Riemann Hypothesis and Random Walks: the Zeta case

André LeClair[a]

Cornell University, Physics Department, Ithaca, NY 14850

Abstract

In previous work it was shown that if certain series based on sums over primes of non-principal Dirichlet characters have a conjectured random walk behavior, then the Euler product formula for its L-function is valid to the right of the critical line $\Re(s) > 1/2$, and the Riemann Hypothesis for this class of L-functions follows. Building on this work, here we propose how to extend this line of reasoning to the Riemann zeta function and other principal Dirichlet L-functions. We use our results to argue that $S_\delta(t) \equiv \lim_{\delta \to 0^+} \dfrac{1}{\pi} \arg \zeta(\frac{1}{2} + \delta + it) = O(1)$, and that it is nearly always on the principal branch. We conjecture that a 1-point correlation function of the Riemann zeros has a normal distribution. This leads to the construction of a probabilistic model for the zeros. Based on these results we describe a new algorithm for computing very high Riemann zeros as a kind of stochastic process, and we calculate the 10^{100}-th zero to over 100 digits.

[a] andre.leclair@gmail.com

I. INTRODUCTION

There are many generalizations of Riemann's zeta function to other Dirichlet series, which are also believed to satisfy a Riemann Hypothesis. A common opinion, based largely on counterexamples, is that the L-functions for which the Riemann Hypothesis is true enjoy both an Euler product formula and a functional equation. However a direct connection between these properties and the Riemann Hypothesis has not been formulated in a precise manner. In [1, 2] a concrete proposal making such a connection was presented for Dirichlet L-functions, and those based on cusp forms, due to the validity of the Euler product formula to the right of the critical line. In contrast to the non-principal case, in this approach the case of principal Dirichlet L-functions, of which Riemann zeta is the simplest, turned out to be more delicate, and consequently it was more difficult to state precise results. In the present work we address further this special case.

Let $\chi(n)$ be a Dirichlet character modulo k and $L(s,\chi)$ its L-function with $s=\sigma+it$. It satisfies the Euler product formula

$$L(s,\chi)=\sum_{n=1}^{\infty}\frac{\chi(n)}{n^s}=\prod_{n=1}^{\infty}\left(1-\frac{\chi(p_n)}{p_n^s}\right)^{-1} \tag{1}$$

where p_n is the n-th prime. The above formula is valid for $\Re(s)>1$ since both sides converge absolutely. The important distinction between principal verses non-principal characters is the following. For non-principal characters the L-function has no pole at $s=1$, thus there exists the possibility that the Euler product is valid partway inside the strip, i.e. has abscissa of convergence $\sigma_c<1$. It was proposed in [1, 2] that $\sigma_c=1/2$ for this case. In contrast, now consider L-functions based on principal characters. The latter character is defined as $\chi(n)=1$ if n is coprime to k and zero otherwise. The Riemann zeta function is the trivial principal character of modulus $k=1$ with all $\chi(n)=1$. L-functions based on principal characters do have a pole at $s=1$, and therefore have abscissa of convergence $\sigma_c=1$, which implies the Euler product in the form given above cannot be valid inside the critical strip $0<\sigma<1$. Nevertheless, in this paper we will show how a truncated version of the Euler product formula is valid for $\sigma>1/2$.

The primary aim of the work [1, 2] was to determine what specific properties of the prime numbers would imply that the Riemann Hypothesis is true. This is the opposite of the more well-studied question of what the validity of the Riemann Hypothesis implies for the fluctuations in the distribution of primes. The answer proposed was simply based on the multiplicative independence of the primes, which to a large extent underlies their pseudo-random behavior. To be more specific, let $\chi(n) = e^{i\theta_n}$. In [1, 2] it was proven that if the series

$$B_N(t,\chi) = \sum_{n=1}^{N} \cos\left(t \log p_n + \theta_{p_n}\right) \tag{2}$$

is $O(\sqrt{N})$, then the Euler product converges for $\sigma > 1/2$ and the formula (1) is valid to the right of the critical line. For non-principal characters the allowed angles θ_n are equally spaced on the unit circle, and it was conjectured in [2] that the above series with $t = 0$ behaves like a random walk due to the multiplicative independence of the primes, and this is the origin of the $O(\sqrt{N})$ growth. Furthermore, this result extends to all t since domains of convergence of Dirichlet series are always half-planes. Taking the logarithm of (1), one sees that $\log L$ is never infinite to the right of the critical line and thus has no zeros there. This, combined with the functional equation that relates $L(s)$ to $L(1-s)$, implies there are also no zeros to the left of the critical line, so that all zeros are on the line. The same reasoning applies to cusp forms if one also uses a non-trivial result of Deligne [2].

In this article we reconsider the principal Dirichlet case, specializing to Riemann zeta itself since identical arguments apply to all other principal cases with $k > 1$. Here all angles $\theta_n = 0$, so one needs to consider the series

$$B_N(t) = \sum_{n=1}^{N} \cos(t \log p_n) \tag{3}$$

which now strongly depends on t. On the one hand, whereas the case of principal Dirichlet L-functions is complicated by the existence of the pole, and, as we will see, one consequently needs to truncate the Euler product to make sense of it, on the other hand B_N can be estimated using the prime number theorem since it does not involve sums over non-trivial characters χ, and this aids the analysis. This is in contrast to the non-principal case, where, however well-motivated, we had to conjecture the random walk behavior alluded to above, so in this respect the principal case is potentially simpler. To this end, a theorem of Kac (Theorem 1 below) nearly does the job: $B_N(t) = O(\sqrt{N})$ in the limit t and $N \to \infty$, which is also a consequence of the multiplicative independence of the primes. This suggests that one can also make sense of the Euler product formula in the limit $t \to \infty$. However this is not enough for our main purpose, which is to have a similar result for finite t which we will develop.

This article is mainly based on our previous work [1, 2] but provides a more detailed analysis and extends it in several ways. It was suggested in [1] that one should truncate the series at an N that depends on t. In the next section we explain how a simple group structure underlies a finite Euler product which relates it to a generalized Dirichlet series which is a subseries of the Riemann zeta function. Subsequently we estimate the error under truncation, which shows explicitly how this error is related to the pole at $s = 1$, as expected. The remainder of the paper presents various applications of these ideas. We use these results to study the argument of the zeta function and calculate very high zeros. We also conjecture that the statistical fluctuations of individual zeros have a normal distribution.

In some aspects, our work is related to the work of Gonek et. al. [4, 5], which also considers a truncated Euler product. The important difference is that the starting point in [4] is a hybrid version of the Euler product which involves both primes and zeros of zeta. Only after assuming the Riemann Hypothesis can one explain in that approach why the truncated product over primes is a good approximation to zeta. In contrast, here we do not assume anything about the zeros of zeta.

II. ALGEBRAIC STRUCTURE OF FINITE EULER PRODUCTS

The aim of this section is to define properly the objects we will be dealing with. In particular we will place finite Euler products on the same footing as other generalized Dirichlet series. The results are straightforward and are mainly definitions.

Definition 1. Fix a positive integer N and let $\{p_1, p_2, \ldots p_N\}$ denote the first N primes where $p_1 = 2$. From this set one can generate an abelian group $\mathbb{Q}_N$ of rank N with elements

$$\mathbb{Q}_N = \left\{ p_1^{n_1} p_2^{n_2} \cdots p_N^{n_N}, \quad n_i \in \mathbb{Z} \ \forall_i \right\} \tag{4}$$

where the group operation is ordinary multiplication. Clearly $\mathbb{Q}_N \subset \mathbb{Q}^+$ where $\mathbb{Q}^+$ are the positive rational numbers. There are an infinite number of integers in $\mathbb{Q}_N$ which form a subset of the natural numbers $\mathbb{N} = \{1, 2, \ldots\}$. We will denote this set as $\mathbb{N}_N \subset \mathbb{N}$, and elements of this set simply as $\mathfrak{n}$.

Definition 2. Fix a positive integer N. For every integer $n \in \mathbb{N}$ we can define the character $c(n)$:

$$\begin{aligned} c(n) &= 1 \qquad \text{if } n \in \mathbb{N}_N \subset \mathbb{Q}_N \\ &= 0 \qquad \text{otherwise} \end{aligned} \tag{5}$$

Clearly, for a prime p, $c(p) = 0$ if $p > p_N$.

Definition 3. Fix a positive integer N and let s be a complex number. Based on $\mathbb{Q}_N$ we can define the infinite series

$$\zeta_N(s) = \sum_{n=1}^{\infty} \frac{c(n)}{n^s} = \sum_{\mathfrak{n} \in \mathbb{N}_N} \frac{1}{\mathfrak{n}^s} \tag{6}$$

which is a generalized Dirichlet series. There are an infinite number of terms in the above series since $\mathbb{N}_N$ is infinite dimensional.

Example 1. For instance

$$\zeta_2(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{6^s} + \frac{1}{8^s} + \frac{1}{9^s} + \frac{1}{12^s} + \dots$$

Because of the group structure of $\mathbb{Q}_N$, ζ_N satisfies a finite Euler product formula:

Proposition 1. *Let σ_c be the abscissa of convergence of the series $\zeta_N(s)$ where $s = \sigma + it$, namely $\zeta_N(s)$ converges for $\Re(s) > \sigma_c$. Then in this region of convergence, ζ_N satisfies a finite Euler product formula:*

$$\zeta_N(s) = \prod_{n=1}^{N} \left(1 - \frac{1}{p_n^s}\right)^{-1} \tag{7}$$

Proof. Based on the completely multiplicative property of the characters,

$$c(nm) = c(n)c(m) \tag{8}$$

one has

$$\zeta_N(s) = \prod_{n=1}^{\infty} \left(1 - \frac{c(p_n)}{p_n^s}\right)^{-1}$$

The result follows then from the fact that $c(p_n) = 0$ if $n > N$. □

Example 2. Let $N = 1$, so that $\{\mathfrak{n}\} = \{1, 2, 2^2, 2^3 \ldots\}$. Then the above Euler product formula (7) is simply the standard formula for the sum of a geometric series:

$$\zeta_1(s) = \sum_{n=0}^{\infty} \frac{1}{2^{ns}} = \frac{1}{1 - 2^{-s}} \tag{9}$$

Here the abscissa of convergence is $\sigma_c = 0$.

The series $\zeta_N(s)$ defined in (6) has some interesting properties:

(i) For finite N the product is finite for $s \neq 0$, thus the infinite series $\zeta_N(s)$ converges for $\Re(s) > 0$ for any finite N.

(ii) Since the logarithm of the product is finite, for finite N, $\zeta_N(s)$ has no zeros nor poles for $\Re(s) > 0$. Thus the Riemann zeros and the pole at $s = 1$ arise from the primes at infinity p_∞, i.e. in the limit $N \to \infty$. In this limit all integers are included in the sum (6) that defines ζ_N since $\mathbb{N}_\infty = \mathbb{N}$. This is in accordance to the fact that the pole is a consequence of there being an infinite number of primes.

The property (ii) implies that, in some sense, the Riemann zeros condense out of the primes at infinity p_∞. Formally one has

$$\lim_{N\to\infty} \zeta_N(s) = \zeta(s) \tag{10}$$

However since N is going to infinity, the above is true only where the series formally converges, which, as discussed in the Introduction, is $\Re(s) > 1$. Nevertheless, for very large but finite N, the function ζ_N can still be a good approximation to $\zeta(s)$ *inside the critical strip* since for N finite there is convergence of $\zeta_N(s)$ for $\Re(s) > 0$. This is the subject of the next section, where we show that a finite Euler product formula is valid for $\Re(s) > 1/2$ in a manner that we will specify.

III. FINITE EULER PRODUCT FORMULA AT LARGE *N* TO THE RIGHT OF THE CRITICAL LIINE.

In this section we propose that the Euler product formula can be a very good approximation to $\zeta(s)$ for $\Re(s) > 1/2$ and large t if N is chosen to depend on t in a specific way which was already proposed in [1, 2]. The new result presented here is an estimate of the error due to the truncation.

The random walk property we will build upon is based on a central limit theorem of Kac, which largely follows from the multiplicative independence of the primes:

Theorem 1. *(**Kac**) Let u be a random variable uniformly distributed on the interval $u \in [T, 2T]$, and define the series*

$$B_N(u) = \sum_{n=1}^{N} \cos(u \log p_n) \tag{11}$$

Then in the limit $N \to \infty$ and $T \to \infty$, $B_N/\sqrt{N}$ approaches the normal distribution $\mathcal{N}(0,1)$, namely

$$\lim_{N\to\infty} \lim_{T\to\infty} P\left\{ \frac{x_1}{\sqrt{2}} < \frac{B_N(u)}{\sqrt{N}} < \frac{x_2}{\sqrt{2}} \right\} = \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-x^2/2} dx \tag{12}$$

where P denotes the probability for the set.

We wish to use the above theorem to conclude something about $B_N(t)$ for a fixed, non-random t. Based on Theorem 1, we first conclude the following for non-random, but large t:

Corollary 1.

$$\lim_{N\to\infty} \lim_{t\to\infty} B_N(t) = O(\sqrt{N}) \tag{13}$$

Proof. This is straightforward: as $T \to \infty$, even though u is random, all u in the range $[T, 2T]$ are tending to ∞. □

A consequence of the above Theorem 1 and the Corollary 13 is that the Euler product formula is valid to the right of the critical line in the limit of large t:

Theorem 2. *For $\sigma > 1/2$,*

$$\lim_{t\to\infty} \zeta(\sigma + it) = \lim_{N\to\infty} \lim_{t\to\infty} \prod_{n=1}^{N} \left(1 - \frac{1}{p_n^{\sigma+it}}\right)^{-1} \tag{14}$$

Proof. The proof is essentially the same as in [1, 2], so we just sketch the main steps involved. Taking the logarithm of the above equation, one concludes that the Euler product converges with $\sigma > 1/2$ if the series $X_N(s) = \sum_{n=1}^{N} 1/p_n^s$ converges as $N \to \infty$. It is enough to consider $\mathcal{S}_N = \Re(X_N)$:

$$\mathcal{S}_N(s) = \sum_{n=1}^{N} a_n b_n, \qquad a_n = \frac{1}{p_n^\sigma}, \qquad b_n = \cos(t \log p_n) \tag{15}$$

The latter can be reorganized using integration by parts:

$$\mathcal{S}_N = a_N B_N + \sum_{n=1}^{N-1} B_n (a_n - a_{n+1}), \quad B_n \equiv \sum_{k=1}^{n} \cos(t \log p_n) \tag{16}$$

The sum above is bounded

$$|\mathcal{S}_N| \leq \sigma \sum_{n=1}^{N-1} |B_n| \frac{g_n}{p_n^{\sigma+1}} + O(1) \tag{17}$$

where $g_n = p_{n+1} - p_n$ is the gap between primes. One then performs another summation by parts using a summed version of the Cramér-Granville conjecture

$$\sum_{n=1}^{N} g_n < \sum_{n=1}^{N} \log^2 p_n \tag{18}$$

Now if $\lim_{t\to\infty} B_N(t) = O(\sqrt{N})$ for large N, as far as convergence is concerned, the sum in (17) behaves as $\sum_n \log^2 n / n^{\sigma+1/2}$ which converges for $\sigma > 1/2$. □

It is desirable to have a version of Theorem 2 where N and t are taken to infinity simultaneously. Namely, we wish to truncate the product at an $N(t)$ that depends on t with the property that $\lim_{t\to\infty} N(t) = \infty$. One can then replace the double limit on the RHS of (14) with one limit $t \to \infty$, or equivalently $N(t) \to \infty$. There is no unique choice for $N(t)$, but there is an optimal upper limit, $N(t) < N_{\max}(t)$, which we now describe. We need the following [1, 2]:

Proposition 2.

$$B_N(t) = O(\sqrt{N}), \qquad \text{for } N < N_{\max}(t) \equiv [t^2] \tag{19}$$

where $[t^2]$ *denotes its integer part.*

Proof. Using the prime number theorem,

$$\begin{aligned} B_N(t) &\approx \int_2^{p_N} \frac{dx}{\log x} \cos(t \log x) = \Re\left(\mathrm{Ei}\left((1+it)\log p_N\right)\right) \\ &\approx \frac{p_N}{\log p_N} \left(\frac{t}{1+t^2}\right) \sin\left(t \log p_N\right) \end{aligned} \tag{20}$$

where Ei is the usual exponential-integral function, and we have used

$$\mathrm{Ei}(z) = \frac{e^z}{z}\left(1 + O\left(\frac{1}{z}\right)\right) \tag{21}$$

The prime number theorem implies $p_N \approx N \log N$. Using this in (20) and imposing $B_N(t) < \sqrt{N}$ proves the proposition. □

Based on the above proposition, henceforth we will always assume the following properties of $N(t)$:

$$N(t) \sim N_{\max}(t) \equiv [t^2] \quad \text{with} \quad \lim_{t\to\infty} N(t) = \infty \tag{22}$$

Repeating the arguments in the proof of Theorem 2, we now have the following:

Proposition 3. *Let $N(t)$ satisfy* (22)*. Then*

$$\lim_{t\to\infty} \zeta(s) = \lim_{t\to\infty} \prod_{n=1}^{N(t)} \left(1 - \frac{1}{p_n^s}\right)^{-1}, \qquad \text{for } \Re(s) > 1/2 \tag{23}$$

Extensive and compelling numerical evidence supporting the above formula was already presented in [1].

Based on the above results we are now in a position to study the following important question. If we fix a finite but large t, and truncate the Euler product at $N(t)$, which is finite, what is the error in the approximation to ζ to the right of the critical line? We estimate this error as follows:

Theorem 3. *Let $N(t)$ satisfy* (22)*. Then for $\Re(s) > 1/2$ and large t,*

$$\zeta(s) = \prod_{n=1}^{N(t)} \left(1 - \frac{1}{p_n^s}\right)^{-1} \exp\left(R_{N(t)}(s)\right) \tag{24}$$

where

$$R_N(s) = \frac{1}{(s-1)}\, O\left(\frac{N^{1-s}}{\log^s N}\right) \tag{25}$$

is finite (except at the pole $s = 1$) and satisfies

$$\lim_{t\to\infty} R_{N(t)}(s) = 0 \tag{26}$$

Proof. From (23), one concludes that (24) must hold in the limit of large t with R_N satisfying (26). The logarithm of (24) reads

$$\log \zeta(s) = -\sum_{n=1}^{N} \log\left(1 - \frac{1}{p_n^s}\right) + R_N(s) \tag{27}$$

In the limit of large t, the error upon truncation is the part that is neglected in (23):

$$R_N(s) = -\sum_{n=N+1}^{\infty} \log\left(1 - \frac{1}{p_n^s}\right) \tag{28}$$

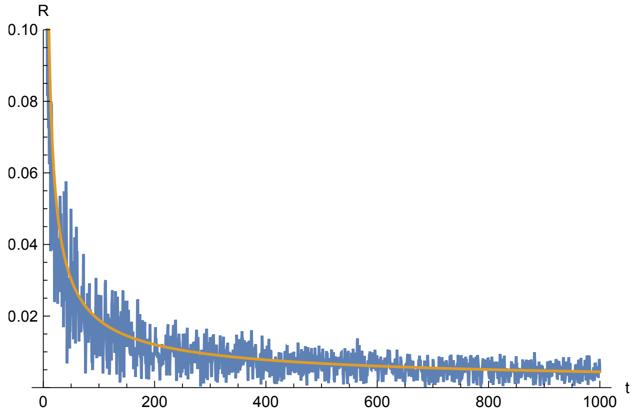

FIG. 1. The error term $|R_N(s)|$ with $N(t) = N_{\max}(t) = [t^2]$ for $\Re(s) = 3/4$ inside the critical strip as a function of t. The fluctuating (blue) curve is $|R_N|$ computed directly from the definition (27) with $\zeta(s)$ the usual analytic continuation into the strip. The smooth (yellow) curve is the approximation $R_N(s) = \frac{1}{(s-1)} \frac{N^{1-s}}{\log^s N}$ based on (25).

Expanding out the logarithm, one has

$$
\begin{aligned}
R_N(s) &\approx \sum_{n=N}^{\infty} \frac{1}{p_n^s} \\
&\approx \int_{p_N}^{\infty} \frac{dx}{\log x} \frac{1}{x^s} \approx \frac{1}{(s-1)} \frac{p_N^{1-s}}{\log p_N}
\end{aligned}
\tag{29}
$$

Next using $p_N \approx N \log N$, one obtains (25). In the above integral, the reason the upper limit of integration $x = \infty$ gives zero is that the lower limit behaves as $N(t)^{1-s}/t < N^{1/2-s}$ which goes to zero as $N \to \infty$ if $\Re(s) > 1/2$. The latter also implies (26). □

Theorem 3 makes it clear that the need for a cut-off $N < N_{\max}$ originates from the pole at $s = 1$, since as long as $s \neq 1$, the error $R_N(s)$ in (25) is finite. The error becomes smaller and smaller the further one is from the pole, i.e. as $t \to \infty$. In Figure 1 we numerically illustrate Theorem 3 inside the critical strip.

Remark 1. For estimating errors at large t the following formula is useful:

$$
|R_{N(t)}(s)| \sim \frac{N(t)^{1-\sigma}}{t} \sim \frac{1}{t^{2\sigma-1}} \tag{30}
$$

Theorem 4. *Assuming Theorem 3, all non-trivial zeros of $\zeta(s)$ are on the critical line.*

Proof. Taking the logarithm of the truncated Euler product, one obtains (27). If there were a zero ρ with $\Re(\rho) > 1/2$, then $\log \zeta(\rho) = -\infty$. However the right hand side of (27) is always finite, thus there are no zeros to the right of the critical line. The functional equation relating $\zeta(s)$ to $\zeta(1-s)$ shows there are also no zeros to the left of the critical line. □

Remark 2. Interestingly Theorems 3 and 4 imply that proving the validity of the Riemann Hypothesis is under better control the higher one moves up the critical line. For instance, it is known that all zeros are on the line up to $t \sim 10^{13}$, and beyond this, the error R_N is too small to spoil the validity of the Riemann Hypothesis. Henceforth, we assume Theorem 4.

IV. IMPLICATIONS FOR THE ARGUMENT OF THE ζ-FUNCTION

The function $S(T)$ is conventionally defined as

$$S(T) = \frac{1}{\pi} \arg \zeta \left(\tfrac{1}{2} + iT \right) \tag{31}$$

for T not the ordinate of a zero. The argument is usually defined by piecewise integration from $s = 2$ to $s = 2 + it$, then to $\frac{1}{2} + it$. The importance of $S(T)$ is its role in the function $\mathcal{N}(T)$ which counts the number of zeros inside the critical strip with ordinate $0 < t < T$. There is an exact formula due to Backlund: $\mathcal{N}(T) = \frac{1}{\pi}\vartheta(T) + 1 + S(T)$, where $\vartheta(T)$ is the Riemann-Siegel ϑ-function. Consequently, there is a large literature concerning $S(T)$; see for instance [7]. $S(t)$ also plays an important role in the transcendental equations satisfied by individual zeros obtained in [6]. The main properties of $S(T)$ that are well-known are: (i) At each zero of ζ it jumps by the multiplicity of the zero; (ii) Between zeros $dS(t)/dt < 0$ since $d\mathcal{N}(t)/dt = 0$ there and $\vartheta(t)$ is monotonically increasing; (iii) The average of $S(t)$ is zero.

The current best bound on $S(t)$ was proven by Goldston and Gonek [8] $|S(t)| \leq \left(\frac{1}{2} + o(1)\right) \log t / \log\log t$ in the limit $t \to \infty$. It should be kept in mind that this is an upper bound so that $|S(t)|$ may actually be much smaller, as numerical evidence would suggest. Below, based on the results of the last section, we will propose that $S(t)$ is actually $O(1)$, and in fact is nearly always on the principal branch.

In the work [6], $S_\delta(t)$ was defined as follows:

$$S_\delta(t) = \lim_{\delta \to 0^+} \frac{1}{\pi} \arg \zeta \left(\tfrac{1}{2} + \delta + it \right) \tag{32}$$

The above $S_\delta(t)$ is actually also well-defined for t equal to the ordinate of a zero. It should be kept in mind that this definition is not necessarily equivalent to other definitions in the

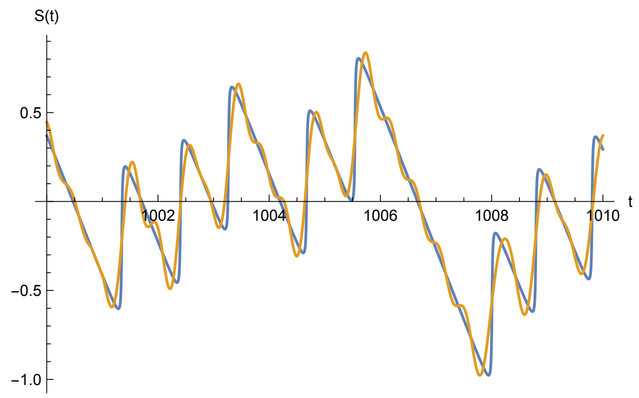

FIG. 2. The exact $S(t)$ (blue line) verses $S(t)$ calculated from the Euler product formula (33) (yellow line). Here we took $\delta = 0.01$ and $N = 10^5 < N_{\max}(t)$.

literature. However, if the Riemann Hypothesis is true, they are expected to be the same away from the ordinate of a zero since no singularities are encountered in the final integration to the critical line. Henceforth, $S(t)$ refers to $S_\delta(t)$ as defined above. Based on Theorem 3, we have

$$S_\delta(t) = \frac{1}{\pi} \Im \log \zeta \left(\tfrac{1}{2} + \delta + it\right) = -\frac{1}{\pi} \sum_{n=1}^{N(t)} \log \left(1 - \frac{1}{p_n^{1/2+\delta+it}}\right) + \frac{1}{\pi} \Im R_N(\tfrac{1}{2} + it) \tag{33}$$

Recall that as $t \to \infty$, R_N actually goes to zero. One can check numerically that the above formula works rather well with R_N disregarded; see Figure 2. From this figure one clearly sees that the above formula knows about all the Riemann zeros, where it jumps by one at each.

It is clear that based on (33), $S(t) = O(1)$ because it is finite. We can state something more precise as follows:

Proposition 4. *Under the assumption of Theorem 3, which implies the Euler product formula* (33) *for* $S(t)$*, then* $S(t)$ *is well-defined for all* t *and* $\lim_{t\to\infty} S(t) = O(1)$.

Proof. Let us fix $N = N(t)$ satisfying (22). Expanding the logarithm, one has

$$S(t) = \lim_{\delta \to 0^+} \frac{1}{\pi} \Im \sum_{n=1}^{N} \frac{1}{p_n^{1/2+\delta+it}} + O(1) \tag{34}$$

We neglected the R_N error since it is also $O(1)$ by (30). As for other functions defined by sums over primes, such as the prime number counting function $\pi(x)$, there is a leading smooth part which is determined by the prime number theorem, and a subleading fluctuating part that depends on the exact locations of the primes. We can therefore write

$$S(t) = S_{\rm pnt}(t) + \delta S(t) \tag{35}$$

where $S_{\rm pnt}(t)$ is the smooth part coming from the prime number theorem, and $\delta S(t)$ are the fluctuating corrections. Consider first the smooth part:

$$\begin{aligned} S_{\rm pnt}(t) &= \frac{1}{\pi}\Im \int_2^{p_N} \frac{dx}{\log x}\frac{e^{-it\log x}}{\sqrt{x}} \\ &= \frac{1}{\pi}\Im\left(\mathrm{Ei}\left[(\tfrac{1}{2}-it)\log p_N\right] - \mathrm{Ei}\left[(\tfrac{1}{2}-it)\log 2\right]\right) \end{aligned} \tag{36}$$

For $y > 0$:

$$\Im(\mathrm{Ei}(-iy)) = -\pi + \frac{\cos y}{y} + O\left(\frac{1}{y^2}\right) \tag{37}$$

Thus $\lim_{y\to\infty}\Im\left(\mathrm{Ei}(-iy)\right) = -\pi$. Now, as $t \to \infty$, in (36) one can replace $\frac{1}{2} - it$ with $-it$, and the two terms cancel:

$$\lim_{t\to\infty} S_{\rm pnt}(t) = 0 \tag{38}$$

Let us now turn to the fluctuating term $\delta S(t)$ which actually knows about the locations of the zeros since at each zero it jumps by its multiplicity. Since the leading contribution $S_{\rm pnt}$ goes to zero, $\delta S(t)$ has no growth and consists only of these jumps, all occurring around $S = 0$. Thus $S(t) = O(1)$. □

If one assumes all zeros of ζ are simple then one can further argue that $S(t)$ is nearly always on the principal branch:

$$-1 \lesssim S(t) \lesssim 1 \tag{39}$$

If all zeros are simple, then $S(t)$ jumps by only 1 at each zero. Thus the largest value of $|\delta S(t)|$ is approximately 1 corresponding to a jump beginning at $t \approx 0$. In other words, $S(t)$ is never very far from zero so that most of the jumps pass through $S = 0$ as seen in Figure 2.

Figure 2 provides numerical evidence for Proposition 4. Simply stated, the above Proposition shows that there is no change in behavior of $S(t)$ as t increases to infinity, such that the pattern in Figure 2 persists. We checked its validity all the way up to $t = 10^{12}$. Only rarely is $|S(t)|$ slightly above 1. Over this whole range we found $|S(t)| < 1.2$.

V. 1-POINT CORRELATION FUNCTION OF THE RIEMANN ZEROS

Montgomery conjectured that the pair correlation function of ordinates of the Riemann zeros on the critical line satisfy GUE statistics [9]. Being a 2-point correlation function, it is a reasonably complicated statistic. In this section we propose a simpler 1-point correlation function that captures the statistical fluctuations of individual zeros.

Let t_n be the exact ordinate of the n-th zero on the critical line, with $t_1 = 14.1347...$ and so forth. The single equation $\zeta(\rho) = 0$ has an infinite number of solutions $\rho = \frac{1}{2} + it_n$. In [6], by placing the zeros in one-to-one correspondence with the zeros of the cosine function, the single equation $\zeta(\rho) = 0$ was replaced by an infinite number of equations, one for each t_n that depends only on n:

$$\vartheta(t_n) + \lim_{\delta \to 0^+} \arg \zeta(\tfrac{1}{2} + \delta + it_n) = (n - \tfrac{3}{2})\pi \tag{40}$$

where ϑ is the Riemann-Siegel function:

$$\vartheta(t) = \Im \log \Gamma(\tfrac{1}{4} + \tfrac{it}{2}) - t \log \sqrt{\pi} \tag{41}$$

This equation was used to calculate zeros very accurately in [6], up to thousands of digits. There is no need for a cut-off $N_{\max}$ in the above equation since the $\arg \zeta$ term can in principle be calculated for arbitrarily high t without the Euler product formula using standard analytic continuation. One aspect of this equation is the following theorem:

Theorem 5. *(***França-LeClair [6]***) If there is a unique solution to the equation* (40) *for every n, then the Riemann Hypothesis is true, and furthermore, all zeros are simple.*

If the $\arg \zeta$ term is ignored, then there is indeed a unique solution for all n since $\vartheta(t)$ is a monotonically increasing function of t. Using its asymptotic expansion for large t, (48) below, then the solution is approximately

$$\widetilde{t}_n = \frac{2\pi(n - \frac{11}{8})}{W\left((n - \frac{11}{8})/e\right)} \tag{42}$$

where W is the Lambert W-function. The only way there would fail be a solution is if $S(t)$ is not well defined for all t. However this appears to be ruled out by the analysis of the last section, in particular Proposition 4.

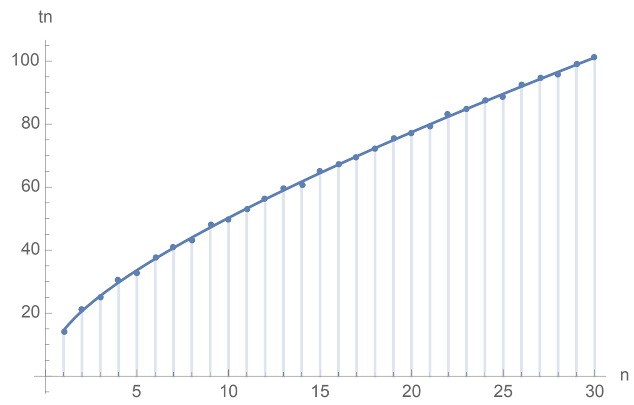

FIG. 3. The first 30 Riemann zeros. The smooth curve is the approximation $\widetilde{t}_n$ in (42), whereas the dots are the actual zeros t_n.

The fluctuations in the zeros obviously come from $S(t)$ since $\widetilde{t}_n$ is a smooth function of n. These small fluctuations are shown in Figure 3. Let us define

$$\delta t_n = t_n - \widetilde{t}_n \tag{43}$$

One needs to properly normalize δt_n, taking into account that the spacing between zeros decreases as $1/\log n$. To this end we expand the equation (40) around $\widetilde{t}_n$. Using $\vartheta(\widetilde{t}_n) \approx (n - \frac{3}{2})\pi$, one obtains $\delta t_n \approx \pi S(t_n)/\vartheta'(\widetilde{t}_n)$ where $\vartheta'(t)$ is the derivative with respect to t. Using $\vartheta'(t) \approx \frac{1}{2}\log(t/2\pi e)$, this leads us to define

$$\delta_n \equiv \frac{\delta t_n}{2\pi} \log\left(\frac{\widetilde{t}_n}{2\pi e}\right) \approx S(t_n) \tag{44}$$

One can then study the probability distribution of the set

$$\Delta_M \equiv \left\{\delta_1, \delta_2, \ldots, \delta_M\right\} \tag{45}$$

for large M. The equation (44) together with (33) makes it clear that the origin of the statistical fluctuations of Δ_M is the fluctuations in the primes.

Let us make the hypothesis that Δ_M satisfies a normal distribution $\mathcal{N}(\mu, \sigma_1)$. Using the properties of $S(t_n)$ described in the last section, together with the equation (44), we can propose the following. First, one expects that the average of δ_n is zero since it is known that the average of $S(t)$ is zero, thus $\mu = 0$. Secondly, if $S(t)$ is nearly always on the principal branch, as argued in the last section, then at each jump by 1 at t_n, on average $S(t_n)$ passes through zero. This implies that the average $\overline{|S(t_n)|} \approx 1/4$. For a normal

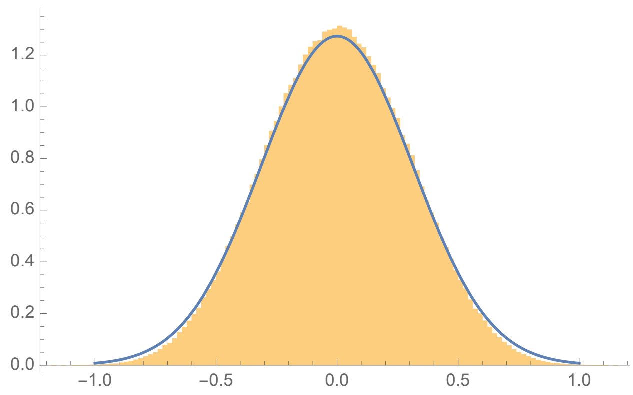

FIG. 4. The probability distribution for the set Δ_M defined in (45) for $M = 10^6$. The smooth curve is the normal distribution $\mathcal{N}(0, \sigma_1)$ with $\sigma_1 = \sqrt{\pi/32}$.

distribution $\overline{|S(t_n)|} = \sqrt{\frac{2}{\pi}}\,\sigma_1$. Thus one expects the standard deviation σ_1 of Δ_M to be $\sigma_1 \approx \sqrt{\pi/32} = 0.313...$ In Figure 4 we present results for the first 10^6-th known zeros. The distribution function fits a normal distribution with $\sigma_1 = \sqrt{\pi/32}$ very well. Performing a fit, one finds $\sigma_1 \approx 0.2966$. This leads us to conjecture:

Conjecture 1. *In the limit of large M the set Δ_M has a normal distribution $\mathcal{N}(0, \sigma_1)$ with $\sigma_1 \approx \sqrt{\pi/32}$.*

Based on Conjecture 1 we can construct a probabilistic model of the Riemann zeros:

Definition 4. **A probabilistic model of the Riemann zeros.** Let $\mathfrak{r}$ be a random variable with normal distribution $\mathcal{N}(0, \sigma_1)$. Then a probabilistic model of the zeros t_n can be defined as the set $\{\widehat{t}_n\}$, where

$$\widehat{t}_n \equiv \widetilde{t}_n + \frac{2\pi\,\mathfrak{r}}{\log(\widetilde{t}_n/2\pi e)} \tag{46}$$

and $\widetilde{t}_n$ is defined in (42).

Such a model could have a variety of applications, similarly to Cramér's probabilistic model of the primes. For instance, one could use it to define a randomized zeta function from a Hadamard product $\widehat{\zeta}(s) = \prod_{\widehat{\rho}_n} (s - \widehat{\rho}_n)$ where $\widehat{\rho}_n = \frac{1}{2} \pm i\widehat{t}_n$. Furthermore, if we assume an Euler product for this randomized ζ, this would define a random model of the primes. We do not pursue this further here, but rather we investigate the following question.

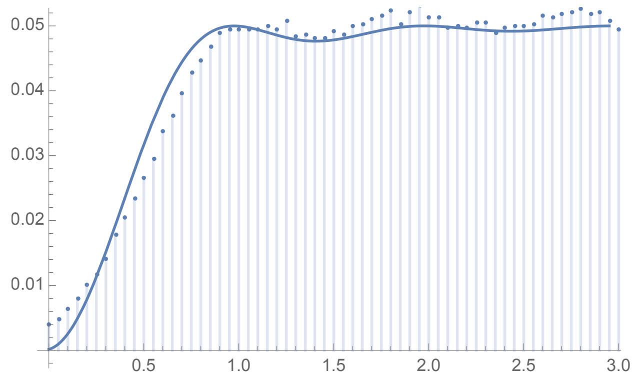

FIG. 5. The pair correction function of $\{\widehat{t}_n\}$ defined in (46) for n up to 10^5 where the standard deviation of $\mathfrak{r}$ was taken to be $\sigma_1 = 0.274$. The solid curve is the GUE prediction. The parameters in (47) are $\beta = \alpha + 0.05$ with $\alpha = (0, 0.05, 0.10, \ldots, 3)$ and the x-axis is given by $x = (\alpha + \beta)/2$.

The statistical model (46) is rather simplistic since it is just based on a normal distribution for $\mathfrak{r}$ and $\widetilde{t}_n$ is smooth and completely deterministic. A natural question then arises. Does the pair correlation function of $\{\widehat{t}_n\}$ satisfy GUE statistics as does the actual zeros $\{t_n\}$? We expect the answer is no, since the only correlation between pairs of $\widehat{t}_n$'s is the smooth, predictable part $\widetilde{t}_n$. Nevertheless, it is interesting to study the 2-point correlation function of $\{\widehat{t}_n\}$. Montgomery's pair correlation conjecture can be stated as follows. Let $\mathcal{N}(T)$ denote the number of zeros up to height T, where $\mathcal{N}(T) \sim \frac{T}{2\pi} \log\left(\frac{T}{2\pi e}\right)$. Let t, t' denote zeros in the range $[0, T]$. Then in the limit of large T:

$$\frac{1}{\mathcal{N}(T)} \sum_{\alpha < d(t,t') < \beta} 1 \sim \int_{\alpha}^{\beta} du \left(1 - \frac{\sin^2(\pi u)}{\pi^2 u^2}\right) \tag{47}$$

where $d(t, t')$ is a normalized distance between zeros $d(t, t') = \frac{1}{2\pi} \log\left(\frac{T}{2\pi e}\right)(t - t')$.

In Figure 5 we plot the pair correction function for the first 10^5-th $\widehat{t}_n$'s. We chose $\sigma_1 = 0.274$ since in this range of n this gives a better fit to the normal distribution of the 1-point function. The results are surprisingly close to the GUE prediction (47), especially considering that for just the first 10^5 true zeros the fit to the GUE prediction is not perfect; for much higher zeros it is significantly better [10].

VI. COMPUTING VERY HIGH ZEROS FROM THE PRIMES

This section can be viewed as providing numerical evidence for some of the previous results. Since we will be calculating $S(t)$ from the primes using (33), which requires $\Re(s) \to$

1/2, this is pushing the limit of the validity of the Euler product formula, nevertheless we will obtain reasonable results.

Many very high zeros of ζ have been computed numerically, beginning with the work of Odlyzko. All zeros up to the 10^{13}-th have been computed and are all on the critical line [11]. Beyond this the computation of zeros remains a challenging open problem. However some zeros around the 10^{21}-st and 10^{22}-nd are known [12]. In this section we describe a new and simple algorithm for computing very high zeros based on the results of Section IV. It will allow us to go much higher than the known zeros since it does not require numerical implementation of the ζ function itself, but rather only requires knowledge of some of the lower primes.

Let us first discuss the numerical challenges involved in computing high zeros from the equation (40) based on the standard Mathematica package. The main difficulty is that one needs to implement the $\arg\zeta$ term. Mathematica computes $\mathrm{Arg}\,\zeta$, i.e. on the principal branch, however near a zero this is likely to be valid due to (39). The main problem is that Mathematica can only compute ζ for t below some maximum value around $t = 10^{10}$. This was sufficient to calculate up to the $n = 10^9$-th zero from (40) in [6]. The $\log\Gamma$ term must also be implemented to very high t, which is also limited in Mathematica.

We deal with these difficulties first by computing $\arg\zeta$ from the formula (33) involving a finite sum over primes. Then, the $\log\Gamma$ term can be accurately computed using corrections to Stirling's formula:

$$\vartheta(t) = \frac{t}{2}\log\left(\frac{t}{2\pi e}\right) - \frac{\pi}{8} + \frac{1}{48\,t} + O(1/t^3) \tag{48}$$

Let $t_{n;N}$ denote the ordinate of the n-th zero computed using the first N primes. For high zeros, it is approximately the solution to the following equation

$$\frac{t_{n;N}}{2}\log\left(\frac{t_{n;N}}{2\pi e}\right) - \frac{\pi}{8} + \frac{1}{48 t_{n;N}} - \lim_{\delta\to 0^+} \Im \sum_{k=1}^{N} \log\left(1 - \frac{1}{p_k^{1/2+\delta+it_{n;N}}}\right) = (n - \tfrac{3}{2})\pi \tag{49}$$

The important property of this equation is that it not longer makes any reference to ζ itself. It is straightforward to solve the above equation with standard root-finder software.

One can view the computation of t_n as a kind of stochastic process. If one includes no primes, i.e. $N = 0$, and drops the next to leading $1/t$ corrections, then the solution is unique and explicitly given by $t_{n;0} = \widetilde{t}_n$ in (42). One then goes from $t_{n;0}$ to $t_{n;1}$ by finding the root to the equation for $t_{n;1}$ in the vicinity of $t_{n;0}$, then similarly $t_{n;2}$ is calculated based on $t_{n;1}$

n	$t_{n;N}$	Odlyzko
$10^{21}-1$	144176897509546973538.205	$\sim .225$
10^{21}	144176897509546973538.301	$\sim .291$
$10^{21}+1$	144176897509546973538.505	$\sim .498$
$10^{22}-1$	1370919909931995308226.498	$\sim .490$
10^{22}	1370919909931995308226.614	$\sim .627$
$10^{22}+1$	1370919909931995308226.692	$\sim .680$

TABLE I. Zeros around the $n = 10^{21}$-first and 10^{22}-nd computed from (49) with $N = 5 \times 10^6$ primes. We fixed $\delta = 10^{-6}$. Above, $\sim$ denotes the integer part of the second column.

and so forth. At each step in the process one includes one additional prime, and this slowly approaches t_n, so long as $N(t) < N_{\max}(t)$.

For very high t, $N_{\max}(t) = [t^2]$ is extremely large and it is not possible in practice to work with such a large number of primes. This is the primary limitation to the accuracy we can obtain. We will limit ourselves to $N = 5 \times 10^6$ primes. Let us very roughly estimate the error in computing a zero at a given $N < N_{\max}$. Since according to Proposition 4, $S(t) = O(1)$, the integer part of $t_{n;0}$ is expected to be correct, and we verify this below. Now, in each step $t_{n;N-1} \to t_{n;N}$ one includes an additional term in (49) that is approximately $1/\sqrt{p_N}$ and this can be used to roughly estimate the error. For example, if we work with only the first $N = 10^6$ primes, we expect to get the first 2 to 3 digits beyond the decimal point correct. Let us verify this by comparing with some known zeros around $n = 10^{21}$ and 10^{22}. The results are shown in Table I. As predicted, we have accuracy to about 2 digits beyond the decimal point. Odlyzko was able to calculate a few more digits; our accuracy can be improved by increasing N of course.

Having made this check, let us now go far beyond this and compute the $n = 10^{100}$-th zero by the same method. At higher t_n one does not need to include more primes if one is still only interested in 2 digits beyond the decimal point. Using $N = 5 \times 10^6$ primes, we found the following t_n:

10^{100}th zero :

$$
\begin{aligned}
&280690383842894069903195445838256400084548030162846\\
&\quad 0451923600592249309223490730430603356531092524 73.244....
\end{aligned}
$$

We are confident that the last 3 digits $\sim .244$ are accurate since we checked that they didn't change beyond $N = 10^6$. We calculated the next zero to be $\sim .273$. By the same procedure, we were easily able to calculate the 10^{1000}-th zero to the same accuracy.

ACKNOWLEDGMENTS

We wish to thank Guilherme França for discussions.

---

[1] G. França and A. LeClair, "On the validity of the Euler product inside the critical strip", arXiv:1410.3520 [math.NT].

[2] G. França and A. LeClair, "Some Riemann Hypotheses from Random Walks over Primes", arXiv:1509.03643 [math.NT].

[3] M. Kac, *Statistical Independence in Probability, Analysis and Number Theory*, The Mathematical Association of America, New Jersey, 1959.

[4] S. M. Gonek, C. P. Hughes, and J. P. Keating, Duke Math. J **136** (2007) 507.

[5] S. M. Gonek, Trans. Amer. Math. Soc. **364** (2011) 2157.

[6] G. França and A. LeClair, Commun. Number Theory and Phys. **9** (2015) 1-50.

[7] A. A. Karatsuba and M. A. Korolev, Russian Math. Surveys **60** (2005) 433.

[8] D. A. Goldston and S. M. Gonek, Bull. London Math. Soc. **39** (2007) 482.

[9] H. Montgomery, in *Analytic number theory, Proc. Sympos. Pure Math. XXIV* (Providence, RI: AMS, 1973).

[10] A. M. Odlyzko, Math. Comp. **48**. 273 (1987).

[11] X. Gourdon, "The 10^{13} First Zeros of the Riemann Zeta Function, and Zeros Computation at Very Large Height," 2004, http://numbers.computation.free.fr/Constants/Miscellaneous/zetazeros1e13-1e24.pdf.

[12] A. M. Odlyzko, The 10^{21}-st zero of the Riemann zeta function, www.research.att.com/$\sim$amo, 1998.