The Annals of Probability
2021, Vol. 49, No. 6, 3106–3141
https://doi.org/10.1214/21-AOP1524

MOMENTS OF THE RIEMANN ZETA FUNCTION ON SHORT INTERVALS OF THE CRITICAL LINE

BY LOUIS-PIERRE ARGUIN[1], FRÉDÉRIC OUIMET[2,*] AND MAKSYM RADZIWIŁŁ[2,†]

[1]*Department of Mathematics, Baruch College and Graduate Center (CUNY), louis-pierre.arguin@baruch.cuny.edu*
[2]*The Division of Physics, Mathematics and Astronomy, California Institute of Technology,*
[*]*ouimetfr@caltech.edu;* [†]*maksym@caltech.edu*

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

$$\int_{-\log^\theta T}^{\log^\theta T} \left|\zeta\left(\frac{1}{2} + it + ih\right)\right|^\beta dh = (\log T)^{f_\theta(\beta)+o(1)},$$

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 (*Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **372** (2014) 20120503, 32). In particular, it shows 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 $o(T)$,

$$\max_{|h| \le \log^\theta T} \left|\zeta\left(\frac{1}{2} + it + ih\right)\right| = (\log T)^{m(\theta)+o(1)},$$

for some explicit $m(\theta)$. This generalizes earlier results of Najnudel (*Probab. Theory Related Fields* **172** (2018) 387–452) and Arguin et al. (*Comm. Pure Appl. Math.* **72** (2019) 500–535) for $\theta = 0$. The proofs are unconditional, except for the upper bounds when $\theta > 3$, where the Riemann hypothesis is assumed.

1. Introduction.

1.1. *Maxima and moments over large intervals.* Understanding the growth of the Riemann zeta function $\zeta(s)$ on the critical line $\operatorname{Re} s = \frac{1}{2}$ is a central problem in number theory, due, among other things, to its relationship with the distribution of the zeros of $\zeta(s)$; see, for example, Theorem 9.3 in Titchmarsh (1986); regarding the more general subconvexity problem, see, for example, Michel and Venkatesh (2010), Venkatesh (2010), and see Iwaniec and Sarnak (2000) for a general discussion.

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

$$\left|\zeta\left(\frac{1}{2} + it\right)\right| = \mathcal{O}\left(\exp\left(\left(\frac{\log 2}{2} + o(1)\right)\frac{\log t}{\log\log t}\right)\right), \quad \text{as } t \to \infty; \tag{1.1}$$

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 $|\zeta(\frac{1}{2} + it)| = \mathcal{O}((1 +$

Received January 2021.
MSC2020 subject classifications. Primary 60G70; secondary 11M06, 60F10, 60G60.
Key words and phrases. Extreme value theory, Riemann zeta function, moments.

$|t|)^{13/84+\varepsilon}$) for any given $\varepsilon > 0$ and all $t \in \mathbb{R}$. Currently, the best unconditional lower bound,

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

is established in de la Bretèche and Tenenbaum (2019) and building on a method from Bondarenko and Seip (2017).

The true order of the maximum of $|\zeta(\frac{1}{2}+\mathrm{i}t)|$ remains elusive 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

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

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

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

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 $|\zeta(\frac{1}{2}+\mathrm{i}t)|$, we are in possession of widely believed conjectures regarding the behavior of moments. Following the work Keating and Snaith (2000), it is expected that, for all $\beta > 0$,

$$\frac{1}{T}\int_T^{2T}\left|\zeta\left(\frac{1}{2}+\mathrm{i}t\right)\right|^{\beta}\mathrm{d}t \sim C_\beta(\log T)^{\beta^2/4}, \quad \text{as } T\to\infty, \tag{1.5}$$

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), Soundararajan (2009) 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 et al. (2004), 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 $[T, 2T]$, they conjectured that, for any $0 < \delta < 1$, there exists $C = C(\delta) > 0$ large enough and independent of T such that with probability $1-\delta$,

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

They also conjectured weak convergence with a limiting tail of the form Cye^{-2y}. 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. (2019) unconditionally. The sharp upper bound was recently established in Arguin, Bourgade and Radziwiłł (2020).

It is also conjectured in Fyodorov et al. (2004), Fyodorov and Keating (2014) (see equations (14) and (2.30), respectively) that the moments in a short interval undergo a *freezing phase transition*, that is, the event

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

has probability $1-\mathrm{o}(1)$ as $T\to\infty$. Fyodorov and Keating (2014) also state corresponding conjectures for mesoscopic intervals of length $\log^{\theta} T$, 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 $\log^{\theta} T$ with $\theta>-1$. We also obtain the corresponding results for local maxima over such intervals. The following functions will be crucial to our analysis:

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

THEOREM 1.1 (Moments). *Let $\theta>-1$, $\beta>0$ and $\varepsilon>0$ be given. Let τ be a random variable uniformly distributed on $[T,2T]$. Then, as $T\to\infty$, we have*

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

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

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

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

When $\beta>\beta_c(\theta)$, the moments exhibit *freezing*, that is, they are dominated by a few large values at the level of the local maximum of $|\zeta(\frac{1}{2}+\mathrm{i}\tau+\mathrm{i}h)|$, $|h|\le\log^{\theta} T$. 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 $\theta\to\infty$. We note that recently a sharp upper bound in the case $(\theta=0,\beta=2)$ has been established in Harper (2019), thus refining the $(\log T)^{\varepsilon}$ factor appearing in (1.10), when $\theta=0$ and $\beta=2$.

THEOREM 1.2 (Local maximum). *Let $\theta>-1$ and $\varepsilon>0$ be given. Let τ be a random variable uniformly distributed on $[T,2T]$. Then, as $T\to\infty$, we have*

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

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

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

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 Selberg (1946), Selberg (1992) or the simple proof in Radziwiłł and Soundararajan (2017), states that, for any given $a < b$,

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

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 $\zeta(\frac{1}{2}+\mathrm{i}t)$ varies on the scale of $(\log T)^{-1}$ for $T \le t \le 2T$, the analysis of large values should be reducible to a discrete set of $(\log T)^{1+\theta}$ points. Putting these two facts together, one expects that the statistics of extreme values of $\log|\zeta(\frac{1}{2}+\mathrm{i}\tau+\mathrm{i}h)|$, $|h| \le \log^\theta T$, 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 $\theta \ge 0$, 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 $\log|\zeta(\frac{1}{2}+\mathrm{i}\tau+\mathrm{i}h)|$, $|h| \le \log^\theta T$ are correlated. In fact, they are log-correlated if $|h-h'| \le 1$, as first noticed by Bourgade (2010). A good probabilistic model for the extreme values in the case $\theta = 0$ is, therefore, a branching random walk. This is explained in more details in Section 1.4 and illustrated in Figure 1. 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 $\log^\theta T$ independent branching random walks. One implication is that 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 $\theta \ge 0$ be given and let $m(\theta)$ be as in (1.8). Let τ be a random variable uniformly distributed on $[T, 2T]$. For any $0 < \delta < 1$, there exists $C = C(\delta) > 0$ large enough and independent of T, such that with probability $1-\delta$,*

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

where

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

In particular, we expect a discontinuity of $r(\theta)$ as $\theta \downarrow 0$. An analysis of a model of the Riemann zeta function shows that the discontinuity can be resolved by approaching 0 at a suitable rate. Namely, if $\theta \sim (\log\log T)^{-\alpha}$, it is expected that $r(\theta) = \frac{1+2\alpha}{4}$, interpolating between $1/4$ and $3/4$ for $0 < \alpha < 1$; see Arguin, Dubach and Hartung (2021). Such hybrid statistics have been studied in the context of branching random walks; see Kistler and Schmidt (2015) and Bovier and Hartung (2020).

For $\theta < 0$, our analysis suggests that the correct model consists of a single random walk up to *time* $|\theta| \log\log T$, followed by a branching random walk. The maximum on such intervals would then be consistent with the level proposed in Section 2(c)(ii) of Fyodorov and Keating (2014),

$$\begin{aligned}&\max_{|h|\le \log^\theta T} \log\left|\zeta\left(\frac{1}{2}+\mathrm{i}\tau+\mathrm{i}h\right)\right| \\ &\quad = m(\theta)\log\log T - \frac{3}{4}\log\log\log T + \sqrt{\frac{|\theta|}{2}\log\log T}\cdot \mathcal{Z} + \mathcal{O}_{\mathbb{P}}(1),\end{aligned} \tag{1.15}$$

where $\mathcal{Z}$ is a standard Gaussian random variable. As explained in Section 1.4, the additional fluctuation would represent the contribution of the Dirichlet polynomial $\sum_{\log p \le \log^{|\theta|} T} \operatorname{Re} p^{-1/2-\mathrm{i}(\tau+h)}$, which is essentially the same random variable for all h's in the interval $|h| \le \log^\theta T$.

1.3. *Relations to other models.* When $-1 < \theta \le 0$, 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

$$\mathbb{E}\left[\left(\frac{1}{2\pi}\int_0^{2\pi} |\det(\mathbb{I}-e^{-\mathrm{i}h}M_N)|^{2\beta}\,\mathrm{d}h\right)^k\right], \quad k>0,\ \beta>0. \tag{1.16}$$

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 (2019) 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).

From the analysis of a particular variant of the log-correlated REM model, Fyodorov and Bouchaud (2008) conjectured an exact formula for the density of the total mass of the subcritical Gaussian multiplicative chaos (GMC) measure, associated to the Gaussian free field (GFF) on the unit circle, cf. Rhodes and Vargas (2014). In the critical case they conjectured that the fluctuations of the maximum can be captured by a sum of two Gumbel variables. Both results were proved in Remy (2020). Naturally, these results are expected to hold in the CUE setting, where the GMC measure is the limit of

$$\frac{|\det(\mathbb{I}-e^{-\mathrm{i}h}M_N)|^{2\beta}}{\mathbb{E}[|\det(\mathbb{I}-e^{-\mathrm{i}h}M_N)|^{2\beta}]}\,\frac{\mathrm{d}h}{2\pi}, \tag{1.17}$$

as proved by Webb (2015), when $-1/4 < \beta < 1/\sqrt{2}$, and by Nikula, Saksman and Webb (2020), when $1/\sqrt{2} \le \beta < 1$. Such a random measure can also be considered in the context of the Riemann zeta function for mesoscopic intervals of length $\log^\theta T$, $-1 < \theta \le 0$, 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 (2020), 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 $X_h = \operatorname{Re}\sum_{p\le T} p^{-1/2-\mathrm{i}h}U_p$, where $(U_p,\ p \text{ primes})$ are i.i.d.

uniform random variables on the unit circle, cf. Arguin, Belius and Harper (2017), Arguin and Ouimet (2019), Harper (2013b). 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 (2016), Saksman and Webb (2020). A proof of the freezing phase transition was given in Arguin and Tai (2019). 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. For the rest of the paper, τ denotes a uniform random variable on $[T, 2T]$. For any event $A_T \subseteq [T, 2T]$ and a random variable $X_T : [T, 2T] \to \mathbb{C}$, we write

$$\mathbb{P}(A_T) = \frac{1}{T}\mathrm{Leb}(A_T) \quad \text{and} \quad \mathbb{E}[X_T] = \frac{1}{T}\int_T^{2T} X_T(t)\,\mathrm{d}t.$$

We also use the standard o and $\mathcal{O}$ notations: thus, $f(T) = \mathrm{o}(g(T))$ if $|f(T)/g(T)|$ tends to 0 as $T \to \infty$, when the parameters θ, β and ε are fixed. Similarly, we write $f(T) = \mathcal{O}(g(T))$ if $\limsup |f(T)/g(T)|$ is bounded for θ, β and ε fixed. We 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. In some statements we write $f(T) \ll_A g(T)$ or $f(T) = \mathcal{O}_A(g(T))$ to highlight the dependence on a specific parameter A in the implicit constant. In some of the proofs, we use the common convention that ε denotes an arbitrarily small positive quantity that may vary from line to line. We will also encounter some arithmetical functions familiar in number theory. These include $\omega(n)$ (which counts the number of distinct primes dividing n), $\Omega(n)$ (which counts with multiplicity the number of primes dividing n) and the Möbius function $\mu(n)$ (which equals 0 if n is divisible by the square of a prime and equals $(-1)^{\omega(n)}$ if n is square-free). Throughout the paper, $x \vee y$ and $x \wedge y$ refer to $\max\{x, y\}$ and $\min\{x, y\}$, respectively.

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

$$\mathbb{E}\left[\left|\zeta\left(\frac{1}{2} + \mathrm{i}\tau\right)\right|^{\beta}\right] \ll (\log T)^{\beta^2/4+\varepsilon}, \tag{1.18}$$

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

$$\max_{|h|\leq \log^{\theta} T}\left|D\left(\frac{1}{2} + \mathrm{i}\tau + \mathrm{i}h\right)\right|^{\beta} \ll \sum_{|k|\leq \log^{1+\theta} T}\left|D\left(\frac{1}{2} + \mathrm{i}\tau + \frac{2\pi \mathrm{i}k}{\log T}\right)\right|^{\beta}. \tag{1.19}$$

Equation (1.19) tells us that the process $(\zeta(\frac{1}{2} + \mathrm{i}\tau + \mathrm{i}h),\ |h| \leq \log^{\theta} T)$ varies on a $(\log T)^{-1}$ scale so that the maximum and moments of $\log|\zeta|$ on an interval of length $\mathcal{O}(\log^{\theta} T)$ behave as those of $\mathcal{O}(\log^{1+\theta} T)$ i.i.d. Gaussian random variables of variance $\frac{1}{2}\log\log T$.[1] The limita-

[1] As in the branching random walk setting, the log-correlations are important in the proof of the first-order asymptotics of the maximum, high points and moments, but they do not influence the results. When comparing Gaussian fields, Slepian's lemma tells us that, at equal variance, the field with no correlations will have, on average, the highest maximum and the highest number of points above any fixed proportion of the maximum (the asymptotics of the moments are derived directly from these two quantities). Therefore, the asymptotics of the maximum and moments for i.i.d. Gaussians are always an upper bound for those of log-correlated Gaussian fields. It turns out that we get a matching lower bound by a coarse-graining of the scales following Kistler (2015). This is why our heuristic here is phrased in terms of i.i.d. Gaussians, because the correlations ultimately only matters for the proof, not the actual results.

tion to $\theta \le 3$ comes from the fact that the upper bounds (1.18) 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

$$(\zeta \cdot e^{-\mathcal{P}_{|\theta|}})\Big(\frac{1}{2} + \mathrm{i}\tau\Big), \quad \text{where } \mathcal{P}_\alpha(s) = \sum_{\log p \le \log^\alpha T} \frac{1}{p^s} \quad \text{for } \alpha > 0, \tag{1.20}$$

instead of $\zeta(\frac{1}{2} + \mathrm{i}\tau)$. The reason is that, when $\theta < 0$, the contribution of the primes up to scale $|\theta|$ is negligible with high probability. Namely, with probability $1 - \mathrm{o}(1)$,

$$\max_{|h| \le \log^\theta T} \Big| \mathcal{P}_{|\theta|}\Big(\frac{1}{2} + \mathrm{i}\tau + \mathrm{i}h\Big)\Big| = \mathrm{o}(\log\log T). \tag{1.21}$$

When τ is restricted to a specific event $\mathcal{A}(T)$ on which (1.20) can be discretized as in (1.19), we can show that

$$\mathbb{E}\Big[\Big|(\zeta \cdot e^{-\mathcal{P}_{|\theta|}})\Big(\frac{1}{2} + \mathrm{i}\tau\Big)\Big|^\beta\Big] \ll (\log T)^{(\beta^2/4)\cdot(1+\theta)+\varepsilon} \tag{1.22}$$

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

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 (3.74)), 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 $\frac{1}{2} \le \sigma \le \frac{1}{2} + (\log T)^{\theta-\varepsilon}$ and for any given $\varepsilon > 0$, that with probability $1 - \mathrm{o}(1)$,

$$\int_{-\log^\theta T}^{\log^\theta T} |\zeta(\sigma + \mathrm{i}\tau + \mathrm{i}h)|^\beta \,\mathrm{d}h \ll \int_{-3\log^\theta T}^{3\log^\theta T} \Big|\zeta\Big(\frac{1}{2} + \mathrm{i}\tau + \mathrm{i}h\Big)\Big|^\beta \,\mathrm{d}h + \frac{1}{(\log T)^{96}}. \tag{1.23}$$

This is accomplished 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 $\mathcal{R} = \{\sigma + \mathrm{i}u : |u| \le (\log T)^\theta, \frac{1}{2} \le \sigma \le \frac{1}{2} + (\log T)^{\theta-\varepsilon}\}$ in the whole strip $\frac{1}{2} \le \mathrm{Re}\, s$. 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

$$\int_{-\log^\theta T}^{\log^\theta T} |\zeta(\sigma_0 + \mathrm{i}\tau + \mathrm{i}h)|^\beta \,\mathrm{d}h, \quad \text{with } \sigma_0 = \frac{1}{2} + \frac{1}{(\log T)^{1-\delta}}, \tag{1.24}$$

for some sufficiently small $\delta > 0$. We adapt mollification results from Arguin et al. (2019) to show that, outside of an event of probability $\mathrm{o}(1)$, the problem can be reduced to understanding

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

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

$$\big(\mathrm{Re}\, \mathcal{P}_{1-\delta}(\sigma_0 + \mathrm{i}\tau + \mathrm{i}h), |h| \le \log^\theta T\big). \tag{1.26}$$

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 (1.26) can be computed

using Lemma A.3 with $a(p) = p^{-\sigma_0}(p^{-\mathrm{i}h} + p^{-\mathrm{i}h'})$,

$$\begin{aligned}&\mathbb{E}\big[\operatorname{Re}\mathcal{P}_{1-\delta}(\sigma_0+\mathrm{i}\tau+\mathrm{i}h)\cdot\operatorname{Re}\mathcal{P}_{1-\delta}(\sigma_0+\mathrm{i}\tau+\mathrm{i}h')\big]\\ &\quad=\frac{1}{2}\sum_{\log p\le(\log T)^{1-\delta}}\frac{\cos(|h-h'|\log p)}{p^{2\sigma_0}}+\mathcal{O}(1).\end{aligned}\tag{1.27}$$

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. (This is also true for $\log|\zeta|$ in the sense of finite-dimensional distributions, as shown in Bourgade (2010).)

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*

$$P_k(h)=\sum_{e^{k-1}<\log p\le e^k}\operatorname{Re}\frac{1}{p^{\sigma_0+\mathrm{i}\tau+\mathrm{i}h}},\quad 1\le k\le\log\log T.\tag{1.28}$$

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 (1.27). 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 on a grid with mesh $(\log T)^{-1}$. 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 (1.26) 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 $\sum_k P_k(h)$, $h\in[-1,1]$ can be seen as a branching random walk on a Galton–Watson tree with an average number of offspring e, cf. Figure 1.

When $\theta<0$, the tree structure suggests that the primes up to $\exp(\log^{|\theta|}T)$ 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. Corollary 2.12. This is equivalent to restricting to a subtree of the one on $[-1,1]$ with $(1+\theta)\log\log T$ increments and $\log^{1+\theta}T$ 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 when $|h-h'|>1$. Therefore, the right probabilistic model seems to consist of $\asymp\log^{\theta}T$ 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), Bovier (2017) for a review. It turns out that the large values at leading order correspond to the ones of a REM with $\log^{1+\theta}T$ variables of 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.

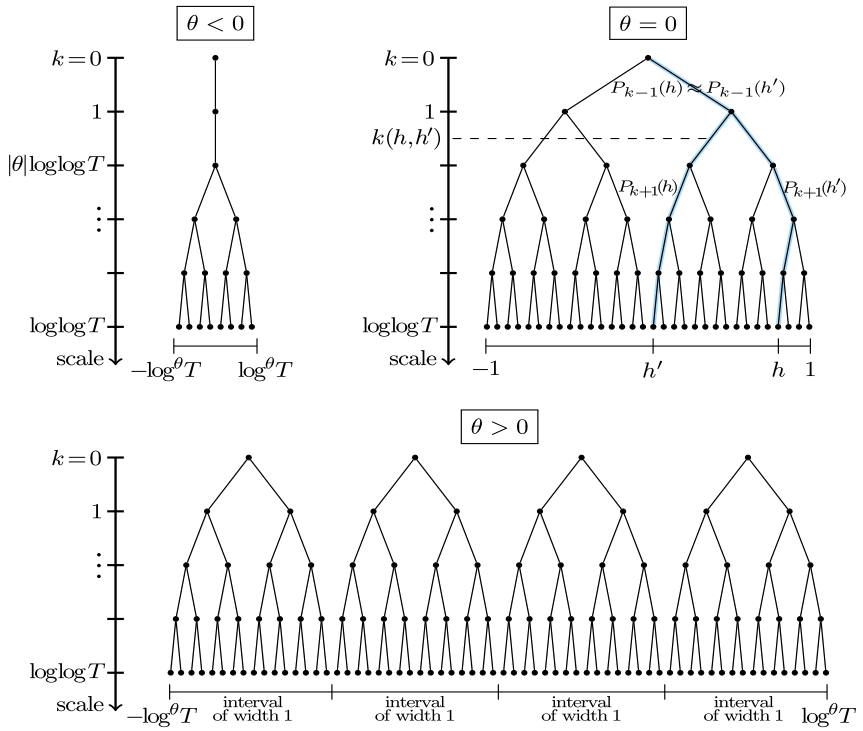


FIG. 1. *An illustration of the branching structure of* $\sum_k P_k$ *for the interval* $[-\log^\theta T, \log^\theta T]$ *with* $\theta < 0$, $\theta = 0$ *and* $\theta > 0$. *When* $\theta < 0$, *the increments* P_k *approximately coincide at each scale up to scale* $|\theta| \log\log T$ (*so one branch*), *followed by the structure of a branching random walk. When* $\theta = 0$, *we have one branching random walk starting at scale* 0. *When* $\theta > 0$, *there is a forest of* $\asymp \log^\theta T$ *approximately independent branching random walks*.

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*,

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

PROOF. See Corollary A in Soundararajan (2009). □

PROPOSITION 2.2. *Let* $0 < \beta \le 4$ *be given. Then*,

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

PROOF. See Theorem 1 in Heap, Radziwiłł and Soundararajan (2019). □

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

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

see Proposition and Lemma 2 in Soundararajan (2009). On the Riemann hypothesis the upper bounds in Theorem 1.1 and Theorem 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 2.3, 2.4, 2.5 and Proposition 2.6 below.

In order to compute the moments of $\zeta \cdot e^{-\mathcal{P}_{|\theta|}}$, we will need to express $e^{-\mathcal{P}_{|\theta|}}$ as a finite Dirichlet polynomial. To this aim, notice that if $|z| \le \nu/10$ for some $\nu \in \mathbb{N}$, we have $|e^z - \sum_{j=0}^{\nu} \frac{z^j}{j!}| \le e^{-\nu}$. Consider, more generally, $e^{\lambda \mathcal{P}(s)}$ with $\lambda \in \mathbb{C}$ and $\mathcal{P}(s) = \sum_{p \le X} a(p) p^{-s}$ for some completely multiplicative function a. We have by the above, assuming $|\lambda \mathcal{P}(s)| \le \nu/10$ for some $\nu \in \mathbb{N}$ and by the multinomial formula, that

$$\left| e^{\lambda \mathcal{P}(s)} - \sum_{k=0}^{\nu} \frac{\lambda^k}{k!} \Big(\sum_{p \le X} \frac{a(p)}{p^s} \Big)^k \right| = \left| e^{\lambda \mathcal{P}(s)} - \sum_{\substack{\Omega(n) \le \nu \\ p|n \Longrightarrow p \le X}} \frac{\lambda^{\Omega(n)} a(n) \mathfrak{g}(n)}{n^s} \right| \le e^{-\nu}, \tag{2.4}$$

where $\Omega(n)$ is the number of prime factors of n with multiplicity. Here, $\mathfrak{g}$ is the multiplicative function defined by $\mathfrak{g}(p^k) = 1/k!$ for all integers k and primes p.

The relevant function a for $e^{-\mathcal{P}_{|\theta|}}$ will be of the following form: Given $\alpha, \beta \in \mathbb{R}$ and $\theta > -1$, let $\mathfrak{F}_{\alpha,\beta,\theta}(n)$ denote a completely multiplicative function such that

$$\mathfrak{F}_{\alpha,\beta,\theta}(p) := \begin{cases} \alpha, & \text{if } \log p \le \log^{|\theta|} T, \\ \beta, & \text{if } \log^{|\theta|} T < \log p. \end{cases} \tag{2.5}$$

In the next three lemmas, we control various terms with the aim of proving the moment estimate in Proposition 2.6, which we will need in the case of short intervals.

LEMMA 2.3. *Let $-1 < \theta < 0$, $\beta > 0$ and $\varepsilon > 0$ be given. Then,*

$$\mathbb{E}\Big[\Big| \sum_{\substack{\Omega(n) \le 100 \lfloor \log\log T \rfloor \\ p|n \Longrightarrow \log p \le \log^{1-\varepsilon} T}} \frac{\mathfrak{F}_{0,\beta/2,\theta}(n) \mathfrak{g}(n)}{n^{1/2+i\tau}} \Big|^2 \Big] \ll (\log T)^{\beta^2(1+\theta)/4}. \tag{2.6}$$

PROOF. Notice that the Dirichlet polynomial in (2.6) has length $\ll T^{\delta}$ for any fixed $\delta > 0$. In particular, by the mean-value formula (Lemma A.2),

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

Dropping the restriction on $\Omega(n)$ and expressing the sum as an Euler product yield

$$\sum_{p|n \Longrightarrow \log p \le \log^{1-\varepsilon} T} \frac{\mathfrak{F}_{0,\beta/2,\theta}(n)^2 \mathfrak{g}(n)^2}{n} = \prod_{\log p \le \log^{1-\varepsilon} T} \Big(1 + \sum_{k \ge 1} \frac{\mathfrak{F}_{0,\beta/2,\theta}(p)^{2k} \mathfrak{g}(p^k)^2}{p^k} \Big). \tag{2.7}$$

The logarithm of the right-hand side is easily evaluated using the prime number theorem (see Lemma A.1) and is $(\beta^2(1+\theta)/4) \log\log T + \mathcal{O}(1)$. This proves the claimed bound. □

LEMMA 2.4. *Let $-1 < \theta < 0$, $0 < \beta \le 2$ and $\varepsilon > 0$ be given. Then,*

$$\mathbb{E}\Big[\Big| \zeta\Big(\frac{1}{2} + i\tau\Big) \Big|^2 \cdot \Big| \sum_{\substack{\Omega(n) \le 100 \lfloor \log\log T \rfloor \\ p|n \Longrightarrow \log p \le \log^{1-\varepsilon} T}} \frac{\mathfrak{F}_{-1,\beta/2-1,\theta}(n) \mathfrak{g}(n)}{n^{1/2+i\tau}} \Big|^2 \Big] \ll (\log T)^{\beta^2(1+\theta)/4+\varepsilon}. \tag{2.8}$$

PROOF. By Theorem 1 in Bettin, Chandee and Radziwiłł (2017), the left-hand side of (2.8) is

$$\begin{aligned}
&\le \frac{1}{T}\int_{\mathbb{R}} \sum_{\substack{\Omega(n)\le 100\lfloor \log\log T\rfloor \\ \Omega(m)\le 100\lfloor \log\log T\rfloor \\ p|n \Longrightarrow \log p\le \log^{1-\varepsilon} T \\ p|m \Longrightarrow \log p\le \log^{1-\varepsilon} T}} \frac{\mathfrak{F}_{-1,\beta/2-1,\theta}(nm)\mathfrak{g}(n)\mathfrak{g}(m)}{[n,m]} \log\Big(\frac{(n,m)^2}{2\pi nm}\Big)\Phi\Big(\frac{t}{T}\Big)\,\mathrm{d}t \\
&+ \frac{1}{T}\int_{\mathbb{R}} \sum_{\substack{\Omega(n)\le 100\lfloor \log\log T\rfloor \\ \Omega(m)\le 100\lfloor \log\log T\rfloor \\ p|n \Longrightarrow \log p\le \log^{1-\varepsilon} T \\ p|m \Longrightarrow \log p\le \log^{1-\varepsilon} T}} \frac{\mathfrak{F}_{-1,\beta/2-1,\theta}(nm)\mathfrak{g}(n)\mathfrak{g}(m)}{[n,m]} (\log t + 2\gamma)\Phi\Big(\frac{t}{T}\Big)\,\mathrm{d}t \qquad (2.9) \\
&+ \mathcal{O}\big(T^{-\varepsilon}\big),
\end{aligned}$$

where Φ is a smooth nonnegative function such that $\Phi(x)\ge 1$ for all $1\le x\le 2$, with support contained in say $[0,3]$, and (n,m) and $[n,m]$ stand for the greatest common divisor and the least common multiple, respectively.

We first note that if n, m have the prime factorization $n=\prod_{i=1}^r p_i^{\alpha_i}$ and $m=\prod_{i=1}^r p_i^{\beta_i}$, where some α_i's and β_i's are possibly 0, then $[n,m]=\prod_{i=1}^r p_i^{\alpha_i\vee\beta_i}$. This means that if $a(n)$ and $b(m)$ are two bounded multiplicative functions, we have

$$\sum_{\substack{p|n\Longrightarrow p\le X \\ p|m\Longrightarrow p\le X}} \frac{a(n)b(m)}{[n,m]} = \prod_{p\le X}\Big(1+\sum_{k\ge 1}p^{-k}\sum_{i,j:\max(i,j)=k} a(p^i)b(p^j)\Big). \qquad (2.10)$$

Using Chernoff's bound, we can get rid of the restriction $\Omega(n)\le 100\lfloor\log\log T\rfloor$ in (2.9). It suffices to notice that the contribution of each sum over n with $\Omega(n)>100\lfloor\log\log T\rfloor$ is

$$\begin{aligned}
&\ll \log T \sum_{\substack{p|n\Longrightarrow p\le T \\ p|m\Longrightarrow p\le T}} \frac{|\mathfrak{F}_{-1,\beta/2-1,\theta}(nm)|\mathfrak{g}(n)\mathfrak{g}(m)}{[n,m]} e^{\Omega(n)-100\log\log T} \\
&\ll (\log T)^{-99}\prod_{p\le T}\Big(1+\frac{(1+e)|\mathfrak{F}_{-1,\beta/2-1,\theta}(p)|}{p}+\frac{e\cdot|\mathfrak{F}_{-1,\beta/2-1,\theta}(p)|^2}{p}\Big) \qquad (2.11) \\
&\ll (\log T)^{-99}\cdot(\log T)^{1+2e} = \mathrm{o}(1),
\end{aligned}$$

where we used (2.10) with $a(n)=|\mathfrak{F}_{-1,\beta/2-1,\theta}(n)|\mathfrak{g}(n)e^{\Omega(n)}$, $b(m)=|\mathfrak{F}_{-1,\beta/2-1,\theta}(m)|\mathfrak{g}(m)$. The contribution of each sum over m with $\Omega(m)>100\lfloor\log\log T\rfloor$ can be removed in the same manner.

Considering the sums in (2.9) without the restriction on $\Omega(n)$ and $\Omega(m)$, we get, by (2.10) and Lemma A.1,

$$\begin{aligned}
&\sum_{\substack{p|n\Longrightarrow \log p\le \log^{1-\varepsilon} T \\ p|m\Longrightarrow \log p\le \log^{1-\varepsilon} T}} \frac{\mathfrak{F}_{-1,\beta/2-1,\theta}(nm)\mathfrak{g}(n)\mathfrak{g}(m)}{[n,m]} \\
&\asymp \prod_{\log p\le \log^{1-\varepsilon} T}\Big(1+\frac{2\mathfrak{F}_{-1,\beta/2-1,\theta}(p)+\mathfrak{F}_{-1,\beta/2-1,\theta}(p)^2}{p}\Big) \qquad (2.12) \\
&\asymp (\log T)^{-|\theta|}\cdot(\log T)^{(\beta^2/4-1)\cdot(1+\theta-\varepsilon)} \\
&\ll (\log T)^{\beta^2(1+\theta)/4-1+\varepsilon}.
\end{aligned}$$

In particular, this means that the second integral in (2.9) is $\ll (\log T)^{\beta^2(1+\theta)/4+\varepsilon}$.

To evaluate the first integral in (2.9), write

$$\log\Big(\frac{(m,n)^2}{mn}\Big) = \frac{1}{2\pi \mathrm{i}} \oint_{|z|=1/\log T} \Big(\frac{(m,n)^2}{mn}\Big)^z \cdot \frac{\mathrm{d}z}{z^2}. \tag{2.13}$$

Then, we end up having to evaluate

$$\frac{1}{2\pi \mathrm{i}} \oint_{|z|=1/\log T} \sum_{\substack{p|n \Longrightarrow \log p \le \log^{1-\varepsilon} T \\ p|m \Longrightarrow \log p \le \log^{1-\varepsilon} T}} \frac{\mathfrak{F}_{-1,\beta/2-1,\theta}(mn)\mathfrak{g}(m)\mathfrak{g}(n)}{[m,n]} \Big(\frac{(m,n)^2}{mn}\Big)^z \cdot \frac{\mathrm{d}z}{z^2}. \tag{2.14}$$

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

$$= \prod_{\log p \le \log^{1-\varepsilon} T} \left(1 + \frac{2\mathfrak{F}_{-1,\beta/2-1,\theta}(p)}{p^{1+z}} + \frac{\mathfrak{F}_{-1,\beta/2-1,\theta}(p)^2}{p} + \sum_{\substack{i,j\ge 0: \\ (i\vee j)\ge 2}} \frac{\mathfrak{F}_{-1,\beta/2-1,\theta}(p)^{i+j}}{i!j!} \frac{p^{(i\wedge j)z}}{p^{(i\vee j)(1+z)}} \right). \tag{2.15}$$

For $|z| = 1/\log T$, note that

$$\Big| \sum_{\substack{i,j\ge 0: \\ (i\vee j)\ge 2}} \frac{\mathfrak{F}_{-1,\beta/2-1,\theta}(p)^{i+j}}{i!j!} \frac{p^{(i\wedge j)z}}{p^{(i\vee j)(1+z)}} \Big| \le \frac{1}{p^2} \sum_{i,j\ge 0} \frac{1}{i!j!} \le \frac{e^2}{p^2}, \tag{2.16}$$

and a Taylor expansion yields

$$\begin{aligned} &\sum_{\log p \le \log^{1-\varepsilon} T} \frac{\mathfrak{F}_{-1,\beta/2-1,\theta}(p)}{p^{1+z}} \\ &= \sum_{\log p \le \log^{1-\varepsilon} T} \frac{\mathfrak{F}_{-1,\beta/2-1,\theta}(p)}{p} + \mathcal{O}\Big(\frac{1}{\log T} \sum_{\log p \le \log^{1-\varepsilon} T} \frac{\log p}{p}\Big). \end{aligned} \tag{2.17}$$

Since the error term in (2.17) is $\mathrm{o}(1)$ by Lemma A.1, the Euler product in (2.15) is

$$\begin{aligned} &\asymp \prod_{\log p \le \log^{1-\varepsilon} T} \Big(1 + \frac{2\mathfrak{F}_{-1,\beta/2-1,\theta}(p) + \mathfrak{F}_{-1,\beta/2-1,\theta}(p)^2}{p} + \mathcal{O}(p^{-2})\Big) \\ &\ll (\log T)^{\beta^2(1+\theta)/4-1+\varepsilon}. \end{aligned} \tag{2.18}$$

By putting this estimate back in the contour integral and using a trivial bound on z^{-2}, equation (2.14) is $\ll (\log T)^{\beta^2(1+\theta)/4+\varepsilon}$, as required. □

LEMMA 2.5. *Let $\varepsilon > 0$ be given. For $\ell = 2\lfloor \log\log T \rfloor$, we have*

$$\mathbb{E}\Big[\Big|\zeta\Big(\frac{1}{2}+\mathrm{i}\tau\Big)\Big|^2 \cdot \Big|\frac{\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau)}{5\log\log T}\Big|^{2\ell}\Big] \ll (\log T)^{-2} \tag{2.19}$$

and

$$\mathbb{E}\Big[\Big|\frac{\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau)}{5\log\log T}\Big|^{2\ell}\Big] \ll (\log T)^{-4}. \tag{2.20}$$

PROOF. First, we apply a moment estimate (Lemma A.4) followed by a prime number theorem estimate (Lemma A.1) to obtain

$$\mathbb{E}\biggl[\biggl|\frac{\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau)}{5\log\log T}\biggr|^{4\ell}\biggr] \ll \frac{(2\ell)!(\sum_{p\le T}p^{-1})^{2\ell}}{5^{4\ell}(\log\log T)^{4\ell}} \ll \sqrt{\ell}\biggl(\frac{4\cdot 2}{25e}\biggr)^{2\ell} \ll e^{-4\ell} \ll (\log T)^{-8}. \tag{2.21}$$

The estimate (2.19) then follows by applying the Cauchy–Schwarz inequality, the fourth moment bound $\mathbb{E}[|\zeta(\frac{1}{2}+\mathrm{i}\tau)|^4] \ll (\log T)^4$; see, for example, Ingham (1927) and (2.21). For (2.20), the same reasoning as in (2.21), yields the estimate $\ll e^{-2\ell} \ll (\log T)^{-4}$. □

The last three lemmas show a moment bound of the right order for $\zeta\cdot e^{-\mathcal{P}_{|\theta|}}$.

PROPOSITION 2.6. *Let $-1<\theta<0$, $0<\beta\le 2$ and $\varepsilon>0$ be given. Then, as $T\to\infty$,*

$$\mathbb{E}\biggl[\biggl|\bigl(\zeta\cdot e^{-\mathcal{P}_{|\theta|}}\bigr)\biggl(\frac{1}{2}+\mathrm{i}\tau\biggr)\biggr|^{\beta}\mathbf{1}_{\mathcal{A}(T)}\biggr] \ll (\log T)^{\beta^2(1+\theta)/4+\varepsilon}, \tag{2.22}$$

with the event

$$\mathcal{A}(T)=\biggl\{\biggl|\mathcal{P}_{|\theta|}\biggl(\frac{1}{2}+\mathrm{i}\tau\biggr)\biggr|\le 2\log\log T\biggr\}. \tag{2.23}$$

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

$$\begin{aligned}\biggl|\zeta\biggl(\frac{1}{2}+\mathrm{i}\tau\biggr)\biggr|^{\beta} &\le \frac{1}{p}\cdot\biggl|\zeta\biggl(\frac{1}{2}+\mathrm{i}\tau\biggr)\biggr|^{2}\cdot e^{-\frac{2}{q}\operatorname{Re}\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau)}+\frac{1}{q}\cdot e^{\frac{2}{p}\operatorname{Re}\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau)}\\ &=\frac{\beta}{2}\cdot\biggl|\zeta\biggl(\frac{1}{2}+\mathrm{i}\tau\biggr)\biggr|^{2}\cdot e^{-(2-\beta)\operatorname{Re}\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau)}+\frac{2-\beta}{2}\cdot e^{\beta\operatorname{Re}\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau)}.\end{aligned} \tag{2.24}$$

Note that (2.24) holds trivially for $\beta=2$. Hence, for $0<\beta\le 2$,

$$\begin{aligned}\biggl|\bigl(\zeta\cdot e^{-\mathcal{P}_{|\theta|}}\bigr)\biggl(\frac{1}{2}+\mathrm{i}\tau\biggr)\biggr|^{\beta} &\le \frac{\beta}{2}\biggl|\zeta\biggl(\frac{1}{2}+\mathrm{i}\tau\biggr)\biggr|^{2}\cdot e^{-(2-\beta)\operatorname{Re}\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau)-\beta\operatorname{Re}\mathcal{P}_{|\theta|}(\frac{1}{2}+\mathrm{i}\tau)}\\ &\quad+\frac{2-\beta}{2}e^{\beta\operatorname{Re}\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau)-\beta\operatorname{Re}\mathcal{P}_{|\theta|}(\frac{1}{2}+\mathrm{i}\tau)}.\end{aligned} \tag{2.25}$$

On the event $\mathcal{A}(T)\cap\{|\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau)|\le 5\log\log T\}$, we get, by the truncation (2.4) with $\nu=100\lfloor\log\log T\rfloor$ and the identity $|z+w|^2\le 2(|z|^2+|w|^2)$, that

$$\begin{aligned}&e^{-(2-\beta)\operatorname{Re}\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau)-\beta\operatorname{Re}\mathcal{P}_{|\theta|}(\frac{1}{2}+\mathrm{i}\tau)}\\ &\ll\biggl|\sum_{\substack{\Omega(n)\le 100\lfloor\log\log T\rfloor\\ p|n\Longrightarrow\log p\le\log^{1-\varepsilon}T}}\frac{\mathfrak{F}_{-1,\beta/2-1,\theta}(n)\mathfrak{g}(n)}{n^{1/2+\mathrm{i}\tau}}\biggr|^{2}+(\log T)^{-200},\end{aligned} \tag{2.26}$$

where $\mathfrak{F}_{\alpha,\beta,\theta}(n)$ is the completely multiplicative function defined in (2.5). Likewise, on the same event we have

$$\begin{aligned}&e^{\beta\operatorname{Re}\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau)-\beta\operatorname{Re}\mathcal{P}_{|\theta|}(\frac{1}{2}+\mathrm{i}\tau)}\\ &\ll\biggl|\sum_{\substack{\Omega(n)\le 100\lfloor\log\log T\rfloor\\ p|n\Longrightarrow\log p\le\log^{1-\varepsilon}T}}\frac{\mathfrak{F}_{0,\beta/2,\theta}(n)\mathfrak{g}(n)}{n^{1/2+\mathrm{i}\tau}}\biggr|^{2}+(\log T)^{-200}.\end{aligned} \tag{2.27}$$

Finally, on the event $\mathcal{A}(T) \cap \{|\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau)| > 5\log\log T\}$, we get, for any $\ell \geq 1$,

$$\left|(\zeta \cdot e^{-\mathcal{P}_{|\theta|}})\left(\frac{1}{2}+\mathrm{i}\tau\right)\right|^{\beta} \leq (\log T)^4 \cdot \left(1+\left|\zeta\left(\frac{1}{2}+\mathrm{i}\tau\right)\right|^2\right) \cdot \left|\frac{\mathcal{P}_{1-\varepsilon}(\frac{1}{2}+\mathrm{i}\tau)}{5\log\log T}\right|^{2\ell}, \tag{2.28}$$

since for $\beta \leq 2$, $|\zeta|^{\beta}$ is bounded by $(1+|\zeta|^2)$ and $|e^{-\mathcal{P}_{|\theta|}}|^{\beta}$ is bounded by $(\log T)^4$ on $\mathcal{A}(T)$. We choose $\ell = 2\lfloor \log\log T \rfloor$. Now take the expectation with τ restricted to $\mathcal{A}(T)$ in (2.25), then split the terms on the right-hand side over the associated events in (2.26), (2.27) and (2.28). We use Lemmas 2.3, 2.4 and 2.5 to bound the expectations. □

2.2. *Discretization.* The analysis of the maximum of zeta on an interval can often be restricted to h's on a grid with mesh of order $(\log T)^{-1}$. 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.

PROPOSITION 2.7. *Let $\theta > -1$, $\beta \geq 1$ and $\varepsilon > 0$ be given. Let $D(s) = \sum_{n \leq T^{1+\varepsilon}} a(n) n^{-s}$ be a Dirichlet polynomial of length $T^{1+\varepsilon}$, where $\sup_{n \leq T^{1+\varepsilon}} |a(n)| \leq B$ for some $B > 0$ possibly depending on ε and T. Then, for all $A > 10(1+\varepsilon)\beta$, $T \leq t \leq 2T$ and $\sigma \geq 1/2$,*

$$\begin{aligned} &\sup_{|h| \leq \log^{\theta} T} \left|D(\sigma+\mathrm{i}t+\mathrm{i}h)\right|^{\beta} \\ &\ll_A \sum_{|k| \leq 2\log^{1+\theta} T} \left|D\left(\sigma+\mathrm{i}t+\frac{2\pi\mathrm{i}k}{(1+2\varepsilon)\log T}\right)\right|^{\beta} \\ &\quad + \sum_{2\log^{1+\theta} T < |k| \leq T} \left|D\left(\sigma+\mathrm{i}t+\frac{2\pi\mathrm{i}k}{(1+2\varepsilon)\log T}\right)\right|^{\beta} \cdot \frac{1}{1+|k|^A} + B^{\beta} T^{-A/2}. \end{aligned} \tag{2.29}$$

PROOF. Let V be a smooth function with $V(x) = 1$ for $x \in [-(1+\varepsilon), 0]$ and compactly supported in $[-(1+2\varepsilon), \varepsilon]$. We show

$$D(\sigma+\mathrm{i}t+\mathrm{i}h) = \frac{1}{1+2\varepsilon} \sum_{k \in \mathbb{Z}} D\left(\sigma+\mathrm{i}t+\frac{2\pi\mathrm{i}k}{(1+2\varepsilon)\log T}\right) \widehat{V}\left(\frac{k}{1+2\varepsilon} - \frac{h\log T}{2\pi}\right). \tag{2.30}$$

By taking the complex norm and applying Hölder's inequality with $\beta \geq 1$, this yields

$$\begin{aligned} \left|D(\sigma+\mathrm{i}t+\mathrm{i}h)\right| &\leq \left(\frac{1}{1+2\varepsilon} \sum_{k \in \mathbb{Z}} \left|D\left(\sigma+\mathrm{i}t+\frac{2\pi\mathrm{i}k}{(1+2\varepsilon)\log T}\right)\right|^{\beta}\right. \\ &\quad \left. \times \left|\widehat{V}\left(\frac{k}{1+2\varepsilon} - \frac{h\log T}{2\pi}\right)\right|\right)^{1/\beta} \\ &\quad \times \left(\frac{1}{1+2\varepsilon} \sum_{k \in \mathbb{Z}} \left|\widehat{V}\left(\frac{k}{1+2\varepsilon} - \frac{h\log T}{2\pi}\right)\right|\right)^{1-1/\beta}. \end{aligned} \tag{2.31}$$

This proves (2.29) after taking the supremum over h, using the rapid decay of $\widehat{V}$ and noticing that $\sup_{n \leq T^{1+\varepsilon}} |a(n)| \leq B$ and our assumption on A imply that

$$\begin{aligned} &\sum_{|k|>T} \left|D\left(\sigma+\mathrm{i}t+\frac{2\pi\mathrm{i}k}{(1+2\varepsilon)\log T}\right)\right|^{\beta} \cdot \frac{1}{1+|k|^A} \\ &\ll B^{\beta} T^{(1+\varepsilon)\beta/2} \sum_{|k|>T} \frac{1}{1+|k|^A} \ll B^{\beta} T^{-A/2}. \end{aligned} \tag{2.32}$$

Since $D(s)$ is of the form $\sum_{n\le T^{1+\varepsilon}} a(n)n^{-s}$, it suffices by linearity to establish (2.30) for a single $n \le T^{1+\varepsilon}$, that is,

$$n^{-\mathrm{i}h} = \frac{1}{1+2\varepsilon}\sum_{k\in\mathbb{Z}} n^{-\frac{2\pi \mathrm{i}k}{(1+2\varepsilon)\log T}} \cdot \widehat{V}\Big(\frac{k}{1+2\varepsilon} - \frac{h\log T}{2\pi}\Big), \quad 1 \le n \le T^{1+\varepsilon}. \tag{2.33}$$

Using the Poisson summation formula, the right-hand side can be rewritten as

$$\begin{aligned}
&\sum_{\ell\in\mathbb{Z}} \int_{\mathbb{R}} \frac{1}{1+2\varepsilon} \exp\Big(2\pi \mathrm{i} y\Big(-\ell - \frac{1}{1+2\varepsilon}\frac{\log n}{\log T}\Big)\Big) \cdot \widehat{V}\Big(\frac{y}{1+2\varepsilon} - \frac{h\log T}{2\pi}\Big)\,\mathrm{d}y \\
&= n^{-\mathrm{i}h} \sum_{\ell\in\mathbb{Z}} e^{-\mathrm{i}\ell(1+2\varepsilon)h\log T} \int_{\mathbb{R}} \exp\Big(2\pi \mathrm{i} u\Big(-\ell(1+2\varepsilon) - \frac{\log n}{\log T}\Big)\Big) \cdot \widehat{V}(u)\,\mathrm{d}u \qquad (2.34)\\
&= n^{-\mathrm{i}h} \sum_{\ell\in\mathbb{Z}} e^{-\mathrm{i}\ell(1+2\varepsilon)h\log T} \cdot V\Big(-\ell(1+2\varepsilon) - \frac{\log n}{\log T}\Big),
\end{aligned}$$

where we made the change of variable $y = (1+2\varepsilon)(u + \frac{h\log T}{2\pi})$. The term $\ell = 0$ is equal to $n^{-\mathrm{i}h}$ since $V(-\frac{\log n}{\log T}) = 1$ for $1 \le n \le T^{1+\varepsilon}$ by the choice of V. The other terms ($\ell \ne 0$) are all equal to 0 since $-\ell(1+2\varepsilon) - \frac{\log n}{\log T}$ falls outside the support of V for $1 \le n \le T^{1+\varepsilon}$. This proves (2.33) and the proposition. □

Proposition 2.7 implies five important corollaries to tackle the maximum of ζ and of Dirichlet polynomials. We first observe that the discretization applies to ζ in Corollary 2.9. This is a consequence of the following approximation.

LEMMA 2.8 (Approximation of ζ). *Let $\varepsilon > 0$ and $\sigma \ge 1/2$ be given, and let $k > \max\{5, 10/\varepsilon\}$ be an integer. Then, as $T \to \infty$ and for $t \asymp T$, we have*

$$\zeta(\sigma+\mathrm{i}t) = \begin{cases} \displaystyle\sum_{n\le T^{1+\varepsilon}} n^{-\sigma-\mathrm{i}t} w_k\Big(\log\frac{e^k n}{T^{1+\varepsilon}}\Big) + \mathcal{O}_k(T^{-k\varepsilon/2}), & \text{if } 1/2 \le \sigma \le 2, \\ \displaystyle\sum_{n\le T} n^{-\sigma-\mathrm{i}t} + \mathcal{O}(T^{-1}), & \text{if } \sigma > 2, \end{cases} \tag{2.35}$$

where the smoothing w_k is defined by setting

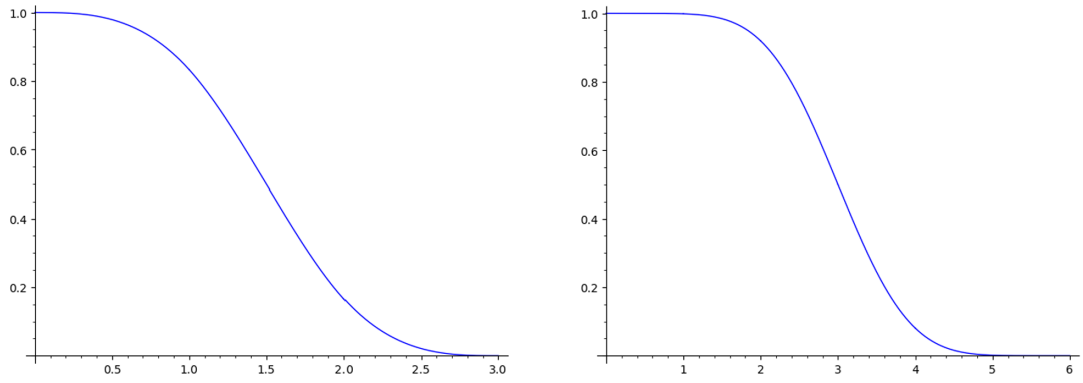
$$w_k(x) := \begin{cases} 1, & \text{if } x < 0, \\ (-1)^k \displaystyle\sum_{\ell=0}^{k} \binom{k}{\ell} \frac{(-1)^\ell}{k!} (\ell - x)_+^k, & \text{if } 0 \le x < k, \\ 0, & \text{if } x \ge k, \end{cases} \tag{2.36}$$

where $(y)_+ := \max\{y, 0\}$. Examples of graphs for $w_k(x)$ are provided in Figure 2.

PROOF. The case $\sigma > 2$ is a trivial consequence of the fact that $|\sum_{n>T} n^{-\sigma-\mathrm{i}t}| \le |\sum_{n>T} n^{-2}| \le T^{-1}$. Therefore, assume $1/2 \le \sigma \le 2$. We claim that

$$w_k(x) = \frac{1}{2\pi\mathrm{i}} \int_{2-\mathrm{i}\infty}^{2+\mathrm{i}\infty} e^{-xz} \Big(\frac{e^z - 1}{z}\Big)^k \frac{\mathrm{d}z}{z}. \tag{2.37}$$

First, it is easy to check that this formula holds for $x < 0$ and $x > k$: if $x < 0$, then we shift the contour toward $\mathrm{Re}\, z = -\infty$ and collect a single pole with residue 1 at $z = 0$, while if



(a) Graph of $w_3(x)$ (b) Graph of $w_6(x)$

FIG. 2. *Examples of graphs for* $w_k(x)$.

$x > k$, then we shift the contour toward $\operatorname{Re} z = \infty$, and we see that the integral is zero. In the remaining intermediate range $0 \le x < k$, we expand

$$\left(\frac{e^z-1}{z}\right)^k = \frac{(-1)^k}{z^k}\sum_{\ell=0}^{k}\binom{k}{\ell}(-1)^\ell e^{\ell z}, \tag{2.38}$$

and we use the fact that

$$\frac{1}{2\pi \mathrm{i}}\int_{2-\mathrm{i}\infty}^{2+\mathrm{i}\infty} e^{-xz}e^{\ell z}\frac{\mathrm{d}z}{z^{k+1}} = \frac{1}{k!}\begin{cases}(\ell-x)^k, & \text{if } \ell - x > 0,\\ 0, & \text{if } \ell - x \le 0.\end{cases} \tag{2.39}$$

Therefore,

$$\sum_{n\ge 1}\frac{1}{n^{\sigma+\mathrm{i}t}}w_k\left(\log\frac{n}{T^{1+\varepsilon}}\right) = \frac{1}{2\pi \mathrm{i}}\int_{2-\mathrm{i}\infty}^{2+\mathrm{i}\infty}\zeta(\sigma+\mathrm{i}t+z)T^{(1+\varepsilon)z}\left(\frac{e^z-1}{z}\right)^k\frac{\mathrm{d}z}{z}. \tag{2.40}$$

We now shift the contour to the line $\operatorname{Re} z = -(k-2)$. We collect a pole at $z=0$ with residue $\zeta(\sigma+\mathrm{i}t)$. On the line $\operatorname{Re} z = -(k-2)$, we bound the integral using the estimate $|\zeta(r+\mathrm{i}t)| \ll (1+|t|)^{1/2-r}$ which is valid for any fixed $r < -\frac{1}{100}$ and all $t \in \mathbb{R}$.[2] Specifically, the contribution of the line $\operatorname{Re} z = -(k-2)$ is bounded by

$$T^{-(1+\varepsilon)(k-2)}\int_{\mathbb{R}}(T^{k-2}+|u|^{k-2})\frac{2^k\,\mathrm{d}u}{|2-k+\mathrm{i}u|^{k+1}} \ll_k T^{-\varepsilon k/2}. \tag{2.41}$$

This proves that

$$\zeta(\sigma+\mathrm{i}t) = \sum_{n\ge 1}\frac{1}{n^{\sigma+\mathrm{i}t}}w_k\left(\log\frac{n}{T^{1+\varepsilon}}\right) + \mathcal{O}_k(T^{-k\varepsilon/2}), \quad \text{for } t \asymp T. \tag{2.42}$$

The conclusion follows by a simple rescaling. □

From Lemma 2.8 we derive the following discretization result.

[2]This estimate follows from applying the functional equation for $\zeta(r+\mathrm{i}t)$, bounding the ratio of Gamma factors using Stirling's formula and bounding $\zeta(1-r-\mathrm{i}t)$ trivially by $O(1)$.

COROLLARY 2.9. *Let $\theta > -1$, $\beta \geq 1$ and $\varepsilon > 0$ be given. For any $A > 10(\varepsilon^{-1}+1)\beta$ and all $T \leq t \leq 2T$,*

$$\begin{aligned}
&\max_{|h|\leq \log^{\theta} T} \left|\zeta\left(\frac{1}{2}+\mathrm{i}t+\mathrm{i}h\right)\right|^{\beta} \\
&\ll_A \sum_{|\ell|\leq 2\log^{1+\theta} T} \left|\zeta\left(\frac{1}{2}+\mathrm{i}t+\frac{2\pi\mathrm{i}\ell}{(1+2\varepsilon)\log T}\right)\right|^{\beta} \\
&\quad + \sum_{2\log^{1+\theta} T<|\ell|\leq T} \left|\zeta\left(\frac{1}{2}+\mathrm{i}t+\frac{2\pi\mathrm{i}\ell}{(1+2\varepsilon)\log T}\right)\right|^{\beta} \cdot \frac{1}{1+|\ell|^{A}} + T^{-A/2}.
\end{aligned} \tag{2.43}$$

PROOF. This is a consequence of the ζ approximation in Lemma 2.8 with $\sigma = 1/2$ and $k = \lfloor (A+1)/(\beta\varepsilon)\rfloor$ as well as the discretization in Proposition 2.7 with $B = 1$. □

As a consequence we get a suboptimal upper bound for $\theta > -1$ using the second moment. Note that this bound also works for θ dependent on T.

COROLLARY 2.10. *Let $0 < \varepsilon \leq 1$ be given, and let $k > 10/\varepsilon$ be an integer. Then, for any $\theta > -1$, possibly dependent on T, we have*

$$\mathbb{P}\left(\max_{|h|\leq \log^{\theta} T}\left|\sum_{n\leq T^{1+\varepsilon}} \frac{1}{n^{\frac{1}{2}+\mathrm{i}\tau+\mathrm{i}h}} w_k\left(\log\frac{e^k n}{T^{1+\varepsilon}}\right)\right| > 2^{\theta}(\log T)^{2+\theta}\right) \ll \frac{2^{-\theta}}{\log T}. \tag{2.44}$$

PROOF. The Dirichlet polynomial in (2.44) is $\ll \sum_{n\leq T^{1+\varepsilon}} n^{-1/2} \ll T^{(1+\varepsilon)/2} \ll T$, so the probability is just zero when $\theta > \log T/\log\log T$. Therefore, we assume that $\theta \leq \log T/\log\log T$. By the ζ approximation in Lemma 2.8, it suffices to prove

$$\mathbb{P}\left(\max_{|h|\leq \log^{\theta} T}\left|\zeta\left(\frac{1}{2}+\mathrm{i}\tau+\mathrm{i}h\right)\right| > 2^{\theta/2}(\log T)^{2+\theta}\right) \ll \frac{2^{-\theta}}{\log T}. \tag{2.45}$$

By applying Markov's inequality and Corollary 2.9 with $A = 100(\varepsilon^{-1}+1)$, the probability in (2.45) is

$$\begin{aligned}
&\leq 2^{-\theta}(\log T)^{-4-2\theta}\,\mathbb{E}\left[\max_{|h|\leq \log^{\theta} T}\left|\zeta\left(\frac{1}{2}+\mathrm{i}\tau+\mathrm{i}h\right)\right|^2\right] \\
&\ll 2^{-\theta}(\log T)^{-4-2\theta} \sum_{|\ell|\leq 2\log^{1+\theta} T} \mathbb{E}\left[\left|\zeta\left(\frac{1}{2}+\mathrm{i}\tau+\frac{2\pi\mathrm{i}\ell}{(1+2\varepsilon)\log T}\right)\right|^2\right] \\
&\quad + 2^{-\theta}(\log T)^{-4-2\theta} \sum_{2\log^{1+\theta} T<|\ell|\leq T} \mathbb{E}\left[\left|\zeta\left(\frac{1}{2}+\mathrm{i}\tau+\frac{2\pi\mathrm{i}\ell}{(1+2\varepsilon)\log T}\right)\right|^2\right]\cdot\frac{1}{1+|\ell|^{100}} \\
&\quad + T^{-50}.
\end{aligned} \tag{2.46}$$

Using a standard second moment bound, see, for example, Titchmarsh (1986), p. 141, the last two expectations are $\ll \log T$. We conclude that the right-hand side of (2.46) is

$$\ll 2^{-\theta}(\log T)^{-2-\theta} \ll \frac{2^{-\theta}}{\log T}, \tag{2.47}$$

since $\theta > -1$. □

A similar reasoning, using Markov's inequality, can be applied to get an upper bound for the maximum of $\mathcal{P}_\alpha$, $0 < \alpha < 1$. The bound below is suboptimal for $\theta < 0$ and optimal for $\theta \geq 0$.

COROLLARY 2.11. *Let $\theta > -1$, $\varepsilon > 0$ and $\sigma \geq 1/2$ be given. Then,*

$$\mathbb{P}\Big(\max_{|h|\leq \log^\theta T} |\mathcal{P}_\alpha(\sigma + i\tau + ih)| > \big(\sqrt{\alpha(1+\theta)} + \varepsilon\big)\log\log T\Big) = o(1). \tag{2.48}$$

PROOF. We apply Markov's inequality with exponent 2ℓ and discretize, as in (2.46), using Proposition 2.7 with $B = 1$. We then use moment estimates from Lemma A.4, with $\ell = \lfloor (1+\theta)\log\log T \rfloor$, to bound the expectations. □

When $\theta < 0$ and $\alpha > |\theta|$, the bound (2.48) (and its analogue for ζ) needs to be refined by discarding the contribution of small primes. The result below directly implies that for $\theta < 0$ and $\alpha > |\theta|$, the sharp upper bound for $\operatorname{Re}\mathcal{P}_\alpha$ is $\sqrt{(\alpha+\theta)(1+\theta)}\log\log T$ since the effective variance is $\frac{(\alpha+\theta)}{2}\log\log T$.

COROLLARY 2.12. *Let $-1 < \theta < 0$ and $\sigma \geq 1/2$ be given. Then, for any $0 < \varepsilon < C$ and $V = V(T)$ that satisfies $\varepsilon \log\log T \leq V \leq C\log\log T$, we have*

$$\mathbb{P}\Big(\max_{|h|\leq \log^\theta T} |\mathcal{P}_{|\theta|}(\sigma + i\tau + ih)| > V\Big) \ll e^{-cV} \tag{2.49}$$

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

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

$$\begin{aligned}\mathbb{P}\Big(\max_{|h|\leq \log^\theta T} |S(h)| > V\Big) &\leq \mathbb{P}\Big(\max_{|h|\leq \log^\theta T} |S(h) - S(0)| > V/2\Big)\\ &\quad + \mathbb{P}\big(|S(0)| > V/2\big).\end{aligned} \tag{2.50}$$

Let ℓ denote a generic natural integer. By Markov's inequality, a moment estimate (Lemma A.4) and a prime number theorem estimate (Lemma A.1), we have

$$\mathbb{P}\big(|S(0)| > V/2\big) \leq \frac{\mathbb{E}[|S(0)|^{2\ell}]}{(V/2)^{2\ell}} \ll \ell!\Big(\frac{\sum_{p\leq T} p^{-2\sigma}}{(V/2)^2}\Big)^\ell \ll \Big(\frac{4\ell\log\log T}{\varepsilon^2(\log\log T)^2}\Big)^\ell. \tag{2.51}$$

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 (2.50). Let ℓ denote another natural integer to be chosen later. By applying Proposition 2.7, we get

$$\mathbb{E}\Big[\max_{|h|\leq \log^\theta T} |S(h) - S(0)|^{2\ell}\Big] \ll \log^{1+\theta} T \cdot \max_{|h|\leq \log^\theta T} \mathbb{E}\big[|S(h) - S(0)|^{2\ell}\big]. \tag{2.52}$$

A short calculation, using moment estimates (Lemma A.4) followed by prime number theorem estimates (Lemma A.1), yields

$$\max_{|h|\leq \log^\theta T} \mathbb{E}\big[|S(h) - S(0)|^{2\ell}\big] \ll \ell!\Big(\sum_{\log p \leq \log^{|\theta|} T} \frac{2 - 2\cos(|h|\log p)}{p}\Big)^\ell \ll (\ell d)^\ell \tag{2.53}$$

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

Then, by Markov's inequality and the choice $\ell = \lfloor \frac{\varepsilon^2}{8d} \log\log T \rfloor$, we deduce

$$\mathbb{P}\Big(\max_{|h|\le \log^\theta T} |S(h) - S(0)| > V/2 \Big) \ll \log^{1+\theta} T \cdot \Big(\frac{4\ell d}{V^2} \Big)^\ell \ll e^{-bV} \tag{2.54}$$

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

As before, the maximum of $\zeta \cdot e^{-\mathcal{P}_{|\theta|}}$ can be discretized by truncating the exponential.

COROLLARY 2.13. *Let $-1 < \theta \le 0$ and $\varepsilon > 0$ be given. Then, there exists a constant $C = C(\theta, \varepsilon) > 0$ such that the event*

$$\begin{aligned} &\max_{|h|\le \log^\theta T} \Big| \big(\zeta \cdot e^{-\mathcal{P}_{|\theta|}}\big)\Big(\frac12 + \mathrm{i}\tau + \mathrm{i}h\Big)\Big|^2 \\ &\le C \sum_{|\ell|\le 2\log^{1+\theta} T} \Big| \big(\zeta \cdot e^{-\mathcal{P}_{|\theta|}}\big)\Big(\frac12 + \mathrm{i}\tau + \frac{2\pi \mathrm{i}\ell}{(1+2\varepsilon)\log T}\Big)\Big|^2 \\ &\quad + C \sum_{2\log^{1+\theta} T < |\ell| \le T} \Big| \big(\zeta \cdot e^{-\mathcal{P}_{|\theta|}}\big)\Big(\frac12 + \mathrm{i}\tau + \frac{2\pi \mathrm{i}\ell}{(1+2\varepsilon)\log T}\Big)\Big|^2 \cdot \frac{1}{1+|\ell|^{100}} + T^{-50} \end{aligned} \tag{2.55}$$

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

PROOF. Define the event

$$\widetilde{\mathcal{A}}(T) = \Big\{ \max_{|h|\le \log^\theta T} \Big| \mathcal{P}_{|\theta|}\Big(\frac12 + \mathrm{i}\tau + \mathrm{i}h\Big)\Big| \le 2\log\log T \Big\}. \tag{2.56}$$

By Corollary 2.12 we have $\mathbb{P}(\widetilde{\mathcal{A}}(T)) = 1 - \mathrm{o}(1)$. By (2.4), for all $\tau \in \widetilde{\mathcal{A}}(T)$, we also have

$$\begin{aligned} \Big| \sum_{\substack{\Omega(n)\le 20\lfloor \log\log T\rfloor \\ p|n \Longrightarrow \log p \le \log^{|\theta|} T}} \frac{(-1)^{\Omega(n)} \mathfrak{g}(n)}{n^{1/2+\mathrm{i}\tau+\mathrm{i}h}} \Big| &= \big| e^{-\mathcal{P}_{|\theta|}(\frac12+\mathrm{i}\tau+\mathrm{i}h)} \big| + \mathcal{O}\big((\log T)^{-20}\big) \\ &\asymp \big| e^{-\mathcal{P}_{|\theta|}(\frac12+\mathrm{i}\tau+\mathrm{i}h)} \big|. \end{aligned} \tag{2.57}$$

Combining this with the ζ approximation in Lemma 2.8 with $\sigma = 1/2$ and $k = 102/\varepsilon$, we conclude that, for all $\tau \in \widetilde{\mathcal{A}}(T)$ and uniformly for $y \asymp T$,

$$\Big| \big(\zeta \cdot e^{-\mathcal{P}_{|\theta|}}\big)\Big(\frac12 + \mathrm{i}\tau + \mathrm{i}y\Big)\Big| \asymp \Big| D\Big(\frac12 + \mathrm{i}\tau + \mathrm{i}y\Big)\Big| + \mathcal{O}\Big(\frac{\log^2 T}{T^{51}}\Big), \tag{2.58}$$

where D is a Dirichlet polynomial of length $T^{1+2\varepsilon}$. Proposition 2.7 implies

$$\begin{aligned} &\max_{|h|\le \log^\theta T} \Big| D\Big(\frac12 + \mathrm{i}\tau + \mathrm{i}h\Big)\Big|^2 \\ &\ll \sum_{|\ell|\le 2\log^{1+\theta} T} \Big| D\Big(\frac12 + \mathrm{i}\tau + \frac{2\pi \mathrm{i}\ell}{(1+2\varepsilon)\log T}\Big)\Big|^2 \\ &\quad + \sum_{2\log^{1+\theta} T < |\ell| \le T} \Big| D\Big(\frac12 + \mathrm{i}\tau + \frac{2\pi \mathrm{i}\ell}{(1+2\varepsilon)\log T}\Big)\Big|^2 \cdot \frac{1}{1+|\ell|^{100}} + T^{-50}. \end{aligned} \tag{2.59}$$

Together with (2.58), this concludes the proof. □

2.3. *Proofs of the upper bounds.*

2.3.1. *The case $\theta \geq 0$.* PROOF OF THEOREM 1.2 FOR $\theta \geq 0$. By Markov's inequality with exponent $\beta > 0$, we have

$$\begin{aligned}\mathbb{P}\Big(\max_{|h|\leq \log^\theta T}\Big|\zeta\Big(\frac{1}{2}+\mathrm{i}\tau+\mathrm{i}h\Big)\Big| > (\log T)^{m(\theta)+\varepsilon}\Big)\\ \ll (\log T)^{-\beta m(\theta)-\beta\varepsilon}\cdot\mathbb{E}\Big[\max_{|h|\leq \log^\theta T}\Big|\zeta\Big(\frac{1}{2}+\mathrm{i}\tau+\mathrm{i}h\Big)\Big|^\beta\Big].\end{aligned} \tag{2.60}$$

If we choose $\beta = 2m(\theta) \geq 2$, we get, by picking A large enough in Corollary 2.9, that the right-hand side of the above equation is

$$\ll (\log T)^{-2m(\theta)^2-2\varepsilon+1+\theta}\cdot\mathbb{E}\Big[\Big|\zeta\Big(\frac{1}{2}+\mathrm{i}\tau\Big)\Big|^\beta\Big]. \tag{2.61}$$

By applying Proposition 2.2 if $\beta \leq 4$ (i.e., if $\theta \leq 3$) and Proposition 2.1 if $\beta > 4$ (i.e., if $\theta > 3$), the expectation is bounded by $(\log T)^{m(\theta)^2+\varepsilon}$. Therefore, the claim follows. □

PROOF OF THEOREM 1.1 FOR $\theta \geq 0$. For all $\beta > 0$, Markov's inequality yields

$$\begin{aligned}\mathbb{P}\Big(\int_{|h|\leq \log^\theta T}\Big|\zeta\Big(\frac{1}{2}+\mathrm{i}\tau+\mathrm{i}h\Big)\Big|^\beta \mathrm{d}t \geq (\log T)^{f_\theta(\beta)+\varepsilon}\Big)\\ \ll (\log T)^{-f_\theta(\beta)-\varepsilon}\log^\theta T\cdot\mathbb{E}\Big[\Big|\zeta\Big(\frac{1}{2}+\mathrm{i}\tau\Big)\Big|^\beta\Big].\end{aligned} \tag{2.62}$$

When $\beta \leq 2\sqrt{1+\theta}$, we have $f_\theta(\beta) = \beta^2/4 + \theta$, so the right-hand side of (2.62) is $\ll (\log T)^{-\varepsilon/2}$ by Proposition 2.2 for $\theta \leq 3$ and by Proposition 2.1 for $\theta > 3$.

It remains to sharpen the bound in the case $\beta > 2\sqrt{1+\theta}$. We use the Lebesgue measure of high points. Let $a, b > 0$. Two successive applications of Markov's inequality yield

$$\begin{aligned}\mathbb{P}\Big(\mathrm{Leb}\Big\{|h| \leq \log^\theta T : \Big|\zeta\Big(\frac{1}{2}+\mathrm{i}\tau+\mathrm{i}h\Big)\Big| > (\log T)^a\Big\} \geq (\log T)^{-a^2+\theta+\varepsilon}\Big)\\ \ll (\log T)^{a^2-\varepsilon}\cdot(\log T)^{-ba}\cdot\mathbb{E}\Big[\Big|\zeta\Big(\frac{1}{2}+\mathrm{i}\tau\Big)\Big|^b\Big].\end{aligned} \tag{2.63}$$

Again, the optimal bound is at $b = 2a$. Using Proposition 2.2 for $\theta \leq 3$ and Proposition 2.1 for $\theta > 3$ and choosing $b = 2a$, we conclude that this is $\ll (\log T)^{-\varepsilon/2}$ for $0 < a \leq m(\theta)$.

We now partition the integral according to the value of the integrand. Let $M \geq 1$ be an integer and $0 \leq j \leq M$. Theorem 1.2 (for $\theta \geq 0$) and the above imply that, with probability $1 - o(1)$,

$$\begin{aligned}\int_{|h|\leq \log^\theta T}\Big|\zeta\Big(\frac{1}{2}+\mathrm{i}\tau+\mathrm{i}h\Big)\Big|^\beta \mathrm{d}h\\ \ll \sum_{0\leq j\leq M}(\log T)^{\beta((j+1)/M)m(\theta)}\cdot(\log T)^{-(j/M)^2 m(\theta)^2+\theta+\varepsilon}.\end{aligned} \tag{2.64}$$

For $\beta > 2\sqrt{1+\theta} \geq 2m(\theta)$, the last term $j = M$ dominates, and, in particular, the above is bounded by

$$\ll (\log T)^{\beta m(\theta)-m(\theta)^2+\theta+2\varepsilon} = (\log T)^{\beta m(\theta)-1+2\varepsilon}, \tag{2.65}$$

provided that M is chosen sufficiently large with respect to θ, β and ε. □

REMARK. In the above proof we could have handled all β's using the Lebesgue measure of high points in the spirit of a Gibbs variational principle. We chose to prove the case $\beta \le 2\sqrt{1+\theta}$ directly, as the proof is straightforward.

2.3.2. *The case* $\theta < 0$. PROOF OF THEOREM 1.2 FOR $\theta < 0$. We notice that

$$\begin{aligned}
&\mathbb{P}\Big(\max_{|h|\le \log^\theta T}\Big|\zeta\Big(\frac12+i\tau+ih\Big)\Big| > (\log T)^{m(\theta)+\varepsilon}\Big)\\
&\le \mathbb{P}\Big(\max_{|h|\le \log^\theta T}\Big|(\zeta\cdot e^{-\mathcal{P}_{|\theta|}})\Big(\frac12+i\tau+ih\Big)\Big| > (\log T)^{m(\theta)+\varepsilon/2}\Big)\\
&\quad+\mathbb{P}\Big(\max_{|h|\le \log^\theta T}\big|e^{\mathcal{P}_{|\theta|}(\frac12+i\tau+ih)}\big| > (\log T)^{\varepsilon/2}\Big).
\end{aligned}\tag{2.66}$$

By Corollary 2.12 the last term is o(1), as $T\to\infty$. As in (2.56), let

$$\widetilde{\mathcal{A}}(T) = \Big\{\max_{|h|\le \log^\theta T}\Big|\mathcal{P}_{|\theta|}\Big(\frac12+i\tau+ih\Big)\Big| \le 2\log\log T\Big\}.\tag{2.67}$$

By Corollary 2.12 again, the probability of $\widetilde{\mathcal{A}}(T)$ is $1-o(1)$. We let $\mathcal{A}_0(T)$ denote the subset of $\widetilde{\mathcal{A}}(T)$ for which the conclusion of Corollary 2.13 holds. The probability of $\mathcal{A}_0(T)$ is $1-o(1)$. Then, by Markov's inequality we have

$$\begin{aligned}
&\mathbb{P}\Big(\Big\{\max_{|h|\le \log^\theta T}\Big|(\zeta\cdot e^{-\mathcal{P}_{|\theta|}})\Big(\frac12+i\tau+ih\Big)\Big| > (\log T)^{m(\theta)+\varepsilon/2}\Big\}\cap\mathcal{A}_0(T)\Big)\\
&\le (\log T)^{-2m(\theta)-\varepsilon}\cdot\mathbb{E}\Big[\max_{|h|\le \log^\theta T}\Big|(\zeta\cdot e^{-\mathcal{P}_{|\theta|}})\Big(\frac12+i\tau+ih\Big)\Big|^2\mathbf{1}_{\mathcal{A}_0(T)}\Big].
\end{aligned}\tag{2.68}$$

By Corollary 2.13, and since $m(\theta)=1+\theta$, this is

$$\ll (\log T)^{-(1+\theta)-\varepsilon}\cdot\mathbb{E}\Big[\Big|(\zeta\cdot e^{-\mathcal{P}_{|\theta|}})\Big(\frac12+i\tau\Big)\Big|^2\mathbf{1}_{\widetilde{\mathcal{A}}(T)}\Big].\tag{2.69}$$

By Proposition 2.6 this is

$$\ll (\log T)^{-(1+\theta)-\varepsilon}\cdot(\log T)^{(1+\theta)+\varepsilon/2} \ll (\log T)^{-\varepsilon/2},\tag{2.70}$$

as needed. □

PROOF OF THEOREM 1.1 FOR $\theta < 0$. Similarly to (2.66), we can restrict the integrand to $\zeta\cdot e^{-\mathcal{P}_{|\theta|}}$ as follows:

$$\begin{aligned}
&\mathbb{P}\Big(\int_{|h|\le \log^\theta T}\Big|\zeta\Big(\frac12+i\tau+ih\Big)\Big|^\beta dh > (\log T)^{f_\theta(\beta)+\varepsilon}\Big)\\
&\le \mathbb{P}\Big(\int_{|h|\le \log^\theta T}\Big|(\zeta\cdot e^{-\mathcal{P}_{|\theta|}})\Big(\frac12+i\tau+ih\Big)\Big|^\beta dh > (\log T)^{f_\theta(\beta)+\varepsilon/2}\Big)+o(1).
\end{aligned}\tag{2.71}$$

As in (2.67), the probability is $\mathbb{P}(\widetilde{\mathcal{A}}(T))=1-o(1)$, and by Markov's inequality we have

$$\begin{aligned}
&\mathbb{P}\Big(\Big\{\int_{|h|\le \log^\theta T}\Big|(\zeta\cdot e^{-\mathcal{P}_{|\theta|}})\Big(\frac12+i\tau+ih\Big)\Big|^\beta dh > (\log T)^{f_\theta(\beta)+\varepsilon/2}\Big\}\cap\widetilde{\mathcal{A}}(T)\Big)\\
&\ll (\log T)^{-f_\theta(\beta)-\varepsilon/2}\cdot\log^\theta T\cdot\mathbb{E}\Big[\Big|(\zeta\cdot e^{-\mathcal{P}_{|\theta|}})\Big(\frac12+i\tau\Big)\Big|^\beta\mathbf{1}_{\widetilde{\mathcal{A}}(T)}\Big].
\end{aligned}\tag{2.72}$$

By Proposition 2.6 the above is

$$\ll (\log T)^{-(\beta^2/4)(1+\theta)-\varepsilon/2} \cdot (\log T)^{(\beta^2/4)\cdot(1+\theta)+\varepsilon/4} \ll (\log T)^{-\varepsilon/4}. \tag{2.73}$$

This bound proves the claim for $\beta \le 2$.

It remains to refine the bound for the case $\beta > 2$. This proceeds in the same way as in the proof of Theorem 1.1 in the case $\theta \ge 0$, with ζ replaced by $\zeta \cdot e^{-\mathcal{P}_{|\theta|}}$ restricted on the event $\widetilde{\mathcal{A}}(T)$. Namely, we have, for $0 < a \le m(\theta)$,

$$\begin{aligned} &\mathbb{P}\left(\left\{\begin{array}{l}\operatorname{Leb}\{|h| \le \log^\theta T : \left|(\zeta \cdot e^{-\mathcal{P}_{|\theta|}})(\frac{1}{2}+\mathrm{i}\tau+\mathrm{i}h)\right| > (\log T)^a\} \\ \ge (\log T)^{-a^2+\theta+\varepsilon}\end{array}\right\} \cap \widetilde{\mathcal{A}}(T)\right) \\ &\quad \ll (\log T)^{a^2-\varepsilon} \cdot (\log T)^{-ba} \cdot \mathbb{E}\left[\left|(\zeta \cdot e^{-\mathcal{P}_{|\theta|}})\left(\frac{1}{2}+\mathrm{i}\tau\right)\right|^b \mathbf{1}_{\widetilde{\mathcal{A}}(T)}\right]. \end{aligned} \tag{2.74}$$

This is o(1) by Proposition 2.6 with the optimal choice $b = 2a/(1+\theta) \le 2$. The remainder is done exactly as in the proof of Theorem 1.1 in the case $\theta \ge 0$, by partitioning the integral over values of the integrand in the range $[0, m(\theta)+\varepsilon]$. □

3. Lower bounds. In this section, we prove:

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

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

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

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

The lower bound for the maximum will be an easy consequence of the lower bound for the moments. The idea is to approximate zeta by an appropriate Dirichlet polynomial. This can be done with good precision off-axis, cf. Section 3.1. 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. (2019) 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 of Gabriel (1927) in the special case $a = b = 1$). *Let F be a complex valued function which is analytic in the strip $\alpha \le \operatorname{Re} z \le \beta$. Suppose that $|F(z)|$ tends to zero as $|\operatorname{Im} z| \to \infty$ uniformly for $\alpha \le \operatorname{Re} z \le \beta$. Then, for any $\gamma \in [\alpha, \beta]$ and any $p > 0$,*

$$I(\gamma) \le I(\alpha)^{(\beta-\gamma)/(\beta-\alpha)} \cdot I(\beta)^{(\gamma-\alpha)/(\beta-\alpha)}, \tag{3.3}$$

where

$$I(\sigma) := \int_{\mathbb{R}} |F(\sigma+\mathrm{i}t)|^p \,\mathrm{d}t. \tag{3.4}$$

This theorem has the following useful consequence.

COROLLARY 3.4. *Let F be a complex valued function which is analytic in the strip $\frac{1}{2} \le \operatorname{Re} z$. Suppose that $|F(z)|$ tends to zero as $|\operatorname{Im} z| \to \infty$ uniformly for $\frac{1}{2} \le \operatorname{Re} z$. Suppose also that $I(\sigma) \to 0$ as $\sigma \to \infty$. Then, for any $\sigma > \frac{1}{2}$ and any $p > 0$,*

$$I(\sigma) \le I\left(\frac{1}{2}\right). \tag{3.5}$$

PROOF. Let $\sigma^\star$ be such that

$$I(\sigma^\star) = \sup_{\sigma \ge 1/2} I(\sigma). \tag{3.6}$$

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 $\gamma = \sigma^\star$, $\alpha = \frac{1}{2}$ and $\beta = \sigma^\star + \varepsilon$, we get

$$I(\sigma^\star) \le I\left(\frac{1}{2}\right)^{\lambda} \cdot I(\sigma^\star + \varepsilon)^{\mu}, \tag{3.7}$$

for some appropriate $\lambda, \mu > 0$ that satisfy $\lambda + \mu = 1$.

Therefore, by definition of $\sigma^\star$ in (3.6),

$$I(\sigma^\star) \le I\left(\frac{1}{2}\right)^{\lambda} \cdot I(\sigma^\star)^{\mu}, \tag{3.8}$$

and hence $I(\sigma^\star)^{\lambda} \le I(\frac{1}{2})^{\lambda}$. Since $\lambda > 0$, we get $I(\sigma^\star) \le I(\frac{1}{2})$. The claim follows from (3.6). □

We now construct a special analytic approximation for the indicator function of the rectangle $\mathcal{R} = \{\sigma + \mathrm{i}v : \frac{1}{2} \le \sigma \le \frac{1}{2} + K, |v| \le L\}$ for $K, L > 0$. The effective width of the indicator function will be $K \approx L/\Delta$ in the statement below.

LEMMA 3.5. *Let $b_1 \in (0,1)$ and $\Delta, L, A, b_2 > 0$ 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 $v \in \mathbb{R}$:*

(i) *For $|v| \ge (1+b_2)L$, uniformly in $\sigma \ge \frac{1}{2}$, $\Phi_{\Delta,L}(z) \ll_A b_2^{-A} \Delta^{1-A}$.*
(ii) *For any $|v| \le (1-b_1)L$, $|\Phi_{\Delta,L}(z)| = 1 + \mathcal{O}_{b_1,A}(\Delta^{-A}) + \mathcal{O}((\sigma - \frac{1}{2})\frac{\Delta^2}{L})$.*
(iii) *For any $|v| \le (1+b_2)L$, $|\Phi_{\Delta,L}(z)| \ll 1 + (\sigma - \frac{1}{2})\frac{\Delta^2}{L}$.*
(iv) *$\Phi_{\Delta,L}(z) \to 0$ 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 $\eta > 0$ and given $z \in \mathbb{C}$ with $\operatorname{Re} z \ge \frac{1}{2}$ and $u \in \mathbb{R}$, consider the following function:

$$\delta_\eta(z) = \int_0^\infty e^{-2\pi(z-\frac{1}{2})x} \cdot V(\eta x)\eta \, \mathrm{d}x. \tag{3.9}$$

Then, $\delta_\eta(z)$ defines an entire function of exponential type. By integration by parts, we see that

$$\delta_\eta(z) \ll_A \left(1 + \left|z - \frac{1}{2}\right|\eta^{-1}\right)^{-A}, \tag{3.10}$$

for any $A > 0$ and uniformly in $\operatorname{Re} z \geq \frac{1}{2}$. Therefore, we may think of $\delta_\eta(z)$ as localizing to $z = \frac{1}{2} + \mathcal{O}(\eta)$. Furthermore, notice that if $z = \frac{1}{2} + \mathrm{i}v$ and $u \in \mathbb{R}$, then

$$\delta_\eta(z - \mathrm{i}u) = \widehat{V}\big((v-u)\eta^{-1}\big), \tag{3.11}$$

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

$$\begin{aligned}\delta_\eta(z - \mathrm{i}u) &= \int_0^\infty e^{-2\pi(\sigma - \frac{1}{2} + \mathrm{i}(v-u))x} \cdot V(\eta x)\eta \,\mathrm{d}x \\ &= \int_0^\infty e^{-2\pi \mathrm{i}(v-u)x} \cdot \left(1 + \mathcal{O}\left(\left(\sigma - \frac{1}{2}\right)x\right)\right) \cdot V(\eta x)\eta \,\mathrm{d}x \\ &= \widehat{V}\big((v-u)\eta^{-1}\big) + \mathcal{O}\left(\left(\sigma - \frac{1}{2}\right)\eta^{-1}\right).\end{aligned} \tag{3.12}$$

Finally, for $z = \sigma + \mathrm{i}v$ with $\sigma \geq \frac{1}{2}$, we have from (3.10) that

$$\big|\delta_\eta(z - \mathrm{i}u)\big| \ll_A \frac{1}{1 + (|v-u|\eta^{-1})^A}. \tag{3.13}$$

The candidate function is, for $\eta = L/\Delta$,

$$\Phi_{\Delta,L}(z) = \frac{\Delta}{L}\int_{-L}^{L} e^{-2\pi \mathrm{i}u\frac{\Delta}{L}} \cdot \delta_{L/\Delta}(z - \mathrm{i}u)\,\mathrm{d}u. \tag{3.14}$$

We will now describe some of the features of this function. Write $z = \sigma + \mathrm{i}v$ with $\sigma \geq \frac{1}{2}$. Using the bound (3.13), we see that, if $|v| > (1+b_2)L$ with $b_2 > 0$, then

$$\Phi_{\Delta,L}(z) \ll_A \frac{\Delta}{L}\int_{-L}^{L} \frac{1}{1 + (|v-u|\frac{\Delta}{L})^A}\,\mathrm{d}u \ll_A b_2^{-A}\Delta^{1-A}. \tag{3.15}$$

This gives the first claim.

If $|v| \leq (1-b_1)L$, then by (3.14) and (3.12) we have

$$\Phi_{\Delta,L}(z) = \frac{\Delta}{L}\int_{-L}^{L} e^{-2\pi \mathrm{i}u\frac{\Delta}{L}} \cdot \widehat{V}\left((v-u)\frac{\Delta}{L}\right)\mathrm{d}u + \mathcal{O}\left(\left(\sigma - \frac{1}{2}\right)\frac{\Delta^2}{L}\right). \tag{3.16}$$

It follows that if $\frac{1}{2} \leq \sigma$ and $|v| \leq (1-b_1)L$, then, due to the rapid decay of $\widehat{V}$, we have

$$\begin{aligned}\Phi_{\Delta,L}(z) &= e^{-2\pi \mathrm{i}v\frac{\Delta}{L}}\int_{v\frac{\Delta}{L}-\Delta}^{v\frac{\Delta}{L}+\Delta} e^{2\pi \mathrm{i}u} \cdot \widehat{V}(u)\,\mathrm{d}u + \mathcal{O}\left(\left(\sigma - \frac{1}{2}\right)\frac{\Delta^2}{L}\right) \\ &= e^{-2\pi \mathrm{i}v\frac{\Delta}{L}} + \mathcal{O}_{b_1,A}\big(\Delta^{-A}\big) + \mathcal{O}\left(\left(\sigma - \frac{1}{2}\right)\frac{\Delta^2}{L}\right)\end{aligned} \tag{3.17}$$

by Fourier inversion and the assumption that $V(1) = 1$. This proves the second claim. If $\frac{1}{2} \leq \sigma \ll 1$ and $|v| \leq (1+b_2)L$, then we have the bound

$$\big|\Phi_{\Delta,L}(z)\big| \ll \int_{\mathbb{R}} \big|\widehat{V}(u)\big|\,\mathrm{d}u + \mathcal{O}\left(\left(\sigma - \frac{1}{2}\right)\frac{\Delta^2}{L}\right) \tag{3.18}$$

which proves the third claim.

Finally, notice that $\delta_{L/\Delta}(z - \mathrm{i}u) \to 0$ uniformly as $\sigma \to \infty$ by (3.10), which implies the last claim that $\Phi_{\Delta,L}(z) \to 0$ uniformly in $v \in \mathbb{R}$ as $\sigma \to \infty$. $\square$

The following proposition relates the moments off and on axis.

PROPOSITION 3.6. *Let $\theta > -1$, $\beta > 0$, $0 < \varepsilon \le 1$ and $T \ge 10^9$ be given. Then, for all $\frac{1}{2} \le \sigma \le \frac{1}{2} + (\log T)^{\theta-3\varepsilon}$, the event*

$$\int_{-\log^\theta T}^{\log^\theta T} \bigl|\zeta(\sigma + \mathrm{i}\tau + \mathrm{i}u)\bigr|^\beta \,\mathrm{d}u \ll \int_{-3\log^\theta T}^{3\log^\theta T} \left|\zeta\left(\frac{1}{2} + \mathrm{i}\tau + \mathrm{i}u\right)\right|^\beta \mathrm{d}u + \frac{1}{(\log T)^{96}} \tag{3.19}$$

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

PROOF. Let

$$D(\sigma + \mathrm{i}\tau) = \begin{cases} \displaystyle\sum_{n \le T^{1+\varepsilon}} n^{-\sigma-\mathrm{i}\tau} w_k\left(\frac{e^k n}{T^{1+\varepsilon}}\right), & \text{if } 1/2 \le \sigma \le 2, \\ \displaystyle\sum_{n \le T} n^{-\sigma-\mathrm{i}\tau}, & \text{if } \sigma > 2, \end{cases} \tag{3.20}$$

with $0 < \varepsilon \le 1$ and $k > 10/\varepsilon$ a fixed integer. Using the ζ approximation in Lemma 2.8, we have, for $T \le \tau \le 2T$ and $\frac{1}{2} \le \sigma \le \frac{1}{2} + (\log T)^{\theta-3\varepsilon}$,

$$\zeta(\sigma + \mathrm{i}\tau) = D(\sigma + \mathrm{i}\tau) + \mathcal{O}\bigl(T^{-1}\bigr). \tag{3.21}$$

Therefore, it suffices to establish (3.19) for ζ replaced by D,

$$\int_{-\log^\theta T}^{\log^\theta T} \bigl|D(\sigma + \mathrm{i}\tau + \mathrm{i}u)\bigr|^\beta \,\mathrm{d}u \ll \int_{-3\log^\theta T}^{3\log^\theta T} \left|D\left(\frac{1}{2} + \mathrm{i}\tau + \mathrm{i}u\right)\right|^\beta \mathrm{d}u + \frac{1}{(\log T)^{96}}. \tag{3.22}$$

Consider

$$I(\sigma) = \int_{\mathbb{R}} \bigl|D(\sigma + \mathrm{i}\tau + \mathrm{i}u)\bigr|^\beta \cdot \bigl|\Phi_{\Delta,L}(\sigma + \mathrm{i}u)\bigr|^\beta \,\mathrm{d}u, \tag{3.23}$$

with $\Delta = \log^\varepsilon T$ and $L = 1.5\log^\theta T$. Then, by Lemma 3.5(i) and (iv), Corollary 3.4 can be applied and yields

$$\begin{aligned} &\int_{\mathbb{R}} \bigl|D(\sigma + \mathrm{i}\tau + \mathrm{i}u)\bigr|^\beta \cdot \bigl|\Phi_{\Delta,L}(\sigma + \mathrm{i}u)\bigr|^\beta \,\mathrm{d}u \\ &\quad\ll \int_{\mathbb{R}} \left|D\left(\frac{1}{2} + \mathrm{i}\tau + \mathrm{i}u\right)\right|^\beta \cdot \left|\Phi_{\Delta,L}\left(\frac{1}{2} + \mathrm{i}u\right)\right|^\beta \mathrm{d}u. \end{aligned} \tag{3.24}$$

Now it remains to unsmooth both sides of this expression. Lemma 3.5(ii) (with $b_1 = 1/3$) implies that $\Phi_{\Delta,L}(\sigma + \mathrm{i}u) \gg 1$ for $|u| \le \log^\theta T$. We thus have

$$\int_{-\log^\theta T}^{\log^\theta T} \bigl|D(\sigma + \mathrm{i}\tau + \mathrm{i}u)\bigr|^\beta \,\mathrm{d}u \ll \int_{\mathbb{R}} \bigl|D(\sigma + \mathrm{i}\tau + \mathrm{i}u)\bigr|^\beta \cdot \bigl|\Phi_{\Delta,L}(\sigma + \mathrm{i}u)\bigr|^\beta \,\mathrm{d}u, \tag{3.25}$$

settling the left-hand side of (3.22). For the right-hand side, note that the choice $\Delta = \log^\varepsilon T$ and $L = 1.5\log^\theta T$ ensures that the error term $(\sigma - \frac{1}{2})\frac{\Delta^2}{L}$ in Lemma 3.5 is $(\log T)^{-\varepsilon}$ for $\sigma - \frac{1}{2} \le (\log T)^{\theta-3\varepsilon}$. Lemma 3.5(iii) (with $b_2 = 1$) shows that the right-hand side of (3.24) is

$$\begin{aligned} &\int_{\mathbb{R}} \left|D\left(\frac{1}{2} + \mathrm{i}\tau + \mathrm{i}u\right)\right|^\beta \cdot \left|\Phi_{\Delta,L}\left(\frac{1}{2} + \mathrm{i}u\right)\right|^\beta \mathrm{d}u \\ &\quad\ll \int_{-3\log^\theta T}^{3\log^\theta T} \left|D\left(\frac{1}{2} + \mathrm{i}\tau + \mathrm{i}u\right)\right|^\beta \mathrm{d}u \\ &\qquad+ \sum_{\ell=0}^{\infty} \int_{\mathcal{U}_\ell} \left|D\left(\frac{1}{2} + \mathrm{i}\tau + \mathrm{i}u\right)\right|^\beta \cdot \left|\Phi_{\Delta,L}\left(\frac{1}{2} + \mathrm{i}u\right)\right|^\beta \mathrm{d}u, \end{aligned} \tag{3.26}$$

where $\mathcal{U}_\ell = \{3(\log T)^{\theta+\ell} \le |u| \le 3(\log T)^{\theta+\ell+1}\}$. By Corollary 2.10 and a union bound, the event

$$\mathcal{S}(T) = \bigcap_{\ell=0}^{\infty} \left\{ \max_{|u|\le \log^\ell T} \left| D\left(\frac{1}{2} + \mathrm{i}\tau + \mathrm{i}u\right)\right| \le 2^\ell (\log T)^{2+\ell} \right\} \tag{3.27}$$

has probability $1 - o(1)$. Moreover, by Lemma 3.5 (i) with $A = 1 + \frac{100}{\varepsilon}(\lceil\theta\rceil + 1)(1 + 1/\beta)$ and $b_2 = 2(\log T)^\ell - 1$, we have, for all $3(\log T)^{\theta+\ell} \le |u|$,

$$\left|\Phi_{\Delta,L}\left(\frac{1}{2} + \mathrm{i}u\right)\right| \ll (\log T)^{-100\ell(1+1/\beta)} \cdot (\log T)^{-100(\lceil\theta\rceil+1)\cdot(1+1/\beta)}. \tag{3.28}$$

Therefore, on the event $\mathcal{S}(T)$ and for every integer $\ell \ge 0$, the following holds:

$$\begin{aligned} &\int_{\mathcal{U}_\ell} \left| D\left(\frac{1}{2} + \mathrm{i}\tau + \mathrm{i}u\right)\right|^\beta \cdot \left|\Phi_{\Delta,L}\left(\frac{1}{2} + \mathrm{i}u\right)\right|^\beta \mathrm{d}u \\ &\quad \ll (\log T)^{\lceil\theta\rceil+\ell+1} \cdot 2^{\beta\ell} (\log T)^{\beta(2+\ell)} \cdot (\log T)^{-100(\beta+1)(\lceil\theta\rceil+\ell+1)} \\ &\quad \ll (\log T)^{-96(\beta+1)\cdot(\lceil\theta\rceil+\ell+1)}. \end{aligned} \tag{3.29}$$

Thus, on $\mathcal{S}(T)$ the contribution of the sum on the right-hand side of (3.26) is negligible. The claim follows by combining equations (3.24), (3.25) and (3.26). □

3.2. *Mollification.* This step is an adaptation of Section 4.2 of Arguin et al. (2019), which is itself based on the work of Radziwiłł and Soundararajan (2017). The treatment is slightly different as the width of the interval needs to be taken into account. Also, we choose to use the discretization in Proposition 2.7 to obtain a uniform control on the interval, as opposed to a Sobolev inequality.

The main idea is to define a mollifier for the zeta function

$$M(s) = \sum_{\substack{\Omega(n)\le \nu_\theta \\ p|n \Longrightarrow p\le X}} \frac{\mu(n)}{n^s}, \tag{3.30}$$

where

$$X = \exp\big((\log T)^{1-K^{-1}}\big), \quad \text{for } K \ge 2, \text{ and } \nu_\theta = 100 K e^{\theta\vee 0} \log\log T. \tag{3.31}$$

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 nonsquare-free. The estimate will be done slightly off-axis,

$$\sigma_0 = \frac{1}{2} + \frac{(\log T)^{3/(2K)}}{\log T}. \tag{3.32}$$

The parameter K will eventually be assumed to be large enough, depending on θ, β and ε.

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

LEMMA 3.7. *Let $\theta > -1$ and $\varepsilon > 0$ be given. Then,*

$$\mathbb{P}\left(\max_{|h|\le \log^\theta T} \left|(\zeta \cdot M)(\sigma_0 + \mathrm{i}\tau + \mathrm{i}h) - 1\right| > \varepsilon \right) = o(1). \tag{3.33}$$

This was proved in the case $\theta = 0$ in Lemma 4.2 of Arguin et al. (2019). In particular, it also holds *verbatim* for $-1 < \theta < 0$ since the interval is just smaller. 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. *Let $\theta > 0$ be given. Then,*

$$\mathbb{E}\big[\big|(\zeta \cdot M)(\sigma_0 + \mathrm{i}\tau) - 1\big|^2\big] \ll (\log T)^{-100e^{\theta}}. \tag{3.34}$$

PROOF. The proof follows Arguin et al. (2019) with a new error term, due to the choice of ν_θ. (The manipulations are very similar to the ones in Lemma 2.4.) The error appears after equation (4.10) in Arguin et al. (2019) and is given by

$$(\log T)e^{-\nu_\theta} \prod_{p \le X} (1 + 7p^{-1}). \tag{3.35}$$

The Euler product is $\ll (\log T)^7$, using Lemma A.1. Using this and the definition of ν_θ in (3.31) yields

$$(\log T)e^{-\nu_\theta} \prod_{p \le X} (1 + 7p^{-1}) \ll (\log T)^8 \cdot (\log T)^{-100Ke^{\theta}}. \tag{3.36}$$

Since $K \ge 2$, 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 in Arguin et al. (2019), is

$$\ll \sum_{p>X} p^{-2\sigma_0} \ll X^{-(\sigma_0 - 1/2)} = \exp\big(-(\log T)^{\frac{1}{2K}}\big) \ll (\log T)^{-100e^{\theta}}. \tag{3.37}$$

This ends the proof. □

PROOF OF LEMMA 3.7 FOR $\theta > 0$. By Lemma 2.8, ζ is well approximated by a Dirichlet polynomial of length $T^{1+\varepsilon}$ for any given $\varepsilon > 0$. Moreover, M is a Dirichlet polynomial of length less than T^{ε} for any given $\varepsilon > 0$. Therefore, an application of Markov's inequality and Proposition 2.7 yield that the probability in (3.33) is

$$\ll \log^{1+\theta} T \cdot \mathbb{E}\big[\big|(\zeta \cdot M)(\sigma_0 + \mathrm{i}\tau) - 1\big|^2\big]. \tag{3.38}$$

The conclusion follows from Lemma 3.8. □

3.3. *Approximation of the mollifier.* We now approximate the mollifier M by the exponential of a Dirichlet polynomial. If we let

$$\widetilde{\mathcal{P}}_{1-K^{-1}}(s) = \sum_{k \ge 1} \sum_{p \le X} \frac{1}{kp^{ks}}, \tag{3.39}$$

then the following relation between $\exp(-\widetilde{\mathcal{P}}_{1-K^{-1}}(s))$ and $M(s)$ holds for all $\operatorname{Re} s \ge 1/2$:

$$\exp\big(-\widetilde{\mathcal{P}}_{1-K^{-1}}(s)\big) = \exp\Big(\log \prod_{p \le X} (1 - p^{-s})\Big) = M(s) + \sum_{\substack{\Omega(n) > \nu_\theta \\ p|n \Longrightarrow p \le X}} \frac{\mu(n)}{n^s}. \tag{3.40}$$

In particular, we see that $\exp(-\widetilde{\mathcal{P}}_{1-K^{-1}}(s))$ and $M(s)$ only differ for integers n with more than ν_θ prime factors ($\Omega(n) > \nu_\theta$) and all their prime factors $\le X$. The following lemma make use of this fact to estimate how close they are when $s = \sigma_0 + \mathrm{i}\tau + \mathrm{i}h$.

LEMMA 3.9. *Let $\theta > -1$ be given. Then, for any $K \ge 2$, we have*

$$\mathbb{P}\Big(\max_{|h| \le \log^{\theta} T} \big|(M - \exp(-\widetilde{\mathcal{P}}_{1-K^{-1}}))(\sigma_0 + \mathrm{i}\tau + \mathrm{i}h)\big| > (\log T)^{-10}\Big) = \mathrm{o}(1). \tag{3.41}$$

PROOF. The discretization in Proposition 2.7, together with the mean value theorem in Lemma A.2, yield

$$\begin{aligned}\mathbb{E}\Big[\max_{|h|\le \log^\theta T} \big|M-\exp(-\widetilde{\mathcal{P}}_{1-K^{-1}})\big|^2(\sigma_0+\mathrm{i}\tau+\mathrm{i}h)\Big]\\ \ll \log^{1+\theta} T\cdot \sum_{\substack{\Omega(n)>\nu_\theta\\ p|n\Longrightarrow p\le X}} n^{-1}.\end{aligned}\tag{3.42}$$

The right-hand side is $\ll (\log T)^{-100}$ by Rankin's trick and Lemma A.1,

$$\begin{aligned}\log^{1+\theta} T\cdot \sum_{\substack{\Omega(n)>\nu_\theta\\ p|n\Longrightarrow p\le X}} n^{-1} &\ll \log^{1+\theta} T\cdot e^{-\nu_\theta} \sum_{p|n\Longrightarrow p\le X} e^{\Omega(n)} n^{-1}\\ &\ll \log^{1+\theta} T\cdot e^{-\nu_\theta} \prod_{p|n\Longrightarrow p\le X}\Big(1+\sum_{k\ge 1}\frac{e^k}{p^k}\Big)\\ &\ll (\log T)^{-100}.\end{aligned}\tag{3.43}$$

The result follows by Markov's inequality. □

3.4. *Proofs of the lower bounds.* Consider, for $0\le j\le K-2$, the Dirichlet polynomials

$$P_j(h)=\operatorname{Re}\sum_{p\in J_j}\frac{1}{p^{\sigma_0+\mathrm{i}\tau+\mathrm{i}h}},\qquad J_j=\big(\exp\big((\log T)^{\frac{j}{K}}\big),\exp\big((\log T)^{\frac{j+1}{K}}\big)\big].\tag{3.44}$$

We choose a probabilistic notation for the increments P_j's seen as random variable, omitting the dependence on the random τ. We first prove a lower bound for the moments of Dirichlet polynomials.

PROPOSITION 3.10. *Let $\theta>-1$ and $\varepsilon>0$ be given. Then,*

$$\mathbb{P}\bigg(\int_{-\log^\theta T}^{\log^\theta T}\exp\Big(\beta\sum_{j=1}^{K-3}P_j(h)\Big)\,\mathrm{d}h>(\log T)^{f_\theta(\beta)-\varepsilon}\bigg)=1-\mathrm{o}(1).\tag{3.45}$$

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 $|h|\le\log^\theta T$ and $j\le K-3$, an application of (A.6) yields

$$s_j^2=\mathbb{E}\big[P_j(h)^2\big]=\frac{1}{2K}\log\log T+\mathcal{O}\big((\log T)^{-\frac{1}{2K}}\big),\tag{3.46}$$

since $\sigma_0-\frac12=(\log T)^{-1+3/(2K)}$. 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 (3.63) in the proof of Proposition 3.10.

PROOF OF PROPOSITION 3.10. 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

$$\mathcal{E}_\theta(\gamma):=\begin{cases}\theta-\dfrac{\gamma^2}{1+\theta}, & \text{if } \theta\le 0,\\ \theta-\gamma^2, & \text{if } \theta>0.\end{cases}\tag{3.47}$$

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

(3.48)
$$E = \bigcap_{j=1}^{M} \left\{ \mathrm{Leb}\left\{ |h| \le \log^\theta T : \exp\left(\sum_{\ell=1}^{K-3} P_\ell(h) \right) > (\log T)^{\gamma_{j-1}} \right\} \ge (\log T)^{\mathcal{E}_\theta(\gamma_{j-1}) - \varepsilon/2} \right\}$$
$$\bigcap \left\{ \max_{|h| \le \log^\theta T} \exp\left(\sum_{\ell=1}^{K-3} P_\ell(h) \right) \le (\log T)^{m(\theta)+\varepsilon} \right\}.$$

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

$$\frac{\log \int_{-\log^\theta T}^{\log^\theta T} \exp(\beta \sum_{j=1}^{K-3} P_j(h))\, \mathrm{d}h}{\log\log T} \ge \max_{1 \le j \le M} \{ \beta \gamma_{j-1} + \mathcal{E}_\theta(\gamma_{j-1}) \} - \varepsilon/2. \tag{3.49}$$

By the continuity of the function $\gamma \mapsto \beta\gamma + \mathcal{E}_\theta(\gamma)$, equation (3.49) implies that, on the event E and for M large enough with respect to ε and β,

$$\frac{\log \int_{-\log^\theta T}^{\log^\theta T} \exp(\beta \sum_{j=1}^{K-3} P_j(h))\, \mathrm{d}h}{\log\log T} > \max_{\gamma \in [\varepsilon, m(\theta)]} \{ \beta\gamma + \mathcal{E}_\theta(\gamma) \} - \varepsilon. \tag{3.50}$$

When $0 < \beta \le 2m(\theta)/(1 + (\theta \wedge 0))$, 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 (3.50) 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 (3.50) is equal to $(\beta m(\theta) - 1) - \varepsilon$. Thus, on the event E and for M large enough, the lower bound in (3.45) 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 $\sum_{j=1}^{K-3} P_j(h)$ in (2.48) (and the remark below it for $\theta < 0$), it is sufficient to prove that, for all $\eta > 0$ and all $0 < \gamma < m(\theta)$, the event

$$\left\{ \mathrm{Leb}\left\{ |h| \le \log^\theta T : \sum_{j=1}^{K-3} P_j(h) > \gamma \log\log T \right\} \ge (\log T)^{\mathcal{E}_\theta(\gamma) - \eta} \right\} \tag{3.51}$$

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

Consider

$$\mathcal{J}(\theta) = \begin{cases} 1, & \text{if } \theta \ge 0, \\ \lfloor K|\theta| \rfloor + 1, & \text{if } -1 < \theta < 0. \end{cases} \tag{3.52}$$

For $\theta < 0$, Corollary 2.12 ensures that the primes up to $\exp(\log^{|\theta|} T)$ only make a very small contribution, namely, the event

$$\left\{ \max_{|h| \le \log^\theta T} \left| \sum_{j=1}^{\mathcal{J}(\theta)-1} P_j(h) \right| \le \frac{\gamma}{(1+\theta)K} \log\log T \right\} \tag{3.53}$$

has probability $1 - \mathrm{o}(1)$. We consider the random variable

$$\mathcal{N} = \mathrm{Leb}\{ |h| \le \log^\theta T : P_j(h) > x_j, \text{ for } \mathcal{J}(\theta) \le j \le K - 3 \}, \tag{3.54}$$

where

$$x_j = \left(1 + \frac{100}{(1 + (\theta \wedge 0))K} \right) \cdot \frac{\gamma}{(1 + (\theta \wedge 0))K} \log\log T. \tag{3.55}$$

By summing the x_j's, it is not hard to check that the intersection of the events $\{\mathcal{N} \geq (\log T)^{\mathcal{E}_\theta(\gamma)-\eta}\}$ and the one in (3.53) is included in the event in (3.51). Therefore, the proof of the proposition is reduced to show

$$\mathbb{P}\big(\mathcal{N} \geq (\log T)^{\mathcal{E}_\theta(\gamma)-\eta}\big) = 1 - \mathrm{o}(1). \tag{3.56}$$

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

$$\begin{aligned} A(h) &= \big\{P_j(h) > x_j, \text{ for } \mathcal{J}(\theta) \leq j \leq K-3\big\}, \\ \theta &> -1, h, h' \in \big[-\log^\theta T, \log^\theta T\big]. \end{aligned} \tag{3.57}$$

The next two propositions are stated as Propositions 5.4 and 5.5 in Arguin et al. (2019). They are consequences of the Gaussian moments in Lemma A.3.

PROPOSITION 3.11 (One-point large deviation estimates). *Consider the event $A(h)$ in (3.57). For any choices of $\sqrt{\log\log T} \ll_K x_j \leq \log\log T$, where $1 \leq j \leq K-3$ and uniformly for $h, h' \in [-\log^\theta T, \log^\theta T]$, we have*

$$\mathbb{P}\big(A(h)\big) = \big(1+\mathrm{o}(1)\big) \prod_{j=\mathcal{J}(\theta)}^{K-3} \int_{x_j/s_j}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}}\,\mathrm{d}y \asymp \prod_{j=\mathcal{J}(\theta)}^{K-3} \frac{s_j}{x_j} \cdot e^{-x_j^2/(2s_j^2)}. \tag{3.58}$$

In the case of two points h, h', the primes are essentially correlated up to $\exp(|h-h'|^{-1})$ and quickly decorrelate afterward. 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 $\mathcal{J}(\theta)-1$. Therefore, the P_j's are essentially independent whenever $|h-h'| > (\log T)^{\theta-\frac{1}{2K}}$. We get the following.

PROPOSITION 3.12 (Two-point large deviation estimates). *Consider the event $A(h)$ in (3.57). For any choices of $0 < x_j \leq \log\log T$ and uniformly for $h, h' \in [-\log^\theta T, \log^\theta T]$ such that $|h-h'| > (\log T)^{-\frac{\mathcal{J}(\theta)}{K}+\frac{1}{2K}}$, we have*

$$\mathbb{P}\big(A(h) \cap A\big(h'\big)\big) = \big(1+\mathrm{o}(1)\big)\mathbb{P}\big(A(h)\big)\mathbb{P}\big(A\big(h'\big)\big). \tag{3.59}$$

Furthermore, let $0 \leq \ell \leq K-3$. Then, uniformly for $h, h' \in [-\log^\theta T, \log^\theta T]$ such that $|h-h'| \leq (\log T)^{-\ell/K}$, we have

$$\mathbb{P}\big(A(h) \cap A\big(h'\big)\big) \ll \exp\Bigg(-\sum_{j=\mathcal{J}(\theta)}^{\ell} \frac{x_j^2}{2s_j^2} - \sum_{j=(\ell+1)\vee\mathcal{J}(\theta)}^{K-3} \frac{x_j^2}{s_j^2}\Bigg). \tag{3.60}$$

Now, in order to prove (3.56), we start by finding a lower bound on $\mathbb{E}[\mathcal{N}]$. By (3.58) the x_j's in (3.55) and the s_j's in (3.46), we have

$$\mathbb{E}[\mathcal{N}] = \int_{-\log^\theta T}^{\log^\theta T} \mathbb{P}\big(A(h)\big)\,\mathrm{d}h \gg \log^\theta T \prod_{j=\mathcal{J}(\theta)}^{K-3} \frac{s_j}{x_j} \cdot e^{-x_j^2/(2s_j^2)} \gg (\log T)^{\mathcal{E}_\theta(\gamma)-\eta/3}, \tag{3.61}$$

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

$$\begin{aligned} \mathbb{P}\big(\mathcal{N} \geq (\log T)^{\mathcal{E}_\theta(\gamma)-\eta}\big) &\geq \mathbb{P}\big(\mathcal{N} \geq (\log T)^{-\eta/3}\mathbb{E}[\mathcal{N}]\big) \\ &\geq \big(1-(\log T)^{-\eta/3}\big)\frac{(\mathbb{E}[\mathcal{N}])^2}{\mathbb{E}[\mathcal{N}^2]}. \end{aligned} \tag{3.62}$$

It remains to show $\mathbb{E}[\mathcal{N}^2] = (1+o(1))(\mathbb{E}[\mathcal{N}])^2$. With $I = [-\log^\theta T, \log^\theta T]$, Fubini's theorem yields

$$\mathbb{E}[\mathcal{N}^2] = \int_{I\times I} \mathbb{P}\big(A(h)\cap A(h')\big)\,\mathrm{d}h\,\mathrm{d}h'. \tag{3.63}$$

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

$$\begin{aligned}
B &= \big\{(h,h') : |h-h'| > (\log T)^{-\frac{\mathcal{J}(\theta)}{K}+\frac{1}{2K}}\big\};\\
B_0 &= \big\{(h,h') : (\log T)^{-\frac{\mathcal{J}(\theta)}{K}} < |h-h'| \le (\log T)^{-\frac{\mathcal{J}(\theta)}{K}+\frac{1}{2K}}\big\};\\
B_\ell &= \big\{(h,h') : (\log T)^{-(\ell+1)/K} < |h-h'| \le (\log T)^{-\ell/K}\big\},\\
&\quad \text{for } \ell = \mathcal{J}(\theta),\dots,K-3;\\
B_{K-2} &= \big\{(h,h') : |h-h'| \le (\log T)^{-(K-2)/K}\big\}.
\end{aligned} \tag{3.64}$$

The dominant term will be the one on B. Note that $\mathrm{Leb}(B) = \mathrm{Leb}(I)^2(1+o(1))$. Hence, by (3.59) we have

$$\int_B \mathbb{P}\big(A(h)\cap A(h')\big)\,\mathrm{d}h\,\mathrm{d}h' = \big(1+o(1)\big)\big(\mathbb{E}[\mathcal{N}]\big)^2. \tag{3.65}$$

By (3.60) and the estimate (3.61), the integral on B_0 is

$$\ll (\log T)^{\theta-\frac{\mathcal{J}(\theta)}{K}+\frac{1}{2K}} \exp\Bigg(\sum_{j=\mathcal{J}(\theta)}^{K-3} -\frac{x_j^2}{s_j^2}\Bigg) \ll (\log T)^{-(\theta\vee 0)-\frac{1}{3K}}\big(\mathbb{E}[\mathcal{N}]\big)^2, \tag{3.66}$$

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 is, by (3.60) and the estimate (3.61),

$$\begin{aligned}
&\ll (\log T)^{\theta-\ell/K}\exp\Bigg(-\sum_{j=\mathcal{J}(\theta)}^{\ell}\frac{x_j^2}{2s_j^2} - \sum_{j=\ell+1}^{K-3}\frac{x_j^2}{s_j^2}\Bigg)\\
&= (\log T)^{-\theta-\ell/K}\exp\Bigg(\sum_{j=\mathcal{J}(\theta)}^{\ell}\frac{x_j^2}{2s_j^2}\Bigg)\cdot(\log T)^{2\theta}\exp\Bigg(-\sum_{j=\mathcal{J}(\theta)}^{K-3}\frac{x_j^2}{s_j^2}\Bigg)\\
&\ll (\log T)^{-\theta-\ell/K+(\ell/K+(\theta\wedge 0))\frac{\gamma^2}{(1+(\theta\wedge 0))^2}+\eta}\big(\mathbb{E}[\mathcal{N}]\big)^2,
\end{aligned} \tag{3.67}$$

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 (3.67) is $o((\mathbb{E}[\mathcal{N}])^2)$ if we fix $\eta > 0$ small enough with respect to θ and γ. Similarly, by (3.58) and the estimate (3.61), the integral on B_{K-2} is

$$\le \int_{B_{K-2}} \mathbb{P}\big(A(h)\big)\,\mathrm{d}h\,\mathrm{d}h' \ll (\log T)^{-1+2/K+\eta/3}\cdot\mathbb{E}[\mathcal{N}] = o\big(\big(\mathbb{E}[\mathcal{N}]\big)^2\big), \tag{3.68}$$

provided that η is small enough with respect to θ and γ and K is large enough with respect to θ, γ and η. This concludes the proof of Proposition 3.10. □

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 the probability in (3.2) is

$$\ge \mathbb{P}\Bigg(\int_{-\frac{1}{3}\log^\theta T}^{\frac{1}{3}\log^\theta T} \big|\zeta(\sigma_0+\mathrm{i}\tau+\mathrm{i}h)\big|^\beta\,\mathrm{d}h > (\log T)^{f_\theta(\beta)-\varepsilon}\Bigg) - o(1). \tag{3.69}$$

By Lemma 3.7 and Lemma 3.9, the above is

$$\geq \mathbb{P}\Big(\int_{-\frac{1}{3}\log^\theta T}^{\frac{1}{3}\log^\theta T} \exp\big(\beta \operatorname{Re} \widetilde{\mathcal{P}}_{1-K^{-1}}(\sigma_0+\mathrm{i}\tau+\mathrm{i}h)\big)\,\mathrm{d}h > (\log T)^{f_\theta(\beta)-2\varepsilon}\Big) - \mathrm{o}(1). \tag{3.70}$$

Now notice that the (double) sum for $k \geq 3$ in $\widetilde{\mathcal{P}}_{1-K^{-1}}(\sigma_0+\mathrm{i}\tau+\mathrm{i}h)$ is of order one (uniformly for $|h| \leq \log^\theta T$) and that the sum for $k=2$ is of negligible order,

$$\mathbb{P}\Big(\max_{|h|\leq \log^\theta T}\Big|\sum_{p\leq X}\frac{1}{2}p^{-2(\sigma_0+\mathrm{i}\tau+\mathrm{i}h)}\Big| > A\Big) \ll A^{-2\ell}\big(\log^{1+\theta} T\big)\cdot \ell!, \tag{3.71}$$

where we use the discretization from Proposition 2.7 and the moment estimates from Lemma A.4. Indeed, the right-hand side of (3.71) is $\mathrm{o}(1)$ with the choice $A=\sqrt{\nu_\theta}$ and $\ell = \lfloor (1+\theta)\log\log T\rfloor$. Hence, $\widetilde{\mathcal{P}}_{1-K^{-1}}$ can be replaced by $\mathcal{P}_{1-K^{-1}}$ with an error less than $\log^\varepsilon T$ with probability $1-\mathrm{o}(1)$, meaning that the right-hand side of (3.70) is

$$\geq \mathbb{P}\Big(\int_{-\frac{1}{3}\log^\theta T}^{\frac{1}{3}\log^\theta T} \exp\big(\beta \operatorname{Re} \mathcal{P}_{1-K^{-1}}(\sigma_0+\mathrm{i}\tau+\mathrm{i}h)\big)\,\mathrm{d}h > (\log T)^{f_\theta(\beta)-3\varepsilon}\Big) - \mathrm{o}(1). \tag{3.72}$$

By (2.48) we may discard the terms with $j=0$ and $j=K-2$ with a similar error. For K large enough with respect to ε, β and θ, the probability in (3.72) is, therefore,

$$\geq \mathbb{P}\Big(\int_{-\frac{1}{3}\log^\theta T}^{\frac{1}{3}\log^\theta T} \exp\Big(\beta\sum_{j=1}^{K-3} P_j(h)\Big)\,\mathrm{d}h > (\log T)^{f_\theta(\beta)-4\varepsilon}\Big) - \mathrm{o}(1). \tag{3.73}$$

Finally, the probability in (3.73) tends to 1, as $T\to\infty$ by Proposition 3.10. □

We now prove the lower bound in Theorem 1.2.

PROOF OF PROPOSITION 3.1. From (1.8) we have that $f_\theta(\beta) = \beta m(\theta) - 1$ 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 $1-\mathrm{o}(1)$) and for β large enough with respect to ε and θ, we have

$$\begin{aligned}\max_{|h|\leq\log^\theta T}\Big|\zeta\Big(\frac{1}{2}+\mathrm{i}\tau+\mathrm{i}h\Big)\Big| &\geq \Big(\frac{1}{2\log^\theta T}\int_{-\log^\theta T}^{\log^\theta T}\Big|\zeta\Big(\frac{1}{2}+\mathrm{i}\tau+\mathrm{i}h\Big)\Big|^\beta \mathrm{d}h\Big)^{1/\beta}\\ &\gg (\log T)^{m(\theta)-\frac{(1+\varepsilon+\theta)}{\beta}}\\ &\geq (\log T)^{m(\theta)-\varepsilon}.\end{aligned} \tag{3.74}$$

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*

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

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

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

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

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

LEMMA A.2 (Lemma 3.3 in Arguin et al. (2019)). *For any complex numbers $a(n)$, $b(n)$ and for $N \le T$, we have*

$$\begin{aligned}&\mathbb{E}\Big[\Big(\sum_{m\le N} a(m)m^{-i\tau}\Big)\Big(\sum_{n\le N} b(n)n^{i\tau}\Big)\Big]\\ &\quad=\sum_{n\le N} a(n)b(n)+\mathcal{O}\Big(\frac{N\log N}{T}\sum_{n\le N}\big(|a(n)|^2+|b(n)|^2\big)\Big).\end{aligned}\tag{A.3}$$

LEMMA A.3 (Lemma 3.4 in Arguin et al. (2019)). *Let $x \ge 2$ be a real number, and suppose that, for primes $p \le x$, $a(p)$ is a complex number with $|a(p)| \le 1$. Then, for any $k \in \mathbb{N}$,*

$$\mathbb{E}\Big[\Big(\frac{1}{2}\sum_{p\le x}\big(a(p)p^{-i\tau}+\overline{a(p)}p^{i\tau}\big)\Big)^k\Big]=\frac{\partial^k}{\partial z^k}\Big(\prod_{p\le x} I_0\big(|a(p)|z\big)\Big)\Big|_{z=0}+\mathcal{O}\Big(\frac{x^{2k}}{T}\Big),\tag{A.4}$$

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

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

$$\prod_{p\le x} I_0\big(|a(p)|z\big)=F(z)\cdot\exp\Big(\frac{z^2}{2}\cdot\frac{1}{2}\sum_{p\le x}p^{-2\sigma}\Big),\tag{A.5}$$

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

$$\mathbb{E}\Big[\Big(\sum_{p\le x}\mathrm{Re}\,p^{-\sigma-i\tau-ih}\Big)^{2k}\Big]=(1+\mathrm{o}(1))\frac{(2k)!}{2^k\cdot k!}\Big(\frac{1}{2}\sum_{p\le x}p^{-2\sigma}\Big)^k.\tag{A.6}$$

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

$$\mathbb{P}\Big(\sum_{p\le x}\mathrm{Re}\,p^{-\sigma-i\tau-ih}>V\Big)\ll\exp(-V^2/(2v^2)),\tag{A.7}$$

by picking the moment $k=\lfloor V^2/2v^2\rfloor$ with $v^2=\frac{1}{2}\sum_{p\le x}p^{-2\sigma}$, for V not too large.

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

LEMMA A.4 (Lemma 3 in Soundararajan (2009)). *Let T be large, and let $2 \le x \le T$. Let ℓ be a natural number such that $x^\ell \ll T/\log T$. For any complex numbers $a(p)$, we have*

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

Funding. L.-P. A. is supported in part by NSF Grant DMS-1513441 and by NSF CAREER DMS-1653602. F. O. is supported by postdoctoral fellowships from the NSERC (PDF) and the FRQNT (B3X). M. R. acknowledges support of a Sloan fellowship and NSF Grant DMS-1902063.

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 https://doi.org/10.1007/s00220-016-2740-6

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 https://doi.org/10.1214/16-AAP1201

ARGUIN, L. P., BOURGADE, P. and RADZIWIŁŁ, M. (2020). The Fyodorov-Hiary-Keating conjecture I. 1–49. Preprint. Available at arXiv:2007.00988.

ARGUIN, L. P., DUBACH, G. and HARTUNG, L. (2021). Maxima of a random model of the Riemann zeta function over intervals of varying length. 1–26. Preprint. Available at arXiv:2103.04817.

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. MR3906393 https://doi.org/10.1016/j.jmaa.2018.11.044

ARGUIN, L. P. and TAI, W. (2019). Is the Riemann zeta function in a short interval a 1-RSB spin glass? In *Sojourns in Probability Theory and Statistical Physics—I. Springer Proceedings in Mathematics & Statistics* 63–88. Springer, Singapore. https://doi.org/10.1007/978-981-15-0294-1.

ARGUIN, L.-P., BELIUS, D., BOURGADE, P., RADZIWIŁŁ, M. and SOUNDARARAJAN, K. (2019). Maximum of the Riemann zeta function on a short interval of the critical line. *Comm. Pure Appl. Math.* **72** 500–535. MR3911893 https://doi.org/10.1002/cpa.21791

BAILEY, E. C. and KEATING, J. P. (2019). On the moments of the moments of the characteristic polynomials of random unitary matrices. *Comm. Math. Phys.* **371** 689–726. MR4019917 https://doi.org/10.1007/s00220-019-03503-7

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 https://doi.org/10.1515/crelle-2014-0133

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 https://doi.org/10.1215/00127094-0000005X

BOURGADE, P. (2010). Mesoscopic fluctuations of the zeta zeros. *Probab. Theory Related Fields* **148** 479–500. MR2678896 https://doi.org/10.1007/s00440-009-0237-3

BOURGAIN, J. (2017). Decoupling, exponential sums and the Riemann zeta function. *J. Amer. Math. Soc.* **30** 205–224. MR3556291 https://doi.org/10.1090/jams/860

BOVIER, A. (2006). *Statistical Mechanics of Disordered Systems*: *A Mathematical Perspective*. *Cambridge Series in Statistical and Probabilistic Mathematics* **18**. Cambridge Univ. Press, Cambridge. MR2252929 https://doi.org/10.1017/CBO9780511616808

BOVIER, A. (2017). *Gaussian Processes on Trees*: *From Spin Glasses to Branching Brownian Motion*. *Cambridge Studies in Advanced Mathematics* **163**. Cambridge Univ. Press, Cambridge. MR3618123 https://doi.org/10.1017/9781316675779

BOVIER, A. and HARTUNG, L. (2020). From 1 to 6: A finer analysis of perturbed branching Brownian motion. *Comm. Pure Appl. Math.* **73** 1490–1525. MR4156608 https://doi.org/10.1002/cpa.21893

BOVIER, A. and KURKOVA, I. (2004). Derrida's generalized random energy models. II. Models with continuous hierarchies. *Ann. Inst. Henri Poincaré Probab. Stat.* **40** 481–495. MR2070335 https://doi.org/10.1016/j.anihpb.2003.09.003

CHANDEE, V. and SOUNDARARAJAN, K. (2011). Bounding $|\zeta(\frac{1}{2}+it)|$ on the Riemann hypothesis. *Bull. Lond. Math. Soc.* **43** 243–250. MR2781205 https://doi.org/10.1112/blms/bdq095

CHHAIBI, R., MADAULE, T. and NAJNUDEL, J. (2018). On the maximum of the CβE field. *Duke Math. J.* **167** 2243–2345. MR3848391 https://doi.org/10.1215/00127094-2018-0016

DERRIDA, B. (1981). Random-energy model: An exactly solvable model of disordered systems. *Phys. Rev. B* (3) **24** 2613–2626. MR0627810 https://doi.org/10.1103/physrevb.24.2613

DE LA BRETÈCHE, R. and TENENBAUM, G. (2019). Sommes de Gál et applications. *Proc. Lond. Math. Soc.* (3) **119** 104–134. MR3957832 https://doi.org/10.1112/plms.12224

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 https://doi.org/10.1515/CRELLE.2007.064

FYODOROV, Y. V. and BOUCHAUD, J.-P. (2008). Freezing and extreme-value statistics in a random energy model with logarithmically correlated potential. *J. Phys. A* **41** 372001, 12. MR2430565 https://doi.org/10.1088/1751-8113/41/37/372001

FYODOROV, Y. V., GNUTZMANN, S. and KEATING, J. P. (2018). Extreme values of CUE characteristic polynomials: A numerical study. *J. Phys. A* **51** 464001, 22. MR3876698 https://doi.org/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. https://doi.org/10.1103/PhysRevLett.108.170601

FYODOROV, Y. V. and KEATING, J. P. (2014). Freezing transitions and extreme values: Random matrix theory, and disordered landscapes. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **372** 20120503, 32. MR3151088 https://doi.org/10.1098/rsta.2012.0503

GABRIEL, R. M. (1927). Some Results Concerning the Integrals of Moduli of Regular Functions Along Certain Curves. *J. Lond. Math. Soc.* **2** 112–117. MR1574583 https://doi.org/10.1112/jlms/s1-2.2.112

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

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

HARPER, A. (2019). On the partition function of the Riemann zeta function, and the Fyodorov–Hiary–Keating conjecture. 1–40. Preprint. Available at arXiv:1906.05783.

HEAP, W., RADZIWIŁŁ, M. and SOUNDARARAJAN, K. (2019). Sharp upper bounds for fractional moments of the Riemann zeta function. *Q. J. Math.* **70** 1387–1396. MR4045106 https://doi.org/10.1093/qmathj/haz027

HEATH-BROWN, D. R. (1981). Fractional moments of the Riemann zeta function. *J. Lond. Math. Soc.* (2) **24** 65–78. MR0623671 https://doi.org/10.1112/jlms/s2-24.1.65

INGHAM, A. E. (1927). Mean-Value Theorems in the Theory of the Riemann Zeta-Function. *Proc. Lond. Math. Soc.* (2) **27** 273–300. MR1575391 https://doi.org/10.1112/plms/s2-27.1.273

IWANIEC, H. and SARNAK, P. (2000). Perspectives on the analytic theory of L-functions. 705–741. Geom. Funct. Anal., Special Volume, Part II, GAFA 2000 (Tel Aviv, 1999). MR1826269 https://doi.org/10.1007/978-3-0346-0425-3_6

KEATING, J. P. and SNAITH, N. C. (2000). Random matrix theory and $\zeta(1/2+it)$. *Comm. Math. Phys.* **214** 57–89. MR1794265 https://doi.org/10.1007/s002200000261

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 https://doi.org/10.1007/978-3-319-17674-1_3

KISTLER, N. and SCHMIDT, M. A. (2015). From Derrida's random energy model to branching random walks: From 1 to 3. *Electron. Commun. Probab.* **20** no. 47, 12. MR3358969 https://doi.org/10.1214/ecp.v20-4189

MICHEL, P. and VENKATESH, A. (2010). The subconvexity problem for GL_2. *Publ. Math. Inst. Hautes Études Sci.* **111** 171–271. MR2653249 https://doi.org/10.1007/s10240-010-0025-8

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 https://doi.org/10.1007/s00440-017-0812-y

NIKULA, M., SAKSMAN, E. and WEBB, C. (2020). Multiplicative chaos and the characteristic polynomial of the CUE: The L^1-phase. *Trans. Amer. Math. Soc.* **373** 3905–3965. MR4105514 https://doi.org/10.1090/tran/8020

OUIMET, F. (2018). Poisson-Dirichlet statistics for the extremes of a randomized Riemann zeta function. *Electron. Commun. Probab.* **23** Paper No. 46, 15. MR3841407 https://doi.org/10.1214/18-ECP154

PAQUETTE, E. and ZEITOUNI, O. (2018). The maximum of the CUE field. *Int. Math. Res. Not. IMRN* **2018** 5028–5119. MR3848227 https://doi.org/10.1093/imrn/rnx033

RADZIWIŁŁ, M. and SOUNDARARAJAN, K. (2013). Continuous lower bounds for moments of zeta and L-functions. *Mathematika* **59** 119–128. MR3028174 https://doi.org/10.1112/S0025579312001088

RADZIWIŁŁ, M. and SOUNDARARAJAN, K. (2017). Selberg's central limit theorem for $\log|\zeta(1/2+it)|$. *Enseign. Math.* **63** 1–19. MR3832861 https://doi.org/10.4171/LEM/63-1/2-1

REMY, G. (2020). The Fyodorov-Bouchaud formula and Liouville conformal field theory. *Duke Math. J.* **169** 177–211. MR4047550 https://doi.org/10.1215/00127094-2019-0045

RHODES, R. and VARGAS, V. (2014). Gaussian multiplicative chaos and applications: A review. *Probab. Surv.* **11** 315–392. MR3274356 https://doi.org/10.1214/13-PS218

SAKSMAN, E. and WEBB, C. (2016). Multiplicative chaos measures for a random model of the Riemann zeta function. 1–36. Preprint. Available at arXiv:1604.08378.

SAKSMAN, E. and WEBB, C. (2020). The Riemann zeta function and Gaussian multiplicative chaos: Statistics on the critical line. *Ann. Probab.* **48** 2680–2754. MR4164452 https://doi.org/10.1214/20-AOP1433

SELBERG, A. (1946). Contributions to the theory of the Riemann zeta-function. *Arch. Math. Naturvidensk.* **48** 89–155. MR0020594

SELBERG, A. (1992). Old and new conjectures and results about a class of Dirichlet series. In *Proceedings of the Amalfi Conference on Analytic Number Theory* (*Maiori*, 1989) 367–385. Univ. Salerno, Salerno. MR1220477

SOUNDARARAJAN, K. (2009). Moments of the Riemann zeta function. *Ann. of Math.* (2) **170** 981–993. MR2552116 https://doi.org/10.4007/annals.2009.170.981

TITCHMARSH, E. C. (1986). *The Theory of the Riemann Zeta-Function*, 2nd ed. The Clarendon Press, Oxford University Press, New York. Edited and with a preface by D. R. Heath-Brown. MR0882550

VENKATESH, A. (2010). Sparse equidistribution problems, period bounds and subconvexity. *Ann. of Math.* (2) **172** 989–1094. MR2680486 https://doi.org/10.4007/annals.2010.172.989

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 https://doi.org/10.1214/EJP.v20-4296