SAMPLING GOLDBACH NUMBERS AT RANDOM

Ljuben Mutafchiev
Department of Mathematics and Science, American University in Bulgaria and
Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences,
Blagoevgrad, 2700, Bulgaria
ljuben@aubg.edu

Abstract
Let \( \Sigma_2^n \) be the set of all partitions of the even integers from the interval \((4, 2n]\), \(n > 2\), into two odd prime parts. We select a partition from the set \( \Sigma_2^n \) uniformly at random. Let \( 2G_n \) be the number partitioned by this selection. \( 2G_n \) is sometimes called a Goldbach number. In [6] we showed that \( G_n/n \) converges weakly to the maximum \( T \) of two random variables which are independent copies of a uniformly distributed random variable in the interval \((0, 1)\). In this note we show that the mean and the variance of \( G_n/n \) tend to the mean \( \mu_T = 2/3 \) and variance \( \sigma_T^2 = 1/18 \) of \( T \), respectively. Our method of proof is based on generating functions and on a Tauberian theorem due to Hardy-Littlewood-Karamata.

1. Introduction and Statement of the Main Result

A Goldbach number is an even positive integer that can be represented as the sum of two primes without regard to order. The representation itself is called a Goldbach partition of the underlying even positive integer. Let \( P = \{p_1, p_2, \ldots\} \) be the sequence of all odd primes arranged in increasing order. For any integer \( k > 2 \), by \( Q_2(2k) \) we denote the number of Goldbach partitions of the number \( 2k \). In 1742 Goldbach conjectured that \( Q_2(2k) \geq 1 \). This problem remains still unsolved (for more details, see e.g. [4], Section 2.8 and p. 594, [8], Section 4.6 and [9], Chapter VI). Let \( \Sigma_2^n \) be the set of all Goldbach partitions of the even integers from the interval \((4, 2n]\), \( n > 2 \). The cardinality of this set is obviously

\[
| \Sigma_2^n | = \sum_{2 < k \leq n} Q_2(2k).
\]  

(1)

We recently established in [6] that

\[
| \Sigma_2^n | \sim \frac{2n^2}{\log^2 n}, \quad n \to \infty.
\]  

(2)

Consider now a random experiment. Suppose that we select a partition uniformly at random from the set \( \Sigma_2^n \), i.e. we assign the probability \( 1/ | \Sigma_2^n | \) to each
Goldbach partition of an even integer from the interval \((4, 2n]\). We denote by \(\mathbb{P}\) the uniform probability measure on \(\Sigma_{2n}\). Let \(2G_n \in (4, 2n]\) be the Goldbach number that is partitioned by this random selection. Using (2), it is not difficult to show that \(G_n/n\) converges weakly, as \(n \to \infty\), to a random variable whose cumulative distribution function is

\[
F(u) = \begin{cases} 
0 & \text{if } u \leq 0, \\
u^2 & \text{if } 0 < u < 1, \\
1 & \text{if } u \geq 1
\end{cases} \tag{3}
\]

(see [6], Theorem 2). It can be easily seen that \(F(u)\) is the distribution function of the random variable \(T = \max \{U_1, U_2\}\), where \(U_1\) and \(U_2\) are two independent copies of a uniformly distributed random variable in the interval \((0, 1)\).

The goal of this present note is to determine asymptotically the first two moments of a random Goldbach number \(2G_n\). We state our main result in terms of the expectation \(\mathbb{E}\) and variance \(\mathbb{V}ar\) of the random variable \(G_n\), both taken with respect to the probability measure \(\mathbb{P}\). By \(\mu_T\) and \(\sigma_T^2\) we denote the expected value and the variance of the random variable \(T\), respectively. (Its cumulative distribution function is given by (3).)

**Theorem 1.** We have

\[
(i) \lim_{n \to \infty} \frac{1}{n} \mathbb{E}(G_n) = \mu_T = \frac{2}{3};
\]

\[
(ii) \lim_{n \to \infty} \frac{1}{n^2} \mathbb{V}ar(G_n) = \sigma_T^2 = \frac{1}{18}.
\]

Our method of proof is similar to that in [6]. It is essentially based on a generating function identity and on a classical Tauberian theorem due to Hardy, Littlewood and Karamata (see e.g. [7], Theorem 8.7).

**Remark 1.** Our method of proof and further but similar computations yield also asymptotic expressions for moments of higher order of the variable \(G_n/n\).

**Remark 2.** The proof of Hardy-Littlewood-Karamata Tauberian theorem may be found e.g. in [3], Chapter 7.

Our paper is organized as follows. Section 2 contains some preliminaries. The proof of Theorem 1 is given in Section 3.

### 2. Preliminary Results

By a prime partition of the positive integer \(n\), we mean a way of writing it as a sum of primes from the set \(\mathcal{P}\) without regard to order; the summands are called parts. Clearly, Goldbach partitions are prime partitions in which the number of parts is 2.
Consider now the number \( Q_m(n) \) of prime partitions of \( n \) into \( m \) parts \((1 \leq m \leq n)\). The bivariate generating function of the numbers \( Q_m(n) \) is of Euler’s type, namely,

\[
1 + \sum_{n=1}^{\infty} z^n \sum_{m=1}^{n} Q_m(n)x^m = \prod_{p_k \in P} (1 - xz^{p_k})^{-1}.
\]

(4)

(the proof may be found in [1], Section 2.1). It is clear that, for \( n > 4 \), \( Q_2(n) \) counts the number of Goldbach partitions of \( n \) and that \( Q_2(n) = 0 \) if \( n \) is odd.

For any real variable \( z \) with \(| z | < 1\), we also set

\[
f(z) = \sum_{p_k \in P} z^{p_k}.
\]

(5)

Differentiating the left- and right-hand sides of (4) twice with respect to \( z \) and setting then \( z = 0 \) and \( m = 2 \), we obtain the following identity.

**Lemma 1.** (See [6], Lemma 3.) For \(| z | < 1\), we have

\[
2 \sum_{k>2} Q_2(2k)z^{2k} = f^2(z) + f(z^2),
\]

(6)

where \( f(z) \) is the function defined by (5).

Further, we will use a Tauberian theorem by Hardy-Littlewood-Karamata [3], Chapter 7. We use it in the form given by Odlyzko [7], Section 8.2.

**Hardy-Littlewood-Karamata Theorem.** (See [7; Theorem 8.7, p. 1225].) Suppose that \( a_k \geq 0 \) for all \( k \), and that

\[
g(x) = \sum_{k=0}^{\infty} a_kx^k
\]

converges for \( 0 \leq x < r \). If there is a \( \rho > 0 \) and a function \( L(t) \) that varies slowly at infinity such that

\[
g(x) \sim (r - x)^{-\rho}L \left( \frac{1}{r - x} \right), \quad x \to r^-,
\]

(7)

then

\[
\sum_{k=0}^{n} a_k r^k \sim \left( \frac{n}{r} \right)^{\rho} \frac{L(n)}{\Gamma(\rho + 1)}, \quad n \to \infty.
\]

(8)

**Remark.** A function \( L(t) \) varies slowly at infinity if, for every \( u > 0 \), \( L(ut) \sim L(t) \) as \( t \to \infty \).
3. Proof of the Main Result

Proof of Theorem 1(i). First, we notice that the definition of the random variable $G_n$ implies that

$$
\frac{1}{n} \mathbb{E}(G_n) = \sum_{2 < k \leq n} \left( \frac{k}{n} \right) \mathbb{P}(G_n = k) = \sum_{2 < k \leq n} \left( \frac{2k}{2n} \right) \sum_{2 < j \leq n} Q_2(2j) Q_2(2k).
$$

The asymptotic behavior of the denominator in the right-hand side of (9) is completely determined by (1) and (2). Differentiating both sides of (6), we will show next that the series

$$
\sum_{k > 2} (2k) Q_2(2k) z^{2k-1} = f(z)f'(z) + z f'(z^2)
$$

and

$$
\sum_{k > 2} (2k)(2k - 1) Q_2(2k) z^{2k-2} = f'^2(z) + f(z) f''(z) + f(z^2) + 2z^2 f'(z^2)
$$

satisfy the conditions of Hardy-Littlewood-Karamata theorem. We need the following lemma.

Lemma 2. Let $f(z)$ be the series defined by (5). Then, as $z \to 1^-$,

$$
f(z) \sim \frac{1}{(1-z) \log \frac{1}{1-z}},
$$

$$
f'(z) \sim \frac{2}{(1-z)^2 \log \frac{1}{1-z}},
$$

$$
f''(z) \sim \frac{2}{(1-z)^3 \log \frac{1}{1-z}}.
$$

Proof. The proof of (12) is given in [6], Lemma 4. Here we present a complete proof of (13). The proof of (14) is similar.

As usual, by $\pi(y)$ we denote the number of primes which do not exceed the positive number $y$. In (5) we set $z = e^{-t}, t > 0$, and apply an argument similar to that given by Stong [10] (see also [2]). We have

$$
f'(z) \big|_{z=e^{-t}} = \sum_{p_k \in \mathcal{P}} p_k z^{p_k-1} \big|_{z=e^{-t}} = \int_0^\infty y e^{-yt} d\pi(y) = I(t) - f(e^{-t}),
$$

where

$$
I(t) = t \int_0^\infty y e^{yt} \pi(y) dy.
$$
In [6] we established that
\[ f(e^{-t}) \sim \frac{1}{t \log \frac{1}{t}}, \quad t \to 0^+. \quad (17) \]

To find the asymptotic of \( I(t) \), we set in (16) \( y = s/t \) and obtain
\[ I(t) = \frac{1}{t} \int_0^{t^{1/2}} s e^{-s} \pi(s/t) ds = \frac{1}{t} (I_1(t) + I_2(t)), \quad (18) \]

where
\[ I_1(t) = \int_0^1 s e^{-s} \pi(s/t) ds, \]
\[ I_2(t) = \int_{t^{1/2}}^{\infty} s e^{-s} \pi(s/t) ds. \]

For \( I_1(t) \) we use the bound \( \pi(s/t) \leq s/t \). Hence, for enough small \( t > 0 \), we have
\[
0 \leq I_1(t) \leq \frac{1}{t} \int_0^{t^{1/2}} s^2 e^{-s} ds = \frac{1}{t} \left( -s^2 e^{-s} \Big|_0^{t^{1/2}} + 2 \int_0^{t^{1/2}} s e^{-s} ds \right)
= -e^{-t^{1/2}} + O(t^{-1/2}) = O(t^{-1/2}). \quad (19)
\]

The estimate of \( I_2(t) \) follows from the Prime Number Theorem with an error term given by Ingham [5], Theorem 23, p.65. Thus, for \( y > 1 \), we have
\[
\pi(y) = \frac{y}{\log y} + O \left( \frac{y}{\log^2 y} \right).
\]

Hence, for \( s \geq t^{1/2} \), we have \( \log s \geq -\frac{1}{2} \log \frac{1}{t} \) and therefore,
\[
\pi(s/t) = \left( \frac{s}{t} \right) \frac{1}{\log \frac{1}{t} + \log s} + O \left( \frac{s}{t \left( \log \frac{1}{t} + \log s \right)^2} \right)
= \frac{s}{t \log \frac{1}{t}} \left( 1 + O \left( \frac{|\log s|}{\log \frac{1}{t}} \right) \right) + O \left( \frac{s}{t \log \frac{1}{t}} \right)
= \frac{s}{t \log \frac{1}{t}} + O \left( \frac{s(1 + |\log s|)}{t \log^2 \frac{1}{t}} \right).
\]
Consequently,

\[ I_2(t) = \frac{1}{t \log \frac{1}{t}} \int_0^{\infty} s^2 e^{-s} ds + O \left( \frac{1}{t \log \frac{1}{t}} \int_{t^{1/2}}^{\infty} s^2 \left(1 + \log s \right) e^{-s} ds \right) \]

\[ = \frac{1}{t \log \frac{1}{t}} \left( \int_0^{\infty} s^2 e^{-s} ds - \int_0^{t^{1/2}} s^2 e^{-s} ds \right) + O \left( \frac{1}{t \log \frac{1}{t}} \right) \]

\[ = \frac{1}{t \log \frac{1}{t}} \left( 2 - O(t^{1/2}) \right) + O \left( \frac{1}{t \log \frac{1}{t}} \right) \]

\[ = \frac{2}{t \log \frac{1}{t}} + O \left( \frac{1}{t^{1/2} \log \frac{1}{t}} \right) + O \left( \frac{1}{t \log \frac{1}{t}} \right) \sim \frac{2}{t \log \frac{1}{t}}, \quad t \to 0^+. \tag{20} \]

Combining (18)-(20), we get

\[ I(t) \sim \frac{2}{t^2 \log \frac{1}{t}}, \quad t \to 0^+. \]

Replacing this asymptotic equivalence and (17) into (15), we conclude that

\[ f'(z) \big|_{z=e^{-t}} = I(t) + O \left( \frac{1}{t \log \frac{1}{t}} \right) \sim \frac{2}{t^2 \log \frac{1}{t}}, \quad t \to 0^+. \]

Returning to the variable \( z \) by \( t = \log \frac{1}{z} \) (\( z < 1 \)), we obtain

\[ f'(z) \sim -\frac{2}{(\log \frac{2}{z}) \log \log \frac{1}{z}}, \quad z \to 1^- . \]

Now, (13) follows from the obvious observation that

\[ \log \frac{1}{z} = -\log z = -\log (1 - (1 - z)) \sim 1 - z, \quad z \to 1^- . \]

The asymptotic equivalence (14) can be obtained in the same way using the representation

\[ f''(z) \big|_{z=e^{-t}} = \int_0^{\infty} y(y - 1)e^{-yt} d\pi(y) . \]

One can show that this last integral is \( \sim \frac{2}{t^2 \log \frac{1}{t}} \) as \( t \to 0^+ \) and then substitute again \( t \) by \( -\log z \). We omit further details. \( \square \)

To complete the proof of Theorem 1(i) we recall (10) and observe that (13) implies that

\[ f'(z^2) \sim \frac{2}{(1 - z^2) \log \frac{1}{1-z^2}} = O \left( \frac{1}{(1 - z)^2 \log \frac{1}{1-z}} \right), \quad z \to 1^- . \]
So, the main contribution to the asymptotic of the right-hand side of (11) is given by the product \( f(z)f'(z) \). Hence, from (12) and (13) we find that

\[
\sum_{k>2}(2k)Q_2(2k)z^{2k-1} = \frac{2}{(1-z)^3 \log^2 \frac{1}{1-z}} + O\left(\frac{1}{(1-z)^2 \log \frac{1}{1-z}}\right)
\]

\[
\sim \frac{2}{(1-z)^3 \log^2 \frac{1}{1-z}}, \quad z \to 1^-,
\]

which implies that the series \( \sum_{k>2}(2k)Q_2(2k)z^{2k-1} \) satisfies condition (17) of Hardy-Littlewood-Karamata Tauberian theorem with \( r=1, \rho=3 \) and \( L(t) = \frac{2}{\log t} \). By (8) we obtain

\[
\sum_{2<k\leq n}(2k)Q_2(2k) \sim \left(\frac{8}{3}\right)\frac{n^3}{\log^2 n}, \quad n \to \infty.
\]  

(22)

Furthermore, (11) and (4) imply that

\[
(2n)\sum_{2<k\leq n}Q_2(2k) \sim \frac{4n^3}{\log^2 n}, \quad n \to \infty.
\]

(23)

Dividing (22) by (23) we see that the limit of the right-hand side of (11) is \( 2/3 \) as \( n \to \infty \).

Remark. Theorem 1(i) also presents a solution of the following curious problem. Consider a sampling procedure that consists of two steps. In the first step we select, as previously, a Goldbach partition from the set \( \Sigma_{2n} \), and, in the second step, we select an even number \( 2R_n \) from the interval \( (4, 2n] \). It is then easy to see that the middle part of (11) represents the probability \( Pr(R_n \leq G_n) \) by the total probability formula. Therefore, Theorem 1(i) implies that

\[
\lim_{n \to \infty} Pr(R_n \leq G_n) = \frac{2}{3}.
\]

Proof of Theorem 1(ii). In the same way, we establish that the leading term in the asymptotic of the right-hand side of (11) is given by the first two terms \( f'(z) + f(z)f''(z) \). By (12), (14), we have

\[
f'(z) + f(z)f''(z) \sim \frac{6}{(1-z)^4 \log^2 \frac{1}{1-z}}, \quad z \to 1^-.
\]

(24)

Furthermore, by (21),

\[
\sum_{k>2}(2k)(2k-1)Q_2(2k)z^{2k-2} = \sum_{k>2}(2k)^2Q_2(2k)z^{2k-2} - \sum_{k>2}(2k)Q_2(2k)z^{2k-2} = \sum_{k>2}(2k)^2Q_2(2k)z^{2k-2} + O\left(\frac{1}{(1-z)^3 \log^2 \frac{1}{1-z}}\right).
\]

(25)
Combining (11), (24) and (21), we obtain
\[ \sum_{k > 2} (2k)^2 Q_2(2k) z^{2k-2} \sim \frac{6}{(1-z)^4 \log^2 \frac{1}{1-z}}, \quad z \to 1^- . \]

Applying again Hardy-Littlewood-Karamata Tauberian theorem with \( r = 1, \rho = 4, L(t) = \frac{6}{\log^2 t}, \) we observe that
\[ \sum_{2 < k \leq n} (2k)^2 Q_2(2k) \sim \frac{4n^4}{\log^2 n}, \quad n \to \infty. \]  
(26)

Similarly to (9), for the second moment of \( G_n \) we have
\[ \frac{1}{n^2} \mathbb{E}(G_n^2) = \sum_{2 < k \leq n} \left( \frac{k}{n} \right)^2 \mathbb{P}(G_n = k) = \sum_{2 < k \leq n} \left( \frac{2k}{2n} \right)^2 \frac{Q_2(2k)}{\sum_{2 < k \leq n} Q_2(2k)} . \]  
(27)

Now, from (26), (27), (1) and (2) it follows that
\[ \lim_{n \to \infty} \frac{1}{n^2} \mathbb{E}(G_n^2) = \frac{1}{2} , \]
which, in combination with the result of Theorem 1(i), completes the proof of part (ii).

References

[1] G.E. Andrews. The Theory of Partitions. Encyclopedia Math. Appl., no. 2, Addison-Wesley, Reading, MA, 1976.

[2] P. Erdős and P. Turán. On some problems of a statistical group theory, IV, Acta Math. Acad. Sci. Hungar. 19 (1968), 413-435.

[3] G.H. Hardy. Divergent Series. Oxford Univ. Press, Oxford, 1949.

[4] G.H. Hardy and E.M. Wright. An Introduction to the Theory of Numbers. Oxford Univ. Press, Oxford, 2008.

[5] A.E. Ingham. The Distribution of the Prime Numbers. Cambridge Univ. Press, Cambridge. 1990.

[6] L. Mutafchiev. A note on Goldbach partitions of large even integers. Electron. J. Combin. 22 (2015), # P.1.40.

[7] A.M. Odlyzko. Asymptotic Enumeration Methods. In Handbook of Combinatorics (R. Graham, M. Grötschel and L. Lovász, Eds.), Vol. II, Elsevier, 1995, pp.1063-1229.
[8] M. Overholt. *A Course in Analytic Number Theory*. Graduate Studies in Mathematics, Vol. 160, American Mathematical Society, Providence, RI, 2014.

[9] K. Prachar. *Primzahlverteilung*. Springer, Berlin, 1957.

[10] R. Stong. *The average order of a permutation*. Electron. J. Combin. **5** (1998), R41.