Large Deviation Estimates of Selberg's Central Limit Theorem and Applications

Louis-Pierre Arguin*[1,2] and Emma Bailey†[2]

[1]Department of Mathematics, Baruch College, CUNY, New York, NY
[2]Department of Mathematics, CUNY Graduate Center, New York, NY

July 13, 2023

Abstract

For $V \sim \alpha \log\log T$ with $0 < \alpha < 2$, we prove

$$\frac{1}{T}\text{meas}\{t \in [T, 2T] : \log|\zeta(1/2 + \mathrm{i}t)| > V\} \ll \frac{1}{\sqrt{\log\log T}} e^{-V^2/\log\log T}.$$

This improves prior results of Soundararajan and of Harper on the large deviations of Selberg's Central Limit Theorem in that range, without the use of the Riemann hypothesis. The result implies the sharp upper bound for the fractional moments of the Riemann zeta function proved by Heap, Radziwiłł and Soundararajan. It also shows a new upper bound for the maximum of the zeta function on short intervals of length $(\log T)^\theta$, $0 < \theta < 3$, that is expected to be sharp for $\theta > 0$. Finally, it yields a sharp upper bound (to order one) for the moments on short intervals, below and above the freezing transition. The proof is an adaptation of the recursive scheme introduced by Bourgade, Radziwiłł and one of the authors to prove fine asymptotics for the maximum on intervals of length 1.

1 Introduction

1.1 Main Result

Selberg's Central Limit Theorem [Sel46, Sel92] states that the logarithm of the Riemann zeta function $\zeta(s)$ at a typical point on the critical line Re $s = 1/2$ behaves like a complex Gaussian random variable of mean 0 and variance $\log\log T$. Specifically, if τ is uniformly distributed on $[T, 2T]$, then for the real part of the logarithm we have

$$\mathbf{P}\Big(\log|\zeta(1/2 + \mathrm{i}\tau)| > \sqrt{\tfrac{1}{2}\log\log T} \cdot y\Big) \sim \int_y^\infty \frac{e^{-z^2/2}}{\sqrt{2\pi}}\mathrm{d}z, \quad y \in \mathbb{R}, \text{ as } T \to \infty.$$

*Email: `louis-pierre.arguin@baruch.cuny.edu`
†Email: `ebailey@gc.cuny.edu`

See [RS17] for an elegant self-contained proof of this, and [Sou21] for a survey on the distribution of values of L-functions in general. In this paper, we prove that the above Gaussian decay persists in the large deviation regime:

Theorem 1.1. *Let $V \sim \alpha \log\log T$ with $0 < \alpha < 2$. We have for T large enough*

$$\mathbf{P}(\log|\zeta(1/2 + \mathrm{i}\tau)| > V) \ll \frac{1}{\sqrt{\log\log T}} \exp\left(\frac{-V^2}{\log\log T}\right).$$

The implicit constant in the inequality can be taken uniform in α in any compact subset of $(0, 2)$.

Throughout the paper, the notation $\ll$ means that the left is O of the right side as $T \to \infty$, and that the implicit constant is possibly α-dependent.

In the interval $0 < V < 2\log\log T$, Theorem 1.1 is an improvement of a more general theorem of Soundararajan [Sou09], which states for this particular range that

$$\mathbf{P}(\log|\zeta(1/2 + \mathrm{i}\tau)| > V) \ll (\log T)^{\mathrm{o}(1)} \cdot \exp\left(\frac{-V^2}{\log\log T}\right). \tag{1}$$

Harper [Har13] also proved sharp bounds for the moments of the zeta function, which by Markov's inequality imply

$$\mathbf{P}(\log|\zeta(1/2 + \mathrm{i}\tau)| > V) \ll \exp\left(\frac{-V^2}{\log\log T}\right). \tag{2}$$

Both results assume the Riemann hypothesis, whereas Theorem 1.1 is unconditional. Equations (1) and (2) do hold conditionally on a wider range of V, for example $V \sim k \log\log T$ for any $k > 0$.

Heap, Radziwiłł and Soundararajan proved sharp upper bounds for the moments between 0 and 4, cf. Corollary 1.2, which imply Equation (2) unconditionally. For $\sqrt{\log\log T}\log\log\log T \leq V \leq 2\log\log T - 2\sqrt{\log\log T}\log\log\log T$, Heap and Soundararajan [HS20] also proved the following behavior unconditionally

$$\mathbf{P}(\log|\zeta(1/2 + \mathrm{i}\tau)| > V) = \exp\left(\frac{-V^2}{\log\log T} + \mathrm{O}\left(\frac{V\log\log\log T}{\sqrt{\log\log T}}\right)\right).$$

It was conjectured by Radziwiłł [Rad11] that the Gaussian behavior actually extends to the whole range $V \sim k\log\log T$, $k > 0$, up to a multiplicative factor

$$\mathbf{P}\Big(\log|\zeta(1/2 + \mathrm{i}\tau)| > V\Big) \sim C_k \int_V^\infty \frac{e^{-y^2/\log\log T}}{\sqrt{\pi \log\log T}} \mathrm{d}y,$$

where C_k is the conjectured leading coefficient of the $2k$-moment (cf. [KS00]). If we write $V = \alpha\log\log T + \sigma y$ for $\sigma = \mathrm{o}(\sqrt{\log\log T})$, then Theorem 1.1 also gives an upper bound to order one for a local version of Selberg's Central Limit Theorem, as proposed in [DBMN19]. (See Proposition 4.8 there for a more precise result for a random model of zeta.) Finally, we also remark that for characteristic polynomials of random unitary matrices, large deviations in the equivalent regime to Theorem 1.1 were proved in [HKO01] and precise asymptotics (including the constant) were proved in [FMN16].

Theorem 1.1 is proved in Section 2. The method is an adaptation of a recursive scheme introduced in [ABR20] to prove a sharp upper bound to the Fyodorov-Hiary-Keating Conjecture, cf. Equation (7). Consider the Dirichlet polynomials

$$S_k = \sum_{\log 2 \leq \log p \leq e^k} \frac{\operatorname{Re} p^{-\mathrm{i}\tau}}{p^{1/2}}, \quad k \geq 1. \tag{3}$$

These partial sums are a good proxy for $\log|\zeta(1/2 + \mathrm{i}\tau)|$ for k close to $\log\log T$. Moreover, the moments of S_k are very close to Gaussian, see for example Lemma A.2 or [ABB+19, Lemma 3.4]. However, the error for these moments is too large to handle simultaneously k close to $\log\log T$ as well as moments of order $\log\log T$.

The idea is to restrict the estimate of the probability to good events where the partial sums (3) takes values in a narrow interval. The implementation of this recursive scheme is much simpler here than in [ABR20], where restrictions at every k were needed. Namely, for Theorem 1.1, the partial sums only need to be constrained on a sparse collection of k's of the form

$$t_\ell = \log\log T - \mathfrak{s} \log_{\ell+2} T, \quad \ell \geq 1, \tag{4}$$

for some (α-dependent) $\mathfrak{s}$, where $\log_\ell$ stands for the logarithm iterated ℓ times. Moreover, since Theorem 1.1 only concerns large values of ζ at a single point, no discretization is needed here compared to [ABR20] where the authors considered the maximum of ζ over a short O(1)-range. This simplifies the statements and proofs of various foundational results (cf. Lemmas 2.4, 2.6, and 2.7) regarding second and twisted fourth moments of Dirichlet polynomials. As a corollary to Theorem 1.1, we prove an upper bound on the maximum of ζ over a growing window, cf. Corollary 1.3.

The restriction is on good events of the form

$$\{S_{t_\ell} \in [L_\ell, U_\ell]\}, \quad \ell \geq 1,$$

where L_ℓ is slightly below the linear interpolation αt_ℓ and U_ℓ is slightly above. These barriers must be chosen carefully and dependent on α. Also, U_ℓ must be much higher than the upper barrier picked in [ABR20] as the fluctuations here can be greater. It turns out that the dominant term of the probability in Theorem 1.1 comes from the intersection of all the good events above. On these events, the increments $S_{t_{\ell+1}} - S_{t_\ell}$ are restricted to a range where large deviations can be estimated.

Theorem 1.1 must be restricted to $\alpha < 2$ since we rely on a twisted fourth moment estimate (Lemma 2.10). More generally, large deviations in the range $\alpha \log\log T$ are controlled by the 2α-moment of zeta. This suggests that the method of proof should be adaptable to prove Theorem 1.1 for any $\alpha > 0$ assuming the Riemann hypothesis, where all such moments can be sharply bounded. This would improve the bounds (1) and (2) in the full range $\alpha \log\log T$, $\alpha > 0$, conditionally. We also expect that a matching lower bound (up to constant) can be found using the techniques of [ABR23]. In [Rad11], it was proved that Selberg's theorem holds up to V of the order of $(\log\log T)^{3/5-\varepsilon}$. Subsequently, Inoue [Ino19] improved the range of V up to $(\log\log T)^{2/3}$. The techniques involved in the proof of Theorem 1.1 do not seem to be applicable to the range $V = \mathrm{o}(\log\log T)$. Interestingly, this leaves a gap between $V \ll (\log\log T)^{2/3}$ and $V \sim \alpha \log\log T$ where the Gaussian decay remains open.

1.2 Applications

The first corollary of Theorem 1.1 is an alternative proof of a sharp upper bound for fractional moments of the zeta function, proved unconditionally by Heap, Radziwiłł and Soundararajan.

Corollary 1.2 (Theorem 1 in [HRS19])**.** *Let $0 < \beta < 4$. We have for T large enough*

$$M_\beta = \frac{1}{T}\int_T^{2T} |\zeta(1/2 + \mathrm{i}t)|^\beta \mathrm{d}t \ll (\log T)^{\beta^2/4}, \tag{5}$$

where the implicit constant depends on β.

The proof in [HRS19] depends on twisted fourth moment estimates, as for Theorem 1.1. Hence, their proof might be considered at the same conceptual level as the proof of Corollary 1.2. Corollary 1.2 is proved in Section 3.1. Note that, via Markov's inequality, Equation (5) shows in particular the Gaussian decay (2) unconditionally, for $V \sim \frac{\beta}{2}\log\log T$ and $\beta \in [0, 4]$.

In short intervals, of size $(\log T)^\theta$ for $0 \le \theta < 3$, Theorem 1.1 implies an upper bound for the maximum up to order one precision:

Corollary 1.3. *Let $0 \le \theta < 3$ and $y > 0$ such that $y = \mathrm{o}\left(\frac{\log\log T}{\log\log\log T}\right)$. We have*

$$\max_{|h| \le (\log T)^\theta} |\zeta(1/2 + \mathrm{i}t + \mathrm{i}h)| \le e^y \frac{(\log T)^{\sqrt{1+\theta}}}{(\log\log T)^{1/(4\sqrt{1+\theta})}}, \tag{6}$$

for all $t \in [T, 2T]$ except on a set of Lebesgue measure $\ll e^{-2\sqrt{1+\theta}y} e^{-y^2/\log\log T}$.

The restriction to $\theta < 3$ is due to the limitations in the range of large deviation, up to $2\log\log T$, in Theorem 1.1. The result also gives a precise decay for the right tail of the maximum, which is exponential for small y's and Gaussian for large ones. The condition on the size of y in the statement of the corollary can be relaxed at the expense of a different decay rate, as can be easily observed within the proof. Upper and lower bounds for the maximum with error $(\log T)^\varepsilon$ were proved in [AOR21]. Corollary 1.3 proves the fine asymptotics up to order one as given in Conjecture 1.3 of [AOR21]. The proof of Corollary 1.3 is given in Section 3.2. It is a simple union bound after suitably discretizing the interval on $(\log T)^{1+\theta}$ points. It is expected that the bound is sharp for $\theta > 0$, see [AAB+21] for numerical evidence of this. This is because for $\theta > 0$, the values of zeta at the $(\log T)^{1+\theta}$ points should each behave like IID Gaussians of variance $\frac{1}{2}\log\log T$, see for example [AOR21] [1]. This is in contrast with the case $\theta = 0$. Corollary 1.3 holds for this case, but it is not sharp. It was conjectured by Fyodorov, Hiary & Keating and Fyodorov & Keating, that the maximum

[1] Closely related is a class of models called 'continuous random energy models', cf. [BKL02, Bov06, Bov17, BH15] that exhibit similar extreme value statistics for a suitable choice of parameters.

of $\log|\zeta|$ on intervals of size one should behave exactly like the maximum of log-correlated stochastic processes [FHK12, FK14]. It was shown in [ABR20] that

$$\max_{|h|\leq 1}|\zeta(1/2+\mathrm{i}t+\mathrm{i}h)|\leq e^{y}\frac{\log T}{(\log\log T)^{3/4}}, \tag{7}$$

for all $t\in[T,2T]$ except on a set of Lebesgue measure $\ll ye^{-2y}e^{-y^2/\log\log T}$. Upper and lower bounds with error $(\log T)^{\varepsilon}$ were proved in [Naj18, ABB$^+$19]. A hybrid regime interpolating between IID and log-correlated statistics was also proposed in [ADH21]. For more on recent developments in extreme values of log-correlated processes, see for example [BK22].

Theorem 1.1 can also be applied to improve current bounds for the moments of ζ in short intervals.

Corollary 1.4. *Let $0\leq\theta<3$. For all $\beta\geq 0$, we have for $A>1$*

$$\int_{|h|\leq(\log T)^{\theta}}|\zeta(1/2+\mathrm{i}t+\mathrm{i}h)|^{\beta}\mathrm{d}h\leq A(\log T)^{\frac{\beta^2}{4}+\theta}, \tag{8}$$

for all $t\in[T,2T]$ except possibly on a subset of Lebesgue measure $\ll 1/A$.

For $\beta>\beta_c=2\sqrt{1+\theta}$, a sharper bound holds:

$$\int_{|h|\leq(\log T)^{\theta}}|\zeta(1/2+\mathrm{i}t+\mathrm{i}h)|^{\beta}\mathrm{d}h\leq C_{A,\beta}\cdot(\log\log T)^{-\frac{\beta}{2\beta_c}}\cdot(\log T)^{\frac{\beta_c}{2}\beta-1}, \tag{9}$$

for all $t\in[T,2T]$ except possibly on a subset of Lebesgue measure $\ll 1/A$, where $C_{A,\beta}$ is an explicit constant dependent on A and β.

Equation (8) was proved in [AOR21]. It follows easily by Markov's inequality and the bound (5). Nevertheless, we provide another proof of this using the Lebesgue measure of high points. This is helpful in understanding the proof of the sharper bound for the moments above β_c. Equation (9) is an improvement on [AOR21], where the result was given with a $(\log T)^{\varepsilon}$ error. Interestingly, Equation (9) is exactly the behavior expected for the moments of $(\log T)^{1+\theta}$ IID Gaussian random variables of variance $\frac{1}{2}\log\log T$ as computed by Bovier, Kurkova & Löwe [BKL02, Theorem 1.6] for large β.

Equations (8) and (9) exhibit a *freezing transition* (also referred to as *intermittency*) where the moments transition from quadratic to linear growth. In view of this, it is natural to ask if the bound (8) at criticality $\beta=\beta_c$ is sharp. At $\theta=0$, where the system seems to behave like a log-correlated process, it can be improved as shown by Harper:

Theorem 1.5 (Theorem 1 and Corollary 1 in [Har19])**.** *We have*

$$\int_{|h|\leq 1}|\zeta(1/2+\mathrm{i}t+\mathrm{i}h)|^{2}\mathrm{d}h\leq A\frac{\log T}{\sqrt{\log\log T}},$$

for all $t\in[T,2T]$ except possibly on a subset of Lebesgue measure $\ll\frac{(\log A)\wedge\sqrt{\log\log T}}{A}$.

The presence of the correction $1/\sqrt{\log\log T}$ is related to the phenomenon of critical Gaussian multiplicative chaos, see [Pow18]. In Section 4, we explain how this correction appears in view of the Lebesgue measure of high points. For $\theta > 0$, where the IID heuristic prevails, such a correction should be absent as predicted by Theorem 1.6 (i) of [BKL02]. Hence, Equation (8) is expected to be sharp to order one at $\beta = \beta_c$.

Notation. Throughout the proofs, τ designates a uniform random variable on $[T, 2T]$. To lighten the notation, we will often use the probabilistic convention for random variables and drop the dependence on τ. Most dramatically, we will simply write

$$\zeta \text{ for the random variable } \zeta(1/2 + \mathrm{i}\tau).$$

Another convenient notation is

$$t = \log\log T.$$

It turns out that log log is the correct scale for the primes in the considered problems. This is because the Dirichlet sums considered, see for example (3) and (10) below, behave like a random walk on that scale.

Acknowledgements We thank Paul Bourgade and Maksym Radziwiłł for insightful discussions on the subject. The research of LPA is supported in part by the awards NSF CAREER DMS-1653602 and NSF DMS-2153803. Part of this work was conducted whilst EB participated in a program during the Fall 2021 semester hosted by the Mathematical Sciences Research Institute in Berkeley, California, which was supported by the NSF Grant DMS-1928930.

2 Proof of Theorem 1.1

The proof is an adaptation of the recursive scheme of [ABR20]. First, we introduce some notations. Consider the partial Dirichlet sums

$$S_k = \sum_{2 \leq p \leq \exp(e^k)} \frac{\mathrm{Re}\, p^{-\mathrm{i}\tau}}{p^{1/2}} + \frac{\mathrm{Re}\, p^{-2\mathrm{i}\tau}}{2p}, \quad k \geq 1, \tag{10}$$

with $S_0 = 0$. (As opposed to the simpler Equation (3), we include here the square of primes within the definition. This simplifies the application of Lemma 2.5 below.) For S_k to be a good approximation for $\log|\zeta|$, the parameter k must be taken close to t. With this in mind, set ~~t is approached in a finite number of steps by iterated logarithms as in (4)~~:

$$t_\ell = t - \mathfrak{s} \log_\ell t, \quad \ell \geq 1, \tag{11}$$

with the convention that $t_0 = 0$. The parameter $\mathfrak{s}$ here depends on α. A good choice (reflecting the symmetry in α) is

$$\mathfrak{s} = \frac{2 \cdot 10^6}{(2-\alpha)^2 \alpha^2}. \tag{12}$$

We will say more on this choice below Equation (19), but as a first remark notice that $\mathfrak{s} \geq 10^6$ since $\alpha \in (0,2)$. The last ℓ, denoted by $\mathcal{L}$, is defined as the largest ℓ such that

$$\exp(10^6(t-t_\ell)^{10^5} e^{t_{\ell+1}}) \leq \exp\left(\frac{1}{100}e^t\right) = T^{1/100}. \tag{13}$$

Note that the left-hand side is

$$\exp\left(10^6(\mathfrak{s}\log_\ell t)^{10^5} \cdot \frac{e^t}{(\log_\ell t)^{\mathfrak{s}}}\right),$$

therefore the choice of $\mathfrak{s}$ ensures that such a $\mathcal{L}$ exists if T is large enough. The finite sequence $t_1, \ldots, t_{\mathcal{L}}$ approaches t such that $t - t_{\mathcal{L}} = \mathfrak{s}\log_{\mathcal{L}} t = \mathrm{O}(1)$ and $\log_{\mathcal{L}} t > 0$.

The corresponding complex partial sums are also needed and are denoted by

$$\widetilde{S}_k = \sum_{2\leq p\leq \exp(e^k)} \frac{p^{-\mathrm{i}\tau}}{p^{1/2}} + \frac{p^{-2\mathrm{i}\tau}}{2p}, \quad k \geq 1, \tag{14}$$

and $\widetilde{S}_0 = 0$. We stress that only the values of the partial sums at t_ℓ, $1 \leq \ell \leq \mathcal{L}$, are necessary.

To approximate $\exp(-S_{t_\ell})$, we use the mollifiers:

$$\mathcal{M}_\ell = \sum_{\substack{p|m \implies \log\log p \in (t_{\ell-1}, t_\ell] \\ \Omega_\ell(m) \leq (t_\ell - t_{\ell-1})^{10^5}}} \frac{\mu(m)}{m^{\frac{1}{2}+i\tau}}, \tag{15}$$

where $\Omega_\ell(m)$ is the number of prime factors of m in $(\exp(e^{t_{\ell-1}}), \exp(e^{t_\ell})]$ with multiplicity, and $\mu(m)$ is the Möbius function. The proof will show that product $\mathcal{M}_1 \cdots \mathcal{M}_\ell$ is typically a good approximation for $\exp(-S_{t_\ell})$.

The idea of the proof is to partition the event

$$H = \{\log|\zeta(1/2 + \mathrm{i}\tau)| > V\}$$

into recursively defined events that greatly restrict the values of the Dirichlet sums (10) and (14). It is expected that, if $\log|\zeta(1/2+\mathrm{i}\tau)| > V$ and $V \sim \alpha t$, then the partial sum S_{t_ℓ} should be close to κt_ℓ where

$$\kappa = \frac{V}{t} \sim \alpha. \tag{16}$$

More precisely, consider for $1 \leq \ell \leq \mathcal{L}$, the decreasing events

$$\begin{aligned} A_\ell &= A_{\ell-1} \cap \{|\widetilde{S}_{t_\ell} - \widetilde{S}_{t_{\ell-1}}| \leq \mathcal{A}(t_\ell - t_{\ell-1})\} \\ B_\ell &= B_{\ell-1} \cap \{S_{t_\ell} \leq \kappa t_\ell + \mathcal{B}\log_\ell t\} \\ C_\ell &= C_{\ell-1} \cap \{S_{t_\ell} \geq \kappa t_\ell - \mathcal{C}\log_\ell t\} \\ D_\ell &= D_{\ell-1} \cap \{|\zeta e^{-S_{t_\ell}}| \leq c_\ell |\zeta \mathcal{M}_1 \cdots \mathcal{M}_\ell| + e^{-\mathcal{D}(t-t_{\ell-1})}\}, \end{aligned} \tag{17}$$

where $c_\ell = \prod_{j=1}^{\ell}(1+e^{-t_{j-1}})$, and $A_0, B_0, C_0, D_0 = [T, 2T]$ (the full sample space). The parameters $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ will be chosen carefully as discussed below. For now, we simply observe that on the *good* event

$$G_\ell = A_\ell \cap B_\ell \cap C_\ell \cap D_\ell,$$

the partial sums are restricted in a narrow corridor between an upper and lower barrier:

$$U_\ell = \kappa t_\ell + \mathcal{B} \log_\ell t \qquad L_\ell = \kappa t_\ell - \mathcal{C} \log_\ell t, \quad 1 \le \ell \le \mathcal{L}. \tag{18}$$

The auxiliary event D_ℓ ensures that $\exp(-S_{t_\ell})$ is well approximated by the mollifier, and A_ℓ is an *a priori estimate* needed for the estimates involving C_ℓ and D_ℓ. The probability of $H = \{\log|\zeta(1/2 + i\tau)| > V\}$ can then be decomposed over the G_ℓ's. The dominant contribution comes from $H \cap G_\mathcal{L}$ where the sums are restricted up to order one away from t. The precise estimates are:

Proposition 2.1. *Let $V \sim \alpha t$ with $0 < \alpha < 2$. With the notation above, we have for some $\delta > 0$ (dependent on α) and t large enough*

$$\mathbf{P}(H \cap G_1^c) \ll \frac{e^{-V^2/t}}{\sqrt{t}} \cdot t^{-\delta}.$$

Proposition 2.2. *Let $V \sim \alpha t$ with $0 < \alpha < 2$. With the notation above, we have for $1 \le \ell \le \mathcal{L} - 1$, some $\delta > 0$ (dependent on α but not ℓ), and t large enough*

$$\mathbf{P}(H \cap G_\ell \cap G_{\ell+1}^c) \ll \frac{e^{-V^2/t}}{\sqrt{t}} \cdot (\log_\ell t)^{-\delta}.$$

Proposition 2.3. *Let $V \sim \alpha t$ with $0 < \alpha < 2$. With the notation above, we have for t large enough*

$$\mathbf{P}(H \cap G_\mathcal{L}) \ll \frac{1}{\sqrt{t}} e^{-V^2/t}.$$

The theorem is a simple consequence of the three propositions.

Proof of Theorem 1.1. It suffices to notice that

$$\mathbf{P}(H) = \mathbf{P}(H \cap G_1^c) + \sum_{\ell=1}^{\mathcal{L}-1} \mathbf{P}(H \cap G_\ell \setminus G_{\ell+1}) + \mathbf{P}(H \cap G_\mathcal{L}).$$

The result follows by applying Propositions 2.1, 2.1, 2.3. □

As mentioned above, the parameters in (17) need to be chosen in a delicate manner. As we shall see from the proof (cf. Equations (29) and (35)), the choice of $\mathcal{B}$ must satisfy the following restrictions.

$$\begin{aligned} 1 + \alpha^2 \mathfrak{s} - 2\alpha\mathcal{B} < 0 \\ \mathcal{B} - \alpha\mathfrak{s} < 0. \end{aligned} \tag{19}$$

The first equation forces $\mathcal{B}$ to be proportional to $1/\alpha$ to handle small α's. In turn, the second equation leads to $\mathfrak{s} > 1/\alpha^2$, motivating in part the choice of $\mathfrak{s}$ in (12). With this choice, the defining inequalities for $\mathcal{B}$ becomes

$$\frac{1}{2\alpha} + \frac{10^6}{\alpha(2-\alpha)^2} < \mathcal{B} < \frac{10^6}{\alpha(2-\alpha)^2} + \frac{10^6}{\alpha(2-\alpha)^2}.$$

This is a non-empty interval since $\alpha > 0$. Therefore, a valid choice is

$$\mathcal{B} = \frac{3 \cdot 10^6}{2\alpha(2-\alpha)^2} + \frac{1}{4\alpha}. \tag{20}$$

The restrictions on $\mathcal{C}$ (cf. Equations (32) and (39)) will be

$$\mathcal{C} > \frac{1}{2(2-\alpha)}\Big\{1 + (2-\alpha)^2\mathfrak{s}\Big\}. \tag{21}$$

(We note in passing that this is the first constraint for $\mathcal{B}$ in (19), after the transformation $\alpha \mapsto 2-\alpha$.) Therefore, a valid choice for $\mathcal{C}$ is

$$\mathcal{C} = \frac{3 \cdot 10^6}{2\alpha^2(2-\alpha)} + \frac{1}{4(2-\alpha)}. \tag{22}$$

This choice implies the upper bound $\mathcal{C} < (2-\alpha)\mathfrak{s}$.

The parameter $\mathcal{A}$ will need to satisfy (cf. Equations (34) and (43)):

$$\mathcal{A} > \frac{\alpha^2}{4} + \frac{\alpha\mathcal{C}}{2\mathfrak{s}} + 2. \tag{23}$$

This choice implies in particular

$$\mathcal{A}^2 > \alpha^2 + \frac{2\alpha\mathcal{C}}{\mathfrak{s}} + 4. \tag{24}$$

For example, one can take

$$\mathcal{A} = 10^3, \tag{25}$$

since, with the choices of $\mathcal{C}$ and $\mathfrak{s}$ above, and for $0 < \alpha < 2$, we have

$$\begin{aligned} \mathcal{A} = 10^3 &> 4 + \frac{\alpha(2-\alpha)}{8}(\alpha^2+3) \\ &> \frac{\alpha^2}{4} + \frac{\alpha\mathcal{C}}{2\mathfrak{s}} + 2. \end{aligned} \tag{26}$$

Finally, the conditions on $\mathcal{D}$ will be as in [ABR20]

$$\mathcal{D} = 10^4. \tag{27}$$

2.1 Proof of Proposition 2.1

First, notice that

$$H \cap G_1^c \subset \quad A_1^c \ \cup \ B_1^c \ \cup \ (H \cap C_1^c \cap A_1 \cap D_1) \ \cup \ (D_1^c \cap A_1).$$

We estimate the probability of the four events in the union on the right individually.

We first evaluate A_1^c:

$$\mathbf{P}(A_1^c) = \mathbf{P}(|\widetilde{S}_{t_1}| > \mathcal{A}t_1).$$

Equation (79) of the appendix is applicable with the choice $q = \lceil \mathcal{A}^2 t_1 \rceil$, and implies that this is

$$\ll \sqrt{t_1} \cdot \exp(-\mathcal{A}^2 t_1).$$

Since $\mathcal{A} = 10^3$, for some $\delta > 0$ this is clearly

$$\ll \frac{e^{-\kappa^2 t}}{\sqrt{t}} t^{-\delta}. \tag{28}$$

Turning to B_1^c, and applying Markov's inequality for some $q > 1$ yields

$$\mathbf{P}(B_1^c) \leq \mathbf{P}(S_{t_1} > U_1) \leq U_1^{-2q} \mathbf{E}[|S_{t_1}|^{2q}].$$

Equation (80) then applies with $q = \lceil U_1^2 / t_1 \rceil$, with U_1 as in (18) giving

$$\mathbf{P}(B_1^c) \ll \sqrt{t_1} e^{-U_1^2/t_1} \ll \frac{e^{-\kappa^2 t}}{\sqrt{t}} \cdot t^{1+\kappa^2 \mathfrak{s} - 2\kappa \mathcal{B}}. \tag{29}$$

By the choice of $\mathcal{B}$ in Equation (19), one has $1 + \alpha^2 \mathfrak{s} - 2\alpha \mathcal{B} < 0$. Since $\kappa = \alpha + \mathrm{o}(1)$ by Equation (16), the above is

$$\ll \frac{e^{-\kappa^2 t}}{\sqrt{t}} \cdot t^{-\delta} \tag{30}$$

for some $\delta > 0$, depending on α and different from (28).

To evaluate $H \cap C_1^c \cap A_1 \cap D_1$, we require the following lemma, proved in Section 2.5.

Lemma 2.4. *For w with $|w| \leq 4t_1$, we have*

$$\mathbf{E}\left[|\zeta \mathcal{M}_1|^4 \, \mathbf{1}\left(S_{t_1} \in (w, w+1]\right)\right] \ll e^{4(t-t_1)} \cdot \frac{e^{-w^2/t_1}}{\sqrt{t}}. \tag{31}$$

Let us explain the intuition behind the result. One should think of $\zeta \mathcal{M}_1$ as a random Euler product involving primes larger than $\exp(e^{t_1})$. Furthermore, Selberg's result suggests its logarithm should be distributed like a Gaussian random variable of variance $t - t_1$. This explains the first factor $e^{4(t-t_1)}$ as the contribution from the moment generating function of such a variable. As explained in Section 2.4, the indicator function can be approximated by a suitable Dirichlet polynomial involving primes less than S_{t_1}. Since primes should behave independently, it is not surprising to see the decoupling between the factors. Most importantly, we obtain a Gaussian behavior for the variable S_{t_1} in a large deviation regime.

The estimate $\mathbf{P}(H \cap C_1^c \cap A_1 \cap D_1)$ is done by first partitioning on the value of S_{t_1} using the restrictions given by A_1 and C_1^c:

$$\begin{aligned} \mathbf{P}(H \cap C_1^c \cap A_1 \cap D_1) &\leq \sum_{-\mathcal{A}(t_1 - t_0) < u < L_1} \mathbf{P}(\{S_{t_1} \in (u, u+1], |\zeta| > e^V\} \cap D_1) \\ &\leq \sum_{-\mathcal{A}(t_1 - t_0) < u < L_1} \mathbf{P}(\{S_{t_1} \in (u, u+1], |\zeta e^{-S_{t_1}}| > e^{V-u-1}\} \cap D_1), \end{aligned}$$

where we recall that $V = \kappa t$ and L_1 is given by Equation (18). The event D_1 implies that

$$|\zeta e^{-S_{t_1}}| \leq 2|\zeta \mathcal{M}_1| + e^{-\mathcal{D}(t-t_0)}.$$

Therefore if $|\zeta e^{-S_{t_1}}| > e^{V-u-1}$, then it must be that either

$$2|\zeta \mathcal{M}_1| > \frac{1}{2} e^{V-u-1}$$

or

$$e^{-\mathcal{D}(t-t_0)} > \frac{1}{2} e^{V-u-1}.$$

The latter case is impossible, since the exponent on the left side is negative, whereas on the right side we have on the range of u

$$V - u - 1 > \kappa t - \kappa t_1 + \mathcal{C} \log t - 1 > 0.$$

This implies that

$$\begin{aligned} \mathbf{P}(A_1 \cap D_1 \cap H \cap C_1^c) &\leq \sum_{-\mathcal{A}(t_1-t_0)<u<L_1} \mathbf{P}(|\zeta \mathcal{M}_1| > \frac{1}{100} e^{V-u} \cap \{S_{t_1} \in (u, u+1]\}) \\ &\ll \sum_{u<L_1} e^{-4(V-u)} \cdot \mathbf{E}[|\zeta \mathcal{M}_1|^4 \mathbf{1}(S_{t_1} \in (u, u+1])]. \end{aligned}$$

The sum over $u < 0$ is $\ll e^{-4V} \cdot e^{4(t-t_1)}$ which is much smaller than $\frac{e^{-\kappa^2 t}}{\sqrt{t}} t^{-\delta}$ for the range of V considered. Lemma 2.4 can be applied on the range $0 \leq u < L_1$ This gives

$$\ll e^{4(t-t_1)} \sum_{u<L_1} e^{-4(V-u)} \frac{e^{-u^2/t_1}}{\sqrt{t_1}}.$$

After the change of variable $w = \kappa t_1 - u$, this becomes

$$\begin{aligned} e^{(4-4\kappa)(t-t_1)} \sum_{u<L_1} e^{-4(\kappa t_1 - u)} \frac{e^{-u^2/t_1}}{\sqrt{t_1}} &\ll \frac{e^{-\kappa^2 t_1}}{\sqrt{t_1}} e^{(4-4\kappa)(t-t_1)} \sum_{w>\mathcal{C}\log t} e^{-(4-2\kappa)w} \\ &\ll \frac{e^{-\kappa^2 t}}{\sqrt{t}} \cdot t^{\mathfrak{s}\kappa^2 + \mathfrak{s}(4-4\kappa) - 2(2-\kappa)\mathcal{C}} \\ &\ll \frac{e^{-\kappa^2 t}}{\sqrt{t}} t^{-\delta}, \end{aligned} \tag{32}$$

for some $\delta > 0$, by the choice of $\mathcal{C}$ in (21).

Finally we estimate D_1^c. In order to proceed, we need the following lemma from [ABR20]. The proof follows by expressing $e^{-(S_{t_{\ell+1}} - S_{t_\ell})}$ in terms of an Euler product, and by bounding the contribution of integers m with $\Omega_\ell(m) > (t_\ell - t_{\ell-1})^{10^5}$ using Rankin's trick.

Lemma 2.5 (Lemma 23 in [ABR20])**.** *Suppose $\ell \geq 0$ and that $|\widetilde{S}_{t_{\ell+1}} - \widetilde{S}_{t_\ell}| \leq 10^3(t_{\ell+1} - t_\ell)$. Then we have*

$$e^{-(S_{t_{\ell+1}} - S_{t_\ell})} \leq (1 + e^{-t_\ell})|\mathcal{M}_\ell| + e^{-10^5(t_{\ell+1} - t_\ell)}.$$

Now, observe that the event $A_1 \cap \{|\zeta| \leq e^{2t}\}$ is contained in $A_1 \cap D_1$. Indeed, since $|\widetilde{S}_{t_1} - \widetilde{S}_{t_0}| \leq 10^3(t_1 - t_0)$ on A_1, Lemma 2.5 implies

$$|\zeta e^{-(S_{t_1}-S_{t_0})}| \leq 2|\zeta \mathcal{M}_1| + |\zeta| e^{-10^5(t_1-t_0)} \leq 2|\zeta \mathcal{M}_1| + e^{2t-10^5(t_1-t_0)},$$

which implies D_1 since $\mathcal{D} = 10^4$. Hence, to estimate $\mathbf{P}(D_1^c \cap A_1)$, it suffices to estimate $\mathbf{P}(|\zeta| > e^{2t})$:

$$\mathbf{P}(D_1^c \cap A_1) \leq \mathbf{P}(|\zeta| > e^{2t}) \leq e^{-4t}\mathbf{E}[|\zeta|^2] \ll \frac{1}{\sqrt{t}} e^{-\kappa^2 t} e^{-100t}, \tag{33}$$

since $\mathbf{E}[|\zeta|^2] \ll e^t$ [Theorem 2.41 in [HL18]].

Summarising, we have by a union bound and successively applying Equations (28), (30), (32), and (33),

$$\begin{aligned}\mathbf{P}(H \cap G_1^c) &\leq \mathbf{P}(A_1^c) + \mathbf{P}(B_1^c) + \mathbf{P}(A_1 \cap D_1 \cap H \cap C_1^c) + \mathbf{P}(A_1 \cap D_1^c) \\ &\ll \frac{1}{\sqrt{t}} e^{-\kappa^2 t} t^{-\delta},\end{aligned}$$

for some $\delta > 0$ dependent on α.

2.2 Proof of Proposition 2.2

Notice that

$$H \cap G_\ell \cap G_{\ell+1}^c \subset (A_{\ell+1}^c \cap G_\ell) \cup (B_{\ell+1}^c \cap G_\ell) \cup (H \cap C_{\ell+1}^c \cap A_{\ell+1} \cap D_{\ell+1} \cap G_\ell) \cup (D_{\ell+1}^c \cap A_{\ell+1} \cap G_\ell).$$

The probability of each event in the union on the right side are now evaluated. In order to handle the event involving $A_{\ell+1}^c$ we will need the following lemma, proved in Section 2.4.

Lemma 2.6. *Let $\ell \geq 1$ be such that $10^6(t - t_\ell)^{10^5} e^{t_{\ell+1}} \leq \frac{1}{100} e^t$. Let $\mathcal{Q}$ be a Dirichlet polynomial of length $N \leq \exp(\frac{1}{100} e^t)$, supported on integers all of whose prime factors are greater than $\exp(e^{t_\ell})$. Then for $w \in [L_\ell, U_\ell]$, we have*

$$\mathbf{E}\left[|\mathcal{Q}(\tfrac{1}{2} + \mathrm{i}\tau)|^2 \mathbf{1}\left(B_\ell \cap C_\ell \cap \{S_{t_\ell} \in (w, w+1]\}\right)\right] \ll \mathbf{E}\left[|\mathcal{Q}(\tfrac{1}{2} + \mathrm{i}\tau)|^2\right] \cdot \frac{e^{-w^2/t_\ell}}{\sqrt{t_\ell}}.$$

As in Lemma 2.7, the decoupling is due to the fact that the Dirichlet polynomials involve primes in different intervals. Though the events $B_\ell \cap C_\ell$ do not appear explicitly in the result, their presence here is crucial to obtain the Gaussian behavior of S_{t_ℓ} in a large deviation regime.

We first show that for $\ell \geq 1$

$$\mathbf{P}(A_{\ell+1}^c \cap G_\ell) \ll \frac{e^{-V^2/t}}{\sqrt{t}} \cdot (\log_\ell t)^{-\delta}.$$

For any $q > 1$, the probability $\mathbf{P}(A_{\ell+1}^c \cap G_\ell)$ is smaller than

$$\sum_{u \in [L_\ell, U_\ell]} \mathbf{E}\left[\frac{|\widetilde{S}_{t_{\ell+1}} - \widetilde{S}_{t_\ell}|^{2q}}{(\mathcal{A}(t_{\ell+1} - t_\ell))^{2q}} \mathbf{1}(B_\ell \cap C_\ell \cap \{S_{t_\ell} \in (u, u+1]\})\right].$$

With the choice $q = \lceil \mathcal{A}^2(t_{\ell+1} - t_\ell) \rceil$, the polynomial $\mathcal{Q} = |\widetilde{S}_{t_{\ell+1}} - \widetilde{S}_{t_\ell}|^{2q}$ both satisfies the assumptions of Lemma 2.6 and Lemma A.1. Therefore, the above is

$$\begin{aligned} &\ll \sum_{u \in [L_\ell, U_\ell]} (t_{\ell+1} - t_\ell)^{1/2} e^{-\mathcal{A}^2(t_{\ell+1}-t_\ell)} \frac{e^{-u^2/t_\ell}}{\sqrt{t_\ell}} \\ &\ll (t_{\ell+1} - t_\ell)^{1/2} \cdot e^{-\mathcal{A}^2(t_{\ell+1}-t_\ell)} \cdot \frac{e^{-L_\ell^2/t_\ell}}{\sqrt{t_\ell}}, \end{aligned}$$

where the last inequality is by estimating the sum over u trivially. Since $L_\ell = \kappa t_\ell - \mathcal{C} \log_\ell t$, this is

$$\ll \frac{e^{-\kappa^2 t_\ell}}{\sqrt{t}} \cdot (\log_{\ell-1} t)^{-\mathcal{A}^2\mathfrak{s}+2\kappa\mathcal{C}} \ll \frac{e^{-\kappa^2 t}}{\sqrt{t}} \cdot (\log_{\ell-1} t)^{\kappa^2\mathfrak{s}-\mathcal{A}^2\mathfrak{s}+2\kappa\mathcal{C}}. \tag{34}$$

The choice of parameters in Equation (24) guarantees that the exponent is negative.

Now we show that for $\ell \geq 1$,

$$\mathbf{P}(B_{\ell+1}^c \cap G_\ell) \ll \frac{1}{\sqrt{t}} e^{-\kappa^2 t} \cdot (\log_\ell t)^{-\delta}.$$

By partitioning on the position of S_{t_ℓ}, we have

$$\begin{aligned} \mathbf{P}(B_{\ell+1}^c \cap G_\ell) &\ll \mathbf{P}(B_{\ell+1}^c \cap B_\ell \cap C_\ell) \\ &\ll \sum_{u \in [L_\ell, U_\ell]} \mathbf{P}(\{S_{t_{\ell+1}} - S_{t_\ell} > U_{\ell+1} - u\} \cap \{S_{t_\ell} \in (u, u+1]\} \cap B_{\ell-1} \cap C_{\ell-1}) \\ &\ll \sum_{u \in [L_\ell, U_\ell]} \mathbf{E}\left[\frac{(S_{t_{\ell+1}} - S_{t_\ell})^{2q}}{(U_{\ell+1} - u)^{2q}} \mathbf{1}(S_{t_\ell} \in (u, u+1], S_{t_k} \in [L_k, U_k] \ \forall k < \ell) \right], \end{aligned}$$

where the final line holds for any $q > 1$ by an application of Markov's inequality, provided that $U_{\ell+1} - U_\ell > 0$. This holds by the choice of $\mathcal{B}$ and $\mathfrak{s}$ in Equations (12) and (19).

Choosing $q = \lceil (U_\ell - u)^2/(t_{\ell+1} - t_\ell) \rceil$, then the Dirichlet polynomial $(S_{t_{\ell+1}} - S_{t_\ell})^q$ has length at most $\exp(2qe^{t_{\ell+1}})$ so the the conditions of Lemma 2.6 and Lemma A.2 are satisfied. An application of Equation (81) then yields

$$\mathbf{P}(B_{\ell+1}^c \cap G_\ell) \ll \sum_{u \in [L_\ell, U_\ell]} e^{-\frac{(U_{\ell+1}-u)^2}{t_{\ell+1}-t_\ell}} \cdot \frac{e^{-u^2/t_\ell}}{\sqrt{t_\ell}} \ll \sqrt{t_{\ell+1} - t_\ell} \cdot \frac{e^{-U_{\ell+1}^2/t_\ell}}{\sqrt{t}}.$$

The last bound follows by bounding the sum over u by the Gaussian integral. Since $U_{\ell+1} = \kappa t_{\ell+1} + \mathcal{B} \log_{\ell+1} t$, this is bounded by

$$\ll \frac{e^{-\kappa^2 t}}{\sqrt{t}} \cdot (\log_\ell t)^{1/2+\mathfrak{s}\kappa^2-2\kappa\mathcal{B}}. \tag{35}$$

The choice of $\mathcal{B}$ in Equation (19) ensures that $1/2 + \mathfrak{s}\kappa^2 - 2\kappa\mathcal{B} < 0$.

The next estimate is

$$\mathbf{P}(H \cap C_{\ell+1}^c \cap A_{\ell+1} \cap D_{\ell+1} \cap G_\ell). \tag{36}$$

For this, we need a more detailed version of Lemma 2.4.

Lemma 2.7. *Let $\ell \geq 1$ such that $10^6(t-t_\ell)^{10^5}e^{t_{\ell+1}} \leq \frac{1}{100}e^t$. For $w \in [L_\ell, U_\ell]$, we have*

$$\mathbf{E}\left[|\zeta\mathcal{M}_1\cdots\mathcal{M}_\ell|^4\ \mathbf{1}\left(B_\ell\cap C_\ell, S_{t_\ell}\in[w,w+1]\right)\right] \ll e^{4(t-t_\ell)}\cdot\frac{e^{-w^2/t_\ell}}{\sqrt{t}}. \tag{37}$$

Moreover, let $\gamma(m)$ be a sequence of complex coefficients with $|\gamma(m)| \leq \exp(\frac{1}{1000}e^t)$ for all $m \geq 1$. Set

$$\mathcal{Q}_\ell = \sum_{\substack{p|m \implies \log\log p\in(t_\ell,t_{\ell+1}]\\ \Omega_{\ell+1}(m)\leq(t_{\ell+1}-t_\ell)^{10^4}}}\frac{\gamma(m)}{m^{\frac{1}{2}+i\tau}}.$$

We have

$$\mathbf{E}\left[|\zeta\mathcal{M}_1\cdots\mathcal{M}_{\ell+1}|^4\ |\mathcal{Q}_\ell|^2\mathbf{1}\left(B_\ell\cap C_\ell, S_{t_\ell}\in[w,w+1]\right)\right] \ll e^{4(t-t_{\ell+1})}\cdot\mathbf{E}\left[|\mathcal{Q}_\ell|^2\right]\cdot\frac{e^{-w^2/t_\ell}}{\sqrt{t}}. \tag{38}$$

We now partition the values of $S_{t_\ell}=u$ for $L_\ell\leq u\leq U_\ell$ (on the event G_ℓ) as well as the values of the increments $S_{t_{\ell+1}}-S_{t_\ell}=v$ with the restrictions $u+v<L_{\ell+1}$ (on the event $C^c_{\ell+1}$) and $|v|\leq\mathcal{A}(t_{\ell+1}-t_\ell)$ (on the event $A_{\ell+1}$). The probability in (36) is then smaller than

$$\sum_{\substack{u\in[L_\ell,U_\ell]\\ u+v\leq L_{\ell+1}\\ |v|\leq\mathcal{A}(t_{\ell+1}-t_\ell)}}\mathbf{P}(\{S_{t_\ell}\in(u,u+1], S_{t_{\ell+1}}-S_{t_\ell}\in(v,v+1], |\zeta e^{-S_{t_{\ell+1}}}|>e^{V-(u+v)}\}\cap B_\ell\cap C_\ell\cap D_{\ell+1}).$$

The definition of the event $D_{\ell+1}$ together with the fact that $|\zeta e^{-S_{t_{\ell+1}}}|>e^{V-(u+v+1)}$ implies that either

$$c_{\ell+1}|\zeta\mathcal{M}_1\cdots\mathcal{M}_{\ell+1}|>\frac{1}{2}e^{V-(u+v+1)},$$

or

$$e^{-\mathcal{D}(t-t_\ell)}>\frac{1}{2}e^{V-(u+v+1)}.$$

The last case cannot occur, since the exponent on the left side is negative whereas the one on the right is

$$V-(u+v)-1>V-L_{\ell+1}-1=(\mathfrak{s}\kappa+\mathcal{C})\log_{\ell+1}t-1>0,$$

for $1\leq\ell\leq\mathcal{L}-1$ as $\log_{\mathcal{L}}t>0$ by definition. This reduces this estimate to

$$\sum_{\substack{u\in[L_\ell,U_\ell]\\ u+v\leq L_{\ell+1}\\ |v|\leq\mathcal{A}(t_{\ell+1}-t_\ell)}}\mathbf{P}(\{S_{t_\ell}\in(u,u+1], S_{t_{\ell+1}}-S_{t_\ell}\in(v,v+1], |\zeta\mathcal{M}_1\cdots\mathcal{M}_{\ell+1}|>\frac{1}{100}e^{V-(u+v)}\}\cap B_\ell\cap C_\ell)$$

$$\ll\sum_{\substack{u\in[L_\ell,U_\ell]\\ u+v\leq L_{\ell+1}\\ |v|\leq\mathcal{A}(t_{\ell+1}-t_\ell)}}e^{-4V+4(u+v)}\mathbf{E}\Big[|\zeta\mathcal{M}_1\cdots\mathcal{M}_{\ell+1}|^4\cdot\frac{|S_{t_{\ell+1}}-S_{t_\ell}|^{2q}}{|v|^{2q}}\mathbf{1}(B_\ell\cap C_\ell\cap\{S_{t_\ell}\in(u,u+1]\})\Big],$$

by Markov's inequality with $q = \lceil |v|^2/(t_{\ell+1}-t_\ell)\rceil \leq \mathcal{A}^2(t_{\ell+1}-t_\ell)$. Applications of Lemma 2.7, Equation (38) with $\mathcal{Q}_\ell = (S_{t_{\ell+1}} - S_{t_\ell})^q$ and Equation (81) then implies that this is

$$\ll \sum_{\substack{u\in[L_\ell,U_\ell]\\ u+v\leq L_{\ell+1}\\ |v|\leq \mathcal{A}(t_{\ell+1}-t_\ell)}} e^{-4V+4(u+v)} \cdot e^{4(t-t_{\ell+1})} \cdot e^{-v^2/(t_{\ell+1}-t_\ell)} \frac{e^{-u^2/t_\ell}}{\sqrt{t_\ell}}.$$

The change of variables $\bar{u} = u - \kappa t_\ell$ and $\bar{v} = v - \kappa(t_{\ell+1}-t_\ell)$ and dropping some conditions on the sum gives

$$\ll \frac{e^{-\kappa^2 t_{\ell+1}}}{\sqrt{t}} e^{(4-4\kappa)(t-t_{\ell+1})} \cdot \sum_{\substack{\bar{v}\in\mathbb{Z}\\ \bar{u}+\bar{v}\leq -\mathcal{C}\log_{\ell+1} t}} e^{(4-2\kappa)(\bar{u}+\bar{v})} e^{-\bar{v}^2/(t_{\ell+1}-t_\ell)},$$

where we dropped the term $e^{-\bar{u}^2/t_\ell}$ since it is of order one by the restriction on $\bar{u}$. It remains to sum over $\bar{u}+\bar{v}$ first, then do the Gaussian sum on $\bar{v}$ to get

$$\frac{e^{-\kappa^2 t}}{\sqrt{t}} \cdot (\log_\ell t)^{\mathfrak{s}(2-\kappa)^2 - 2(2-\kappa)\mathcal{C} + 1/2}. \tag{39}$$

Again, the last term is $(\log_\ell t)^{-\delta}$ by the choice of parameters in (21).

Finally we consider $D_{\ell+1}^c \cap A_{\ell+1} \cap G_\ell$. We claim that it is enough to evaluate

$$\mathbf{P}(\{|\zeta\mathcal{M}_1\cdots\mathcal{M}_\ell| > e^{\mathcal{A}(t-t_\ell)}\} \cap G_\ell).$$

To see this, it suffices to notice that $A_{\ell+1} \cap \{|\zeta\mathcal{M}_1\cdots\mathcal{M}_\ell| \leq e^{\mathcal{A}(t-t_\ell)}\} \cap D_\ell$ is in $A_{\ell+1} \cap D_{\ell+1}$. Indeed, on the event $A_{\ell+1} \cap \{|\zeta\mathcal{M}_1\cdots\mathcal{M}_\ell| \leq e^{\mathcal{A}(t-t_\ell)}\} \cap D_\ell$, we have

$$|S_{t_{\ell+1}} - S_{t_\ell}| \leq \mathcal{A}(t_{\ell+1}-t_\ell) \tag{40}$$

$$|\zeta e^{-S_{t_\ell}}| \leq c_\ell |\zeta\mathcal{M}_1\cdots\mathcal{M}_\ell| + e^{-\mathcal{D}(t-t_{\ell-1})} \tag{41}$$

$$|\zeta\mathcal{M}_1\cdots\mathcal{M}_\ell| \leq e^{\mathcal{A}(t-t_\ell)}. \tag{42}$$

Equations (40) and (41) imply that

$$\begin{aligned}|\zeta e^{-S_{t_{\ell+1}}}| &\leq \left(c_\ell|\zeta\mathcal{M}_1\cdots\mathcal{M}_\ell| + e^{-10^4(t-t_{\ell-1})}\right) e^{-(S_{t_{\ell+1}}-S_{t_\ell})}\\ &\leq c_\ell|\zeta\mathcal{M}_1\cdots\mathcal{M}_\ell| e^{-(S_{t_{\ell+1}}-S_{t_\ell})} + e^{-10^3(t-t_{\ell-1})},\end{aligned}$$

for t large enough. Then, Lemma 2.5 gives

$$|\zeta e^{-S_{t_{\ell+1}}}| \leq c_\ell|\zeta\mathcal{M}_1\cdots\mathcal{M}_{\ell+1}| + c_\ell e^{10^3(t-t_\ell)-10^5(t_{\ell+1}-t_\ell)} + e^{-10^3(t-t_{\ell-1})}.$$

We conclude that $D_{\ell+1}$ holds.

It remains to estimate $\mathbf{P}(\{|\zeta\mathcal{M}_\ell| > e^{\mathcal{A}(t-t_\ell)}\} \cap G_\ell)$. We have by subsequent applications of Markov's inequality and Lemma 2.7, Equation (37),

$$\begin{aligned}\mathbf{P}(\{|\zeta\mathcal{M}_\ell| > e^{\mathcal{A}(t-t_\ell)}\} \cap G_\ell) &\ll e^{-4\mathcal{A}(t-t_\ell)}\mathbf{E}[|\zeta\mathcal{M}_\ell|^4 \mathbf{1}(G_\ell)]\\ &\ll e^{-4(\mathcal{A}-1)(t-t_\ell)} \frac{e^{-L_\ell^2/t_\ell}}{\sqrt{t_\ell}}\\ &\ll \frac{1}{\sqrt{t}} e^{-\kappa^2 t} e^{-4\mathfrak{s}(\mathcal{A}-1)\log_\ell t} e^{\mathfrak{s}\kappa^2\log_\ell t + 2\kappa\mathcal{C}\log_\ell t}.\end{aligned}$$

We conclude that

$$\mathbf{P}(D_{\ell+1}^c \cap A_{\ell+1} \cap G_\ell) \ll \frac{1}{\sqrt{t}} e^{-\kappa^2 t} \cdot (\log_{\ell-1} t)^{\mathfrak{s}\kappa^2 + 2\kappa\mathcal{C} - 4\mathfrak{s}(\mathcal{A}-1)}. \tag{43}$$

The exponent is negative by Equation (23).

2.3 Proof of Proposition 2.3

Finally we establish that

$$\mathbf{P}(H \cap G_{\mathcal{L}}) \ll \frac{1}{\sqrt{t}} e^{-\kappa^2 t}.$$

After partitioning on the value of $S_{t_{\mathcal{L}}}$, applying Markov's inequality, and subsequently Lemma 2.7 we have

$$\mathbf{P}(H \cap G_{\mathcal{L}}) \ll \sum_{v \in [L_{\mathcal{L}}, U_{\mathcal{L}}]} e^{4(t-t_{\mathcal{L}})} e^{4v} \frac{e^{-v^2/t_{\mathcal{L}}}}{\sqrt{t_{\mathcal{L}}}}.$$

Applying the transformation $w = v - \kappa t_{\mathcal{L}}$, the probability is bounded by

$$\ll \frac{1}{\sqrt{t}} e^{-\kappa^2 t} e^{(2-\kappa)^2 \log_{\mathcal{L}} t} \sum_{-\mathcal{C} \log_{\mathcal{L}} t < w < \mathcal{B} \log_{\mathcal{L}} t} e^{2(2-\kappa)w}.$$

Since $\alpha < 2$, the sum is bounded by $\exp(2(2-\kappa)\mathcal{B} \log_{\mathcal{L}} t)$, so after grouping we find

$$\ll \frac{1}{\sqrt{t}} e^{-\kappa^2 t} \cdot e^{(2-\kappa+2\mathcal{B})(2-\kappa) \log_{\mathcal{L}} t}.$$

By the choice of $\mathcal{L}$, this is $\ll \frac{1}{\sqrt{t}} e^{-\kappa^2 t}$.

2.4 Proof of Lemma 2.6

We express the event $B_\ell \cap C_\ell \cap \{S_{t_\ell} \in (w, w+1]\}$ in terms of the increments

$$Y_j = S_{t_j} - S_{t_{j-1}}, \quad 1 \le j \le \ell. \tag{44}$$

Recall that $w \in [L_\ell, U_\ell]$. Therefore the event implies that $S_{t_j} \in [L_j, U_j]$ for all $j < \ell$, and $S_{t_\ell} \in [L_\ell, U_\ell + 1]$. We partition these intervals into subintervals of width Δ_j^{-1} where

$$\Delta_j = (t_j - t_{j-1}),$$

so $\Delta_j \le \mathfrak{s} \log_{j-1} t$ for $j > 1$, and Δ_1 is $t_1 = t - \mathfrak{s} \log t$. Note that Δ_j is of the same order as the variance of Y_j. Moreover, we have

$$\sum_{j \ge 1} \Delta_j^{-1} \le 1.$$

Consider the set $\mathcal{I}$ of ℓ-tuples $\mathbf{u} = (u_1, \dots, u_\ell)$ such that

$$\sum_{i=1}^{j} u_i \in [L_j - 1, U_j + 1], \quad j \le \ell, \qquad \sum_{i=1}^{\ell} u_i \in [w-1, w+1]. \tag{45}$$

As a consequence of the definition, we have for all $j > 1$

$$L_j - 1 - (U_{j-1} + 1) \le u_j \le U_j + 1 - (L_{j-1} - 1)$$

which implies $|u_j| \le (\kappa\mathfrak{s} + \mathcal{B} + \mathcal{C}) \log_{j-1} t + 2$. We will also shortly require the following estimate. Since $\mathcal{B} < \alpha\mathfrak{s}$ and $\mathcal{C} < (2-\alpha)\mathfrak{s}$ (by (20) and (22)), we conclude from $\alpha < 2$ that

$$|u_j| < 4\Delta_j + 2. \tag{46}$$

With these definitions, it is straightforward to check that we have the following inclusion of events

$$B_\ell \cap C_\ell \cap \{S_{t_\ell} \in (w, w+1]\} \subset \bigcup_{\mathbf{u} \in \mathcal{I}} \{Y_j \in [u_j, u_j + \Delta_j^{-1}], 1 \le j \le \ell\}. \tag{47}$$

In particular, this implies

$$\mathbf{1}(B_\ell \cap C_\ell \cap \{S_{t_\ell} \in (w, w+1]\}) \le \sum_{\mathbf{u} \in \mathcal{I}} \prod_j \mathbf{1}(Y_j \in [u_j, u_j + \Delta_j^{-1}]). \tag{48}$$

We first prove:

Lemma 2.8. *In the above notation, we have for $A \ge 10$ and $j \le \ell$,*

$$\mathbf{1}(Y_j \in [u_j, u_j + \Delta_j^{-1}]) \le |\mathcal{D}_{\Delta_j, A}(Y_j - u_j)|^2 (1 + c e^{-\Delta_j^{A-1}}), \tag{49}$$

where c is an absolute constant and $\mathcal{D}_{\Delta_j,A}(Y_j - u_j)$ is a Dirichlet polynomial on integers n whose prime factors are in $(\exp(e^{t_{j-1}}), \exp(e^{t_j})]$ with $\Omega(n) \le \Delta_j^{10A}$. In particular, its length is less than $\exp(2e^{t_j}\Delta_j^{10A})$.

Proof. Lemma 6 in [ABR20] states that for any $\Delta, A \ge 3$, there exists an entire function $G_{\Delta,A}(x) \in L^2(\mathbb{R})$ such that for some absolute constant $c > 0$:

1. the Fourier transform $\widehat{G}_{\Delta,A}(x)$ is supported on $[-\Delta^{2A}, \Delta^{2A}]$;
2. $0 \le G_{\Delta,A}(x) \le 1$ for all $x \in \mathbb{R}$;
3. $\mathbf{1}(x \in [0, \Delta^{-1}]) \le G_{\Delta,A}(x) \cdot (1 + c e^{-\Delta^{A-1}})$;
4. $G_{\Delta,A}(x) \le \mathbf{1}(x \in [-\Delta^{-A/2}, \Delta^{-1} + \Delta^{-A/2}]) + c e^{-\Delta^{A-1}}$;
5. $\int_{\mathbb{R}} |\widehat{G}_{\Delta,A}(x)| \mathrm{d}x \le 2\Delta^{2A}$.

From the property (3), we get

$$\mathbf{1}(Y_j \in [u_j, u_j + \Delta_j^{-1}]) \le |G_{\Delta_j,A}(Y_j - u_j)|^2 (1 + c e^{-\Delta_j^{A-1}}). \tag{50}$$

Writing $G_{\Delta_j,A}$ in terms of its Fourier transform, we have by truncating the exponential at $\nu = \Delta_j^{10A}$ (this choice will be motivated by the estimate (57) below):

$$\begin{aligned} G_{\Delta_j,A}(x) &= \int_{\mathbb{R}} e^{2\pi \mathrm{i} \xi x} \widehat{G}_{\Delta_j,A}(\xi) \mathrm{d}\xi \\ &= \sum_{k \le \nu} \frac{(2\pi \mathrm{i} x)^k}{k!} \int_{\mathbb{R}} \xi^k \widehat{G}_{\Delta_j,A}(\xi) \mathrm{d}\xi + \mathrm{O}^\star\Big(\frac{(2\pi)^\nu x^\nu}{\nu!} \int_{\mathbb{R}} \xi^\nu \widehat{G}_{\Delta_j,A}(\xi) \mathrm{d}\xi\Big), \end{aligned} \tag{51}$$

where O* means that implicit constant is smaller than 1 in absolute value. The polynomial term in (51) is our definition of the polynomial $\mathcal{D}_{\Delta_j,A}(x)$ in (49). Since the Y_j is a sum over primes in $(\exp(e^{t_{j-1}}), \exp(e^{t_j})]$, it is clear that $\mathcal{D}_{\Delta_j,A}(Y_j - u_j)$ is a Dirichlet polynomial involving integers with prime factors in that interval and that its length is at most $\exp(2e^{t_j}\Delta_j^{10A})$. (The factor 2 comes from the fact that Y_j includes squares of primes.) It remains to estimate the error term. Since Equation (49) is trivial if $|Y_j - u_j| > \Delta_j^{-1}$, we assume without loss of generality that $|Y_j - u_j| \le \Delta_j^{-1}$. Therefore the error term is

$$\frac{(2\pi)^\nu}{\nu!}\int_{\mathbb{R}} \xi^\nu \widehat{G}_{\Delta_j,A}(\xi)\mathrm{d}\xi \le \frac{(2\pi)^\nu}{\nu!}\int_{\mathbb{R}} |\xi|^\nu |\widehat{G}_{\Delta_j,A}(\xi)|\mathrm{d}\xi \quad \le \frac{(2\pi)^\nu}{\nu!}\cdot 2\Delta_j^{2A(\nu+1)} \le \frac{(100)^\nu}{\nu^\nu}\Delta_j^{3A\nu}, \tag{52}$$

where we use properties (1) and (5) above. This is $e^{-\Delta^{4A}}$ for the choice $\nu = \Delta_j^{10A}$. Putting this back in (50) yields

$$\mathbf{1}(Y_j \in [u_j, u_j + \Delta_j^{-1}]) \le |\mathcal{D}_{\Delta_j,A}(Y_j - u_j) + \mathrm{O}^\star(e^{-\Delta_j^{4A}})|^2 \cdot (1 + ce^{-\Delta_j^{A-1}}).$$

The term $\mathrm{O}^\star(e^{-\Delta_j^{4A}})$ can be absorbed in the multiplicative error by adjusting c. The choice $A \ge 10$ ensures a decay much better than Gaussian. □

It follows from Equation (48) and Lemma 2.8 that

$$\mathbf{1}(B_\ell \cap C_\ell \cap \{S_{t_\ell} \in (w, w+1]\}) \le \sum_{\mathbf{u}\in\mathcal{I}} \prod_j |\mathcal{D}_{\Delta_j,A}(Y_j - u_j)|^2 (1 + ce^{-\Delta_j^{A-1}}). \tag{53}$$

We choose $A = 20$ for the rest of the proof. The product over j of $|\mathcal{D}_{\Delta_j,A}(Y_j - u_j)|^2$ is a Dirichlet polynomial of length at most

$$\exp\big(2\sum_{j=1}^{\ell} e^{t_j}\Delta_j^{10A}\big) \le \exp\Big(2e^{t_\ell}\Delta_\ell^{10A}\sum_{j=1}^{\ell}\Big(\frac{\log_{\ell-1} t}{\log_{j-1} t}\Big)^{\mathfrak{s}-10A}\Big) \le \exp(2e^{t_\ell}\Delta_\ell^{10A}), \tag{54}$$

since $\mathfrak{s} \ge 10^6 > 10A$ by the choice of $\mathfrak{s}$ in (12) and the choice $A = 20$. The mean-value theorem for Dirichlet polynomials, see Lemma A.3 (which applies by the assumption on ℓ), implies

$$\mathbf{E}\Big[\prod_j |\mathcal{D}_{\Delta_j,A}(Y_j - u_j)|^2\Big] = (1 + \mathrm{o}(1))\mathbf{E}\Big[\prod_j |\mathcal{D}_{\Delta_j,A}(\mathcal{Y}_j - u_j)|^2\Big], \tag{55}$$

where $(\mathcal{Y}_j, j \le \ell)$ are independent random variables of the form

$$\mathcal{Y}_j = \sum_{e^{t_{j-1}} < \log p \le e^{t_j}} \frac{\cos\theta_p}{p^{1/2}} + \frac{\cos^2\theta_p}{2p}, \tag{56}$$

and $(\theta_p, p$ primes) are independent random variables uniform on $[0, 2\pi]$. It remains to estimate $\mathbf{E}\Big[|\mathcal{D}_{\Delta_j,A}(\mathcal{Y}_j - u_j)|^2\Big]$ for each j.

Lemma 2.9. *With the above notation, we have for $j \le \ell$ and an absolute constant $c > 0$,*

$$\mathbf{E}\Big[|\mathcal{D}_{\Delta_j,A}(\mathcal{Y}_j - u_j)|^2\Big] \le \mathbf{P}(\mathcal{Y}_j - u_j \in [-\Delta_j^{-A/2}, \Delta_j^{-1} + \Delta_j^{-A/2}]) + ce^{-\Delta_j^{A-1}}.$$

Proof. The idea is to use the approximation with $G_{\Delta_j,A}$ in reverse. For this, it is necessary to re-introduce the error term in Equation (51), assuming it is small enough. On the event $|\mathcal{Y}_j - u_j| \le \Delta_j^{6A}$, the estimate (52) becomes

$$\frac{(2\pi)^\nu \Delta_j^{6A\nu}}{\nu!} \int_{\mathbb{R}} \xi^\nu \widehat{G}_{\Delta_j,A}(\xi) \mathrm{d}\xi \le \frac{(2\pi)^\nu}{\nu!} \cdot \Delta_j^{2A(4\nu+1)} \le \frac{(100)^\nu}{\nu^\nu} \Delta_j^{9A\nu}. \tag{57}$$

This is $e^{-\Delta_j^{4A}}$ for the choice $\nu = \Delta_j^{10A}$. On the event $|\mathcal{Y}_j - u_j| > \Delta_j^{6A}$, Cauchy-Schwarz inequality yields

$$\mathbf{E}\Big[|\mathcal{D}_{\Delta_j,A}(\mathcal{Y}_j - u_j)|^2 \mathbf{1}(|\mathcal{Y}_j - u_j| > \Delta_j^{6A})\Big] \le \mathbf{E}\Big[|\mathcal{D}_{\Delta_j,A}(\mathcal{Y}_j - u_j)|^4\Big]^{1/2} \cdot \mathbf{P}\big(|\mathcal{Y}_j - u_j| > \Delta_j^{6A})\big)^{1/2}.$$

The fourth moment of $\mathbf{E}[|\mathcal{D}_{\Delta_j,A}(\mathcal{Y}_j(h) - u_j)|^4]$ is bounded by

$$\mathbf{E}\Big[\Big(\sum_{\ell \le \Delta_j^{10A}} \frac{(2\pi)^\ell}{\ell!} 2\Delta_j^{2A(\ell+1)} (|\mathcal{Y}_j| + |u_j|)^\ell\Big)^4\Big] \ll \Delta_j^{2A}\, \mathbf{E}[\exp(9\pi \Delta_j^{2A}(|\mathcal{Y}_j| + 4\Delta_j))] \ll e^{\Delta_j^{5A}}, \tag{58}$$

where we used Equation (46) and the fact that $\mathbf{E}[e^{\lambda \mathcal{Y}_j}] \ll \exp(\lambda^2 \Delta_j)$ by Lemma A.5. The probability is bounded by Chernoff's inequality using the same lemma

$$\mathbf{P}\big(|\mathcal{Y}_j - u_j| > \Delta_j^{6A}\big) \ll \exp(-\frac{1}{2}\Delta_j^{6A}). \tag{59}$$

Equations (58) and (59) together imply

$$\mathbf{E}\Big[|\mathcal{D}_{\Delta_j,A}(\mathcal{Y}_j - u_j)|^2 \mathbf{1}(|\mathcal{Y}_j - u_j| > \Delta_j^{6A})\Big] \le e^{-\frac{1}{8}\Delta_j^{6A}}.$$

Altogether, we have shown

$$\mathbf{E}\Big[|\mathcal{D}_{\Delta_j,A}(\mathcal{Y}_j - u_j)|^2\Big] \le \mathbf{E}[|G_{\Delta_j,A}(\mathcal{Y}_j - u_j) + \mathrm{O}(e^{-\Delta_j^{4A}})|^2] + e^{-\frac{1}{8}\Delta_j^{6A}}.$$

Since $G_{\Delta_j,A}$ is in $[0,1]$ the error inside the expectation can be made additive. The statement of the lemma then follows from property (4) of the function $G_{\Delta_j,A}$. ☐

The proof of Lemma 2.6 can now be concluded.

Proof of Lemma 2.6. We write $\mathcal{Q}$ for $\mathcal{Q}(\frac{1}{2} + \mathrm{i}\tau)$. Following Equation (53), we have

$$\mathbb{E}\Big[|\mathcal{Q}|^2 \mathbf{1}(B_\ell \cap C_\ell \cap \{S_{t_\ell} \in (w, w+1]\})\Big] \ll \sum_{\underline{u} \in \mathcal{I}} \mathbb{E}\left[|\mathcal{Q}|^2 \prod_j |\mathcal{D}_{\Delta_j,A}(Y_j - u_j)|^2 (1 + ce^{-\Delta_j^{A-1}})\right].$$

The Dirichlet polynomials $\mathcal{D}_{\Delta_j,A}$ are supported on integers whose prime factors at most $\exp(e^{t_j})$, so the product features primes as large as $\exp(e^{t_\ell})$. As for $\mathcal{Q}$, it is supported on integers whose prime factors are at least $\exp(e^{t_\ell})$. Recall from (54) that the length of the product over j of $|\mathcal{D}_{\Delta_j,A}|^2$ is at most $\exp(2e^{t_\ell}(\mathfrak{s}\log_{\ell-1} t)^{10A}) \leq \exp(\frac{1}{100}e^t)$. By assumption the length of $\mathcal{Q}$ is less than $\exp(\frac{1}{100}e^t)$. Thus, Lemma A.4 applies and the expectation splits:

$$
\begin{aligned}
&\mathbb{E}\left[|\mathcal{Q}|^2 \prod_j |\mathcal{D}_{\Delta_j,A}(Y_j - u_j)|^2(1 + ce^{-\Delta_j^{A-1}})\right] \\
&\qquad \ll \mathbb{E}\left[|\mathcal{Q}|^2\right] \prod_{j=1}^{\ell} \mathbb{E}\left[\mathcal{D}_{\Delta_j,A}(Y_j - u_j)|^2(1 + ce^{-\Delta_j^{A-1}})\right] \\
&\qquad \ll \mathbb{E}\left[|\mathcal{Q}|^2\right] \prod_{j=1}^{\ell} \left(\mathbf{P}(\mathcal{Y}_j - u_j \in [-\Delta_j^{-A/2}, \Delta_j^{-1} + \Delta_j^{-A/2}]) + ce^{-\Delta_j^{A/2}}\right)
\end{aligned}
$$

which follows from applying Lemma 2.9.

Now notice that by a direct application of Berry-Esseen theorem, see Lemma A.6, we have for any $j \geq 2$

$$
\mathbf{P}(\mathcal{Y}_j - u_j \in [-\Delta_j^{-A/2}, \Delta_j^{-1} + \Delta_j^{-A/2}]) = \mathbf{P}(\mathcal{N}_j - u_j \in [-\Delta_j^{-A/2}, \Delta_j^{-1} + \Delta_j^{-A/2}]) + \mathrm{O}(e^{-ce^{t_{j-1}}}). \tag{60}
$$

where $\mathcal{N}_j$ is a Gaussian random variable of mean 0 with variance $\frac{1}{2}(t_j - t_{j-1}) + \mathrm{o}(1)$. For $j = 1$, we use the less accurate estimate in Lemma A.7:

$$
\mathbf{P}(\mathcal{Y}_1 - u_1 \in [-\Delta_1^{-A/2}, \Delta_1^{-1} + \Delta_1^{-A/2}]) \ll \mathbf{P}(\mathcal{N}_1 - u_1 \in [-\Delta_1^{-A/2}, \Delta_1^{-1} + \Delta_1^{-A/2}]).
$$

Since $|u_j| < 4\Delta_j$ by Equation (46), we have that $u_j \cdot \Delta_j^{-A/2}$ is very small, and therefore by using a Gaussian estimate, we get for all $j \geq 2$,

$$
\mathbf{P}(\mathcal{N}_j - u_j \in [-\Delta_j^{-A/2}, \Delta_j^{-1} + \Delta_j^{-A/2}]) = \mathbf{P}(\mathcal{N}_j - u_j \in [0, \Delta_j^{-1}])(1 + \mathrm{O}(\Delta_j)^{-A/4}).
$$

For $j = 1$, the corresponding estimate holds with $\ll$ instead of $=$. We also notice that the error term in (60) is much smaller than the probability and can be absorbed in the multiplicative error above. Therefore we have shown for $j \geq 2$ that

$$
\mathbf{P}(\mathcal{Y}_j - u_j \in [-\Delta_j^{-A/2}, \Delta_j^{-1} + \Delta_j^{-A/2}]) = \mathbf{P}(\mathcal{N}_j - u_j \in [0, \Delta_j^{-1}])(1 + \mathrm{O}(\Delta_j^{-A/4})),
$$

and for $j = 1$

$$
\mathbf{P}(\mathcal{Y}_1 - u_1 \in [-\Delta_1^{-A/2}, \Delta_1^{-1} + \Delta_1^{-A/2}]) \ll \mathbf{P}(\mathcal{N}_1 - u_1 \in [0, \Delta_1^{-1}])(1 + \mathrm{O}(\Delta_1^{-A/4})).
$$

It follows that (noticing again that the additive error from Lemma 2.9 can be made multiplicative),

$$
\begin{aligned}
&\mathbb{E}[|\mathcal{Q}(\tfrac{1}{2} + \mathrm{i}\tau)|^2 \mathbf{1}(B_\ell \cap C_\ell \cap \{S_{t_\ell} \in (w, w+1]\})] \\
&\qquad \ll \sum_{\mathbf{u} \in \mathcal{I}} \mathbb{E}\left[|\mathcal{Q}(\tfrac{1}{2} + \mathrm{i}\tau)|^2\right] \prod_{j=1}^{\ell} \mathbf{P}(\mathcal{N}_j \in [u_j, u_j + \Delta_j^{-1}])(1 + \mathrm{O}(\Delta_j^{-A/4})).
\end{aligned}
$$

It remains to re-express the events in terms of the partial sums of $\mathcal{N}_j$, exactly as we did in Equation (47) but in reverse. By the definition of $\mathcal{I}$ and the summability of Δ_j^{-1}, we conclude that

$$\mathbf{P}(\mathcal{N}_j \in [u_j, u_j + \Delta_j^{-1}]) \ll \mathbf{P}\Big(\sum_{j=1}^{\ell} \mathcal{N}_j \in (w-1, w+2]\Big).$$

Here, we dropped the intermediate restrictions on the partial sums that are no longer needed. The right side is $\ll \frac{1}{\sqrt{t_\ell}} e^{-w^2/t_\ell}$ as claimed.

□

2.5 Proof of Lemma 2.4 and Lemma 2.7

We prove Equation (38). The proof of Lemma 2.4 and of Equation (37) are similar and simpler. The proof follows closely the one of Lemma 2.6 with an additional tool from [ABR20]. Given $\ell \geq 1$, a Dirichlet polynomial $\mathcal{Q}$ is said to be degree-e^{t_ℓ} well-factorable if it can be expressed as

$$\prod_{1 \leq \lambda \leq \ell} \mathcal{Q}_\lambda(s), \quad \text{where} \quad \mathcal{Q}_\lambda(s) = \sum_{\substack{p|m \implies \log p \in (e^{t_{\lambda-1}}, e^{t_\lambda}] \\ \Omega_\lambda(m) \leq 10(t_\lambda - t_{\lambda-1})^{10^4}}} \frac{\gamma(m)}{m^s},$$

and γ are arbitrary coefficients such that $|\gamma(m)| \leq \exp(\frac{1}{500} e^t)$ for every $m \geq 1$. We need the following twisted fourth moment estimate.

Lemma 2.10 (Lemma 9 in [ABR20])**.** *Let* $\ell \geq 0$ *be such that* $\exp(10^6 (t_{\ell+1} - t_\ell)^{10^5} e^{t_{\ell+1}}) \leq \exp(\frac{1}{100} e^t)$*. Let* $\mathcal{Q}$ *be a degree-*$e^{t_{\ell+1}}$ *well-factorable Dirichlet polynomial. Then, we have*

$$\mathbf{E}\Big[|\zeta \mathcal{M}_1 \cdots \mathcal{M}_{\ell+1}|^4 \cdot |\mathcal{Q}(1/2 + \mathrm{i}\tau)|^2 (1/2 + \mathrm{i}\tau)\Big] \ll e^{4(t - t_{\ell+1})} \, \mathbf{E}\Big[|\mathcal{Q}(1/2 + \mathrm{i}\tau)|^2\Big].$$

Proof of Lemma 2.7. We proceed as in the proof of Lemma 2.6 by approximating the indicator function by a Dirichlet polynomial. More precisely, using Equations (48) and (49), the left-hand side of (38) becomes

$$\ll \sum_{\mathbf{u} \in \mathcal{I}} \mathbf{E}\Big[|\zeta \mathcal{M}_1 \cdots \mathcal{M}_{\ell+1}|^4 |\mathcal{Q}_\ell|^2 \prod_j \mathcal{D}_{\Delta_j, A}(Y_j - u_j)|^2\Big].$$

We choose $A = 20$. The polynomial $Q = Q_\ell \prod_j \mathcal{D}_{\Delta_j, A}(Y_j - u_j)$ is well-factorable, and $\mathcal{Q}_\ell$ is as defined in the statement of Lemma 2.7. Since the coefficients of $\mathcal{D}_{\Delta_j, A}$ are bounded by $\Delta_j^{2A(\nu+1)}$, the coefficients of $\mathcal{Q}$ are bounded by $\exp(\frac{1}{500} e^t)$. Moreover, its length is

$$\leq \exp(10 e^{t_{\ell+1}} (t_{\ell+1} - t_\ell)^{10^4}) \cdot \exp(2 e^{t_\ell} \Delta_\ell^{200}) < \exp(e^t / 100),$$

since $\mathfrak{s} \geq 10^6$ and by the assumption on ℓ. This implies by Lemma 2.10 that the above is

$$\ll e^{4(t - t_{\ell+1})} \sum_{\mathbf{u} \in \mathcal{I}} \mathbf{E}\Big[|\mathcal{Q}_\ell|^2 \prod_j \mathcal{D}_{\Delta_j, A}(Y_j - u_j)|^2\Big].$$

The expectation splits by Lemma A.4. It remains to proceed as before from Equation (55) to get Equation (38). □

3 Proofs of Corollaries

3.1 Proof of Corollary 1.2

Consider the CDF of the random variable $\log|\zeta(1/2+\mathrm{i}\tau)|$, i.e., $F(V) = \mathbf{P}(\log|\zeta(1/2+\mathrm{i}\tau)| \leq V)$. Write for short

$$S(V) = \mathbf{P}(\log|\zeta(1/2+\mathrm{i}\tau)| > V).$$

Recall that τ is distributed uniformly on $[T, 2T]$, and we write $t = \log\log T$. Clearly, the moments (cf. Equation (5)) can be written as

$$M_\beta = \int_{-\infty}^{+\infty} e^{\beta V} \mathrm{d}F(V).$$

Integration by parts yields

$$M_\beta = -e^{\beta V} S(V)\Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} \beta e^{\beta V} S(V)\mathrm{d}V. \tag{61}$$

Since $S(V)$ is bounded by one, the boundary term at $-\infty$ is zero. Moreover, Markov's inequality with the fourth moment of zeta [Theorem B [Ing26]] gives

$$S(V) \leq \frac{1}{2\pi^2} e^{4t} e^{-4V}. \tag{62}$$

In particular, this implies that the boundary term at $+\infty$ is zero for $\beta < 4$. The contribution to negative V's in the integral in Equation (61) is also negligible since

$$\int_{-\infty}^{0} \beta e^{\beta V} S(V)\mathrm{d}V \leq \int_{-\infty}^{0} \beta e^{\beta V}\mathrm{d}V = 1. \tag{63}$$

It remains to estimate $\int_0^{+\infty} \beta e^{\beta V} S(V)\mathrm{d}V$. Consider β_- and β_+ such that $0 < \beta_- < \beta < \beta_+ < 4$. These have to be chosen close enough to 0 and to 4 respectively. It turns out that the choices

$$\beta_- = \frac{\beta}{4}$$
$$\beta_+ = \beta + \frac{3}{4}(4-\beta) = 3 + \frac{\beta}{4}$$

are adequate. The dominant contribution to the β-moment comes from the interval $[\frac{\beta_-}{2}t, \frac{\beta_+}{2}t]$. Indeed, by Theorem 1.1, we have

$$\int_{\frac{\beta_-}{2}t}^{\frac{\beta_+}{2}t} e^{\beta V} S(V)\mathrm{d}V \ll \int_{\frac{\beta_-}{2}t}^{\frac{\beta_+}{2}t} e^{\beta V} \frac{e^{-V^2/t}}{\sqrt{t}}\mathrm{d}V = e^{\frac{\beta^2}{4}t} \int_{\frac{\beta_-}{2}t}^{\frac{\beta_+}{2}t} \frac{e^{-(\frac{\beta}{2}t - V)^2/t}}{\sqrt{t}}\mathrm{d}V \ll e^{\frac{\beta^2}{4}t}.$$

The contribution of the intervals $[0, \frac{\beta_-}{2}t]$ is less since it is smaller than

$$\int_0^{\frac{\beta_-}{2}t} \beta e^{\beta V}\mathrm{d}V \leq e^{\frac{\beta^2 t}{8}}, \tag{64}$$

by the choice of β_-. For the interval $[\frac{\beta_+}{2}t, \infty]$, we use the bound (62) to get that the contribution is

$$\leq e^{4t} \int_{\frac{\beta_+}{2}t}^{\infty} e^{(\beta-4)V} \mathrm{d}V \leq \frac{1}{4-\beta} e^{t(\frac{\beta\beta_+}{2} - 2\beta_+ + 4)}.$$

This is $\ll e^{\beta^2 t/4}$ by the choice of β_+.

3.2 Proof of Corollary 1.3

We will require the following discretization result of [FGH07]. Effectively, this shows that the maximum of concern in Corollary 1.3 can be restricted to those h lying $1/\log T$ apart. Corollary 1.3 may also be deduced from a more general discretization result of [AOR21], applicable to Dirichlet polynomials.

Lemma 3.1 (Lemma 2.2 of [FGH07]). *Let $t^\star$ be such that $|\zeta(1/2+\mathrm{i}t^\star)| = \max_{t\in[T,2T]} |\zeta(1/2+\mathrm{i}t)|$. There is an absolute constant $A > 0$ such that if $|t - t^\star| < A/\log T$ then $2|\zeta(1/2+\mathrm{i}t)| > |\zeta(1/2+\mathrm{i}t^\star)|$.*

Thus, as u ranges over a window of size $A/\log T$, the value of $|\zeta(1/2+\mathrm{i}u)|$ is close to the maximum within the window. Hence, we deduce via a union bound that, for some universal positive constant $C > 0$,

$$\mathbf{P}\Big(\max_{|h|\leq \log^\theta T} |\zeta(1/2+\mathrm{i}u+\mathrm{i}h)| > e^V\Big) \leq e^{(1+\theta)t} \cdot \mathbf{P}\Big(|\zeta(1/2+\mathrm{i}u)| > \frac{1}{C} e^V\Big). \quad (65)$$

Corollary 1.3 now follows by setting $V = \sqrt{1+\theta}t - \frac{1}{4\sqrt{1+\theta}} \log t + y$ (for $y = \mathrm{o}(t/\log t)$, $\theta \in [0,3)$), and applying Theorem 1.1.

3.3 Proof of Corollary 1.4

Case $\beta \geq 0$: We write

$$\mathcal{Z}_\beta(\tau) = \frac{1}{2e^{\theta t}} \int_{|h|\leq e^{\theta t}} |\zeta(1/2+\mathrm{i}\tau+\mathrm{i}h)|^\beta \mathrm{d}h,$$

i.e., the left-hand side of Equation (8) normalized by $2e^{\theta t}$ and with the identification $t = \log\log T$. The moment $\mathcal{Z}_\beta$ is a random variable dependent on τ. From now on, we use the probabilistic convention and drop the dependence on τ from the notation. Consider also the (normalized) Lebesgue measure of high points in the interval $[-e^{\theta t}, e^{\theta t}]$ around τ:

$$\mathcal{S}(V) = \frac{1}{2e^{\theta t}} \mathrm{meas}\{|h| \leq e^{\theta t} : \log|\zeta(1/2+\mathrm{i}\tau+\mathrm{i}h)| > V\}.$$

Proceeding as in the proof of Corollary 1.2, we have by integration by parts:

$$\mathcal{Z}_\beta = -e^{\beta V}\mathcal{S}(V)\Big|_{-\infty}^{+\infty} + \beta \int_{-\infty}^{\infty} e^{\beta V}\mathcal{S}(V)\mathrm{d}V.$$

Again, since $\mathcal{S}(V) \leq 1$ for all V, we have that the boundary term at $V = -\infty$ is 0.

For $V = +\infty$, it is necessary to restrict the estimate to a good event. Define

$$E = \left\{ \max_{|h| \leq e^{\theta t}} \log |\zeta(1/2 + \mathrm{i}\tau + \mathrm{i}h)| \leq m(t) + A \right\}, \tag{66}$$

where

$$m(t) = \sqrt{1+\theta} t - \frac{1}{4\sqrt{1+\theta}} \log t = \frac{\beta_c}{2} t - \frac{1}{2\beta_c} \log t, \tag{67}$$

and $\beta_c = 2\sqrt{1+\theta}$. In view of Corollary 1.3 with the choice $y = A$ (and since $\theta \in [0, 3)$ by assumption), the probability of E^c is

$$\mathbf{P}(E^c) \ll e^{-\beta_c A}. \tag{68}$$

This handles the upper limit $V = +\infty$.

On the event E, there are clearly no values of V beyond $m(t) + A$. Moreover, as in the proof of Corollary 1.2, the contribution of negative values is of order one (cf. Equation (63)). Finally, the bound (64) still holds. The problem is therefore reduced to finding a good event on which to bound

$$\int_{\beta t/8}^{m(t)+A} e^{\beta V} \mathcal{S}(V) \mathrm{d}V. \tag{69}$$

The idea now is that $\mathcal{S}(V)$ should behave like $e^{-V^2/t - 1/2 \log t}$, thanks to Theorem 1.1. In particular, as can be seen easily in the proof, the dominant contribution to the integral should come from V's around $\beta t/2$. Hence, the specifics of the interval of integration do not matter much as long as it contains this optimizer. The main technical difficulty in implementing this idea is to control $\mathcal{S}(V)$ on a range of V simultaneously.

Consider $(V_j, 1 \leq j \leq J)$ the set of V's in $[\frac{\beta}{8}t, m(t) + A] \cap \sqrt{t}\mathbb{Z}$, and additionally define $V_0 = V_1 - \sqrt{t}$ and $V_{J+1} = V_J + \sqrt{t}$. (The choice of the mesh size $\sqrt{t}$ is informed by the typical fluctuation of $\log |\zeta|$.) Define

$$I_j = \int_{V_j}^{V_{j+1}} e^{\beta V} \mathcal{S}(V) \mathrm{d}V, \quad 0 \leq j \leq J.$$

Consider the events

$$E_j = \left\{ I_j \leq a_j \int_{V_j}^{V_{j+1}} e^{\beta V} \frac{e^{-V^2/t}}{\sqrt{t}} \mathrm{d}V \right\}, \tag{70}$$

for a collection of a_j's to be fixed later.

We have $\mathbf{P}(E_j^c) \ll a_j^{-1}$, since by linearity and Theorem 1.1

$$\mathbf{E}[I_j] = \int_{V_j}^{V_{j+1}} e^{\beta V} \mathbf{E}[\mathcal{S}(V)] \mathrm{d}V \ll \int_{V_j}^{V_{j+1}} e^{\beta V} \frac{e^{-V^2/t}}{\sqrt{t}} \mathrm{d}V. \tag{71}$$

The good event to consider is

$$G = E \cap \left(\bigcap_j E_j \right),$$

so that by (68)

$$\mathbf{P}(G^c) \ll \sum_j a_j^{-1} + e^{-\beta_c A}. \tag{72}$$

On the event G, we have

$$\int_{\beta t/8}^{m(t)+A} e^{\beta V} \mathcal{S}(V) \mathrm{d}V \leq e^{\beta^2 t/4} \sum_j a_j \int_{V_j}^{V_{j+1}} \frac{e^{-(\frac{\beta}{2}t - V)^2/t}}{\sqrt{t}} \mathrm{d}V. \tag{73}$$

Since the quadratic form is maximized at $\beta t/2$, we pick for a_j:

$$a_j = A \cdot \begin{cases} \left(\frac{\beta}{2}\sqrt{t} - \frac{V_j}{\sqrt{t}}\right)^2 + \frac{1}{100} & \text{if } V_j > \beta t/2, \\ \left(\frac{\beta}{2}\sqrt{t} - \frac{V_{j+1}}{\sqrt{t}}\right)^2 + \frac{1}{100} & \text{if } V_j \leq \beta t/2 \text{ and } V_j < V_{j+1} \leq \beta t/2 \\ \frac{1}{100} & \text{if } V_j \leq \beta t/2 \text{ and } V_{j+1} > \beta t/2. \end{cases}$$

(The term 1/100 is simply there to make sure a_j is bounded away from 0.) This choice ensures that $a_j \leq A\left(\frac{1}{100} + (\frac{\beta}{2}\sqrt{t} - \frac{V}{\sqrt{t}})^2\right)$ for $V \in [V_j, V_{j+1}]$.

Thus, on one hand from Equation (73), we have on G

$$\begin{aligned} \int_{\beta t/8}^{m(t)+A} e^{\beta V} \mathcal{S}(V) \mathrm{d}V &\leq A e^{\beta^2 t/4} \int_{\beta t/8}^{m(t)+A} \left(\frac{1}{100} + \left(\frac{\beta}{2}\sqrt{t} - \frac{V}{\sqrt{t}} \right)^2 \right) \cdot \frac{e^{-(\frac{\beta}{2}t - V)^2/t}}{\sqrt{t}} \mathrm{d}V \\ &\leq A e^{\beta^2 t/4} \int_{\frac{(\beta-\beta_c)}{2}\sqrt{t}+\mathrm{o}(1)}^{\frac{3\beta}{8}\sqrt{t}} \left(\frac{1}{100} + u^2 \right) e^{-u^2} \mathrm{d}u \\ &\leq A e^{\beta^2 t/4}, \end{aligned}$$

where the last bound follows by integrating over the whole line. On the other hand, from Equation (72) the probability of G^c is

$$\mathbf{P}(G^c) \ll \sum_j a_j^{-1} + e^{-\beta_c A} \ll \frac{1}{A}.$$

The a_j's are summable since $V_j \in \sqrt{t}\mathbb{Z}$. This proves Equation (8).

Case $\beta > \beta_c$: We can use a reduction as in the previous case. We use the same event E in (66) for the maximum. For a lower bound on the values of V, we take $\beta_c t/4$ since

$$\int_0^{\beta_c t/4} e^{\beta V} \mathcal{S}(V) \mathrm{d}V \leq e^{\frac{\beta_c}{4}\beta t},$$

which is much smaller than the the desired bound. Therefore, it remains to estimate

$$\int_{\beta_c t/4}^{m(t)+A} e^{\beta V} \mathcal{S}(V) \mathrm{d}V. \tag{74}$$

The partitioning of the interval of integration is more delicate as it is close to the level of the maximum. A mesh size of 1 instead of $\sqrt{t}$ is needed. More precisely, we take $(V_j, 1 \leq$

$j \leq J$) to be $[\frac{\beta_c}{4}t, m(t)+A] \cap \mathbb{Z}$. The events E_j are defined as in (70). As before, we take $G = E \cap (\bigcap_j E_j)$. The difference here is that the optimizer lies outside the interval, so the bound can be sharpened. On the event G, the above becomes

$$\leq \sum_j a_j \int_{V_j}^{V_{j+1}} e^{\beta V} \frac{e^{-V^2/t}}{\sqrt{t}} \mathrm{d}V.$$

The change of variable $V = m(t) + y$ yields (with $y_j = V_j - m(t)$)

$$\begin{aligned} e^{\beta m(t)} \sum_j a_j \int_{y_j}^{y_{j+1}} e^{\beta y} \frac{e^{-m(t)^2/t}}{\sqrt{t}} e^{-\frac{2m(t)y}{t}} e^{-y^2/t} \mathrm{d}y \\ \leq e^{\beta m(t)-(1+\theta)t} \sum_j a_j \int_{y_j}^{y_{j+1}} e^{(\beta-\beta_c)y} e^{y\frac{(\log t)^2}{4\beta_c^2 t}} \mathrm{d}y, \end{aligned} \tag{75}$$

since $m(t)^2 = (1+\theta)t - \frac{1}{2}\log t + \frac{(\log t)^2}{4\beta_c^2}$ and $e^{-y^2/t} \leq 1$. We pick $a_j = A(1+y_j^2)$ if y_j is positive, and $a_j = A(1+y_{j+1}^2)$ if y_{j+1} is negative. If $y_j < 0 < y_{j+1}$ then set $a_j = A$. This choice ensures that $a_j \leq A(2+y^2)$ for $y \in [y_j, y_{j+1}]$, the term 2 taking care of the values close to 0.

This gives that Equation (75) is bounded by

$$\begin{aligned} &\leq A e^{\beta m(t)-(1+\theta)t} \int_{-\infty}^{A} (2+y^2) e^{(\beta-\beta_c)y} e^{y\frac{(\log t)^2}{4\beta_c^2 t}} \mathrm{d}y \\ &\leq \left(\tfrac{2A}{(\beta-\beta_c)^3} + \tfrac{A(A^2+2)}{\beta-\beta_c} \right) e^{(\beta-\beta_c+1)A} \cdot e^{\beta m(t)-(1+\theta)t} \end{aligned}$$

since $e^{y\frac{(\log t)^2}{4\beta_c^2 t}} \leq e^A$, and by direct integration of $(2+y^2)e^{(\beta-\beta_c)y}$. The probability of G^c is then

$$\mathbf{P}(G^c) = e^{-\beta_c A} + \sum_j a_j^{-1} \ll \frac{1}{A}.$$

This proves the corollary in the case $\beta > \beta_c$.

Remark (Case $\beta = \beta_c$)**.** Since it is possible to improve the bound (8) in the range $\beta > \beta_c$, one might hope to do the same at $\beta = \beta_c$. This is possible in the case $\theta = 0$, as discussed in the next section, but it is not expected to be possible for $\theta > 0$. Indeed, in this range of θ, the above proof should be optimal. In fact, Equation (9) would become (dropping the a_j's for simplicity)

$$e^{\beta_c m(t)} \int_{-\infty}^{A} e^{\beta y} \frac{e^{-m(t)^2/t}}{\sqrt{t}} e^{-\frac{2m(t)y}{t}} e^{-y^2/t} \mathrm{d}t \leq e^A \cdot e^{\frac{\beta_c^2}{4}t} \int_{-\infty}^{A} \frac{e^{-y^2/t}}{\sqrt{t}} \mathrm{d}y. \tag{76}$$

This is because $e^{-(1+\theta)t} = e^{-\frac{\beta_c^2}{4}t}$ and $t^{-\frac{\beta}{2\beta_c}} = t^{-1/2}$. The integral is now finite, so one recovers the bound (8) up to a factor of order one.

4 Relation to Theorem 1.5 for $\theta = 0$

We briefly explain an alternative approach to proving a sharp upper bound to the β_c-moment in the case $\theta = 0$. It is based on the measure of the level sets in the spirit of the proof of Corollary 1.4.

The deterministic level of the maximum is now by Equation (7)

$$m(t) = t - \frac{3}{4}\log t = \frac{\beta_c}{2}t - \frac{3}{2\beta c}\log t.$$

There is a factor 3 in the logarithmic correction and not 1 as in (67). The important observation is that the typical measure of the level sets $\mathcal{S}(m(t)+y)$ is no longer $e^{-t}e^{-2y}e^{-y^2/t}$ as for the case $\theta > 0$. In fact, the proof of (7) in [ABR20] also shows that

$$\mathcal{S}(m(t) + y) \leq Ae^{-t} \cdot |y|e^{-2y}e^{-y^2/2t}, \quad |y| = \mathrm{o}(t), \tag{77}$$

except on an event of probability A. This is what is expected from the study of the extreme values of log-correlated processes, see for example Theorem 1.1 and Lemma 4.2 in [CHL19]. We explain how the additional y in the decay is responsible for the extra $1/\sqrt{t}$ factor in the size of the moment. The integral (74) with $\beta = \beta_c = 2$ becomes

$$\begin{aligned}
\int_{t/2}^{m(t)+A} e^{2V}\mathcal{S}(V)\mathrm{d}V &\leq A\frac{e^{2t}}{t^{3/2}}\int_{-t/2+\frac{3}{4}\log t}^{A} e^{-t}|y|e^{-y^2/t}\mathrm{d}y \\
&= A\frac{e^{t}}{t^{1/2}}\int_{-t/2+\frac{3}{4}\log t}^{A} \frac{|y|}{\sqrt{t}}\frac{e^{-y^2/t}}{\sqrt{t}}\mathrm{d}y \\
&= A\frac{e^{t}}{t^{1/2}}\int_{-\sqrt{t}/2+\mathrm{o}(1)}^{A/\sqrt{t}} |u|e^{-u^2}\mathrm{d}u.
\end{aligned} \tag{78}$$

The last integral is of order one. At criticality, there is now an extra factor $1/\sqrt{t}$ coming from $t^{3/2}$ that is left, thereby giving the overall magnitude of $\frac{e^t}{t^{1/2}}$ for the moment. It is also important to observe that, because of the $\sqrt{t}$-normalization in the integral, it is not necessary to know the level of the maximum up to order one as in Equation (7).

A Appendix

The appendix gathers known results on moments of Dirichlet polynomials and probability estimates of random models.

A.1 Moments of Dirichlet Polynomials

Lemma A.1. *Let $\widetilde{S}_j$ as in Equation (14). For any integers $j \leq k$ and $2q \leq e^{t-k}$, we have*

$$\mathbf{E}[|\widetilde{S}_k - \widetilde{S}_j|^{2q}] \ll q!(k-j+1)^q.$$

Proof. This is the content of [Sou09, Lemma 3]. □

With the choice $q = \lceil \frac{V^2}{k-j+1} \rceil$, Markov's inequality, Lemma A.1 and Stirling's formula imply

$$\mathbf{P}\Big(|\widetilde{S}_k - \widetilde{S}_j| > V \Big) \ll \frac{V+1}{(k-j)^{1/2}} \exp\Big(- \frac{V^2}{k-j+1} \Big). \tag{79}$$

Lemma 16 of [ABR20]] gives a more precise estimate for the moments of the real part S_j.

Lemma A.2. *For any integers $t/2 \le j < k$ and $2q \le e^{t-k}$ we have*

$$\mathbf{E}[|S_k - S_j|^{2q}] \ll \frac{(2q)!}{2^q q!} \left(\frac{k-j}{2} \right)^q.$$

Moreover, there exists $C > 0$ such that for any $j < k$, and $2q \le e^{t-k}$ such that

$$\mathbf{E}[|S_k - S_j|^{2q}] \ll \sqrt{q} \frac{(2q)!}{2^q q!} \left(\frac{k-j+C}{2} \right)^q. \tag{80}$$

As in Equation (79), one gets a Gaussian decay from Lemma A.2 for the choice $q = \lceil \frac{V^2}{2(k-j+1)} \rceil$

$$\mathbf{P}\left(|S_k - S_j| > V \right) \ll e^{-\frac{V^2}{k-j}}, \tag{81}$$

when $j > t/2$ and $V^2 \le \frac{k-j}{2} e^{t-k}$.

We now explain the link between Dirichlet polynomials and the random model (55). We consider the following general setup. Let $(\theta_p, p \text{ prime})$ be a sequence of IID random variables, uniformly distributed on $[0, 2\pi]$. For an integer n with prime factorization $n = p_1^{\alpha_1} \dots p_k^{\alpha_k}$ with $p_1, \dots, p_k$ all distinct, define the random variable

$$Z_n = \prod_{j=1}^{k} \exp(\mathrm{i}\alpha_j \theta_{p_j}).$$

By construction, we have the orthogonality relation $\mathbf{E}[Z_n \overline{Z}_m] = \mathbf{1}_{n=m}$. Therefore, for an arbitrary sequence $a(n)$ of complex numbers, the following holds

$$\sum_{n \le N} |a(n)|^2 = \mathbf{E}\Big[\Big| \sum_{n \le N} a(n) Z_n \Big|^2 \Big].$$

The expectation for the random variable is directly related to the mean-value of the square of Dirichlet polynomial, see [MV07, Corollary 3].

Lemma A.3. *We have*

$$\mathbf{E}\Big[\Big| \sum_{n \le N} a(n) n^{\mathrm{i}\tau} \Big|^2 \Big] = \Big(1 + \mathrm{O}\left(\frac{N}{T} \right) \Big) \sum_{n \le N} |a(n)|^2 = \Big(1 + \mathrm{O}\left(\frac{N}{T} \right) \Big) \mathbf{E}\Big[\Big| \sum_{n \le N} a(n) Z_n \Big|^2 \Big].$$

A direct consequence of Lemma A.3 is the splitting of the expectation for Dirichlet polynomials involving different range of primes, see for example [ABR20, Lemma 14].

Lemma A.4. *Let*

$$A(s) = \sum_{\substack{n \le N \\ p|n \implies p \le w}} \frac{a(n)}{n^s} \text{ and } B(s) = \sum_{\substack{n \le N \\ p|n \implies p > w}} \frac{b(n)}{n^s}$$

be two Dirichlet polynomials with $N \le T^{1/4}$. Then, we have

$$\mathbf{E}[|A(\tfrac{1}{2} + \mathrm{i}\tau)|^2 \, |B(\tfrac{1}{2} + \mathrm{i}\tau)|^2] = (1 + \mathrm{O}(T^{-1/2})) \mathbf{E}[|A(\tfrac{1}{2} + \mathrm{i}\tau)|^2] \, \mathbf{E}[|B(\tfrac{1}{2} + \mathrm{i}\tau)|^2].$$

A.2 Estimates for the random model

Recall the definition of the random model in Equation (55).

$$\mathcal{Y}_j = \sum_{e^{t_{j-1}} < \log p \le e^{t_j}} \frac{\cos\theta_p}{p^{1/2}} + \frac{\cos^2\theta_p}{2p}. \tag{82}$$

The moment generating function is easily estimated using the independence between the θ_p's.

Lemma A.5. *For $\lambda < \exp(\frac{1}{2}e^{t_j})$, we have*

$$\mathbf{E}[\exp(\lambda\mathcal{Y}_j)] \ll \exp\Big(\frac{\lambda^2}{4}(t_j - t_{j-1})\Big).$$

Proof. See for example [ABR20, Lemma 15]. □

The comparison between the random model and the Gaussian model can be made more precise at the level of the probabilities. A version was proved in [ABH17, Proposition 2.11] using a Berry-Esseen estimate. See also [ABR20, Lemma 20].

Lemma A.6. *For $j \ge 2$, let $\mathcal{N}_j$ be a Gaussian random variable of mean 0 and variance $\frac{1}{2}(t_j - t_{j-1})$. There exists a constant $c > 0$ such that, for any interval A and $j \ge 2$,*

$$\mathbf{P}\Big(\mathcal{Y}_j \in A\Big) = \mathbf{P}\Big(\mathcal{N}_j \in A\Big) + \mathrm{O}(e^{-ce^{j/2}}).$$

In the case $j = 1$ above, the variable $\mathcal{Y}_1$ is not asymptotically Gaussian because of the small primes. Nevertheless, the following estimate holds by a saddle-point method [ABR20, Lemma 18].

Lemma A.7. *Let $|v| \le 100r$. Then, for $r > 1000$ and for all $\Delta \ge 1$, we have*

$$\mathbf{P}(\mathcal{Y}_1 \in [v, v + \Delta^{-1}]) \asymp \frac{1}{\Delta} \cdot \frac{1}{\sqrt{r}} \exp\Big(-\frac{v^2}{r}\Big).$$

References

[AAB+21] E. Amzallag, L.-P. Arguin, E. Bailey, K. Hui, and R. Rao. Evidence of Random Matrix Corrections for the Large Deviations of Selberg's Central Limit Theorem. Exp. Math., 0(0):1–13, 2021.

[ABB+19] L.-P. Arguin, D. Belius, P. Bourgade, M. Radziwiłł, and K. Soundararajan. Maximum of the Riemann zeta function on a short interval of the critical line. Comm. Pure Appl. Math., 72(3):500–535, 2019.

[ABH17] Louis-Pierre Arguin, David Belius, and Adam J Harper. Maxima of a randomized Riemann zeta function, and branching random walks. Ann. Appl. Probab., 27(1):178–215, 2017.

[ABR20] L.-P. Arguin, P. Bourgade, and M. Radziwiłł. The Fyodorov-Hiary-Keating Conjecture. I. Preprint arXiv:2007.00988, 2020.

[ABR23] L.-P. Arguin, P. Bourgade, and M. Radziwiłł. The Fyodorov-Hiary-Keating Conjecture. II. Preprint arXiv:2307.00982, 2023.

[ADH21] L.-P. Arguin, G. Dubach, and L. Hartung. Maxima of a Random Model of the Riemann Zeta Function over Intervals of Varying Length. to appear in Ann. Inst. Henri Poincaré Probab. Stat., 2021.

[AOR21] L.-P. Arguin, F. Ouimet, and M. Radziwiłł. Moments of the Riemann zeta function on short intervals of the critical line. Ann. Prob., 49(6):3106–3141, 2021.

[BH15] Anton Bovier and Lisa Hartung. Variable speed branching Brownian motion 1. Extremal processes in the weak correlation regime. ALEA Lat. Am. J. Probab. Math. Stat., 12(1):261–291, 2015.

[BK22] E. C. Bailey and J. P. Keating. Maxima of log-correlated fields: some recent developments. J. Phys. A, 55(5):053001, jan 2022.

[BKL02] Anton Bovier, Irina Kurkova, and Matthias Löwe. Fluctuations of the free energy in the REM and the p-spin SK models. Ann. Probab., 30(2):605–651, 2002.

[Bov06] Anton Bovier. Statistical mechanics of disordered systems, volume 18 of Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2006. A mathematical perspective.

[Bov17] Anton Bovier. Gaussian processes on trees: From spin glasses to branching Brownian motion, volume 163. Cambridge University Press, 2017.

[CHL19] Aser Cortines, Lisa Hartung, and Oren Louidor. The structure of extreme level sets in branching Brownian motion. Ann. Probab., 47(4):2257–2302, 2019.

[DBMN19] Martina Dal Borgo, Pierre-Loïc Méliot, and Ashkan Nikeghbali. Local limit theorems and mod-ϕ convergence. ALEA Lat. Am. J. Probab. Math. Stat., 16(1):817–853, 2019.

[FGH07] D. W. Farmer, S. M. Gonek, and C. P. Hughes. The maximum size of L-functions. J. Reine Angew. Math., 609:215–236, 2007.

[FHK12] Y. V. Fyodorov, G. A Hiary, and J. P. Keating. Freezing transition, characteristic polynomials of random matrices, and the Riemann zeta function. Phys. Rev. Lett., 108(17):170601, 2012.

[FK14] Y. V. Fyodorov and J. P. Keating. Freezing transitions and extreme values: random matrix theory, and disordered landscapes. Philos. Trans. Roy. Soc. A, 372(2007):20120503, 2014.

[FMN16] V. Féray, P.-L. Méliot, and A. Nikeghbali. Mod-ϕ convergence. Springer Briefs in Probability and Mathematical Statistics. Springer, Cham, 2016. Normality zones and precise deviations.

[Har13] A. J. Harper. Sharp conditional bounds for moments of the Riemann zeta function. Preprint arXiv:1305.4618, 2013.

[Har19] A. J. Harper. On the partition function of the Riemann zeta function, and the Fyodorov–Hiary–Keating conjecture. Preprint arXiv:1906.05783, 2019.

[HKO01] C. P. Hughes, J. P. Keating, and N. O'Connell. On the characteristic polynomial of a random unitary matrix. Comm. Math. Phys., 220(2):429–451, 2001.

[HL18] G. H. Hardy and J. E. Littlewood. Contributions to the theory of the Riemann zeta-function and the theory of the distribution of primes. Acta Math., 41:119–196, 1918.

[HRS19] W. Heap, M. Radziwiłł, and K. Soundararajan. Sharp upper bounds for fractional moments of the Riemann zeta function. Q. J. Math., 70(4):1387–1396, 2019.

[HS20] Winston Heap and Kannan Soundararajan. Lower bounds for moments of zeta and l-functions revisited. arXiv preprint arXiv:2007.13154, 2020.

[Ing26] A. E. Ingham. Mean-value theorems in the theory of the Riemann zeta-function. Proc. Lond. Math. Soc., 2(1):273–300, 1926.

[Ino19] Shōta Inoue. On the logarithm of the Riemann zeta-function and its iterated integrals. arXiv preprint arXiv:1909.03643, 2019.

[KS00] J. P. Keating and N. C. Snaith. Random matrix theory and $\zeta(1/2 + it)$. Communications in Mathematical Physics, 214(1):57–89, 2000.

[MV07] Hugh L. Montgomery and Robert C. Vaughan. Multiplicative number theory. I. Classical theory, volume 97 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2007.

[Naj18] J. Najnudel. On the extreme values of the Riemann zeta function on random intervals of the critical line. Probab. Theory Related Fields, 172(1-2):387–452, 2018.

[Pow18] Ellen Powell. Critical Gaussian chaos: convergence and uniqueness in the derivative normalisation. Electronic Journal of Probability, 23(none):1 – 26, 2018.

[Rad11] M. Radziwiłł. Large deviations in Selberg's central limit theorem. Preprint arXiv:1108.5092, 2011.

[RS17] Maksym Radziwiłł and Kannan Soundararajan. Selberg's central limit theorem for $\log|\zeta(1/2 + it)|$. Enseign. Math., 63(1-2):1–19, 2017.

[Sel46] Atle Selberg. Contributions to the theory of the Riemann zeta-function. Arch. Math. Naturvid., 48(5):89–155, 1946.

[Sel92] Atle Selberg. Old and new conjectures and results about a class of Dirichlet series. In Proceedings of the Amalfi Conference on Analytic Number Theory (Maiori, 1989), pages 367–385. Univ. Salerno, Salerno, 1992.

[Sou09] K. Soundararajan. Moments of the Riemann zeta function. Ann. of Math. (2), 170(2):981–993, 2009.

[Sou21] K. Soundararajan. The distribution of values of zeta and L-functions. arXiv:2112.03389, 2021.