Probability to obtain a Goldbach decomposition of an even number (Denise Vella-Chemla, august 2022)

1. An illustrative example

Let us take the example of looking for Goldbach decompositions of the even number $n = 98$.

$$S_{98} = \begin{cases} 98 \equiv 0 \pmod 2 \\ 98 \equiv 2 \pmod 3 \\ 98 \equiv 3 \pmod 5 \\ 98 \equiv 0 \pmod 7 \end{cases}$$

Let us call d_{98} a potential Goldbach decomponent of $n = 98$. d_{98} can be congruent, except 0, to every number that $n = 98$ isn't congruent to. The $\vee$ symbol in the above congruence system is to be read as an exclusive or, its extended use is to be understood as the fact to respect as many congruence systems as combinatorics permits to have (combining one possibility of each exclusive or).

$$S_{d_{98}} = \begin{cases} d_{98} \equiv 1 \pmod 2 \\ d_{98} \equiv 1 \pmod 3 \\ d_{98} \equiv 1 \vee 2 \vee 4 \pmod 5 \\ d_{98} \equiv 1 \vee 2 \vee 3 \vee 4 \vee 5 \vee 6 \pmod 7 \end{cases}$$

Remark : we notice that obeying the system of systems of congruences is a sufficient condition but not a necessary condition to obtain a Goldbach decomponent of n. The demonstration of the validity of this characterization of Goldbach decomponents of an even number n that are greater than $\sqrt{n}$ is provided in section 2.

What can be easily understood is that modules that don't divide n "eliminate more congruence classes" (2 by each prime module lesser than $\sqrt{n}$) than modules that divide n. Let us take the worst case, where two congruence classes are eliminated for each prime module lesser than $\sqrt{n}$ (for n an even number of the form $2^k p$ with p prime for instance), we find all the same

$$\prod_{\substack{p \text{ prime} \\ 3 \leq p \leq \sqrt{n}}} (p-2)$$

different congruence classes by applying the chinese remainder theorem to each of the congruence systems combinatorialy found (see $S_{d_{98}}$ above). All those solutions are lesser than $D = \prod_{\substack{p \text{ prime} \\ 3 \leq p \leq \sqrt{n}}} p$.

Could it be possible to miss the targeted interval, i.e. that all solutions should be greater than n, between n and D ? In section 3, we will see that the probability to obtain at least one solution lesser than n tends to 1 very quickly.

2. Characterization of the Goldbach decomponents of n greater than $\sqrt{n}$[1]

Let $n \in 2\mathbb{N} + 6$ be an even number greater than 6.

[1] Leila Schneps wrote this proof that the characterization of those Goldbach decomponents was valid.

For each $p \in \mathbb{P}^*$ an odd prime number lesser than $\sqrt{n}$ (*i.e.* $3 \leq p \leq \sqrt{n}$), let us define the set :

$$F_n(p) = \{m \in 2\mathbb{N}+1 : 3 \leq m \leq n/2,\ m \not\equiv 0\ [p],\ m \not\equiv n\ [p]\}$$

The intersection of the sets $F_n(p)$ for each p prime between 3 and $\sqrt{n}$ is denoted as :

$$D_n = \bigcap_{\substack{p \in \mathbb{P} \\ 3 \leq p \leq \sqrt{n}}} F_n(p)$$

We are going to show that D_n and its complementary $n - D_n$ contain only prime numbers.

Lemma 1 : Let $m \in 2\mathbb{N}+1$ be an odd integer. If m is divisible by no prime number between 3 and $\sqrt{m}$, then it is prime.

Proof : If m is composite, we have $m = pq$, where p is the smallest prime number in m's canonical factorization in prime numbers and where q is the product of all the other factors. Since m is odd, $p \geq 3$, and since $q \geq p$ (q being the product of integers $\geq p$), $m = pq \geq pp = p^2$ and so $\sqrt{m} \geq p$ (the *sqrt* function being increasing). We have thus shown that if m an odd number is composite, it is divisible by a prime number between 3 and $\sqrt{m}$. The lemma is obtained by contraposition. □

Lemma 2 : $D_n \subseteq \mathbb{P}$

Proof : Let $m \in D_n$. Then $m \in F_n(p)$ for all prime number p between 3 and $\sqrt{n}$. Thus, m is odd and m is divisible by no prime number p between 3 and $\sqrt{n}$ (since $m \not\equiv 0\ [p]$), and so *a fortiori* by no prime number between 3 and $\sqrt{m}$ (since $m \leq n/2 \implies m \leq n \implies \sqrt{m} \leq \sqrt{n}$). By lemma 1, m is therefore prime. □

Lemma 3 : $n - D_n \subseteq \mathbb{P}$

Proof : Let $m \in D_n$. Then $m \in F_n(p)$ for all prime number p between 3 and $\sqrt{n}$. Therefore, $n - m$ is odd (since m is odd and n is even) and $n - m$ is divisible by no prime number p between 3 and $\sqrt{n}$ (since $m \not\equiv n\ [p]$), and thus *a fortiori* by no prime number between 3 and $\sqrt{n-m}$ (because $n - m \leq n \implies \sqrt{n-m} \leq \sqrt{n}$). From lemma 1, $n - m$ is thus prime. □

The sets D_n contain only Goldbach decomponents of n.

Lemma 4 : Let $n \in 2\mathbb{N}+6$. If $D_n \neq \varnothing$, then n verify Goldbach's conjecture.

Proof : If $D_n \neq \varnothing$, it contains an integer p necessarily prime (from lemma 1), such that $q = n - p$ is also prime (from lemma 2), and so $n = p + q$ verify Goldbach's conjecture. □

3. Probability $P(n,k,p)$ to pick a number lesser than or equal to k, without replacement, when we pick uniformly p integers among the n first integers.

The probability[2] $P(n,k,p)$ to pick an integer lesser than or equal to k, without replacement, when we pick uniformly p integers among the n first integers is computed using the following formula :

$$P \quad = \frac{k}{n} + \frac{n-k}{n}\left(\frac{k}{n-1} + \frac{n-k-1}{n-1}\left(\frac{k}{n-2} + \frac{n-k-2}{n-2}\left(\dots\left(\frac{k}{n-p+1}\right)\dots\right)\right)\right)$$

The first term of the sum corresponds to the fact the first number picked is lesser than k. The second term of the sum corresponds to the fact that the first number is greater than k at the first pick, this number is not replaced (we don't have the possibility to pick it another time) and the chance is tried on the remaining numbers, probability being uniform on remaining numbers, etc.

This probability is computed for

$$p = \prod_{\substack{x \text{ premier} \\ 3 \leqslant x \leqslant \sqrt{k}}} (x-2)$$

and

$$n = \prod_{\substack{x \text{ premier} \\ 3 \leqslant x \leqslant \sqrt{k}}} x.$$

The following python program was used :

```
import math

def P(n, k, p):
    assert(1 <= p and p <= n and k <= n-p)
    s, t = 0, 1
    for i in range(p):
        s += t*(k/(n-i))
        t *= (n-k-i)/(n-i)
    return s

for n, k, p in [(30, 26, 3),
                (210, 50, 15),
                (2310, 122, 135),
                (30030, 170, 1485),
                (510510, 290, 22275),
                (9699690, 362, 378675),
                (223092870, 530 , 7952175),
                (6469693230, 842, 214708725),
                (200560490130, 962, 6226553025)]:
    print(f'n = {n}, k = {k}, p = {p} : P_n(k,p) = {P(n, k, p)}')
```

```
n = 30, k = 26, p = 3 : P_n(k,p) = 0.9990147783251231
n = 210, k = 50, p = 15 : P_n(k,p) = 0.9856514594832753
n = 2310, k = 122, p = 135 : P_n(k,p) = 0.9994752040784769
n = 30030, k = 170, p = 1485 : P_n(k,p) = 0.999824267526177
n = 510510, k = 290, p = 22275 : P_n(k,p) = 0.9999976037996607
n = 9699690, k = 362, p = 378675 : P_n(k,p) = 0.9999994514468453
n = 223092870, k = 530, p = 7952175 : P_n(k,p) = 0.9999999955788792
n = 6469693230, k = 842, p = 214708725 : P_n(k,p) = 0.9999999997119475
n = 200560490130, k = 962, p = 6226553025 : P_n(k,p) = 0.9999999921336346
```

[2] Thanks Jacques.

Let us provide its results in the following table :

k	n	k^2+1	p	$P(n,k,p)$
5	30	26	3	0.9990147783251231
7	210	50	15	0.9856514594832753
11	2310	122	135	0.9994752040784769
13	30030	170	1485	0.999824267526177
17	510510	290	22275	0.9999976037996607
19	9699690	362	378675	0.9999994514468453
23	223092870	530	7952175	0.9999999955788792
29	6469693230	842	214708725	0.9999999997119475
31	200560490130	962	6226553025	0.9999999921336346

To confirm the program results, we use a function that calculate the complementary events probabilities, i.e. the probability that during the p pickings without replacement realised under a uniform discrete law in interval $1..n$, all picked numbers would be greater than k according to the formula

$$\overline{P(n,p,k)} = 1 - P(n,p,k) = \frac{n-p}{n}.\frac{n-p-1}{n-1}\dots\frac{n-p-k+1}{n-k+1}.$$

The probability to obtain a Goldbach decomponent of an even number is equal to 1 above prime number 37 if computation precision is fixed to 20 decimal digits.

```
import functools

def prod(iterable):
    return functools.reduce(lambda x, y: x*y, iterable, 1)

P = [2]
Pmax = 100
for n in range(3, Pmax+1, 2):
    if all(n%p for p in P):
        P.append(n)

def Q(n, p, k):
    assert(1 <= p and p <= n and 1 <= k and k <= n-p)
    t, a, b = 1, n-p, n
    for i in range(k):
        t, a, b = t*a/b, a-1, b-1
    return t

print(f'{"N":>3} | {"n":>40s} | {"p":>40s} | {"k":>6} | {"1 - P(n,p,k)":>28} | {"P(n,p,k)":>28}')
print(f'{"-"*3}-+-{"-"*40}-+-{"-"*40}-+-{"-"*6}-+-{"-"*28}-+-{"-"*28}')
for N in P:
    if N > 3:
        n, p, k = prod(x for x in P if x <= N), prod(x-2 for x in P if 2 < x and x <= N), N*N+1
        q = Q(n, p, k)
        print(f'{N:3} | {n:40d} | {p:40d} | {k:6} | {q:28.26f} | {1-q:28.26f}')
```

This program results are provided in the following table :

N	n	p	k	P(n,p,k)
5	30	3	26	0.99901477832512319832147796
7	210	15	50	0.98565145948327537173128121
11	2310	135	122	0.99947520407847800782974446
13	30030	1485	170	0.99982426752617770127073982
17	510510	22275	290	0.99999760379966495804637816
19	9699690	378675	362	0.99999945144687962805818415
23	223092870	7952175	530	0.99999999557885699275061597
29	6469693230	214708725	842	0.99999999999954458651529876
31	200560490130	6226553025	962	0.99999999999993338661852249
37	7420738134810	217929355875	1370	1.00000000000000000000000000
41	304250263527210	8499244879125	1682	1.00000000000000000000000000
43	13082761331670030	348469040044125	1850	1.00000000000000000000000000
47	614889782588491410	15681106801985625	2210	1.00000000000000000000000000
53	32589158477190044730	799736446901266875	2810	1.00000000000000000000000000
59	1922760350154212639070	45584977473372211875	3482	1.00000000000000000000000000
61	117288381359406970983270	2689513670928960500625	3722	1.00000000000000000000000000
67	7858321551080267055879090	174818388610382432540625	4490	1.00000000000000000000000000
71	557940830126698960967415390	12062468814116387845303125	5042	1.00000000000000000000000000
73	40729680599249024150621323470	856435285802263537016521875	5330	1.00000000000000000000000000
79	3217644767340672907899084554130	65945517006774292350272184375	6242	1.00000000000000000000000000
83	267064515689275851355624017992790	5341586877548717680372046934375	6890	1.00000000000000000000000000
89	23768741896345550770650537601358310	464718058346738438192368083290625	7922	1.00000000000000000000000000
97	2305567963945518424753102147331756070	44148215542940151628274967912609375	9410	1.00000000000000000000000000