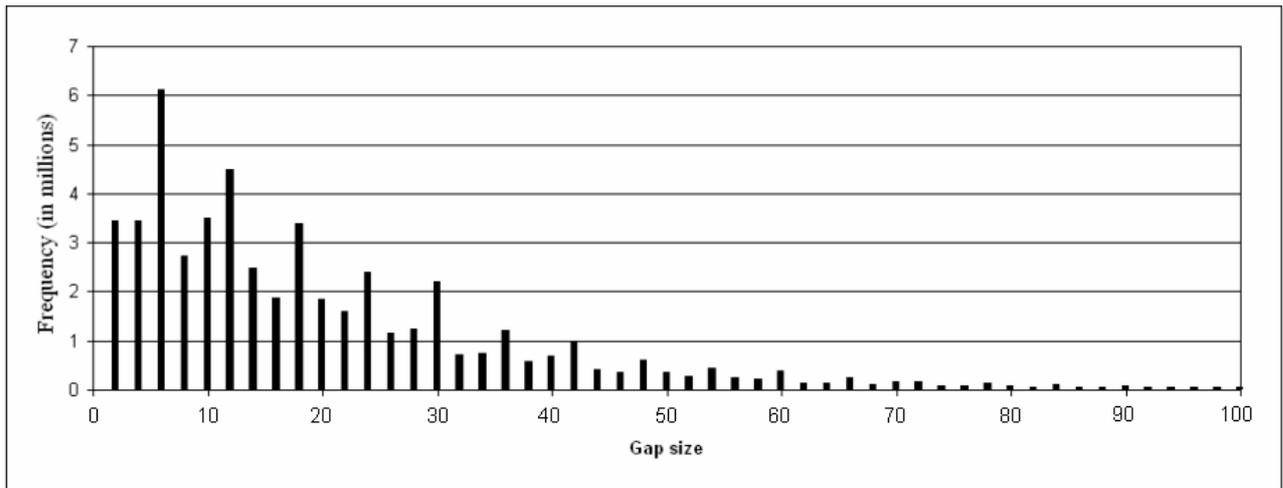


Fig. 8a: Distribution of the prime gaps for all prime numbers up to 10^9 .

Fig. 8a shows the distribution of the prime gaps (difference between two consecutive prime numbers) for all prime numbers $< 10^9$. The bar chart gives local maxima at multiples of 6.

The $P_2\#$ -sieve divides possible prime numbers into two groups above the struts S_1 and S_2 , based on $6a \pm 1$ with $a \in \mathbf{N}^+$ (fig. 2c). Each group is split $p_3 = 5$ ways when building the $P_3\#$ -sieve. One of the new columns is (a multiple of) p_3 and contains no prime numbers $> p_3$ (fig. 3 and 4a).

Fig. 8b shows the combined distribution of the prime gaps per separate group, with each group contributing equally. The ratio of prime numbers in $f(a) = 6a + 1$ corresponds with three times the ratio of prime numbers in $f(n) = n$. Clearly visible is the $\varphi(p_2\#) / \varphi(p_3\#)$ increase at (multiples of) $p_3\# = 30$. The higher frequency at gap = 42 is credited to a higher sieve.

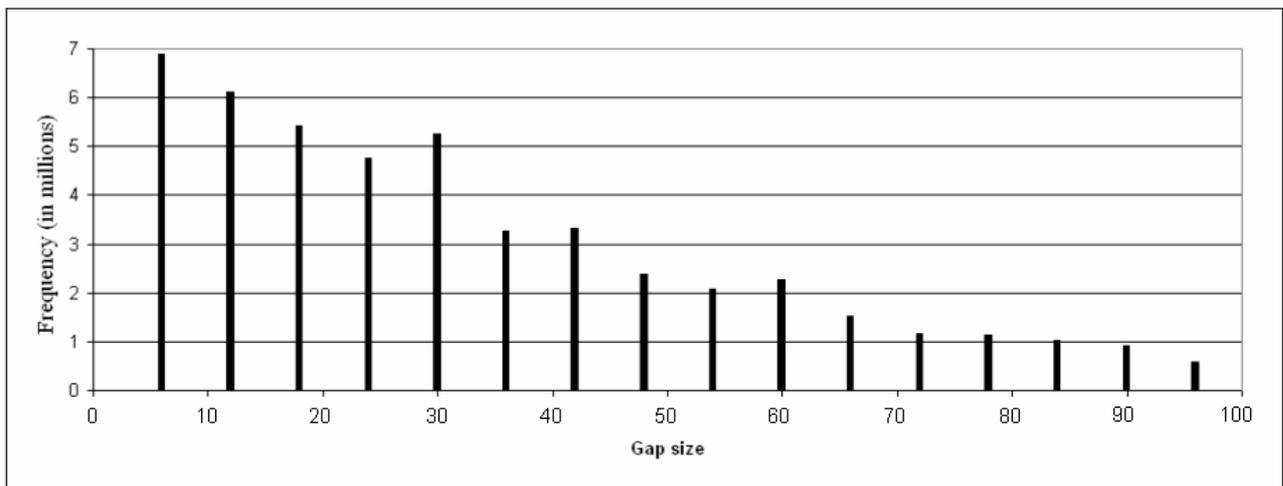


Fig. 8b: Distribution of the prime gaps for all prime numbers up to 10^9 , when split into the $6a \pm 1$ groups.

Fig. 8c gives the distribution of prime numbers $< 10^9$ above the struts of the $P_4\#$ -sieve. The columns are set apart based on $6a + 1$ and $6a - 1$ with $a \in \mathbf{N}^+$ (see also Fig. 6). Together the struts show no apparent pattern, simultaneously they are uniquely mirrored in $p_4\# / 2$. Dividing the prime numbers into two groups supports the conjecture that twin primes are not related by a common denominator.

At the same time the $P_2\#$ -sieve stipulates that all twin primes $> p_2$ are of the form $6a \pm 1$ (fig. 2c). Based on the $P_3\#$ -sieve twin primes $> p_3$ are narrowed down to the form $(30a + 12m) \pm 1$ with $a \in \mathbf{N}_0$ and $-1 \leq m \leq 1$ (Fig. 4b).

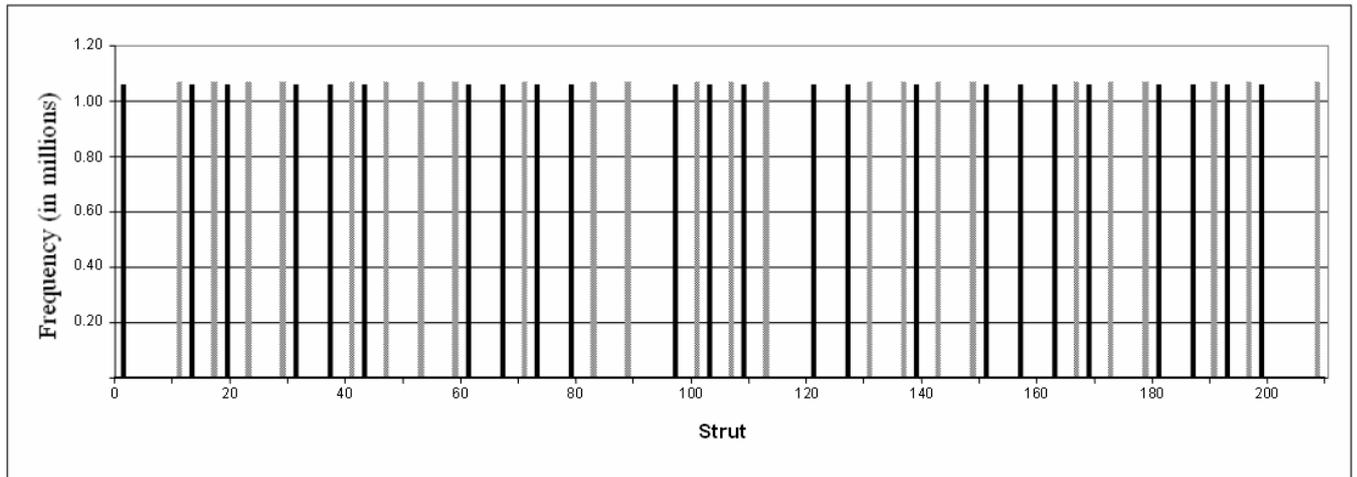


Fig. 8c: $P_4\#$ -sieve: distribution of prime numbers $< 10^9$ based on $6a \pm 1$.

3.3. The Last digit gap conundrum.

In 2016, Stanford number theorists, Robert Lemke Oliver and Kannan Soundararajan discovered that the first hundred million consecutive prime numbers, e.g. $\pi(m) = 10^8$, end less frequently with the same digit than other digits. The last digit gap (9, 1) is favored (Fig. 9a).

This find is exceptional because prime numbers have no last digit preference.

	$ld_2 = 1$	$ld_2 = 3$	$ld_2 = 7$	$ld_2 = 9$
$(1, ld_2)$	4,6%	7,4%	7,5%	5,4%
$(3, ld_2)$	6,0%	4,4%	7,0%	7,5%
$(7, ld_2)$	6,4%	6,8%	4,4%	7,4%
$(9, ld_2)$	8,0%	6,4%	6,0%	4,6%

Fig. 9a: Last digit gap occurrence in the first hundred million consecutive prime numbers.

3.3.1. An explanation for the Last digit gap uneven distribution.

An explanation for this uneven distribution of the last digit gap among prime numbers follows from the build of the $P_4\#$ -sieve. Fig. 9b shows the last digit gap occurrence in the struts of the $P_3\#$ -sieve and the $P_4\#$ -sieve. Included are the (9, 1)-combinations of the last strut with the repetition of the first strut, that follows when generating the next primorial sieve.

The $P_3\#$ -sieve has no consecutive struts that end with the same last digit. The symmetrical distributed $\varphi(p_3\#) = 8$ struts of the $P_3\#$ -sieve are repeated p_4 times. Multiples of p_4 are removed, generating newly last digit combinations (1, 1) and (9, 9) (Fig. 5).

	$P_3\#$ -sieve				$P_4\#$ -sieve			
	$ld_2 = 1$	$ld_2 = 3$	$ld_2 = 7$	$ld_2 = 9$	$ld_2 = 1$	$ld_2 = 3$	$ld_2 = 7$	$ld_2 = 9$
$(1, ld_2)$	0,0%	12,5%	12,5%	0,0%	2,1%	10,4%	12,5%	0,0%
$(3, ld_2)$	0,0%	0,0%	12,5%	12,5%	2,1%	0,0%	10,4%	12,5%
$(7, ld_2)$	12,5%	0,0%	0,0%	12,5%	10,4%	4,2%	0,0%	10,4%
$(9, ld_2)$	12,5%	12,5%	0,0%	0,0%	10,4%	10,4%	2,1%	2,1%

Fig. 9b: Last digit gap occurrence within the struts of the $P_3\#$ -sieve and the $P_4\#$ -sieve.

Building the struts of the $P_5\#$ -sieve using the struts of the $P_4\#$ -sieve gives eight times the unique last digit sequence {9, 9, 1, 1} (Fig. 9c). Prime numbers above struts with this sequence have in the beginning a higher chance at a last digit gap combination (9, 1). The influence of this {9, 9, 1, 1} sequence is inherited by every next sieve.

	199	409	619	829	1039	1249	1459	1669	1879	2089	2299
	209	419	-629	839	1049	1259	-1469	-1679	1889	2099	2309
1	211	421	631	-841	1051	-1261	1471	-1681	-1891	2101	2311
11	-221	431	641	-851	1061	-1271	1481	-1691	1901	2111	2321

1 "Prime number" $p_0 = 1$

(Multiples of) $p_5 = 11$

Prime number

Composite number

Fig. 9c: The last digit gap sequence {9, 9, 1, 1} in the struts of the $P_5\#$ -sieve.

3.3.2. Investigating the difference in the Last digit gap frequency.

In 2016, Robert Lemke Oliver and Kannan Soundararajan were caught off guard by quirks in the final digits of primes.

In the first hundred million consecutive prime numbers, e.g. $\pi(m) = 10^8$ with $m \approx 2 \cdot 10^9$, the number theorists found variations that far exceeded expectations.

See the table which is also displayed in Fig. 9a above.

# of prime numbers	100,000,000			
	$ld_2 = 1$	$ld_2 = 3$	$ld_2 = 7$	$ld_2 = 9$
$(1, ld_2)$	4.6%	7.4%	7.5%	5.4%
$(3, ld_2)$	6.0%	4.4%	7.0%	7.5%
$(7, ld_2)$	6.4%	6.8%	4.4%	7.4%
$(9, ld_2)$	8.0%	6.4%	6.0%	4.6%

Investigations into the uniqueness of this find led to Fig. 9d.

The graph shows the frequencies of the Last digit gaps for several values of m , with m the number of consecutive natural numbers. The displayed curves stabilize after the erratic behaviour up to around $m = 10^5$, due to the buildup of the $P_n\#\text{-sieves}$.

It is the conjecture that all curves ultimately converge to the expected value of 6,25% (e.g. $1/16$) and thus:

Conjecture: "prime numbers have no last digit preference".

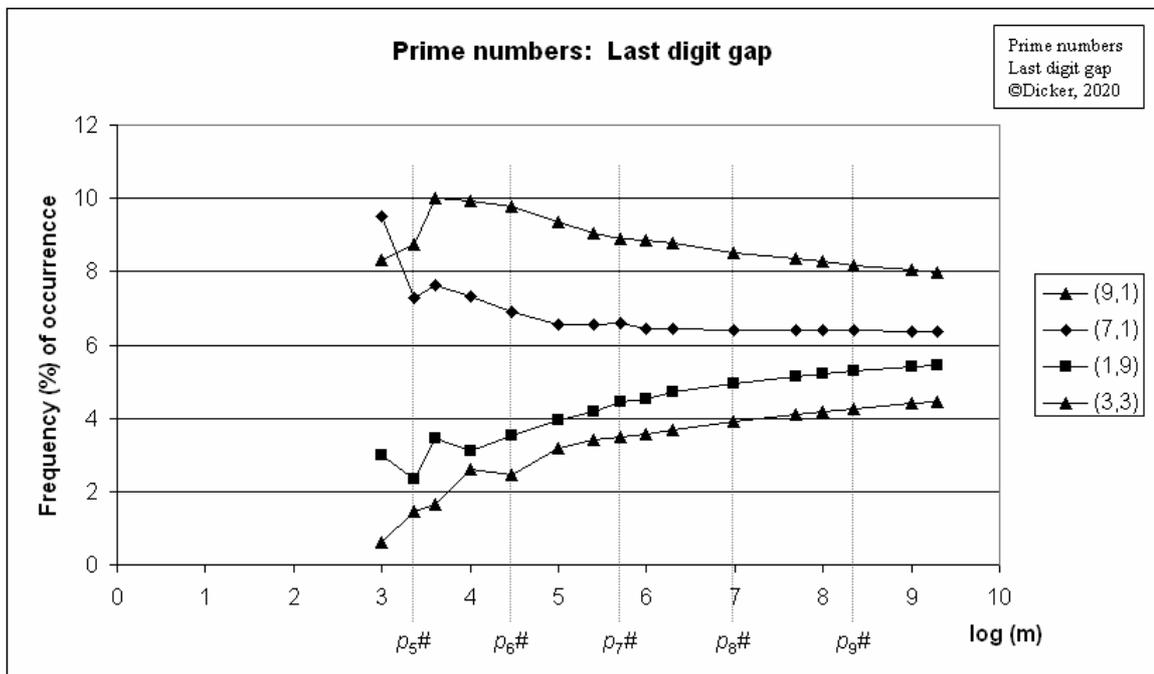


Fig. 9d: Curves with the Last digit gap occurrence up to the first hundred million consecutive prime numbers.

4. Results of the (double) primorial sieve.

The primorial sieve consists of the infinite set of $P_n\#$ -sieves. The width of the sieve is equal to the primorial $p_n\#$, the product of the first n prime numbers. All natural numbers sequential arranged on top of the base of the sieve form together a matrix of infinite height.

The $\varphi(p_n\#)$ struts $S(p_n\#)_j$ of the sieve support the columns above which potential prime numbers $g > p_n$ are located, that comply with $g \bmod(p_n\#) \in \{ S(p_n\#)_j \mid 1 \leq j \leq \varphi(p_n\#) \}$. Non-prime numbers with $\gcd(g, p_n\#) \neq 1$ are filtered through the holes in the sieve.

The base of the primorial sieve contains the list of consecutive prime numbers. The discrete upper limit can be extended infinitely far by building ever increasing $P_n\#$ -sieves from the previous sieve. The full list $> p_n$ with prime numbers $< p_n\#$ appears when in the last sieve all composite struts are omitted.

This cleanup is done via $S(p_n\#)_j \cdot S(p_n\#)_i < p_n\#$ with $i, j > 1$ and $S(p_n\#)_j \leq S(p_n\#)_i$ and not as with the Sieve of Eratosthenes via multiples of all prime numbers $< \sqrt{p_n\#}$.

The double primorial sieve is a method for preliminary filtering of potential prime numbers within all natural numbers. Of the infinite set of natural numbers only $\varphi(p_n\#) / p_n\#$ numbers are left that could be a prime number. For the final check of a potential prime number $g > p_n$ the division by prime divisors $d \leq \sqrt{g}$ can be approximated by $d \in \{ S(p_n\#)_j + m \cdot p_n\# \mid 1 \leq j \leq \varphi(p_n\#) \wedge m \in \mathbb{N}_0 \}$.

From the $P_4\#$ -sieve onwards the $\varphi(p_n\#)$ struts are almost directly proportional split over the base. There is no need for an indexed array, or binary search.

The double primorial sieves (specifically the $P_4\#$ -sieve) offers a platform to further investigate the distribution of prime numbers. Dividing prime numbers into separate groups based on $6a \pm 1$ with $a \in \mathbb{N}^+$ might offer an opening.

The $P_9\#$ -sieve with a base of $p_9\# = 223,092,870$ is the last sieve where 4 Byte integers suffices in the calculations. By making the $P_9\#$ -sieve five times wider all prime numbers $< 10^9$ can be efficiently loaded in memory.

The stretched sieve uses one sequential array with length $5 \cdot p_9 \cdot \varphi(p_8\#) \approx 200 \cdot 10^6$ and 800 MB internal memory.

The effectiveness of the double primorial sieve is the ratio $\varphi(p_n\#) / p_n\#$, the number of struts (or spokes in the wheel).

n	p_n	$p_n\#$	Number of struts $\varphi(p_n\#)$	Ratio $\varphi(p_n\#) / p_n\#$	Comment
1	2	2	1	0.5000	odd integers
2	3	6	2	0.3333	integers $6a \pm 1$
3	5	30	8	0.2667	
4	7	210	48	0.2296	
...	...				
9	23	223,092,870	36,495,360	0.1636	$P_{10}\#$ -sieve: ratio is 0.1579

References:

- [1] <http://en.wikipedia.org> & <http://www.wikipedia.nl>
- [2] Dubner, H. (1987). Factorial and Primorial Primes. *J. Rec. Math.* **19**, 197-203.
- [3] Dubner, H. (1989), A new primorial prime, *J. Rec. Math.*, **21**, 276.
- [4] Croft; www.primesdemystified.com
- [5] Glöggler, G. (2007-2008). Prime numbers in groups; www.g-gloeggler.de
- [6] Diedrich, K.: Primzahl-Lücken-Verteilung ohne Geheimnisse? www.primzahlen.de
- [7] Dupaquis, V.: Computer-implemented method for fast generation and testing of probable prime numbers for cryptographic applications; Publication Number: US20030235299
- [8] Pritchard, P. (1981). A sublinear additive sieve for finding prime numbers. *Communications of the ACM*, **24**(1):18-23.
- [9] Pritchard, P. (1982). Explaining the wheel sieve. *Acta Informatica*, **17**:477-485.
- [10] Atkin, A. & Bernstein, D. (2004). Prime sieves using binary quadratic forms. *Math. Comp.*, **73**:1023-1030.
- [11] Lemke Oliver, R. & Soundararajan, K. (2016). Unexpected Biases in the Distribution of Consecutive Primes. *arXiv*, March 15.