Discovering

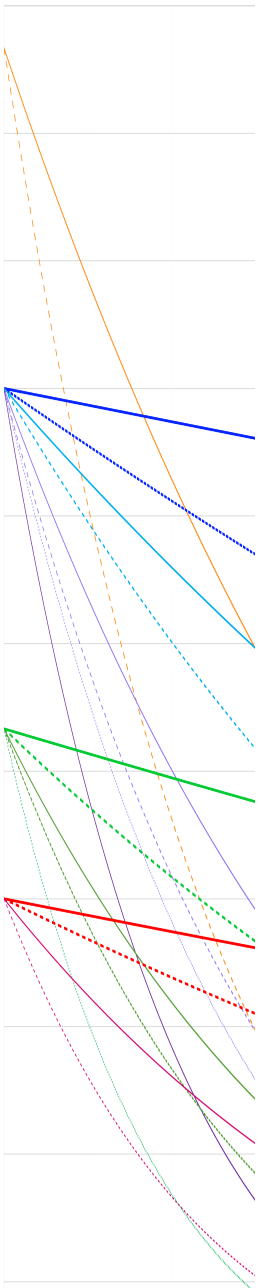
Patterns among the Primes

A book of exercises

Fred B. Holt

https://www.primegaps.info

May 2024



Objective: to study Eratosthenes sieve as a discrete dynamic system in order to develop models for the relative populations of the gaps among the candidate prime numbers, across stages of the sieve.

Edition 9 May 2024

Patterns among the Primes
Fred B. Holt
https://www.primegaps.info

1. Eratosthenes Sieve

List the natural numbers (up to some bound *N*).

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 …

Mark '*1*' as the unit and repeat these two steps:

E1. Mark the smallest unmarked number as the next prime *p*, and

E2. Cross off all multiples of *p* from the list

Until every number on the list has been marked as a prime or crossed off.

This is the stage of Eratosthenes sieve for *p=2.*

1 2 3 ~~4~~ 5 ~~6~~ 7 ~~8~~ 9 ~~10~~ 11 ~~12~~ 13 ~~14~~ 15 ~~16~~ 17 ~~18~~ 19 ~~20~~ 21 ~~22~~ 23 ~~24~~ 25 ~~26~~ 27 ~~28~~ 29 ~~30~~ 31 ~~32~~ 33 …

This is the stage of Eratosthenes sieve for *p=3.*

1 2 3 ~~4~~ 5 ~~6~~ 7 ~~8~~ ~~9~~ ~~10~~ 11 ~~12~~ 13 ~~14~~ ~~15~~ ~~16~~ 17 ~~18~~ 19 ~~20~~ ~~21~~ ~~22~~ 23 ~~24~~ 25 ~~26~~ ~~27~~ ~~28~~ 29 ~~30~~ 31 ~~32~~ ~~33~~ …

Unit

Confirmed primes

Remaining candidate primes

Exercises:

1. Perform Eratosthenes sieve for the bound *N=100.*
2. In the stage of the sieve for *p,* show that p^2 is the smallest multiple of *p* to be crossed off in step E2.

2. Cycles of gaps $\mathcal{G}(p\#)$

At each stage of Eratosthenes sieve, there is a corresponding *cycle of gaps* $\mathcal{G}(p\#)$ among the remaining candidate primes.

For $p=2$, the cycle $\mathcal{G}(2\#) = 2$.

1 ~~2~~ 3 ~~4~~ 5 ~~6~~ 7 ~~8~~ 9 ~~10~~ 11 ~~12~~ 13 ~~14~~ 15 ~~16~~ 17 ~~18~~ 19 ~~20~~ 21 ~~22~~ 23 ~~24~~ 25 ~~26~~ 27 ~~28~~ 29 ~~30~~ 31 ~~32~~ 33 …

2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2

That is, after removing all the multiples of 2 (all the even numbers), the remaining candidate primes (the odd numbers) are separated by a gap of $g=2$.

The ***length*** of this cycle $\mathcal{G}(2\#)$ is 1 gap and its ***span*** (the sum of its gaps) is 2.

The first gap in the cycle goes from the unit 1 to the next candidate prime 3, since the number 2 has been confirmed as a prime number.

The notation $p\#$ denotes the *primorial of p.* This is the product of all the prime numbers up to and including p.

$$2\# = 2$$
$$3\# = 6$$
$$5\# = 30$$
$$7\# = 210$$

For *p=3*, the cycle $\mathcal{G}(3\#) = 4\ 2$.

1 2 3 ~~4~~ 5 ~~6~~ 7 ~~8~~ ~~9~~ ~~10~~ 11 ~~12~~ 13 ~~14~~ ~~15~~ ~~16~~ 17 ~~18~~ 19 ~~20~~ ~~21~~ ~~22~~ 23 ~~24~~ 25 ~~26~~ ~~27~~ ~~28~~ 29 ~~30~~ 31 ~~32~~ ~~33~~ …

4 2 4 2 4 2 4 2 4 2 4

The ***length*** of this cycle $\mathcal{G}(3\#)$ is 2 gaps and its ***span*** (the sum of its gaps) is 6.

At the stage of Eratosthenes sieve for *p=3*, the remaining candidate primes are separated by alternating gaps of *g=4* and *g=2*.

The first gap in the cycle goes from the unit 1 to the next candidate prime 5, passing over the confirmed primes.

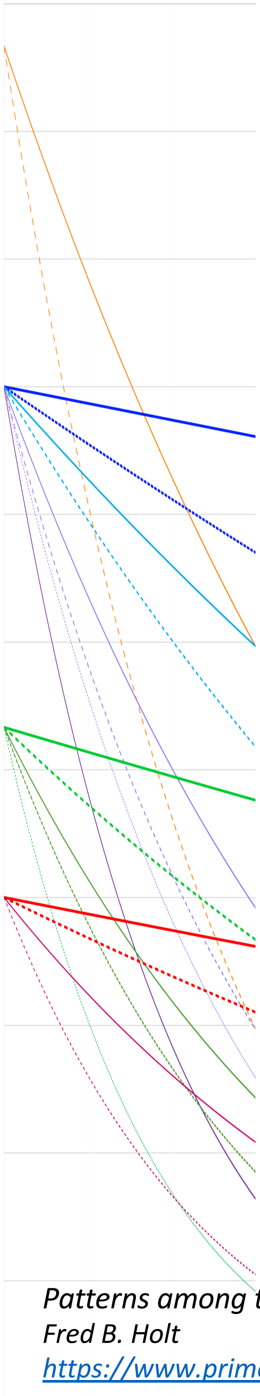
For *p=5*, the cycle $\mathcal{G}(5\#) = 6\ 4\ 2\ 4\ 2\ 4\ 6\ 2$.

1 2 3 ~~4~~ 5 ~~6~~ 7 ~~8~~ ~~9~~ ~~10~~ 11 ~~12~~ 13 ~~14~~ ~~15~~ ~~16~~ 17 ~~18~~ 19 ~~20~~ ~~21~~ ~~22~~ 23 ~~24~~ ~~25~~ ~~26~~ ~~27~~ ~~28~~ 29 ~~30~~ 31 ~~32~~ ~~33~~ …

6 4 2 4 2 4 6 2 6

The ***length*** of this cycle $\mathcal{G}(5\#)$ is 8 gaps and its ***span*** (the sum of its gaps) is 30.

Exercise: show that the span of the cycle $\mathcal{G}(p\#)$ is *p#*.



3. Recursion across the cycles of gaps $\mathcal{G}(p\#)$

There is a 3-step recursion that produces the next cycle of gaps from the current one.

$$\mathcal{G}(p_k\#) \longrightarrow \mathcal{G}(p_{k+1}\#)$$

Recursion. For p_k we have the cycle of gaps $\mathcal{G}(p_k\#) = g_1\, g_2\, g_3 \ldots g_N$

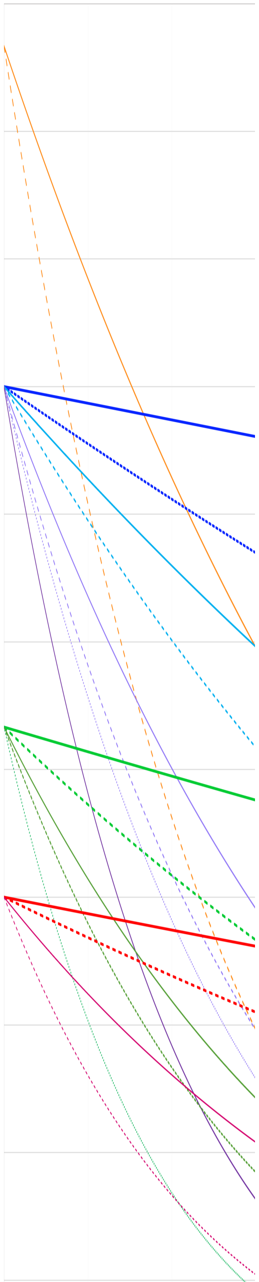
R1. The next prime $p_{k+1} = g_1+1$

R2. Concatenate p_{k+1} copies of $\mathcal{G}(p_k\#)$

R3. Add together the gaps g_1+g_2 and thereafter add adjacent gaps at the running sums indicated by the elementwise product $p_{k+1} * \mathcal{G}(p_k\#)$

The discrete dynamic system consists of the cycles of gaps under this recursion

$$\mathcal{G}(p_0\#) \longrightarrow \mathcal{G}(p_1\#) \longrightarrow \ldots \longrightarrow \mathcal{G}(p_k\#) \longrightarrow \mathcal{G}(p_{k+1}\#) \longrightarrow \ldots$$



<u>Recursion from $\mathcal{G}(p_k\#) \longrightarrow \mathcal{G}(p_{k+1}\#)$</u>

For p_k we have the cycle of gaps $\mathcal{G}(p_k\#) = g_1\, g_2\, g_3 \ldots g_N$

R1. The next prime $p_{k+1} = g_1{+}1$

R2. Concatenate p_{k+1} copies of $\mathcal{G}(p_k\#)$

R3. Add together the gaps $g_1{+}g_2$ and thereafter add adjacent gaps at the running sums indicated by the elementwise product $p_{k+1} * \mathcal{G}(p_k\#)$

As an example, let's use the recursion to produce $\mathcal{G}(5\#)$ from $\mathcal{G}(3\#)$.

R1. The next prime $p_{k+1} = g_1{+}1 = 4{+}1 = 5$

R2. Concatenate 5 copies of $\mathcal{G}(3\#)$

R3. Add together adjacent gaps using the elementwise product $5 * \mathcal{G}(3\#)$

	$p_k=3$ and $\mathcal{G}(3\#) = 4\ 2$
R1	$p_{k+1} = 4+1 = 5$
R2	$4 \overset{+}{\ } 2\ 4\ 2\ 4\ 2\ 4 \overset{+}{\ } 2\ 4\ 2$
R3	$5*\mathcal{G}(3\#) =$ 20 10
	$\mathcal{G}(5\#) = 6\ 4\ 2\ 4\ 2\ 4\ 6\ 2$

(Note that the last running sum 10 wraps around the end of the cycle, back to the first addition.)

In step R3, we call those additions of adjacent gaps *fusions.*

Each fusion corresponds to crossing a multiple of p_{k+1} off the list of candidate primes. When we eliminate a candidate prime, we form a new gap that is the sum of the gaps on either side of this former candidate.

As a second example, let's produce *𝒢(7#)* from *𝒢(5#)* by recursion.

p_k=5 and 𝒢(5#) = 6 4 2 4 2 4 6 2

R1. p_{k+1} = 6+1 = 7

R2. 7 copies of 𝒢(5#)

6 4 2 4 2 4 6 2 6 4 2 4 2 4 6 2 6 4 2 4 2 4 6 2 6 4 2 4 2 4 6 2 6 4 2 4 2 4 6 2 6 4 2 4 2 4 6 2 6 4 2 4 2 4 6 2

R3. Fusions

*7*𝒢(5#) = 42 28 14 28 14 28 42 14*

Elementwise product for running sums between fusions

𝒢(7#)= **10** 2 4 2 4 6 2 6 4 2 4 **6** 6 2 6 4 2 **6** 4 6 **8** 4 2 4 2 4 **8** 6 4 **6** 2 4 6 2 6 **6** 4 2 4 6 2 6 4 2 4 2 **10** 2

(Remember – the last running sum 14 wraps around the end of the cycle, back to the first fusion.)

The cycle of gaps *𝒢(7#)* has length 48 gaps and span 210.

Exercises: 1. To become familiar with the recursion, create *𝒢(5#)* and *𝒢(7#)* by hand

2. Show that the length of *𝒢(p#)* is $\phi(p\#) = \prod_{q \le p}(q-1)$.

Let’s take a closer look at <u>step R3</u> of the recursion.
We stay with the example of producing $\mathcal{G}(7\#)$ from $\mathcal{G}(5\#)$.

$p_k=5$ and $\mathcal{G}(5\#) = 6\ 4\ 2\ 4\ 2\ 4\ 6\ 2$

R1. $p_{k+1} = 6+1 = 7$

R2. 7 copies of $\mathcal{G}(5\#)$

6 4 2 4 2 4 6 2 6 4 2 4 2 4 6 2 6 4 2 4 2 4 6 2 6 4 2 4 2 4 6 2 6 4 2 4 2 4 6 2 6 4 2 4 2 4 6 2 6 4 2 4 2 4 6 2

+ + + + + + + +

R3. Fusions

7$\mathcal{G}$(5#) : 7***6** = 42 7***4** = 28 7***2** =14 7***4** = 28 7***2** =14 7***4** = 28 7***6** = 42 7***2** =14*

Elementwise product for running sums between fusions

Fusion #1 Fusion #2 Fusion #3 *… etc. …*

$\mathcal{G}$(7#)= **10** 2 4 2 4 6 2 6 4 2 4 **6** 6 2 6 4 2 **6** 4 6 **8** 4 2 4 2 4 **8** 6 4 **6** 2 4 6 2 6 **6** 4 2 4 6 2 6 4 2 4 2 **10** 2

The fusions occur where we have marked the next prime p_{k+1} and crossed its multiples off the list of candidate primes.

i. The first fusion is after the first gap, where we have marked p_{k+1} (here 7) as the next prime.

ii. We reset our running sum to 0 at this marker + and add up the gaps in step R2 until we reach the total 7*6=42. We reach this total with the gaps 4 2 4 2 4 6 2 6 4 2 4 2. So our second fusion occurs after this gap 2.

iii. We reset our running sum to 0 at this new marker + and add up the gaps in step R2 until we reach the total 7*4=28. We reach this total with the gaps 4 6 2 6 4 2 4. And our third fusion occurs after this gap 4.

Continue for the rest of the running sums provided by 7*$\mathcal{G}$(5#). Note that the last running sum wraps around the end of the cycle back to the first fusion. It had better align this way! These are cycles after all.

Check these sums

The cycle of gaps $\mathcal{G}(p\#)$ has length $\phi(p\#)$ gaps and span $p\#$.
These values grow very quickly.

prime p	Length of $\mathcal{G}(p\#)$ *#gaps* $\phi(p\#)$	Span of $\mathcal{G}(p\#)$ *Sum of gaps* $p\#$
2	1	2
3	2	6
5	8	30
7	48	210
11	480	2310
13	5760	30030
17	92160	510510
19	1658880	9699690
23	36495360	223092870
29	1021870080	6469693230
31	30656102400	200560490130

$\phi(n)$ is Euler's phi-function.

For primorials $p\#$

$$\phi(p\#) = \prod_{q \leq p} (q-1)$$

Where the product is over all primes up to and including *p*.

Exercises:

3. Use software to create the cycles $\mathcal{G}(p\#)$
4. What is the largest complete cycle $\mathcal{G}(p\#)$ you can create?

4. Two observations

We make two very useful observations about the recursive construction of $\mathcal{G}(p_{k+1}\#)$

Observation 1. The minimum distance between fusions in step R3 is $2p_{k+1}$.

The fusions are separated by the running sums $p_{k+1} * \mathcal{G}(p_k)$, and the smallest gap in $\mathcal{G}(p_k)$ is $g=2$. So the smallest running sum is $2p_{k+1}$.

Observation 2. Each possible fusion in $\mathcal{G}(p_k\#)$ occurs exactly once in step R3 of creating $\mathcal{G}(p_{k+1})$.

This is a result of the Chinese Remainder Theorem.

Recursion $\mathcal{G}(p_k\#) \rightarrow \mathcal{G}(p_{k+1}\#)$

R1. The next prime $p_{k+1} = g_1+1$

R2. Concatenate p_{k+1} copies of $\mathcal{G}(p_k\#)$

R3. Add together adjacent gaps using the elementwise product $p_{k+1} * \mathcal{G}(p_k\#)$

$p_k=5$ and $\mathcal{G}(5\#)$ = 6 4 2 4 2 4 6 2

R1. p_{k+1} = 6+1 = 7

R2. 7 copies of $\mathcal{G}(5\#)$

6 4 2 4 2 4 6 2 6 4 2 4 2 4 6 2 6 4 2 4 2 4 6 2 6 4 2 4 2 4 6 2 6 4 2 4 2 4 6 2 6 4 2 4 2 4 6 2 6 4 2 4 2 4 6 2

\+ \+ \+ \+ \+ \+ \+ \+

R3. Fusions

7$\mathcal{G}(5\#)$ = 42 28 14 28 14 28 42 14*

Elementwise product for running sums between fusions

$\mathcal{G}(7\#)$= **10** 2 4 2 4 6 2 6 4 2 4 **6** 6 2 6 4 2 **6** 4 6 **8** 4 2 4 2 4 **8** 6 4 **6** 2 4 6 2 6 **6** 4 2 4 6 2 6 4 2 4 2 **10** 2

Exercises:

1. Verify Observation 1 for the eight running sums in the construction from $\mathcal{G}(5\#)$ to $\mathcal{G}(7\#)$.

2. Verify Observation 2 for the eight fusions in the construction from $\mathcal{G}(5\#)$ to $\mathcal{G}(7\#)$.

From these two observations we derive an important lemma for building our population models for gaps.

$\mathcal{G}(5\#) = 6\ 4\ 2\ 4\ 2\ 4\ 6\ 2$

$\mathcal{G}(7\#) = 10\ 2\ 4\ 2\ 4\ 6\ 2\ 6\ 4\ 2\ 4\ 6\ 6\ 2\ 6\ 4\ 2\ 6\ 4\ 6\ 8\ 4\ 2\ 4\ 2\ 4\ 8\ 6\ 4\ 6\ 2\ 4\ 6\ 2\ 6\ 6\ 4\ 2\ 4\ 6\ 2\ 6\ 4\ 2\ 4\ 2\ 10\ 2$

A *constellation* is a sequence of consecutive gaps in $\mathcal{G}(p\#)$.

The *length* of a constellation is the number of gaps in the constellation, and the *span* of a constellation is the sum of the gaps in the constellation.

A gap is a constellation of length 1.

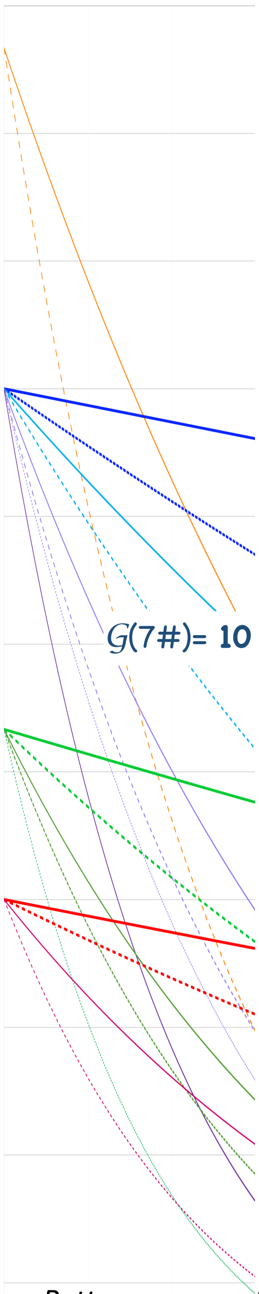
<u>Lemma</u>. Let *s* be a constellation of gaps in $\mathcal{G}(p_k\#)$ of length *J*.

If the span of *s* is less than $2p_{k+1}$, then each of the $J+1$ possible fusions in *s* occurs exactly once in forming $\mathcal{G}(p_{k+1}\#)$, and these fusions occur in $J+1$ distinct images of *s*.

Thus $p_{k+1}-J-1$ images of *s* survive into $\mathcal{G}(p_{k+1}\#)$ intact.

5. The number of gaps g=2 in $\mathcal{G}(p\#)$

With that lemma, we can exactly model the number of gaps $g=2$ that occur in the cycles of gaps $\mathcal{G}(p\#)$ across the stages of Eratosthenes sieve.



Let $n_2(p\#)$ be the number of gaps $g=2$ in the cycle $\mathcal{G}(p\#)$. We see that

$\mathcal{G}(3\#) = 4\ 2$ $\qquad$ $n_2(3\#) = 1.$

$\mathcal{G}(5\#) = 6\ 4\ 2\ 4\ 2\ 4\ 6\ 2$ $\qquad$ $n_2(5\#) = 3.$

$\mathcal{G}(7\#) = 10\ 2\ 4\ 2\ 4\ 6\ 2\ 6\ 4\ 2\ 4\ 6\ 6\ 2\ 6\ 4\ 2\ 6\ 4\ 6\ 8\ 4\ 2\ 4\ 2\ 4\ 8\ 6\ 4\ 6\ 2\ 4\ 6\ 2\ 6\ 6\ 4\ 2\ 4\ 6\ 2\ 6\ 4\ 2\ 4\ 2\ 10\ 2$

$n_2(7\#) = 15.$

By the Lemma, the gap $g=2$ is a constellation of length $J=1$, and $g < 2p_{k+1}$ for all $p_k > 2$.

So the factor $p-J-1 = p-2$ and $n_2(p_{k+1}\#) = (p_{k+1} - 2)\, n_2(p_k\#)$.

Starting from $p_0=3$, we have

$$n_2(p\#) = \prod_{3 \le q \le p} (q-2)$$

$n_2(p\#)$ is an exact count of the number of gaps $g=2$ in the cycle $\mathcal{G}(p\#)$.

Although we are very limited in the cycles $\mathcal{G}(p\#)$ that we can explicitly construct, we can calculate this population $n_2(p\#)$ for cycles for very large primes.

5 (cont'd). The number of gaps $g=4$ in $\mathcal{G}(p\#)$

Similarly let $n_4(p\#)$ be the number of gaps $g=4$ in the cycle $\mathcal{G}(p\#)$. We see that

$\mathcal{G}(3\#) = 4\ \ 2$ $\qquad$ $n_4(3\#) = 1.$

$\mathcal{G}(5\#) = 6\ 4\ 2\ 4\ 2\ 4\ 6\ 2$ $\qquad$ $n_4(5\#) = 3.$

$\mathcal{G}(7\#) =$ 10 2 4 2 4 6 2 6 4 2 4 6 6 2 6 4 2 6 4 6 8 4 2 4 2 4 8 6 4 6 2 4 6 2 6 6 4 2 4 6 2 6 4 2 4 2 10 2

$n_4(7\#) = 15.$

By the Lemma, the gap $g=4$ is a constellation of length $J=1$, and $g < 2p_{k+1}$ for all $p_k > 2$.

So $n_4(p_{k+1}\#) = (p_{k+1} - 2)\, n_4(p_k\#)$.

Starting from $p_0=3$, we have

$$n_4(p\#) = \prod_{3 \le q \le p} (q-2)$$

The number of gaps $g=4$ is the same as the number of gaps $g=2$ across all stages of Eratosthenes sieve for $p \geq 3$.

$$n_4(p\#) = n_2(p\#) \ \text{ for } p \geq 3$$

Exercises:

1. Calculate the number of gaps $g=2$ and $g=4$ in the cycle $\mathcal{G}(89\#)$
2. Calculate the number of gaps $g=2$ and $g=4$ in the cycle $\mathcal{G}(499\#)$

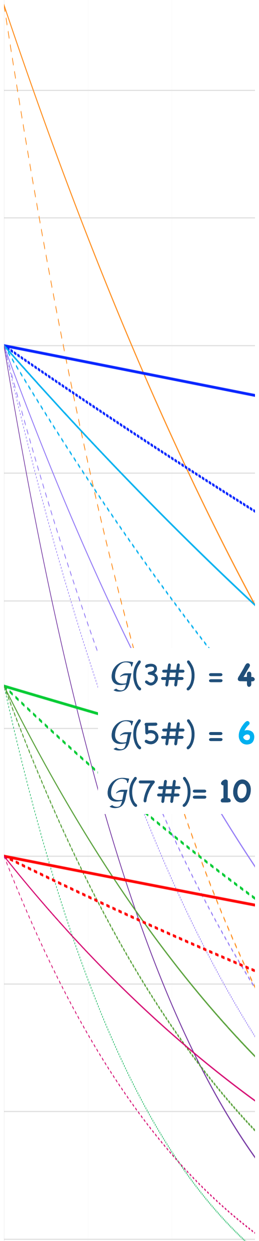
You may need to use logarithms.

6. Driving terms : the number of gaps g=6 in $\mathcal{G}(p\#)$

For the gap $g=6$ the population model gets more interesting.

The gap $g=6$ satisfies the condition $g < 2p_{k+1}$ for $p_k \geq 3$.

Under fusions, gaps $g=6$ are created by replication with the factor $(p-2)$, but they are also created by the interior fusions to the constellations $s= 2\ 4$ and $s = 4\ 2$.



We start a table

p	$n_{6,1}(p\#)$	$n_{6,2}(p\#)$	$n_2(p\#)$
3	0	2	1
5	2	4	3
7	14	16	15
11	142	128	135
13	1690	1280	1485

$\mathcal{G}(3\#)$ = 4 2

$\mathcal{G}(5\#)$ = 6 4 2 4 2 4 6 2

$\mathcal{G}(7\#)$= 10 2 4 2 4 6 2 6 4 2 4 6 6 2 6 4 2 6 4 6 8 4 2 4 2 4 8 6 4 6 2 4 6 2 6 6 4 2 4 6 2 6 4 2 4 2 10 2

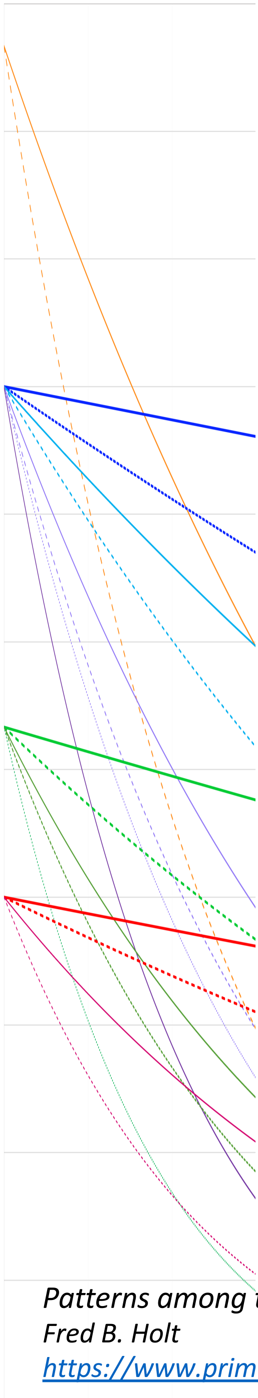
These constellations $s= 2\ 4$ and $s = 4\ 2$ are the *driving terms* of length 2 for the gap $g=6$.

We extend the notation $n_{g,j}(p\#)$ to mean the number of driving terms for the gap g of length j in the cycle $\mathcal{G}(p\#)$. $n_{g,1}(p\#)$ is the population of the gap g itself.

From our lemma, the population of gaps g=6 is given by the system

$$n_{6,1}(p_{k+1}\#) = (p_{k+1} - 2)\, n_{6,1}(p_k\#) + n_{6,2}(p_k\#)$$

$$n_{6,2}(p_{k+1}\#) = (p_{k+1} - 3)\, n_{6,2}(p_k\#)$$



We can rewrite the model for the population of the gap *g=6* in $\mathcal{G}(p\#)$ as a linear system

$$\begin{bmatrix} n_{6,1} \\ n_{6,2} \end{bmatrix} (p_{k+1}\#) = \begin{bmatrix} p_{k+1}-2 & 1 \\ 0 & p_{k+1}-3 \end{bmatrix} \begin{bmatrix} n_{6,1} \\ n_{6,2} \end{bmatrix} (p_k\#)$$

Exercises:

1. Calculate the number of gaps *g=6* in the cycle $\mathcal{G}(89\#)$
2. Calculate the number of gaps *g=6* in the cycle $\mathcal{G}(199\#)$
3. What percentages of the cycle $\mathcal{G}(199\#)$ are the gaps *g=2*, *g=4*, and *g=6*?

7. The number of gaps g=8 in $\mathcal{G}(p\#)$

The gap *g=8* satisfies the condition $g < 2p_{k+1}$ for $p_k \geq 3$, although the gap itself doesn't appear until $\mathcal{G}(7\#)$.

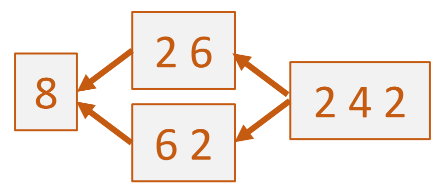
$\mathcal{G}(3\#)$ = 4 2

$\mathcal{G}(5\#)$ = 6 4 2 4 2 4 6 2

$\mathcal{G}(7\#)$= 10 2 4 2 4 6 2 6 4 2 4 6 6 2 6 4 2 6 4 6 8 4 2 4 2 4 8 6 4 6 2 4 6 2 6 6 4 2 4 6 2 6 4 2 4 2 10 2

For the gap *g=8*, we have driving terms of length *j=2*: *s= 6 2* and *s= 2 6;* and this gap has driving terms of length *j=3*: *s= 2 4 2*.

At each stage of Eratosthenes sieve, the interior fusions in *s=6 2* and *s=2 6* produce new gaps *g=8;*



and each of the two interior fusions in *s= 2 4 2* produce driving terms for *g=8* of length *j=2*, either *s=2 6* or *s=6 2*.

We can write the population model for the gap *g=8* in $\mathcal{G}(p\#)$ as this linear system

$$\begin{bmatrix} n_{8,1} \\ n_{8,2} \\ n_{8,3} \end{bmatrix} (p_{k+1}\#) = \begin{bmatrix} p_{k+1}-2 & 1 & 0 \\ 0 & p_{k+1}-3 & 2 \\ 0 & 0 & p_{k+1}-4 \end{bmatrix} \begin{bmatrix} n_{8,1} \\ n_{8,2} \\ n_{8,3} \end{bmatrix} (p_k\#)$$

To use this model we just need the initial counts in some $\mathcal{G}(p_0\#)$ such that $8 < 2p_1$.
We can use any $p_0 \geq 3$.

p	$n_{8,1}(p\#)$	$n_{8,2}(p\#)$	$n_{8,3}(p\#)$
3	0	0	1
5	0	2	1
7	2	10	3
11	28	86	21
13	394	902	189

Exercises:

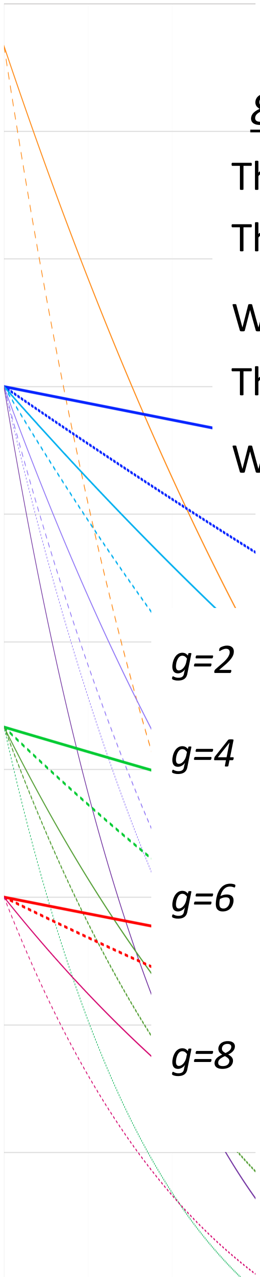
1. Identify the driving terms for *g=8* in $\mathcal{G}(3\#)$
2. Identify the driving terms for *g=8* in $\mathcal{G}(5\#)$
3. Track the driving terms, under fusions, from $\mathcal{G}(5\#)$ into $\mathcal{G}(7\#)$

Remember that the $\mathcal{G}(p\#)$ are *cycles.* The end wraps around to the beginning.

$\mathcal{G}(3\#)$ = 4 2

$\mathcal{G}(5\#)$ = 6 4 2 4 2 4 6 2

$\mathcal{G}(7\#)$= 10 2 4 2 4 6 2 6 4 2 4 6 6 2 6 4 2 6 4 6 8 4 2 4 2 4 8 6 4 6 2 4 6 2 6 6 4 2 4 6 2 6 4 2 4 2 10 2



8. Moving to models of relative populations

The population of every gap grows predominantly by factors of *(p-2)*.

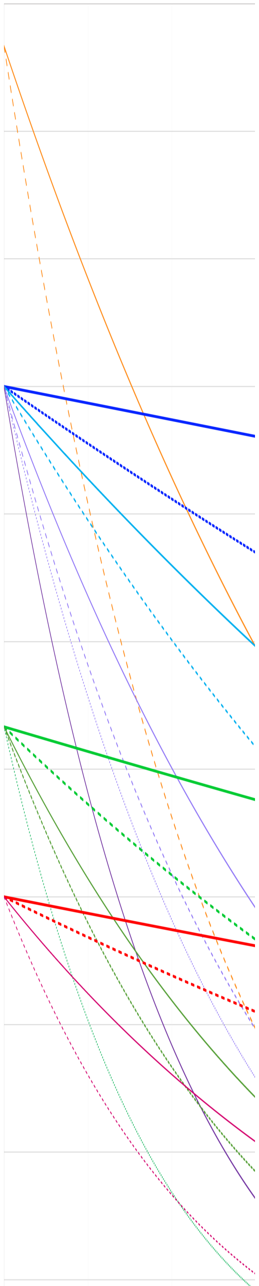
This is **super-exponential growth**, which quickly becomes unmanageable.

We normalize the population models by dividing by this factor of *(p-2)* at every stage.

The resulting population models are the *relative populations* of the gaps.

We denote the relative population of driving terms for gap *g* of length *j* in $\mathcal{G}(p\#)$ by $w_{g,j}(p\#)$.

	Population in $\mathcal{G}(p\#)$	Relative population in $\mathcal{G}(p\#)$
g=2	$n_2(p_{k+1}\#) = (p_{k+1}-2)\, n_2(p_k\#) = \prod_{3\le q\le p}(q-2)$	$w_2(p_{k+1}\#) = w_2(p_k\#) = 1$
g=4	$n_4(p_{k+1}\#) = (p_{k+1}-2)\, n_4(p_k\#) = \prod_{3\le q\le p}(q-2)$	$w_4(p_{k+1}\#) = w_4(p_k\#) = 1$
g=6	$\begin{bmatrix} n_{6,1} \\ n_{6,2} \end{bmatrix}(p_{k+1}\#) = \begin{bmatrix} p_{k+1}-2 & 1 \\ 0 & p_{k+1}-3 \end{bmatrix} \begin{bmatrix} n_{6,1} \\ n_{6,2} \end{bmatrix}(p_k\#)$	$\begin{bmatrix} w_{6,1} \\ w_{6,2} \end{bmatrix}(p_{k+1}\#) = \begin{bmatrix} 1 & \frac{1}{p_{k+1}-2} \\ 0 & \frac{p_{k+1}-3}{p_{k+1}-2} \end{bmatrix} \begin{bmatrix} w_{6,1} \\ w_{6,2} \end{bmatrix}(p_k\#)$
g=8	$\begin{bmatrix} n_{8,1} \\ n_{8,2} \\ n_{8,3} \end{bmatrix}(p_{k+1}\#) = \begin{bmatrix} p_{k+1}-2 & 1 & 0 \\ 0 & p_{k+1}-3 & 2 \\ 0 & 0 & p_{k+1}-4 \end{bmatrix} \begin{bmatrix} n_{8,1} \\ n_{8,2} \\ n_{8,3} \end{bmatrix}(p_k\#)$	$\begin{bmatrix} w_{8,1} \\ w_{8,2} \\ w_{8,3} \end{bmatrix}(p_{k+1}\#) = \begin{bmatrix} 1 & \frac{1}{p_{k+1}-2} & 0 \\ 0 & \frac{p_{k+1}-3}{p_{k+1}-2} & \frac{2}{p_{k+1}-2} \\ 0 & 0 & \frac{p_{k+1}-4}{p_{k+1}-2} \end{bmatrix} \begin{bmatrix} w_{8,1} \\ w_{8,2} \\ w_{8,3} \end{bmatrix}(p_k\#)$



We can interpret the relative population for a gap *g* as the ratio of populations of gaps *g* to gaps *2* in $\mathcal{G}(p\#)$: $w_{g,1}(p\#) = n_{g,1}(p\#) / n_2(p\#)$

The relative population models $w_{g,j}(p\#)$ are still exact, and the values are more manageable.

p	$w_2(p\#)$	$w_4(p\#)$	$w_{6,1}(p\#)$	$w_{8,1}(p\#)$
3	1	1	0.	0.
5	1	1	0.666667	0.
7	1	1	0.933333	0.133333
11	1	1	1.051852	0.207407
13	1	1	1.138047	0.265320
17	1	1	1.195511	0.305814
19	1	1	1.242834	0.340160
…				
199	1	1	1.549497	0.583563
499	1	1	1.611811	0.637093
997	1	1	1.649362	0.669986
2503	1	1	1.690123	0.706229
4999	1	1	1.715031	0.728652
10007	1	1	1.736414	0.748067
49999	1	1	1.775411	0.783871
100003	1	1	1.788920	0.796393
500009	1	1	1.814773	0.820527
1000003	1	1	1.824052	0.829245
4999999	1	1	1.842411	0.846576
9999991	1	1	1.849183	0.852998
15250009	1	1	1.853031	0.856654

By comparison,

$n_2(15250009\#) \approx 6.178 \times 10^{6620813}$

YIKES !!

9. Easier and easier calculations -- eigenstructure

The models $w_{g,j}(p\#)$ for the exact relative populations of the gap g and its driving terms simplify in three ways.

A. The models have simple eigenstructures.

The eigenstructure provides a basis of right eigenvectors R that isolate the scalings of a matrix along these directions. The scalings are given by the eigenvalues Λ, and the left eigenvectors L act as filters to extract the coefficients for a vector in the basis R. $RL = LR = I$.

Here's the eigenstructure

$$\begin{bmatrix} 1 & \frac{1}{q-2} & 0 \\ 0 & \frac{q-3}{q-2} & \frac{2}{q-2} \\ 0 & 0 & \frac{q-4}{q-2} \end{bmatrix} = R\,\Lambda\,L = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{q-3}{q-2} & 0 \\ 0 & 0 & \frac{q-4}{q-2} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

For our models of the relative populations $w_{g,j}(p\#)$, the eigenvectors R and L do not change with the prime p. So the model 'telescopes' and we can more easily calculate values for large primes.

Here’s the telescoping product

$$\prod_{p_1}^{p_k} \begin{bmatrix} 1 & \frac{1}{q-2} & 0 \\ 0 & \frac{q-3}{q-2} & \frac{2}{q-2} \\ 0 & 0 & \frac{q-4}{q-2} \end{bmatrix} = R\,\Lambda^k\,L = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \prod_{p_1}^{p_k} \frac{q-3}{q-2} & 0 \\ 0 & 0 & \prod_{p_1}^{p_k} \frac{q-4}{q-2} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

From the simple eigenstructure, we can isolate the top row to extract exact models of the relative populations for a gap g with $g < 2p_1$. In the cycle $\mathcal{G}(p_0\#)$ the gap g itself may not appear, but it will have driving terms up to some length J. Then

$$w_{g,1}(p_k\#) = l_1 - l_2 \prod_{p_1}^{p_k} \frac{q-3}{q-2} + l_3 \prod_{p_1}^{p_k} \frac{q-4}{q-2} - \ldots + (-1)^{J+1} l_J \prod_{p_1}^{p_k} \frac{q-J-1}{q-2}$$

Where the coefficients are given by

$$\begin{bmatrix} l_1 \\ \vdots \\ l_J \end{bmatrix} = L\, w_g(p_0\#) \quad \text{and} \quad w_{g,j}(p_0\#) = \frac{n_{g,j}(p_0\#)}{n_2(p_0\#)} = \frac{n_{g,j}(p_0\#)}{\prod_3^{p_0}(q-2)}$$

For example, to compute the relative population of the gap *g=8* in $\mathcal{G}$*(15485863#),* instead of computing one million 3x3 matrix multiplications, we calculate

$$\prod_{q=7}^{15485863} \frac{q-3}{q-2} = 0.1101249 \qquad \prod_{q=7}^{15485863} \frac{q-4}{q-2} = 0.0108483$$

We have the initial conditions $w_{8,1}(5\#) = 0/3$; $w_{8,2}(5\#) = 2/3$; $w_{8,3}(5\#) = 1/3$.

And

$$w_8(15485863\#) = R\,\Lambda^k\,L \begin{bmatrix} 0 \\ 2/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} 1 - \frac{4}{3}\,0.1101249 + \frac{1}{3}\,0.0108483 \\ \frac{4}{3}\,0.1101249 - \frac{2}{3}\,0.0108483 \\ \frac{1}{3}\,0.0108483 \end{bmatrix}$$

The top row provides the model of the relative population for the gap *g=8* itself

$$w_{8,1}(p_k\#) = 1 - \frac{4}{3}\prod_{p_1}^{p_k}\frac{q-3}{q-2} + \frac{1}{3}\prod_{p_1}^{p_k}\frac{q-4}{q-2}$$

Exercise. Using $\mathcal{G}$*(5#)*, determine the model for the relative population of the gap *g=10.*

$$w_{g,1}(p_k\#) = l_1 - \; l_2 \prod_{p_1}^{p_k} \frac{q-3}{q-2} + l_3 \prod_{p_1}^{p_k} \frac{q-4}{q-2} \; - \; \ldots + (-1)^{J+1} l_J \prod_{p_1}^{p_k} \frac{q-J-1}{q-2}$$

B. The parameters for these models are almost polynomial.

We denote the first parameter in the model as λ

$$\lambda = \prod_{p_1}^{p_k} \frac{q-3}{q-2}$$

Then the other parameters are approximately powers of λ,

$$\prod_{p_1}^{p_k} \frac{q-j-1}{q-2} \; \approx \; \lambda^{j-1}$$

and the models are approximately polynomial

$$w_{g,1}(p_k\#) \approx l_1 - \; l_2\lambda + l_3\lambda^2 \; - \; \ldots + (-1)^{J+1} l_J \lambda^{J-1}$$

Exercise. Using the models for *g=8* and *g=10,* analyze the error between the exact model for the relative population and the polynomial approximation.

$$w_{g,1}(p_k\#) \approx l_1 - \; l_2\lambda + l_3\lambda^2 \; - \; \ldots + (-1)^{J+1} l_J \lambda^{J-1} \;\; \text{with } \lambda = \prod_{p_1}^{p_k} \frac{q-3}{q-2}$$

We can explore these models in the parameter λ and then tie the parameter λ to the prime p_k. We could tabulate the correspondence $\lambda \leftrightarrow \; p_k$ up through a few million primes but how do we estimate this correspondence for very large primes?

C. Merten's Third Theorem

Franz Merten's third theorem from 1874 provides the approximation

$$\prod_{2}^{p_k} \frac{q-1}{q} \approx \frac{e^{-\gamma}}{\ln p_k}$$

We can use this approximation to associate really large primes with a correspondingly small value of λ.

$$\lambda = \prod_{p_1}^{p_k} \frac{q-3}{q-2} \approx \prod_{p_1}^{p_k} \frac{q-1}{q} \approx C \frac{e^{-\gamma}}{\ln p_k}$$

Where the constant C adjusts for the product from 2 through p_0 and also for the quality of the approximation $\frac{q-3}{q-2} \approx \frac{q-1}{q}$ for small primes.

Exercises:

3. Estimate a good value of C when $p_0 = 5$.

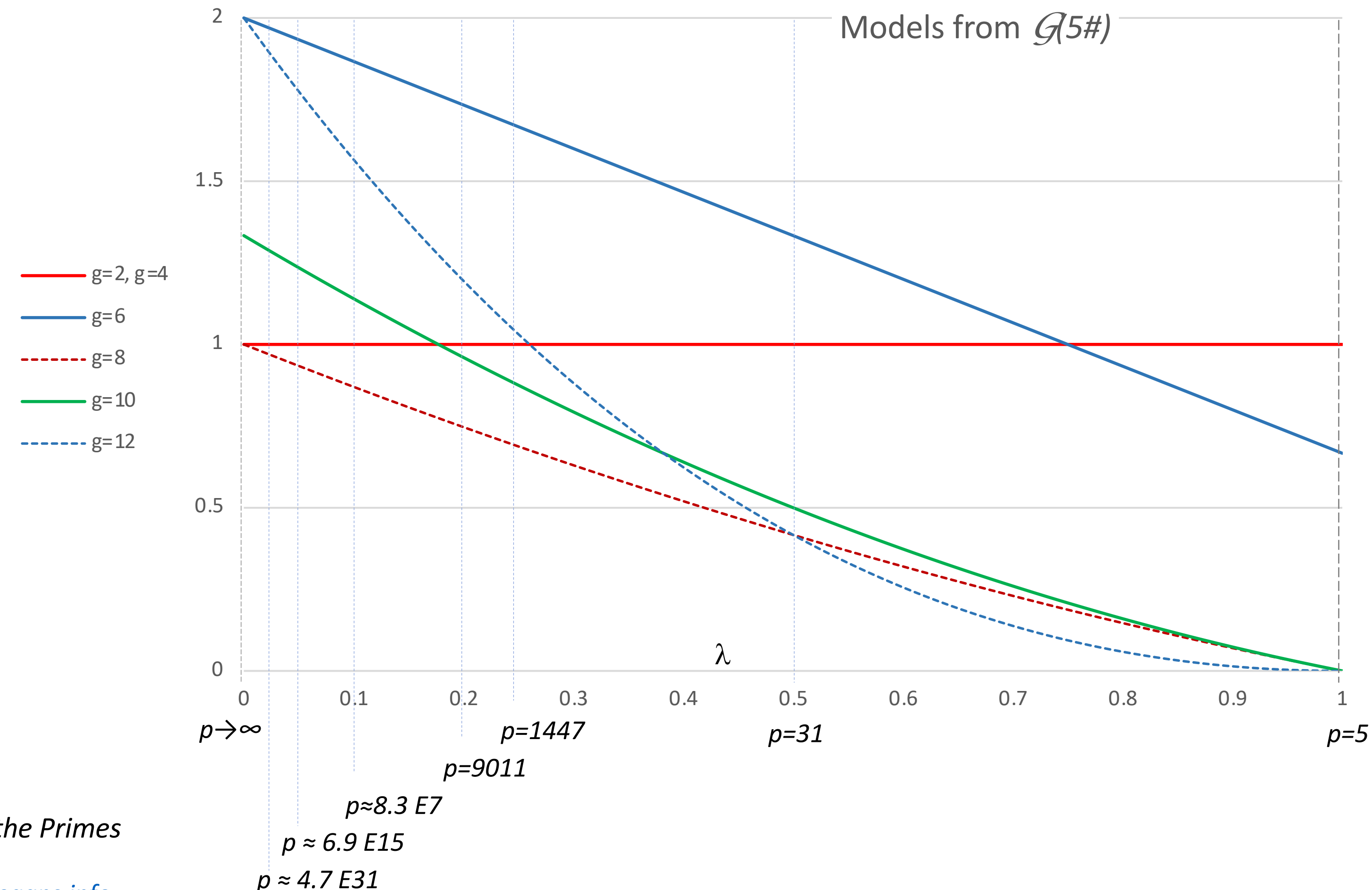
For $p_0=5$ we estimate
$C = 3.24719$

4. Suppose $\lambda = \lambda_1$ corresponds to a prime p. Show that if we want to reduce λ by half, to $\lambda = \frac{\lambda_1}{2}$, we have to use primes q near p^2.

Now we have models $w_{g,1}(\lambda)$ that are approximately polynomial in the parameter λ, and we have a map from λ to p_k that enables us to use the models for primes well beyond the computational range.

$$\lambda \to 0 \text{ as } p_k \to \infty$$

For $p_0=5$, we have the following models.



10. Surviving the sieve

The gaps at the front of the cycle $\mathcal{G}(p_k\#)$ survive to be confirmed as gaps among primes. Every gap from p_{k+1} to p_{k+1}^2 will survive the further stages of the sieve.

In this interval $[p_{k+1}, p_{k+1}^2]$ all of the gaps up to p_k^2 were confirmed in the previous stage of the sieve. So the new *interval of survival* of gaps confirmed by $\mathcal{G}(p_k\#)$ is $[p_k^2, p_{k+1}^2]$.

$\mathcal{G}(7\#)$= 10 2 4 2 4 6 2 6 4 2 4 6 6 2 6 4 2 6 4 6 8 4 2 4 2 4 8 6 4 6 2 4 6 2 6 6 4 2 4 6 2 6 4 2 4 2 10 2 • • •

Next prime p_1=11

horizon of survival for p=7

121 *143* *187* *209*

horizon of survival for p=11

$\mathcal{G}(11\#)$= 12 4 2 4 6 2 6 4 2 4 6 6 2 6 4 2 6 4 6 8 4 2 4 2 4 14 4 6 2 10 2 6 6 4 2 4 6 2 10 2 4 2 12 10 2 4 • • •

Next prime p_2=13

169 *221*

14 4 6 2 10 2 6 6 4

Interval of survival $[p_1^2, p_2^2]$ for p=11

This interval contains the gaps newly confirmed as gaps between primes at this stage of the sieve.

horizon of survival for p=13

$\mathcal{G}(13\#)$= 16 4 2 . . . 2 4 14 4 6 2 10 2 6 6 4 6 6 2 10 2 4 2 12 12 4 2 4 6 2 10 6 6 6 2 6 4 2 6 4 14 4 2 4 • • •

Next prime p_3=17

289

6 6 2 10 2 4 2 12 12 4 2 4 6 2 10 6 6 6 2 6 4 2

Interval of survival for p=13
$[p_2^2, p_3^2]$

We expect the gaps among the primes in this interval $[p_k^2, p_{k+1}^2]$ to be statistically consistent with the ratios $w_2(p_k\#)$, $w_4(p_k\#)$, …, $w_g(p_k\#)$.

The *interval of survival* of gaps confirmed by $\mathcal{G}(p_k\#)$ is $[p_k^2, p_{k+1}^2]$.

We expect the gaps among the primes in this interval $[p_k^2, p_{k+1}^2]$ to be statistically consistent with the ratios $w_2(p_k\#), w_4(p_k\#), \ldots, w_g(p_k\#)$.

As an example, for $\lambda \approx 0.25$ we tabulate the populations of gaps $2 \leq g \leq 12$ in a series of intervals of survival $[p_k^2, p_{k+1}^2]$ for some of the p_k corresponding to $\lambda \approx 0.25$.

Primes p_k	gap	p_k^2	**g=2**	**4**	**6**	**8**	**10**	**12**	
1427		2036329							
1429	2	2042041	33	34	56	24	28	37	**212**
1433	4	2053489	71	73	111	51	70	70	**446**
1439	6	2070721	115	108	189	74	113	123	**722**
1447	8	2093809	147	142	235	104	131	137	**896**
1451	4	2105401	56	59	128	57	73	74	**447**
1453	2	2111209	36	37	64	29	31	39	**236**
1459	6	2128681	103	111	196	80	87	111	**688**
1471	12	2163841	229	211	385	161	209	235	**1430**
			790	**775**	**1364**	**580**	**742**	**826**	**5077**
		$w_g(\lambda\text{=}.25)$	*1*	*1*	*1.6667*	*0.6875*	*0.875*	*1.03125*	

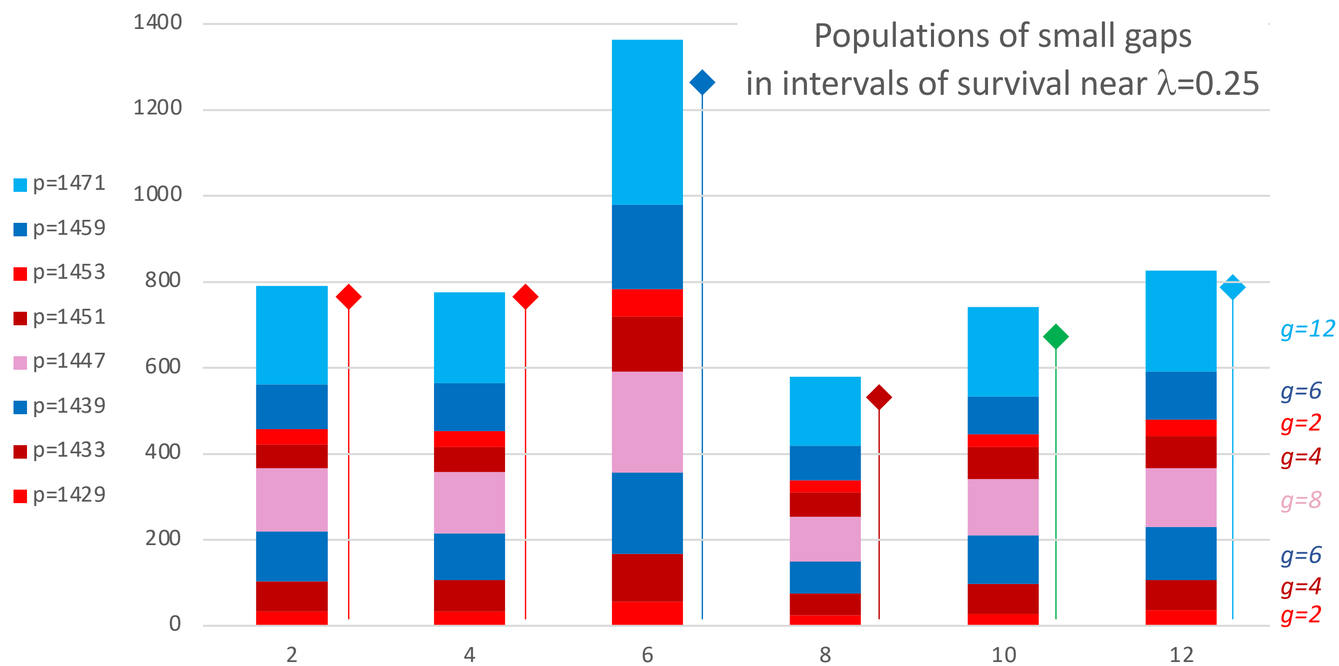
Of the *8761* prime gaps between $1427^2 = 2036329$ and $1471^2 = 2163841$, *5077* are counted in the table above.

Let $g = p_{k+1} - p_k$.

Then the length of the interval $[p_k^2, p^2_{k+1}]$ is $\quad p^2_{k+1} - p_k^2 = (p_k+g)^2 - p_k^2$

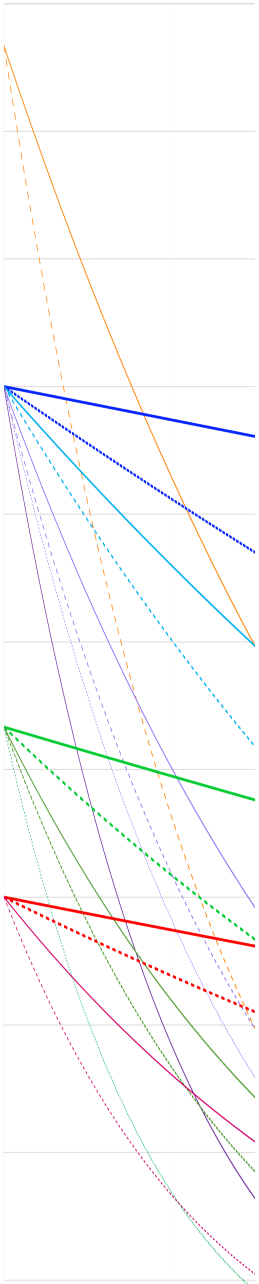
$$= g(2p_k+g)$$

Here we plot the data tabulated above. Each column represents the counts of the gaps in the intervals of survival, and these are color-coded by the gap $g = p_{k+1} - p_k$ for the interval of survival $[p_k^2, p_{k+1}^2]$.



The diamonds mark the relative populations for these gaps in $\mathcal{G}(p\#)$ at $\lambda=0.25$, with $p_0=5$.

Patterns among the Primes
Fred B. Holt
https://www.primegaps.info



With the condition $g < 2p_1$

If we use $p_0 = 5$ we can use the initial conditions in $\mathcal{G}(5\#)$ to calculate the relative population models $w_g(\lambda)$ for gaps up to $g=12$.

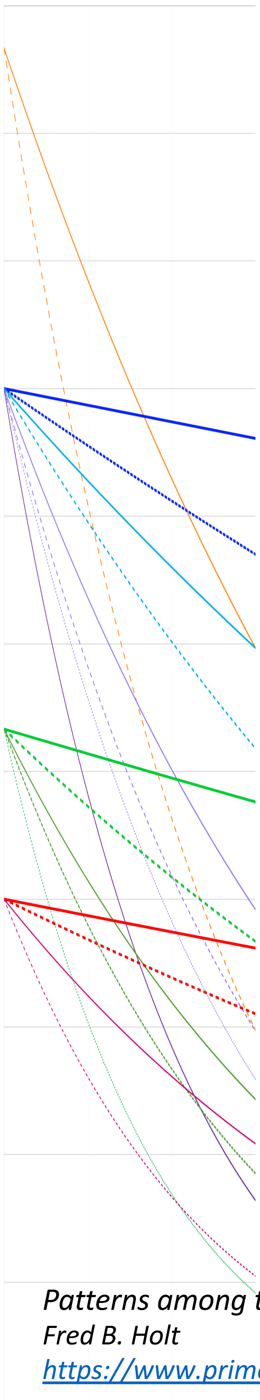
If we use $p_0 = 7$ we can use the initial conditions in $\mathcal{G}(7\#)$ to calculate the relative population models for gaps up to $g=20$.

Exercises

1. What is the largest prime p_0 for which you could create the complete cycle $\mathcal{G}(p_0\#)$, i.e. for Exercise 4 in Section 3? With this cycle, for what range of gaps g can you calculate the relative population models?

2. For this p_0 calculate the initial conditions for gaps $g < 2p_1$, and the models for the relative populations of these gaps.

3. For the gaps $g=6$ and $g=8$, how do the models with $p_0=5$ compare to the models for this larger p_0? The models are all exact, so how do we reconcile the variations in the models for different choices of p_0?



Patterns among the Primes
Fred B. Holt
https://www.primegaps.info

F.B. Holt, *Patterns among the Primes,* KDP 2022.

https://primegaps.info