Moments of the Riemann zeta function on short intervals of the critical line

Louis-Pierre Arguin[*], **Frédéric Ouimet**[†] and **Maksym Radziwiłł**[‡]

Baruch College and Graduate Center (CUNY)
New-York, NY 10010, USA.
e-mail: louis-pierre.arguin@baruch.cuny.edu

Université de Montréal
Montreal, QC H3T 1J4, Canada.
e-mail: ouimetfr@dms.umontreal.ca

California Institute of Technology
Pasadena, CA 91125, USA.
e-mail: maksym@caltech.edu

Abstract: We show that as $T \to \infty$, for all $t \in [T, 2T]$ outside of a set of measure $\mathrm{o}(T)$,

$$\int_{-(\log T)^\theta}^{(\log T)^\theta} |\zeta(\tfrac{1}{2} + \mathrm{i}t + \mathrm{i}h)|^\beta \mathrm{d}h = (\log T)^{f_\theta(\beta) + \mathrm{o}(1)},$$

for some explicit exponent $f_\theta(\beta)$, where $\theta > -1$ and $\beta > 0$. This proves an extended version of a conjecture of Fyodorov and Keating (2014). In particular, it show that, for all $\theta > -1$, the moments exhibit a phase transition at a critical exponent $\beta_c(\theta)$, below which $f_\theta(\beta)$ is quadratic and above which $f_\theta(\beta)$ is linear. The form of the exponent f_θ also differs between mesoscopic intervals ($-1 < \theta < 0$) and macroscopic intervals ($\theta > 0$), a phenomenon that stems from an approximate tree structure for the correlations of zeta. We also prove that, for all $t \in [T, 2T]$ outside a set of measure $\mathrm{o}(T)$,

$$\max_{|h| \leq (\log T)^\theta} |\zeta(\tfrac{1}{2} + \mathrm{i}t + \mathrm{i}h)| = (\log T)^{m(\theta) + \mathrm{o}(1)},$$

for some explicit $m(\theta)$. This generalizes earlier results of Najnudel (2018) and Arguin et al. (2018) for $\theta = 0$. The proofs are unconditional, except for the upper bounds when $\theta > 3$, where the Riemann hypothesis is assumed.

MSC 2010 subject classifications: Primary 11M06, 60F10, 60G60, 60G70.
Keywords and phrases: Extreme value theory, Riemann zeta function.

1. Introduction

1.1. Maxima and moments over large intervals

Understanding the growth of the Riemann zeta function $\zeta(s)$ on the critical line $\mathrm{Re}\, s = \frac{1}{2}$ is a central problem in number theory due, among other things, to its relationship with the distribution of the zeros of $\zeta(s)$ and the more general subconvexity problem.

The Lindelöf hypothesis predicts that, for any $\varepsilon > 0$ and all $t \in \mathbb{R}$, we have $|\zeta(\frac{1}{2} + \mathrm{i}t)| = \mathcal{O}((1 + |t|)^\varepsilon)$, whereas it follows from the Riemann hypothesis that

$$|\zeta(\tfrac{1}{2} + \mathrm{i}t)| = \mathcal{O}\left(\exp\left(\left(\frac{\log 2}{2} + \mathrm{o}(1)\right)\frac{\log t}{\log\log t}\right)\right), \tag{1.1}$$

[*] L.-P. Arguin is supported in part by NSF Grant DMS-1513441 and by NSF CAREER DMS-1653602.
[†] F. Ouimet is supported by a NSERC Doctoral Program Alexander Graham Bell scholarship (CGS D3).
[‡] M. Radziwiłł acknowledges support of a Sloan fellowship.

as $t \to \infty$; see Chandee and Soundararajan (2011).

Unfortunately, there is a large gap between these conditional results and the best unconditional upper bounds, such as Bourgain (2017), which shows that $|\zeta(\frac{1}{2}+\mathrm{i}t)| = \mathcal{O}\left((1+|t|)^{13/84+\varepsilon}\right)$ for any given $\varepsilon > 0$ and all $t \in \mathbb{R}$. Currently, the best unconditional lower bound,

$$\max_{t\in[0,T]} |\zeta(\tfrac{1}{2}+\mathrm{i}t)| \geq \exp\left((\sqrt{2}+\mathrm{o}(1))\sqrt{\frac{\log T \log\log\log T}{\log\log T}}\right), \tag{1.2}$$

as $T \to \infty$, is established in de la Bretèche and Tenenbaum (2018) building on a method from Bondarenko and Seip (2017).

The true order of the maximum of $|\zeta(\frac{1}{2}+it)|$ remains a subject of dispute to this day. A conjecture that we find plausible is stated in Farmer, Gonek and Hughes (2007), where it is conjectured based on probabilistic models that

$$\max_{t\in[0,T]} |\zeta(\tfrac{1}{2}+\mathrm{i}t)| = \exp\left(\left(\frac{1}{\sqrt{2}}+\mathrm{o}(1)\right)\sqrt{\log T \cdot \log\log T}\right), \quad \text{as } T \to \infty. \tag{1.3}$$

Another set of central objects in the theory of the Riemann zeta function are the *moments*

$$\frac{1}{T}\int_T^{2T} |\zeta(\tfrac{1}{2}+\mathrm{i}t)|^\beta \mathrm{d}t, \quad \beta > 0. \tag{1.4}$$

Their importance comes from their relationship to the size and zero-distribution of $\zeta(s)$. However, unlike the problem of understanding the size of the global maximum of $|\zeta(\frac{1}{2}+it)|$, we are in possession of widely believed conjectures as to the behavior of moments. Following the work Keating and Snaith (2000), it is expected that, for all $\beta > 0$,

$$\frac{1}{T}\int_T^{2T} |\zeta(\tfrac{1}{2}+\mathrm{i}t)|^\beta \mathrm{d}t \sim C_\beta (\log T)^{\beta^2/4}, \tag{1.5}$$

as $T \to \infty$, and that the constant $C_\beta > 0$ factors into a product of two constants: one is computed from the moments of the characteristic polynomial of random unitary matrices, and the other is an arithmetic factor coming from the small primes.

There are a few results supporting (1.5). First, the conjecture (1.5) is known for $\beta = 2$ and $\beta = 4$ following the classical work of Hardy-Littlewood and Ingham. Upper bounds of the correct order of magnitude are established in Heap, Radziwiłł and Soundararajan (2019) for $0 < \beta \leq 4$. Meanwhile, lower bounds of the correct order of magnitude have been established for all $\beta \geq 2$ in Radziwiłł and Soundararajan (2013). Conditionally on the Riemann hypothesis, the correct order of magnitude of (1.5) is known for all $\beta > 0$ (see Harper (2013a) for the upper bounds and Heath-Brown (1981) for the lower bounds).

1.2. Maxima and moments over short intervals

Motivated by the problem of understanding the global maximum, Fyodorov, Hiary and Keating (2012); Fyodorov and Keating (2014) initiated the question of understanding the true size of *the local maximum of* $\zeta(\frac{1}{2}+\mathrm{i}t)$ by establishing a connection with log-correlated processes. If τ is sampled uniformly on $[T, 2T]$ under $\mathbb{P}$, they conjectured that for any $0 < \delta < 1$, there exists $C = C(\delta) > 0$ large enough and independent of T, such that with probability $1-\delta$,

$$\max_{h\in[-1,1]} \log|\zeta(\tfrac{1}{2}+\mathrm{i}\tau+\mathrm{i}h)| - \left(\log\log T - \frac{3}{4}\log\log\log T\right) \in [-C, C]. \tag{1.6}$$

They also conjectured the type of fluctuations around the recentering term.

The leading order $\log\log T$ was proved in Najnudel (2018) (conditionally on the Riemann hypothesis for the lower bound) and in Arguin et al. (2018) unconditionally. In Fyodorov, Hiary and Keating (2012), it is also argued that the moments in a short interval undergo a *freezing phase transition*, that is, as $T \to \infty$, the event,

$$\int_{[-1,1]} |\zeta(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)|^\beta \mathrm{d}h = \begin{cases} (\log T)^{\beta^2/4+1+\mathrm{o}(1)}, & \text{if } \beta \leq 2, \\ (\log T)^{\beta-1+\mathrm{o}(1)}, & \text{if } \beta > 2, \end{cases} \tag{1.7}$$

has $\mathbb{P}$-probability $1 - \mathrm{o}(1)$ as $T \to \infty$. Fyodorov and Keating (2014) state corresponding conjectures for mesoscopic intervals of length $(\log T)^\theta$ when $\theta \in (-1, 0)$, as well as finer asymptotics for the moments.

In view of Equations (1.5) and (1.7), an obvious question is to determine up to which interval size the freezing phase transition persists. In this paper, we establish that freezing transitions occur exactly for interval sizes of order $(\log T)^\theta$ with $\theta > -1$, including large intervals with $\theta > 0$. We also obtain the corresponding results for local maxima over such intervals. The following functions will be crucial to our analysis :

$$\begin{aligned} &\theta \leq 0: \qquad m(\theta) := 1 + \theta, \qquad f_\theta(\beta) := \begin{cases} \frac{\beta^2}{4}(1+\theta) + \theta, & \text{if } \beta \leq \beta_c := 2, \\ \beta m(\theta) - 1, & \text{if } \beta > \beta_c, \end{cases} \\ &\theta > 0: \qquad m(\theta) := \sqrt{1+\theta}, \qquad f_\theta(\beta) := \begin{cases} \frac{\beta^2}{4} + \theta, & \text{if } \beta \leq \beta_c := 2\sqrt{1+\theta}, \\ \beta m(\theta) - 1, & \text{if } \beta > \beta_c. \end{cases} \end{aligned} \tag{1.8}$$

Theorem 1.1 (Moments)**.** *Let $\theta > -1$, $\beta > 0$ and $\varepsilon > 0$ be given. Let τ be a random variable uniformly distributed on $[T, 2T]$ under the probability measure $\mathbb{P}$. Then, as $T \to \infty$, we have*

$$\mathbb{P}\Big(\int_{-(\log T)^\theta}^{(\log T)^\theta} |\zeta(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)|^\beta \mathrm{d}h < (\log T)^{f_\theta(\beta)-\varepsilon}\Big) = \mathrm{o}(1). \tag{1.9}$$

Moreover, if $\theta \leq 3$ or if the Riemann hypothesis holds, then as $T \to \infty$,

$$\mathbb{P}\Big(\int_{-(\log T)^\theta}^{(\log T)^\theta} |\zeta(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)|^\beta \mathrm{d}h > (\log T)^{f_\theta(\beta)+\varepsilon}\Big) = \mathrm{o}(1). \tag{1.10}$$

Proof. For the upper bound, see Section 2.3, and for the lower bound, see Proposition 3.2. □

When $\beta > \beta_c$, the moments exhibit *freezing*, i.e. they are dominated by just one large value corresponding to the local maximum of $|\zeta(\frac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)|$, $|h| \leq (\log T)^\theta$. Theorem 1.1 also suggests that freezing does not occur for intervals larger than any fixed power of $\log T$, since $\beta_c(\theta) \to \infty$ as $\theta \to \infty$.

Theorem 1.2 (Local maximum)**.** *Let $\theta > -1$ and $\varepsilon > 0$ be given. Let τ be a random variable uniformly distributed on $[T, 2T]$ under the probability measure $\mathbb{P}$. Then, as $T \to \infty$, we have*

$$\mathbb{P}\Big(\max_{|h| \leq (\log T)^\theta} |\zeta(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)| < (\log T)^{m(\theta)-\varepsilon}\Big) = \mathrm{o}(1). \tag{1.11}$$

Moreover, if $\theta \leq 3$ or if the Riemann hypothesis holds, then as $T \to \infty$,

$$\mathbb{P}\Big(\max_{|h| \leq (\log T)^\theta} |\zeta(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)| > (\log T)^{m(\theta)+\varepsilon}\Big) = \mathrm{o}(1). \tag{1.12}$$

Proof. For the upper bound, see Section 2.3, and for the lower bound, see Proposition 3.1. □

It is instructive to put these results in the context of two well-known facts on ζ. First, Selberg's central limit theorem, see for example Radziwiłł and Soundararajan (2017), states that, for any given $a < b$,

$$\mathbb{P}\left(\frac{\log|\zeta(\frac{1}{2}+\mathrm{i}\tau)|}{\sqrt{\frac{1}{2}\log\log T}} \in (a,b)\right) \overset{T\to\infty}{\longrightarrow} \int_a^b \frac{e^{-u^2/2}}{\sqrt{2\pi}}\,\mathrm{d}u. \tag{1.13}$$

In other words, a typical value of $\log|\zeta(\frac{1}{2}+\mathrm{i}\tau)|$ is a Gaussian random variable of variance $\frac{1}{2}\log\log T$. This is consistent with the moment conjecture (1.5) which gives a precise expression for the Laplace transform of $\log|\zeta(\frac{1}{2}+\mathrm{i}\tau)|$. Second, since $\zeta(\frac{1}{2}+it)$ with $T \leq t \leq 2T$ varies on the scale of $(\log T)^{-1}$, the statistics of extreme values of $\log|\zeta(\frac{1}{2}+\mathrm{i}\tau+\mathrm{i}h)|$, $|h| \leq (\log T)^{\theta}$, should be similar to the ones of $(\log T)^{1+\theta}$ Gaussian random variables of variance $\frac{1}{2}\log\log T$. If the random variables were independent, this is the so-called *Random Energy Model* (REM) in statistical mechanics introduced in Derrida (1981). For $\theta \geq 0$, it is not hard to check, using basic Gaussian tail estimates, that the expression (1.8) corresponds to the free energy of the model, and the results of Theorem 1.2, to the maximum of the REM. For more on this, we refer to Kistler (2015), where many techniques from REM were introduced to analyze log-correlated processes.

The REM heuristic is of course limited as the values of $\log|\zeta(\frac{1}{2}+\mathrm{i}\tau+\mathrm{i}h)|$, $|h| \leq (\log T)^{\theta}$, are correlated, as explained in Section 1.4. For $\theta < 0$, the correct model is a branching random walk which accurately predicts the changes in $m(\theta)$ and $f_\theta(\beta)$. For $\theta > 0$, our results show that the correlations do not affect large values at leading order (though the proofs must take them into account). As argued in Section 1.4, we believe that the correct probabilistic model for large values in this case is $(\log T)^{\theta}$ independent branching random walks. One implication is that, unlike the case $\theta \leq 0$, the REM heuristic should persist to subleading order (but fail at the level of fluctuations). In view of this, we believe that conjecture (1.6) needs to be expanded as follows to include large intervals:

Conjecture 1.3. *Let τ be sampled uniformly on $[T, 2T]$ under $\mathbb{P}$. Let $\theta > -1$ be given and let $m(\theta)$ be as in (1.8). For any $0 < \delta < 1$, there exists $C = C(\delta) > 0$ large enough and independent of T, such that with probability $1-\delta$,*

$$\max_{|h|\leq(\log T)^{\theta}} \log|\zeta(\tfrac{1}{2}+\mathrm{i}\tau+\mathrm{i}h)| - \big(m(\theta)\log\log T - r(\theta)\log\log\log T\big) \in [-C, C], \tag{1.14}$$

where

$$r(\theta) = \frac{3}{4} \quad \textit{if } \theta \leq 0 \qquad \textit{and} \qquad r(\theta) = \frac{1}{4\sqrt{1+\theta}} \quad \textit{if } \theta > 0.$$

1.3. Relations to other models

When $-1 < \theta \leq 0$, Conjecture 1.3 is based on modelling ζ by the characteristic polynomial of a random unitary matrix (CUE). More precisely, if M_N is a random matrix sampled from the Haar measure on the unitary group $\mathcal{U}(N)$, one can consider the moments

$$\mathbb{E}\left[\left(\frac{1}{2\pi}\int_0^{2\pi} |\det(\mathbb{I} - e^{-\mathrm{i}h}M_N)|^{2\beta}\mathrm{d}h\right)^k\right], \quad k > 0,\ \beta > 0. \tag{1.15}$$

These can be computed in the limit $N \to \infty$, at least heuristically, using Selberg integrals and the Fisher-Hartwig formula, cf. Fyodorov and Keating (2014). Exact expressions were recently obtained in Bailey and Keating (2018) in the regime $k, \beta \in \mathbb{N}$. The statistics of $\log \int_0^{2\pi} |\det(\mathbb{I} - e^{-\mathrm{i}h} M_N)|^{2\beta} \mathrm{d}h$ and of $\max_{h\in[0,2\pi]} |\det(\mathbb{I} - e^{-\mathrm{i}h} M_N)|$ in the limit $N \to \infty$ can be inferred from the asymptotics of the moments by comparison with log-correlated processes, cf. Fyodorov, Gnutzmann and Keating (2018) for a numerical study. In the CUE setting, the freezing analogue of (1.7) and the leading order as in (1.6) were proved in Arguin, Belius and Bourgade (2017). The subleading order of the maximum was proved in Paquette and Zeitouni (2018), and up to constant C in Chhaibi, Madaule and Najnudel (2018).

In the subcritical regime $\beta < \frac{1}{2}$, it is expected from the analysis of log-correlated processes, cf. Fyodorov and Bouchaud (2008), that the fluctuations of the maximum can be captured by a sum of two Gumbel variables. This was proved in Rémy (2018) for a specific log-correlated model by computing the moments in the range $k < \frac{1}{4\beta^2}$ of a random measure related to the theory of Gaussian multiplicative chaos, cf. Rhodes and Vargas (2014). In the CUE setting, this measure is the limit of

$$\frac{|\det(\mathbb{I} - e^{-\mathrm{i}h} M_N)|^{2\beta}}{\mathbb{E}[|\det(\mathbb{I} - e^{-\mathrm{i}h} M_N)|^{2\beta}]} \frac{\mathrm{d}h}{2\pi}. \tag{1.16}$$

The limit of the above was shown to be non-degenerate for $\beta < 1$ in Webb (2015); Nikula, Saksman and Webb (2018). Such a random measure can also be considered in the context of the Riemann zeta function for mesoscopic intervals of length $(\log T)^{\theta}$, $-1 < \theta \leq 0$, with $|\zeta(\frac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)|$ in place of $|\det(\mathbb{I} - e^{-\mathrm{i}h} M_N)|$. (There does not seem to be any obvious equivalent for macroscopic intervals, $\theta > 0$, in the CUE model.) A step in this direction was made in Saksman and Webb (2018) where $\zeta(\frac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)$, $h \in \mathbb{R}$, was shown to converge as $T \to \infty$ when considered as a random variable on the space of tempered distributions.

Another model for the large values of $\log|\zeta(\frac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)|$, $h \in [-1, 1]$, is to consider a random Dirichlet polynomial $X_h = \mathrm{Re} \sum_{p \leq T} p^{-1/2 - \mathrm{i}h} U_p$, where ($U_p$, p primes) are i.i.d. uniform random variables on the unit circle, cf. Harper (2013b); Arguin, Belius and Harper (2017); Arguin and Ouimet (2019). The analogue of conjecture (1.6) for this model was proved up to second-order corrections in Arguin, Belius and Harper (2017), and large deviations and continuity estimates for the derivative were found in Arguin and Ouimet (2019). The limit of the corresponding multiplicative chaos measure was obtained in Saksman and Webb (2018). A proof of the freezing phase transition was given in Arguin and Tai (2018). In the latter, the limit of the Gibbs measure $\exp(\beta X_h)\mathrm{d}h$ is also studied in the supercritical regime $\beta > 2$, showing that it is supported on h's that are at a relative distance of order one or order $(\log T)^{-1}$ of each other. This result was used in Ouimet (2018) to prove that the normalized Gibbs weights converge to a Poisson-Dirichlet distribution.

Notation. Throughout the article, the notation τ will denote a random variable uniformly distributed on $[T, 2T]$ under $\mathbb{P}$. Expectations under $\mathbb{P}$ are denoted by $\mathbb{E}$. We write $f(T) = \mathrm{o}(g(T))$ if $|f(T)/g(T)|$ tends to 0 as $T \to \infty$ when the parameters β, θ and ε are fixed. Similarly, we write $f(T) = \mathcal{O}(g(T))$ if $\limsup |f(T)/g(T)|$ is bounded for β, θ and ε fixed. Finally, we will sometimes write for conciseness $f(T) \ll g(T)$ if $f(T) = \mathcal{O}(g(T))$, and also $f(T) \asymp g(T)$ if both $f(T) \ll g(T)$ and $g(T) \ll f(T)$ hold.

1.4. Outline of the proof

For $\theta > 0$, the upper bound part of Theorem 1.1 and Theorem 1.2 follows from the moment estimates

$$\mathbb{E}\Big[|\zeta(\tfrac{1}{2} + \mathrm{i}\tau)|^{\beta}\Big] \ll (\log T)^{\beta^2/4+\varepsilon}, \tag{1.17}$$

and from a discretization result which roughly shows that for a Dirichlet polynomial D that approximates zeta, and for $\beta \geq 1$, we have

$$\max_{|h| \leq (\log T)^{\theta}} |D(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)|^{\beta} \ll \sum_{|k| \leq (\log T)^{1+\theta}} \big|D\big(\tfrac{1}{2} + \mathrm{i}\tau + \tfrac{2\pi \mathrm{i}k}{\log T}\big)\big|^{\beta}. \tag{1.18}$$

Equation (1.18) tells us that the process $(\zeta(\frac{1}{2} + \mathrm{i}\tau + \mathrm{i}h),\ |h| \leq (\log T)^{\theta})$ varies on a $(\log T)^{-1}$ scale, so that the maximum and moments on an interval of length $\mathcal{O}((\log T)^{\theta})$ behave as those of $\mathcal{O}((\log T)^{1+\theta})$ i.i.d. Gaussian random variables of variance $\frac{1}{2}\log\log T$. The limitation to $\theta \leq 3$ comes from the fact that the upper bounds (1.17) are not known unconditionally for $\beta > 4$.

When $\theta < 0$, the upper bounds in Theorem 1.1 and Theorem 1.2 are a bit more delicate. We follow essentially the same strategy, but we apply it to the function

$$(\zeta \cdot e^{-\mathcal{P}_{|\theta|}})(\tfrac{1}{2} + \mathrm{i}\tau), \quad \text{where } \mathcal{P}_{|\theta|}(s) := \sum_{\log p \leq (\log T)^{|\theta|}} \frac{1}{p^s}, \tag{1.19}$$

instead of $\zeta(\frac{1}{2} + \mathrm{i}\tau)$. As discussed in more detail below, the reason for this is that when $\theta < 0$, the contribution of the primes up to scale $|\theta|$ is negligible with high probability, namely, with probability $1 - \mathrm{o}(1)$,

$$\max_{|h| \leq (\log T)^{\theta}} \big|\mathcal{P}_{|\theta|}(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)\big| = \mathrm{o}(\log\log T). \tag{1.20}$$

When τ is restricted to a specific event $\mathcal{A}(T)$ on which (1.19) can be discretized as in (1.18), we can show that

$$\mathbb{E}\Big[\big|(\zeta \cdot e^{-\mathcal{P}_{|\theta|}})(\tfrac{1}{2} + \mathrm{i}\tau)\big|^{\beta}\Big] \ll (\log T)^{(\beta^2/4)\cdot(1+\theta)+\varepsilon} \tag{1.21}$$

for $\beta \leq 2$. This explains the additional factor $(\beta^2/4)\theta$ in $f_{\theta}(\beta)$ when $-1 < \theta < 0$ and $\beta \leq 2$.

We then turn to the lower bound part of Theorem 1.1 and Theorem 1.2. The lower bounds in Theorem 1.2 follow directly from Theorem 1.1 (see (3.90)), so it is enough to discuss Theorem 1.1.

The problem is first reduced to obtaining lower bounds for moments off the critical line. In particular, it is shown, uniformly in $\frac{1}{2} \leq \sigma \leq \frac{1}{2} + (\log T)^{\theta - 3\varepsilon}$ and for any given $\varepsilon > 0$, that with probability $1 - \mathrm{o}(1)$,

$$\int_{-(\log T)^{\theta}}^{(\log T)^{\theta}} |\zeta(\sigma + \mathrm{i}t + \mathrm{i}h)|^{\beta} \mathrm{d}h \ll \int_{-2(\log T)^{\theta}}^{2(\log T)^{\theta}} |\zeta(\tfrac{1}{2} + \mathrm{i}t + \mathrm{i}h)|^{\beta} \mathrm{d}h + \frac{1}{(\log T)^7}. \tag{1.22}$$

This is accomplished by using a result of Gabriel (1927) for subharmonic functions, and the construction of an explicit entire function which is a good approximation to the indicator function of the rectangle $\mathcal{R} = \{\sigma + \mathrm{i}u : |u| \leq (\log T)^{\theta}, \frac{1}{2} \leq \sigma \leq \frac{1}{2} + (\log T)^{\theta - 3\varepsilon}\}$ in the whole strip $\frac{1}{2} \leq \mathrm{Re}\, s$. The fact that the interval can be very small when $\theta < 0$ makes this part rather technical. We believe that this result might be useful in other applications as well.

The problem is therefore reduced to obtaining a good lower bound for

$$\int_{|h|\leq(\log T)^{\theta}} |\zeta(\sigma_0+\mathrm{i}\tau+\mathrm{i}h)|^{\beta}\mathrm{d}h, \quad \text{with } \sigma_0=\frac{1}{2}+\frac{1}{(\log T)^{1-\delta}}, \tag{1.23}$$

for some sufficiently small $\delta>0$. We adapt mollification results from Arguin et al. (2018) to show that, outside of an event of probability $\mathrm{o}(1)$, the problem can be reduced to understanding

$$\int_{|h|\leq(\log T)^{\theta}} \exp\big(\beta\,\mathrm{Re}\,\mathcal{P}_{1-\delta}(\sigma_0+\mathrm{i}\tau+\mathrm{i}h)\big)\mathrm{d}h. \tag{1.24}$$

The proof of the lower bound is now restricted to the problem of understanding the correlation structure of the process

$$\big(\mathrm{Re}\,\mathcal{P}_{1-\delta}(\sigma_0+\mathrm{i}\tau+\mathrm{i}h),\ |h|\leq(\log T)^{\theta}\big). \tag{1.25}$$

The remaining part of the argument is done in Section 3.4 by a multiscale second moment method introduced in Kistler (2015). The covariance of the process (1.25) can be computed using Lemma A.4 with $a(p)=p^{-\sigma_0}(p^{-\mathrm{i}h}+p^{-\mathrm{i}h'})$:

$$\begin{aligned}\mathbb{E}\Big[&\mathrm{Re}\,\mathcal{P}_{1-\delta}(\sigma_0+\mathrm{i}\tau+\mathrm{i}h)\cdot\mathrm{Re}\,\mathcal{P}_{1-\delta}(\sigma_0+\mathrm{i}\tau+\mathrm{i}h')\Big]\\ &=\frac{1}{2}\sum_{\log p\leq(\log T)^{1-\delta}}\frac{\cos(|h-h'|\log p)}{p^{2\sigma_0}}+\mathcal{O}(1).\end{aligned} \tag{1.26}$$

The cosine factor implies that primes smaller than $\exp(|h-h'|^{-1})$ are almost perfectly correlated, whereas primes greater than $\exp(|h-h'|^{-1})$ decorrelate quickly. In fact, the covariance can be evaluated precisely using the prime number theorem and equals $\frac{1}{2}\log|h-h'|^{-1}+\mathcal{O}(1)$. This shows that the process is approximatively a log-correlated Gaussian process.

The identification with a log-correlated process is useful as it suggests that the Dirichlet polynomials have an underlying tree structure. To see this, consider the *increments*

$$P_k(h)=\sum_{e^{k-1}<\log p\leq e^k}\mathrm{Re}\,\frac{1}{p^{\sigma_0+\mathrm{i}\tau+\mathrm{i}h}}, \quad 1\leq k\leq\log\log T. \tag{1.27}$$

The range of primes is chosen so that each P_k has variance $\frac{1}{2}+\mathrm{o}(1)$. In this framework, the Dirichlet polynomial at h can be seen as a random walk with independent and identically distributed increments. However, the random walks for different h's are not independent by (1.26). In fact, the walks are almost perfectly correlated until they *branch out* around the prime $p\approx\exp(|h-h'|^{-1})$, corresponding to the increment $k(h,h')=\log|h-h'|^{-1}$. Since k goes to essentially $\log\log T$, the analysis can be restricted to h's at a distance $(\log T)^{-1}$ of each other. Furthermore, the h's in an interval of size $(\log T)^{-\alpha}$ for $0<\alpha<1$ will share the same increments up to $k\approx\alpha\log\log T$.

The above observations have important consequences for the probabilistic analysis. For $\theta=0$, this means that the process (1.25) on an interval of order one is well approximated by a Gaussian process indexed by a tree of average degree $e=2.718\ldots$, where the independent increments $P_k(h)$ are identified with the edges of the tree. Note that the number of leaves on the interval $[-1,1]$ is then $\approx e^{\log\log T}=\log T$. Equivalently, the walks $\sum_k P_k(h)$, $h\in[-1,1]$,

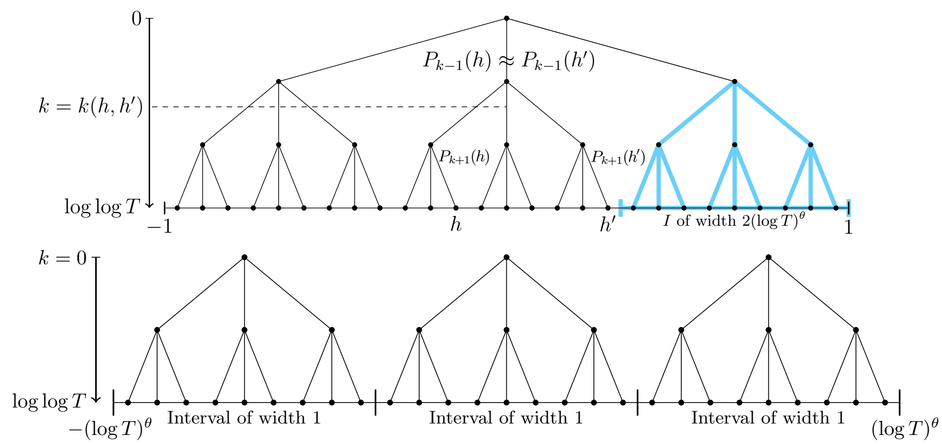

Fig 1: **(Top)** An illustration of the branching random walk $\sum_k P_k$ for the interval I with $\theta = 0$. The one for a subinterval with $\theta < 0$ is depicted in blue. **(Bottom)** An illustration of the independent branching random walks $\sum_k P_k$ for disjoint intervals of width 1 inside I of length $2(\log T)^{\theta}$ with $\theta > 0$.

can be seen as a branching random walk on a Galton-Watson tree with an average number of offspring e, cf. Figure 1.

For $\theta < 0$, the tree structure suggests that the primes up to $\exp((\log T)^{|\theta|})$ do not contribute to large values, since they should be essentially the same for all h's in the interval . Therefore these primes can be cutoff at a low cost, cf. Lemma 2.3. This is equivalent to restricting to a subtree of the one on $[-1, 1]$ with $(1 + \theta) \log\log T$ increments and $(\log T)^{1+\theta}$ leaves, yielding a maximum at leading order of $(1 + \theta) \log\log T$ by the REM heuristic.

The case $\theta > 0$ stands out as the analogy with branching random walks fails. This is because the random walks for h and h' are essentially independent for $|h - h'| > 1$. Therefore the right probabilistic model seems to consist of $(\log T)^{\theta}$ independent branching random walks corresponding to different intervals of order one, see Figure 1. A large class of similar models (called CREM's for *Continuous Random Energy Models*) have been studied in Bovier and Kurkova (2004), see Bovier (2006, 2017) for a review. It turns out that the large values at leading order correspond to the ones of a REM with $(\log T)^{1+\theta}$ variables with variance $\frac{1}{2} \log\log T$. This yields a maximum of $\sqrt{1+\theta} \log\log T$ at leading order. In fact, in view of the extreme value statistics of CREM's, we expect that the REM heuristic holds for subleading corrections. This is the motivation for Conjecture 1.3.

2. Upper bounds

2.1. Moment estimates

We will need a number of moment estimates which we state below.

Proposition 2.1. *Assume the Riemann hypothesis. Let $\beta > 0$ and $\varepsilon > 0$ be given. Then,*

$$\mathbb{E}\big[|\zeta(\tfrac{1}{2} + \mathrm{i}\tau)|^{\beta}\big] \ll (\log T)^{\beta^2/4+\varepsilon}. \tag{2.1}$$

Proof. See Soundararajan (2009). □

Proposition 2.2. *Let $0 < \beta \leq 4$ be given. Then,*

$$\mathbb{E}\big[|\zeta(\tfrac{1}{2} + \mathrm{i}\tau)|^{\beta}\big] \ll (\log T)^{\beta^2/4}. \tag{2.2}$$

Proof. See Heap, Radziwiłł and Soundararajan (2019). □

Remark. *The proof of Proposition 2.1 is based on the following deterministic upper bound for ζ: Suppose that T is large. Let $T \leq t \leq 2T$, and let $2 \leq x \leq T^2$. Then, as $T \to \infty$, we have*

$$\log|\zeta(\frac{1}{2} + \mathrm{i}t)| \leq Re \sum_{p \leq x} \frac{1}{p^{\frac{1}{2} + \frac{1}{\log x} + \mathrm{i}t}} \frac{\log(x/p)}{\log x} + \frac{\log T}{\log x} + \mathcal{O}(\log\log\log T), \tag{2.3}$$

see Proposition and Lemma 2 in Soundararajan (2009). On the Riemann hypothesis, the upper bounds to Theorems 1.1 and 1.2 could be proved in a simpler way by using this deterministic bound, and by proving the corresponding results for the Dirichlet polynomials. For unconditional results, such a deterministic upper bound is not available. We need to work on average to discard the contribution of large primes. This is the purpose of Lemmas 2.4, 2.5, 2.6 and Proposition 2.7 below.

Everywhere in Section 2, we will denote, for $\alpha > 0$ and $s \in \mathbb{C}$,

$$\mathcal{P}_{\alpha}(s) := \sum_{\log p \leq (\log T)^{\alpha}} \frac{1}{p^s}. \tag{2.4}$$

The following auxiliary estimate will be repeatedly useful in the case of short intervals to discard the contribution of small primes, which should be negligible as explained at the end of Section 1.4.

Lemma 2.3. *Let $-1 < \theta < 0$ and $\sigma \geq 1/2$. For any $0 < \varepsilon < C$ and $V = V(T)$ that satisfies $\varepsilon \log\log T \leq V \leq C \log\log T$, we have*

$$\mathbb{P}\Big(\max_{|h| \leq (\log T)^{\theta}} \big|\mathcal{P}_{|\theta|}(\sigma + \mathrm{i}\tau + \mathrm{i}h)\big| > V\Big) \ll e^{-cV}, \tag{2.5}$$

for some constant $c = c(\varepsilon, C) > 0$.

Proof. For a lighter notation, write $S(h) := \mathcal{P}_{|\theta|}(\sigma + \mathrm{i}\tau + \mathrm{i}h)$. (We keep the dependence on τ implicit, consistent with the probabilistic notation for random variables.) We have

$$\begin{aligned}\mathbb{P}\Big(\max_{|h| \leq (\log T)^{\theta}} |S(h)| > V\Big) \leq \mathbb{P}\Big(&\max_{|h| \leq (\log T)^{\theta}} |S(h) - S(0)| > V/2\Big) \\ &+ \mathbb{P}\big(|S(0)| > V/2\big).\end{aligned} \tag{2.6}$$

Let ℓ denote a generic natural integer. By Chebyshev's inequality, a moment estimate (Lemma A.5) and a prime number theorem estimate (Lemma A.1), we have

$$\mathbb{P}\big(|S(0)| > V/2\big) \leq \frac{\mathbb{E}\big[|S(0)|^{2\ell}\big]}{(V/2)^{2\ell}} \ll \ell! \left(\frac{\sum_{p \leq T} p^{-2\sigma}}{(V/2)^2}\right)^{\ell} \ll \left(\frac{4\ell \log\log T}{\varepsilon^2 (\log\log T)^2}\right)^{\ell}. \tag{2.7}$$

With the choice $\ell = \lfloor \frac{\varepsilon^2}{8} \log\log T \rfloor$, this probability is $\ll \exp(-aV)$ for some constant $a = a(\varepsilon, C) > 0$.

It remains to control the first probability on the right-hand side of (2.6). Let ℓ denote another natural integer to be chosen later. By applying the Sobolev inequality (Lemma A.2) to the function $h \mapsto (S(h) - S(0))^{\ell}$ and the Cauchy-Schwarz inequality, we have

$$\mathbb{E}\Big[\max_{|h|\leq(\log T)^{\theta}} |S(h)-S(0)|^{2\ell}\Big] \ll \max_{|h|\leq(\log T)^{\theta}} \mathbb{E}\big[|S(h)-S(0)|^{2\ell}\big] + \ell\,(\log T)^{\theta} \max_{|h|\leq(\log T)^{\theta}} \Big(\mathbb{E}\big[|S(h)-S(0)|^{4\ell-2}\big]\cdot\mathbb{E}\big[|S'(h)|^2\big]\Big)^{1/2}. \tag{2.8}$$

A short calculation, using moment estimates (Lemma A.5) followed by prime number theorem estimates (Lemma A.1), gives, for all $|h| \leq (\log T)^{\theta}$,

$$\begin{aligned}\mathbb{E}\big[|S(h)-S(0)|^{4\ell-2}\big] &\ll (2\ell-1)!\Bigg(\sum_{\log p\leq(\log T)^{|\theta|}} \frac{2-2\cos(|h|\log p)}{p}\Bigg)^{2\ell-1} \\ &\ll (\ell\,c)^{2\ell-1},\end{aligned} \tag{2.9}$$

for some constant $c > 0$ (to obtain the last inequality, note that $|h|\cdot(\log T)^{|\theta|} \leq 1$). Similarly, for all $|h| \leq (\log T)^{\theta}$,

$$\mathbb{E}\big[|S'(h)|^2\big] \ll \sum_{\log p\leq(\log T)^{|\theta|}} \frac{(\log p)^2}{p^{2\sigma}} \ll (\log T)^{2|\theta|}. \tag{2.10}$$

Combining (2.9) and (2.10) shows that the right-hand side of (2.8) is $\ll \ell^{1/2}(\ell\,c)^{\ell}$. Then, by Chebyshev's inequality and the choice $\ell = \lfloor \frac{\varepsilon^2}{8c}\log\log T\rfloor$, we deduce

$$\mathbb{P}\Big(\max_{|h|\leq(\log T)^{\theta}} |S(h)-S(0)| > V/2\Big) \ll \ell^{1/2}\bigg(\frac{4\,\ell\,c}{V^2}\bigg)^{\ell} \ll e^{-bV}, \tag{2.11}$$

for some constant $b = b(\varepsilon, C) > 0$. □

Given $\alpha, \beta \in \mathbb{R}$ and $\theta > -1$, let $\mathfrak{F}_{\alpha,\beta,\theta}(n)$ denote a completely multiplicative function such that

$$\mathfrak{F}_{\alpha,\beta,\theta}(p) := \begin{cases} \alpha, & \text{if } \log p \leq \log^{|\theta|} T, \\ \beta, & \text{if } \log^{|\theta|} T \leq \log p, \end{cases} \tag{2.12}$$

and let $\mathfrak{g}$ be a multiplicative function such that,

$$\mathfrak{g}(p^{\alpha}) = \frac{1}{\alpha!} \tag{2.13}$$

for all $\alpha \geq 1$ and all primes p.

In the next three lemmas, we control various terms with the aim of proving the moment estimate in Proposition 2.7, which we will need in the case of short intervals.

Lemma 2.4. *Let $\varepsilon > 0$, $\beta > 0$ and $-1 < \theta < 0$ be given. Then,*

$$\mathbb{E}\Bigg[\Bigg|\sum_{\substack{\Omega(n)\leq 100\lfloor\log\log T\rfloor \\ p|n \implies \log p\leq\log^{1-\varepsilon} T}} \frac{\mathfrak{F}_{0,\beta/2,\theta}(n)\mathfrak{g}(n)}{n^{1/2+\mathrm{i}\tau}}\Bigg|^2\Bigg] \ll (\log T)^{\beta^2(1+\theta)/4}. \tag{2.14}$$

Proof. Notice that the Dirichlet polynomial in (2.14) has length $\ll T^{\varepsilon}$ for any fixed $\varepsilon > 0$. In particular, by the mean-value formula (Lemma A.3),

$$\mathbb{E}\Big[\Big|\sum_{\substack{\Omega(n)\leq 100\lfloor\log\log T\rfloor\\ p\mid n \implies \log p\leq \log^{1-\varepsilon} T}} \frac{\mathfrak{F}_{0,\beta/2,\theta}(n)\mathfrak{g}(n)}{n^{1/2+\mathrm{i}\tau}}\Big|^2\Big] \ll \sum_{\substack{\Omega(n)\leq 100\lfloor\log\log T\rfloor\\ p\mid n \implies \log p\leq \log^{1-\varepsilon} T}} \frac{\mathfrak{F}_{0,\beta/2,\theta}(n)^2\mathfrak{g}(n)^2}{n} \leq \prod_{\log^{|\theta|} T\leq \log p\leq \log^{1-\varepsilon} T}\Big(1+\frac{\mathfrak{F}_{0,\beta/2,\theta}(p)}{p}\Big), \tag{2.15}$$

and this is $\ll (\log T)^{\beta^2(1+\theta)/4}$ as needed. □

Lemma 2.5. *Let $\varepsilon > 0$, $\beta > 0$ and $-1 < \theta < 0$ be given. Then,*

$$\mathbb{E}\Big[|\zeta(\tfrac{1}{2}+\mathrm{i}\tau)|^2\cdot\Big|\sum_{\substack{\Omega(n)\leq 100\lfloor\log\log T\rfloor\\ p\mid n \implies \log p\leq \log^{1-\varepsilon} T}} \frac{\mathfrak{F}_{-1,\beta/2-1,\theta}(n)\mathfrak{g}(n)}{n^{1/2+\mathrm{i}\tau}}\Big|^2\Big] \ll (\log T)^{\beta^2(1+\theta)/4+\varepsilon}. \tag{2.16}$$

Proof. By Bettin, Chandee and Radziwiłł (2017), the left-hand side of (2.16) is

$$\leq \sum_{\substack{\Omega(n)\leq 100\lfloor\log\log T\rfloor\\ \Omega(m)\leq 100\lfloor\log\log T\rfloor\\ p\mid n \implies \log p\leq \log^{1-\varepsilon} T\\ p\mid m \implies \log p\leq \log^{1-\varepsilon} T}} \frac{\mathfrak{F}_{-1,\beta/2-1,\theta}(nm)\mathfrak{g}(n)\mathfrak{g}(m)}{[n,m]}\cdot\frac{1}{T}\int_{\mathbb{R}}\Big(\log\Big(\frac{t(n,m)^2}{2\pi nm}\Big)+2\gamma\Big)\Phi\Big(\frac{t}{T}\Big)\mathrm{d}t+\mathcal{O}(T^{-\varepsilon}),$$

with Φ a smooth non-negative function such that $\Phi(x)\geq 1$ for $1\leq x\leq 2$.

Using Chernoff's bound we can get rid of the restriction $\Omega(n)\leq 100\lfloor\log\log T\rfloor$. Indeed, it suffices to notice that the contribution of the above sum over n with $\Omega(n) > 100\lfloor\log\log T\rfloor$ is bounded by

$$\begin{aligned}\log T &\sum_{\substack{p\mid n \implies p\leq T\\ p\mid m \implies p\leq T}} \frac{|\mathfrak{F}_{-1,\beta/2-1,\theta}(nm)|}{[n,m]} e^{\alpha\Omega(n)-\alpha 100\log\log T}\\ &\ll (\log T)^{1-100\alpha}\prod_{p\leq T}\Big(1+\frac{(1+e^{\alpha})|\mathfrak{F}_{-1,\beta/2-1,\theta}(p)|}{p}+\frac{|\mathfrak{F}_{-1,\beta/2-1,\theta}(p^2)|e^{\alpha}}{p}\Big)\\ &\ll (\log T)^{1-100\alpha}\cdot(\log T)^{1+2e^{\alpha}},\end{aligned} \tag{2.17}$$

and this is completely negligible if we choose $\alpha = 1$, for example.

We now notice that

$$\begin{aligned}\sum_{\substack{p\mid n \implies \log p\leq \log^{1-\varepsilon} T\\ p\mid m \implies \log p\leq \log^{1-\varepsilon} T}} \frac{\mathfrak{F}_{-1,\beta/2,\theta}(nm)\mathfrak{g}(n)\mathfrak{g}(m)}{[n,m]} &\asymp \prod_{\log p\leq \log^{1-\varepsilon} T}\Big(1+\frac{2\mathfrak{F}_{-1,\beta/2-1,\theta}(p)+\mathfrak{F}_{-1,\beta/2-1,\theta}(p)^2}{p}\Big)\\ &\asymp (\log T)^{-|\theta|}\cdot(\log T)^{(\beta^2/4-1)\cdot(1+\theta-\varepsilon)}\\ &\ll (\log T)^{\beta^2(1+\theta)/4-1+\varepsilon}.\end{aligned} \tag{2.18}$$

To evaluate the remaining part of the sum, write

$$\log\Big(\frac{(m,n)^2}{mn}\Big) = \frac{1}{2\pi \mathrm{i}} \oint_{|z|=1/\log T} \Big(\frac{(m,n)^2}{mn}\Big)^z \cdot \frac{\mathrm{d}z}{z^2}. \tag{2.19}$$

Then, we end up having to evaluate,

$$\frac{1}{2\pi \mathrm{i}} \oint_{|z|=1/\log T} \sum_{\substack{p|n \implies \log p \leq \log^{1-\varepsilon} T \\ p|m \implies \log p \leq \log^{1-\varepsilon} T}} \frac{\mathfrak{F}_{-1,\beta/2,\theta}(mn)\mathfrak{g}(m)\mathfrak{g}(n)}{[m,n]} \cdot \Big(\frac{(m,n)^2}{mn}\Big)^z \cdot \frac{\mathrm{d}z}{z^2}. \tag{2.20}$$

As above, the sum over m and n factors into an Euler product which is

$$\asymp \prod_{\log p \leq \log^{1-\varepsilon} T} \Big(1 + \frac{2\mathfrak{F}_{-1,\beta/2,\theta}(p)}{p^{1+z}} + \frac{\mathfrak{F}_{-1,\beta/2,\theta}(p)^2}{p} + \mathcal{O}\Big(\frac{1}{p^2}\Big)\Big). \tag{2.21}$$

Note that since $|z| = 1/\log T$ by a Taylor expansion,

$$\sum_{\log p \leq \log^{1-\varepsilon} T} \frac{\mathfrak{F}_{-1,\beta/2,\theta}(p)}{p^{1+z}} = \sum_{\log p \leq \log^{1-\varepsilon} T} \frac{\mathfrak{F}_{-1,\beta/2,\theta}(p)}{p} + \mathcal{O}\Big(\frac{1}{\log T} \sum_{\log p \leq \log^{1-\varepsilon} T} \frac{\log p}{p}\Big),$$

and since the above error term is o(1), the Euler product in (2.21) is

$$\asymp \prod_{\log p \leq \log^{1-\varepsilon} T} \Big(1 + \frac{2\mathfrak{F}_{-1,\beta/2,\theta}(p) + \mathfrak{F}_{-1,\beta/2,\theta}(p)^2}{p}\Big) \asymp (\log T)^{\beta^2(1+\theta)/4 - 1 + \varepsilon}. \tag{2.22}$$

Therefore (2.20) is $\ll (\log T)^{\beta^2(1+\theta)/4+\varepsilon}$ as required. □

Lemma 2.6. *Let $\varepsilon > 0$ be given. For $\ell = 50\lfloor \log\log T \rfloor$, we have*

$$\mathbb{E}\Big[|\zeta(\tfrac{1}{2} + \mathrm{i}\tau)|^2 \cdot \Big|\frac{\mathcal{P}_{1-\varepsilon}(\frac{1}{2} + \mathrm{i}\tau)}{100 \log\log T}\Big|^{2\ell}\Big] \ll (\log T)^{-21}, \tag{2.23}$$

and

$$\mathbb{E}\Big[\Big|\frac{\mathcal{P}_{1-\varepsilon}(\frac{1}{2} + \mathrm{i}\tau)}{100 \log\log T}\Big|^{2\ell}\Big] \ll (\log T)^{-21}. \tag{2.24}$$

Proof. For (2.23), we apply the Cauchy-Schwarz inequality, a fourth moment bound on zeta, and a moment estimate (Lemma A.5) followed by a prime number theorem estimate (Lemma A.1) on the remaining term to conclude that the expectation is

$$\ll (\log T)^2 \cdot \mathbb{E}\Big[\Big|\frac{\mathcal{P}_{1-\varepsilon}(\frac{1}{2} + \mathrm{i}\tau)}{100 \log\log T}\Big|^{4\ell}\Big]^{1/2} \ll (\log T)^2 \cdot (\log T)^{-50}. \tag{2.25}$$

The proof of (2.24) is even more straightforward. □

Finally, we will need the following moment estimate for the case of short intervals.

Proposition 2.7. *Let $-1 < \theta < 0$, $0 < \beta \leq 2$ and $\varepsilon > 0$ be given. Then, as $T \to \infty$,*

$$\mathbb{E}\Big[\big|(\zeta \cdot e^{-\mathcal{P}_{|\theta|}})(\tfrac{1}{2} + \mathrm{i}\tau)\big|^{\beta} \mathbf{1}_{\mathcal{A}(T)}\Big] \ll (\log T)^{\beta^2(1+\theta)/4+\varepsilon}, \tag{2.26}$$

with the event

$$\mathcal{A}(T) := \big\{|\mathcal{P}_{|\theta|}(\tfrac{1}{2} + \mathrm{i}\tau)| \leq 2 \log\log T\big\}. \tag{2.27}$$

Proof. Let $0 < \beta < 2$. By Young's inequality with $p = 2/\beta$ and $q = 2/(2-\beta)$,

$$\begin{aligned} \left|\zeta(\tfrac{1}{2} + \mathrm{i}\tau)\right|^\beta &\leq \tfrac{1}{p} \cdot |\zeta(\tfrac{1}{2} + \mathrm{i}\tau)|^2 \cdot e^{-\frac{2}{q}\mathrm{Re}\,\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau)} + \tfrac{1}{q} \cdot e^{\frac{2}{p}\mathrm{Re}\,\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau)} \\ &= \tfrac{\beta}{2} \cdot |\zeta(\tfrac{1}{2} + \mathrm{i}\tau)|^2 \cdot e^{-(2-\beta)\mathrm{Re}\,\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau)} + \tfrac{2-\beta}{2} \cdot e^{\beta\mathrm{Re}\,\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau)}. \end{aligned} \tag{2.28}$$

Note that (2.28) holds trivially for $\beta = 2$. Hence, for $0 < \beta \leq 2$,

$$\begin{aligned} \left|(\zeta \cdot e^{-\mathcal{P}_{|\theta|}})(\tfrac{1}{2} + \mathrm{i}\tau)\right|^\beta \leq \tfrac{\beta}{2} \, |\zeta(\tfrac{1}{2} + \mathrm{i}\tau)|^2 \cdot e^{-(2-\beta)\mathrm{Re}\,\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau) - \beta\mathrm{Re}\,\mathcal{P}_{|\theta|}(\frac{1}{2}+\mathrm{i}\tau)} \\ + \tfrac{2-\beta}{2} \, e^{\beta\mathrm{Re}\,\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau) - \beta\mathrm{Re}\,\mathcal{P}_{|\theta|}(\frac{1}{2}+\mathrm{i}\tau)}. \end{aligned} \tag{2.29}$$

On the event $\mathcal{A}(T) \cap \{|\mathcal{P}_{1-\varepsilon}(\frac{1}{2} + \mathrm{i}\tau)| \leq 100 \log\log T\}$,

$$e^{-(2-\beta)\mathrm{Re}\,\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau) - \beta\mathrm{Re}\,\mathcal{P}_{|\theta|}(\frac{1}{2}+\mathrm{i}\tau)} \ll \Bigg| \sum_{\substack{\Omega(n) \leq 100\lfloor \log\log T \rfloor \\ p|n \implies \log p \leq \log^{1-\varepsilon} T}} \frac{\mathfrak{F}_{-1,\beta/2-1,\theta}(n)\mathfrak{g}(n)}{n^{1/2+\mathrm{i}t}} \Bigg|^2 \tag{2.30}$$

where $\mathfrak{F}_{\alpha,\beta,\theta}(n)$ is the completely multiplicative function defined in (2.12) and $\mathfrak{g}(n)$ is a multiplicative function such that $\mathfrak{g}(p^\alpha) = 1/\alpha!$ for all integers $\alpha \geq 1$ and primes p. Likewise, on the event $\mathcal{A}(T) \cap \{|\mathcal{P}_{1-\varepsilon}(\frac{1}{2} + \mathrm{i}\tau)| \leq 100 \log\log T\}$,

$$e^{\beta\mathrm{Re}\,\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau) - \beta\mathrm{Re}\,\mathcal{P}_{|\theta|}(\frac{1}{2}+\mathrm{i}\tau)} \ll \Bigg| \sum_{\substack{\Omega(n) \leq 100\lfloor \log\log T \rfloor \\ p|n \implies \log p \leq \log^{1-\varepsilon} T}} \frac{\mathfrak{F}_{0,\beta/2,\theta}(n)\mathfrak{g}(n)}{n^{1/2+\mathrm{i}t}} \Bigg|^2 \tag{2.31}$$

Finally, on the event $\mathcal{A}(T) \cap \{|\mathcal{P}_{1-\varepsilon}(\frac{1}{2} + \mathrm{i}\tau)| > 100 \log\log T\}$,

$$\left|(\zeta \cdot e^{-\mathcal{P}_{|\theta|}})(\tfrac{1}{2} + \mathrm{i}\tau)\right|^\beta \leq (\log T)^4 \cdot (1 + |\zeta(\tfrac{1}{2} + \mathrm{i}\tau)|^2) \cdot \left|\frac{\mathcal{P}_{1-\varepsilon}(\frac{1}{2} + \mathrm{i}\tau)}{100 \log\log T}\right|^{2\ell} \tag{2.32}$$

for any $\ell \geq 1$. Choose $\ell = 50\lfloor \log\log T \rfloor$. By taking the expectation over the respective events in (2.30), (2.31) and (2.32), and then summing the three equations, the result follows from (2.29) and by applying Lemma 2.4, Lemma 2.5 and Lemma 2.6 to bound the expectations. □

2.2. *Discretization*

The analysis of large values of zeta on an interval can often be reduced to the analysis on a discrete set of points at a distance of roughly $(\log T)^{-1}$ of each other. This can be proved for the maximum using the functional equation for zeta, see for example Lemma 2.2 in Farmer, Gonek and Hughes (2007). We will need a more elaborate variant for general Dirichlet polynomials. This formulation will also allow us to directly derive the result for the maximum from the one for the moments.

Lemma 2.8. *Let $\delta > 0$, $\theta > -1$ and $\beta \geq 1$ be given. Let D be a Dirichlet polynomial of length $T^{1+\delta}$ with $\delta \geq 0$. Then, for all $A > 0$ and $T \leq t \leq 2T$,*

$$\begin{aligned} \sup_{|h| \leq (\log T)^\theta} |D(\tfrac{1}{2} + \mathrm{i}t + \mathrm{i}h)|^\beta \ll_A \sum_{|k| \leq 2(\log T)^{1+\theta}} \left|D\big(\tfrac{1}{2} + \mathrm{i}t + \tfrac{2\pi \mathrm{i}k}{(2+3\delta)\log T}\big)\right|^\beta \\ + \sum_{|k| > 2(\log T)^{1+\theta}} \left|D\big(\tfrac{1}{2} + \mathrm{i}t + \tfrac{2\pi \mathrm{i}k}{(2+3\delta)\log T}\big)\right|^\beta \cdot \frac{1}{1 + |k|^A}. \end{aligned} \tag{2.33}$$

Proof. Let V be a smooth compactly supported function with $V(x) = 1$ for $0 \leq x \leq 1+\delta$ and compactly supported in $[-\delta, 1+2\delta]$. Let $G(x) = V(2\pi x/\log T)$, so that $G\big(\frac{1}{2\pi}\log n\big) = 1$ for $1 \leq n \leq T^{1+\delta}$. Moreover,

$$\widehat{G}(x) = \tfrac{\log T}{2\pi} \cdot \widehat{V}\big(\tfrac{x \log T}{2\pi}\big), \tag{2.34}$$

and $\widehat{G}(x)D(\frac{1}{2}+\mathrm{i}t+\mathrm{i}h+\mathrm{i}x)$ is a function of exponential type in x with exponent $\leq (2+3\delta)\log T$. Now, consider

$$\sum_{k\in\mathbb{Z}} D\big(\tfrac{1}{2}+\mathrm{i}t+\tfrac{2\pi \mathrm{i}k}{(2+3\delta)\log T}\big)\widehat{G}\big(\tfrac{2\pi k}{(2+3\delta)\log T}-h\big). \tag{2.35}$$

By Poisson summation, the above is equal to

$$\sum_{\ell\in\mathbb{Z}} \int_{\mathbb{R}} D\big(\tfrac{1}{2}+\mathrm{i}t+\mathrm{i}h+\tfrac{2\pi \mathrm{i}x}{(2+3\delta)\log T}\big)\widehat{G}\big(\tfrac{2\pi x}{(2+3\delta)\log T}-h\big)e^{-2\pi \mathrm{i}\ell x}\mathrm{d}x. \tag{2.36}$$

By a change of variable, this is equal to

$$(2+3\delta)\cdot\frac{\log T}{2\pi}\sum_{\ell\in\mathbb{Z}}\int_{\mathbb{R}} D\big(\tfrac{1}{2}+\mathrm{i}t+\mathrm{i}x\big)\widehat{G}(x-h)e^{-\mathrm{i}x\ell(2+3\delta)\log T}\mathrm{d}x. \tag{2.37}$$

By Paley-Wiener, since $\widehat{G}(x)D(\frac{1}{2}+\mathrm{i}t+\mathrm{i}h+\mathrm{i}x)$ is, as a function of x in the complex plane, bounded by $\exp((2+3\delta)\log T\cdot|x|)$, it follows that all the terms with $\ell\neq 0$ are equal to zero. Finally, the term $\ell = 0$ is equal to $D(\frac{1}{2}+\mathrm{i}t+\mathrm{i}h)$ since $G(\frac{1}{2\pi}\log n) = 1$ for $1\leq n\leq T^{1+\delta}$. It follows that

$$D(\tfrac{1}{2}+\mathrm{i}t+\mathrm{i}h) = \frac{1}{2+3\delta}\sum_{k\in\mathbb{Z}} D\big(\tfrac{1}{2}+\mathrm{i}t+\tfrac{2\pi \mathrm{i}k}{(2+3\delta)\log T}\big)\widehat{V}\big(\tfrac{k}{2+3\delta}-\tfrac{h\log T}{2\pi}\big). \tag{2.38}$$

Taking absolute values and applying Holder's inequality, we obtain

$$\begin{aligned}|D(\tfrac{1}{2}+\mathrm{i}t+\mathrm{i}h)| \leq \Big(&\frac{1}{2+3\delta}\sum_{k\in\mathbb{Z}}\big|D\big(\tfrac{1}{2}+\mathrm{i}t+\tfrac{2\pi \mathrm{i}k}{(2+3\delta)\log T}\big)\big|^{\beta}\cdot\big|\widehat{V}\big(\tfrac{k}{2+3\delta}-\tfrac{h\log T}{2\pi}\big)\big|\Big)^{1/\beta}\\ &\times\Big(\frac{1}{2+3\delta}\sum_{k\in\mathbb{Z}}\big|\widehat{V}\big(\tfrac{k}{2+3\delta}-\tfrac{h\log T}{2\pi}\big)\big|\Big)^{1-1/\beta}.\end{aligned} \tag{2.39}$$

Using the rapid decay of $\widehat{V}$, we conclude that, for $\beta\geq 1$,

$$|D(\tfrac{1}{2}+\mathrm{i}t+\mathrm{i}h)|^{\beta} \ll \sum_{k\in\mathbb{Z}}\big|D\big(\tfrac{1}{2}+\mathrm{i}t+\tfrac{2\pi \mathrm{i}k}{(2+3\delta)\log T}\big)\big|^{\beta}\cdot\big|\widehat{V}\big(\tfrac{k}{2+3\delta}-\tfrac{h\log T}{2\pi}\big)\big|. \tag{2.40}$$

Taking a supremum over $|h|\leq(\log T)^{\theta}$, and using the rapid decay of $\widehat{V}$, we get (2.33). □

As an immediate corollary, we find:

Corollary 2.9. *Let $\theta > -1$ and $\beta \geq 1$ be given. Then, for any $A, B > 0$ and $T \leq t \leq 2T$,*

$$\begin{aligned}\max_{|h|\leq(\log T)^{\theta}}|\zeta(\tfrac{1}{2}+\mathrm{i}t+\mathrm{i}h)|^{\beta} \ll_{A,B} &\sum_{|k|\leq 2(\log T)^{1+\theta}}\big|\zeta\big(\tfrac{1}{2}+\mathrm{i}t+\tfrac{\pi \mathrm{i}k}{\log T}\big)\big|^{\beta}\\ &+\sum_{|k|>2(\log T)^{1+\theta}}\big|\zeta\big(\tfrac{1}{2}+\mathrm{i}t+\tfrac{\pi \mathrm{i}k}{\log T}\big)\big|^{\beta}\cdot\frac{1}{1+|k|^{A}}+T^{-B}.\end{aligned} \tag{2.41}$$

Proof. Notice that, for any $A > 0$,

$$\sum_{n \le T} \frac{1}{n^{1/2+\mathrm{i}t}} \cdot \Big(1 - \frac{n}{T}\Big)^A = \zeta(\tfrac{1}{2} + \mathrm{i}t) + \mathcal{O}_A(T^{-A/2}). \tag{2.42}$$

We apply Lemma 2.8 to conclude. □

The above results imply two more corollaries.

Corollary 2.10. *For any $A \ge 0$ integer,*

$$\mathbb{P}\Big(\max_{|h|<(\log T)^A} |\zeta(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)| > 2^A (\log T)^{2+A}\Big) \ll \frac{1}{\log T} \cdot 2^{-A}. \tag{2.43}$$

Proof. Without loss of generality, we can assume that $A \le \log T / \log\log T$. Then, by applying Chebyshev's inequality and Corollary 2.9, the probability in (2.43) is bounded above by

$$\begin{aligned} &2^{-2A}(\log T)^{-4-2A}\, \mathbb{E}\Big[\max_{|h|\le(\log T)^A} |\zeta(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)|^2\Big] \\ &\ll 2^{-2A}(\log T)^{-4-2A} \sum_{|k|\le 2(\log T)^{1+A}} \mathbb{E}\Big[\big|\zeta\big(\tfrac{1}{2} + \mathrm{i}\tau + \tfrac{\pi \mathrm{i}k}{\log T}\big)\big|^2\Big] \\ &\quad + 2^{-2A}(\log T)^{-4-2A} \sum_{|k|> 2(\log T)^{1+A}} \mathbb{E}\Big[\big|\zeta\big(\tfrac{1}{2} + \mathrm{i}\tau + \tfrac{\pi \mathrm{i}k}{\log T}\big)\big|^2\Big] \cdot \frac{1}{1+|k|^{100}} + T^{-101}. \end{aligned} \tag{2.44}$$

We bound the first expectation using the standard second moment bound. We bound the second expectation by enlarging the integration to $|t| \le T|k|$ and then applying the second moment bound, i.e.

$$\mathbb{E}\Big[\big|\zeta\big(\tfrac{1}{2} + \mathrm{i}\tau + \tfrac{\pi \mathrm{i}k}{\log T}\big)\big|^2\Big] \le |k| \cdot \frac{1}{T|k|} \int_{|t|\le T|k|} \big|\zeta(\tfrac{1}{2} + \mathrm{i}t)\big|^2 \mathrm{d}t \ll |k| \cdot \log(T|k|). \tag{2.45}$$

We conclude that the right-hand side of (2.44) is

$$\ll 2^{-2A}(\log T)^{-2-A} \ll \frac{1}{\log T} \cdot 2^{-A}. \tag{2.46}$$

This ends the proof. □

Corollary 2.11. *Let $0 \ge \theta > -1$ be given. Then, the event*

$$\begin{aligned} &\max_{|h|\le(\log T)^\theta} \Big|(\zeta \cdot e^{-\mathcal{P}_{|\theta|}})(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)\Big|^2 \\ &\ll \sum_{|k|\le 2(\log T)^{1+\theta}} \Big|(\zeta \cdot e^{-\mathcal{P}_{|\theta|}})\big(\tfrac{1}{2} + \mathrm{i}\tau + \tfrac{2\pi \mathrm{i}k}{(2+2\varepsilon)\log T}\big)\Big|^2 + \mathrm{o}(1) \end{aligned} \tag{2.47}$$

has $\mathbb{P}$-probability $1 - \mathrm{o}(1)$.

Proof. Define the event

$$\mathcal{B}(T) := \left\{ \begin{array}{l} \max_{|h|\le(\log T)^\theta} \log|\zeta(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)| \asymp \log\log T \\ \max_{|h|\le(\log T)^\theta} \mathcal{P}_{|\theta|}(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h) \le 2\log\log T \end{array} \right\}. \tag{2.48}$$

By Corollary 2.10, the result of Section 3.4 (which is logically independent of this section) and Lemma 2.3, we have $\mathbb{P}(\mathcal{B}(T)) = 1 - \mathrm{o}(1)$. Furthermore, for all $|h| \leq (\log T)^\theta$ and $\tau \in \mathcal{B}(T)$,

$$\Big|\sum_{n \leq T} \frac{1}{n^{1/2+\mathrm{i}\tau+\mathrm{i}h}}\Big| = |\zeta(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)| + \mathcal{O}(T^{-1/2}) \asymp |\zeta(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)|, \tag{2.49}$$

and

$$\begin{aligned} \Big|\sum_{\substack{\Omega(n) \leq 10\lfloor \log\log T \rfloor \\ p | n \implies \log p \leq \log^{|\theta|} T}} \frac{(-1)^{\Omega(n)} \mathfrak{g}(n)}{n^{1/2+\mathrm{i}\tau+\mathrm{i}h}}\Big| &= \Big|e^{-\mathcal{P}_{|\theta|}(\frac{1}{2}+\mathrm{i}\tau+\mathrm{i}h)}\Big| + \mathcal{O}\big((\log T)^{-10}\big) \\ &\asymp \Big|e^{-\mathcal{P}_{|\theta|}(\frac{1}{2}+\mathrm{i}\tau+\mathrm{i}h)}\Big|. \end{aligned} \tag{2.50}$$

Combining (2.49) and (2.50), we get for all $\tau \in \mathcal{B}(T)$,

$$\big|(\zeta \cdot e^{-\mathcal{P}_{|\theta|}})(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)\big| \asymp |D(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)|, \tag{2.51}$$

with D a Dirichlet polynomial of length $\ll T^{1+\varepsilon}$ for every fixed $\varepsilon > 0$.

Applying Lemma 2.8, it follows that there exists a subset $\mathcal{B}_0(T)$ of $\mathcal{B}(T)$ of probability $1 - \mathrm{o}(1)$ such that for all $\tau \in \mathcal{B}_0(T)$

$$\max_{|h| \leq (\log T)^\theta} |D(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)|^2 \ll \sum_{|k| \leq 2(\log T)^{1+\theta}} \big|D\big(\tfrac{1}{2} + \mathrm{i}\tau + \tfrac{2\pi \mathrm{i}k}{(2+2\varepsilon)\log T}\big)\big|^2 + \mathrm{o}(1). \tag{2.52}$$

Together with (2.51), this concludes the proof. □

2.3. Proofs of the upper bounds

2.3.1. The case of $\theta \geq 0$

Proof of Theorem 1.2 for $\theta \geq 0$. By Chernoff's inequality, for any $\beta > 0$, we have

$$\begin{aligned} &\mathbb{P}\Big(\max_{|h| \leq (\log T)^\theta} |\zeta(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)| > (\log T)^{m(\theta)+\varepsilon}\Big) \\ &\qquad \ll (\log T)^{-\beta m(\theta) - \beta\varepsilon} \cdot \mathbb{E}\Big[\max_{|h| \leq (\log T)^\theta} |\zeta(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)|^\beta\Big]. \end{aligned} \tag{2.53}$$

By Corollary 2.9, the above is

$$\ll (\log T)^{-\beta m(\theta) - \beta\varepsilon + 1 + \theta} \cdot \mathbb{E}\Big[|\zeta(\tfrac{1}{2} + \mathrm{i}\tau)|^\beta\Big]. \tag{2.54}$$

Choosing $\beta = 2m(\theta)$ and applying Proposition 2.2 if $\theta \leq 3$ and Proposition 2.1 if $\theta > 3$, the claim follows. □

Proof of Theorem 1.1 for $\theta \geq 0$. The case of $\beta \leq 2\sqrt{1+\theta}$ is immediate from the first moment bound, since

$$\begin{aligned} &\mathbb{P}\Big(\int_{|h| \leq (\log T)^\theta} |\zeta(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)|^\beta \mathrm{d}t \geq (\log T)^{f_\theta(\beta)+\varepsilon}\Big) \\ &\qquad \ll (\log T)^{-f_\theta(\beta)-\varepsilon} (\log T)^\theta \cdot \mathbb{E}\big[|\zeta(\tfrac{1}{2} + \mathrm{i}\tau)|^\beta\big] \ll (\log T)^{-\varepsilon/2}, \end{aligned} \tag{2.55}$$

by Proposition 2.2 for $\theta \leq 3$ and by Proposition 2.1 for $\theta > 3$.

It remains to deal with the case of $\beta > 2\sqrt{1+\theta}$. By the same reasoning as above, for any $0 \le j \le M$ and integer $M \ge 1$,

$$\begin{aligned} &\mathbb{P}\Big(\mathrm{Leb}\Big\{|h| \le (\log T)^{\theta} : |\zeta(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)| > (\log T)^{(j/M)m(\theta)}\Big\} \ge (\log T)^{-(j/M)^2 m(\theta)^2 + \theta + \varepsilon}\Big) \\ &\quad \le (\log T)^{(j/M)^2 m(\theta)^2 - \varepsilon} \cdot (\log T)^{\theta} \cdot \mathbb{P}\Big(|\zeta(\tfrac{1}{2} + \mathrm{i}\tau)| > (\log T)^{(j/M)m(\theta)}\Big). \end{aligned}$$

By Chernoff's inequality this is

$$\ll (\log T)^{(j/M)^2 m(\theta)^2 - \varepsilon} \cdot (\log T)^{-\beta (j/M) m(\theta)} \cdot \mathbb{E}\Big[|\zeta(\tfrac{1}{2} + \mathrm{i}\tau)|^{\beta}\Big]. \tag{2.56}$$

Using Proposition 2.2 for $\theta \le 3$ and Proposition 2.1 for $\theta > 3$ and choosing $\beta = 2(j/M)m(\theta)$, we conclude that this is $\ll (\log T)^{-\varepsilon/2}$. In particular, it follows that if $\beta > 2\sqrt{1+\theta}$, then in probability,

$$\int_{|h| \le (\log T)^{\theta}} |\zeta(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)|^{\beta} \mathrm{d}h \ll \sum_{0 \le j \le M} (\log T)^{\beta((j+1)/M)m(\theta)} \cdot (\log T)^{-(j/M)^2 m(\theta)^2 + \theta + \varepsilon}. \tag{2.57}$$

The above bound corresponds to partitioning the integral according to the value distribution of the integrand. Since $\beta > 2\sqrt{1+\theta} \ge m(\theta)$, the last term $j = M$ dominates and in particular the above is bounded by

$$\ll (\log T)^{\beta m(\theta) - m(\theta)^2 + \theta + 2\varepsilon} = (\log T)^{\beta m(\theta) - 1 + 2\varepsilon}, \tag{2.58}$$

provided that M is chosen sufficiently large. □

2.3.2. The case of $\theta < 0$

Proof of Theorem 1.2 for $\theta < 0$. We notice that

$$\begin{aligned} &\mathbb{P}\Big(\max_{|h| \le (\log T)^{\theta}} |\zeta(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)| > (\log T)^{m(\theta) + \varepsilon}\Big) \\ &\le \mathbb{P}\Big(\max_{|h| \le (\log T)^{\theta}} \big|(\zeta \cdot e^{-\mathcal{P}_{|\theta|}})(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)\big| > (\log T)^{m(\theta) + \varepsilon/2}\Big) \\ &\quad + \mathbb{P}\Big(\max_{|h| \le (\log T)^{\theta}} \Big|e^{\mathcal{P}_{|\theta|}(\frac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)}\Big| > (\log T)^{\varepsilon/2}\Big) \end{aligned} \tag{2.59}$$

By Lemma 2.3 the last term is $\mathrm{o}(1)$ as $T \to \infty$. Let

$$\widetilde{\mathcal{A}}(T) := \Big\{\max_{|h| \le (\log T)^{\theta}} \big|\mathcal{P}_{|\theta|}(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)\big| \le 2 \log\log T\Big\}. \tag{2.60}$$

By Lemma 2.3, the probability of $\widetilde{\mathcal{A}}(T)$ is $1 - \mathrm{o}(1)$. We let $\mathcal{A}_0(T)$ denote the subset of $\widetilde{\mathcal{A}}(T)$ for which the conclusion of Corollary 2.11 holds, so that the probability of $\mathcal{A}_0(T)$ is $1 - \mathrm{o}(1)$. Then, by Chebyshev's inequality, we have

$$\begin{aligned} &\mathbb{P}\Big(\Big\{\max_{|h| \le (\log T)^{\theta}} \big|(\zeta \cdot e^{-\mathcal{P}_{|\theta|}})(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)\big| > (\log T)^{m(\theta) + \varepsilon/2}\Big\} \cap \mathcal{A}_0(T)\Big) \\ &\le (\log T)^{-2m(\theta) - \varepsilon} \cdot \mathbb{E}\Big[\max_{|h| \le (\log T)^{\theta}} \big|(\zeta \cdot e^{-\mathcal{P}_{|\theta|}})(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)\big|^2 \mathbf{1}_{\mathcal{A}_0(T)}\Big]. \end{aligned} \tag{2.61}$$

By Corollary 2.11, and since $m(\theta) = 1 + \theta$, this is

$$\ll (\log T)^{-(1+\theta)-\varepsilon} \cdot \mathbb{E}\Big[\big|(\zeta \cdot e^{-\mathcal{P}_{|\theta|}})(\tfrac{1}{2} + \mathrm{i}\tau)\big|^2 \mathbf{1}_{\widetilde{\mathcal{A}}(T)}\Big]. \tag{2.62}$$

By Proposition 2.7, this is

$$\ll (\log T)^{-(1+\theta)-\varepsilon} \cdot (\log T)^{(1+\theta)+\varepsilon/2} \ll (\log T)^{-\varepsilon/2}, \tag{2.63}$$

as needed. □

Proof of Theorem 1.1 for $\theta < 0$. We begin with the case $\beta < 2$. Similarly to (2.59), we have

$$\begin{aligned} &\mathbb{P}\Big(\int_{|h|\le(\log T)^\theta} |\zeta(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)|^\beta \mathrm{d}h \ge (\log T)^{f_\theta(\beta)+\varepsilon}\Big) \\ &\quad\le \mathbb{P}\Big(\int_{|h|\le(\log T)^\theta} \big|(\zeta \cdot e^{-\mathcal{P}_{|\theta|}})(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)\big|^\beta \mathrm{d}h > (\log T)^{f_\theta(\beta)+\varepsilon/2}\Big) + \mathrm{o}(1). \end{aligned} \tag{2.64}$$

As in (2.60), $\mathbb{P}(\widetilde{\mathcal{A}}(T)) = 1 - \mathrm{o}(1)$, and by Markov's inequality, we have

$$\begin{aligned} &\mathbb{P}\Big(\Big\{\int_{|h|\le(\log T)^\theta} \big|(\zeta \cdot e^{-\mathcal{P}_{|\theta|}})(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)\big|^\beta \mathrm{d}h > (\log T)^{f_\theta(\beta)+\varepsilon/2}\Big\} \cap \widetilde{\mathcal{A}}(T)\Big) \\ &\quad\ll (\log T)^{-f_\theta(\beta)-\varepsilon/2} \cdot (\log T)^\theta \cdot \mathbb{E}\Big[\big|(\zeta \cdot e^{-\mathcal{P}_{|\theta|}})(\tfrac{1}{2} + \mathrm{i}\tau)\big|^\beta \mathbf{1}_{\widetilde{\mathcal{A}}(T)}\Big]. \end{aligned} \tag{2.65}$$

By Proposition 2.7, the above is

$$\ll (\log T)^{-(\beta^2/4)(1+\theta)-\varepsilon/2} \cdot (\log T)^{(\beta^2/4)\cdot(1+\theta)+\varepsilon/4} \ll (\log T)^{-\varepsilon/4}, \tag{2.66}$$

and the claim follows.

Now it remains to handle the case of $\beta \ge 2$. This proceeds in the same way as in the proof of Theorem 1.1 in the case of $\theta > 0$ upon collecting the distribution information provided by the above estimate for moments with $\beta < 2$ and the estimate for the local maximum. We leave the details to the interested reader. □

3. Lower bounds

In this section, we prove:

Proposition 3.1. *Let $\theta > -1$ and $\varepsilon > 0$ be given. Then,*

$$\mathbb{P}\Big(\max_{|h|\le(\log T)^\theta} |\zeta(1/2 + \mathrm{i}\tau + \mathrm{i}h)| > (\log T)^{m(\theta)-\varepsilon}\Big) = 1 - \mathrm{o}(1). \tag{3.1}$$

Proposition 3.2. *Let $\theta > -1$ and $\varepsilon > 0$ be given. Then,*

$$\mathbb{P}\Big(\int_{-(\log T)^\theta}^{(\log T)^\theta} |\zeta(1/2 + \mathrm{i}\tau + \mathrm{i}h)|^\beta \mathrm{d}h > (\log T)^{f_\theta(\beta)-\varepsilon}\Big) = 1 - \mathrm{o}(1). \tag{3.2}$$

The lower bound for the maximum will be an easy consequence of the lower bound for the moments. As for the upper bounds, the idea is to approximate zeta by an appropriate Dirichlet polynomial. This can be done with good precision off-axis. In Section 3.1, we reduce the problem to ζ on the line

$$\sigma_0 = \frac{1}{2} + \frac{(\log T)^{3/(2K)}}{\log T}. \tag{3.3}$$

We shall assume that $K > 3$ is large enough so that $3/(2K) - 1 - \theta < 0$. Moreover, K will also be assumed to be large enough depending on β and ε. The approximation to a Dirichlet polynomial is then shown in Section 3.2. The lower bound for the moments of the Dirichlet polynomials is proved in Section 3.3 using Kistler's multiscale second moment method. Finally, the two propositions above are proved in Section 3.4.

3.1. Reduction off-axis

In Arguin et al. (2018), the maximum on a short interval of the critical line was compared to the one on a short interval away from the critical line by exploiting the analyticity of ζ away from its pole. More precisely, a value off-axis can be seen as an average of zeta over the critical line weighed by the corresponding Poisson kernel. This approach could also be used in the case of the moments by using the subharmonicity of the function $z \mapsto |z|^\beta$. We choose to apply a different method based on the following convexity theorem of Gabriel, which handles error terms more efficiently.

Proposition 3.3 (Theorem 2 in Gabriel (1927))**.** *Let F be a complex valued function which is regular in the strip $\alpha < Re\, z < \beta$ and continuous for $\alpha \le Re\, z \le \beta$. Suppose that $|F(z)|$ tends to zero uniformly as $|Im\, z| \to \infty$, uniformly for $\alpha \le Re\, z \le \beta$. Then, for any $\gamma \in [\alpha, \beta]$, and any $k > 0$,*

$$I(\gamma) \le I(\alpha)^{(\beta-\gamma)/(\beta-\alpha)} \cdot I(\beta)^{(\gamma-\alpha)/(\beta-\alpha)} \tag{3.4}$$

where

$$I(\sigma) := \int_{\mathbb{R}} |F(\sigma + \mathrm{i}t)|^k \mathrm{d}t. \tag{3.5}$$

This theorem has the following useful consequence.

Corollary 3.4. *Let F be a complex valued function which is regular in the strip $\frac{1}{2} \le Re\, z$. Suppose that $|F(z)|$ tends to zero uniformly as $|Im\, z| \to \infty$. Given $k > 0$ real, let*

$$I(\sigma) := \int_{\mathbb{R}} |F(\sigma + \mathrm{i}t)|^k \mathrm{d}t. \tag{3.6}$$

Suppose that $I(\sigma) \to 0$ as $\sigma \to \infty$. Then, for any $\sigma > \frac{1}{2}$,

$$I(\sigma) \le I(\tfrac{1}{2}). \tag{3.7}$$

Proof. Let $\sigma^\star$ be such that,

$$I(\sigma^\star) = \sup_{\sigma \ge 1/2} I(\sigma). \tag{3.8}$$

Note that because of the assumption that $I(\sigma) \to 0$ as $\sigma \to \infty$, the above $\sigma^\star$ has a finite value. Let $\varepsilon > 0$ be given. If $\sigma^\star = \frac{1}{2}$ then we are done. If $\sigma^\star \ne \frac{1}{2}$, then by Proposition 3.3 applied with $\gamma = \sigma^\star$ and $\beta = \sigma^\star + \varepsilon$ and $\alpha = \frac{1}{2}$, we get,

$$I(\sigma^\star) \le I(\tfrac{1}{2})^\alpha \cdot I(\sigma^\star + \varepsilon)^\beta, \tag{3.9}$$

with $\alpha, \beta > 0$ such that $\alpha + \beta = 1$.

Therefore, by definition of $\sigma^\star$,

$$I(\sigma^\star) \le I(\tfrac{1}{2})^\alpha \cdot I(\sigma^\star)^\beta, \tag{3.10}$$

and hence $I(\sigma^\star)^\alpha \le I(\frac{1}{2})^\alpha$. Since $\alpha > 0$, we get $I(\sigma^\star) \le I(\frac{1}{2})$. By definition of $\sigma^\star$, the claim follows. ☐

We now construct a special analytic approximation for the indicator function of an interval.

Lemma 3.5. *Let $0 \le \Delta \le L$ and $\varepsilon > 0$ be given. There exists an entire function $\Phi_{\Delta,L}(z)$ such that, for $z = \sigma + \mathrm{i}v$ with $\sigma \ge \frac{1}{2}$ and $v \in \mathbb{R}$,*

1. *For $K > 1 + \varepsilon$ and $|v| > KL$, uniformly in $\sigma \ge \frac{1}{2}$, $\Phi_{\Delta,L}(z) \ll_A ((K-1)\Delta)^{-A}$.*
2. *For any $|v| \le (1-\varepsilon)L$, $|\Phi_{\Delta,L}(z)| = 1 + \mathcal{O}_A(\Delta^{-A}) + \mathcal{O}((\sigma - \frac{1}{2})\Delta^2/L)$.*
3. *For any $(1-\varepsilon)L \le |v| \le (1+\varepsilon)L$, $|\Phi_{\Delta,L}(z)| \ll 1 + (\sigma - \frac{1}{2})\Delta^2/L$.*
4. *$\Phi_{\Delta,L}(z) \to 0$ uniformly in v as $\sigma \to \infty$.*

Proof. Let V be a smooth function, compactly supported in $[0,\infty)$ and such that $V(1) = 1$. Given a parameter $T > 0$ and given $z \in \mathbb{C}$ with $\mathrm{Re}\, z \ge \frac{1}{2}$ and $u \in \mathbb{R}$, consider the following function :

$$\delta_{1/T}(z) = \frac{1}{T}\int_0^\infty e^{-2\pi(z-1/2)x} \cdot V\big(\tfrac{x}{T}\big)\mathrm{d}x. \tag{3.11}$$

Then $\delta_{1/T}(z)$ defines an entire function of exponential type. By integration by parts, we see that $\delta_{1/T}(z) \ll_A (1 + |z - \frac{1}{2}|T)^{-A}$ for any $A > 0$ and uniformly in $\mathrm{Re}\, z \ge \frac{1}{2}$. Therefore, we can think of $\delta_{1/T}(z)$ as localizing to $z = \frac{1}{2} + \mathcal{O}(1/T)$. Furthermore, notice that if $z = \frac{1}{2} + \mathrm{i}v$ and $u \in \mathbb{R}$, then

$$\delta_{1/T}(z - \mathrm{i}u) = \widehat{V}(T(v-u)), \tag{3.12}$$

and for $z = \sigma + \mathrm{i}v$, we have by a Taylor expansion of the exponential,

$$\begin{aligned}\delta_{1/T}(z - \mathrm{i}u) &= \frac{1}{T}\int_0^\infty e^{-2\pi(\sigma - \frac{1}{2} + \mathrm{i}(v-u))x} \cdot V\big(\tfrac{x}{T}\big)\mathrm{d}x \\ &= \frac{1}{T}\int_0^\infty e^{-2\pi\mathrm{i}(v-u)x} \cdot \Big(1 + \mathcal{O}\big((\sigma - \tfrac{1}{2})T\big)\Big) \cdot V\big(\tfrac{x}{T}\big)\mathrm{d}x \\ &= \widehat{V}(T(v-u)) + \mathcal{O}\big((\sigma - \tfrac{1}{2})T\big).\end{aligned} \tag{3.13}$$

Finally, we also recall that for $z = \sigma + \mathrm{i}v$ with $\frac{1}{2} \le \sigma$, we have

$$|\delta_{1/T}(z - \mathrm{i}u)| \ll_A \frac{1}{1 + (|u-v|T)^A}. \tag{3.14}$$

Thus, set

$$\Phi_{\Delta,L}(z) := \frac{\Delta}{L}\int_{-L}^{L} e^{-2\pi\mathrm{i}u(\Delta/L)} \cdot \delta_{L/\Delta}(z - \mathrm{i}u)\mathrm{d}u. \tag{3.15}$$

We will now describe some of the features of this function. Write $z = \sigma + \mathrm{i}v$ with $\sigma \ge \frac{1}{2}$. Using the bound (3.14), we see that if $|v| > KL$ with $K > (1+\varepsilon)$ and $\frac{1}{2} \le \sigma$ then

$$\Phi_{\Delta,L}(z) \ll_A \frac{\Delta}{L}\int_{-L}^{L} \frac{1}{1 + (\Delta/L \cdot |u-v|)^A}\,\mathrm{d}u \ll_A (K-1)^{-A}\Delta^{1-A}. \tag{3.16}$$

This gives the first claim.

If $|v| \leq (1+\varepsilon)L$, then by (3.13), we have

$$\begin{aligned}\Phi_{\Delta,L}(u) &= \frac{\Delta}{L}\int_{-L}^{L} e^{-2\pi \mathrm{i}u(\Delta/L)} \cdot \delta_{L/\Delta}(z-\mathrm{i}u)\mathrm{d}u \\ &= \frac{\Delta}{L}\int_{-L}^{L} e^{-2\pi \mathrm{i}u(\Delta/L)} \cdot \widehat{V}\big(\tfrac{\Delta}{L}\cdot(v-u)\big)\mathrm{d}u + \mathcal{O}\big((\sigma-\tfrac{1}{2})\Delta^2/L\big). \end{aligned} \tag{3.17}$$

In particular, it follows that if $\frac{1}{2} \leq \sigma$ and $|v| \leq (1-\varepsilon)\cdot L$, then due to the rapid decay of $\widehat{V}$,

$$\begin{aligned}\Phi_{\Delta,L}(u) &= e^{-2\pi \mathrm{i}v(\Delta/L)}\int_{v\Delta/L-\Delta}^{v\Delta/L+\Delta} e^{2\pi \mathrm{i}u}\cdot \widehat{V}(u)\mathrm{d}u + \mathcal{O}\big((\sigma-\tfrac{1}{2})\Delta^2/L\big) \\ &= e^{-2\pi \mathrm{i}v(\Delta/L)} + \mathcal{O}_A(\Delta^{-A}) + \mathcal{O}\big((\sigma-\tfrac{1}{2})\Delta^2/L\big), \end{aligned} \tag{3.18}$$

by Fourier inversion and the assumption that $V(1) = 1$. This proves the second claim. If $\frac{1}{2} \leq \sigma \ll 1$ and $|v| \leq 2L$, then we have the bound

$$|\Phi_{\Delta,L}(u)| \ll \int_{\mathbb{R}} |\widehat{V}(u)|\mathrm{d}u + \mathcal{O}\big((\sigma-\tfrac{1}{2})\Delta^2/L\big) \tag{3.19}$$

and this proves the third claim. Finally, we also notice that $\delta_L(z-\mathrm{i}u) \to 0$ uniformly as $\sigma \to \infty$ and that likewise,

$$\Phi_{\Delta,L}(z-\mathrm{i}u) \to 0, \tag{3.20}$$

uniformly in $u \in \mathbb{R}$, as $\sigma \to \infty$, which proves the last claim. □

The following proposition relates the moments off and on axis.

Proposition 3.6. *Let $\theta > -1$, $\beta > 0$, $\varepsilon > 0$, $T \geq 10^9$. Then, for all $\frac{1}{2} \leq \sigma \leq \frac{1}{2} + (\log T)^{\theta - 3\varepsilon}$,*

$$\mathbb{P}\Big(\int_{-(\log T)^\theta}^{(\log T)^\theta} |\zeta(\sigma+\mathrm{i}t+\mathrm{i}u)|^\beta \mathrm{d}u \ll_{\theta,\varepsilon,\beta} \int_{-2(\log T)^\theta}^{2(\log T)^\theta} |\zeta(\tfrac{1}{2}+\mathrm{i}t+\mathrm{i}u)|^\beta \mathrm{d}u + \frac{1}{(\log T)^7}\Big) = 1 - \mathrm{o}(1). \tag{3.21}$$

Proof. Let

$$D(\sigma+\mathrm{i}t) := \sum_{n\leq T} \frac{1}{n^{\sigma+\mathrm{i}t}}\cdot\Big(1-\frac{n}{T}\Big)^A, \tag{3.22}$$

with $A > 100$ fixed. Then, for $T \leq t \leq 2T$ and $\sigma \geq \frac{1}{2}$,

$$\zeta(\sigma+\mathrm{i}t) = D(\sigma+\mathrm{i}t) + \mathcal{O}_A(T^{-A/2}). \tag{3.23}$$

Consider now,

$$I(\sigma) := \int_{\mathbb{R}} |D(\sigma+\mathrm{i}t+\mathrm{i}u)|^\beta \cdot |\Phi_{\Delta,L}(\sigma+\mathrm{i}u)|^\beta \mathrm{d}u, \tag{3.24}$$

with $\Delta = (\log T)^\varepsilon$ and $L = 2(\log T)^\theta$. Then, by Lemma 3.5 and Corollary 3.4, we have

$$\begin{aligned}&\int_{\mathbb{R}} |D(\sigma+\mathrm{i}t+\mathrm{i}u)|^\beta \cdot |\Phi_{\Delta,L}(\sigma+\mathrm{i}u)|^\beta \mathrm{d}u \\ &\qquad \ll \int_{\mathbb{R}} |D(\tfrac{1}{2}+\mathrm{i}t+\mathrm{i}u)|^\beta \cdot |\Phi_{\Delta,L}(\tfrac{1}{2}+\mathrm{i}u)|^\beta \mathrm{d}u. \end{aligned} \tag{3.25}$$

Now it remains to unsmooth. By Lemma 3.5 provided that $\sigma - \frac{1}{2} \leq (\log T)^{\theta - 3\varepsilon}$, we have

$$\int_{-(\log T)^\theta}^{(\log T)^\theta} |D(\sigma + \mathrm{i}t + \mathrm{i}u)|^\beta \mathrm{d}u \ll \int_{\mathbb{R}} |D(\sigma + \mathrm{i}t + \mathrm{i}u)|^\beta \cdot |\Phi_{\Delta,L}(\sigma + \mathrm{i}u)|^\beta \mathrm{d}u. \tag{3.26}$$

On the other hand, by Lemma 3.5, we have

$$\begin{aligned} &\int_{\mathbb{R}} |D(\tfrac{1}{2} + \mathrm{i}t + \mathrm{i}u)|^\beta \cdot |\Phi_{\Delta,L}(\tfrac{1}{2} + \mathrm{i}u)|^\beta \mathrm{d}u \\ &\quad \leq \int_{2I} |D(\tfrac{1}{2} + \mathrm{i}t + \mathrm{i}u)|^\beta \mathrm{d}u + \sum_{A \geq 0} \int_{\mathcal{U}_A} |D(\tfrac{1}{2} + \mathrm{i}t + \mathrm{i}u)|^\beta \cdot |\Phi_{\Delta,L}(\tfrac{1}{2} + \mathrm{i}u)|^\beta \mathrm{d}u, \end{aligned} \tag{3.27}$$

where $\mathcal{U}_A := \{2(\log T)^{\theta + A} \leq |u| \leq 2(\log T)^{\theta + A + 1}\}$. By Corollary 2.10 and a union bound,

$$\mathcal{S}(T) := \Big\{ \max_{A \in \mathbb{N} \cup \{0\}} \max_{|u| \leq (\log T)^A} |D(\tfrac{1}{2} + \mathrm{i}t + \mathrm{i}u)| \leq 2^A (\log T)^{2+A} \Big\}. \tag{3.28}$$

as probability $1 - \mathrm{o}(1)$. Moreover, by Lemma 3.5 for all $2(\log T)^{\theta + A} \leq |u|$, we have

$$|\Phi_{\Delta,L}(\tfrac{1}{2} + \mathrm{i}u)| \ll_{\theta,\varepsilon,\beta} (\log T)^{-4A(1+1/\beta)} \cdot (\log T)^{-(10\lceil\theta\rceil + 10)\cdot(1+1/\beta)}. \tag{3.29}$$

Therefore, for each $A \geq 0$,

$$\begin{aligned} &\int_{2(\log T)^{\theta+A} \leq |u| \leq 2(\log T)^{\theta+A+1}} |D(\tfrac{1}{2} + \mathrm{i}t + \mathrm{i}u)|^\beta \cdot |\Phi_{\Delta,L}(\tfrac{1}{2} + \mathrm{i}u)|^\beta \mathrm{d}u \\ &\qquad \ll_{\theta,\varepsilon,\beta} (\log T)^{(\beta+1)(\lceil\theta\rceil + A + 3)} \cdot 2^{A\beta} \cdot (\log T)^{-(\beta+1)\cdot(10\lceil\theta\rceil + 4A + 10)} \\ &\qquad \ll_{\theta,\varepsilon,\beta} (\log T)^{-(\beta+1)\cdot(A+7)}. \end{aligned} \tag{3.30}$$

Thus, the contribution of the sum over $A \geq 0$ in (3.27) is negligible. The claim follows. □

3.2. Mollification

This step is an adaptation of Section 4.2 of Arguin et al. (2018), which is itself based on the work of Radziwiłł and Soundararajan (2017). The treatment has to be slightly generalized to account for the width of the interval. The idea is to define a mollifier for the zeta function

$$M(s) = \sum_n \frac{\mu(n) a(n)}{n^s}. \tag{3.31}$$

Here μ denotes the Möbius function $\mu(n) = (-1)^{\omega(n)}$ if n is square-free where $\omega(n)$ is the number of distinct prime factors, and $\mu(n) = 0$ if n is non-square free. The term $a(n)$ equals 1 if all primes factors of n are smaller than

$$X = \exp((\log T)^{1 - 1/K}), \quad K \geq 2, \tag{3.32}$$

and if

$$\Omega(n) \leq 100 K e^{\theta \vee 0} \log\log T =: \nu, \tag{3.33}$$

where $\Omega(n)$ counts the number of prime factors of n (with multiplicity), and $a(n) = 0$ otherwise.

The goal of this section is to prove that M is an approximate inverse of ζ:

Lemma 3.7 (Mollification)**.** *Let $\theta > -1$ and $\varepsilon > 0$ be given. Then,*

$$\mathbb{P}\bigg(\max_{|\tau-u|\leq(\log T)^\theta} |M(\sigma_0+\mathrm{i}u)\zeta(\sigma_0+\mathrm{i}u)-1| > \varepsilon\bigg) = \mathrm{o}(1). \tag{3.34}$$

This was proved in the case $\theta = 0$ in Lemma 4.2 of Arguin et al. (2018). In particular, it also holds *verbatim* for $-1 < \theta < 0$, since the interval is just smaller. The particular choice of ν is needed for large θ to cancel the factor $(\log T)^\theta$ coming from the integral in the Sobolev inequality. It will also be used in the proof of Lemma 3.10. The proof of Lemma 3.7 also holds in the case $\theta > 0$ with slight modifications that we highlight. The key idea is the following L^2-control:

Lemma 3.8 (Adaptation of Lemma 4.2 of Arguin et al. (2018))**.** *Let $\theta > -1$ be given. Then,*

$$\mathbb{E}\Big[\big|M(\sigma_0+\mathrm{i}\tau)\zeta(\sigma_0+\mathrm{i}\tau)-1\big|^2\Big] \ll (\log T)^{-(100+e^\theta)}. \tag{3.35}$$

Proof. We only have to prove the case $\theta > 0$. The proof is exactly as in Arguin et al. (2018) with a new error term due to the choice of ν. The error appears after Equation (4.10) in Arguin et al. (2018) and is given by

$$(\log T)e^{-\nu}\prod_{p\leq X}(1+7p^{-1}). \tag{3.36}$$

The Euler product is bounded by $\ll (\log T)^7$ using Lemma A.1. Using this and the definition of ν yields

$$(\log T)e^{-\nu}\prod_{p\leq X}(1+7p^{-1}) \ll (\log T)^8 \cdot (\log T)^{-100K} \cdot (\log T)^{-e^\theta}. \tag{3.37}$$

Since $K \geq 2$, this gives the correct estimate. Note that the expression $\sum_{p>X}\log(1-p^{-2\sigma_0})^{-1}$ entering in the remainder of the proof of Lemma 4.2 is in fact $\ll \exp(-(\log T)^{\frac{1}{2K}}) \ll (\log T)^{-(100+e^\theta)}$ since

$$\sum_{p>X}\log(1-p^{-2\sigma_0})^{-1} \ll \sum_{p>X}p^{-2\sigma_0} \ll X^{-(\sigma_0-1/2)} = \exp(-(\log T)^{\frac{1}{2K}}). \tag{3.38}$$

This ends the proof. □

Proof of Lemma 3.7. Again, it suffices to address the case $\theta > 0$. A direct application of Sobolev's inequality (A.3) yields

$$\begin{aligned}&\mathbb{E}\Big[\max_{|\tau-u|\leq(\log T)^\theta}\big|M(\sigma_0+\mathrm{i}u)\zeta(\sigma_0+\mathrm{i}u)-1\big|^2\Big] \\ &\quad\ll \int_{-(\log T)^\theta}^{(\log T)^\theta}\mathbb{E}\Big[(|M\zeta-1|\cdot|M'\zeta+M\zeta'|)(\sigma_0+\mathrm{i}\tau+\mathrm{i}h)\Big]\mathrm{d}h \\ &\qquad+\mathcal{O}\Big((\log T)^{\theta-(100+e^\theta)}\Big),\end{aligned} \tag{3.39}$$

where the error term stems from Lemma 3.8 applied to the boundary terms, and integrating over the interval. The expectation in the integral can be bounded by the Cauchy-Schwarz inequality and is smaller than

$$\Big(\mathbb{E}\big[|M\zeta(\sigma_0+\mathrm{i}\tau+\mathrm{i}h)-1|^2\big]\Big)^{1/2}\cdot\Big(\mathbb{E}\big[|M'\zeta+M\zeta'|(\sigma_0+\mathrm{i}\tau+\mathrm{i}h)|^2\big]\Big)^{1/2}. \tag{3.40}$$

The first term is $\ll (\log T)^{-(100+e^{\theta})}$ by Lemma 3.8. The second term can be estimated by another Cauchy-Schwarz inequality yielding that it is

$$\ll \Big(\mathbb{E}\big[(|\zeta|^4 + |\zeta'|^4)(\sigma_0 + \mathrm{i}\tau + \mathrm{i}h)\big]\Big)^{1/4} \cdot \Big(\mathbb{E}\big[(|M|^4 + |M'|^4)(\sigma_0 + \mathrm{i}\tau + \mathrm{i}h)\big]\Big)^{1/4}. \tag{3.41}$$

The first term can be bounded unconditionally by $\ll (\log T)^2$ using the work of Conrey (1988). The second factor is also bounded by $\ll (\log T)^2$ using elementary estimates on Dirichlet polynomials, see Lemma A.3. Putting this back in (3.39), this gives after integrating over the interval

$$\mathbb{E}\left[\max_{|\tau-u|\leq(\log T)^{\theta}} \big|M(\sigma_0+\mathrm{i}u)\zeta(\sigma_0+\mathrm{i}u)-1\big|^2\right] = \mathrm{o}(1). \tag{3.42}$$

The statement then follows by Chebyshev's inequality. □

3.3. Bounds for Dirichlet polynomials

The aim of this section is to approximate the mollifier M by the exponential of a Dirichlet polynomial. This again relies on Section 4.3 in Arguin et al. (2018) with slight modification to account for the length of the interval. To do so, we shall need some intermediary polynomials. Consider, for $0 \leq j \leq K-2$, the Dirichlet polynomials

$$\mathcal{P}_j(h) = \sum_{n\in J_j} \frac{\Lambda(n)}{n^{\sigma_0+\mathrm{i}\tau+\mathrm{i}h}\log n}, \quad \widetilde{P}_j(h) = \sum_{p\in J_j} \frac{1}{p^{\sigma_0+\mathrm{i}\tau+\mathrm{i}h}}, \quad P_j(h) = \mathrm{Re}\, \widetilde{P}_j(h), \tag{3.43}$$

where

$$J_j = (\exp((\log T)^{\frac{j}{K}}, \exp((\log T)^{\frac{j+1}{K}})]. \tag{3.44}$$

The function $\Lambda(n)$ is the von Mangoldt's function, which equals $\log p$ if p is a prime power, and is 0 otherwise. The difference between $\mathcal{P}_j$ and $\widetilde{P}_j$ is only in the prime powers. We can estimate the contribution of powers larger than 2 trivially. We get

$$Q(h) := \sum_{j=0}^{K-2} (\mathcal{P}_j(h) - \widetilde{P}_j(h)) = \frac{1}{2}\sum_{p\leq\sqrt{X}} \frac{1}{p^{2\sigma_0+2\mathrm{i}\tau+\mathrm{i}h}} + \mathcal{O}(1). \tag{3.45}$$

In Lemma 3.10, we will show that $\max_{|h|\leq(\log T)^{\theta}} |M(\sigma_0+\mathrm{i}\tau+\mathrm{i}h) - \exp(-\sum_{j=0}^{K-2}\mathcal{P}_j(h))|$ is small for most τ's. We need the following lemma, which is a generalization of Lemma 4.4 in Arguin et al. (2018).

Lemma 3.9. *Let $\theta > -1$ be given. With the notation as above,*

$$\mathbb{P}\Big(\max_{|h|\leq(\log T)^{\theta}} |Q(h)| \geq Ce^{\theta\vee 0}\sqrt{\log\log T}\Big) = \mathrm{o}(1), \tag{3.46}$$

for some large enough universal constant $C>0$, and

$$\mathbb{P}\Big(\max_{|h|\leq(\log T)^{\theta}} \max_{0\leq j\leq K-2} |\widetilde{P}_j(h)| \geq 10\, e^{\theta\vee 0}K^{-1/2}\log\log T\Big) = \mathrm{o}(1). \tag{3.47}$$

Proof. The case $\theta \leq 0$ was proved in Arguin et al. (2018), so we will assume that $\theta > 0$. We start with (3.46). Let ℓ denote a natural number to be chosen later. Applying the Sobolev inequality from Lemma A.2 to the function $h \mapsto Q(h)^\ell$, we obtain

$$\max_{|h|\leq(\log T)^\theta} |Q(h)|^{2\ell} \ll |Q((\log T)^\theta)|^{2\ell} + |Q(-(\log T)^\theta)|^{2\ell} + \ell \int_{-(\log T)^\theta}^{(\log T)^\theta} |Q(v)|^{2\ell-1}|Q'(v)|\mathrm{d}v. \tag{3.48}$$

Combining this with Chebyshev's inequality and the Cauchy-Schwarz inequality, we may bound the probability in (3.46) by

$$\ll \frac{\max_{|h|\leq(\log T)^\theta} \mathbb{E}\big[|Q(h)|^{2\ell}\big] + \ell\,(\log T)^\theta \max_{|h|\leq(\log T)^\theta} \big(\mathbb{E}\big[|Q(h)|^{4\ell-2}\big]\mathbb{E}\big[|Q'(h)|^2\big]\big)^{\frac{1}{2}}}{(C^2\, e^{2\theta} \log\log T)^\ell}. \tag{3.49}$$

Now, an application of Lemma A.5 (assuming $X^{(2\ell-1)/2} \ll T/\log T$) shows that, for all $|h| \leq (\log T)^\theta$,

$$\begin{aligned}
&\mathbb{E}\big[|Q(h)|^{2\ell}\big] \ll \ell! \left(\frac{1}{4}\sum_{p\leq\sqrt{X}} \frac{1}{p^{4\sigma_0}}\right)^\ell \ll (\ell\, c)^\ell, \\
&\mathbb{E}\big[|Q(h)|^{4\ell-2}\big] \ll (2\ell-1)! \left(\frac{1}{4}\sum_{p\leq\sqrt{X}} \frac{1}{p^{4\sigma_0}}\right)^{2\ell-1} \ll (\ell\, c)^{2\ell-1}, \\
&\mathbb{E}\big[|Q'(h)|^2\big] \ll \sum_{p\leq\sqrt{X}} \frac{(\log p)^2}{p^{4\sigma_0}} \ll c,
\end{aligned} \tag{3.50}$$

for some universal constant $c > 0$. If we take $\ell = \lfloor \log\log T \rfloor$ in (3.49), then $X^{(2\ell-1)/2} \ll T/\log T$ and the probability in (3.46) is

$$\ll \left(\frac{c}{C^2\, e^{2\theta}}\right)^\ell + \ell^{1/2}\left(\frac{c}{C^2}\right)^\ell. \tag{3.51}$$

Taking C large enough with respect to c proves (3.46).

For the second claim of the lemma, let ℓ (again) denote a natural number to be chosen later. By applying the same argument as above to $h \mapsto \widetilde{P}_j(h)^\ell$, we have, for all $0 \leq j \leq K-2$,

$$\begin{aligned}
&\mathbb{P}\Big(\max_{|h|\leq(\log T)^\theta} |\widetilde{P}_j(h)| \geq 10\, e^\theta K^{-1/2} \log\log T\Big) \\
&\ll \frac{\max_{|h|\leq(\log T)^\theta} \mathbb{E}\big[|\widetilde{P}_j(h)|^{2\ell}\big] + \ell\,(\log T)^\theta \max_{|h|\leq(\log T)^\theta} \big(\mathbb{E}\big[|\widetilde{P}_j(h)|^{4\ell-2}\big]\mathbb{E}\big[|\widetilde{P}_j'(h)|^2\big]\big)^{\frac{1}{2}}}{(100\, e^{2\theta} K^{-1} (\log\log T)^2)^\ell}.
\end{aligned} \tag{3.52}$$

An application of Lemma A.5 (assuming $X^{2\ell-1} \ll T/\log T$) and Lemma A.1 shows that, for all $|h| \leq (\log T)^\theta$,

$$\begin{aligned}
&\mathbb{E}\big[|\widetilde{P}_j(h)|^{2\ell}\big] \ll \ell! \left(\sum_{p\in J_j} \frac{1}{p^{2\sigma_0}}\right)^\ell \ll (\ell K^{-1} \log\log T)^\ell, \\
&\mathbb{E}\big[|\widetilde{P}_j(h)|^{4\ell-2}\big] \ll (2\ell-1)! \left(\sum_{p\in J_j} \frac{1}{p^{2\sigma_0}}\right)^{2\ell-1} \ll (\ell K^{-1} \log\log T)^{2\ell-1}, \\
&\mathbb{E}\big[|\widetilde{P}_j'(h)|^2\big] \ll \sum_{p\in J_j} \frac{(\log p)^2}{p^{2\sigma_0}} \ll (\log X)^2.
\end{aligned} \tag{3.53}$$

If we take $\ell = \lfloor \log\log T \rfloor$, then $X^{2\ell-1} \ll T/\log T$ and the probability in (3.52) is

$$\ll \left(\frac{1}{100\, e^{2\theta}}\right)^{\ell} + K^{1/2}\left(\frac{1}{100}\right)^{\ell} \log X = \mathrm{o}(1). \tag{3.54}$$

The claim (3.47) follows from a union bound. □

The next lemma is the extension of Lemma 4.5 in Arguin et al. (2018) to $\theta > 0$.

Lemma 3.10. *Let $\theta > -1$ be given. Then,*

$$\mathbb{P}\bigg(\max_{|h|\leq(\log T)^{\theta}} \Big|M(\sigma_0 + \mathrm{i}\tau + \mathrm{i}h) - \exp\Big(-\sum_{j=0}^{K-2} \mathcal{P}_j(h)\Big)\Big| > (\log T)^{-2}\bigg) = \mathrm{o}(1). \tag{3.55}$$

Proof. Assume that $\theta > 0$. We have the following fact: if z is a complex number that satisfies $1 \leq |z| \leq \frac{n}{10}$ for some large natural integer n, then

$$\Big|e^{z} - \sum_{j=0}^{n} \frac{z^j}{j!}\Big| \leq \sum_{j>n} \frac{|z|^j}{j!} \leq \frac{|z|^n}{n!} \leq e^{-n}. \tag{3.56}$$

Recall from (3.33) that $\nu = 100\, e^{\theta} K \log\log T$, and define the truncated exponential

$$\mathcal{M}(h) := \sum_{k\leq\nu} \frac{(-1)^k}{k!}\Big(\sum_{j=0}^{K-2} \mathcal{P}_j(h)\Big)^k. \tag{3.57}$$

Lemma 3.9 ensures that, on an event of probability $1 - \mathrm{o}(1)$, we have

$$\max_{|h|\leq(\log T)^{\theta}} \Big|\sum_{j=0}^{K-2} \mathcal{P}_j(h)\Big| \leq \max_{|h|\leq(\log T)^{\theta}} \Big(|Q(h)| + \sum_{j=0}^{K-2} |\widetilde{P}_j(h)|\Big) \leq \frac{\nu}{10}. \tag{3.58}$$

On this event, the truncated exponential is a good approximation by (3.56) :

$$\max_{|h|\leq(\log T)^{\theta}} \Big|\mathcal{M}(h) - \exp\Big(-\sum_{j=0}^{K-2} \mathcal{P}_j(h)\Big)\Big| \leq e^{-\nu} \ll (\log T)^{-100}. \tag{3.59}$$

Hence, the proof of (3.55) is reduced to showing

$$\mathbb{P}\Big(\max_{|h|\leq(\log T)^{\theta}} |M(\sigma_0 + \mathrm{i}\tau + \mathrm{i}h) - \mathcal{M}(h)| > (\log T)^{-3}\Big) = \mathrm{o}(1). \tag{3.60}$$

The rest of the proof is virtually identical to the one of Lemma 4.5 in Arguin et al. (2018). The only difference is that the choice $\nu = 100\, e^{\theta} K \log\log T$ changes the bounds $\ll (\log T)^{-50}$ to $\ll (\log T)^{-50e^{\theta}}$. Now, since $50e^{\theta} \geq 50 + 50\theta$, the factor $(\log T)^{-50\theta}$ that we gain easily cancels out the additional factor $(\log T)^{\theta}$ from the Sobolev inequality. □

The following upper bound for the maximum of a short Dirichlet polynomial is needed to bound the moments in Proposition 3.12. It can be proved by discretizing as we did in the upper bound proof of Theorem 1.2, using Lemma 2.8 and applying Chernoff's bound with a suitable exponent.

Proposition 3.11. *Let $\theta > -1$ and $\varepsilon > 0$ be given. Then,*

$$\mathbb{P}\Big(\max_{|h|\leq(\log T)^{\theta}} \sum_{j=1}^{K-3} P_j(h) > (m(\theta) + \varepsilon)\log\log T\Big) = \mathrm{o}(1). \tag{3.61}$$

3.4. Proofs of the lower bounds

We first prove a lower bound for the moments of Dirichlet polynomials.

Proposition 3.12. *Let $\theta > -1$ and $\varepsilon > 0$ be given. Then,*

$$\mathbb{P}\Big(\int_{-(\log T)^\theta}^{(\log T)^\theta} \exp\big(\beta \sum_{j=1}^{K-3} P_j(h)\big) \mathrm{d}h > (\log T)^{f_\theta(\beta)-\varepsilon}\Big) = 1 - \mathrm{o}(1). \tag{3.62}$$

The polynomial P_{K-2} is not included in the sum to ensure that the variances of the P_j's are almost equal. Indeed, for all $|h| \leq (\log T)^\theta$ and $j \leq K-3$, a simple application of the prime number theorem (see e.g. Theorem 6.9 in Montgomery and Vaughan (2007)) yields

$$\begin{aligned} s_j^2 := \mathbb{E}[P_j(h)^2] &= \frac{1}{2} \sum_{p \in J_j} p^{-2\sigma_0} = \frac{1}{2} \int_{\inf J_j}^{\sup J_j} \frac{1}{u^{2\sigma_0} \log u} \mathrm{d}u + \mathcal{O}\Big(e^{-c\sqrt{\log(\inf J_j)}}\Big) \\ &= \frac{1}{2} \log \frac{\log(\sup J_j)}{\log(\inf J_j)} + \mathcal{O}\Big((\sigma_0 - \tfrac{1}{2}) \log(\sup J_j) + e^{-c\sqrt{\log(\inf J_j)}}\Big) \\ &= \frac{1}{2K} \log\log T + \mathcal{O}((\log T)^{-\frac{1}{2K}}), \end{aligned} \tag{3.63}$$

where the last line holds because $\sigma_0 - \frac{1}{2} = (\log T)^{-1+3/(2K)}$ from (3.3) and $\log \sup J_{K-3} = (\log T)^{1-2/K}$ from (3.44). The polynomial P_0 is ignored to ensure that the polynomials $\sum_{j=1}^{K-3} P_j(h)$ are almost independent for h's that are far apart, which will be crucial for the second-moment method to go through; see below (3.80) in the proof of Proposition 3.12.

Proof of Proposition 3.12. This is similar to the upper bound proof of Theorem 1.1. We first relate the moments to the measure of high points. Let $\varepsilon > 0$ and $M \in \mathbb{N}$, and set

$$\mathcal{E}_\theta(\gamma) := \begin{cases} \theta - \frac{\gamma^2}{1+\theta}, & \text{if } \theta \leq 0, \\ \theta - \gamma^2, & \text{if } \theta > 0. \end{cases} \tag{3.64}$$

Consider $\gamma_j = \frac{j}{M} m(\theta) + \varepsilon$ for $1 \leq j \leq M$, and the good event

$$\begin{aligned} E = \bigcap_{j=1}^{M} &\Big\{ \mathrm{Leb}\big\{|h| \leq (\log T)^\theta : \exp\big(\sum_{j=1}^{K-3} P_j(h)\big) > (\log T)^{\gamma_{j-1}}\big\} \geq (\log T)^{\mathcal{E}_\theta(\gamma_{j-1}) - \varepsilon/2} \Big\} \\ &\bigcap \Big\{ \max_{|h| \leq (\log T)^\theta} \exp\big(\sum_{j=1}^{K-3} P_j(h)\big) \leq (\log T)^{m(\theta)+\varepsilon} \Big\}. \end{aligned} \tag{3.65}$$

We will show below that $\mathbb{P}(E)$ is $1 - \mathrm{o}(1)$. First, we prove the lower bound on the moments on the event E. We have

$$\frac{\log \int_{-(\log T)^\theta}^{(\log T)^\theta} \exp\big(\beta \sum_{j=1}^{K-3} P_j(h)\big) \mathrm{d}h}{\log\log T} \geq \max_{1 \leq j \leq M} \{\beta \gamma_{j-1} + \mathcal{E}_\theta(\gamma_{j-1})\} - \varepsilon/2. \tag{3.66}$$

By the continuity of the function $\gamma \mapsto \beta\gamma + \mathcal{E}_\theta(\gamma)$, Equation (3.66) implies that, on the event E and for M large enough with respect to ε and β,

$$\frac{\log \int_{-(\log T)^\theta}^{(\log T)^\theta} \exp\big(\beta \sum_{j=1}^{K-3} P_j(h)\big) \mathrm{d}h}{\log\log T} > \max_{\gamma \in [\varepsilon, m(\theta)]} \big\{\beta\gamma + \mathcal{E}_\theta(\gamma)\big\} - \varepsilon. \tag{3.67}$$

When $0 < \beta \leq 2m(\theta)/(1 + (\theta \wedge 0))$, take $\varepsilon > 0$ small enough so that $\beta > 2\varepsilon/(1 + (\theta \wedge 0))$. The maximum is attained at $\gamma = \frac{\beta}{2}(1 + (\theta \wedge 0))$, in which case the right-hand side of (3.67) is equal to $\frac{\beta^2}{4}(1 + (\theta \wedge 0)) + \theta - \varepsilon$. When $\beta > 2m(\theta)/(1 + (\theta \wedge 0))$, the maximum is attained at $\gamma = m(\theta)$, in which case the right-hand side of (3.67) is equal to $(\beta m(\theta) - 1) - \varepsilon$. Thus, on the event E and for M large enough, the lower bound in (3.62) is satisfied.

To conclude the proof of the proposition, it remains to show that $\mathbb{P}(E) \to 1$ as $T \to \infty$. By the upper bound on the maximum of $\sum_{j=1}^{K-3} P_j(h)$ from Proposition 3.11, it is sufficient to prove that, for all $\eta > 0$ and all $0 < \gamma < m(\theta)$, the event

$$\left\{ \text{Leb}\left\{ |h| \leq (\log T)^{\theta} : \sum_{j=1}^{K-3} P_j(h) > \gamma \log\log T \right\} \geq (\log T)^{\mathcal{E}_\theta(\gamma) - \eta} \right\} \tag{3.68}$$

has probability $1 - \mathrm{o}(1)$. For $\theta < 0$, Lemma 2.3 ensures that the primes up to $\exp((\log T)^{|\theta|})$ only make a very small contribution, namely the event

$$\left\{ \left| \sum_{j=1}^{\mathcal{J}(\theta)-1} P_j(h) \right| \leq \tfrac{\gamma}{(1+(\theta\wedge 0))\,K} \log\log T \right\} \tag{3.69}$$

has probability $1 - \mathrm{o}(1)$. In view of this, we consider

$$\mathcal{J}(\theta) = \begin{cases} 1, & \text{if } \theta \geq 0, \\ \lfloor K|\theta| \rfloor + 1, & \text{if } \theta < 0, \end{cases} \tag{3.70}$$

and the random variable

$$\mathcal{N} = \text{Leb}\left\{ |h| \leq (\log T)^{\theta} : P_j(h) > x_j, \text{ for } \mathcal{J}(\theta) \leq j \leq K - 3 \right\}, \tag{3.71}$$

where

$$x_j = \left(1 + \frac{100}{(1 + (\theta \wedge 0))\,K} \right) \cdot \frac{\gamma}{(1 + (\theta \wedge 0))\,K} \log\log T. \tag{3.72}$$

By summing the x_j's, it is not hard to check that $\{\mathcal{N} \geq (\log T)^{\mathcal{E}_\theta(\gamma)-\eta}\} \cap (3.69) \subseteq (3.68)$. Therefore, the proof of the proposition is reduced to show

$$\mathbb{P}\big(\mathcal{N} \geq (\log T)^{\mathcal{E}_\theta(\gamma)-\eta}\big) = 1 - \mathrm{o}(1). \tag{3.73}$$

This is established by the Paley-Zygmund inequality.

To this aim, we shall need one-point and two-point large deviation estimates for the event

$$A(h) = \left\{ P_j(h) > x_j, \text{ for } \mathcal{J}(\theta) \leq j \leq K - 3 \right\}. \tag{3.74}$$

The next two propositions are stated as Propositions 5.4 and 5.5 in Arguin et al. (2018). They are consequences of the Gaussian moments in Lemma A.4.

Proposition 3.13 (One-point large deviation estimates)**.** *Let $\theta > -1$ be given. For any choices of $0 < x_j \leq \log\log T$, where $1 \leq j \leq K - 3$, we have*

$$\mathbb{P}(A(h)) = (1 + \mathrm{o}(1)) \prod_{j=\mathcal{J}(\theta)}^{K-3} \int_{x_j/s_j}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}} \mathrm{d}y \asymp \prod_{j=\mathcal{J}(\theta)}^{K-3} \frac{s_j}{x_j} \cdot e^{-x_j^2/(2s_j^2)}. \tag{3.75}$$

In the case of two points h, h', the primes are essentially correlated up to $\exp(|h-h'|^{-1})$ and quickly decorrelate afterwards. For $\theta \geq 0$, this means that the P_j's are essentially independent whenever $|h-h'| > (\log T)^{-\frac{1}{2K}}$, since $j = 0$ is excluded. For $\theta < 0$, we must exclude the j's up to $\mathcal{J}(\theta)$. Therefore, the P_j's are essentially independent whenever $|h-h'| > (\log T)^{\theta - \frac{1}{2K}}$. We get:

Proposition 3.14 (Two-point large deviation estimates)**.** *Let $\theta > -1$ be given. Let $h, h' \in [-(\log T)^\theta, (\log T)^\theta]$ be such that $|h-h'| > (\log T)^{-\frac{\mathcal{J}(\theta)}{K} + \frac{1}{2K}}$. Then,*

$$\mathbb{P}(A(h) \cap A(h')) = (1 + \mathrm{o}(1))\, \mathbb{P}(A(h))\, \mathbb{P}(A(h')). \tag{3.76}$$

If $|h-h'| \leq 1$, let $0 \leq \ell \leq K-3$ denote the largest integer in this range with $|h-h'| \leq (\log T)^{-\ell/K}$. Then, for any choices of $\sqrt{\log\log T} \ll x_j \leq \log\log T$, we have

$$\mathbb{P}(A(h) \cap A(h')) \ll \exp\Big(- \sum_{j=\mathcal{J}(\theta)}^{\ell} \frac{x_j^2}{2s_j^2} - \sum_{j=(\ell+1)\vee\mathcal{J}(\theta)}^{K-3} \frac{x_j^2}{s_j^2} \Big). \tag{3.77}$$

First, we find a lower bound on $\mathbb{E}[\mathcal{N}]$. By (3.75), the x_j's in (3.72) and the s_j's in (3.63), we have

$$\mathbb{E}[\mathcal{N}] = \int_{-(\log T)^\theta}^{(\log T)^\theta} \mathbb{P}(A(h))\mathrm{d}h \gg (\log T)^\theta \prod_{j=\mathcal{J}(\theta)}^{K-3} \frac{s_j}{x_j} \cdot e^{-x_j^2/(2s_j^2)} \gg (\log T)^{\mathcal{E}_\theta(\gamma) - \eta/3}, \tag{3.78}$$

assuming that K is large enough with respect to θ, γ and η. By the Paley-Zygmund inequality, this implies

$$\begin{aligned} \mathbb{P}\big(\mathcal{N} \geq (\log T)^{\mathcal{E}_\theta(\gamma)-\eta}\big) &\geq \mathbb{P}\big(\mathcal{N} \geq (\log T)^{-\eta/3}\mathbb{E}[\mathcal{N}]\big) \\ &\geq \big(1 - (\log T)^{-\eta/3}\big) \frac{(\mathbb{E}[\mathcal{N}])^2}{\mathbb{E}[\mathcal{N}^2]}. \end{aligned} \tag{3.79}$$

It remains to show $\mathbb{E}[\mathcal{N}^2] = (1+\mathrm{o}(1))(\mathbb{E}[\mathcal{N}])^2$. Linearity yields

$$\mathbb{E}[\mathcal{N}^2] = \int_{I \times I} \mathbb{P}(A(h) \cap A(h'))\ \mathrm{d}h\mathrm{d}h'. \tag{3.80}$$

The integral can be divided into $(K - \mathcal{J}(\theta) + 1)$ parts:

$$\begin{aligned} &B = \{(h,h') : |h-h'| > (\log T)^{-\frac{\mathcal{J}(\theta)}{K} + \frac{1}{2K}}\}; \\ &B_0 = \{(h,h') : (\log T)^{-\frac{\mathcal{J}(\theta)}{K}} < |h-h'| \leq (\log T)^{-\frac{\mathcal{J}(\theta)}{K} + \frac{1}{2K}}\}; \\ &B_\ell = \{(h,h') : (\log T)^{-(\ell+1)/K} < |h-h'| \leq (\log T)^{-\ell/K}\}, \quad \text{for } \ell = \mathcal{J}(\theta), \dots, K-3; \\ &B_{K-2} = \{(h,h') : |h-h'| \leq (\log T)^{-(K-2)/K}\}. \end{aligned} \tag{3.81}$$

The dominant term will be the one on B. Note that $\mathrm{Leb}(B) = \mathrm{Leb}(I)^2(1+\mathrm{o}(1))$. Hence, by (3.76), we have

$$\int_B \mathbb{P}(A(h) \cap A(h'))\mathrm{d}h\mathrm{d}h' = (1+\mathrm{o}(1))(\mathbb{E}[\mathcal{N}])^2. \tag{3.82}$$

By (3.77) and the estimate (3.78), the integral on B_0 is

$$\begin{aligned} &\ll (\log T)^{\theta} (\log T)^{-\frac{\mathcal{J}(\theta)}{K} + \frac{1}{2K}} \exp\Big(\sum_{j=\mathcal{J}(\theta)}^{K-3} -\frac{x_j^2}{s_j^2} \Big) \\ &\ll (\log T)^{-(\theta \vee 0) - \frac{1}{3K}} (\mathbb{E}[\mathcal{N}])^2, \end{aligned} \tag{3.83}$$

assuming that K is large enough with respect to θ and γ. For $\ell = \mathcal{J}(\theta), \dots, K-3$, the integral on B_ℓ yields, by (3.77) and the estimate (3.78),

$$\begin{aligned} &\ll (\log T)^{\theta - \ell/K} \exp\Big(- \sum_{j=\mathcal{J}(\theta)}^{\ell} \frac{x_j^2}{2 s_j^2} - \sum_{j=\ell+1}^{K-3} \frac{x_j^2}{s_j^2} \Big) \\ &= (\log T)^{-\theta - \ell/K} \exp\Big(\sum_{j=\mathcal{J}(\theta)}^{\ell} \frac{x_j^2}{2 s_j^2} \Big) \cdot (\log T)^{2\theta} \exp\Big(- \sum_{j=\mathcal{J}(\theta)}^{K-3} \frac{x_j^2}{s_j^2} \Big) \\ &\ll (\log T)^{-\theta - \ell/K + (\ell/K + (\theta \wedge 0)) \frac{\gamma^2}{(1 + (\theta \wedge 0))^2} + \eta} (\mathbb{E}[\mathcal{N}])^2, \end{aligned} \tag{3.84}$$

assuming again that K is large enough with respect to θ, γ and η. Since $\gamma^2 < m(\theta)^2 = (1+\theta)(1 + (\theta \wedge 0))$, the right-hand side of (3.84) is $\mathrm{o}\big((\mathbb{E}[\mathcal{N}])^2\big)$ if we fix $\eta > 0$ small enough with respect to θ. Similarly, by (3.75) and the estimate (3.78), the integral on B_{K-2} is

$$\leq \int_{B_{K-2}} \mathbb{P}(A(h)) \mathrm{d}h \mathrm{d}h' \ll (\log T)^{-\theta - 1 + 2/K + \eta/3} \cdot \mathbb{E}[\mathcal{N}] = \mathrm{o}\big((\mathbb{E}[\mathcal{N}])^2\big). \tag{3.85}$$

This concludes the proof of Proposition 3.12. □

Putting all the work of Section 3 together, we can prove the lower bound in Theorem 1.1.

Proof of Proposition 3.2. By Proposition 3.6 (reduction off-axis), the probability in (3.2) is

$$\geq \mathbb{P}\Big(\int_{-\frac{1}{2}(\log T)^{\theta}}^{\frac{1}{2}(\log T)^{\theta}} |\zeta(\sigma_0 + \mathrm{i}\tau + \mathrm{i}h)|^{\beta} \mathrm{d}h > (\log T)^{f_{\theta}(\beta) - \varepsilon/2} \Big) - \mathrm{o}(1). \tag{3.86}$$

By Lemma 3.7 (mollification), the above is

$$\geq \mathbb{P}\Big(\int_{-\frac{1}{2}(\log T)^{\theta}}^{\frac{1}{2}(\log T)^{\theta}} |M(\sigma_0 + \mathrm{i}\tau + \mathrm{i}h)|^{-\beta} \mathrm{d}h > (\log T)^{f_{\theta}(\beta) - \varepsilon/3} \Big) - \mathrm{o}(1), \tag{3.87}$$

and, by Lemma 3.10, the above is

$$\geq \mathbb{P}\Big(\int_{-\frac{1}{2}(\log T)^{\theta}}^{\frac{1}{2}(\log T)^{\theta}} \exp\big(\beta \sum_{j=0}^{K-2} \mathrm{Re}\, \mathcal{P}_j(h)\big) \mathrm{d}h > (\log T)^{f_{\theta}(\beta) - \varepsilon/4} \Big) - \mathrm{o}(1). \tag{3.88}$$

By Lemma 3.9, we may replace $\mathrm{Re}\, \mathcal{P}_j(h)$ by $P_j(h)$ with a negligible error, and also discard the terms with $j = 0$ and $j = K-2$. For K large enough with respect to ε, β and θ, the probability in (3.88) is

$$\geq \mathbb{P}\Big(\int_{-\frac{1}{2}(\log T)^{\theta}}^{\frac{1}{2}(\log T)^{\theta}} \exp\big(\beta \sum_{j=1}^{K-3} P_j(h)\big) \mathrm{d}h > (\log T)^{f_{\theta}(\beta) - \varepsilon/5} \Big) - \mathrm{o}(1). \tag{3.89}$$

Finally, the probability in (3.89) tends to 1 as $T \to \infty$ by Proposition 3.12. □

We now prove the lower bound in Theorem 1.2.

Proof of Proposition 3.1. From (1.8), $f_\theta(\beta) = \beta m(\theta) - 1$ when $\beta > \beta_c(\theta) = 2\sqrt{1 + (\theta \wedge 0)}$. Thus, on the event in the statement of Proposition 3.2 (which has probability $1 - o(1)$), and for β large enough with respect to ε and θ, we have

$$\begin{aligned}
\max_{|h| \leq (\log T)^\theta} |\zeta(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)| &\geq \left(\frac{1}{2(\log T)^\theta} \int_{-(\log T)^\theta}^{(\log T)^\theta} |\zeta(\tfrac{1}{2} + \mathrm{i}\tau + \mathrm{i}h)|^\beta \mathrm{d}h \right)^{1/\beta} \\
&\gg (\log T)^{m(\theta) - \frac{(1+\varepsilon+\theta)}{\beta}} \geq (\log T)^{m(\theta) - \varepsilon}.
\end{aligned} \tag{3.90}$$

This ends the proof. □

Appendix A: Useful estimates

The prime number theorem yields estimates on the sum of primes with a good error.

Lemma A.1. *Let $1 \leq P \leq Q$, then*

$$\sum_{P < p \leq Q} \frac{(\log p)^m}{p} = \begin{cases} \frac{(\log Q)^m}{m} - \frac{(\log P)^m}{m} + \mathcal{O}_m(1), & \text{if } m \geq 1, \\ \log\log Q - \log\log P + \mathcal{O}(e^{-c\sqrt{\log P}}), & \text{if } m = 0. \end{cases} \tag{A.1}$$

Also, for $|\alpha \log Q| \leq 1$,

$$\sum_{P < p \leq Q} \frac{\cos(\alpha \log p)}{p} = \log\log Q - \log\log P + \mathcal{O}(1). \tag{A.2}$$

Proof. For (A.1), see Lemma A.1 of Arguin and Ouimet (2019) and Lemma 2.1 of Arguin, Belius and Harper (2017). For (A.2), see p.20 in Harper (2013b). □

In the remainder of this section, we gather standard estimates on Dirichlet polynomials. The following is useful to get a uniform control on an interval.

Lemma A.2 (Sobolev's inequality, Equation (6) in Arguin et al. (2018))**.** *Let $f : [a, b] \to \mathbb{R}$ be a continuously differentiable function on a finite interval $[a, b] \subseteq \mathbb{R}$, then*

$$\max_{h \in [a,b]} |f(h)|^2 \leq \frac{|f(b)|^2 + |f(a)|^2}{2} + \int_a^b |f'(v) f(v)| \mathrm{d}v. \tag{A.3}$$

The next three results yields moment estimates of Dirichlet polynomials. The first one is an elementary bound. The second ensures that moments of Dirichlet polynomials that are not too high are Gaussian.

Lemma A.3 (Lemma 3.3 in Arguin et al. (2018))**.** *For any complex numbers $a(n)$ and $b(n)$, and for $N \leq T$, we have*

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\sum_{m \leq N} a(m) m^{-\mathrm{i}\tau} \Big) \Big(\sum_{n \leq N} b(n) n^{\mathrm{i}\tau} \Big) \Big] \\
&\quad = \sum_{n \leq N} a(n) b(n) + \mathcal{O}\Big(\frac{N \log N}{T} \sum_{n \leq N} (|a(n)|^2 + |b(n)|^2) \Big).
\end{aligned} \tag{A.4}$$

Lemma A.4 (Lemma 3.4 in Arguin et al. (2018)). *Let $x \geq 2$ be a real number, and suppose that for primes $p \leq x$, $a(p)$ are complex numbers with $|a(p)| \leq 1$. Then, for any $k \in \mathbb{N}$,*

$$\begin{aligned}\mathbb{E}\Big[\Big(\frac{1}{2}\sum_{p\leq x}(a(p)p^{-\mathrm{i}\tau} + a^\star(p)p^{\mathrm{i}\tau})\Big)^k\Big] \\ = \frac{\partial^k}{\partial z^k}\Big(\prod_{p\leq x} I_0(\sqrt{a(p)a^\star(p)}z)\Big)\Big|_{z=0} + \mathcal{O}\Big(\frac{x^{2k}}{T}\Big),\end{aligned} \tag{A.5}$$

where $I_0(z) = \sum_{n\geq 0} z^{2n}/(2^{2n}(n!)^2)$ denotes the modified Bessel function of the first kind. In particular, the expression is $\mathcal{O}\left(x^{2k}/T\right)$ for odd k.

The relations with Gaussian moments in the case where $a(p) = p^{-\sigma-\mathrm{i}h}$ is obtained by expanding the product to get

$$\prod_{p\leq x} I_0(|a(p)|z) = F(z)\cdot \exp\left(\frac{z^2}{2}\cdot\frac{1}{2}\sum_{p\leq x}p^{-2\sigma}\right) \tag{A.6}$$

where $F(z)$ is analytic in a neighborhood of 0 with $F(0) = 1$ and derivatives uniformly bounded by $\sum_{p\leq x} p^{-4\sigma}$. In particular, this implies that, for $\sigma \geq 1/2$ and k small enough so that $x^{2k}/T = \mathrm{o}(1)$,

$$\mathbb{E}\Big[\Big(\sum_{p\leq x}\mathrm{Re}\, p^{-\sigma-\mathrm{i}\tau-\mathrm{i}h)}\Big)^{2k}\Big] = (1+\mathrm{o}(1))\,\frac{(2k)!}{2^k\cdot k!}\left(\frac{1}{2}\sum_{p\leq x}p^{-2\sigma}\right)^k. \tag{A.7}$$

The above also holds if $a(p) = 0$ for $p \leq y$ (say) with the sum over primes restricted to $y < p \leq x$. In particular, the error $\sum_{y<p\leq x} p^{-4\sigma}$ can be made $\mathrm{o}(1)$ by taking y large.

Finally the third estimate is a cruder version of the Gaussian moment estimates that yields quick upper bounds on moments.

Lemma A.5 (Lemma 3 in Soundararajan (2009)). *Let T be large, and let $2 \leq x \leq T$. Let k be a natural number such that $x^k \ll T/\log T$. For any complex numbers $a(p)$, we have*

$$\mathbb{E}\Big[\Big|\sum_{p\leq x}\frac{a(p)}{p^{1/2+i\tau}}\Big|^{2k}\Big] \ll k!\left(\sum_{p\leq x}\frac{|a(p)|^2}{p}\right)^k. \tag{A.8}$$

References

Arguin, L. P., Belius, D. and Bourgade, P. (2017). Maximum of the characteristic polynomial of random unitary matrices. *Comm. Math. Phys.* **349** 703–751. MR3594368

Arguin, L. P., Belius, D. and Harper, A. J. (2017). Maxima of a randomized Riemann zeta function, and branching random walks. *Ann. Appl. Probab.* **27** 178–215. MR3619786

Arguin, L. P. and Ouimet, F. (2019). Large deviations and continuity estimates for the derivative of a random model of $\log|\zeta|$ on the critical line. *J. Math. Anal. Appl.* **472** 687–695. doi:10.1016/j.jmaa.2018.11.044.

Arguin, L. P. and Tai, W. (2018). Is the Riemann zeta function in a short interval a 1-RSB spin glass ? *To appear in Sojourns in Probability and Statistical Physics (Springer), in honor of the 70th birthday of C. Newman.* arXiv:1706.08462.

Arguin, L. P., Belius, D., Bourgade, P., Radziwiłł, M. and Soundararajan, K. (2018). Maximum of the Riemann zeta function on a short interval of the critical line. *To appear in Comm. Pure Appl. Math.* doi:10.1002/cpa.21791.

Bailey, E. C. and Keating, J. P. (2018). On the moments of the moments of the characteristic polynomials of random unitary matrices. *Preprint* 1–26. arXiv:1807.06605.

Bettin, S., Chandee, V. and Radziwiłł, M. (2017). The mean square of the product of the Riemann zeta-function with Dirichlet polynomials. *J. Reine Angew. Math.* **729** 51–79. MR3680370

Bondarenko, A. and Seip, K. (2017). Large greatest common divisor sums and extreme values of the Riemann zeta function. *Duke Math. J.* **166** 1685–1701. MR3662441

Bourgain, J. (2017). Decoupling, exponential sums and the Riemann zeta function. *J. Amer. Math. Soc.* **30** 205–224. MR3556291

Bovier, A. (2006). *Statistical mechanics of disordered systems. Cambridge Series in Statistical and Probabilistic Mathematics* **18**. Cambridge University Press, Cambridge. MR2252929

Bovier, A. (2017). *Gaussian processes on trees. Cambridge Studies in Advanced Mathematics* **163**. Cambridge University Press, Cambridge. MR3618123

Bovier, A. and Kurkova, I. (2004). Derrida's generalized random energy models. II. Models with continuous hierarchies. *Ann. Inst. H. Poincaré Probab. Statist.* **40** 481–495. MR2070335

Chandee, V. and Soundararajan, K. (2011). Bounding $|\zeta(\frac{1}{2}+it)|$ on the Riemann hypothesis. *Bull. Lond. Math. Soc.* **43** 243–250. MR2781205

Chhaibi, R., Madaule, T. and Najnudel, J. (2018). On the maximum of the CβE field. *Duke Math. J.* **167** 2243–2345. MR3848391

Conrey, J. B. (1988). The fourth moment of derivatives of the Riemann zeta-function. *Quart. J. Math. Oxford Ser. (2)* **39** 21–36. MR929792

de la Bretèche, R. and Tenenbaum, G. (2018). Sommes de Gál et applications. *Preprint.* arXiv:1804.01629.

Derrida, B. (1981). Random-energy model: an exactly solvable model of disordered systems. *Phys. Rev. B (3)* **24** 2613–2626. MR627810

Farmer, D. W., Gonek, S. M. and Hughes, C. P. (2007). The maximum size of L-functions. *J. Reine Angew. Math.* **609** 215–236. MR2350784

Fyodorov, Y. V. and Bouchaud, J. P. (2008). Freezing and extreme-value statistics in a random energy model with logarithmically correlated potential. *Journal of Physics A: Mathematical and Theoretical* **41** 372001. MR2430565

Fyodorov, Y. V., Gnutzmann, S. and Keating, J. P. (2018). Extreme values of CUE characteristic polynomials: a numerical study. *Journal of Physics A: Mathematical and Theoretical* **51** 464001. doi:10.1088/1751-8121/aae65a.

Fyodorov, Y. V., Hiary, G. A. and Keating, J. P. (2012). Freezing Transition, Characteristic Polynomials of Random Matrices, and the Riemann Zeta-Function. *Phys. Rev. Lett.* **108** 1–4. doi:10.1103/PhysRevLett.108.170601.

Fyodorov, Y. V. and Keating, J. P. (2014). Freezing transitions and extreme values: random matrix theory, $\zeta(\frac{1}{2}+it)$ and disordered landscapes. *Philos. Trans. R. Soc. A* **372** 1–32. MR3151088

Gabriel, R. M. (1927). Some Results Concerning the Integrals of Moduli of Regular Functions Along Certain Curves. *J. London Math. Soc.* **2** 112–117. MR1574583

Harper, A. J. (2013a). Sharp conditional bounds for moments of the Riemann zeta function. *Preprint* 1–20. arXiv:1305.4618.

Harper, A. J. (2013b). A note on the maximum of the Riemann zeta function, and log-correlated random variables. *Preprint* 1–26. arXiv:1304.0677.

Heap, W., Radziwiłł, M. and Soundararajan, K. (2019). Upper bounds for fractional moments of the Riemann zeta-function. *Preprint*.

Heath-Brown, D. R. (1981). Fractional moments of the Riemann zeta function. *J. London Math. Soc. (2)* **24** 65–78. MR623671

Keating, J. P. and Snaith, N. C. (2000). Random matrix theory and $\zeta(1/2+it)$. *Comm. Math. Phys.* **214** 57–89. MR1794265

Kistler, N. (2015). Derrida's random energy models. From spin glasses to the extremes of correlated random fields. In *Correlated random systems: five different methods. Lecture Notes in Math.* **2143** 71–120. Springer, Cham. MR3380419

Montgomery, H. L. and Vaughan, R. C. (2007). *Multiplicative number theory. I. Classical theory. Cambridge Studies in Advanced Mathematics* **97**. Cambridge University Press, Cambridge. MR2378655

Najnudel, J. (2018). On the extreme values of the Riemann zeta function on random intervals of the critical line. *Probab. Theory Related Fields* **172** 387–452. MR3851835

Nikula, M., Saksman, E. and Webb, C. (2018). Multiplicative chaos and the characteristic polynomial of the CUE: the L^1-phase. *Preprint* 1–49. arXiv:1806.01831.

Ouimet, F. (2018). Poisson-Dirichlet statistics for the extremes of a randomized Riemann zeta function. *Electron. Commun. Probab.* **23** Paper No. 46, 15. MR3841407

Paquette, E. and Zeitouni, O. (2018). The maximum of the CUE field. *Int. Math. Res. Not. IMRN* **2018** 5028–5119. MR3848227

Radziwiłł, M. and Soundararajan, K. (2013). Continuous lower bounds for moments of zeta and L-functions. *Mathematika* **59** 119–128. MR3028174

Radziwiłł, M. and Soundararajan, K. (2017). Selberg's central limit theorem for $\log|\zeta(1/2+it)|$. *Enseign. Math.* **63** 1–19. MR3832861

Rémy, G. (2018). The Fyodorov-Bouchaud formula and Liouville conformal field theory. *Preprint* 1–24. arXiv:1710.06897.

Rhodes, R. and Vargas, V. (2014). Gaussian multiplicative chaos and applications: a review. *Probab. Surv.* **11** 315–392. MR3274356

Saksman, E. and Webb, C. (2018). The Riemann zeta function and Gaussian multiplicative chaos: statistics on the critical line. *Preprint*. arXiv:1609.00027.

Soundararajan, K. (2009). Moments of the Riemann zeta function. *Ann. of Math. (2)* **170** 981–993. MR2552116

Webb, C. (2015). The characteristic polynomial of a random unitary matrix and Gaussian multiplicative chaos—the L^2-phase. *Electron. J. Probab.* **20** no. 104, 21. MR3407221