A Statistician’s Approach to Goldbach’s Conjecture

KEYWORDS:
Teaching;
Goldbach’s conjecture;
Prime number;
Probability.

Neil Sheldon
The Manchester Grammar School, England.
e-mail: nas@mgs.org

Summary
Goldbach’s conjecture is explored by means of probability.

In 1742, Christian Goldbach conjectured that every even number greater than 2 can be written as the sum of two primes. Although Goldbach’s conjecture has been verified for numbers up to 4×10^{14}, it has never been proved. This paper takes a statistical, or probabilistic, look at the problem.

Let us define the function G as the number of ways that a number can be written as the sum of two primes. Then, for example, $G(4) = 1$ because $4 = 2 + 2$ and there are no other ways of writing 4 as the sum of two primes. But $G(10) = 2$ because there are two ways of decomposing 10, namely $3 + 7$ and $5 + 5$. (1, of course, is not a prime number.)

Goldbach’s conjecture may now be expressed as

$G(n) > 0$ for all n in $\{4, 6, 8, \ldots\}$.

It is interesting to see how quickly G grows as n increases.

n	100	1000	10,000
$G(n)$	6	28	127

In fact, except when n is very small, $G(n)$ seems to be comfortably positive. There appears to be no chance of $G(n)$ being zero – that is, no chance of Goldbach’s conjecture being false.

In order to make this last idea more precise, we need to use Gauss’s law for the distribution of primes. In 1793 Gauss gave the approximate formula

$$\pi(n) = \frac{n}{\ln n}$$

for the number of primes less than or equal to n.

Gauss’s formula can be used to produce something like the ‘probability’ that a randomly chosen odd number is prime. (There is no need to consider even numbers as the only even prime is 2. The anomalous existence of a single even prime is ignored from now on.)

The quantity $\pi(2n + 1) - \pi(2n - 1)$ which, for large n, is almost exactly $\frac{2}{\ln(2n)}$, can be interpreted as the expected number of primes in the set $\{2n, 2n + 1\}$. Since there is only one candidate for being a prime, namely $2n + 1$, this is equivalent to saying that $\frac{2}{\ln(2n)}$ is the probability that the number $2n + 1$ is prime. (At this point, some will be horrified at the notion of attaching probability to statements of pure mathematics. Such sensitive souls are advised to stop reading.)

Once we have a formula for the number of primes and a formula for the probability that a number is prime, it is reasonably straightforward to tackle Goldbach’s conjecture.

Consider first just one number: $n = 2{,}000{,}000$, say. Every Goldbach decomposition of this number is a pair of numbers P and Q such that $P + Q = 2{,}000{,}000$ with both P and Q prime. If we take $P \le Q$ then $P \le 1{,}000{,}000$ and $1{,}000{,}000 \le Q < 2{,}000{,}000$.

To evaluate $G(2{,}000{,}000)$ we have to find each prime P and establish whether or not the corresponding Q is prime. Gauss’s formula gives us an estimate of the number of primes P:

$$\pi(1{,}000{,}000) \approx 72{,}382$$

Each of the 72,382 values of P generates a Q that may or may not be prime. Now, if Q is an odd number near 1,000,000, the probability that it is prime, using the formula $\frac{2}{\ln(2n)}$, is 0.1448. For Q near to 2,000,000, the probability is 0.1378. Clearly the probability does not change much in this range. We can say that the expected number of primes Q is at least

$$72{,}382 \times 0.1378 \approx 9978$$

And, assuming independence, we can also say that the probability that none of the Q is prime is at most

$$(1 - 0.1378)^{72{,}382} \approx 10^{-4663}$$

So a conservative estimate of the probability that Goldbach’s conjecture fails for $n = 2{,}000{,}000$ is astonishingly small: less than 1 in 10^{4663}.

It is now a simple matter to extend the argument from a single number to a block of numbers. Let us investigate the probability that Goldbach’s conjecture fails for some number in the range 2,000,000 to 20,000,000.

Again being conservative, we can say that the probability of failure is no greater than

$$\begin{aligned} &1 - (1 - 10^{-4663})^{(10{,}000{,}000-1{,}000{,}000)} \\ &\approx (10{,}000{,}000 - 1{,}000{,}000) \times 10^{-4663} \\ &\approx 10^{-4656} \end{aligned}$$

because there are (10,000,000 – 1,000,000) even numbers to test and each has a probability of 10^{-4663} or less of causing Goldbach’s conjecture to fail. This, again, is a very small probability.

(A more sophisticated analysis would take into account the fact that, when n is a multiple of 6, $G(n)$ is, on average, twice as big as it is when n is not a multiple of 6. This curious result – which is easy to prove – has only a very tiny effect on the subsequent probabilities.)

Finally, consider the sequence of blocks of numbers:

$$2 \times 10^6 \text{ to } 2 \times 10^7,\ 2 \times 10^7 \text{ to } 2 \times 10^8,\\ 2 \times 10^8 \text{ to } 2 \times 10^9, \ldots$$

The probabilities that Goldbach’s conjecture fails in these blocks, calculated conservatively as above, are:

$$10^{-4656},\ 10^{-34{,}100},\ 10^{-261{,}000}, \ldots$$

So the probability that Goldbach’s conjecture fails for some $n > 2 \times 10^6$ is

$$\begin{aligned} &1 - (1 - 10^{-4656})(1 - 10^{-34{,}100})(1 - 10^{-261{,}000}) \ldots \\ &\approx 10^{-4656} + 10^{-34{,}100} + 10^{-261{,}000} + \ldots \\ &\approx 10^{-4656} \end{aligned}$$

Now, it is quite arbitrary that this argument began at $n = 2{,}000{,}000$. Goldbach’s conjecture has been verified up to $n = 4 \times 10^{14}$. With that starting point, we obtain the probability that Goldbach’s conjecture fails for some $n > 4 \times 10^{14}$ as roughly $10^{-150{,}000{,}000{,}000}$.

This probability is 1 in a million million million . . . , where you have to say ‘million’ 25,000,000,000 times. At a reasonable rate of three words per second that would take about 264 years – almost exactly the length of time that has elapsed since Goldbach first made his conjecture.