Generalization of the Goldbach-Polignac conjectures and estimation of counting functions

Generalization of the Goldbach-Polignac Conjectures and Estimation of Counting Functions

F. Gassmann[(1)] and A. Marques Filho[(2)]

[(1)]*Project Complex Systems, Paul Scherrer Institut, CH-5232 Villigen PSI, Switzerland, Tel. +41 310 26 47; Fax. +41 310 44 16, E-mail* gassmann@psi.ch *corresponding author*

[(2)] *Instituto Nacional de Pesquisas da Amazônia, P.O. Box 478, MCT/INPA, Manaus, Brazil, E-mail* ari@inpa.gov.br

Abstract

We develop an asymptotic prime counting function based on the sieve process of Erathostenes. This function can be generalized to include a finite set of conditions imposed on sums (generalized Goldbach conditions) or differences (generalized Polignac conditions) of M primes. With one additive condition for two primes p_1, p_2 (M=2) with $p_1+p_2=E$, the generalized counting function approximates the number of Goldbach partitions and its average lower bound. Its relative errors decrease with increasing even number E, and so enable it to explain the structure of the Goldbach-pair counting function. With one subtractive condition for two primes (M=2), the number of prime pairs with a given distance (e.g. twin primes with distance 2) in the interval 2...Z (Z=integer) can be estimated. For more than two primes (M>2), additive and subtractive conditions can be mixed. As examples, counting functions are considered to investigate strings of M=5 or more equidistant primes, twin-Goldbach pairs (M=4), and well known triplets (M=3) or quadruplets (M=4). The generalized counting function is based on an equality assumption we cannot prove (a proof would be equivalent to a simultaneous proof for the Goldbach conjecture and other conjectures involving more than 2 primes). However, reasons given for its validity include an observed scaling relation for the quasi-random errors of the generalized counting function, leading to a heuristic "proof" for the Goldbach conjecture.

2000 MSC: 11N36 Applications of sieve methods

The problem

Instead of trying to solve one of the hardest problems in number theory, we make the task even more difficult! Beyond asking for at least one prime pair adding up to a given even number E, we are e.g. interested in the number of prime pairs $p_1+p_2=E$ that are at the same time members of twin primes in such a way that p_1-2 and p_2+2 are also primes. For E=36 the three *twin-Goldbach pairs* would be 7+29, 19+17 and 31+5. For which E≥12 are there solutions and how many? Our generalized counting function answers these and other questions with an accuracy quickly increasing with increasing E. Does our method show a way heading towards a proof of the Goldbach conjecture?

Introduction

Goldbach's conjecture[1] from 1742 states that every even integer E>2 can be expressed as sum of two primes (in the following called *Goldbach pairs* or *Goldbach partitions*). Though computer-based numerical tests[2] have confirmed the theorem up to $4\cdot10^{14}$ (and ongoing computations have increased this limit[3] to over $8\cdot10^{15}$), it resisted any formal mathematical proof up to now[3,4]. Polignac's conjecture[3,5] from 1849 states that there exist infinitely many pairs of consecutive primes with any given even difference D≥2. A necessary condition for the conjectures to be valid is, that new primes steadily appear on the number axis without limit. The respective proof was already given by Euclid around 300 BC and has been confirmed repeatedly in different forms since then[3,6,7].

We performed previous studies on the structure of prime-multiples and their effects on the number of Goldbach partitions (Marques Filho & Walker, 2001[8]). A series of numerical tests up to $E=3\cdot10^7$ confirmed the developed approximation $N_G(E)$ for the Goldbach-pair counting function (Marques Filho, Gassmann & Walker, 2005[9]). The computations showed the relative error of $N_G(E)$ decreasing proportional to approximately $E^{-0.4}$ for increasing E. Further computations to $E = 2.56\cdot10^{10}$ showed converging of this power law towards $E^{-1/2}$.

In the present paper, we will generalize our approach. We begin with a slightly modified sieve process to develop an asymptotic prime counting function. A generalization of this function, on the basis of an *equality assumption* (to be defined later), will unite the sets of primes, Goldbach pairs, twin primes, and higher prime combinations, to members of a family called *generalized Goldbach-Polignac problems*. To demonstrate the validity of our method, we apply it to the special case of Goldbach pairs and develop a calibrated Goldbach-pair counting function $N_G(E)$ for $E=4...2.56\cdot10^{10}$. After showing other applications, we investigate errors of $N_G(E)$ and show them to scale with the square root of $N_G(E)$. We propose reasons for the validity of our equality assumption. Its proof would be equivalent to simultaneous proofs of the Goldbach conjecture and other conjectures involving more than two primes, and so is clearly beyond the scope of the present work. However, a heuristic "proof" of the Goldbach conjecture, based on the observed scaling law, is given at the end of this paper.

Approximation for the prime counting function

Gauss[10] estimated the asymptotic total number $\pi(Z)$ of primes smaller than a given positive integer Z in 1792, at the age of 15, to:

$$\pi(Z) \approx \frac{Z}{\ln Z} \tag{1}$$

ln is the natural logarithm throughout the paper. This nowadays called *prime number theorem* could be strictly proven only in 1896. To introduce our method and our definitions, we give another approximation for $\pi(Z)$, based on the sieve method which Erathostenes[11] proposed around 300 BC as an efficient means to find prime numbers.

Number of P-relative primes up to Z

We consider the sequence of integers after application of a slightly modified Erathostenes sieve up to a prime P: multiples of all primes $P_L \leq P$ are assumed to be set to 0 (including P_L, in contrast to the normal procedure). For P=5 and Z=50, the resulting integer sequence is 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 11, 0, 13, 0, 0, 0, 17, 0, 19, 0, 0, 0, 23, 0, 0, 0, 0, 0, 29, 0, 31, 0, 0, 0, 0, 0, 37, 0, 0, 0, 41, 0, 43, 0, 0, 0, 47, 0, 49, 0. The process shows that all non-zero integers up to P_N^2-1 are real primes, $P_N > P$ being the next higher prime following P. P_N^2 (49 in our example) is the lowest *P-relative prime* not being a real prime, i.e. it cannot be divided by any integer smaller than P_N. By substituting all non-zero integers by 1, we get the *5-relative prime-indicator sequence*

0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0.

To find a generalizable procedure, we introduce the notion of *phases*. To perform the iteration step on the *relative prime-indicator sequence* involving P, we define phases $□_i$ □ {0...P-1} as the set of all positions $□_i + k \cdot P \leq Z$ (k=1,2...). Our central assumption (called *equality assumption*) states that all the ones (and the zeros) are almost equally distributed on all P phases. As the combined set of all phases contains the whole set of ones, a deletion of □ different phases reduces the number of ones by approximately □/P, or multiplies it by (1- □/P) = (P- □)/P. Here, deletion of a phase means setting all its members to 0. For the special case of the Erathostenes method, □=1 for all P, because exactly one phase (phase 0) is deleted in each iteration step. It follows for the approximative number $N_{□,P}$ of P-relative primes (i.e. the number of ones in the *relative prime-indicator sequence*) below Z:

$$N_{□,P} \approx Z \cdot \prod_{P'=2}^{P} \frac{P' - 1}{P'} \tag{2}$$

In this and the following relations, the product term always involves all primes P' within the indicated boundaries.

Number of primes below Z

To find an approximative asymptotic prime counting function $N_□(Z)$, the procedure has to be repeated until $P \approx Z^{1/2}$ giving:

$$N_□(Z)\ \mu\ □ \cdot Z \cdot \prod_{P'=2}^{\sqrt{Z}} \frac{P' - 1}{P'} \tag{3}$$

We use μ instead of the equal sign to indicate asymptotic convergence. Due to omission of all primes $2...P \leq Z^{1/2}$, and also to a slight asymmetry of phase 0 (explanation see later), a correction factor □ has been introduced. It can be determined using the prime number theorem together with the following asymptotic relation found by Landau[12]:

$$\left(\frac{e^{\gamma}}{2}\right)\left(\prod_{P'=2}^{\sqrt{Z}} \frac{P' - 1}{P'}\right)\ \mu\ \frac{1}{\ln(Z)} \tag{4}$$

Here, γ is the Euler constant (0.577215...). Multiplication of (4) with Z and comparing with (1) and (3) leads to the value of □:

$$□ = \tfrac{1}{2} e^{\gamma} \approx 0.890536... \tag{5}$$

□<1 means that phase 0 (the only phase deleted), in the average, contains somewhat more ones than we would expect on the basis of our equality assumption. We will investigate the reason for this asymmetry towards the end of this report. We would like to mention that the product term in (3) converges to the inverse of the Riemann Zeta-function with argument 1 for large □:

$$\frac{1}{\zeta(1)} = \lim_{Z\to\infty} \prod_{P'=2}^{\sqrt{Z}} \frac{P'-1}{P'} = 0 \qquad (6)$$

Generalization of the method for M primes

Imposing M-1 conditions on M-tuples of primes p_0, p_1,...p_{M-1} is possible by defining additive ($p_0 + p_i = Z_i$, generalized Goldbach conditions) or subtractive ($p_0 - p_i = Z_i$, generalized Polignac conditions) or mixed conditions (we use the term "generalized Polignac conditions" for differences between primes that are not necessarily consecutive). Obviously, each condition i=1...M-1 leads to deletion of a certain phase, and therefore (3) has to be generalized for the deletion of a variable number ω of phases in each iteration step:

$$N(Z) \propto \varepsilon(M) \cdot Z \cdot \prod_{P'=2}^{\sqrt{Z}} \frac{P'-\omega(P')}{P'} \geq 0 \qquad (7)$$

The asymptotic number of representations N(Z) is a subset of $\pi(Z)$. $\varepsilon(M)$ is a prefactor depending on the number M of primes involved and the nature of the conditions. Analogous to different calibrations, developed for better approximations of $\pi(Z)$ at small Z, the systematic error of (7) can be reduced by a second prefactor $\varepsilon'(M, Z)$ converging to 1 for large Z (an example for M=2 will be given below). $\omega(P')$ is the number of *different* phases φ_i to be deleted in the iteration step involving P': They always include phase 0 and are defined by the M-1 conditions imposed on the prime M-tuples:

$\varphi_0 = 0$ (8a)

$\varphi_i = Z_i \bmod P'$ (i=1..M-1) (8b)

mod is the modulo-function (*a mod b* is the rest of the integer division a/b, e.g., 162 mod 5 = 2). For Goldbach pairs, M=2 and Z_1=Z. According to (8), e.g. for Z=162, the phases deleted for P'=2,3,5,7,11 are (0,0),(0,0),(0,2),(0,1),(0,8) and the respective numbers of *different* phases $\omega(P')$=1,1,2,2,2.

Restrictions on Z and Z_i implied by M-1 conditions

Possible values for ω are 1...M. M=1 imposes no constraints. Restrictions for Z are imposed for M>1 to prevent factors of zero in the product term of (7). These restrictions are effective for P'≤M. For Goldbach pairs (M=2, Z_1=Z), P'=2 restricts Z to even integers E, because the phase φ_1=1, resulting from odd Z, would imply deletion of two phases, and (7) would evaluate to 0. However, also the zero resulting from odd Z is a correct approximative result, because the sum of two primes is an odd number only in rare exceptional situations when prime 2 is involved (e.g. Goldbach pairs for Z=19 are (2,17) and (17,2)). If (7) evaluates to a negative number (possible, e.g., for M=4), the correct result has to be zero, indicated by the condition ≥0 in (7).

The following restrictions for Z_i are necessary for the applicability of (7). For *additive conditions*, all $Z_i^{1/2}$ should imply the same maximum prime P_{max} to be considered in the product term in (7), giving the condition:

$P_{max}^2 < Z_i < P_N^2$, where P_N is the next higher prime following P_{max} (9a)

If (9a) is valid, the different phases involved have approximately equal lengths, and the equality assumption can be applied. However, (9a) might be too restrictive in some cases and, especially for large $Z_l=Z$, this restriction is of minor importance, because the respective factor $(P_N-\square)/P_N$ in (7) becomes nearly unity.

For *subtractive conditions*, P_{max} is defined by Z, and by definition, all Z_i (i>0) are smaller than Z. For $Z_i<0$, the relevant phase length is Z, imposing no further constraints for the applicability of the equality assumption. However, for $Z_i>0$, the relevant portions of the involved phases are confined between Z_i and Z with lengths $Z-Z_i$. Consider e.g. the extreme case $Z_i=Z$: No prime pairs p_0, p_i exist with $p_0<Z$ and $p_0-p_i=Z_i=Z$. A restriction is therefore:

$$Z_i<<Z \tag{9b}$$

Examples for the application of the generalized counting function (7)

(7) is based on our equality assumption applied for each prime $P'=2...Z^{1/2}$. Before we try to make this assumption plausible, we show different applications of (7), because they convincingly demonstrate its validity.

Calibrated formula for Goldbach partitions for the range 4...2.56·10¹⁰

Numerical tests with Goldbach pairs for even numbers (M=2, Z=E), performed by one of the authors (Gassmann) for E in the range $4...2.56\cdot10^{10}$, give the following calibrated approximation:

$$N_G(E) \square \frac{E}{1 \square \frac{1}{a+b\cdot\ln(E)}} \square^2 \prod_{P'\square 2}^{\sqrt{E}} \frac{P'\square\square(P')}{P'\square} \quad \text{with } a \approx -0.5455,\ b \approx 0.5153 \tag{10}$$

$\square(P')$, according to (7, 8), has the two possible values 1 (P' divides E) and 2 (P' does not divide E). The asymptotic prefactor $\square^2$ (defined by (5)) gave a very good approximation for large E. a and b were calibrated for the range $4...2.56\cdot10^{10}$. Deviations of the approximation (10) from the exact numbers of Goldbach pairs never exceeded $3.61\cdot N_G^{1/2}$ for all E=4...1'000'000. This observation suggests the following *scaling law for the absolute errors e*, leading to a universal distribution function for the normalized errors e_n:

$$e_n(E) = \frac{e(E)}{\sqrt{N_G(E)}} \tag{11a}$$

The variance σ_n^2 for this normalized error distribution is the average of the squares of $e_n(E)$ taken over many different E:

$$\sigma_n^2 = av\left[\frac{e^2(E)}{N_G(E)}\right] \tag{11b}$$

Fig. 1 gives an impression of the astonishingly good performance of (10) around $E=3\cdot10^7$. Fig. 2 shows an almost perfect Gaussian distribution for the normalized errors.

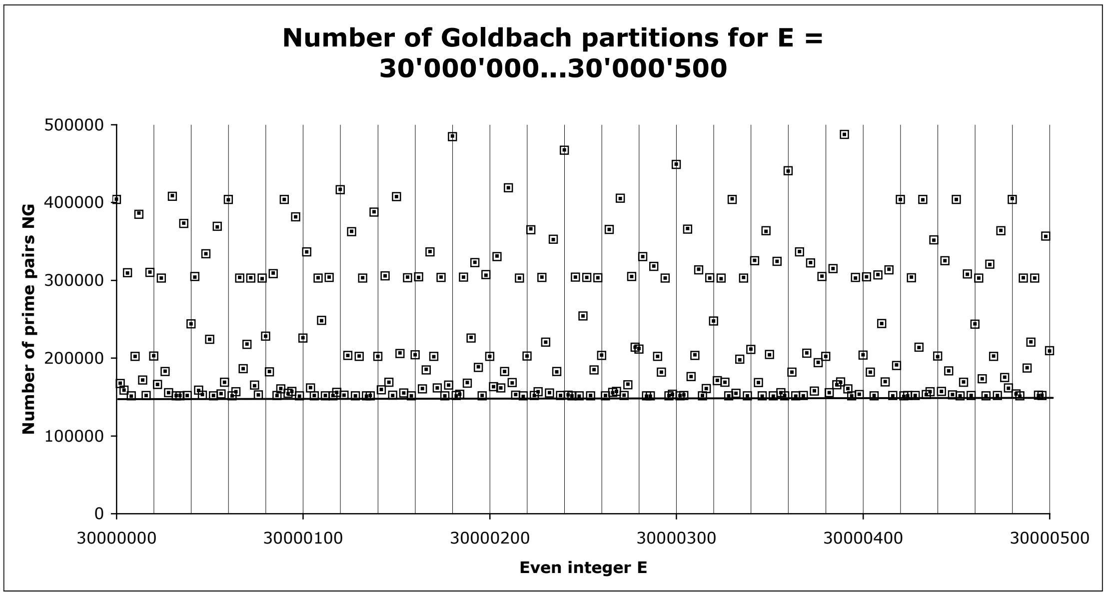

Fig. 1: **"Goldbach Comet" near E=3·10^7.** Our approximation (10) (squares) as compared to the exact numbers of Goldbach partitions (points). The squares encompass about $\pm 10 \cdot N_G^{1/2}$. Maximum errors in this range are $\pm 2.4 \cdot N_G^{1/2}$. The branch near 300'000, stemming from P'=3, contains E's being multiples of 3 and lies a factor of 2 over the well-defined lower boundary, indicated by the horizontal line at 151'869 according to (14a). For detailed explanations see text below.

Structure of $N_G(E)$

The familiar phenomenon we call *resonances*, observed with numerical experiments on Goldbach pairs (see, e.g., highest points in Fig. 1), finds a simple explanation based on (10). Consider, e.g., the well known resonance at the primorial $P!_{17}$ = 510'510 = 2·3·5·7·11·13·17 with 18'986 Goldbach pairs, whereas for 510'508 and 510'512, the respective numbers are 4'998 and 4'534, leading to ratios of 3.80 and 4.19. By definition, 510'510 involves only phase 0 for the lowest 7 primes, and □(P')=1 for P'=2...17. For the other two numbers, respective □(P') are 1, 2, 2, 2, 2, 2, 2 resulting in the ratio:

$$\frac{N_G(510510)}{N_G(510510 \pm 2)} \approx \frac{1}{1}\frac{2}{1}\frac{4}{3}\frac{6}{5}\frac{10}{9}\frac{12}{11} = \frac{2^7}{33} \approx 3.88$$

lying very near to the above given exact ratios. The asymmetry of the two exact numbers stems from higher P', about half of the difference being attributed to P'=23, giving phase 0 for 510'508, but not for 510'512.

In general, the main structure of the counting function $N_G(E)$ results from the □(P') attributed to the *lowest P'*. Higher P' are responsible for finer detail in the distribution. Its most prominent branches result from P'=3 and 5 with ratios 2 (3 divides E, compared to 3 does not divide E), 4/3 (same with 5) and 8/3 (3 and 5 combined) when compared with the well-defined lower boundary (14a). The branch with ratio 2, stemming from P'=3, is clearly recognizable in Fig. 1.

Lower bound of the generalized counting function

If Z fulfils the above explained restrictions (P'-ε(P')≥1 for all P'≤M), we will normally find ε(P')=M for many P'>M. The minimum number of representations $N_{inf}(Z)$ is defined by:

$$N_{inf}(Z) \propto \varepsilon(M) \cdot \alpha' \cdot Z \cdot \prod_{P'>M}^{\sqrt{Z}} \frac{P'-M}{P'} \tag{12}$$

α' contains all factors of the product related to P'=2...M.

For *Goldbach partitions* (M=2, Z=E), (10) in the form of (12) can be transformed using (1), (4), (5) and the asymptotic relation[3,13]

$$c_{TP} = \prod_{P'\geq 3} \frac{P'(P'-2)}{(P'-1)^2} = 0.66016\ldots \text{ (twin-prime constant)} \tag{13}$$

into the following approximation for the lower bound:

$$N_{G,inf}(E) \approx \frac{2 \cdot c_{TP} \cdot E}{(\ln E)^2 \cdot \left(1 + \frac{1}{a+b\cdot\ln E}\right)} \tag{14a}$$

α' was set to 1/2 due to the first term omitted in (12) for P'=2: (P'-1)/P'. The asymptotic minimum density of Goldbach pairs, being proportional to $(\ln E)^{-2}$, suggests a superposition of quasi-randomly distributed primes with density $(\ln E)^{-1}$.

Related to this strong Goldbach conjecture (14a) is a heuristic probabilistic approach given by Max See Chin Woon[14] (2000) involving a sum of products of reciprocal logarithms. Our bound (14a) has the advantage of simplicity, and it is based on a clearly defined method resulting in a high precision estimate of the minimum number of Goldbach partitions. Even nearer to our result (14a) comes Aktay[15] (2000), who describes our approximative relation (7) for the special case of Goldbach pairs in words. Further, he mentions the asymptotic limit of our stronger form of Goldbach's conjecture (14a), but without any clear reasoning and with a lower factor (1/2 instead of $2\cdot c_{TP}$).

(14a) can be used to explain our observation mentioned in the introduction of this report, that relative errors e_{rel} decrease with a power of E approaching -0.5. With the scaling law (11a), we find for the relative errors:

$$e_{rel} = \frac{e}{N_G} = \frac{e_n\sqrt{N_G}}{N_G} = \frac{e_n}{\sqrt{N_G}} \tag{15a}$$

For a fixed $N_G=N_{G,inf}$, the same relation is also true for the respective standard deviation. With σ_n=constant, and the asymptotic lower bound $N_{G,inf}\sim E/(\ln E)^2$, we find the relation:

$$\sigma_{rel} = \frac{\sigma_n}{\sqrt{N_{G,inf}}} \sim \frac{\ln E}{\sqrt{E}} \tag{15b}$$

To investigate the power law for the decrease of the standard deviation of the relative errors σ_{rel} with increasing E, we take the logarithm of (15b) and differentiate:

$$\ln \square_{rel} = a + \ln\{\ln E\} - \tfrac{1}{2}\ln E$$
$$\frac{d\ln \square_{rel}}{d\ln E} = \frac{1}{\ln E} - \frac{1}{2} \tag{15c}$$

For large E, the term 1/lnE vanishes, and we get the asymptotic power law:

$$\square_{rel} \sim E^{-\frac{1}{2}} \tag{15d}$$

in accordance with our calculations.

Special questions regarding the distribution of primes

(7) gives answers to different questions related to the distribution of primes. In the following, we give examples involving twin primes, Goldbach pairs and strings of primes in arithmetic progression.

<u>Twin primes:</u> Twin primes are prime pairs with difference 2, as 11,13 or 17,19. According to (7), twin primes are addressed with one subtractive condition (M=2, Z_1=2). According to (8), $\square_1$=2 gives $\square$(P')=2 for all P'>2. The number of twin primes is therefore identical with the lower bound for the number of Goldbach pairs (14a):

$$N_{TP}(Z) \approx \frac{2 \cdot c_{TP} \cdot Z}{(\ln Z)^2 \cdot \left(1 - \frac{1}{a+b\cdot\ln Z}\right)} \tag{14b}$$

E.g. for $Z=3\cdot 10^6$, there are approximately 20'708 twin primes according to (14b) with their exact number being 20'932, only $1.6\cdot N_{TP}^{1/2}$ away from (14b), though this relation has been calibrated (constants a, b) for Goldbach pairs, not for twin primes. Hardy & Littlewood[13] (equation 5.311 on p. 43) have conjectured the same asymptotic relation in 1923 (i.e. without the term containing a and b). Their result "*Thus there should be approximately equal numbers of prime-pairs differing by 2 and by 4, but twice as many differing by 6*" immediately follows from (8) and (14b), because with difference 4, the same numbers $\square$(P') of phases are deleted for all primes P' and with difference 6, only one phase is deleted for P'=3 instead of 2.

<u>Strings of primes in arithmetic progression:</u> We can ask, e.g., for M-1 Goldbach partitions $p_0+p_i=E+(i-1)\cdot D$ (i=1...M-1) leading to strings of equidistant primes $p_1...p_{M-1}$. We consider M=5: D has to be chosen such that the four conditions Z_1=E (E=even number), Z_2=E+D, Z_3=E+2D, Z_4=E+3D do not result in factors $\square$0 for P'=2, 3 and 5. For P'=2, the resulting phase has to be always 0, i.e. D must be an even number. For P'=3 it suffices that Z_1(=E), Z_2, Z_3, Z_4 result in the same phase, being not necessarily zero. From this follows, according to (8), the condition:

$\square$ = E mod 3 = (E+D) mod 3 = (E+2D) mod 3 = (E+3D) mod 3

Solutions are E=6k or E=6k+2 or E=6k+4 and D=6k' (k, k'=natural numbers). For P'=5, the only restriction is that not all 4 conditions lead to a different phase different from zero. A simple solution is to require that the phase for E be zero, i.e. E=30k''. A numerical test with multiples of 30 for E and D=6 (i.e. k'=k''=1) around E=3'000'000 shows that (10) is a good approximation (though calibrated for Goldbach pairs) with maximum errors of about 2.5 $N^{1/2}$ around the mean positioned at 1.23 $N^{1/2}$ above the value given by (10). Calculations show that there is no single violation (i.e. N_{exact}=0) from E=30 to E=1'000'020. Further, the approximative counting function

N(E) deviates less than 3.8 $N^{1/2}$ (~4.8 standard deviations) from the exact number of representations for all E=30k up to E=1'000'020. Because the probability for a violation decreases with increasing E, and minimum numbers of representations (N≈54 for E≈10^6) are about 7.3 $N^{1/2}$, we conject that *for every E=30k, there exists at least one Goldbach partition $p_0+p_1=E$ with the property that p_1+6, p_1+12, p_1+18 are also prime.*

In the same way, counting functions for *strings of primes in arithmetic progression* ($p_i = p_0+(i-1)\cdot D$, i=1...M) can be investigated on the basis of subtractive conditions. To prevent deletion of all phases for each P'≤M, the factorization of D has to contain all P'≤M, leading to exclusive deletion of phase 0 for all P'≤M. Ribenboim[3] gives the longest known string of primes in arithmetic progression on p. 209 containing M=22 primes with $D=4'609'098'694'200=2^3\cdot3\cdot5^2\cdot7\cdot11\cdot13\cdot17\cdot19\cdot23\cdot1033$. The factorization of D contains all primes ≤22, in accordance with our analysis. The necessary condition P'≤M shows that the much smaller $D=2\cdot3\cdot5\cdot7\cdot11\cdot13\cdot17\cdot19=9'699'690$ (primorial of 19) might give strings with 22 terms also. The factor $f=2^2\cdot5\cdot23\cdot1033$ increases the chance to find long strings by compensating for the very low density of very large primes (the reported p_0=11'410'337'850'553). We think that our relations (7, 8) help to optimize the factor f for highest probability to find such strings in a given range, and so might speed up computations (see the example for M=10 given below).

On p. 210, Ribenboim[3] gives the longest known *string of consecutive primes in arithmetic progression* with M=10 terms and $D=210=2\cdot3\cdot5\cdot7$. As strings of consecutive primes are a subset of the strings we can investigate with our method, a necessary condition to find infinitely many consecutive primes in arithmetic progression with D=210 is M<11 (the next higher prime after 7). We conclude that strings with 10 terms are the longest possible strings with D=210 (not taking into account *exceptions* as mentioned earlier). Further, the above introduced factor f has been set to the lowest possible value of unity, a plausible choice for the search of *consecutive* primes. Our relations (7,8) with the prefactor of (10) show that p_0 must exceed ≈10^9 (giving N≈5±2) for a reasonable chance to find strings of (not necessarily consecutive) primes in arithmetic progression with D=210. Computations for Z=10^8 show for D=210, $210\cdot11$, $210\cdot11\cdot13$ the following numbers of prime-strings with 10 terms: N=2±1 (1), 12±3 (13), 61±7 (57). The numbers in parentheses are exact results from computerized counting. A comparison with the approximations based on relation (10), calibrated for Goldbach pairs, shows that the exact numbers lie within an interval of less than $N^{1/2}$ around the approximations. These examples demonstrate the considerable increase of the probability to find strings of primes in arithmetic progression, when D is increased by prime factors above 7 (i.e. above the largest necessary prime). The only string for D=210 and Z=10^8 begins at p_0=199. Our computations for strings of primes in arithmetic progression make the high p_0 in the order of 10^{92} plausible, where strings of *consecutive* primes for D=210 with 10 terms were found.

Further, the density of the well known *triplets* (p, p+2, p+6) and (p, p+4, p+6) and the *quadruplet* (p, p+2, p+6, p+8) can be estimated using relations (7,8). Numerical tests for these and different other problems show that the prefactor of (10), calibrated for Goldbach pairs, is always a good first approximation.

A last example we give for M=4 are *twin-Goldbach pairs*, consisting of the subset of Goldbach partitions for which the first prime has a lower and the second has a higher twin prime, i.e. $p_1+p_2=E$, $p_1-2=p_3$, $p_2+2=p_4$ (E=even number, p_i=primes). The condition that not all phases of P'=3 are deleted gives E=6k. Calculations show that for all E=6k between 4'212 and 500'004, at least one twin-Goldbach pair exists. E=4'206 is the largest E below 500'004 for which no twin-Goldbach pair exists. We found an almost perfect Gaussian error distribution (as the one shown in Fig. 2) with mean 0.44 and normalized standard deviation $□_n\approx 1.18$. The maximum deviation from the mean in this range is $-4.5□_n$. With a minimum of 81 solutions found for E around 500'000, only an error exceeding 7 standard deviations (occurring with a probability of the order 10^{-15}) would lead to a violation of our conjecture that *for all E=6k larger than 4'206, at least one twin-Goldbach pair exists*. Note that this conjecture is much stronger than the Goldbach conjecture!

Reasons for the validity of the equality assumption

A rigorous proof for the validity of the equality assumption for M=2 would be equivalent to a proof of the Goldbach conjecture, which is beyond the scope of the present paper. Instead, reasons will be given to make the validity of the equality assumption plausible, using the example of Goldbach partitions. For this aim, we will investigate the propagation of the deletions of certain phases during the iteration process beginning with P=2 and ending with $P_{max}\approx E^{1/2}$.

Initially, all relative prime indicators are set to 1, except for the positions 0 and 1 which are put to 0 (0 and 1 are not primes). For Z=E, phase 1 contains one prime indicator with value 1 (later called one occupied position) less than phase 0. As phase 0 is deleted in every iteration step (and for P=2 only phase 0), slightly more ones than 50% are deleted. (10) assumes deletion of exactly 50% of the number of ones, showing already now its approximative character. As all even positions are deleted (phase 0 of P=2), only the odd positions remain (phase 1 of P=2), i.e. E/2-1 positions are occupied, including all odd numbers 3....E-1.

These occupied positions are distributed in the following way on the 3 phases of P=3:

phase 0 of P=3: positions 3, 9, 15,...

phase 1 of P=3: positions 7, 13, 19,...

phase 2 of P=3: positions 5, 11, 17,...

In each phase, occupied positions form a string in arithmetic progression with separation 6. For all E, phase 0 contains the same number of occupied positions as each of the other two phases or it contains one position more. Numerical tests show that phase 0 contains more occupied positions than averaged over all phases for most of the lowest primes up to about $0.7\cdot E^{1/2}$. This fact is the reason for the prefactor $□^2$ in (10) being smaller than one: Because the systematic deletion of the (most of the time) larger phase 0 leads to a reduction of the remaining occupied positions below the average value given by the product term in (10). This asymmetry concerning phase 0 stems from the asymmetry in the iteration process, always deleting phase 0, whereas the other phases are deleted in a pseudo-random way.

In the next iteration step, involving prime 3, there are three possibilities for the function □(3) in (10): Depending on E, only phase 0 is deleted (□(3)=1) or phases 0 and 1 or 0 and 2 are deleted (□(3)=2). We investigate the effect of the deletion of one phase of P=3 (phase i) on the occupation numbers (i.e. numbers of ones) of phases i_n belonging to any P_n>3. Phase i of P=3 has the above specified occupied positions i_0+6k (i_0=3, 7, 5 for phase 0, 1, 2 respectively, k=0,1,2...). The phases i_n of P_n are related to the phase i of 3 to be deleted in the following way:

$$i_0 + 6k = i_n + k_n P_n \qquad (16a)$$

From this follows

$$i_n = (i_0 + 6k) \bmod P_n \qquad (16b)$$

and obviously, with k increasing one by one, all P_n different phases of P_n will be visited before the cyclic process repeats itself, i.e. *the deletion of phase i of 3 will be distributed almost regularly on all the phases of all primes above 3*. Maximum differences due to this distribution step are 1 position. This means that the remaining occupied positions in all phases for all primes above 3 are almost the same, maximum differences being 2 (1 from different numbers of occupied positions in each phase and 1 from deletion of phase i of prime 3). If two phases of prime 3 are deleted, differences can increase to 3.

Continuing with P=5, the occupied positions are distributed in the following way on the 5 phases of P=5:

phase 0 of P=5: positions 5, 25, 35, 55...

phase 1 of P=5: positions 11, 31, 41, 61…

phase 2 of P=5: positions 7, 17, 37, 47…

phase 3 of P=5: positions 13, 23, 43, 53…

phase 4 of P=5: positions 19, 29, 49, 59…

Occupied positions are now separated by 2·3·5=30, and there are two different strings in arithmetic progression with separation 30 in each phase. The above presented reasoning is true for both sequences of each phase, leading to more complicated superposition effects resulting in larger maximum deviations. In addition, phases deleted during the iteration step with P=3 affect occupations of the phases of P=5.

Due to the described more and more involved superposition effects, we leave exact mathematical reasoning and introduce statistical arguments. Our numerical tests showed that the errors can be normalized according to (11a) and suggest a Gaussian error distribution typical for statistical superposition processes. Fig. 2 convincingly demonstrates such an error distribution calculated with all E in the range E=500'000...721'024.

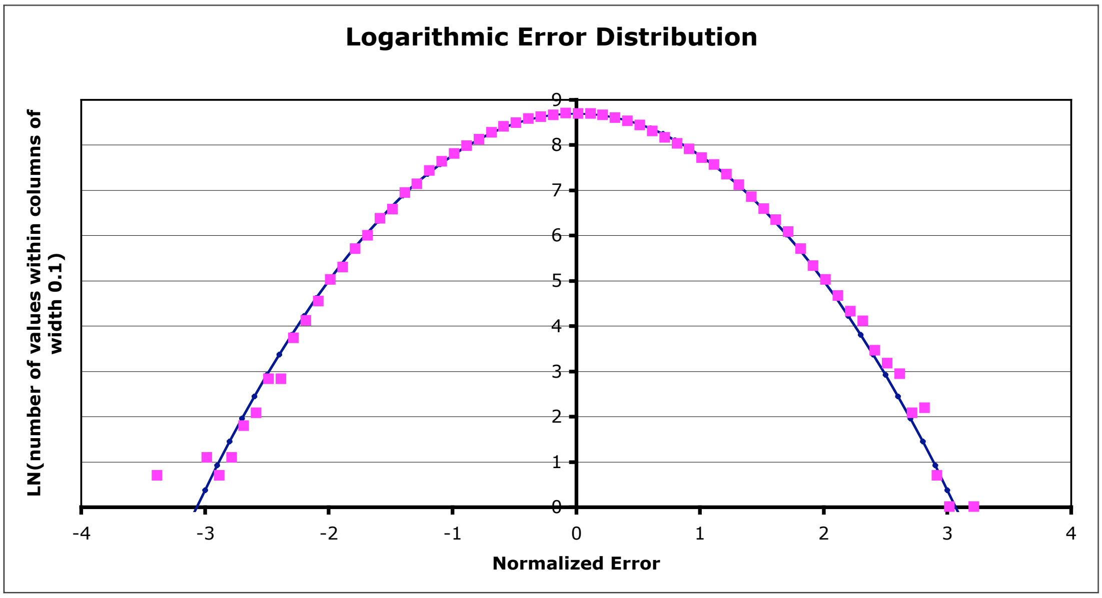

Fig. 2: Natural logarithm of the normalized error distribution for 110'513 even numbers E between 500'000 and 721'024. Errors are defined as deviations (N_G-$N_{G,exact}$) of the number of Goldbach partitions N_G calculated with (10) from the exact number $N_{G,exact}$ based on computerized counting. The abscissa indicates normalized errors $e_n(E)$ according to (11a). The squares are calculated numbers of errors lying within intervals of width 0.1. The line indicates respective values based on a Gaussian distribution with the same normalized standard deviation $\sigma_n \approx 0.736$, calculated according to (11b). The mean lies at -0.216 and was shifted to 0 for comparison with the Gaussian distribution. Higher moments are 0.0564 (skewness) and 0.0188 (excess) which both vanish for a Gaussian distribution. The maximum deviation occurred for $E=712'432=2^4 \cdot 7 \cdot 6361$ and was $4.6\sigma_n$ (a second error was near to this value, both together leading to the square at the bottom left). These two points are the largest normalized errors existing below 10^6. The calculated distribution is not distinguishable from a Gaussian distribution within the range of about 4 standard deviations.

To estimate the probability of a violation of the Goldbach conjecture and also for our heuristic "proof" given below, we approximate the Gaussian error distribution by a binomial distribution $B_{N,p}$ with N steps and probability p for one of two outcomes (in our case decreasing the number of Goldbach pairs by 1). The average μ and the variance σ^2 are:

$$B_{N,p}(k) = \frac{N!}{k! \cdot (N-k)!} p^{N-k} \cdot (1-p)^k$$
$$\mu = Np \quad , \quad \sigma^2 = Np(1-p) = \mu(1-p) \tag{17}$$

Identifying μ with a number N_G of Goldbach partitions, around which we want to investigate the errors, we get the relations:

$$\mu = Np = N_G$$
$$\sigma = \sigma_n \sqrt{N_G} = \sqrt{N_G(1-p)} \tag{18}$$

With σ_n=0.736 (see Fig. 2), we get p≈1/2 and N≈2N_G. A violation of the Goldbach conjecture would therefore have a probability of the order 2^{-N}. For the lower bound (14a), this violation probability p_v would become approximately:

$$p_v \approx 2^{-2E/(\ln E)^2} \tag{19}$$

With increasing E, p_v quickly decreases and reaches a value of about $10^{-200'000'000'000}$ for $E=4\cdot 10^{14}$, the upper end[2] of the range in which numerical tests showed no violation of the Goldbach conjecture.

Heuristic "proof" of the Goldbach conjecture

The above presented estimate was based on a binomial distribution for $N \approx 2N_G$ steps, which is unsatisfactory, because the earlier explained redistribution process involves only far less iteration steps N_S, namely the number of primes in the range $2...E^{1/2}$:

$$N_S \approx \frac{\sqrt{E}}{\ln\sqrt{E}} \approx \sqrt{\frac{2N_G}{c_{TP}}} << 2N_G \tag{20}$$

In (20) the asymptotic lower bound (14a) has been used. We showed that additional deviations from the equality assumption contributed by each step increase during the iteration process. We approximate this fact and describe the build-up of the errors e(E) as a diffusion process involving N_S steps with *average* size d and probabilities of 1/2 for both directions (+d and -d). For our assumed symmetric diffusion process, maximum deviations from its centre are $\pm d\cdot N_S$. Instead of fitting a binomial distribution covering the range $0...N=2N_G$ (with p=1/2 and mean N_G) and adjusting N as we did above, we prescribe now $N_b=2N_S$ and use step size d of the diffusion process to adjust its variance to the observed value:

$$\sigma = \sigma_n\sqrt{N_G} = d\sigma_b = d\sqrt{\frac{N_b}{4}} = d\sqrt{\frac{N_S}{2}} \tag{21}$$

With $\sigma_n^2 \approx 1/2$ (see Fig. 1), we get $N_G=d^2N_S$, and with use of (20):

$$d = \sqrt{\frac{N_G}{N_S}} = \sqrt[4]{\frac{N_G c_{TP}}{2}} \tag{22}$$

Using (20) and (22), maximum deviations around N_G become:

$$d\cdot N_S = \sqrt[4]{\frac{2}{c_{TP}}} N_G^{3/4} \approx 1.32\cdot N_G^{3/4} < N_G \tag{23a}$$

Expressed as multiples of the standard deviation σ (cf. the second term in (21)), maximum deviations around N_G become:

$$\frac{d\cdot N_S}{\sigma} = 2\cdot\sqrt[4]{\frac{N_G}{2c_{TP}}} \approx 1.86\cdot N_G^{1/4} \tag{23b}$$

This estimate shows that maximum deviations $d\cdot N_S$ of the Goldbach counting function below its asymptotic lower bound (14a) are considerably smaller than its lower bound N_G. This heuristic "proof" of the Goldbach conjecture is based on the observed error scaling law (11a) and the resulting binomial error distribution with the assumption that the underlying process is a quasi random diffusion process with a limited number (N_S) of equal steps (d). This conclusion is supported by the fact that among 500'000 even numbers tested in the range 2...1'000'000, only 2 errors deviated more than $4.3\sigma_n$ from the mean, whereas a Gaussian distribution would give a probability of $5.6 \cdot 10^{-5}$ giving 28 values outside this limit. This suggests that the normalized error distribution falls off faster than a Gaussian distribution. A fit of a symmetrical binomial distribution to the normalized error distribution with maximum deviations of $4.3\sigma_n$ from its mean gives:

$$4.3 \cdot \sqrt{N \tfrac{1}{2}\left(1 - \tfrac{1}{2}\right)} = \tfrac{1}{2} N \tag{24}$$

resulting in $N = 4.3^2 \approx 18$. The probability to reach either extreme of this binomial distribution is $2 \cdot 2^{-18} = 2^{-17}$ leading to four values out of 500'000. For comparison with our calculations for all even integers E below 10^6, the range of –3.6...+3.6 has been divided into 18 intervals of width 0.4. In the lowest interval –3.6... –3.2, only the two values on the left side of Fig. 2 (positioned at –3.4) have been found. In the highest interval +3.2...+3.6, only one value on the right side of Fig. 2 (positioned at +3.22) has been found, though another value outside the E-range considered in Fig. 2 was found at +3.19, lying very near to the somewhat arbitrary interval boundary of 3.20. A reasonable estimate is therefore four values as suggested also by the symmetrical binomial distribution with maximum deviations of $4.3\sigma_n$.

Conclusions

Based on a plausible equality assumption, we developed a counting function that can be applied for estimates of the number of representations for generalized Goldbach-Polignac problems involving a finite set of additive and subtractive conditions. Our function encompasses the prime number theorem (no conditions, M=1), the Goldbach conjecture (one additive condition, M=2), twin prime density (one subtractive condition, M=2) and infinitely many higher order problems. The quasi-random errors of the formula (at least for M=1...10) scale with the square root of the number of representations and therefore, relative quasi-random errors vanish for large E. An important exact result of the function is its distinction between sets of conditions allowing only very few exceptional representations or a number of representations monotonically increasing with increasing Z.

A proof of the equality assumption would prove the generalized Goldbach-Polignac conjecture for a finite number of conditions. Though this aim is beyond the scope of the present report, we hope that the equality assumption might be provable in the future. A heuristic "proof", based on statistical properties of the error distribution, together with an assumed diffusion process involving a limited number of equal steps, shows a strategy for combining the statistical nature of the Goldbach problem with absolute limits and so circumventing the nonmathematical way to "prove" the Goldbach conjecture with vanishing probability of violations.

We would like to point to a more general and astonishing fact, namely, that primes do not play an important role in physics, e.g. in quantum mechanics or in the framework of complex systems research. For physical applications, restrictive boundary conditions are always present, because they define the special problem to be investigated. Lacking a general formula involving primes and boundary conditions, primes could not easily penetrate physics. Possibly, our counting function could pave a way into this direction.

Acknowledgment

We would like to express our gratitude to Dra. Ilse Walker from the Instituto Nacional de Pesquisas da Amazônia, Manaus (Brazil) for bringing the two authors together, as well as for her interesting discussions with both of them on prime numbers and possible biological implications.

Literature:

[1] Goldbach, C. (1742) On the margin of a letter to Leonard Euler dated June 7, 1742, Christian Goldbach (1690 - 1764) formulated his famous conjecture.

[2] Richstein, J. (2001) Verifying the Goldbach conjecture up to $4{\cdot}10^{14}$, *Math. Comput.* **70**: 1745-1750.

[3] Ribenboim, P. (2004) The little book of bigger primes, Springer, New York: 356pp.

[4] Wang-Yuan (ed.) (1984) Goldbach's Conjecture, World Scientific Publishing Co. Pte LTd., Singapore: 311pp.

[5] de Polignac, A. (1849) Six propositions arithmologiques déduites de crible d'Ératosthène, *Nouv. Ann. Math.* **8**: 423-429.

[6] Selberg, A. (1949) An elementary proof of the prime number theorem, *Annals of Mathematics* **50**(2): 305-313.

[7] Newman, J. D. (1980) Simple analytical proof of the prime-number theorem, *Amer. Math. Monthly* **87**: 693-696.

[8] Marques-Filho, A. & Walker, I. (2001) The composition of prime numbers and their relation with even numbers: The proof of Goldbach's conjecture, *Série Documentos* **7**, MCT/INPA, Manaus, Brazil: 23pp.

[9] Marques Filho, A., Gassmann, F. & Walker, I. (2005) Approximation for the number of prime pairs adding up to even integers, ETH E-Collection, http://maliera.ethz.ch.

[10] Encyclopedia of Mathematics (1989) Kluwer Academic Publishers, Dordrecht, Boston, London.

[11] Greaves, G. (2000) Sieves in number theory, Springer, Berlin, Heidelberg, New York.

[12] Landau, E. (1912) Gelöste und ungelöste Probleme aus der Theorie der Primzahlverteilung und der Riemann'schen Zetafunktion, *Jahresbericht der Deutschen Math.-Vereinigung* **21**: 218.

[13] Hardy, G.H. & Littlewood, J.E. (1923) On the expression of a number as sum of primes, *Acta Mathematica* **44**: 1-70.

[14] Max See Chin Woon (2000) On partitions of Goldbach's conjecture, arXiv:math.GM/0010027, Cambridge, UK.

[15] Aktay, M. (2000) Goldbach's Rule, arXiv:math.GM/0005191, Cambridge, UK.