Report

Approximation for the number of prime pairs adding up to even integers

Author(s):
Marques Filho, A.; Gassmann, F.; Walker, I.

Publication Date:
2005

Permanent Link:
https://doi.org/10.3929/ethz-a-004953116

Rights / License:
In Copyright - Non-Commercial Use Permitted
Approximation for the number of prime pairs adding up to even integers

A. Marques Filho\textsuperscript{(1)}, F. Gassmann\textsuperscript{(2)} and I. Walker\textsuperscript{(1)}

\textsuperscript{(1)} Instituto Nacional de Pesquisas da Amazônia, P.O. Box 478, MCT/INPA, Manaus, AM-Brazil
\textsuperscript{(2)} Project Complex Systems, Paul Scherrer Institute, CH-5232 Villigen PSI

Tel. +41 310 26 47; Fax. +41 310 44 16, E-mail fritz.gassmann@psi.ch, corresponding author

Abstract

We investigate periodicity of indicators for relative primes after applying the Erathostenes sieve up to a prime \( P \). The notion of \textit{in phase} and \textit{out of phase} of two overlapping mirror symmetric sequences of prime indicators leads to an approximation for the number \( N_E \) of prime pairs adding up to a given even integer \( E \). Due to our assumption of regularly distributed relative prime pairs (density assumption), our approximation cannot be transformed into a hard lower bound for \( N_E \) and therefore cannot prove the Goldbach conjecture in a strict mathematical sense. However, the approximation is remarkably precise and can explain large differences of \( N_E \) for neighbouring \( E \)'s (resonances). Further, the standard deviation of the relative error of the approximation decreases according to a power law at least up to \( E = 3 \times 10^7 \). Based on our approach, a strong form of the Goldbach conjecture is obtained, stating that the number of Goldbach partitions increases with \( E/(\ln E)^2 \). Taken together, these observations suggest a statistical proof of Goldbach's conjecture with very quickly vanishing probability of failure for large \( E \).

Introduction

Goldbach's conjecture\textsuperscript{1}, stating that every even integer \( \geq 2 \) can be expressed as sum of two primes (in the following called \textit{prime pairs}), is keeping mathematicians busy. Though computer-based numerical tests\textsuperscript{2} have confirmed the theorem up to \( 4 \times 10^{14} \), it resisted any formal mathematical proof up to now\textsuperscript{3}. A necessary condition for the conjecture to be valid is, that new primes steadily appear on the sequence of integers without limit. The respective proof was already given by the Greek mathematician Euclid around 300 BC and has been confirmed repeatedly in different forms since then\textsuperscript{4,5}.

As regards the numbers of existing prime pairs for consecutive even integers, one of the puzzels is the occurrence of \textit{resonances} (large differences for consecutive even integers): e.g. 30/030 can be written as the sum of 1810 prime pairs, whereas 30/028 and 30/032 have 474 and 450 prime pairs, respectively (we count symmetric sums as 2 pairs: e.g. 10 = 3+7, 5+5, 7+3). In addition, these resonances seem to be coupled to the distances of consecutive primes. Astonishingly, twin primes or prime tuplets, as e.g. 17/19 and 29/31, occur relatively frequently although prime density continuously diminishes towards large integers. In fact, the continued existence of twin primes is another prime conjecture awaiting mathematical proof\textsuperscript{6}. These mysterious prime patterns are even discussed in the popular press: \textit{The New York Times} presented\textsuperscript{7} the twin primes 98'/711/98'/713 and 98'/729/98'/731 together with the intermediate prime 98'/717 as examples for these strange patterns.

The present paper is a continuation of a previous study on the structure of prime-multiples and its effect on prime pair frequency (Marques Filho & Walker, 2001\textsuperscript{8}). Particularly, the claim of the solution of Goldbach’s conjecture is re-analysed
with more rigour. As a convincing application of our approximation for the number of Goldbach partitions, resonances can be explained and understood. Numerical tests up to \( E = 3 \cdot 10^7 \) suggest the possibility of a unusual probabilistic proof for Goldbach’s conjecture.

### Analysis of periodic prime-indicator patterns

Erathostenes proposed the well-known sieve method\(^9\) as an efficient means to find prime numbers around 300 BC. On this basis, Gauss\(^10\) estimated the total number of primes \( N(Z) \) smaller than a given integer \( Z \):

\[
N(Z) = \left \lceil \frac{Z}{\ln Z} \right \rceil
\]

Our analysis is based on a generalized sieve method. In a first step, we will count the number of relative primes (definition see below) in appropriate periodic intervals. Then, we will generalise this method to find the number of relative prime pairs adding up to an even integer \( E \). This will lead to an approximation for the number of Goldbach partitions \( N_E \) for each even integer \( E \).

### Number of relative primes within periodic intervals

We consider the sequence of integers after application of the Erathostenes sieve up to a prime \( P \): multiples of all primes \( P_i \leq P \) are assumed to be set to 0. For \( P=5 \), the resulting integer sequence would be \( 1, 2, 3, 0, 5, 0, 7, 0, 0, 11, 0, 13, 0, 0, 17, 0, 19, 0, 0, 0, 23, 0, 0, 0, 0, 29, 0, 31, 0, 0, 0, 0, 0, 0, 0, 37, 0, 0, 0, 0, 41, 0, 43, 0, 0, 47, 0, 49, 0, ... \). The process shows that all non-zero integers up to \( P_N^2 - 1 \) are real primes (except 1), \( P_N > P \) being the next higher prime following \( P \). \( P_N^2 \) (49 in our example with \( P=5 \), \( P_N=7 \)) is the lowest \( P \)-relative prime, i.e. it cannot be divided by any integer smaller than \( P_N \). By substituting all non-zero integers \( Z_i=i \) at positions \( i \) by 1 (i.e. for all \( i>0 \): if \( Z_i>0 \) then \( Z_i=1 \)), we get a periodic pattern for \( i>P \) with period \( T_P \). We call this periodic pattern relative prime-indicator pattern. As it is created by setting prime-multiples to zero, we call it also prime-multiple structure. For a given \( T_P \), \( P_N=T_P \) will be the shortest corresponding period for the \( P_N \)-relative prime pattern, leading to the general relation:

\[
T_P = \prod_{P=2}^P P^i
\]

(2)

\( T_P \) is often called primorial\(^11\) in the literature. We have to prove that there is no shorter period than \( T_P \). For \( P=2 \), \( T_P = 2 \) is obviously correct. We assume a \( P_N \)-related period \( T < 2 \cdot 3 \cdot 5 \cdot ... \cdot P_N \), i.e. a certain phase of the smaller period \( T_N \) (two integers with distance \( Z_i-T_P \)) should be visited by multiples of \( P_N \) separated by \( T = Z_i P_N < T_P \):\( P_N \):

\[
T = Z_i T_P = Z_i P_N \quad \text{with} \quad Z_i < P_N, Z_i < T_P
\]

(3)

The prime-factorization of \( T \) does not include the prime \( P_N \), because \( P_N \) is not included in \( T_P \) according to eq.2 and cannot show up in the prime-factorization of \( Z_i \) due to \( Z_i < P_N \). As \( T = Z_i P_N \) according to eq.3, its prime-factorization includes \( P_N \) and so contradicts to the uniqueness of prime-factorization of \( T \). It follows from this proof
Superposition of two periodic sequences gives sums of relative prime pairs

To find the number \( G_{P,E} \) of relative prime pairs adding up to a given even integer \( E \) (i.e. the number of relative Goldbach partitions for \( E \)), we superpose the integer sequence after application of the sieve-process up to \( P \) with its mirrored sequence as shown in the following example for \( E=32 \) with a sieve applied up to \( P=3 \). Only the odd positions \( i \) (normal sequence) and \( j \) (mirrored sequence) are shown together with the according 3-relative prime patterns \( i1 \) and \( j1 \):

\[
\begin{align*}
\text{i1} & : 1 \ 1 \ 1 \ 1 \ 1 \ 0 \ 1 \ 1 \ 0 \ 1 \ 1 \ 0 \ 1 \ 0 \ 1 \ 1 \ 1 \\
\text{j} & : 3 \ 7 \ 35 \ 30 \ 29 \ 0 \ 25 \ 23 \ 0 \ 19 \ 17 \ 0 \ 13 \ 11 \ 0 \ 7 \ 5 \ 3 \ 1 \\
\text{i} & : 1 \ 3 \ 5 \ 7 \ 0 \ 11 \ 13 \ 0 \ 17 \ 19 \ 0 \ 23 \ 25 \ 0 \ 29 \ 31 \ 0 \ 35 \ 37 \\
\text{i1} & : 1 \ 1 \ 1 \ 1 \ 0 \ 1 \ 1 \ 0 \ 1 \ 1 \ 0 \ 1 \ 1 \ 0 \ 1 \ 1 \ 0 \ 1 \ 1 \\
\end{align*}
\]

In the periodic parts of the sequences \( i, j \), we find the 4 underlined 3-relative prime pairs 7+25, 13+19, 19+13, 25+7 characterized by coinciding 1 in the according 3-
relative prime-indicator patterns i1 and j1. The pairs 3+29 and 29+3 are not counted, because they do not belong to the periodic structure considered here.

We ask now for the number of P-relative prime pairs lost when applying the sieve process for P N, assuming the respective number G P, E for P is known. This iteration step depends on the relative position between the two overlapping sequences being defined by the even position E of the origin (0) of the j-sequence measured in the i-sequence. If E is a multiple of P N, we call the sequences in phase relative to P N and the sieve process can only delete paired positions in the i-and j-sequences leading to a relation analogous to eq.4. For out of phase situations relative to P N (E is not a multiple of P N), deletions in sequences i and j never coincide and so, twice as many pairs are lost. These two special cases lead to the general relation:

\[ G_{PN, E} = G_{P, E}P_N - (2 - [\delta_{E, ZPN}])G_{P, E} \quad (7) \]

\( \delta_{E, ZPN} \) is the Kronecker-delta being 1 if E=Z·P N (Z=integer) and 0 otherwise. The recursive relation eq.7 leads to the following general relation for the number of P-relative Goldbach partitions for E in the periodic parts of the sequences i and j:

\[ G_{P, E} = \prod_{p \mid 3} P2 + \delta_{[E, Z P]} \quad (8) \]

As E is always in phase with 2, contributing a factor of 1, the product can begin with 3 instead of 2. The average P-relative pair-density \( \bar{\delta}_{P, E} \) within a period T_p is therefore:

\[ \bar{\delta}_{P, E} = \frac{G_{P, E}}{T_P} = \frac{1}{2} \prod_{p \not| 3} P2 + \delta_{[E, Z P]} \quad (9) \]

In our above example with E=32 being out of phase with 3, eq.9 gives \( \bar{\delta}_{3, 32} = 1/6 \). For the periodic part between P+1=4 and E=32, the total number of relative prime pairs is 28/6 ≈ 4 as indicated in our example.

It can be shown that \( \bar{\delta}_{P, E} > 1/(2P) \):

\[ \bar{\delta}_{P, E} \geq \frac{1}{2} \prod_{p \not| 3} \frac{P2}{P} > \frac{1}{2} \frac{113579}{2357911} \ldots \frac{P2}{P} = \frac{1}{2} \frac{1}{P} \quad \text{for } P > 7 \quad (10) \]

The left unequal sign holds because not all P might be out of phase and the right unequal sign holds because some ratios (all <1) are missing in the product \( \prod \), e.g. 7/9.

**Estimation of the number of prime pairs adding up to an even integer E**

All relative primes <E are primes, if the sieve process has been performed up to P with P²<E and P²>E. For this P, we calculate the density \( \bar{\delta}_{P, E} \) according to eq.9. It refers to P-relative prime pairs within a periodic interval of length T_p. To construct such a periodic interval, we have to shift the zero point of sequence j (being at position i=E) to the right by 2·T_p, i.e. this new situation would be correct to find Goldbach partitions for 2·T_p+E instead of E as intended. Further, this T_p-interval contains numerous P-relative primes that are not prime. We consider the interval E_i: i= T_p... T_p+E in the shifted situation (zero point of j at i=2·T_p+E) and compare it with the interval E_0: i=0...E in the base situation (zero point of j at i=E). E_0 contains all P-relative prime pairs found in E_i as images being T_p smaller. According to our
condition for P, all these images are real primes in $E_0$. In addition, $E_0$ contains all primes 2...P. From these primes $P \leq P$, those being *out of phase* with E (and if $E=2\cdot P^r$ also this *in phase* $P^r$) are also candidates for prime pairs $P+E$, $P+(-E-P)$. The number $N_E$ of Goldbach partitions for $E$ is therefore equal or greater than the number $N_{E_1}$ of $P$-relative prime pairs adding up to $2\cdot T+P$: $N_E \geq N_{E_1}$. As we do not know $N_{E_1}$, we approximate this value using a *homogeneity assumption*: We assume that the local density of $P$-relative prime pairs within the interval of length $E$ is approximately equal to the average density $\square_{P,E}$ over a period $T_P$ according to eq.9: $N_{E_1} \approx \square_{P,E} \cdot E$. Taken together, we write:

$$N_E \geq N_{E_1} \cdot \square_{P,E} \cdot E \geq E \cdot \frac{\sqrt{E}}{2 \cdot \frac{\sqrt{E}}{2}} = \frac{1}{3} \sqrt{E} \quad (11)$$

If $N_{E_1} \approx \square_{P,E} \cdot E$ could be transformed into the form $N_{E_1} \geq \square_{P,E} \cdot E$ for $\square > 0$, Goldbach's conjecture would be proven. As we found no strict possibility for this transformation, we first draw some interesting conclusions from our relations eq.9 and 11.

The number $N_E$ of Goldbach partitions for $E$ is only slightly larger than $N_{E_1}$ in most cases and we approximate $N_E$ with $N_{E_1}$ giving:

$$N_E \cdot \frac{E}{2} \cdot \frac{\sum_{P \in \mathbb{Z}} P \cdot \square_{P,E} \cdot \frac{\sum_{P \in \mathbb{Z}} P \cdot \square_{P,E}}{P \cdot \square_{P,E}}}{P} \quad (12)$$

with $P = \text{primes}$, $P = \text{nearest prime} \cdot \sqrt{E}$

To illustrate the power of this approximation, we consider $E$ around the primorial $P!_{17} = 510'510 = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17$ and compare correct $N_E$ with approximations. Tab. 1 shows the well known resonance at $P!_{17}$ that is described with high accuracy also with our approximation. From this follows that the resonances are due to the fact that a primorial is *in phase* with all lowest primes up to $P$ leading to higher factors in eq.12 than neighbouring $E$: With the exception of $E=510'504$ and $510'516$, being *in phase* with 3 and giving smaller resonances, all other $E$ are *out of phase* for all primes 3...17. With our conception of *in phase* and *out of phase*, the strange resonances of the number of prime pairs for certain even integers $E$ can be explained.

<table>
<thead>
<tr>
<th>$E$</th>
<th>$N_E$ correct</th>
<th>$N_E$ approximated with eq.12</th>
<th>relative deviation in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>510'502</td>
<td>4'641</td>
<td>4'868</td>
<td>4.9</td>
</tr>
<tr>
<td>510'504</td>
<td>9'434</td>
<td>9'890</td>
<td>4.8</td>
</tr>
<tr>
<td>510'506</td>
<td>4'557</td>
<td>4'868</td>
<td>6.8</td>
</tr>
<tr>
<td>510'508</td>
<td>4'998</td>
<td>5'306</td>
<td>6.2</td>
</tr>
<tr>
<td><strong>510'510</strong></td>
<td><strong>18'986</strong></td>
<td><strong>20'142</strong></td>
<td><strong>6.1</strong></td>
</tr>
<tr>
<td>510'512</td>
<td>4'534</td>
<td>4'868</td>
<td>7.4</td>
</tr>
<tr>
<td>510'514</td>
<td>4'730</td>
<td>4'976</td>
<td>5.2</td>
</tr>
<tr>
<td>510'516</td>
<td>9'816</td>
<td>10'160</td>
<td>3.5</td>
</tr>
<tr>
<td>510'518</td>
<td>4'619</td>
<td>4'868</td>
<td>5.4</td>
</tr>
</tbody>
</table>

**Tab. 1**: Number of prime pairs and approximation with eq.12 around $E = P!_{17} = 510'510$
Convergence of the approximation eq.12 with increasing E

The approximation eq.12 is based on our assumption of regularly distributed relative prime pairs (density assumption), i.e. we assume that the local density of P-relative prime pairs measured over the interval E is equal to the mean density over the much larger interval T_p. Due to $P \sim E^{0.5}$ and $T_p = P!$ (P! stands for primorial), the fraction $E/T_p$ very quickly converges to zero. This makes our density assumption a very strong assumption and we would expect that the precision of the approximation decreases with increasing E. Astonishingly, the contrary seems to be true as shown by our calculations in the range up to $3 \cdot 10^7$ summarized in Fig. 1. The standard deviation $\sigma$ (sigma) of the relative error decreases with increasing E, whereas its mean increases. However, the latter can be corrected by increasing P using a smooth function $f(E)$:

$$
P \sigma f(E) \cdot \sqrt{E} \sigma 0.505 \cdot E^{0.0663} \cdot \sqrt{E}\) \tag{13}

The remaining standard deviation $\sigma$ decreases, as shown in Fig. 1, according to:

$$
\sigma 1.494 E^{-0.394}\) \tag{14}

Fig. 2 shows the precision of the approximation eq.12 for $E \approx 3 \cdot 10^7$.

![Graph showing precision of approximation eq.12](image)

**Fig. 1:** Standard deviation sigma (squares) and mean-correction factor $f$ (diamonds) for the relative error between the correct number of Goldbach partitions and its approximation by eq.12. Calculations are based on 100 to 1000 consecutive even integers E for each point.
A strong form of Goldbach's conjecture

The increasing relative accuracy of the approximation eq.12 lets us suggest a much stronger version of the Goldbach conjecture. The average prime density in the interval $1...Z$, according to the Gauss estimate eq.1, is proportional to $(\ln Z)^{-1}$. Considering, according to the spirit of our method, that the density of P-relative prime pairs $\prod_{\mathfrak{p} \in \mathfrak{E}}$ involves (for the general case of out of phase situations) two independent prime sequences, the respective density can be expected to be proportional to $(\ln Z)^{-2}$. This leads us to a much stronger conjecture than stated above (based on eq. 10 and 11):

$$N_E \geq N_{E,\text{all out of phase}} \left( \sum_{\mathfrak{p} \in \mathfrak{E}} \prod_{\mathfrak{p} \in \mathfrak{E}} \frac{P_{\mathfrak{p}}^2}{P} \right) \geq \frac{E}{(\ln E)^2} > \frac{1}{2} \sqrt{E} \quad \text{for } E > 10 \quad (15)$$

The most right inequality strictly holds for all $E>10$. Based on the Gauss estimate eq.1 for the prime density, it is possible to show that also the other inequality holds for the limit $P_{\mathfrak{p}}^2$ (proof to be published later). The position of this strong form of Goldbach's conjecture is indicated in Fig. 2.

Discussion

A literature search resulted in an indication of our factors P-1 and P-2: Hardy and Littlewood\textsuperscript{12} (1923) comment on the term:

---

Fig. 2: Our approximation eq.12 with $P = 1.60 \cdot E^{0.5}$ (squares) compared with the correct numbers of Goldbach partitions (points) around $E \approx 3 \cdot 10^7$. The horizontal line at 101'209 indicates our proposition for a strong form of Goldbach's conjecture.
with reference to Sylvester\textsuperscript{13,14} (1871) with the words:".. to which the irregularities of \( N_2(n) \) are due" (\( N_2(n) \) being prime pairs within sequence \( n \)) and "There is no sufficient evidence to show, how he was led to his results" ("he" refers to Sylvester). Sylvester's approach proved to be inadequate and later, Brun\textsuperscript{15} (1915) introduced the factor again into his analysis of prime pair frequency. More than a century ago, \textit{in phase} and \textit{out of phase} may have played their role in the minds of mathematicians occupied with the Goldbach conjecture.

Further, the term \((\ln E)^2\) can be found in the given reference to Hardy and Littlewood. However, more related to our strong Goldbach conjecture (eq.15) is a heuristic probabilistic approach given by Max See Chin Woon\textsuperscript{16} (2000) involving a sum of the products of reciprocal logarithms. Our proposition has the advantage of simplicity and it is based on a clearly defined method resulting in a high precision estimate of the number of Goldbach partitions. Even nearer to our approach comes Aktay\textsuperscript{17} (2000), who describes our approximation formula (eq.12) in words and gives numerical examples, demonstrating that his procedure is identical to eq.12. Further, he mentions our stronger form of Goldbach's conjecture (eq.15), but without any clear reasoning. However, Aktay's deduction of eq.12 concentrates on the interval \((1...E)\) without involving the much larger intervals \( T_P \) and so, his deduction contains rounding errors. Especially, the essential step from \( T_P \) to \( E \) being the source of statistical fluctuations and our correction function \( f(E) \) is missing and therefore, the result lacks clarity. Then, convergence of the standard deviation to zero with increasing \( E \) (eq.14) as well as \( f(E) \) (eq.13) are not mentioned.

With respect to potential applications of the described process for the construction of relative prime pairs within \( T_P \), its result can be summarized as follows: The process leads to arbitrarily long, aperiodic and palindromic sequences with vanishing density fluctuations. These properties might be useful for a new class of statistical functions and their application in physics.

\textbf{Conclusions}

Our analysis of periodicity of \( P \)-relative prime-indicators after applying the Erathostenes sieve up to a prime \( P \) leads to an approximation for the number \( N_E \) of Goldbach partitions for each even integer \( E \). The basic element in our method is the notion of \textit{in phase} and \textit{out of phase} of two overlapping mirror symmetric sequences of \( P \)-relative prime indicators (1 if position holds a \( P \)-relative prime and 0 otherwise). Due to our assumption of regularly distributed \( P \)-relative prime pairs (density assumption), our approximation cannot be transformed into a hard lower bound in the mathematical sense for \( N_E \) and therefore cannot strictly prove the Goldbach conjecture. However, the power of the approximation is remarkable and it can explain large differences of \( N_E \) for neighbouring \( E \)'s (resonances). Further, the standard deviation of the relative error of the approximation decreases according to a power law with increasing \( E \) (at least up to \( E=3\cdot10^7 \)). Considering that the fraction \( E/T_P \) dramatically decreases with increasing \( E \), we conclude that our density assumption seems to hold to a unexpected high degree. This might be a hint towards a underlying basic number theoretic law that is strictly valid in the limit \( P \rightarrow \). Based on the numerically observed convergence of the relative errors towards zero, a strong form
of the Goldbach conjecture is obtained, stating that the number of Goldbach partitions increases with \( E/(\ln E)^2 \). Taken together, our observations suggest a statistical proof of Goldbach's conjecture with very rapidly vanishing probability of failure for large \( E \): With a standard deviation of roughly \( 2\% \) around \( E=3\cdot10^7 \), failure would mean to find an outlier 500 standard deviations away from the average (with assumed Gauss distribution of relative errors) occurring with a probability below \( 10^{-50000} \).

**Literature:**