Random Matrix Theory and Refined Large Deviations of the Riemann Zeta Function

Matthew James Thompson

PhD

University of York
Mathematics

October 2022

Abstract

The analogy between the Riemann zeta function and the topic of random matrix theory was first established by Keating and Snaith [35]. Following this, many mathematicians have attempted to answer number-theoretic questions using random matrix theory. One such question is the maximum of the Riemann zeta function up to a height T along the critical line, and Farmer, Gonek Hughes established a conjecture for this maximum [21].

Our work builds upon these ideas and goes further in that we compute refined large deviations results for the characteristic polynomial of a random CβE matrix, generalising the CUE results of Farmer, Gonek and Hughes. We then apply these ideas to the Riemann zeta function, where we utilise a Hybrid Euler-Hadamard result of Gonek, Hughes and Keating [26] and compute refined large deviations results for each of these models separately.

Our results are consistent with the original works cited throughout this thesis, however there are some difference between the results of the two models, which we attempt to clarify here. Numerical results are presented to support (where possible) the results in this thesis.

Contents

List of Figures

Acknowledgements

While the completion of this thesis has been an enjoyable one, it hasn't always been easy. Working through Covid and the multitude of lockdowns was not exactly easy on my mental health and led to several hurdles along the way. With that, I would like to express my gratitude to a number of people.

To my friends at home; Curtis, Lois and many others, who I won't list here as I would almost certainly forget someone. Your calls and support during self-isolation and the many lockdowns we had were helpful for keeping me going at times when doing so felt rather difficult.

To my friends and colleagues in York that I've had the privilege of working alongside for these past four years; I've truly loved getting to know each and every one of you, and I hope we do not lose touch.

To Phoebe: your continued support and belief in me in the final stages of my PhD cannot be understated. I'm thankful and lucky to have you by my side as I complete my thesis, and I don't think there's enough words for me to express my gratitude.

Finally, to my supervisor Chris Hughes. I can't imagine I've always been the best PhD student to work with, yet I thank you for your continued support and patience with me as I plough on with my research. The insight and ideas you provide have been monumental in shaping this thesis, and I can't even begin to imagine what this would have looked like had I not had your support and guidance.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

1 Introduction

1.1 The Riemann Zeta Function

The Riemann zeta function[1] is defined for $\mathrm{Re}(s) > 1$ as

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \tag{1}$$

$$= \prod_{p \, \mathrm{prime}} \frac{1}{1 - p^{-s}}. \tag{2}$$

The result in (2) is known as the Euler-product expansion, and is important in that it gives an explicit connection between the zeta function and the prime numbers. By analytic continuation one can extend $\zeta(s)$ to a meromorphic function on the entire complex plane with the exception of a simple pole at $s = 1$; to achieve this Riemann made use of the following expression in his seminal paper of 1859 (a translated version of which is given in [20])[2]:

$$\zeta(s) = \frac{\Gamma(1-s)}{2\pi i} \int_C \frac{(-z)^{s-1}}{e^z - 1} \mathrm{d}z, \quad s \in \mathbb{C} \backslash \{1\}. \tag{3}$$

Here $\Gamma(s)$ denotes Euler's gamma function, while C denotes the complex contour starting at positive infinity and going along the positive real axis, before encircling once the origin at a radius $r < 2\pi$ (to avoid the poles of $e^z - 1$) in a positive direction and returning to positive infinity along the positive real axis.

By deforming this complex contour Riemann obtained the following functional equation:

$$\zeta(s) = 2^s \pi^{s-1} \sin\left(\frac{\pi s}{2}\right) \Gamma(1-s)\zeta(1-s). \tag{4}$$

This gives an explicit connection between points in the complex plane to the left and right of the line $s = \frac{1}{2} + it$: from the Euler-product expansion (2) it is immediate that $\zeta(s)$ is non-zero for $\mathrm{Re}(s) > 1$. By the functional equation (4), $\zeta(s)$ is also non-zero for $\mathrm{Re}(s) < 0$ (save for the trivial zeros, which we discuss below) since the zeros of $\sin(\frac{\pi s}{2})$ cancel the poles of $\Gamma(1-s)$ when $s = 2n,\ n \in \mathbb{N}$. In particular, if we consider $s = 2n + 1,\ n \in \mathbb{N}$ we have that

[1] Although the function is named after Riemann, it was first studied by Euler the century prior. However, Euler did not have access to complex analysis at the time; by virtue of the results Riemann was able to accomplish with access to these new mathematical tools, the function is now synonymous with him.

[2] Note that Riemann uses a slightly modified definition of the Euler-gamma function, $\Pi(s) = \int_0^\infty e^{-x} x^s \mathrm{d}x = \Gamma(s+1)$ which Edwards maintains throughout his publication for consistency. We will not adopt this convention, and instead use the now standard notation for the Euler-gamma function.

$$\zeta(2n+1) = 2^{2n+1}\pi^{2n}\sin\left(\pi n + \frac{\pi}{2}\right)\Gamma(-2n)\zeta(-2n).$$

The left-hand side is analytic and non-zero, therefore the same must also be true of the right-hand side. To ensure this is indeed the case we must have $\zeta(-2n) = 0$ for $n \in \mathbb{N}$ to cancel the poles of the gamma function; these are the so-called "trivial" zeros of the zeta function.

A more tasking problem concerns the location of the "non-trivial" zeros of the Riemann zeta function; from the above calculations $\zeta(s)$ is non-zero for $\mathrm{Re}(s) > 1$ and for $\mathrm{Re}(s) < 0$ by the functional equation, which leaves the strip $0 \leq \mathrm{Re}(s) \leq 1$. The statement that $\zeta(1+it) \neq 0$ for real t is equivalent to the prime number theorem which can be stated as follows: define $\pi(X) = \#\{p \text{ prime} : p \leq X\}$. Then

$$\pi(X) \sim \frac{X}{\log X}$$

in the limit as $X \to \infty$. This was proved independently by Hadamard and de la Valée Poussin in 1896 [28, 17]. From the functional equation, $\zeta(s)$ is also non-zero for $\mathrm{Re}(s) = 0$. Therefore if there are any non-trivial zeros they must lie in the strip $0 < \mathrm{Re}(s) < 1$, and this is commonly known as the "critical strip".

From (1), it is clear that $\zeta(s) = \overline{\zeta(\bar{s})}$, and so we find that the zeros come in conjugate pairs. Furthermore, the functional equation tells us that if s is a zero, then $1-s$ must also be a zero. Thus if we can find one non-trivial zero s, we immediately obtain three more: $\bar{s},\ 1-s,\ 1-\bar{s}$, and these are symmetrically distributed about the line $\mathrm{Re}(s) = 1/2$. The Riemann hypothesis conjectures that all of the non-trivial zeros lie on this line, which is commonly known as the "critical line"; in other words, $\zeta(\frac{1}{2}+it) = 0$ has non-trivial solutions only if $t \in \mathbb{R}$.

Initial progress on this conjecture was made by Hardy and Littlewood (1921) [30] who proved that there are infinitely many zeros on the critical line; work of Selberg [48] (1942) showed that a positive proportion of zeros lie on the critical line. Levinson [37] then proved (1974) that more than one third of the zeros lie on the critical line, which was followed up by Conrey (1989) [13] who proved that at least two fifths of the non-trivial zeros lie on the critical line, a result he subsequently improved (2011) to 41.05% in collaboration with Bui and Young [11]. Using Levinson's method, Roy, Robles and Zaharescu (2016) [47] improved this result to 41.0725%. However, predating this Shaoji Feng (2012) [22] utilised the methods of Selberg to show that 41.28% of the non-trivial zeros are on the critical line. More recently Pratt, Robles, Zaharescu, Zeindler (2018) [46] improved these results to show that more than five twelths of the non-trivial zeros lie on the critical line.

This problem cannot be understated, not only because of the connection to the prime numbers via the Euler product in (2), but also because the Riemann hypothesis is one of the millennium problems, a series of problems in mathematics for which the Clay Mathematics Institute offers a prize of $\$1,000,000$ for each correct solution. More recent efforts to tackle this conjecture have led to the startling connection between number theory and random matrix theory, and we explore this notion in more detail here.

1.2 Connection to Random Matrix Theory

Before we can formally discuss the connection between number theory and random matrix theory we must first discuss the notion of a random matrix, and what the notion of 'random' means in this context.

When we say 'random' here, we mean that the entries of our matrix are chosen with respect to some probability measure[3]. What this measure is will be revealed in 1.2.2, but we proceed in this section with a discussion on the circular β ensemble, which bears the most relevance to our work here. We follow this with a brief reference to other matrix ensembles before subsequently explaining how the worlds of number theory and random matrix come together.

1.2.1 Circular β Ensemble

We begin our look at the classical compact groups by first looking at the circular beta ensemble (or CβE). We introduce the physical construction, before considering explicit choices of β which correspond to the classical circular matrix ensembles.

Its construction is as follows: consider N identically charged particles on the unit circle $\mathbb{U}$ with logarithmic interaction potential and inverse temperature parameter β $(= 1/T)$. This gives rise to a probability distribution

$$P(\theta_1, \cdots, \theta_N) = \frac{1}{(2\pi)^N Z_{N,\beta}} \prod_{1\leq j<k\leq N} |e^{i\theta_j} - e^{i\theta_k}|^\beta. \tag{5}$$

Here $Z_{N,\beta}$ is a constant fixed under normalisation. Explicitly, it is given by (see [35])

$$Z_{N,\beta} = \frac{\Gamma(1+N\beta/2)}{\Gamma(1+\beta/2)^N}. \tag{6}$$

We will not discuss this point much further, but rewriting the probability distribution as

$$P(\theta_1, \cdots, \theta_N) = \frac{1}{(2\pi)^N Z_{N,\beta}} \exp(-\beta W), \tag{7}$$

where $W = -\sum_{1\leq j<k\leq N} \log|e^{i\theta_j - i\theta_k}|$, demonstrates a connection to the two-dimensional Coulomb gas model (see [19, 49] for more); here we see this logarithmic interaction

[3]This notion will be defined in Chapter 2

potential in the variable W, whereby this logarithmic function comes into play if one assumes a two-dimensional universe.

We relate this notion to random matrix theory by observing that the expression (7) demonstrates the repulsion between eigenvalues, as represented by the variable W; the eigenvalues do not like to be close together and as such repel one another logarithmically.

Choosing β explicitly, we arrive at some of the classical circular ensembles with probability measure (resp. probability density function) corresponding to that in (5). Namely:

- Circular Orthogonal Ensemble (COE) - Corresponding to $\beta = 1$, this is the set of Unitary matrices U invariant under Orthogonal transformations $U \to W^T U W$, where W is any $N \times N$ unitary matrix.

- Circular Unitary Ensemble (CUE) - Corresponding to $\beta = 2$, this is the set of all Unitary matrices with Haar measure[4].

- Circular Symplectic Ensemble (CSE) - Corresponding to $\beta = 4$, this is the set of Unitary matrices U invariant under Symplectic transformations $U \to W^R U W$, where W here is any $N \times N$ unitary quaternion matrix. R denotes the dual of W, $W^R = -ZW^T Z$, and we say that W is self-dual if $W^R = W$.

The results that we present here are for the general CβE (see Chapter 3). For the purposes of the Riemann zeta function we will be resticting our attention to the circular unitary ensemble ($\beta = 2$), for reasons which will later become clear. One can think of the circular unitary ensemble as the unitary group with an assigned Haar measure, and the unitary group is an example of a classical compact group, which we outline in the following Section.

1.2.2 Classical Compact Groups

If one considers any of the following classical matrix groups: the orthogonal group $O(N)$, the unitary group $U(N)$ or the symplectic group $Sp(2N)$, each of these forms a compact Lie group. One can then make use of the following theorem[5]:

Theorem 1.1. *If G is a compact topological group, there is a unique probability measure[6] μ which is invariant: that is, $\mu(V\mathcal{A}) = \mu(\mathcal{A})$ for every measurable subset $\mathcal{A} \subseteq G$ and fixed $V \in G$.*

This measure also satisfies $\mu(\mathcal{A}V) = \mu(\mathcal{A}) = \mu(\mathcal{A}^{-1})$.

[4]This is also the set of unitary matrices invariant under unitary transformations $U \to W^* U W$ (W is again unitary), but this coincides with the set of unitary matrices with Haar measure.

[5]The theorem is as stated in [38], with additional clarification where necessary.

[6]The notion of probability measure will formally be defined later.

Here $\mathcal{A}V := \{UV : U \in \mathcal{A}\}$ and similarly for $V\mathcal{A}$. This theorem is true in more generality; here we restrict our attention to the above compact groups, and for these groups the measure in question is the called a Haar measure. The property that $\mu(V\mathcal{A}) = \mu(\mathcal{A})$ is commonly referred to as left translation invariance, with $\mu(\mathcal{A}V) = \mu(\mathcal{A})$ the corresponding result for right translation invariance. For our purposes both left and right translation invariance hold, so we omit these prefixes and refer to this property simply as translation invariance[7].

We conclude this section with a broader picture of random matrix theory by including a discussion of the Gaussian ensembles.

1.2.3 Gaussian Ensembles

We've discussed above what can be thought of as the classical ensembles; we include a small aside here to discuss a collection of different ensembles know as the Gaussian ensembles. To motivate this discussion we should outline some of the physics which underpins these ideas. This will by no means be a thorough overview of the subject, and as a result some details are omitted (although one should consult [39] for more details on these topics).

Initial interest in the Gaussian ensembles comes from the field of nuclear physics; to be more concise, the excitation spectra of various nuclei is of great importance to nuclear physicists. The study of the averages of various energy levels of atomic nuclei are important in the study of nuclear reactions, and the interpretation in quantum mechanics is that the energy levels of some system can (supposedly) be described by the eigenvalues of a Hermitian operator, namely the Hamiltonian.

Since the only interest is the discrete part of the energy level schemes of various quantum systems, one can approximate the Hilbert space by another space having a finite (albeit large) number of dimensions. It was postulated by Wigner in the 1950's that the spacings between the lines in the spectrum of the nucleus of a heavy atom should resemble the spacings between eigenvalues of a random matrix [54, 55], and as such we can replace the Hermitian operator H with that of a Hermitian matrix. For the purposes of this section the matrices considered will be $N \times N$ matrices, where N is large but fixed.

In order to specify precisely the correlations among various elements of our matrices, we need a careful analysis of the consequences of time-reversal invariance.

Time-reversal Invariance The material from this section is primarily taken from a paper of Dyson [19] (although it is also included in Chapter 2 of [39]).

What do we mean when we say time-reversal invariance? A physical system is said to be time-reversal invariant if the underlying laws of the system are not

[7] It is important to note that it is not true in general that left and right translation invariance are equivalent when one looks at compact groups in more generality.

sensitive to the direction of time. A rather simplified example of this would be two billiard balls colliding with one another. If one were to record this event and play the footage both forwards and backwards, to an external viewer it might not be clear which direction the footage should be played. We do not discuss this idea in more detail as it is not relevant to later calculations, but the concept of time-reversal invariance will be key.

From physical considerations, it is required that the time-reversal operator T be anti-unitary, and as such this can be expressed in the form

$$T = KC,$$

where K is a fixed unitary operator and the operator C takes the complex conjugate of the expression which proceeds it. Therefore a state can be seen to transform under time-reversal as

$$\psi^R = T\psi = (KC)\psi = K(C\psi) = K\bar{\psi}.$$

Here ψ^R denotes transformation under time-reversal; $\bar{\psi}$ denotes the complex conjugate of ψ. The operation of time-reversal applied to a matrix is defined as

$$A^R = KA^TK^{-1},$$

where A^T denotes the transpose of A and K^{-1} denotes the inverse of K. A is called self-dual if $A^R = A$, and the physical system is said to be invariant under time-reversal if the Hamiltonian is self-dual: $H^R = H$.

It is clear that applying the time-reversal operator T twice should leave the physical system unchanged; therefore we have

$$T^2 = \alpha \cdot \mathbb{1}, \ |\alpha| = 1.$$

Here $\mathbb{1}$ denotes the unit operator. From our initial equation for T we have

$$T^2\psi = (KCKC)\psi = KC(K\bar{\psi}) = (K\bar{K})\psi = \alpha\psi.$$

Recalling that K is a unitary operator, we therefore have that

$$\bar{K}K^T = \mathbb{1}.$$

From these equations, we have that

$$K\psi = \alpha K^T\psi = \alpha(\alpha K^T)^T\psi = \alpha^2 K\psi.$$

Hence $\alpha^2 = 1$ and so $\alpha = \pm 1$. This then leads to two cases:

$$K\bar{K} = 1,$$
$$K\bar{K} = -1.$$

The first of these equations results in the unitary operator K being symmetric, corresponding to systems with even-spin. The latter of these equations results in K being antisymmetric and this corresponds to odd-spin systems.

We defer the additional details and formally define the joint probability density function for the Gaussian beta ensemble (GβE) for general beta, before considering the ensembles where beta is made explicit.

Theorem 1.2. *The joint probability density function for the eigenvalues of matrices from the Gaussian beta ensemble is given by*

$$P_{\beta N}(x_1, \cdots, x_N) = Z_{N,\beta} \exp\left(-\frac{\beta}{2}\sum_{j=1}^{N} x_j^2\right) \prod_{1\le j<k\le N} |x_j - x_k|^{\beta},$$

where the constant $Z_{N,\beta}$ is defined by

$$Z_{N,\beta} = (2\pi)^{(1/2)N} \beta^{-(1/2)N-(1/4)\beta N(N-1)} \left[\Gamma\left(1+\tfrac{1}{2}\beta\right)\right]^{-N} \prod_{j=1}^{N} \Gamma\left(1+\tfrac{1}{2}\beta j\right).$$

If we choose β explicitly (as with the CβE), we arrive at the Gaussian ensembles. Namely:

- Gaussian Orthogonal Ensemble (GOE) - Corresponding to $\beta = 1$, this is the set of real Hermitian matrices invariant under real Orthogonal transformations. Physically this corresponds to even-spin systems with time-reversal invariance.

- Gaussian Unitary Ensemble (GUE) - Corresponding to $\beta = 2$ this is the set of all Hermitian matrices invariant under unitary transformations. Physically this corresponds to a system without time-reversal invariance.

- Gaussian Symplectic Ensemble (GSE) - Corresponding to $\beta = 4$, this is the set of self-dual Hermitian matrices invariant under symplectic transformations. Physically this corresponds to odd-spin systems with time-reversal invariance but no rotational symmetry.

As with the circular beta ensemble we see in Theorem 1.2 a connection to the two-dimensional Coulomb gas model; writing the expression as

$$P_{\beta N}(x_1, \cdots, x_N) = Z_{N,\beta} \exp\left(-\frac{\beta}{2} \sum_{j=1}^{N} x_j^2 + \beta \sum_{1 \le j < k \le N} \log |x_j - x_k| \right),$$

we see that the logarithmic potential is present in the second-term. The initial term present in the joint PDF represents a Harmonic potential, and this comparison arises when we make the identification

$$\beta = (kT)^{-1}$$

where k denotes the Boltzmann constant.

We conclude our discussion of the Gaussian ensemble here, referring the reader to [39] should more information be desired.

1.2.4 Number Theory Correspondence

We started this Chapter with a brief but formal introduction to the classical compact groups in random matrix theory, but how does this material relate to the field of number theory and the study of the Riemann zeta function? Before delving into the results we present a historical overview which establishes the connection between the two fields.

In the 1970s, Montgomery studied the two-point correlations of the zeros of the zeta function [40]. Utilising the notation $\rho = \frac{1}{2} + i\gamma$ to denote a non-trivial zero of the zeta function, Montgomery's aim was to investigate the distribution of the differences $\gamma - \gamma'$ between the zeros (for which it would be thus desirable for Montgomery to know the Fourier transform of the distribution function of the numbers $\gamma - \gamma'$). To achieve this, he studied the following function:

$$F(\alpha) = F(\alpha, T) = \left(\frac{T}{2\pi} \log T \right)^{-1} \sum_{\substack{0<\gamma\le T \\ 0<\gamma'\le T}} T^{i\alpha(\gamma-\gamma')} w(\gamma - \gamma'), \tag{8}$$

where α and $T \ge 2$ are real, and $w(u) = 4/(4 + u^2)$ is a suitable weighting function. Note that the sum in (7) includes terms where $\gamma = \gamma'$.

Conditionally on the Riemann hypothesis, he asserts the following

Theorem 1.3. *Assume the Riemann Hypothesis. For real α, $T \ge 2$, let $F(\alpha)$ be defined by (8). Then $F(\alpha)$ is real, and $F(\alpha) = F(-\alpha)$. If $T > T_0(\epsilon)$ then $F(\alpha) \ge -\epsilon$ for all α. For fixed α satisfying $0 \le \alpha < 1$,*

$$F(\alpha) = (1 + o(1)) T^{-2\alpha} \log T + \alpha + o(1). \tag{9}$$

In addition, Montgomery makes use of a few heuristic arguments to suggest that

$$F(\alpha) = 1 + o(1) \tag{10}$$

for $\alpha \geq 1$, uniformly in bounded intervals. In order to investigate sums involving $\gamma - \gamma'$, Montgomery need only convolve $F(\alpha)$ (as defined in (8)) with an appropriate kernel $\hat{r}(\alpha)$[8]. Montgomery was able to show that

$$\sum_{\substack{0<\gamma\leq T \\ 0<\gamma'\leq T}} r\left((\gamma - \gamma')\frac{\log T}{2\pi}\right) w(\gamma - \gamma') = \left(\frac{T}{2\pi}\log T\right) \int_{-\infty}^{\infty} F(\alpha)\hat{r}(\alpha)\ \mathrm{d}\alpha \tag{11}$$

by Multiplying (8) by $\hat{r}(\alpha)$ and integrating both sides. Here $\hat{r}$ denotes the Fourier transform of r,

$$\hat{r}(\alpha) = \int_{-\infty}^{\infty} r(u)e(-\alpha u)\ \mathrm{d}u, \quad (e(\theta) = \exp(2\pi i\theta)).$$

With an appropriate use of (10) and a suitable choice of $F(\alpha)$, Montgomery conjectured that for fixed $\alpha < \beta$,

$$\sum_{\substack{0<\gamma\leq T \\ 0<\gamma'\leq T \\ 2\pi\alpha/\log T\leq\gamma-\gamma'\leq 2\pi\beta/\log T}} 1 \sim \left(\int_{\alpha}^{\beta}\left(1 - \left(\frac{\sin \pi u}{\pi u}\right)^2\right)\mathrm{d}u + \delta(\alpha,\beta)\right)\frac{T}{2\pi}\log T, \tag{12}$$

as T tends to infinity; $\delta(\alpha, \beta) = 1$ if $0 \in [\alpha, \beta]$ and 0 otherwise. Here γ, γ' denote the imaginary parts of two zeros on the critical line. The Dirac δ in the above equation is a natural occurrence, since $0 \in [\alpha, \beta]$ means the above sum includes terms where $\gamma = \gamma'$.

The assertion here is that the two-point correlation of the zeros of the zeta function is given by $1 - ((\sin \pi u)/\pi u)^2$. If we consider now the n-level correlation function of matrices in the CUE:

$$R_n(\theta_1, \ldots, \theta_n) = \frac{N!}{(N-n)!}\int_0^{2\pi}\cdots\int_0^{2\pi} P(\theta_1, \ldots, \theta_N)\ \mathrm{d}\theta_{n+1}\cdots\mathrm{d}\theta_N \tag{13}$$

$$= \det[K_N(\theta_i, \theta_j)]_{i,j=1,\ldots,n} \tag{14}$$

where we use Gram's result to rewrite this expression as a determinant. Here, $P(\theta_1, \ldots, \theta_N)$ is the joint PDF of matrices in the CUE as given in (5) with $\beta = 2$; $K_N(\theta_i, \theta_j)$ is defined to be

[8] By 'appropriate', we simply mean a kernel which, when convolved with $F(\alpha)$ produces the desired information.

$$K_N(\theta_i, \theta_j) = \frac{\sin[\frac{1}{2}N(\theta_i - \theta_j)]}{2\pi \sin[\frac{1}{2}(\theta_i - \theta_j)]}. \tag{15}$$

From this, one can compute the level densities to be $K_N(\theta, \theta) = N/2\pi$, and further the two-point correlation function is given by

$$\begin{aligned} R_2(\theta_1, \theta_2) &= \det \begin{bmatrix} \frac{N}{2\pi} & \frac{\sin\left[\frac{1}{2}N(\theta_1-\theta_2)\right]}{2\pi\sin\left[\frac{1}{2}(\theta_1-\theta_2)\right]} \\ \frac{\sin\left[\frac{1}{2}N(\theta_2-\theta_1)\right]}{2\pi\sin\left[\frac{1}{2}(\theta_2-\theta_1)\right]} & \frac{N}{2\pi} \end{bmatrix} \\ &= \left(\frac{N}{2\pi}\right)^2 - \left(\frac{\sin[\frac{1}{2}N(\theta_1 - \theta_2)]}{2\pi \sin[\frac{1}{2}(\theta_1 - \theta_2)]}\right)^2. \end{aligned} \tag{16}$$

Scaling θ_1, θ_2 by their level densities (so that $N\theta_1/2\pi = \xi$ and $N\theta_2/2\pi = \eta$) and taking the limit as the matrix size $N \to \infty$ gives

$$\lim_{N\to\infty} \left(\frac{2\pi}{N}\right)^2 R_2(\theta_1, \theta_2) = \left(1 - \left(\frac{\sin \pi r}{\pi r}\right)^2\right), \tag{17}$$

where $r = |\xi - \eta|$. From these calculations (see [39] for more detail) it can be seen that the two-point correlation function for matrices in the CUE matches that of the Riemann zeta function (12).

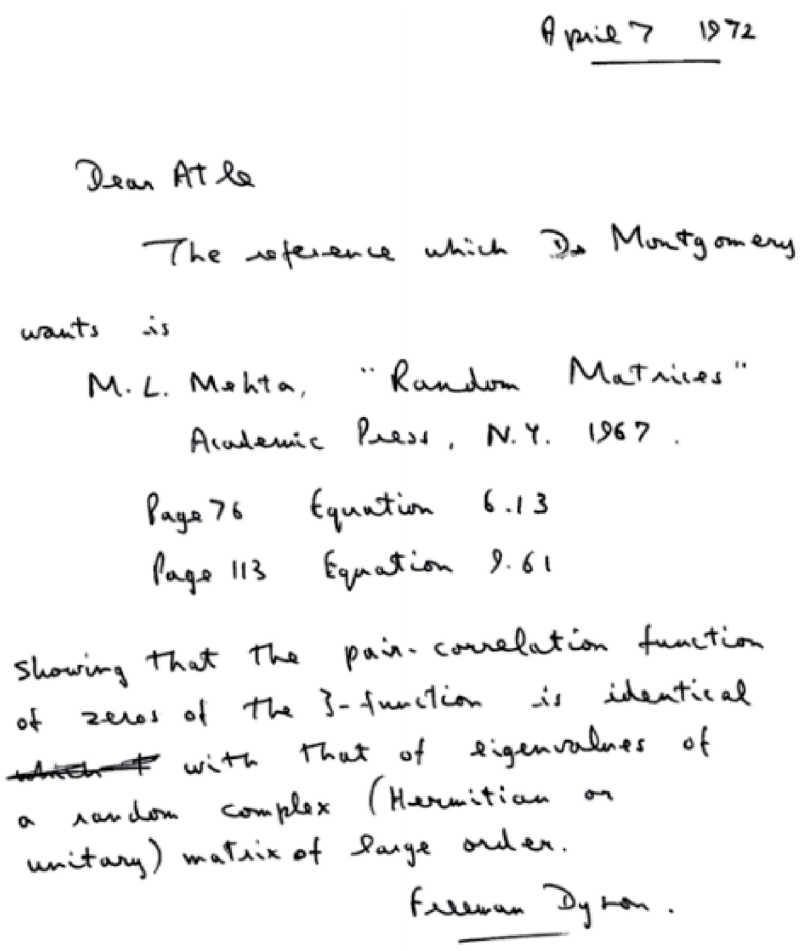
While at the Institute of Advanced Study in April 1972, Sarvadaman Chowla noticed that Freeman Dyson was present, and asked Montgomery if he had met him, to which Montgomery replied that he had not. Chowla insisted on introducing them to one another, and this went back and forth for a while before Montgomery eventually gave in.

After Chowla spoke to Dyson for a while, Dyson turned his attention to Montgomery and asked him “So what are you thinking about?” After initially being caught off guard by this question, Montgomery replied “I think the difference between the zeros of the zeta function are distributed with the density 1 minus the quantity $(\sin(\pi u)/\pi u)^2$”. Almost immediately Dyson responded with “That’s the pair correlation for the eigenvalues of a Hermitian matrix”[9]. Dyson later sent a letter to Selberg, which can be seen in Figure 1 below.

This observation opened up a new line of attack to some unsolved problems regarding the zeta function, with many hoping that by studying the eigenvalues of random complex hermitian or unitary matrices, one might better understand the behaviour of the zeros of the zeta function.

Today, the intricate connection between the seemingly disparate areas of random matrices and the Riemann zeta function has only continued to hold weight, with

[9]These details were given by Montgomery during a talk at the Institute of Advanced Study in June 2022, celebrating 50 years of Number Theory and Random Matrix Theory.



April 7 1972

Dear Atle

The reference which Dr Montgomery wants is

M. L. Mehta, "Random Matrices"
Academic Press, N.Y. 1967.

Page 76 Equation 6.13

Page 113 Equation 9.61

Showing that the pair-correlation function of zeros of the ζ-function is identical ~~which~~ with that of eigenvalues of a random complex (Hermitian or unitary) matrix of large order.

Freeman Dyson.

Figure 1: Letter from Dyson to Selberg referencing Mehta's work, which includes the discussed result for pair-correlation function of a random complex hermitian (or unitary) matrix of large order.

recent results in random matrix theory having number-theoretic analogues. In addition, numerical results of Odlyzko [44] as well as statistical results of Coram and Diaconis [16] further display the connection between these two areas of research. We explore this connection in more detail by looking at the moments of the zeta function.

1.3 Extreme Value Theory

1.3.1 Moments of $\zeta(1/2 + it)$

At an attempt to better understand the behaviour of the zeta function along the critical line, Keating and Snaith (2000) [35] studied the characteristic polynomial

$$\Lambda_U(\theta) = \det\left(I - Ue^{-i\theta}\right) \tag{18}$$

of a random CUE matrix[10] of size N. A key subject of this paper concerned the moments of the zeta function on the critical line:

$$I_k(T) = \frac{1}{T}\int_0^T \left|\zeta\left(\frac{1}{2}+it\right)\right|^{2k} \mathrm{d}t. \tag{19}$$

It was first proved in 1916 [29] by Hardy and Littlewood that

$$I_1(T) \sim \log T$$

as $T \to \infty$. It was subsequently proven by Ingham in 1926 [34] that

$$I_2(T) \sim \frac{1}{2\pi^2}(\log T)^4.$$

Following these results it was conjectured that

$$I_k(T) \sim c_k(\log T)^{k^2}$$

for some positive constant c_k. In a series of lectures given in the 1980's, Conrey and Ghosh expressed the moments in a more precise form; namely,

$$I_k(T) \sim \frac{a(k)g(k)}{\Gamma(k^2+1)}(\log T)^{k^2}, \tag{20}$$

where

$$a(k) = \prod_p \left\{\left(1-\frac{1}{p}\right)^{k^2}\left(\sum_{m=0}^{\infty}\left(\frac{\Gamma(k+m)}{m!\Gamma(k)}\right)^2 p^{-m}\right)\right\} \tag{21}$$

exists. The product here is taken over all prime numbers p, and when k is an integer, $g(k)$ is also an integer. From the above results we see that the values of c_k – and thus $g(k)$ – are known for $k=1$ and 2. Explicitly, we have $g(1)=1$ and $g(2)=2$. However, a plausible conjecture for the other values of $g(k)$, $k>2$ had yet to be established.

Some progress on this problem was made in the early 1990's; Conrey and Ghosh [14] conjectured that $g(3)=42$, and later Conrey and Gonek [15] conjectured that $g(4)=24,024$. It is worth remarking that the methods employed by these authors reproduce the previous values of $g(k)$.

Motivated by the connection to Random matrix theory, Keating and Snaith studied the moments of the characteristic polynomial

$$M_N(k) = \mathbb{E}\left[|\Lambda_U(\theta)|^k\right] = \prod_{j=1}^{N}\frac{\Gamma(j)\Gamma(j+k)}{(\Gamma(j+k/2))^2}, \quad \mathrm{Re}(k) > -1/2, \tag{22}$$

[10] In their original paper they used the notation $Z(U,\theta)$. We instead adopt the notation $\Lambda_U(\theta)$, although this is not to be confused with $\Lambda(\lambda)$ used to denote the Logarithmic MGF in Large Deviations, nor is it to be confused with $\Lambda^*(x)$ used to denote the convex-dual. We later make use of the von-Mangoldt function $\Lambda(n)$, and it should be clear which function is used from the context.

where this expectation is over the CUE of $N \times N$ unitary matrices. Considering even moments, scaling by the degree of the polynomial and taking the limit as the matrix size N tends to infinity, Keating and Snaith determined that

$$f_{\mathrm{CUE}}(k) = \lim_{N\to\infty} \frac{1}{N^{k^2}} \mathbb{E}\left[|\Lambda_U(\theta)|^{2k}\right] = \frac{G^2(k+1)}{G(2k+1)} \tag{23}$$

where G denotes the Barnes-G function [8] (some details about the Barnes G-function can be found in Appendix B). Therefore we have, after rewriting this expression,

$$\mathbb{E}\left[|\Lambda_U(\theta)|^{2k}\right] \sim \frac{G^2(k+1)}{G(2k+1)} N^{k^2}. \tag{24}$$

For the values $k = 1, 2, 3, 4$, Keating and Snaith observed that

$$\frac{G^2(k+1)}{G(2k+1)} = \frac{g(k)}{\Gamma(k^2+1)}.$$

Thus the leading conjecture regarding the moments of the zeta function, otherwise known as the Keating-Snaith conjecture, is the following:

Conjecture 1.1. *For fixed k with* $\mathrm{Re}(k) > -1/2$,

$$\frac{1}{T}\int_0^T \left|\zeta\left(\frac{1}{2}+it\right)\right|^{2k} \mathrm{d}t \sim a(k)\frac{G^2(k+1)}{G(2k+1)}(\log T)^{k^2}. \tag{25}$$

We see from this conjecture that if $N \sim \log T$ the even moments for the zeta function match the results for the characteristic polynomial of a random unitary matrix. We explore this relation between N and T further, but for now we study the value distributions of the characteristic polynomial.

Keating-Snaith studied the value distributions of the real and imaginary parts of the characteristic polynomial. In doing so, they obtained the following results:

$$Q_n(N) = \mathbb{E}\left[(\mathrm{Re}\log\Lambda_U(\theta))^n\right] = \frac{2^{n-1}-1}{2^{n-1}}\sum_{j=1}^{N}\psi^{(n-1)}(j)$$

and

$$R_n(N) = \mathbb{E}\left[(\mathrm{Im}\log\Lambda_U(\theta))^n\right] = \begin{cases} \frac{(-1)^{1+n/2}}{2^{n-1}}\sum_{j=1}^{N}\psi^{(n-1)}(j) & \text{if } n \text{ even,}\\ 0 & \text{if } n \text{ odd.}\end{cases}$$

Here $\psi^{(n-1)}(j)$ denotes the polygamma function of order $n-1$. In particular, formally evaluating $R_n(N)$ for $n = 2$ gives the equation

$$R_2(N) = Q_2(N) = \frac{1}{2}\log N + \frac{1}{2}(\gamma+1) + \frac{1}{24N^2} + O(N^{-4}).$$

Under the assumption of Montgomery's conjecture and the Riemann Hypothesis, Goldston [25] was able to prove that

$$\frac{1}{T}\int_0^T \left(\operatorname{Im}\log\zeta\left(\frac{1}{2}+it\right)\right)^2 \mathrm{d}t = \frac{1}{2}\log\log\frac{T}{2\pi}+\frac{1}{2}(\gamma+1)+\sum_{m=2}^{\infty}\sum_{p}\frac{(1-m)}{m^2}\frac{1}{p^m}+o(1).$$

Making the identication $N=\log\frac{T}{2\pi}$ (which comes from equating the mean density of the eigenangles, $N/2\pi$ with the mean density of the zeros of the zeta function at a height T up the critical line, $\frac{1}{2\pi}\log\frac{T}{2\pi}$), we find that the first two terms are in agreement. The third term demonstrates the prime contribution in the zeta case, and this led Keating and Snaith to speculate the possible splitting of the zeta function into a product of two terms; the first term a product over primes, the second term a product over non-trivial zeros[11]. We return to this notion a little later on.

With the identification $N=\log\frac{T}{2\pi}$, their results about the value distributions of the real and imaginary parts of the characteristic polynomial:

$$\lim_{N\to\infty}\frac{1}{N}\mathbb{P}\left(\frac{\log\Lambda_U(\theta)}{\sqrt{\frac{1}{2}\log N}}\in E\right)=\frac{1}{2\pi}\int\int_E e^{-(x^2+y^2)/2}\,\mathrm{d}x\,\mathrm{d}y$$

align with the following result of Selberg for the logarithm of the zeta function at values on the critical line [44, 53]: for any rectangle $E\in\mathbb{R}^2$

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}&\left|\left\{t : T\le t\le 2T,\frac{\log|\zeta(\frac{1}{2}+it)|}{\sqrt{(1/2)\log\log T}}\in E\right\}\right|\\ &=\frac{1}{2\pi}\int\int_E e^{-(x^2+y^2)/2}\,\mathrm{d}x\,\mathrm{d}y.\end{aligned}$$

In other words, the value distributions of the real and imaginary parts of the zeta function (after scaling by $\sqrt{(1/2)\log\log T}$) converge independently to a Gaussian with mean 0 and variance 1. We return to Selberg's theorem when we later discuss the maximum of the zeta function along the critical line.

By this point we've looked at results concerning the value distribution of the real and imaginary parts of the characteristic polynomial, as well as Selberg's theorem which details the value distribution of the real and imaginary parts of the zeta function, but what about the distribution of the zeros of the zeta function?

If one assumes the Riemann hypothesis, we know everything about the horizontal distribution of the zeros as the hypothesis places all of the non-trivial zeros on the critical line. The vertical distribution of the zeros still remains an open problem; while computation has allowed mathematicians to place the first 100 trillion or so

[11] This was left as a brief comment in their seminal paper on the zeta function and random matrix theory.

zeros on the critical line, the exact positions of the non-trivial zeros are yet to be determined.

Another discussion point is the following: we've seen Selberg's result above which gives results for the value distribution of the real and imaginary parts of $\zeta(\frac{1}{2}+it)$, but what if one were to look at the zeta function over a different range, say $[0,T]$ or a smaller range such as $[T,T+1]$? It is questions like these that motivate some of the main results in this thesis.

Before we can proceed by looking at these results, we continue to delve into the history; the results we list in the sections that follow are key results to keep in mind when looking at the results of Chapter 2.

1.3.2 Fyodorov-Hiary-Keating

Continuing with our previous discussion, Fyodorov, Hiary, Keating [23, 24] continued to explore the connections between random matrix theory and the zeta function. As such, the authors studied the value distribution of (the maximum of) the modulus of the characteristic polynomial of a random unitary matrix; these results were then applied to the topic of the zeta function, leading to a conjectured result for large values taken by the zeta function over stretches of the critical line. In their papers they sought to connect these two research areas to a third: the statistical mechanics of disordered landscapes.

The focus taken by the authors here is the following: they were concerned with the maximum values of the characteristic polynomials of individual matrices (as opposed to a large number of matrices; we discuss this approach in more detail in later chapters), and they obtain the full value distribution of the maxima in the limit as the matrix size N tends to infinity. The outcome of this method was a model for the distribution of the maximum values of $|\zeta(\frac{1}{2}+it+ih)|$ over $0 \le h \le 1$ where $T \le t \le 2T$. This is in the limit $T \to \infty$, where they make the usual identification between N and T, $N = \log \frac{T}{2\pi}$.

In particular, they conjecture the following:

Conjecture 1.2 (Fyodorov-Hiary-Keating)**.** *There exists a cumulative distribution function F such that, for any y, as $T \to \infty$,*

$$\frac{1}{T} meas \left\{ T \le t \le 2T : \max_{0 \le h \le 1} \left| \zeta\left(\frac{1}{2}+it+ih\right)\right| \le e^y \frac{\log T}{(\log\log T)^{3/4}} \right\} \sim F(y). \quad (26)$$

Moreover, as $y \to \infty$ the right-tail decay is $1-F(y) \sim Cye^{-2y}$ for some positive constant C.

We see here a model which conjectures the behaviour of the zeta function up to sub-leading order, as well as give some indication of the decay rate for the right-tail. This is not the only model available for the zeta function however, and we examine now another model which generates an interesting discussion regarding the behaviour of the zeta function along the critical line.

1.3.3 The $1/4$ Model

The model for the zeta function introduced here assumes that the values of the zeta function on the critical line are correlated at short range, and we see that this leads to a noticeable difference between the results of this model and the results from the Fyodorov-Hiary-Keating conjecture. For clarity we include this to illustrate the the differences in comparison with the model introduced by Fyodorov, Hiary and Keating. We will not be making use of this model going forward.

Up to height T on the critical line, there are asymptotically $\frac{T}{2\pi}\log T$ zeros. Therefore the average spacing between zeros is $T/(\frac{T}{2\pi}\log T) = 2\pi/\log T$ for $t \approx T$. We therefore expect that the behaviour of $\zeta(\frac{1}{2}+it)$ changes completely if we shift by $2\pi/\log T$. We also expect that $\zeta(\frac{1}{2}+it)$ is more or less constant if we instead shift by $2\pi\epsilon/\log T$ for small $\epsilon > 0$.

From Selberg's Theorem $\log|\zeta(\frac{1}{2}+it)|$ has a Gaussian distribution. It is thus reasonable to expect in this model that the joint distribution of

$$\left\{\log\left|\zeta\left(\frac{1}{2}+it+\frac{2\pi ik}{\log T}\right)\right|\right\}_{|k|<\frac{1}{2\pi}\log T}$$

is that of $\frac{1}{\pi}\log T$ independent Gaussians with mean 0 and variance $\frac{1}{2}\log\log T$.

Were this a valid model, then for typical $t \in [T, 2T]$

$$\max_{|h|\leq 1}\log|\zeta(\tfrac{1}{2}+it+ih)|$$

would be accurately modelled by

$$\max_i G_i$$

where $i \in \{1, \cdots, \left[\frac{1}{\pi}\log T\right]\}$ and the G_i are independent Gaussians with mean 0 and variance $\frac{1}{2}\log\log T$.

This now becomes a simple exercise in probability. The steps performed here are the same as those we utilise throughout this thesis; for this reason we include these calculations.

Since $M(T)$ is the maximum we seek, one way to interpret this is to say we want the largest $M(T)$ for which

$$\mathbb{P}\left(\forall i \leq \frac{1}{\pi}\log T : G_i < M(T)\right) = o(1),$$

in other words we are looking for the largest $M(T)$ for which there exists a G_i with $i \in \{1, \cdots, \left[\frac{1}{\pi}\log T\right]\}$ such that $G_i > M(T)$ with probability $1-o(1)$. The G_i are independent, so this amounts to finding asymptotically the largest $M(T)$ such that

$$\mathbb{P}(G < M(T))^{[\log T/\pi]} = o(1),$$

where G is a Gaussian with mean 0 and variance $\frac{1}{2}\log\log T$.

Rewriting this amounts to solving

$$\begin{aligned}\mathbb{P}(G < M(T))^{[\log T/\pi]} &= (1 - \mathbb{P}(G > M(T)))^{[\log T/\pi]} \\ &= \left(1 - \frac{1}{2}\mathrm{erfc}\left(\frac{M(T)}{\sqrt{\log\log T}}\right)\right)^{[\log T/\pi]} = o(1).\end{aligned}$$

Using the asymptotic expansion of the complementary error function (assuming $M(T)$ is of order larger than $\sqrt{\log\log T}$) and taking logarithms, the above expression gives

$$\begin{aligned}\left[\frac{\log T}{\pi}\right]\log\left(1 - \frac{1}{2}\mathrm{erfc}\left(\frac{M(T)}{\sqrt{\log\log T}}\right)\right) &\sim -\frac{\log T}{\pi} \times \frac{1}{2}\mathrm{erfc}\left(\frac{M(T)}{\sqrt{\log\log T}}\right) \\ &= -\frac{\log T}{\pi} \times \frac{\sqrt{\log\log T}}{2\sqrt{\pi}M(T)}\exp\left(-\frac{M(T)^2}{\log\log T}\right) + \cdots \\ &= -\frac{\log T\sqrt{\log\log T}}{2\pi^{3/2}M(T)} \times \exp\left(-\frac{M(T)^2}{\log\log T}\right) + \cdots .\end{aligned}$$

It is now a case of balancing this expression (at leading order), which simply involves solving for $M(T)$ the expression

$$\exp\left(\frac{M(T)^2}{\log\log T}\right) = \frac{\log T\sqrt{\log\log T}}{2\pi^{3/2}M(T)}.$$

Taking logarithms and solving on the left for $M(T)$ gives

$$\begin{aligned}M(T) = \Bigg((\log\log T)^2 &+ \frac{1}{2}\log\log T\log\log\log T - \log\log T\log M(T) \\ &- \log 2\pi^{3/2}\log\log T + \cdots\Bigg)^{1/2}.\end{aligned}$$

If we then take the logarithm of this expression, we obtain

$$\log M(T) = \log\log\log T + \frac{1}{4}\frac{\log\log\log T}{\log\log T} + \cdots .$$

Plugging this into the expression for $M(T)$ and simplifying the expression gives the solution

$$
\begin{aligned}
M(T) &= \left((\log\log T)^2 - \frac{1}{2}\log\log T \log\log\log T - \log 2\pi^{3/2}\log\log T + \cdots \right)^{1/2} \\
&= \log\log T \left(1 - \frac{\log\log\log T}{2\log\log T} - \frac{\log 2\pi^{3/2}}{\log\log T} + \cdots \right)^{1/2} \\
&= \log\log T \left[1 - \frac{\log\log\log T}{4\log\log T} - \frac{\log 2\pi^{3/2}}{2\log\log T} + \cdots \right].
\end{aligned}
$$

This suggests that $M(T) = \log\log T - \frac{1}{4}\log\log\log T + O(1)$. However we need to be careful here, since we have taken logarithms in our equation for the probability that G is smaller than $M(T)$, and so the right-hand side is now $\log o(1)$ which approaches negative infinity as T gets large.

If we instead take $M(T) = \log\log T - \frac{1}{4}\log\log\log T + \psi(T)$ where $\psi(T)$ tends to infinity at a slower rate than $\log\log\log T$ then we have, after plugging this into the above expression for the probability,

$$
\begin{aligned}
&-\frac{\log T\sqrt{\log\log T}}{2\pi^{3/2}\log\log T}\left[1 + \frac{\log\log\log T}{4\log\log T} + \frac{\psi(T)}{\log\log T} + \cdots\right] \times \exp(-\log\log T) \\
&\quad \times \exp(\tfrac{1}{2}\log\log\log T) \times \exp(2\psi(T)) \times \cdots
\end{aligned}
$$

and this expression cancels to give (at leading order) the result

$$
-\frac{\exp(2\psi(T))}{2\pi^{3/2}}\left[1 + \frac{\log\log\log T}{4\log\log T} + \frac{\psi(T)}{\log\log T} + \cdots\right].
$$

This implies that the left-hand side of our expression for G (after taking logarithms) also tends to negative infinity. Thus our solution for $M(T)$ can be written in the above form with $\psi(T)$, or alternatively in the form $M(T) = \log\log T - \frac{1}{4}\log\log\log T + o(\log\log\log T)$.

Writing these results in the context of the zeta function we have the following: for t chosen uniformly in $[T, 2T]$,

$$
\max_{|h|\leq 1} \log\left|\zeta\left(\frac{1}{2} + it + ih\right)\right| = \log\log T - \tfrac{1}{4}\log\log\log T + o(\log\log\log T)
$$

as $T \to \infty$. Observe here that in this model for the zeta function, the constant on the subleading term is $-1/4$ as opposed to the $-3/4$ factor predicted by Fyodorov, Hiary, Keating. We now have the question of which model is the correct one, and whether there might be some way to transition between the two models, the first model considering no short-range correlations between the zeros while the second model considers short-range correlations between the values on the critical line.

1.3.4 $3/4$ versus $1/4$

We have two competing models for the maximum value of the zeta function in a short interval on the critical line, and we have a connection to random matrix theory first established by Montgomery and Dyson. It is natural then to consider the random matrix analogue of this problem to see whether this model agrees with the $3/4$ case or the $1/4$ case, and in fact this is what Fyodorov, Hiary, and Keating did.

Precisely, the connection they develop between the fields of random matrix theory and number theory to the statistical eigenvalues of disordered landscapes suggests that the maximum value of the modulus of the characteristic polynomial $\Lambda_U(\theta)$ of a random unitary matrix (a matrix in the CUE) for θ in the interval $[0, 2\pi)$ is given by

$$\max_{\theta\in[0,2\pi)} \log|\Lambda_U(\theta)| = \log N - \frac{3}{4}\log\log N + o(1) \tag{27}$$

as the matrix size $N \to \infty$ (recall from earlier that the results are rendered in agreement with those of the zeta function once the identification $N = \log\frac{T}{2\pi}$ is made). This suggests that the model as conjectured by Fyodorov, Hiary, Keating is likely the correct one, although a better indication of this comes from studying the literature. We outline the key results and contributions below.

Progress towards the conjectured random matrix analogue was first made by Arguin, Belius and Bourgade [1] who established the leading order behaviour

$$\lim_{N\to\infty} \frac{\max_{\theta\in[0,2\pi)} \log|\Lambda_U(\theta)|}{\log N} = 1 \text{ in probability.}$$

The second order,

$$\lim_{N\to\infty} \frac{\max_{\theta\in[0,2\pi)} \log|\Lambda_U(\theta)| - \log N + \frac{3}{4}\log\log N}{\log\log N} = 0 \text{ in probability,}$$

was first proved by Paquette and Zeitouni [45], although they were unable to establish tightness of the o-term. This was later established by means of the following:

$$\max_{\theta\in[0,2\pi)} \log|\Lambda_U(\theta)| = \log N - \frac{3}{4}\log\log N + O(\psi(N))$$

with probability one, as $N \to \infty$ for any $\psi(N)$ tending to infinity arbitrarily slowly. This result was due to Chhaibi, Madaule and Najnudel [12]. With these results firmly established, it is increasingly likely that the result of Fyodorov-Hiary-Keating is the correct result. This is the best we have on the random matrix side, so we now return to the Riemann zeta function to explore the progress made towards the conjecture of Fyodoroy-Hiary-Keating.

The first result was computed in 2016 by Arguin, Belius, Bourgade, Radziwiłł and Soundararajan [2], who established the leading order behaviour

$$\max_{|h|\leq 1} \log \left|\zeta\left(\frac{1}{2}+it+ih\right)\right| \sim \log\log T$$

on a set $t \in [T, 2T]$ of measure $T + o(T)$. In the same year and independently of the above authors, Najnudel also obtained the leading order behaviour, conditionally on the Riemann hypothesis [42].

The upper bound with the sub-leading behaviour was established by Harper in 2019 [32]:

$$\max_{|h|\leq 1} \log \left|\zeta\left(\frac{1}{2}+it+ih\right)\right| \leq \log\log T - \frac{3}{4}\log\log\log T + \frac{3}{2}\log\log\log\log T + g(T)$$

for a set $T \leq t \leq 2T$ of measure $T + o(T)$, where $g(T)$ is any real function tending to infinity with T. While this matches the result of Fyodorov, Hiary, Keating at both leading and sub-leading order, this result is not sharp due to the sub-subleading term.

Recently, Arguin, Bourgade and Radziwiłł [4] were able to express the upper bound of the conjecture in a strong form.

Theorem 1.4 (Arguin-Bourgade-Radziwiłł). *There exists a constant $C > 0$ such that uniformly in $T > 3$ and $y \geq 1$,*

$$\frac{1}{T}\operatorname*{meas}_{T\leq t\leq 2T}\left\{\max_{|h|\leq 1}\left|\zeta\left(\frac{1}{2}+it+ih\right)\right| > \frac{\log T}{(\log\log T)^{3/4}}e^{y}\right\} \leq Cye^{-2y}.$$

This result is strong in that it reproduces the previous results, but goes further in that it gives the limiting behaviour in terms of y which agrees with the conjectured growth of Fyodorov, Hiary, Keating (see Conjecture 1.2).

The result in this theorem is expected to be sharp in the range $y = O(\sqrt{\log\log T})$. For larger y (in the range $y \in [1, \log\log T]$), it is expected that the sharp decay rate is instead

$$\ll ye^{-2y}\exp\left(-\frac{y^2}{\log\log T}\right).$$

This limiting behaviour is interesting in that it suggests the older conjecture as seen in [53] is perhaps closer to the truth, since the result of Arguin, Bourgade and Radziwiłł would suggest that when one conditions on large values as opposed to typical values the growth appears to be exponential.

We see later that this contrasts the results of Farmer, Gonek and Hughes in Section 1.4 since their results suggest that the growth for large values is Gaussian rather than exponential, and it is these results which play a key role in studying the transition between the 1/4 and 3/4 regimes.

1.3.5 Phase Transition

We've seen the arguments made for 1/4 and 3/4, with results strongly suggesting that the 3/4 argument is the correct one.

To add further weight to the 3/4 case we look at the following: in a paper of Harper [31] which builds upon ideas of Soundararajan [52], he includes the following proposition, giving a suitable model (conditional on the Riemann hypothesis) for large height T along the critical line:

Proposition 1.1 (Harper). *Assume the Riemann hypothesis, and let T be large. Then for any $T \leq t \leq 2T$ we have*

$$\log|\zeta(1/2+it)| \leq \Re(\sum_{p\leq T} \frac{1}{p^{1/2+1/\log T+it}} \frac{\log(T/p)}{\log T} + \sum_{p^2\leq T} \frac{(1/2)}{p^{1+2/\log T+2it}} \frac{\log(T^2/p)}{\log T}) + O(1).$$

Moreover, there exists a set $\mathcal{H} \subset \{T, T+2\pi\}$, of measure at least 1.99π such that

$$\log\left|\zeta\left(\frac{1}{2}+it\right)\right| = \mathrm{Re}\left(\sum_{p\leq T} \frac{1}{p^{1/2+it}} \frac{\log(T/p)}{\log T}\right) + O(1) \quad \forall t \in \mathcal{H}.$$

The use of $\mathcal{H}$ corresponds to the initial study of $\max_{T\leq t\leq T+2\pi} |\zeta(1/2+it)|$ by Fyodorov and Keating[12].

With this initial model for zeta, the authors in [3] utilise the argument that the finite-dimensional distribution of the process (p^{-it}, p primes), t sampled uniformly in $[0, T]$, converges as $T \to \infty$ to a sequence of independent random variables uniformly distributed on the unit circle; this motivates the use of the following model to study the large values of $\log|\zeta|$ in some short interval I:

$$W_T(h) = \sum_{p\leq T} \frac{\mathrm{Re}(U_p p^{-ih})}{p^{1/2}}, \ h \in I.$$

Note that we are looking to study the large values of $\log|\zeta|$ here which is real, and so it is necessary that the real part is looked at in the expression $W_T(h)$. Looking at the maximum of $W_T(h)$ Arguin, Belius and Harper were able to verify the leading and subleading terms in the Fyodorov-Hiary-Keating conjecture. What is perhaps interesting about this approach is that the proof of their main result uses an approximate tree structure, which is both present in the model and in the zeta function itself. The idea for this implementation came from the observation that the leading and subleading orders of the maximum in the conjecture of Fyodorov, Hiary, Keating correspond exactly to those of the maximum of a branching random walk (a collection of correlated random walks indexed by the leaves of a tree).

[12]Note that when we referred to the Fyodorov-Hiary-Keating conjecture earlier, we used a different albeit equivalent formulation than what is mentioned here.

These ideas and approach were first developed by Bramson [10] in his seminal work on the maximum of branching Brownian motion, and we will mention these ideas again shortly.

Returning to the 3/4 result, the techniques developed in establishing these results can also be used to investigate large values in intervals whose lengths vary with T. In [6], it was conjectured that for intervals of size $(\log T)^{\theta}$ (where $\theta > 0$ is fixed) the maximum is

$$\max_{|h|\leq(\log T)^{\theta}} \left|\zeta\left(\frac{1}{2}+it+ih\right)\right| = \frac{(\log T)^{\sqrt{1+\theta}}}{(\log\log T)^{\frac{1}{4\sqrt{1+\theta}}}} e^{\mathcal{M}_{\theta}(T)},$$

where $(\mathcal{M}_{\theta}(T),\ T>1)$ is a tight sequence of random variables.

The authors prove the leading order $(\log T)^{\sqrt{1+\theta}}$ behaviour. The interesting thing to note here is the exponent $\frac{1}{4\sqrt{1+\theta}}$ of the $\log\log T$ term here, as it would seem to suggest that as we let $\theta \to 0$ there is a jump-discontinuity in the exponent, as it approaches 1/4 and not 3/4 as previous results would appear to suggest.

A recent publication of Arguin, Dubach and Hartung [5] sheds some light on this issue, as well as demonstrating a smooth transition from the 1/4 regime to the 3/4 regime.

In their publication, rather than consider the model $W_T(h)$ of Arguin, Belius and Harper, they studied a Gaussian variant of this model; here the $U_p's$ are replaced by standard complex Gaussian:

$$X_T(h) = \sum_{p\leq T} \frac{\mathrm{Re}(G_p p^{-ih})}{p^{1/2}}, \quad h \in I.$$

Here I is some short interval, while the $G_p's$ are i.i.d standard complex Gaussian variables.

We include here the three main results of their paper, as well as indicate the important role that these results play in regards to the overall picture. The first of these results concerns intervals of length $(\log T)^{\theta}$ (where $\theta > 0$ is fixed) and is a verification of the leading and subleading behaviour in the conjectured result in [5].

Theorem 1.5 (Arguin-Dubach-Hartung). *For $\theta > 0$ fixed, we have*

$$\lim_{T\to\infty} \max_{|h|\leq(\log T)^{\theta}} \frac{X_T(h)-\sqrt{1+\theta}\log\log T}{\log\log\log T} = \frac{-1}{4\sqrt{1+\theta}} \quad \textit{in probability.}$$

An equivalent way of writing this result is to say that

$$\max_{|h|\leq(\log T)^{\theta}} X_T(h) = \sqrt{1+\theta}\log\log T - \frac{1}{4\sqrt{1+\theta}}\log\log\log T + o(\log\log\log T).$$

The key thing to note in this result is that the sub-leading term has coefficient

$-\frac{1}{4\sqrt{1+\theta}}$ so for small θ this is effectively $-1/4$.

The second of these results allows one to interpolate between the $1/4$ and $3/4$ regimes, provided θ tends to zero at a suitable rate.

Theorem 1.6. *For $\alpha \in (0,1)$ and $\theta = (\log\log T)^{-\alpha}$, we have*

$$\lim_{T\to\infty} \max_{|h|\leq(\log T)^\theta} \frac{X_T(h) - \sqrt{1+\theta}\log\log T}{\log\log\log T} = -\frac{1+2\alpha}{4} \quad \textit{in probability.}$$

An equivalent way of writing this is

$$\max_{|h|\leq(\log T)^\theta} X_T(h) = \sqrt{1+\theta}\log\log T - \frac{1+2\alpha}{4}\log\log\log T + o(\log\log\log T).$$

With this, we observe that if we let $\alpha \to 0$ (so $\theta > 0$) we have the $1/4$ regime seen prior. In the case where $\alpha \to 1$ (so θ is small, since T is large), we have the $3/4$ regime. So this result neatly provides a smooth transition between the two regimes. A key ingredient to their proof is the following result, which harks back to Theorem 1.4 and sheds more light on the conjectured result of Farmer, Gonek and Hughes.

Theorem 1.7. *Let $y > 0$ and $y = o(\log\log T)$. Then we have*

$$\mathbb{P}\left(\max_{|h|\leq 1} X_T(h) > \log\log T - \frac{3}{4}\log\log\log T + y\right) \leq Cye^{-2y}e^{-y^2/\log\log T},$$

for some constant $C > 0$.

This refined result adds more weight to the conjectured result of Farmer, Gonek and Hughes; the reason here being that, as y gets larger we see a transition from an exponential regime to a Gaussian regime. Because of this transition to a Gaussian regime, it adds more weight to the conjectured Gaussian result of Farmer, Gonek and Hughes.

1.3.6 Farmer-Gonek-Hughes

The notions of Large Deviations Theory were later applied by Farmer, Gonek and Hughes [21] to compute the maximum value of the zeta function.

The questions they posed, which we reintroduce here, are the following:

"How does the Riemann zeta function grow on the critical line?
What can we say about the value distribution of $\zeta(1/2+it)$ as we vary t?"

An initial result, conditional on the Riemann hypothesis, is the following (see [53]):

$$\zeta\left(\frac{1}{2}+it\right)=O\left(\exp\left(C\frac{\log t}{\log\log t}\right)\right) \tag{28}$$

for some positive constant C. Additionally, results of the form

$$\zeta\left(\frac{1}{2}+it\right)=\Omega\left(\exp\left(C'\sqrt{\frac{\log t}{\log\log t}}\right)\right) \tag{29}$$

have also been established; as before C' is a positive constant. Various improvements and refinements have been made regarding the possible values C' can take here, and we provide references to these improvements and refinements here [41, 7, 51, 36]. Here, Ω refers to the fact that the function $\zeta(\frac{1}{2}+it)$ takes the argument inside the brackets infinitely often. Therefore the maximum size of the zeta function on the critical line must lie somewhere between these two exponential terms in (28) and (29). A natural question to then ask – given this information – is the following: which of (28) or (29) is closer to the truth?

A more recent result of Bondarenko and Seip [9] provides some clarity regarding this problem:

Theorem 1.8 (Bondarenko-Seip)**.** *Let $0 \leq \beta < 1$ be given and let c be a positive number less than $\sqrt{1-\beta}$. If T is sufficiently large, there exists a t, $T^{\beta} \leq t \leq T$, such that*

$$\left|\zeta\left(\frac{1}{2}+it\right)\right| \geq \exp\left(c\sqrt{\frac{\log T\log\log\log T}{\log\log T}}\right).$$

The implication of this result is that $|\zeta(1/2+it)| = \Omega\left(\exp\left((1+o(1))\sqrt{\frac{\log T\log\log\log T}{\log\log T}}\right)\right)$. This result therefore suggests that equation (29) is closer to the truth.

Farmer, Gonek and Hughes also sought to give a definitive answer to the above question (in fact, their work predates that of Bondarenko and Seip), and their conjectured answer (while also indicating that (29) is closer to the truth) suggests a larger maximum; their calculations utilise the model we discuss below.

1.3.7 Hybrid Euler-Hadamard Product

Recall the Keating-Snaith conjecture given in (25):

$$\frac{1}{T}\int_0^T\left|\zeta\left(\frac{1}{2}+it\right)\right| \mathrm{d}t \sim a(k)\frac{G^2(k+1)}{G(2k+1)}(\log T)^{k^2}.$$

From the product on the right-hand side, it can be observed that there is an arithmetic contribution in the form of the term $a(k)$. However, when we consider the random matrix analogue in the form of the characteristic polynomial, we see that there is no arithmetic contribution. The prime numbers do not appear in the random matrix computations of the moments, and instead must be inserted in an ad hoc manner.

Given the explicit connection between the zeta function and the primes (as illustrated by the Euler-product expansion), it becomes abundantly clear that any model for the zeta function should incorporate the primes in some manner.

From the earlier result of Goldston, Keating and Snaith remarked about the possible splitting of the moments of zeta as a product over the primes times a product over the zeros of zeta. This idea was more properly realised by Gonek, Hughes and Keating [26] in the following product:

Theorem 1.9 (Gonek-Hughes-Keating). *Let $s = \sigma + it$ with $\sigma \geq 0$ and $|t| > 2$, let $X \leq 2$ be a real parameter, and let K be any fixed positive integer. Let $f(x)$ be a non-negative C^∞-function of mass one supported in $[0,1]$, and set $u(x) = Xf(X\log(x/e)+1)/x$. Thus $u(x)$ is a function of mass one supported in $[e^{1-1/X}, e]$. Set*

$$U(z) = \int_0^\infty u(x) E_1(z \log x)\, \mathrm{d}x, \tag{30}$$

where $E_1(z) = \int_z^\infty e^{-w}/w \,\mathrm{d}w$ is the exponential integral. Then

$$\zeta(s) = P_X(s) Z_X(s) \left(1 + O\left(\frac{X^{2-\sigma+K}}{(|t|\log X)^K}\right) + O(X^{-\sigma}\log X)\right), \tag{31}$$

where

$$P_X(s) = \exp\left(\sum_{n \leq X} \frac{\Lambda(n)}{n^s \log n}\right), \tag{32}$$

$\Lambda(n)$ is the von-Mangoldt function, and

$$Z_X(s) = \exp\left(-\sum_{\rho_n} U((s-\rho_n)\log X)\right). \tag{33}$$

The constants implied by the O-terms depend on f and K.

Here ρ_n denotes the non-trivial zeroes of the zeta function.

The intuition behind this model is the following: the P_X term can be viewed as a product over the primes, while the Z_X term can be seen as a product over the zeros, as seen below. Therefore one can derive results from the zeta function by computing separately results for the product over the primes and the product over the zeros and subsequently combining them. This is the approach Farmer, Gonek and Hughes utilised to compute the maximum size of the zeta function, and we adopt this same approach here. For this, some observations were key, and we document these below.

The parameter X in the Hybrid Euler-Hadamard product controls the relative influence of both the primes and the zeros. If the parameter X is large, the number of primes picked up in P_X is larger and the zeros closest to s will affect the product.

The justification here stems from the following: since $u(x)$ has support which is concentrated around e, $U(z)$ is roughly $E_1(z)$, which is asymptotic to $-\gamma - \log z$ as $z \to 0$. If one assumes the Riemann hypothesis, with $s = \frac{1}{2} + it$ we have that

$$Z_X\left(\frac{1}{2} + it\right) = \exp\left(-\sum_{\gamma_n} U(i(t-\gamma_n)\log X)\right) \approx \prod_{\gamma_n} i(t-\gamma_n)e^{\gamma}\log X.$$

From this expression it soon becomes clear that the zeros very close to s will affect the product. If, on the contrary X is small, it is the zeros furthest from s that will most greatly impact $Z_X(s)$, but by reducing the value of X we diminish the contribution from the primes. In particular, as the value of X decreases the prime contribution lessens, with the main contribution coming from the zeros. Thus the model used here approaches that of the characteristic polynomial model in random matrix theory.

In the intermediate range, it is expected that both Z_X and P_X contribute independently, and Gonek, Hughes and Keating provide evidence of this.

1.3.8 Large Deviations for Zeta

With the model above formally introduced, Farmer, Gonek and Hughes used this model to conjecture large values of zeta. In particular, they computed separately the large values of Z_X and P_X, and by performing a convolution argument they computed the large values of zeta in the intermediary region, where both terms contribute to the expression.

For the case where the dominant contribution comes from Z_X, we are looking at a function of the zeros of zeta, and thus the correspondence to random matrix theory provides an effective approach for computing these maximal values.

In this case they are looking at extreme values of the characteristic polynomial, so naturally the tools and techniques from large deviations theory are effective at computing the maximal values.

For the characteristic polynomial of a random unitary matrix (that is, a random CUE matrix), they compute the following result:

Lemma 1.1 (Farmer-Gonek-Hughes)**.** *If $\delta > 0$ is fixed and $\delta \leq \lambda \leq 1 - \delta$, then*

$$\mathbb{P}\left\{\max_{\theta} |\Lambda_U(\theta)| \geq \exp(N^{\lambda})\right\} = \exp\left(-\frac{N^{2\lambda}}{(1-\lambda)\log N}(1 + o(1))\right). \tag{34}$$

We will not discuss the steps here as the techniques will be applied later to obtain refined results for the characteristic polynomial. By refined here, we mean that we give a more explicit expression for the $o(1)$ term present in (36).

Upon applying the results of the characteristic polynomial to the product over zeros Z_X they derive the following conjecture:

Conjecture 1.3 (Farmer-Gonek-Hughes). *If* $2 < X < \log^A T$, *then*

$$\max_{t\in[0,T]} \left| Z_X\left(\frac{1}{2}+it\right)\right| = \exp\left((1+o(1))\sqrt{\frac{1}{2}\log T \log\log T}\right).$$

For the case of P_X, the first observation is that

$$\begin{aligned}P_X\left(\frac{1}{2}+it\right) = \exp\left(\sum_{n\leq X}\frac{\Lambda(n)}{n^{1/2+it}\log n}\right) &= \exp\left(\sum_{p\leq X}\frac{1}{p^{1/2+it}}\right)\times \exp(O(\log\log X))\\ &= \exp\left(P_X^*\left(\frac{1}{2}+it\right)\right)\times \exp(O(\log\log X)).\end{aligned}$$

Here

$$P_X^*\left(s\right) = \exp\left(\sum_{p\leq X}\frac{1}{p^s}\right) = \prod_{p\leq X}\frac{1}{p^s}.$$

For the leading order results of P_X it is therefore sufficient to compute the results for large values of P_X^*. For this the authors use what is essentially a modified central limit theorem to show that the distribution of $P_X^*(\frac{1}{2}+it)$ converges as $X\to\infty$ to a Gaussian with mean 0 and variance $\frac{1}{2}\log\log X$. The randomness in this model comes from treating the p^{-it} as independent random variables due to the logarithm of the primes, $\log p$ being linearly independent for distinct primes.

In the regime where P_X is the dominant contribution to the Hybrid product, taking $X=\exp(\sqrt{\log T})$ and independently choosing $T\log^c T$ values of t yields the following conjecture

Conjecture 1.4. *If* $X=\exp(\sqrt{\log T})$, *then*

$$\max_{t\in[0,T]} \left| P_X\left(\frac{1}{2}+it\right)\right| = \exp\left((1+o(1))\sqrt{\frac{1}{2}\log T \log\log T}\right).$$

What is our motivation for taking $X=\exp(\sqrt{\log T})$? Recall from Section 1.3.1 that $\zeta(\frac{1}{2}+it)$ tends to a Gaussian with mean 0 and variance $\frac{1}{2}\log\log T$ as T gets large (by a simple reformulation of Selberg's result). In the regime where P_X is the dominant contribution, it is reasonable to choose X such that the two results have matching variance. Thus from making the identification $\log\log X \leftrightarrow \frac{1}{2}\log\log T$ we get $X=\exp(\sqrt{\log T})$.

We see from this result that in the two regimes where the dominant contributions to the product come from P_X and Z_X respectively, the conjectured maxima are in agreement. The final step in their calculation was to consider the intermediate regime, where both functions contribute to the product. In the intermediate regime where the value of X is not too large, it is expected that P_X and Z_X behave independently, and Gonek, Hughes and Keating [26] give evidence that this is indeed the case.

We defer some of the details until later, but we remark that in the intermediate case, they obtain the conjectured result that

Conjecture 1.5.

$$\max_{t\in[0,T]}\left|\zeta\left(\frac{1}{2}+it\right)\right| = \exp\left((1+o(1))\sqrt{\frac{1}{2}\log T\log\log T}\right).$$

Therefore in all possible regimes the results are in agreement, thus adding weight to the conjecture.

1.4 Overview of Thesis

This thesis is split into seven chapters. Chapter 2 introduces the necessary tools and material from the theory of large deviations, and it is in this Chapter that we prove a more general result for the large deviations principle, predicated on a number of assumptions. This result is utilised throughout and essentially forms the backbone of this thesis.

In Chapter 3 we apply this large deviations result to the circular beta ensemble, obtaining a general result for the maximum of the characteristic polynomial of a CβE matrix. These results go much further than Farmer, Gonek and Hughes outlined in their results (as seen in Chapter 1.3.6), who focused their attention exclusively on the circular unitary ($\beta = 2$) ensemble.

This leads naturally into the computation of refined large deviations results for the zeta function in Chapter 4, which we separate according to the product over primes and the product over zeros. Here we propose an alternative model for P_X than that in the original literature, and we use this to go further and obtain refined results for the product over primes.

That this alternate model is suitable for our purposes is documented in Chapter 5 where we provide theoretical evidence, as well as numerical evidence, to support our argument.

The use of these two models leads to some inconsistency in the results, with the refined behaviour for these models not in agreement with one another. We discuss this in Chapter 6, where we discuss the possible reasons for this difference as well as discuss which results we align with here.

We provide a number of theoretic results in this paper, and in Chapter 7 we perform some numerics to further demonstrate the validity and consistency of our findings, as well as support the refined results we propose for the zeta function.

Additional calculations and details are provided in the Appendices, where we clarify some of the subtleties that arose during research.

1.5 Things Not Considered Here

The research presented here was completed with the Riemann zeta function being the main focus. As such, there is scope for further research to be done, and there are various directions and branches that this research could take as a result. We highlight some of the notable ones here, but note that this is by no means an exhaustive list.

- In Chapter 3 we present more general results for the large deviations principle, predicated on a series of assumptions which we detail in Chapter 2. Is the result we obtain in Chapter 2 the best we can do given the current mathematical framework, or can we refine this? Can we obtain more terms in our large deviations result? Is it possible to remove some of our assumptions, or replace these with more general assumptions that allow us to apply our results to more settings?

- We computed refined large deviations results here, predicated on a number of assumptions which we then apply to the circular beta ensemble. It is plausible that the results here can be applied to other matrix ensemble and groups: the Gaussian ensembles, orthogonal and symplectic groups (the relevance and significance of such results is unclear). Is this achievable with our current results, or does the computation of results for different matrix ensembles require additional assumptions?

- We have applied our findings to the Riemann zeta function, but is it possible that these results can be applied further? For example, can we utilise some of these ideas and concepts and apply them to other number-theoretic concepts such as other L-functions? What about in a broader context such as number fields or elliptic curves? Is there motivation for doing this, and were we to apply these ideas to other scenarios, what would these results mean in the context of those other mathematical disciplines?

2 Large Deviations Theory

The focus of this Chapter is to lay the groundwork for the improvement of results in [21], and to achieve this we begin by first introducing some basic theory from the theory of large deviations; we later implement this theory to obtain more general large deviations results.

The results in this Chapter are predicated on some assumptions and we outline these below. Before doing so we first introduce the necessary concepts and ideas for obtaining these results, and these concepts stem from probability theory.

2.1 Basic Probability Theory

The following concepts, although not always stated explicitly, will be utilised throughout this thesis.

Definition 2.1. *Let $\mathcal{X}$ be a topological space, and let $\mathcal{P}(\mathcal{X})$ denote the set of all subsets of $\mathcal{X}$. We say that $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$ is a σ-algebra if the following three conditions are satisfied:*

(i) $\emptyset \in \mathcal{A}$*;*

(ii) For all $A \in \mathcal{A}$, A^c also lies in $\mathcal{A}$ (closure under complements);

(iii) If $(A_n)_{n\in\mathbb{N}}$ is a collection of sets such that $A_n \in \mathcal{A}$ for each $n \in \mathbb{N}$, then $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$ (closure under countable unions).

For the context of our results here, $\mathcal{X}$ will denote the topological space $(\mathbb{R}, \tau)$, where the topology τ on $\mathbb{R}$ is the usual topology. That is, open sets are identified with open intervals and closed sets sets with closed intervals.

Definition 2.2. *We define $\mathcal{B}$ as the σ-algebra generated by the collection of open subsets of $\mathbb{R}$. The sets contained in $\mathcal{B}$ are known as the Borel sets.*

Definition 2.3. *A probability measure is a map $\mathbb{P} : \mathcal{X} \to [0, \infty]$ which satisfies the following properties:*

(i) $\mathbb{P}(\emptyset) = 0$ *and* $\mathbb{P}(\mathcal{X}) = 1$*;*

(ii) For a collection $(A_n)_{n\in\mathbb{N}}$ of pairwise-disjoint sets,

$$P\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n).$$

With this definition, it now becomes clear what we mean when we think of Haar measure as being a probability measure.

Next we introduce the necessary tools and techniques from large deviations theory that will be essential to us throughout this thesis.

2.2 Introductory Concepts

Large deviations theory is the study of increasingly unlikely events. These ideas and techniques play an integral role in the work that follows, for our intentions here are to study the maximum values attained by various functions. Here, these maxima are extremely rare, to the point that detecting them numerically is often difficult; large deviations theory is effective for such scenarios. Therefore before we can continue, we must introduce the necessary concepts and ideas.

Note that the following results stated here are for the one-dimensional case. However, they generalise to higher dimensions (see [18] for more details).

Definition 2.4. *A rate function Λ^* is a lower semicontinuous mapping $\Lambda^* : \mathcal{X} \to [0, \infty]$ (such that for all $\alpha \in [0, \infty)$, the level set $\psi_{\Lambda^*}(\alpha) = \{x : \Lambda^*(x) \leq \alpha\}$ is a closed subset of $\mathcal{X}$).*

A good rate function is a rate function for which the level sets $\psi_{\Lambda^*}(\alpha)$ are compact subsets of $\mathcal{X}$.

Definition 2.5. *The sequence of random variables $(X_N)_{N\in\mathbb{N}}$ satisfies the large deviation principle (LDP) with rate function Λ^* and speed $B(N)$ if, for all $\Gamma \in \mathcal{B}$:*

$$-\inf_{x\in\Gamma^o} \Lambda^*(x) \leq \liminf_{N\to\infty} \frac{1}{B(N)} \log \mathbb{P}(X_N \in \Gamma) \leq \limsup_{N\to\infty} \frac{1}{B(N)} \log \mathbb{P}(X_N \in \Gamma) \leq -\inf_{x\in\bar{\Gamma}} \Lambda^*(x).$$

We refer to the function $B(N)$ here as the speed since it determines the rate of convergence of the logarithmic moment generating function (MGF), as highlighted in the following assumption:

Assumption 2.1. *For all $\lambda \in \mathbb{R}$ the logarithmic MGF, defined as the limit*

$$\Lambda(\lambda) = \lim_{N\to\infty} \frac{1}{B(N)} \log \mathbb{E}\left[e^{\lambda B(N) X_N}\right] \tag{35}$$

exists as an extended real number, that is, in $\mathbb{R} \cup \{\infty\}$.

Definition 2.6. *A convex function $\Lambda : \mathbb{R} \to (-\infty, \infty]$ is essentially smooth if the following properties are satisfied:*

(i) *$\mathcal{D}_\Lambda^o$ is non-empty. Here, $\mathcal{D}_\Lambda = \{\lambda \in \mathbb{R} : \Lambda(\lambda) < \infty\}$ is known as the effective domain[13];*

(ii) *Λ is differentiable on $\mathcal{D}_\Lambda^o$;*

(iii) *$\lim_{n\to\infty} |\Lambda'(\lambda_n)| = \infty$ whenever $\{\lambda_n\}$ is a sequence in $\mathcal{D}_\Lambda^o$ converging to a point on the boundary of $\mathcal{D}_\Lambda^o$. If Λ satisfies this property, we say that Λ is a steep function.*

[13] As will be seen from Assumption 2.3, this is automatically satisfied for our choice of Λ.

We can be confident in our statement of the convexity of Λ given the following result (see Lemma 2.3.9 of [18]).

Lemma 2.1. *Let Assumption 2.1 hold.*

(i) $\Lambda(\lambda)$ *is a convex function,* $\Lambda(\lambda) > -\infty$ *everywhere, and* $\Lambda^*(x)$ *is a good convex rate function;*

(ii) Suppose that $x = \Lambda'(\bar{\lambda})$ *for some* $\bar{\lambda} \in \mathcal{D}_\Lambda^o$. *Then*

$$\Lambda^*(x) = \bar{\lambda}x - \Lambda(\bar{\lambda}).$$

We omit the proof, but this result will be essential to later calculations.

Assumption 2.2. $\Lambda(\lambda)$ *is an essentially smooth lower semi-continuous function.*

Note that with Assumption 2.2, part *(ii)* of Lemma 2.1 is now reasonable, since $\Lambda(\lambda)$ being essentially smooth means that it is differentiable on its effective domain, via Definition 2.6.

Assumption 2.3. *0 belongs to the interior of* $\mathcal{D}_\Lambda = \{\lambda \in \mathbb{R} : \Lambda(\lambda) < \infty\}$.

In the case where the probability density functions (PDFs) p_N corresponding to the X_N are independent and identically distributed (IID), Assumption 2.3 tells us that Assumption 2.1 holds (see for example Page 43 of [18]).

In addition, with Lemma 2.1 included here, we see that it is indeed sensible for us to discuss the notion of essentially smooth since our given $\Lambda(\lambda)$ is convex. Further, $\Lambda(\lambda) > -\infty$, so $\Lambda : \mathbb{R} \to (-\infty, \infty]$ makes sense in the statement of Definition 2.6.

With Assumptions 2.1, 2.2 and 2.3, the Gärtner-Ellis theorem [18] tells us that the sequence of random variables $\{X_N\}_{N\in\mathbb{N}}$ satisfies the Large Deviation Principle (LDP) with good rate function

$$\Lambda^*(x) = \sup_{\lambda\in\mathbb{R}}(\lambda x - \Lambda(\lambda)). \tag{36}$$

Given we are working in the one-dimensional case with $\Gamma = (x, \infty]$ as our set in Definition 2.5, we then have that $\Gamma^o = (x, \infty]$ and $\bar{\Gamma} = [x, \infty]$. For the infima in this statement the difference in these two sets is not significant, and these results allow us to write this large deviations principle more concisely as

$$\lim_{N\to\infty} \frac{1}{B(N)} \log \mathbb{P}(X_N \geq x) = -\Lambda^*(x); \tag{37}$$

$\Lambda^*(x)$ is known as the Fenchel-Legendre transform (or convex dual) of $\Lambda(\lambda)$. With this notation, (37) is equivalent to saying that $\mathbb{P}(X_N \geq x) = e^{-B(N)\Lambda^*(x)(1+o(1))}$ as $N \to \infty$. This is the weakest statement we have; our goal is to obtain a much stronger result.

Let us return to (35). One additional assumption needed here is the following:

Assumption 2.4. *There exists a probability density function (pdf) $p_N(x)$ which decays in such a way that it satisfies the following integral relation:*

$$\mathbb{E}\left[e^{\lambda B(N)X_N}\right] = \int_{-\infty}^{\infty} e^{\lambda B(N)x} p_N(x)\,\mathrm{d}x. \tag{38}$$

2.3 Refined Large Deviations Results

With the tools and techniques of Large Deviations Theory formally introduced, we seek to refine and improve the weak result in (37) so that we can obtain something more applicable for our purposes.

The theory of large deviations coupled with Assumptions 2.1 - 2.4 allow us to conclude the following:

Theorem 2.1. *With $\Lambda(\lambda)$ defined as in (35) with corresponding pdf $p_N(x)$, we have that*

$$p_N(x) = \frac{B(N)}{\sqrt{2\pi f''(\bar{\lambda})}} e^{f(\bar{\lambda})} + o_x(1) \tag{39}$$

as $N \to \infty$, where $f(\lambda) = -B(N)\lambda x + \log \mathbb{E}[e^{\lambda B(N)X_N}]$ and $\bar{\lambda}$ satisfies $f'(\bar{\lambda}) = 0$. Further, suppose our logarithmic expectation takes the following form:

$$\log \mathbb{E}\left[e^{\lambda B(N)X_N}\right] = B(N)\Lambda(\lambda) + C(N)\tilde{\Lambda}(\lambda) + o(C(N))$$

where $C(N) = o(B(N))$ and $\tilde{\Lambda}(\lambda)$ is some function of λ. Then we can conclude that

$$\log \mathbb{P}(X_N \geq x) = -B(N)\Lambda^*(x) + C(N)\tilde{\Lambda}(\bar{\lambda}) - \tfrac{1}{2}\log B(N) + O_x(1).$$

It should be noted that it is not clear from the statement of Theorem 2.1 that the second derivative f'' exists. From Lemma 2.1 it follows that the function f is convex, and so by the Alexandrov theorem (see Theorem 3.11.2 of [43] for more) the second derivative exists almost everywhere; that is, except for a null set, which is a set of Lebesgue measure zero. The range we are working with is $\lambda \in \mathbb{R}$ (a non-null set) and so f'' exists.

Proof of Theorem 2.1. From Assumption 2.4 we can write

$$\mathbb{E}\left[e^{\lambda B(N)X_N}\right] = \int_{-\infty}^{\infty} e^{\lambda B(N)x} p_N(x)\mathrm{d}x = \mathcal{B}[p_N(x); -\lambda B(N)]$$

where $\mathcal{B}$ denotes the bilateral Laplace transform. Therefore we can invert this to give

$$
\begin{aligned}
p_N(x) &= \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{-\lambda B(N)x}\mathbb{E}\left[e^{\lambda B(N)X_N}\right] \mathrm{d}B(N)\lambda \\
&= \frac{B(N)}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{-\lambda B(N)x}\mathbb{E}\left[e^{\lambda B(N)X_N}\right] \mathrm{d}\lambda
\end{aligned}
$$

in the common region of convergence. Utilising the saddle-point method, we express the probability density function in the form

$$
p_N(x) = \frac{B(N)}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{f(\lambda)}\mathrm{d}\lambda,
$$

where $f(\lambda) = -B(N)\lambda x + \log \mathbb{E}\left[e^{\lambda B(N)X_N}\right]$. Letting $\bar{\lambda}$ denote the solution to $f'(\lambda) = 0$, it is clear that $\bar{\lambda}$ satisfies the equation

$$
x = \frac{\mathbb{E}\left[X_N e^{\bar{\lambda}B(N)X_N}\right]}{\mathbb{E}\left[e^{\bar{\lambda}B(N)X_N}\right]}, \tag{40}
$$

where the solution $\bar{\lambda}$ is unique by convexity of $\log \mathbb{E}[e^{\lambda B(N)X_N}]$ (recalling Lemma 2.1). Note that we make use of Assumption 2.2 that $\Lambda(\lambda)$ is an essentially smooth function, and is thus differentiable on $\mathcal{D}_\Lambda^o$. The leading contribution to this integral for $p_N(x)$ comes from a small interval of length 2ϵ (where ϵ is to be made explicit) around the saddle point, so we have

$$
p_N(x) = \frac{B(N)}{2\pi i}\int_{\bar{\lambda}-i\epsilon}^{\bar{\lambda}+i\epsilon} e^{f(\lambda)}\,\mathrm{d}\lambda + \frac{B(N)}{2\pi i}\int_{c-i\infty}^{\bar{\lambda}-i\epsilon} e^{f(\lambda)}\,\mathrm{d}\lambda + \frac{B(N)}{2\pi i}\int_{\bar{\lambda}+i\epsilon}^{c+i\infty} e^{f(\lambda)}\,\mathrm{d}\lambda.
$$

Here our interval around the saddle point is vertical in the complex plane because $f''(\lambda) = \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}\log \mathbb{E}[e^{\lambda B(N)X_N}] > 0$ for all λ by convexity, again from Lemma 2.1.

Consider the second integral (the following arguments also apply to the third integral): by rewriting the integrand and performing integration by parts, we have

$$
\begin{aligned}
\frac{B(N)}{2\pi i}\int_{c-i\infty}^{\bar{\lambda}-i\epsilon} e^{f(\lambda)}\mathrm{d}\lambda &= \frac{B(N)}{2\pi i}\int_{c-i\infty}^{\bar{\lambda}-i\epsilon} \frac{f'(\lambda)}{f'(\lambda)}e^{f(\lambda)}\,\mathrm{d}\lambda \\
&= \frac{B(N)}{2\pi i}\left(\left[\frac{e^{f(\lambda)}}{f'(\lambda)}\right]_{c-i\infty}^{\bar{\lambda}-i\epsilon} + \int_{c-i\infty}^{\bar{\lambda}-i\epsilon} \frac{f''(\lambda)}{(f'(\lambda))^2}e^{f(\lambda)}\,\mathrm{d}\lambda\right).
\end{aligned}
$$

We are able to do this as the integrand is analytic, since we are working away from $\lambda = \bar{\lambda}$. To evaluate this we consider only the dominant contribution, which comes from the leading term in these brackets. For the evaluation at $\bar{\lambda} - i\epsilon$, we evaluate the exponential term by performing a Taylor expansion around the point $\bar{\lambda}$:

$$
\begin{aligned}
f(\bar\lambda - i\epsilon) &= f(\bar\lambda) + (\bar\lambda - i\epsilon - \bar\lambda) f'(\bar\lambda) + \frac{(\bar\lambda - i\epsilon - \bar\lambda)^2}{2} f''(\bar\lambda) + \cdots \\
&\ll_{\bar\lambda, N} -\bar\lambda B(N) x + \log \mathbb{E}[e^{\bar\lambda B(N) X_N}] - \frac{\epsilon^2}{2} B(N) \Lambda''(\bar\lambda).
\end{aligned}
$$

By splitting the term inside the integral, we can perform an additional application of integration by parts,

$$
\int_{c-i\infty}^{\bar\lambda - i\epsilon} \frac{f''(\lambda)}{(f'(\lambda))^2} e^{f(\lambda)} \, \mathrm{d}\lambda = \left[\frac{e^{f(\lambda)}}{(f'(\lambda))^3} \right]_{c-i\infty}^{\bar\lambda - i\epsilon} + \int_{c-i\infty}^{\bar\lambda - i\epsilon} e^{f(\lambda)} \left[\frac{f'''(\lambda)^2 - f''(\lambda) f^{(4)}(\lambda)}{(f''(\lambda))^2} \right] \mathrm{d}\lambda
$$

and we see after further applications of integration by parts that the terms of our expression decrease in magnitude. Therefore our expression for the integral can be bounded above by some multiple of the first term. Combining this with the fact that $f'(\bar\lambda)$ is zero (so $f'(\bar\lambda - i\epsilon)$ is of order $B(N)\epsilon$ near $\bar\lambda$) gives the result

$$
\begin{aligned}
\frac{B(N)}{2\pi i} \int_{c-i\infty}^{\bar\lambda - i\epsilon} e^{f(\lambda)} \mathrm{d}\lambda &\ll_{\epsilon, N} \frac{B(N) e^{f(\bar\lambda - i\epsilon)}}{f'(\bar\lambda - i\epsilon)} \\
&= O\left(\frac{\exp(-\bar\lambda B(N) x + \log \mathbb{E}[e^{\bar\lambda B(N) X_N}] - B(N) \Lambda''(\bar\lambda) \epsilon^2 / 2)}{\epsilon} \right).
\end{aligned}
$$

Returning to the original integral, we Taylor expand around the saddle point, giving

$$
\begin{aligned}
\frac{B(N)}{2\pi i} \int_{\bar\lambda - i\epsilon}^{\bar\lambda + i\epsilon} e^{f(\lambda)} \, \mathrm{d}\lambda &= \frac{B(N)}{2\pi i} e^{f(\bar\lambda)} \int_{\bar\lambda - i\epsilon}^{\bar\lambda + i\epsilon} e^{\frac{(\lambda - \bar\lambda)^2}{2} f''(\bar\lambda)} \mathrm{d}\lambda \\
&+ \frac{B(N)}{2\pi i} e^{f(\bar\lambda)} \int_{\bar\lambda - i\epsilon}^{\bar\lambda + i\epsilon} \sum_{l=1}^{\infty} \frac{1}{l!} \left(\sum_{k=3}^{\infty} \frac{(\lambda - \bar\lambda)^k}{k!} f^{(k)}(\bar\lambda) \right)^l e^{\frac{(\lambda - \bar\lambda)^2}{2} f''(\bar\lambda)} \mathrm{d}\lambda.
\end{aligned}
$$

Here the sub-leading term is, after a change of variable $\lambda \to \lambda + \bar\lambda$ followed by another change of variable $\lambda \to i\lambda$,

$$
\begin{aligned}
&\frac{B(N)}{2\pi} e^{f(\bar\lambda)} \int_{-\epsilon}^{\epsilon} \sum_{l=1}^{\infty} \frac{1}{l!} \left(\sum_{k=3}^{\infty} \frac{(i\lambda)^k}{k!} f^{(k)}(\bar\lambda) \right)^l e^{-\frac{\lambda^2}{2} f''(\bar\lambda)} \mathrm{d}\lambda \\
&\qquad = O\left(B(N)^2 |\Lambda^{(4)}(\bar\lambda)| \exp\left(-\bar\lambda B(N) x + \log \mathbb{E}\left[e^{\bar\lambda B(N) X_N} \right] \right) \epsilon^5 \right).
\end{aligned}
$$

We have made use of the fact that, in the case where k is odd, the expression inside the integral is odd and so the resulting integral is 0. So we need to consider the smallest value of k for which we get a non-zero integrand for our error term, which comes from the case $k = 4$. This gives a contribution of ϵ^4, which in addition to the range 2ϵ we are integrating over gives an overall contribution of ϵ^5 to the error term.

It is important to note that we are assuming that $\Lambda^{(4)}(\bar{\lambda})$ exists. If this is zero, we do not have an upper bound on ϵ in our later calculations.

The remaining components inside the big-O term follow from the same ideas utilised earlier. The leading term – following the changes of variable $\lambda \to \lambda + \bar{\lambda}$ and $\lambda \to i\lambda$ – can be written as

$$\frac{B(N)}{2\pi} e^{f(\bar{\lambda})} \int_{-\epsilon}^{\epsilon} e^{-\lambda^2/2} \, \mathrm{d}\lambda.$$

We now let ϵ approach infinity; the reason for this is that the exponential decays rapidly away from $\bar{\lambda}$. Some care is needed however, as we need to ensure that our error terms do not get too large in the limit. Overall, we have

$$\begin{aligned} p_N(x) = & \frac{B(N)}{2\pi} e^{f(\bar{\lambda})} \int_{-\epsilon}^{\epsilon} e^{-\lambda^2/2} \mathrm{d}\lambda \\ & + O\left(\frac{\exp\left(-\bar{\lambda}B(N)x + \log \mathbb{E}\left[e^{\bar{\lambda}B(N)X_N}\right] - B(N)\Lambda''(\bar{\lambda})\epsilon^2/2\right)}{\epsilon} \right) \\ & + O\left(B(N)^2 |\Lambda^{(4)}(\bar{\lambda})| \exp\left(-\bar{\lambda}B(N)x + \log \mathbb{E}\left[e^{\bar{\lambda}B(N)X_N}\right]\right) \epsilon^5\right). \end{aligned}$$

For the first of these error terms, recall that $\log \mathbb{E}[e^{\bar{\lambda}B(N)X_N}]$ behaves roughly like $B(N)\Lambda(\bar{\lambda})$ for large N, so $-\bar{\lambda}B(N)x + \log \mathbb{E}[e^{\bar{\lambda}B(N)X_N}]$ behaves like $-B(N)\Lambda^*(x)$ for large N (note that we are letting N tend to infinity, so this leading order approximation is sufficient for our purposes). With this in mind, Taylor expanding the exponential gives in the O-term

$$\frac{\exp\left(-B(N)\Lambda^*(x)\right)\left(1 - B(N)\Lambda''(\bar{\lambda})\epsilon^2 + \cdots\right)}{\epsilon}.$$

Therefore in order to ensure that both sub-leading terms stay small in ϵ and N, we require that

$$\epsilon \ll \frac{-1 - \sqrt{1 + 2B(N)\Lambda''(\bar{\lambda}) \exp(-2B(N)\Lambda^*(x))}}{\exp(-B(N)\Lambda^*(x))B(N)\Lambda''(\bar{\lambda})}$$

or

$$\epsilon \gg \frac{-1 + \sqrt{1 + 2B(N)\Lambda''(\bar{\lambda}) \exp(-2B(N)\Lambda^*(x))}}{\exp(-B(N)\Lambda^*(x))B(N)\Lambda''(\bar{\lambda})},$$

in addition to the condition that

$$\epsilon \ll \left(\frac{e^{B(N)\Lambda^*(x)}}{B(N)^2|\Lambda^{(4)}(\bar{\lambda})|}\right)^{1/5}.$$

If we therefore choose ϵ such that

$$\frac{-1+\sqrt{1+2B(N)\Lambda''(\bar\lambda)\exp(-2B(N)\Lambda^*(x))}}{\exp(-B(N)\Lambda^*(x))B(N)\Lambda''(\bar\lambda)} \ll \epsilon \ll \left(\frac{e^{B(N)\Lambda^*(x)}}{B(N)^2|\Lambda^{(4)}(\bar\lambda)|}\right)^{1/5}$$

our error terms do not get large as we let ϵ approach infinity, and therefore we can say that the error terms are of order $o(1)$ as we let N approach infinity. So after evaluating the leading order integral we have the result that

$$p_N(x) = \frac{B(N)}{\sqrt{2\pi f''(\bar\lambda)}} e^{f(\bar\lambda)} + o_x(1),$$

thus completing the first part of the Theorem. With this result, we now compute the probability

$$\begin{aligned}\mathbb{P}(X_N \geq x) &= \int_x^\infty p_N(u)\,\mathrm{d}u \\ &= \frac{B(N)}{\sqrt{2\pi f''(\bar\lambda)}}\int_x^\infty e^{f(\bar\lambda)} + o_x(1)\,\mathrm{d}u \\ &= \frac{B(N)}{\sqrt{2\pi f''(\bar\lambda)}}\int_x^\infty e^{-g(u)}\,\mathrm{d}u + o_x(1).\end{aligned}$$

Note that our o-term is with respect to N, and since we are integrating with respect to the variable u here this term is unchanged by the integration. Here we set $-g(u) = f(\bar\lambda) = -B(N)\bar\lambda u + \log\mathbb{E}[e^{\bar\lambda B(N)X_N}]$. Rewriting this equation and performing integration by parts gives

$$\frac{B(N)}{\sqrt{2\pi f''(\bar\lambda)}}\int_x^\infty \frac{-g'(u)}{-g'(u)}e^{-g(u)}\,\mathrm{d}u = \frac{B(N)}{\sqrt{2\pi f''(\bar\lambda)}}\left[-\frac{e^{-g(u)}}{g'(u)}\right]_x^\infty = \frac{B(N)}{\sqrt{2\pi f''(\bar\lambda)}}\frac{e^{-g(x)}}{g'(x)},$$

where the integral vanishes since $g''(u) = 0$. Taking logarithms of both sides of the equation, we have

$$\begin{aligned}\log\mathbb{P}(X_N \geq x) &= \log B(N) - \tfrac{1}{2}\log 2\pi - \tfrac{1}{2}\log f''(\bar\lambda) - g(x) - \log g'(x) + o_x(1) \\ &= \log B(N) - \tfrac{1}{2}\log 2\pi - \tfrac{1}{2}\log B(N) - \tfrac{1}{2}\log\Lambda''(\bar\lambda) \\ &\quad - B(N)\bar\lambda x + \log\mathbb{E}\left[e^{\bar\lambda B(N)X_N}\right] - \log B(N) - \log\left(\tfrac{\mathrm{d}\bar\lambda}{\mathrm{d}x}x + \bar\lambda\right) + O_x(1) \\ &= \log\mathbb{E}\left[e^{\bar\lambda B(N)X_N}\right] - B(N)\bar\lambda x - \tfrac{1}{2}\log B(N) + O_x(1) \\ &= B(N)\Lambda(\bar\lambda) + C(N)\tilde\Lambda(\bar\lambda) + o(C(N)) - B(N)\bar\lambda x - \tfrac{1}{2}\log B(N) + O_x(1) \\ &= -B(N)\Lambda^*(x) + C(N)\tilde\Lambda(\bar\lambda) - \tfrac{1}{2}\log B(N) + O_x(1)\end{aligned}$$

as desired. We have yet to fully justify some of the final steps in the proof, so we include this justification here.

In going from the second line to the third, we've introduced a $O_x(1)$ term into our calculations. This comes from the fact that the terms in x are of constant order, so we absorb them into a $O_x(1)$ term to tidy up the expression. For the $\frac{1}{2}\log f''(\bar{\lambda})$ term, utilising our assumption for the expression in Theorem 2.1 and differentiating this expression gives

$$\begin{aligned}\tfrac{1}{2}\log f''(\bar{\lambda}) &= \tfrac{1}{2}\log(B(N)\Lambda(\bar{\lambda}) + C(N)\Lambda^{(1)}(\bar{\lambda}) + o(C(N))) \\ &= \tfrac{1}{2}\log B(N) + O_x(1)\end{aligned}$$

at leading order. Therefore the steps of the proof are valid. One could also obtain this without having to resort to the logarithmic expectation taking the assumed form, instead attaining this expression for the second derivative of f'' by looking purely at the leading order behaviour. □

3 Refined Large Deviations for the Circular β Ensemble

In this Chapter, we utilise the large deviations result from Chapter 2 and apply this to the circular beta ensemble (CβE). This allows us to compute explicit results for the circular unitary ensemble (CUE) by setting $\beta = 2$, which we utilise in Chapter 4 to compute refined large deviations results for the Riemann zeta function.

To compute generalised results for the CβE we take the result introduced in (5) and average over the eigenphases, leading to the following result as stated in [35]:

$$
\begin{aligned}
M_N(\beta, s) &= \mathbb{E}\left[|\Lambda_{U,\beta}(\theta)|^s\right] \\
&= \frac{1}{(2\pi)^N Z_{N,\beta}} \int_0^{2\pi} \cdots \int_0^{2\pi} \mathrm{d}\theta_1 \dots \mathrm{d}\theta_N \cdot \prod_{1 \le j < m \le N} \left| e^{i\theta_j} - e^{i\theta_m} \right|^\beta \times \left| \prod_{p=1}^{N} \left(1 - e^{i(\theta_p - \theta)}\right) \right|^s \\
&= \prod_{j=0}^{N-1} \frac{\Gamma(1 + j\beta/2)\Gamma(1 + s + j\beta/2)}{(\Gamma(1 + s/2 + j\beta/2))^2},
\end{aligned}
$$

where $\Lambda_{U,\beta}(\theta)$ denotes the characteristic polynomial of a random CβE matrix ($\beta =$ 1, 2, 4). By taking logarithms of this result, we obtain the following result:

Theorem 3.1. *For $M_N(\beta, s) = \mathbb{E}\left[|\Lambda_{U,\beta}(\theta)|^s\right]$ and for $s \ll N$ we have the following result for $\beta = 1,\ 2,\ 4$:*

$$
\log M_N(\beta, s) = \frac{s^2}{2\beta} \log N - \frac{s^2}{2\beta} \log s - \frac{s^2}{\beta} \log 2 + \frac{s^2}{2\beta} \log \beta + \frac{3s^2}{4\beta} + o(s^2).
$$

Proof of Theorem 3.1. We begin by taking logarithms of the above result for $M_N(\beta, s)$, which gives the following result:

$$
\log M_N(\beta, s) = \sum_{j=0}^{N-1} \log \Gamma(1 + j\beta/2) + \log \Gamma(1 + s + j\beta/2) - 2 \log \Gamma(1 + s/2 + j\beta/2).
$$

Isolating the $j = 0$ terms and making use of Stirling's formula:

$$
\log \Gamma(1 + z) = z \log z - z + \frac{1}{2} \log z + \frac{1}{2} \log 2\pi + O\left(\frac{1}{z}\right),
$$

the logarithmic expectation becomes

$$
\begin{aligned}
\log M_N(\beta,s) &= \log\Gamma(1+s) - 2\log\Gamma(1+s/2) + \sum_{j=1}^{N-1}\log\Gamma(1+j\beta/2)\\
&\quad + \log\Gamma(1+s+j\beta/2) - 2\log\Gamma(1+s/2+j\beta/2)\\
&= \left(s\log s - s + \frac{1}{2}\log s + \frac{1}{2}\log 2\pi + O\left(\frac{1}{s}\right)\right) - 2\Bigg(\frac{s}{2}\log\left(\frac{s}{2}\right) - \frac{s}{2}\\
&\quad + \frac{1}{2}\log\left(\frac{s}{2}\right) + \frac{1}{2}\log 2\pi + O\left(\frac{1}{s}\right)\Bigg) + \sum_{j=1}^{N-1}\Bigg(\frac{j\beta}{2}\log\left(\frac{j\beta}{2}\right) - \frac{j\beta}{2}\\
&\quad + \frac{1}{2}\log\left(\frac{j\beta}{2}\right) + \frac{1}{2}\log 2\pi + O\left(\frac{2}{j\beta}\right)\Bigg) + \Bigg(\left(s+\frac{j\beta}{2}\right)\log\left(s+\frac{j\beta}{2}\right)\\
&\quad - \left(s+\frac{j\beta}{2}\right) + \frac{1}{2}\log\left(s+\frac{j\beta}{2}\right) + \frac{1}{2}\log 2\pi + O\left(\frac{1}{s+j\beta/2}\right)\Bigg)\\
&\quad - 2\Bigg(\left(\frac{s}{2}+\frac{j\beta}{2}\right)\log\left(\frac{s}{2}+\frac{j\beta}{2}\right) - \left(\frac{s}{2}+\frac{j\beta}{2}\right) + \frac{1}{2}\log\left(\frac{s}{2}+\frac{j\beta}{2}\right)\\
&\quad + \frac{1}{2}\log 2\pi + O\left(\frac{2}{s+j\beta}\right)\Bigg).
\end{aligned}
$$

The first two terms simplify to give $s\log 2 - \frac{1}{2}\log s + \log 2 - \frac{1}{2}\log 2\pi + O(1/s)$. For the summation we split this expression into three parts, with each part represented by the bracketed expressions above. We then evaluate each part individually before combining them again to obtain the result in the theorem.

3.1 First Part of Sum

The first part of the sum gives (after collecting terms)

$$
\begin{aligned}
&\sum_{j=1}^{N-1}\left(\frac{1}{2}+\frac{j\beta}{2}\right)\log\left(\frac{j\beta}{2}\right) - \frac{j\beta}{2} + \frac{1}{2}\log 2\pi + O\left(\frac{2}{j\beta}\right)\\
&= \frac{1}{2}\log((N-1)!) + \frac{N-1}{2}\log\beta - \frac{N-1}{2}\log 2 + \left(\sum_{j=1}^{N-1}\frac{j\beta}{2}\log j\right) + \frac{N(N-1)}{4}\beta\log\beta\\
&\qquad - \frac{N(N-1)}{4}\beta\log 2 - \frac{N(N-1)}{4}\beta + \frac{N-1}{2}\log 2\pi + O(\log N)\\
&= \frac{1}{2}\log((N-1)!) + \frac{N}{2}\log\beta - \frac{N}{2}\log 2 + \left(\sum_{j=1}^{N-1}\frac{j\beta}{2}\log j\right) + \frac{N(N-1)}{4}\beta\log\beta - \frac{N(N-1)}{4}\beta\log 2\\
&\qquad - \frac{N(N-1)}{4}\beta + \frac{N}{2}\log 2\pi + O(\log N).
\end{aligned}
$$

3.2 Second Part of Sum

For the second part of the sum, we have

$$\sum_{j=1}^{N-1}\left(\frac{1}{2}+s+\frac{j\beta}{2}\right)\log\left(s+\frac{j\beta}{2}\right)-s-\frac{j\beta}{2}+\frac{1}{2}\log 2\pi+O\left(\frac{1}{s+j\beta/2}\right).$$

Here a little more care is needed. For our purposes, s is a function of N which grows at a rate slower than N (so $s \ll N$). However, we need to be careful when evaluating this sum as some terms in j may be smaller than s; when does this occur?

A quick check shows that the change occurs at $[2s/\beta]$, where $[\cdot]$ denotes the integer part. Splitting the sum appropriately gives

$$\begin{aligned}&\sum_{j=1}^{[2s/\beta]}\left(\frac{1}{2}+s+\frac{j\beta}{2}\right)\log s+\left(\frac{1}{2}+s+\frac{j\beta}{2}\right)\left[\frac{j\beta}{2s}-\frac{j^2\beta^2}{8s^2}+\cdots\right]-s-\frac{j\beta}{2}+\frac{1}{2}\log 2\pi\\&\qquad+O\left(\frac{1}{s}\right)\\&+\sum_{[2s/\beta]+1}^{N-1}\left(\frac{1}{2}+s+\frac{j\beta}{2}\right)\log\left(\frac{j\beta}{2}\right)+\left(\frac{1}{2}+s+\frac{j\beta}{2}\right)\left[\frac{2s}{j\beta}-\frac{2s^2}{j^2\beta^2}+\cdots\right]-s-\frac{j\beta}{2}\\&\qquad+\frac{1}{2}\log 2\pi+O\left(\frac{2}{j\beta}\right).\end{aligned}$$

We evaluate both of these sums separately; the same idea applies to either sum. Before we can formally evaluate this first sum, we first need to consider the second term. Our aim is to compute refined large deviations results up to the $o(s^2)$ term[14]. It can be seen that terms of the form $j^{n+1}\beta^{n+1}/s^n$ (after evaluating) give terms of order s^2, while others are absorbed by the $o(s^2)$ term.

Looking at these terms, we have the following expression:

$$\begin{aligned}\left(\frac{1}{2}+s+\frac{j\beta}{2}\right)\left[\frac{j\beta}{2s}-\frac{j^2\beta^2}{8s^2}+\cdots\right]&=\frac{j\beta}{2}+\frac{j^2\beta^2}{4s}-\frac{j^2\beta^2}{8s}-\frac{j^3\beta^3}{16s^2}+\frac{j^3\beta^3}{24s^2}+\cdots\\&=\frac{j\beta}{2}+\frac{j^2\beta^2}{8s}-\frac{j^3\beta^3}{48s^2}+\cdots\\&=\frac{j\beta}{2}+s\sum_{n=2}^{\infty}\frac{(-1)^n}{n(n-1)}\left(\frac{j\beta}{2s}\right)^n.\end{aligned}$$

Here we have two sums: the sum over n and the sum over j. The sum over j is finite and so converges absolutely; we need only check that the sum over n here converges absolutely, and this can be seen by comparison with $\zeta(2)$ (here the sum is for integer

[14] While it may be possible to go further with the techniques on display here it would require a great deal of work which we do not go into in this thesis (we will later see that our results for the maximum of the product over primes P_X do not extend beyond the subleading term, so while we could obtain further terms here we do not have additional terms in P_X to compare against and verify). We hence restrict our results to terms of order exceeding s^2.

j in $[1,[2s/\beta]]$ and so $j < 2s/\beta$, i.e. $j\beta/2s < 1$). We therefore have

$$\begin{aligned}
&\sum_{j=1}^{[2s/\beta]} \frac{j\beta}{2} + s\sum_{n=2}^{\infty} \frac{(-1)^n}{n(n-1)}\left(\frac{j\beta}{2s}\right)^n \\
=&\frac{[2s/\beta]([2s/\beta]+1)}{4}\beta + s\sum_{n=2}^{\infty} \frac{(-1)^n}{n(n-1)}\left(\frac{\beta}{2s}\right)^n \sum_{j=1}^{[2s/\beta]} j^n \\
=&\frac{[2s/\beta]([2s/\beta]+1)}{4}\beta + s\sum_{n=2}^{\infty} \frac{(-1)^n}{n(n-1)}\left(\frac{\beta}{2s}\right)^n \left[\frac{[2s/\beta]^{n+1}}{(n+1)} + \frac{[2s/\beta]^n}{2} + \cdots\right] \\
=&\frac{[2s/\beta]([2s/\beta]+1)}{4}\beta + \frac{2s^2}{\beta}\sum_{n=2}^{\infty} \frac{(-1)^n}{(n-1)n(n+1)} + o(s^2),
\end{aligned}$$

where we make use of Faulhaber's formula in the third line. Therefore the first of these two sums evaluates to give

$$\begin{aligned}
&\frac{[2s/\beta]}{2}\log s + [2s/\beta]s\log s + \frac{[2s/\beta]([2s/\beta]+1)}{4}\beta\log s + \frac{[2s/\beta]([2s/\beta]+1)}{4}\beta \\
+&\frac{2s^2}{\beta}\sum_{n=2}^{\infty} \frac{(-1)^n}{(n-1)n(n+1)} + o(s^2) - [2s/\beta]s - \frac{[2s/\beta]([2s/\beta]+1)}{4}\beta \\
=&[2s/\beta]s\log s + \frac{[2s/\beta]^2}{4}\beta\log s + \frac{2s^2}{\beta}\sum_{n=2}^{\infty} \frac{(-1)^n}{(n-1)n(n+1)} - [2s/\beta]s + o(s^2).
\end{aligned}$$

We now look to evaluate the second of our sums. As before, we need to consider the second term; here the terms of interest are those of the form $s^{n+1}/j^n\beta^n$.

Focusing on these specific terms, we have the following:

$$\begin{aligned}
\left(\frac{1}{2} + s + \frac{j\beta}{2}\right)\left[\frac{2s}{j\beta} - \frac{2s^2}{j^2\beta^2} + \cdots\right] &= s + \frac{2s^2}{j\beta} - \frac{s^2}{j\beta} - \frac{2s^3}{j^2\beta^2} + \frac{4s^3}{3j^2\beta^2} + \cdots \\
&= s + \frac{s^2}{j\beta} - \frac{4s^3}{6j^2\beta^2} + \frac{8s^4}{12j^3\beta^3} + \cdots \\
&= s + \frac{s^2}{j\beta} + s\sum_{n=2}^{\infty} \frac{(-1)^{n+1}}{n(n+1)}\left(\frac{2s}{j\beta}\right)^n.
\end{aligned}$$

As before we are summing over both j and n here, but since both sums converge absolutely by the same argument as before, we can freely interchange summands. This gives the following:

$$
\begin{aligned}
&\sum_{j=[2s/\beta]+1}^{N-1} s + \frac{s^2}{j\beta} + s\sum_{n=2}^{\infty} \frac{(-1)^{n+1}}{n(n+1)} \left(\frac{2s}{j\beta}\right)^n \\
=&s(N-1) - [2s/\beta]s + \frac{s^2}{\beta}\log(N-1) - \frac{s^2}{\beta}\log([2s/\beta]) + s\sum_{n=2}^{\infty} \frac{(-1)^{n+1}}{n(n+1)} \left(\frac{2s}{\beta}\right)^n \sum_{j=[2s/\beta]}^{N-1} \frac{1}{j^n} \\
=&s(N-1) - [2s/\beta]s + \frac{s^2}{\beta}\log(N-1) - \frac{s^2}{\beta}\log([2s/\beta]) \\
&\quad + s\sum_{n=2}^{\infty} \frac{(-1)^{n+1}}{n(n+1)} \left(\frac{2s}{\beta}\right)^n \left[\frac{1}{(n-1)[2s/\beta]^{n-1}} - \frac{1}{(n-1)(N-1)^{n-1}}\right] \\
=&s(N-1) - [2s/\beta]s + \frac{s^2}{\beta}\log(N-1) - \frac{s^2}{\beta}\log([2s/\beta]) + \frac{2s^2}{\beta}\sum_{n=2}^{\infty} \frac{(-1)^{n+1}}{(n-1)n(n+1)} + o(s^2).
\end{aligned}
$$

Here the second term in the sum over n is absorbed by the o-term. Therefore the second sum here gives the result

$$
\begin{aligned}
&\frac{1}{2}\log((N-1)!) + \frac{N-1}{2}\log\beta - \frac{N-1}{2}\log 2 + s\log((N-1)!) - s\log([2s/\beta]!) \\
&+(N-1)s\log\beta - [2s/\beta]s\log\beta - (N-1)s\log 2 + [2s/\beta]s\log 2 + \sum_{[2s/\beta]+1}^{N-1} \frac{j\beta}{2}\log j \\
&+\frac{N(N-1)}{4}\beta\log\beta - \frac{[2s/\beta]([2s/\beta]+1)}{4}\beta\log\beta - \frac{N(N-1)}{4}\beta\log 2 \\
&+\frac{[2s/\beta]([2s/\beta]+1)}{4}\beta\log 2 + s(N-1) - [2s/\beta]s + \frac{s^2}{\beta}\log(N-1) - \frac{s^2}{\beta}\log([2s/\beta]) \\
&+\frac{2s^2}{\beta}\sum_{n=2}^{\infty} \frac{(-1)^{n+1}}{(n-1)n(n+1)} + o(s^2) - s(N-1) + [2s/\beta]s - \frac{N(N-1)}{4}\beta \\
&+\frac{[2s/\beta]([2s/\beta]+1)}{4}\beta \\
=&\frac{1}{2}\log((N-1)!) + \frac{N}{2}\log\beta - \frac{N}{2}\log 2 + s\log((N-1)!) - s\log([2s/\beta]!) + Ns\log\beta \\
&-[2s/\beta]s\log\beta - Ns\log 2 + [2s/\beta]s\log 2 + \sum_{[2s/\beta]+1}^{N-1} \frac{j\beta}{2}\log j + \frac{N(N-1)}{4}\beta\log\beta \\
&-\frac{[2s/\beta]^2}{4}\beta\log\beta - \frac{N(N-1)}{4}\beta\log 2 + \frac{[2s/\beta]^2}{4}\beta\log 2 + \frac{s^2}{\beta}\log N - \frac{s^2}{\beta}\log([2s/\beta]) \\
&+\frac{2s^2}{\beta}\sum_{n=2}^{\infty} \frac{(-1)^{n+1}}{(n-1)n(n+1)} + o(s^2) - \frac{N(N-1)}{4}\beta + \frac{[2s/\beta]^2}{4}\beta.
\end{aligned}
$$

3.3 Third Part of Sum

For the third part of the sum, we have (accounting for the factor of -2)

$$\sum_{j=1}^{N-1}(-1-s-j\beta)\log\left(\frac{s}{2}+\frac{j\beta}{2}\right)+s+j\beta-\log 2\pi+O\left(\frac{2}{s+j\beta}\right).$$

As before, we need to account for the size of j here and for what values $j\beta/2$ is smaller than $s/2$, and in the third part of the sum this change occurs at $[s/\beta]$. So we write this as

$$\sum_{j=1}^{[s/\beta]}(-1-s-j\beta)\log\left(\frac{s}{2}\right)+(-1-s-j\beta)\left[\frac{j\beta}{s}-\frac{j^2\beta^2}{2s^2}+\cdots\right]+s+j\beta-\log 2\pi$$
$$+O\left(\frac{2}{s}\right)+\sum_{j=[s/\beta]+1}^{N-1}(-1-s-j\beta)\log\left(\frac{j\beta}{2}\right)+(-1-s-j\beta)\left[\frac{s}{j\beta}-\frac{s^2}{2j^2\beta^2}+\cdots\right]$$
$$+s+j\beta-\log 2\pi+O\left(\frac{2}{j\beta}\right).$$

As before, in the first of these sums we have to once again consider the second term; those terms of interest are of the form $j^{n+1}\beta^{n+1}/s^n$. Here we have

$$\begin{aligned}(-1-s-j\beta)\left[\frac{j\beta}{s}-\frac{j^2\beta^2}{2s^2}+\cdots\right]&=-j\beta-\frac{j^2\beta^2}{s}+\frac{j^2\beta^2}{2s}+\frac{j^3\beta^3}{2s^2}-\frac{j^3\beta^3}{3s^2}+\cdots\\&=-j\beta-\frac{j^2\beta^2}{2s}+\frac{j^3\beta^3}{6s^2}+\cdots\\&=-j\beta+s\sum_{n=2}^{\infty}\frac{(-1)^{n-1}}{n(n-1)}\left(\frac{j\beta}{s}\right)^n.\end{aligned}$$

As before both sums over j and n converge absolutely, so we can interchange summands in this step, giving the result

$$\begin{aligned}&\sum_{j=1}^{[s/\beta]}-j\beta+s\sum_{n=2}^{\infty}\frac{(-1)^{n-1}}{n(n-1)}\left(\frac{j\beta}{s}\right)^n\\&=-\frac{[s/\beta]([s/\beta]+1)}{2}\beta+s\sum_{n=2}^{\infty}\frac{(-1)^{n-1}}{n(n-1)}\left(\frac{\beta}{s}\right)^n\sum_{j=1}^{[s/\beta]}j^n\\&=-\frac{[s/\beta]([s/\beta]+1)}{2}\beta+s\sum_{n=2}^{\infty}\frac{(-1)^{n-1}}{n(n-1)}\left(\frac{\beta}{s}\right)^n\left[\frac{[s/\beta]^{n+1}}{n+1}+\frac{[s/\beta]^n}{2}+\cdots\right]\\&=-\frac{[s/\beta]([s/\beta]+1)}{2}\beta+\frac{s^2}{\beta}\sum_{n=2}^{\infty}\frac{(-1)^{n-1}}{(n-1)n(n+1)}+o(s^2).\end{aligned}$$

The first sum then evaluates to give

$$-[s/\beta]s\log s+[s/\beta]s\log 2-\frac{[s/\beta]^2}{2}\beta\log s+\frac{[s/\beta]^2}{2}\beta\log 2+\frac{s^2}{\beta}\sum_{n=2}^{\infty}\frac{(-1)^{n-1}}{(n-1)n(n+1)}$$
$$+[s/\beta]s+O(1).$$

We repeat the same procedure for the second sum: for this expression, isolating the $s^n/j^{n-1}\beta^{n-1}$ terms gives

$$\begin{aligned}(-1-s-j\beta)\left[\frac{s}{j\beta}-\frac{s^2}{2j^2\beta^2}+\frac{s^3}{3j^3\beta^3}+\cdots\right]&=-s-\frac{s^2}{j\beta}+\frac{s^2}{2j\beta}+\frac{s^3}{2j^2\beta^2}-\frac{s^3}{3j^2\beta^2}+\cdots\\&=-s-\frac{s^2}{2j\beta}+\frac{s^3}{6j^2\beta^2}-\frac{s^4}{12j^4\beta^4}+\cdots\\&=-s-\frac{s^2}{2j\beta}+s\sum_{n=2}^{\infty}\frac{(-1)^n}{n(n+1)}\left(\frac{s}{j\beta}\right)^n.\end{aligned}$$

As before we are summing over both n and j, and since both sums converge absolutely we can interchange these sums at will, which gives

$$\begin{aligned}&\sum_{j=[s/\beta]+1}^{N-1}-s-\frac{s^2}{2j\beta}+s\sum_{n=2}^{\infty}\frac{(-1)^n}{n(n+1)}\left(\frac{s}{j\beta}\right)^n\\&=-(N-1)s+[s/\beta]s-\frac{s^2}{2\beta}\log(N-1)+\frac{s^2}{2\beta}\log([s/\beta])+s\sum_{n=2}^{\infty}\frac{(-1)^n}{n(n+1)}\left(\frac{s}{\beta}\right)^n\sum_{j=[s/\beta]+1}^{N-1}j^n\\&=-(N-1)s+[s/\beta]s-\frac{s^2}{2\beta}\log(N-1)+\frac{s^2}{2\beta}\log([s/\beta])\\&\quad+s\sum_{n=2}^{\infty}\frac{(-1)^n}{n(n+1)}\left(\frac{s}{\beta}\right)^n\left[\frac{1}{(n-1)[s/\beta]^{n-1}}-\frac{1}{(n-1)(N-1)^{n-1}}+\cdots\right]\\&=-(N-1)s+[s/\beta]s-\frac{s^2}{2\beta}\log(N-1)+\frac{s^2}{2\beta}\log([s/\beta])+\frac{s^2}{\beta}\sum_{n=2}^{\infty}\frac{(-1)^n}{(n-1)n(n+1)}+o(s^2).\end{aligned}$$

The second sum therefore evaluates to

$$-\log((N-1)!)-N\log\beta+N\log 2-s\log((N-1)!)+s\log([s/\beta]!)-Ns\log\beta$$
$$+[s/\beta]s\log\beta+Ns\log 2-[s/\beta]s\log 2-\sum_{j=[s/\beta]+1}^{N-1}j\beta\log j-\frac{N(N-1)}{2}\beta\log\beta$$
$$+\frac{[s/\beta]^2}{2}\beta\log\beta+\frac{N(N-1)}{2}\beta\log 2-\frac{[s/\beta]^2}{2}\beta\log 2-\frac{s^2}{2\beta}\log(N-1)+\frac{s^2}{2\beta}\log([s/\beta])$$
$$+\frac{s^2}{\beta}\sum_{n=2}^{\infty}\frac{(-1)^n}{(n-1)n(n+1)}+o(s^2)+\frac{N(N-1)}{2}\beta-\frac{[s/\beta]^2}{2}\beta.$$

3.4 Combining Cases

Now that we have the results for each individual case, we combine these and simplify where necessary; with some rewriting as well as the use of Stirling's formula once again, we arrive at the following:

$$\log M_N(\beta, s) = \sum_{j=1}^{[s/\beta]} j\beta \log j - \sum_{j=1}^{[2s/\beta]} \frac{j\beta}{2} \log j + \frac{s^2}{2\beta} \log N + \frac{s^2}{2\beta} + o(s^2).$$

Finally, we apply Euler-Maclaurin summation to the two sums which gives

$$\log M_N(\beta, s) = \frac{s^2}{2\beta} \log N - \frac{s^2}{2\beta} \log s - \frac{s^2}{\beta} \log 2 + \frac{s^2}{2\beta} \log \beta + \frac{3s^2}{4\beta} + o(s^2).$$

which is the result in Theorem 3.1.

$\square$

We've obtained here a result for the logarithmic expectation, but with this result we can go further; we would like a result for the maximum of the characteristic polynomial of a random CβE matrix, and for this we utilise a result from [21]: in this paper, Farmer, Gonek and Hughes used Large Deviations Theory to prove in the Appendix a statement regarding the tails of the distribution of $\max_\theta |\Lambda_U(\theta)|$. This is the result in Lemma 1.1, which we recall here: If $\delta > 0$ is fixed and $\delta \leq \lambda \leq 1 - \delta$, then

$$\mathbb{P}\left\{\max_\theta |\Lambda_U(\theta)| \geq \exp(N^\lambda)\right\} = \exp\left(-\frac{N^{2\lambda}}{(1-\lambda)\log N}(1 + o(1))\right).$$

In the proof of this statement, the authors derive the following expression for the maximum of the characteristic polynomial:

$$M_N(2k) \leq \mathbb{E}\left[e^{2k \log \max_\theta |\Lambda_U(\theta)|}\right] \leq \pi(2k+1) N M_N(2k),$$

Note that this result is for the characteristic polynomial of a CUE ($\beta = 2$) matrix, however a clear inspection of the steps involved apply to a general characteristic polynomial, and thus apply to the characteristic polynomial of a CβE matrix. The CβE analogue reads as

$$M_N(\beta, s) \leq \mathbb{E}\left[e^{s \log \max_\theta |\Lambda_{U,\beta}(\theta)|}\right] \leq \pi(s+1) N M_N(\beta, s). \tag{41}$$

Taking logarithms of this expression, we see that the left and right-hand sides of this set of inequalities agree up to order $\log N$ (as our focus is $s \ll N$).

If we choose $s = \lambda B(N)/A(N)$ (where $A(N)$ and $B(N)$ are to be determined), our aim is to ensure a finite limit as we scale by $B(N)$ and take the limit as N tends to infinity. Using the result in Theorem 3.1 this now reads

$$\log \mathbb{E}\left[e^{s \log \max_\theta |\Lambda_{U,\beta}(\theta)|}\right] = \frac{\lambda^2 B(N)^2}{2\beta A(N)^2} \log N - \frac{\lambda^2 B(N)^2}{2\beta A(N)^2} \log\left(\frac{\lambda B(N)}{A(N)}\right) - \frac{\lambda^2 B(N)^2}{\beta A(N)^2} \log 2$$
$$+ \frac{\lambda^2 B(N)^2}{2\beta A(N)^2} \log \beta + \frac{3\lambda^2 B(N)^2}{4\beta A(N)^2} + o\left(\frac{B(N)^2}{A(N)^2}\right).$$

By choosing $A(N) = N^\mu$ with $\delta < \lambda < 1 - \delta$ and scaling by $1/N$, we find that in order for the limit as N tends to infinity to be finite, we must have $B(N) = N^{2\mu}/(1-\mu)\log N$. This gives the result

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[e^{s \log \max_\theta |\Lambda_{U,\beta}(\theta)|}\right] = \frac{\lambda^2}{2\beta}.$$

By comparing with Assumption 2.1, we conclude that the logarithm of (41) satisfies a large deviations principle with rate function $\Lambda(\lambda) = \frac{\lambda^2}{2\beta}$.

As we've obtained a result for $\log M_N(\beta, s)$ with a subleading term of $o(s^2)$, the choice of s ensures we have an exact equality for the expectation in (41).

Large deviations theory tells us that

$$\Lambda^*(x) = \sup_\lambda \left(\lambda x - \frac{\lambda^2}{2\beta}\right) = \frac{1}{2}\beta x^2$$

and this is attained at $\bar{\lambda} = \beta x$. Looking at the standard result from large deviations (37) gives

$$\mathbb{P}\left(\max_\theta |\Lambda_{U,\beta}(\theta)| \geq \exp(xN^\mu)\right) = \exp\left(-\frac{\beta x^2 N^{2\mu}}{2(1-\mu)\log N}(1+o(1))\right).$$

This now allows us to find the maximum over the circular beta ensemble, and the initial steps here follow those seen in [21].

Theorem 3.2. *Fix $\eta > 0$. Let $M = \exp(N^\delta)$, with $\eta < \delta < 2 - \eta$, and set*

$$K_\epsilon(N) = \exp\left(\left(\sqrt{\frac{2}{\beta}\left(1 - \frac{1}{2}\delta\right)} + \epsilon\right)\sqrt{\log M}\sqrt{\log N}\right).$$

If $U_1, \ldots, U_M$ are chosen independently from the $C\beta E$, then as $N \to \infty$,

$$\mathbb{P}\left(\max_{1\leq j\leq M} \max_\theta |\Lambda_{U,\beta}| \leq K_\epsilon(N)\right) \to 1$$

for all $\epsilon > 0$ and for no $\epsilon < 0$. Here, s is chosen as above.

Proof of Theorem 3.2. We have chosen the $U_1, \ldots, U_M$ independently. Therefore we can rewrite the probability to give

$$
\begin{aligned}
&\mathbb{P}\left(\max_{1\le j\le M}\max_{\theta}|\Lambda_{U,\beta}| \le K_\epsilon(N)\right)\\
&\quad = \mathbb{P}\left(\max_{\theta}|\Lambda_{U,\beta}| \le K_\epsilon(N)\right)^M .
\end{aligned}
$$

If one were to take logarithms of this expression, this probability having limit one would be equivalent to the limit as N tends to infinity of the logarithm

$$
\begin{aligned}
\log \mathbb{P}\left(\max_{\theta}|\Lambda_{U,\beta}| \le K_\epsilon(N)\right)^M &= M\log\left(1-\mathbb{P}\left(\max_{\theta}|\Lambda_{U,\beta}| > K_\epsilon(N)\right)\right)\\
&= -M\mathbb{P}\left(\max_{\theta}|\Lambda_{U,\beta}(\theta)| > K_\epsilon(N)\right) + \text{(Lower Order Terms)}
\end{aligned}
$$

being equal to 0. For this to have limit 0, it is sufficient for the leading-order term to approach 0 in the limit. We can use our result from Theorem 3.1 if we make a suitable change of variable here. If we take $\mu = \log\log K_\epsilon(N)/\log N$ and set $x = 1$[15], this gives (neglecting the negative sign[16])

$$
\begin{aligned}
0 = M\exp\Bigg(&-\frac{\beta\log^2 K_\epsilon(N)}{2(\log N - \log\log K_\epsilon(N))} + \frac{\beta\log^2 K_\epsilon(N)}{2(\log N - \log\log K_\epsilon(N))^2}\log\log N\\
&- \frac{\beta\log^2 K_\epsilon(N)}{2(\log N - \log\log K_\epsilon(N))^2}\left(2\log 2 - \frac{3}{2}\right) + \frac{\beta\log^2 K_\epsilon(N)}{2(\log N - \log\log K_\epsilon(N))^2}\log\left(1-\frac{\log\log K_\epsilon(N)}{\log N}\right)\\
&+ o\left(\frac{\log^2 K_\epsilon(N)}{(\log N - \log\log K_\epsilon(N))^2}\right)\Bigg).
\end{aligned}
$$

Balancing the leading-order term gives

$$
\begin{aligned}
M &= \exp\left(\frac{\beta\log^2 K_\epsilon(N)}{2(\log N - \log\log K_\epsilon(N))}\right)\\
\implies \log M &= \frac{\beta\log^2 K_\epsilon(N)}{2(\log N - \log\log K_\epsilon(N))}\\
\implies \log^2 K_\epsilon(N) &= \frac{2}{\beta}\log M(\log N - \log\log K_\epsilon(N))\\
\implies K_\epsilon(N) &= \exp\left(\sqrt{\frac{2}{\beta}\log M(\log N - \log\log K_\epsilon(N))}\right).
\end{aligned}
$$

Iterating this expression and recalling that $M = \exp(N^\delta)$ gives

[15] The justification for why we can do this is provided in the Appendix A.

[16] As this will have limit 0, neglecting the negative sign will not have an impact on the results.

$$\begin{aligned}
2\log\log K_\epsilon(N) &= \log(2/\beta) + \log\log M + \log(\log N - \log\log K_\epsilon(N)) \\
\implies \log\log K_\epsilon(N) &= \frac{1}{2}\log(2/\beta) + \frac{1}{2}\log\log M + \frac{1}{2}\log(\log N - \log\log K_\epsilon(N)) \\
&= \frac{1}{2}\log(2/\beta) + \frac{\delta}{2}\log N + \frac{1}{2}\log\left(\left(1-\frac{\delta}{2}\right)\log N - \frac{1}{2}\log\beta - \cdots\right) \\
&= \frac{\delta}{2}\log N + \frac{1}{2}\log\log N + \frac{1}{2}\log\left(\frac{2}{\beta}\left(1-\frac{\delta}{2}\right)\right) \\
&\quad + \frac{1}{2}\log\left(1 - \frac{\log\beta}{(2-\delta)\log N} - \cdots\right).
\end{aligned}$$

Substituting this expression into our expression for $K_\epsilon(N)$ gives

$$\begin{aligned}
K_\epsilon(N) &= \exp\left(\sqrt{\frac{2}{\beta}\log M(\log N - \log\log K_\epsilon(N))}\right) \\
&= \exp\left(\sqrt{\frac{2}{\beta}\left(1-\frac{\delta}{2}\right)\log M\log N + (\text{Lower Order Terms})}\right).
\end{aligned}$$

Note the necessity for $\eta < \delta < 2-\eta$ in order for the leading term to have a positive coefficient, which here is

$$\exp\left(\sqrt{\frac{2}{\beta}\left(1-\frac{\delta}{2}\right)}\sqrt{\log M}\sqrt{\log N}\right). \tag{42}$$

The result from Theorem 3.2 follows.

□

In fact, our refined large deviations results allow us to go further, and obtain more than simply a leading-order expression for the maximum of the characteristic polynomial of a CβE matrix:

Theorem 3.3. *Fix $\eta > 0$. Let $M = \exp(N^\delta)$, with $\eta < \delta < 2-\eta$ as in Theorem 3.1. If $U_1, \ldots, U_M$ are chosen independently from the CβE, then as the matrix size N tends to infinity,*

$$\mathbb{P}\left(\max_{1\le j\le M}\max_\theta M_N(\beta,s) \le K\right) \to 1,$$

where $M_N(\beta,s) = \mathbb{E}\left[|\Lambda_{U,\beta}(\theta)|^s\right]$ and

$$K = \exp\Bigg(\Bigg(\left(\frac{2}{\beta}\left(1-\frac{\delta}{2}\right)\log M \log N + \frac{1}{\beta}\log M \log\log N\right.$$
$$\left. - \frac{1}{\beta}\log M\left(4\log 2 - 3 + \log\left(\frac{2}{\beta}\right) - \log\left(1-\frac{\delta}{2}\right)\right) + o(\log M)\right)^{1/2}\Bigg)\Bigg).$$

Proof of Theorem 3.3. From the proof of Theorem 2.1 we have the expression

$$\log \mathbb{P}(X_N \geq x) = \log \mathbb{E}\left[e^{\bar{\lambda}B(N)X_N}\right] - B(N)\bar{\lambda}x - \tfrac{1}{2}\log B(N) + O(1).$$

Adapting this expression to account for the scaling factor $A(N)$, we then have

$$\begin{aligned}
\log \mathbb{P}\left(\log \max_\theta |\Lambda_{U,\beta}(\theta)| \geq A(N)x\right) &= \log \mathbb{E}\left[e^{\bar{\lambda}\frac{B(N)}{A(N)}\max_\theta\left|\Lambda_{U,\beta}(\theta)\right|}\right] - B(N)\bar{\lambda}x - \tfrac{1}{2}\log B(N) + O(1) \\
&= \frac{\bar{\lambda}^2 B(N)^2}{2\beta A(N)^2}\log N - \frac{\bar{\lambda}^2 B(N)^2}{2\beta A(N)^2}\log\left(\frac{\bar{\lambda}B(N)}{A(N)}\right) \\
&\quad - \frac{\bar{\lambda}^2 B(N)^2}{\beta A(N)^2}\log 2 + \frac{\bar{\lambda}^2 B(N)^2}{2\beta A(N)^2}\log\beta + \frac{3\bar{\lambda}^2 B(N)^2}{4\beta A(N)^2} \\
&\quad + o\left(\frac{B(N)^2}{A(N)^2}\right) - B(N)\bar{\lambda}x - \tfrac{1}{2}\log B(N) + O(1).
\end{aligned}$$

Using $B(N) = N^{2\mu}/(1-\mu)\log N$, $A(N) = N^\mu$ and $\bar{\lambda} = \beta x$ as before gives

$$\begin{aligned}
\log \mathbb{P}\left(\log \max_\theta |\Lambda_{U,\beta}(\theta)| \geq A(N)x\right) &= -\frac{\beta x^2 N^{2\mu}}{2(1-\mu)\log N} + \frac{\beta x^2 N^{2\mu}}{2(1-\mu)^2\log N}\log\log N \\
&\quad - \frac{\beta x^2 N^{2\mu}}{2(1-\mu)^2\log^2 N}\left(\log x + 2\log 2 - \frac{3}{2}\right) \\
&\quad + \frac{\beta x^2 N^{2\mu}}{2(1-\mu)^2\log^2 N}\log(1-\mu) + o\left(\frac{N^{2\mu}}{(1-\mu)^2\log^2 N}\right).
\end{aligned}$$

We want to find the maximum over the CβE, so we set $\mu = \log\log K/\log N$ and $x = 1$ to give the following[17]:

[17] We again refer to the Appendix A for why this choice of variables is justified.

$$\mathbb{P}\left(\max_\theta |\Lambda_{U,\beta}(\theta)| \geq K\right) = \exp\Bigg(-\frac{\beta \log^2 K}{2(\log N - \log\log K)} + \frac{\beta \log^2 K}{2(\log N - \log\log K)^2} \log\log N$$

$$- \frac{\beta \log^2 K}{2(\log N - \log\log K)^2}\left(2\log 2 - \frac{3}{2}\right)$$

$$+ \frac{\beta \log^2 K}{2(\log N - \log\log K)^2} \log\left(1 - \frac{\log\log K}{\log N}\right)$$

$$+ o\left(\frac{\log^2 K}{(\log N - \log\log K)^2}\right)\Bigg).$$

Again we choose $U_1, \ldots, U_M$ independently from the CβE, and this gives

$$\log \mathbb{P}\left(\max_\theta |\Lambda_{U,\beta}| \leq K\right)^M = -M\mathbb{P}\left(\max_\theta |\Lambda_{U,\beta}(\theta)| > K\right)$$

$$= M \exp\Bigg(-\frac{\beta \log^2 K}{2(\log N - \log\log K)} + \frac{\beta \log^2 K}{2(\log N - \log\log K)^2} \log\log N$$

$$- \frac{\beta \log^2 K}{2(\log N - \log\log K)^2}\left(2\log 2 - \frac{3}{2}\right)$$

$$+ \frac{\beta \log^2 K}{2(\log N - \log\log K)^2} \log\left(1 - \frac{\log\log K}{\log N}\right)$$

$$+ o\left(\frac{\log^2 K}{(\log N - \log\log K)^2}\right)\Bigg).$$

From this, balancing leading order terms gives the result in Theorem 3.2 (minus the epsilon term). We are able to go further due to the addition of more terms in our expression for the large deviations. Taking

$$K = \exp\left(\sqrt{\frac{2}{\beta}\left(1 - \frac{\delta}{2}\right)\log M \log N + \epsilon}\right),$$

this gives

$$\log^2 K = \frac{2}{\beta}\left(1 - \frac{\delta}{2}\right)\log M \log N + \epsilon,$$

$$\log N - \log\log K = \log N - \frac{1}{2}\log\left(\frac{2}{\beta}\left(1-\frac{\delta}{2}\right)\right) - \frac{1}{2}\log\log M - \frac{1}{2}\log\log N$$
$$- \frac{1}{2}\log\left(1 - \frac{\epsilon}{\frac{2}{\beta}(1-\frac{\delta}{2})\log M \log N}\right) + \ldots$$
$$= \left(1-\frac{\delta}{2}\right)\log N - \frac{1}{2}\log\log N - \frac{1}{2}\log\left(\frac{2}{\beta}\left(1-\frac{\delta}{2}\right)\right) + \ldots$$

and therefore

$$\frac{1}{\log N - \log\log K} = \frac{1}{(1-\frac{\delta}{2})\log N\left[1 - \frac{\log\log N}{(2-\delta)\log N} - \frac{\log\left(\frac{2}{\beta}\left(1-\frac{\delta}{2}\right)\right)}{(2-\delta)\log N} + \cdots\right]}$$
$$= \frac{2}{(2-\delta)\log N}\left[1 + \frac{\log\log N}{(2-\delta)\log N} + \frac{\log\left(\frac{2}{\beta}\left(1-\frac{\delta}{2}\right)\right)}{(2-\delta)\log N} + \frac{(\log\log N)^2}{(2-\delta)^2\log^2 N} + \cdots\right].$$

Plugging this all into our refined large deviations result gives the following:

$$1 = M\exp\Bigg(-\beta\frac{\left(\frac{2}{\beta}\left(1-\frac{\delta}{2}\right)\log M\log N + \epsilon\right)}{(2-\delta)\log N}\left[1 + \frac{\log\log N}{(2-\delta)\log N} + \cdots\right]$$
$$+ 2\beta\frac{\left(\frac{2}{\beta}(1-\frac{\delta}{2})\log M\log N + \epsilon\right)}{(2-\delta)^2\log^2 N} + O\left(-\exp\left(-\frac{\log M}{\log N}\right)\right)$$
$$= M\exp\Bigg(-\log M - \frac{\beta\epsilon}{(2-\delta)\log N} - \frac{\log M\log\log N}{(2-\delta)\log N} + \frac{2\log M\log\log N}{(2-\delta)\log N} + \cdots$$
$$+ O\left(\exp\left(-\frac{\log M}{\log N}\right)\right)\Bigg).$$

The M and the $\exp(-\log M)$ terms balance one another, so we balance the next two terms which gives

$$0 = -\frac{\beta\epsilon}{(2-\delta)\log N} + \frac{\log M\log\log N}{(2-\delta)\log N}$$
$$\implies \epsilon = \frac{1}{\beta}\log M\log\log N.$$

This gives us the second term in our result in Theorem 3.3.

We now look to find the next term in this expression so we set

$$K = \exp\left(\sqrt{\frac{2}{\beta}\left(1-\frac{\delta}{2}\right)\log M\log N + \frac{1}{\beta}\log M\log\log N + \epsilon}\right).$$

If we then plug this result into our expression (now factoring in the first three terms) and simplify we get the following:

$$1 = M\exp\left(-\frac{\beta\epsilon}{(2-\delta)\log N} - \frac{\log M\log(\frac{2}{\beta}(1-\frac{\delta}{2}))}{(2-\delta)\log N} - \frac{\log M(4\log 2 - 3)}{(2-\delta)\log N}\right.$$
$$\left. + \frac{2\log M}{(2-\delta)\log N}\log\left(1-\frac{\delta}{2}\right) + O\left(\frac{\log M}{\log N}\right)\right).$$

Balancing terms gives

$$\epsilon = -\frac{1}{\beta}\log M\left(4\log 2 - 3 - \log\left(1-\frac{\delta}{2}\right)\right).$$

We are unable to obtain more terms at this point as we have no terms, but with this final term we arrive at the result in Theorem 3.3.

□

3.4.1 Physical Interpretation

Looking at our result for the maximum for the CβE, K, it is reasonable to ask if this results makes sense in the context of the circular ensembles.

More precisely, should we expect that as we vary β from $\beta = 1$ (COE) to $\beta = 2$ (CUE) and $\beta = 4$ (CSE) the conjectured maximum decreases? We claim that this is indeed to be expected, and for this we implement the following ideas...

Fundamentally, the orthogonal matrices are those matrices U for which the inverse U^{-1} is the transpose of U, U^T, which we immediately recognise as the real-unitary matrices $\mathbb{RU}(N)$ (since $U^* = U^T$ for real U).

Recalling the definition of the circular orthogonal ensemble, this is the set of unitary matrices invariant under orthogonal transformations. From the above reasoning every orthogonal matrix is unitary; therefore the orthogonal group is a subgroup of the unitary group. As such, the set of unitary matrices invariant under unitary transformations is a subset of the set of unitary matrices invariant under orthogonal transformations,

$$\text{CUE} \subseteq \text{COE},$$

and we should thus expect that the maximum of the characteristic polynomial of a matrix from the circular orthogonal ensemble will be greater than or equal to the maximum of the characteristic polynomial of a matrix from the circular unitary ensemble.

Similarly, the unitary group is a subgroup of the symplectic group, and so applying the same reasoning as above we should expect that

$$\text{CSE} \subseteq \text{CUE} \subseteq \text{COE}.$$

Thus the maximum over the circular unitary ensemble is at least that of the maximum over the circular symplectic ensemble. So we conclude that our conjectured result for the maximum K is reasonable.

It is clear that the above conjecture is sensible in the context of random matrix theory, but how does one interpret this result in the context of mathematical physics?

We mentioned briefly before that the model for the CβE aligns with that of the two-dimensional Coulomb gas model with inverse temperature parameter $\beta\,(= 1/T)$. From the viewpoint of particle interactions, increasing the temperature leads to increased energy per particle, which leads to an increased likelihood of collisions occurring.

This increase in temperature also leads to increased repulsion between particles (as seen in the logarithmic interaction parameter, see Chapter 1.2.1) thereby resulting in a larger joint probability distribution function and by extension a larger maximum for K. This is because an increase in $P(\theta_1, \cdots, \theta_N)$ leads to an increase in the expectation (or MGF). This corresponds to our result for K since we have a factor of β^{-1} in our result and β is the inverse temperature parameter. Therefore if the temperature T approaches infinity, we expect that K should approach infinity also.

In the opposing regime (zero temperature, taking the limit $\beta \to \infty$), we expect no collisions to occur between particles in the gas. Therefore the position of the particles would be fixed and $P(\theta_1, \cdots, \theta_N)$ goes from continuous to discrete, with it being equal to 1 when the particle is found at $(\theta_1, \cdots, \theta_N)$ and 0 otherwise. The overall maximum ends up being 1 in this regime, aligning with our results. Therefore in both cases, our results make sense if we look at the physical interpretation.

4 Refined Large Deviations for the Riemann Zeta Function

4.1 The Characteristic Polynomial of a Random Unitary Matrix

In the previous Chapter we computed refined large deviations results for the general circular beta ensemble; here we provide an equivalent method of computing the large deviations results for the circular unitary ($\beta = 2$) ensemble; the steps here align with those of Farmer, Gonek and Hughes in their paper [21].

We recall once again the definition of the characteristic polynomial:

$$\Lambda_U(\theta) = \det\left(I - Ue^{-i\theta}\right) = \prod_{n=1}^{N}\left(1 - e^{i(\theta_n - \theta)}\right).$$

Since we are now looking at the CUE, here U denotes a random CUE matrix.

As was the case in Chapter 3 our aim here is to determine the rate of growth of $\log\max_\theta|\Lambda_U(\theta)|$. In the language of large deviations, with X_N denoting the distribution of $\log\max_\theta|\Lambda_U(\theta)|$, we are looking at the logarithmic moment generating function

$$\log\mathbb{E}\left[e^{\frac{\lambda B(N)}{A(N)}X_N}\right] = \log\mathbb{E}\left[\max_\theta|\Lambda_U(\theta)|^{\frac{\lambda B(N)}{A(N)}}\right]. \tag{43}$$

With this, we prove the following result:

Theorem 4.1. *Let $k = \frac{\lambda B(N)}{A(N)}$ and let X_N denote the distribution of $\log\max_\theta|\Lambda_U(\theta)|$, where $A(N)$ is some scaling factor which ensures that Assumption 2.1 holds; further, assume that $|\frac{k}{N}| < 1$. If $A(N) = N^\mu, B(N) = \frac{N^{2\mu}}{(1-\mu)\log N}$, where $\delta < \mu < 1-\delta$ ($\delta > 0$), the following holds:*

$$\begin{aligned}\log\mathbb{E}\left[e^{\frac{\lambda B(N)}{A(N)}X_N}\right] = {} & \frac{\lambda^2 B(N)^2}{4A(N)^2}\log N - \frac{\lambda^2 B(N)^2}{4A(N)^2}\log\left(\frac{\lambda B(N)}{A(N)}\right) + \frac{3\lambda^2 B(N)^2}{8A(N)^2}\\ & - \frac{\lambda^2 B(N)^2}{4A(N)^2}\log 2 + O\left(\frac{B(N)}{\log^2 N}\right).\end{aligned}$$

The context behind why this result is significant stems from Section 1.3.7: recalling the expression (31) in Theorem 1.9:

$$\zeta(s) = P_X(s)Z_X(s)\left(1 + O\left(\frac{X^{2-\sigma+K}}{(|t|\log X)^K}\right) + O(X^{-\sigma}\log X)\right),$$

we split the zeta function as a product of two terms, Z_X and P_X. As we vary the value of X we vary the influence of the primes and the zeros. For small X the number of primes less than or equal to X is diminished and so the dominant contribution comes from the zeros, namely the Z_X term in the expression.

With the dominant contribution coming from the zeros, the product over zeros Z_X can be effectively modelled via the characteristic polynomial of a random unitary matrix. Hence computing the maximum of the characteristic polynomial gives the maximum of the product over zeros Z_X, therefore when the prime contribution is diminished the maximum for the characteristic polynomial (and thus Z_X) also applies to the maximum of the zeta function.

Note that, as we should expect, the expression in Theorem 4.1 agrees with our result in Theorem 3.2 for $\log M_N(\beta, s)$ when we set $\beta = 2,\ s = \frac{\lambda B(N)}{A(N)}$. In addition to the refined large deviations result of Theorem 2.1, we establish the following:

Corollary 4.1. *We deduce the following expression for x in the tails of the distribution:*

$$\begin{aligned}\log \mathbb{P}\left(\log \max_{\theta} |\Lambda_U(\theta)| \geq N^{\mu} x\right) = &-\frac{x^2 N^{2\mu}}{(1-\mu)\log N} + \frac{x^2 N^{2\mu}}{(1-\mu)^2 \log^2 N} \log\log N \\ &+ \frac{3x^2 N^{2\mu}}{2(1-\mu)^2 \log^2 N} - \frac{2x^2 N^{2\mu}}{(1-\mu)^2 \log^2 N} \log 2 \\ &- \frac{x^2 N^{2\mu}}{(1-\mu)^2 \log^2 N} \log x + \frac{x^2 N^{2\mu}}{(1-\mu)^2 \log^2 N} \log(1-\mu) \\ &+ O\left(\frac{N^{2\mu}}{\log^3 N}\right).\end{aligned}$$

Before proving these results we first clarify again our above line of thinking: we have introduced a scaling term $A(N)$ in (43) because, given the distribution X_N we cannot be sure that Assumption 2.1 holds. What this scaling does is ensure that

$$\Lambda(\lambda) = \lim_{N\to\infty} \frac{1}{B(N)} \log \mathbb{E}\left[e^{\frac{\lambda B(N)}{A(N)} X_N}\right]$$

exists and is indeed finite, as in Chapter 3.4. With this clarification, we now prove the result in Theorem 4.1.

Proof of Theorem 4.1. To prove this we study the asymptotic behaviour of $\log \mathbb{E}\left[|\Lambda_U(\theta)|^{\frac{\lambda B(N)}{A(N)}}\right] = \log \mathbb{E}\left[e^{\frac{\lambda B(N)}{A(N)} \log|\Lambda_U(\theta)|}\right]$, for which Keating and Snaith [33] give the following: for $k > -1/2$,

$$M_N(2k) = \mathbb{E}\left[|\Lambda_U(\theta)|^{2k}\right] = \frac{G^2(k+1)}{G(2k+1)} \times \frac{G(N+1)G(N+2k+1)}{G^2(N+k+1)}. \tag{44}$$

Here G denotes the Barnes-G function. Our reason for introducing the Barnes-G function stems from the following bound first derived in [21]:

$$\log \mathbb{E}\left[|\Lambda_U(\theta)|^{2k}\right] \leq \log \mathbb{E}\left[e^{2k\log\max_\theta|\Lambda_U(\theta)|}\right] \leq \log \mathbb{E}\left[|\Lambda_U(\theta)|^{2k}\right] + \log N + \log(2k+1) + \log \pi. \tag{45}$$

The context behind this expression will be made clear later. Taking logarithms and setting $2k = \lambda B(N)/A(N)$ in (44) gives:

$$\begin{aligned}\log \mathbb{E}\left[e^{\frac{\lambda B(N)}{A(N)}\log|\Lambda_U(\theta)|}\right] &= 2\log G\left(\frac{\lambda B(N)}{2A(N)}+1\right) + \log G(N+1) + \log G\left(N + \frac{\lambda B(N)}{A(N)}+1\right)\\ &\quad - \log G\left(\frac{\lambda B(N)}{A(N)}+1\right) - 2\log G\left(N+\frac{\lambda B(N)}{2A(N)}+1\right);\end{aligned}$$

so knowledge of the asymptotic behaviour of $\log \mathbb{E}[e^{\frac{\lambda B(N)}{A(N)}\log|\Lambda_U(\theta)|}]$ requires knowledge of the asymptotics of the Barnes G-function. For large positive values of N we have the following result for small k [8]:

$$\begin{aligned}\log G(N+k+1) &= \frac{N+k}{2}\log 2\pi + \zeta'(-1) - \frac{3N^2}{4} - kN + \left(\frac{N^2}{2} - \frac{1}{12} + \frac{k^2}{2} + kN\right)\log N\\ &\quad + O\left(\frac{1}{N}\right).\end{aligned}$$

For large $k < N$, the resulting asymptotic expression is necessary (the derivation of which can be found in Appendix B):

$$\begin{aligned}\log G(N+k+1) &= \frac{N+k}{2}\log 2\pi + \zeta'(-1) - \frac{3N^2}{4} - kN + \left(\frac{N^2}{2} - \frac{1}{12} + \frac{k^2}{2} + kN\right)\log N\\ &\quad + O\left(\frac{k^3}{N}\right).\end{aligned}$$

The above formulae combined with (44) gives the following simplified expression:

$$\begin{aligned}\log \mathbb{E}\left[e^{\frac{\lambda B(N)}{A(N)}\log|\Lambda_U(\theta)|}\right] &= \frac{\lambda^2 B(N)^2}{4A(N)^2}\log N - \frac{\lambda^2 B(N)^2}{4A(N)^2}\log\left(\frac{\lambda B(N)}{A(N)}\right) + \frac{3\lambda^2 B(N)^2}{8A(N)^2}\\ &\quad - \frac{\lambda^2 B(N)^2}{4A(N)^2}\log 2 - \frac{1}{12}\log\left(\frac{\lambda B(N)}{A(N)}\right) + \frac{1}{6}\log 2 + \zeta'(-1)\\ &\quad + O\left(\max\left\{\frac{1}{N}, \frac{A(N)}{B(N)}, \frac{B(N)^3}{NA(N)^3}\right\}\right).\end{aligned}$$

We now choose an appropriate $A(N)$ such that

$$\lim_{N\to\infty}\frac{1}{B(N)}\log \mathbb{E}\left[e^{\frac{\lambda B(N)}{A(N)}\log\max_\theta|\Lambda_U(\theta)|}\right]$$

exists and is finite. Hughes, Keating and O'Connell [33] have shown that if $A(N) = N^{\mu}$ with $\delta < \mu < 1-\delta$ and

$$B(N) = \frac{N^{2\mu}}{(1-\mu)\log N},$$

then for $\lambda \geq 0$,

$$\lim_{N\to\infty} \frac{1}{B(N)} \log \mathbb{E}\left[e^{\frac{\lambda B(N)}{A(N)}\log\max_\theta |\Lambda_U(\theta)|}\right] = \frac{1}{4}\lambda^2, \tag{46}$$

and indeed this is the case with the above result for the logarithmic MGF; therefore Assumption 2.1 holds[18], with the implementation of a scaling factor $A(N) = N^{\mu}$.

With these results, we then arrive at the following:

$$\begin{aligned}\log \mathbb{E}\left[e^{\frac{\lambda B(N)}{A(N)}\log|\Lambda_U(\theta)|}\right] &= \frac{\lambda^2 B(N)^2}{4A(N)^2}\log N - \frac{\lambda^2 B(N)^2}{4A(N)^2}\log\left(\frac{\lambda B(N)}{A(N)}\right) + \frac{3\lambda^2 B(N)^2}{8A(N)^2}\\ &\quad - \frac{\lambda^2 B(N)^2}{4A(N)^2}\log 2 - \frac{1}{12}\log\left(\frac{\lambda B(N)}{A(N)}\right)\\ &\quad + O\left(\max\left\{\frac{1}{N}, \frac{A(N)}{B(N)}, \frac{B(N)^3}{NA(N)^3}\right\}\right).\end{aligned}$$

If we combine this with the result from Farmer, Gonek and Hughes' paper (45) this gives the following upper bound:

$$\begin{aligned}\log \mathbb{E}\left[e^{\frac{\lambda B(N)}{A(N)}\log\max_\theta|\Lambda_U(\theta)|}\right] &\leq \frac{\lambda^2 B(N)^2}{4A(N)^2}\log N - \frac{\lambda^2 B(N)^2}{4A(N)^2}\log\left(\frac{\lambda B(N)}{A(N)}\right) + \frac{3\lambda^2 B(N)^2}{8A(N)^2}\\ &\quad - \frac{\lambda^2 B(N)^2}{4A(N)^2}\log 2 - \frac{1}{12}\log\left(\frac{\lambda B(N)}{A(N)}\right) + \log N + \log\pi\\ &\quad + \log\left(\frac{\lambda B(N)}{A(N)} + 1\right) + O\left(\max\left\{\frac{1}{N}, \frac{A(N)}{B(N)}, \frac{B(N)^3}{NA(N)^3}\right\}\right)\\ &= \frac{\lambda^2 B(N)^2}{4A(N)^2}\log N - \frac{\lambda^2 B(N)^2}{4A(N)^2}\log\left(\frac{\lambda B(N)}{A(N)}\right) + \frac{3\lambda^2 B(N)^2}{8A(N)^2}\\ &\quad - \frac{\lambda^2 B(N)^2}{4A(N)^2}\log 2 + O\left(\frac{B(N)}{\log^2 N}\right).\end{aligned}$$

This proves Theorem 4.1 as both upper and lower bounds in (45) are of the form above. Note also that $B(N)/\log^2 N > B(N)^3/NA(N)^3$ for $\delta < \mu < 1-\delta$, and so our choice of O-term here is sensible.

□

From here the proof of Corollary 4.1 is straightforward.

[18]Recall from the introduction that the rate function is denoted $\Lambda(\lambda)$ and so our parameter here would normally be λ. We have renamed λ to μ in $A(N)$ and $B(N)$ here to avoid overuse of the variable λ and to avoid any confusion or ambiguity.

Proof of Corollary 4.1. Making use of Theorem 2.1 we have

$$\log \mathbb{P}(X_N \geq x) = \log \mathbb{E}\left[e^{\bar{\lambda}B(N)X_N}\right] - B(N)\bar{\lambda}x - \tfrac{1}{2}\log B(N) + o(1),$$

where $\bar{\lambda}$ satisfies the equation $\Lambda'(\bar{\lambda}) = x$.

Replacing X_N with $\log \max_\theta |\Lambda_U(\theta)|/A(N)$ and substituting the result from Theorem 4.1 gives a new expression for the large deviations:

$$\begin{aligned}\log \mathbb{P}\left(\log \max_\theta |\Lambda_U(\theta)| \geq A(N)x\right) &= \log \mathbb{E}\left[e^{\frac{\bar{\lambda}B(N)}{A(N)}\log\max_\theta|\Lambda_U(\theta)|}\right] - B(N)\bar{\lambda}x - \tfrac{1}{2}\log B(N) \\ &\quad + o(1) \\ &= \frac{\bar{\lambda}^2 B(N)^2}{4A(N)^2}\log N - \frac{\bar{\lambda}^2 B(N)^2}{4A(N)^2}\log\left(\frac{\bar{\lambda}B(N)}{A(N)}\right) \\ &\quad + \frac{3\bar{\lambda}^2 B(N)^2}{8A(N)^2} - \frac{\bar{\lambda}^2 B(N)^2}{4A(N)^2}\log 2 - B(N)\bar{\lambda}x \\ &\quad + O\left(\frac{B(N)}{\log^2 N}\right).\end{aligned}$$

Setting $A(N) = N^\mu, B(N) = \frac{N^{2\mu}}{(1-\mu)\log N}$ and separating the logarithmic term gives the following:

$$\begin{aligned}\log \mathbb{P}(\log \max_\theta |\Lambda_U(\theta)| \geq N^\mu x) &= \frac{N^{2\mu}}{(1-\mu)\log N}\left(\frac{\bar{\lambda}^2}{4} - \bar{\lambda}x\right) + \frac{\bar{\lambda}^2 N^{2\mu}}{4(1-\mu)^2\log^2 N}\log\log N \\ &\quad + \frac{\bar{\lambda}^2 N^{2\mu}}{(1-\mu)^2\log^2 N}\left(\frac{3}{8} + \frac{\log(1-\mu)}{4} - \frac{\log(\bar{\lambda})}{4} - \frac{\log 2}{4}\right) \\ &\quad + O\left(\frac{N^{2\mu}}{\log^3 N}\right).\end{aligned}$$

Recall that $\Lambda^*(x) = \sup_{\lambda\in\mathbb{R}}(\lambda x - \Lambda(\lambda))$; from the above expression,

$$\lim_{N\to\infty}\frac{1}{B(N)}\log \mathbb{P}\left(\log \max_\theta |\Lambda_U(\theta)| \geq A(N)x\right) = -\left(\bar{\lambda}x - \frac{\bar{\lambda}^2}{4}\right)$$

which is our expression for $\Lambda^*(x)$. Since we have an explicit expression for $\Lambda(\lambda)$ we need only solve the equation

$$\Lambda^*(x) = \sup_{\lambda\in\mathbb{R}}\left(\lambda x - \frac{\lambda^2}{4}\right) = \bar{\lambda}x - \frac{\bar{\lambda}^2}{4} \tag{47}$$

for λ, which solves to give $\Lambda^*(x) = x^2$ for $\bar{\lambda} = 2x$; thus

$$
\begin{aligned}
\log \mathbb{P}\left(\log \max_\theta |\Lambda_U(\theta)| \geq N^\mu x\right) = &-\frac{x^2 N^{2\mu}}{(1-\mu)\log N} + \frac{x^2 N^{2\mu}}{(1-\mu)^2 \log^2 N}\log\log N \\
&+ \frac{3x^2 N^{2\mu}}{2(1-\mu)^2 \log^2 N} - \frac{2x^2 N^{2\mu}}{(1-\mu)^2 \log^2 N}\log 2 \\
&- \frac{x^2 N^{2\mu}}{(1-\mu)^2 \log^2 N}\log x + \frac{x^2 N^{2\mu}}{(1-\mu)^2 \log^2 N}\log(1-\mu) \\
&+ O\left(\frac{N^{2\mu}}{\log^3 N}\right)
\end{aligned}
$$

which is the result in Corollary 4.1.

□

4.2 Refined Large Deviations Results for Z_X

With the results for the characteristic polynomial established, we next apply these results to compute refined large deviations results for Z_X. These results can then be applied to obtain a conjecture for the maximum of the zeta function.

Deriving these refined large deviations results for the zeta function comes by way of the same observation first made by Gonek, Hughes and Keating. We outline the ideas at play below.

So far we have computed refined large deviations results for the characteristic polynomial of a matrix taken from the CβE. For the applications to the zeta function, we restrict our attention here to the CUE ($\beta = 2$).

Before we can continue there is a small matter which we need to address: the large deviations result computed above is for the characteristic polynomial of a single unitary matrix U; we desire an equivalent result for a collection of CUE matrices. To be precise, choosing independently M random unitary matrices $U_1, \ldots, U_M$ we aim to compute the maximum $K_\epsilon(N)$ such that

$$
\mathbb{P}\left\{\max_{1\leq j\leq M} \max_\theta |\Lambda_{U_j}(\theta)| \leq K_\epsilon(N)\right\} \to 1
$$

as $N \to \infty$ for all $\epsilon > 0$ and no $\epsilon < 0$. By independence of the $U_1, \ldots, U_M$ and the Taylor expansion of the logarithm, we can write this as

$$
\begin{aligned}
&M \log\left(1 - \mathbb{P}\left\{\max_\theta |\Lambda_U(\theta)| > K_\epsilon(N)\right\}\right) \\
&= -M\mathbb{P}\left\{\max_\theta |\Lambda_U(\theta)| > K_\epsilon(N)\right\} + \text{(Lower order terms)},
\end{aligned}
$$

where we want this result to approach zero as the matrix dimension N gets large. Farmer, Gonek and Hughes applied this philosophy to Z_X, and we replicate this idea with our refined results for the circular ensembles.

In the publication of Gonek, Hughes and Keating, the key purpose of splitting the zeta function as a product over primes times a product over zeros was to compute the moments of the zeta function by computing separately the moments for Z_X and P_X. For this they conjectured that for $T \to \infty$ (and $X \to \infty$, where X is carefully chosen with respect to T) the moments of zeta split as a product of the moments of P_X and Z_X.

In their calculations for the moments of the product over zeros Z_X they conjectured the following:

Conjecture 4.1 (Gonek-Hughes-Keating)**.** *Suppose that X, $T \to \infty$ with $X = O((\log T)^{2-\epsilon})$. Then for any fixed $k > -1/2$, we have*

$$\frac{1}{T}\int_T^{2T} \left| Z_X\left(\frac{1}{2}+it\right)\right|^{2k} \mathrm{d}t \sim \frac{G^2(k+1)}{G(2k+1)}\left(\frac{\log T}{e^\gamma \log X}\right)^{k^2}.$$

If we compare this with the moments result for the characteristic polynomial (24), we see that there is a correspondence between these results if one makes the identification

$$N = [\log T/e^\gamma \log X]. \tag{48}$$

Proceeding with this line of reasoning, this suggests that if we make the above identification we can model the characteristic polynomial $\Lambda_U(\theta)$ by Z_X and apply our refined large deviations results to Z_X, and subsequently the zeta function; some care is needed however.

The largest value of $|\Lambda_U(\theta)|$ is 2^N, and this occurs when the matrix U is in a small neighbourhood of scalar multiples of the identity matrix. If we take N as in (48), then for $X = \exp(o(\log\log T))$ this violates the maximum as given in (28). Therefore the method utilised by the authors (which is utilised here) must do something different than simply take the maximum of $|\Lambda_U(\theta)|$ over all matrices $U \in U(N)$.

The range of values of t for which we can look at $Z_X(\frac{1}{2}+it)$ is $T/X^{e^\gamma} < t < T$, since the random matrix model allows us to model the zeta function at large height along the critical line with random matrices of large size N. With this choice of t, a slight rewriting gives

$$\frac{\log T}{e^\gamma \log X} - 1 < \frac{\log t}{e^\gamma \log X} < \frac{\log T}{e^\gamma \log X}$$

and so this range of t ensures that the random matrix model is a good fit for the zeta function. Naturally, we are looking for the maximum of Z_X for $t \in [0, T]$ rather than $t \in [T/X^{e^\gamma}, T]$, so we need to take some care here. However, if we let $X \to \infty$ the interval $[T/X^{e^\gamma}, T]$ should cover almost all of $[0, T]$, and so it should cover the maximum.

For now we are focused on the regime where the dominant contribution to the zeta function comes from the product over zeros Z_X. For this we require that X be small to ensure the number of primes in the product P_X is diminished. With this line of reasoning we consider the case where N is of size $\log T$. From our results for the zeta function we know that up to height T there are (neglecting constants) $T\log T$ zeros on the critical line.

From the correspondence between zeros of zeta and eigenvalues of random unitary matrices, we therefore want $T\log T$ eigenvalues. To ensure this is the case we take $M = T\log T/N = T$ matrices.

Making the change of variable $\mu = \log\log K/\log N$ with $N = \log T$ leads to the following results for Z_X:

$$
\begin{aligned}
T\cdot\mathbb{P}\left(\max_{t\in[0,T]}\left|Z_X\left(\frac{1}{2}+it\right)\right|\geq K\right) &= T\cdot\exp\Bigg(-\frac{\log^2 K}{\log\log T-\log\log K}\\
&+\frac{\log^2 K}{(\log\log T-\log\log K)^2}\log\log\log T+\frac{3\log^2 K}{2(\log\log T-\log\log K)^2}\\
&-\frac{2\log^2 K}{(\log\log T-\log\log K)^2}\log 2+O\left(\frac{\log^2 K}{(\log\log T)^3}\right)\Bigg).
\end{aligned}
$$

We now balance this expression, and we explain this notion by showing the principal step in the evaluation: our aim is to find the smallest possible K which satisfies

$$
1 = T\cdot\exp\left(-\frac{\log^2 K}{\log\log T-\log\log K}\right).
$$

The reason for this is as follows: from our earlier calculations we want the left-hand side of our above expression, $T\cdot\mathbb{P}(\max_{t\in[0,T]}|Z_X(\frac{1}{2}+it)|\geq K)$, to vanish in the large T limit; therefore the same must also be true of the right-hand side. K is a function of T and our goal is to determine the smallest value of K for which this expression vanishes, so we begin by equating leading order terms on the right-hand side of our expression. This is the same idea that was employed in Chapter 1.3.3.

If we choose $2 < X < (\log T)^A$, we recover the conjectured result

$$
\max_{t\in[0,T]}\left|Z_X\left(\frac{1}{2}+it\right)\right| = \exp\left((1+o(1))\sqrt{\frac{1}{2}\log T\log\log T}\right)
$$

given by Farmer, Gonek and Hughes; the lower bound on X comes from the hybrid Euler-Hadamard product as given in Theorem 1.7 which is conditional on $X\geq 2$. The upper limit comes from Conjecture 4.1, where $X < \log T^{2-\epsilon}$ and so $\log X < \log T^A$.

In their original paper, due to the limitations of only having the leading order behaviour they were unable to compute beyond the leading order behaviour for the maximum of the product Z_X.

We are now in a position to do so, since we have not only the leading order behaviour in the large devations results, but the subleading and sub-subleading behaviour. If we repeat the ideas above, taking into account the additional terms present in the large deviations results, we arrive at the following:

Conjecture 4.2. *If* $2 < X < (\log T)^A$, *then*

$$\max_{t\in[0,T]}\left|Z_X\left(\frac{1}{2}+it\right)\right| = \exp\left(\left(\frac{1}{2}\log T\log\log T + \frac{1}{2}\log T\log\log\log T - \frac{1}{2}(5\log 2 - 3)\log T + o(\log T)\right)^{1/2}\right).$$

We omit the steps in obtaining this result, as the steps are identical to those in the proof of Theorem 3.3. (Note: that this result aligns with that result for $\beta = 2$, $\delta = 1$.)

With this result we are able to determine that in the regime where Z_X is the dominant contribution, the contribution from P_X is negligible; for $X = O(\log T)$ (as is the case here),

$$\begin{aligned}\left|P_X\left(\frac{1}{2}+it\right)\right| &\leq \exp\left(\sum_{n\leq X}\frac{\Lambda(n)}{\sqrt{n}\log n}\right)\\ &= \exp\left(\sum_{p\leq X}\frac{1}{\sqrt{p}} + \sum_{p\leq X}\frac{1}{2p} + \cdots\right)\\ &= \exp\left(\frac{2\sqrt{X}}{\log X} + \sum_{p\leq X}\frac{1}{2p} + \cdots\right) = O\left(\exp\left(\frac{3\sqrt{X}}{\log X}\right)\right).\end{aligned}$$

Therefore the contribution from P_X does not affect our results up to subsubleading order. This leads to the following:

Consequence 4.1. *Provided our results for* Z_X *in this regime are suitable for modelling zeta, we have that*

$$\max_{t\in[0,T]}\left|\zeta\left(\frac{1}{2}+it\right)\right| = \exp\left(\left(\frac{1}{2}\log T\log\log T + \frac{1}{2}\log T\log\log\log T - \frac{1}{2}(5\log 2 - 3)\log T + o(\log T)\right)^{1/2}\right).$$

We perform some numerics to accompany these results and verify their validity, and these are provided in Chapter 7.

4.3 Refined Large Deviations Results for P_X

We now turn our attention towards P_X. For this, we discuss two approaches for computing refined large deviations results for P_X, one of which is based upon prior results and utilises Gaussian behaviour, while the other approach comes from adopting an alternative model for the prime contribution; we argue later that this is an effective model for P_X, whilst also discussing the shortcomings of this approach.

4.3.1 Gaussian Approach

The first approach towards computing refined large deviations results for P_X comes from looking at the following material of Granville and Soundararajan (as seen in [21]): by isolating the dominant behaviour in the primes, P_X can be expressed as the product of two exponential terms,

$$\begin{aligned} P_X\left(\frac{1}{2}+it\right) &= \exp\left(\sum_{p\le X}\frac{1}{p^{1/2+it}}\right)\times\exp(O(\log\log X)) \\ &= \exp\left(P_X^*\left(\frac{1}{2}+it\right)\right)\times\exp(O(\log\log X)). \end{aligned} \tag{49}$$

It follows immediately from (49) that we can compute refined large deviations results for P_X by computing refined large deviations results for P_X^*, provided of course our results exceed the O-term in the second exponential.

One potential approach might be to utilise the argument of Farmer, Gonek and Hughes, which stems from the following:

Theorem 4.2 (Farmer, Gonek, Hughes [21], Lemma 4.2). *Let $\{z_j\}$ be a sequence of independent random variables distributed uniformly on the unit circle and let $\{a_j\}$ be a sequence of bounded real numbers such that for all $n\ge 3$,*

$$\frac{1}{V_J^{\frac{n}{2}}}\sum_{1\le j\le J}a_j^n\to 0$$

as $J\to\infty$, where

$$V_J := \sum_{1\le j\le J}a_j^2.$$

Then as $J\to\infty$, the distribution of

$$Y_J := \mathrm{Re}\sum_{1\le j\le J}a_jz_j$$

tends to a Gaussian with mean 0 and variance $\frac{1}{2}V_J$.

This result suggests that it might be possible to obtain refined large deviations results for $|P_X(\frac{1}{2}+it)|$ by modelling the distribution as a Gaussian; for taking absolute values in (49) gives

$$\left|P_X\left(\frac{1}{2}+it\right)\right| = \exp\left(\mathrm{Re}\left(P_X^*\left(\frac{1}{2}+it\right)\right)\right) \times \exp(O(\log\log X)),$$

so applying Theorem 4.2 with $z_j = p_j^{-it}$ and $a_j = 1/\sqrt{p_j}$ we see that $V_J \sim \log\log X$. However if we try this approach, we find that the sub-leading contribution from the Gaussian is absorbed by the $O(\log\log X)$ term in the second exponential. Hence we cannot obtain more refined results from this method, and thus a different approach is required.

4.3.2 Alternative Method

With this in mind we try the following: we see from (49) that P_X is given by

$$\left|P_X\left(\frac{1}{2}+it\right)\right| = \exp\left(\mathrm{Re}\left(\sum_{p\le X}\frac{1}{p^{1/2+it}}\right)\right) \times \exp(O(\log\log X)).$$

For the leading term we consider the following model: let R_X denote the distribution of $\mathfrak{R}(\sum_{p\le X}\frac{z_p}{\sqrt{p}})$; here z_p are independent variables uniformly distributed on the unit circle.

The motivation behind this idea comes from the same intuition adopted by Arguin, Belius and Harper in [3]: for distinct primes p we have that $\log p$ are linearly independent, so by computing moments one can show that the finite-dimensional distributions of the process (p^{-it}, p primes) converges as $T\to\infty$ to a sequence of independent random variables uniformly distributed on the unit circle. Is this applicable for our purposes? We argue that this is indeed the case.

For the case where the dominant contribution to the zeta function comes from the prime term P_X, we take $X = \exp(\sqrt{\log T})$, and so as we let $T\to\infty$ we increase the prime contribution in R_X albeit at a sub-linear rate. We argue in Chapter 6 however that this is not an issue.

Considering then $T\log^c T$ (where $c>0$) copies of R_X (our reasoning for this will become clear later), the large deviations are given by

$$\mathbb{P}\left(\max_{j\in\{1,\ldots,T\log^c T\}} R_X^{(j)} \le \log K\right) = \mathbb{P}\left(R_X \le \log K\right)^{T\log^c T}$$

and we want this probability to tend to 1 in the limit as $T\to\infty$. This is equivalent to saying that we want

$$T \log^c T \cdot \log \left(1 - \mathbb{P} \left(\exp \left(\mathrm{Re} \sum_{p \le X} \frac{z_p}{\sqrt{p}} \right) > K \right) \right)$$

to tend to zero as $T \to \infty$.

For this we require an expression for the large deviations of R_X. For this we prove the following result:

Theorem 4.3. *Let* $k = \frac{\lambda B(N)}{A(N)}$ *and let* R_X *denote the distribution of* $\Re(\sum_{p \le X} \frac{z_p}{\sqrt{p}})$*, where* $A(N)$ *is some scaling factor which ensures that Assumption 2.1 holds. If* $A(N) = X^{\mu/2}$*,* $B(N) = X^{\mu}$*, where* $0 < \mu < 1$*. Then we have*

$$\log \mathbb{E}\left[e^{\frac{\lambda B(X)}{A(X)} R_X} \right] = \frac{\lambda^2 B(X)^2}{4A(X)^2} \log\log X - \frac{\lambda^2 B(X)^2}{4A(X)^2} \log\log \left(\frac{\lambda^2 B(X)^2}{A(X)^2} \right) + O\left(\frac{B(X)^2}{A(X)^2 \log(B(X)/A(X))} \right).$$

From this, we derive the following Corollary:

Corollary 4.2. *For* $0 < \mu < 1$ *we have*

$$\log \mathbb{P}\left((R_X \ge x X^{\mu/2} \log(1/\mu) \right) = -x^2 X^{\mu} \log(1/\mu) + O\left(\frac{X^{\mu}}{\mu \log X} \right).$$

Proof of Theorem 4.3. As we have done previously, writing everything in the language of large deviations gives

$$\begin{aligned} \log \mathbb{E}\left[e^{\frac{\lambda B(X)}{A(X)} R_X} \right] &= \log \mathbb{E}\left[e^{\frac{\lambda B(X)}{A(X)} \sum_{p \le X} \frac{\Re(z_p)}{\sqrt{p}}} \right] \\ &= \sum_{p \le X} \log \mathbb{E}\left[e^{\Re\left(\frac{\lambda B(X)}{A(X)\sqrt{p}} z_p \right)} \right], \end{aligned}$$

since the z_p are independent; as before we introduce a scaling term $A(X)$. Hence for $\lambda > 0$ we have

$$\log \mathbb{E}\left[e^{\frac{\lambda B(X)}{A(X)} R_X} \right] = \sum_{p \le X} \log I_0 \left(\frac{\lambda B(X)}{A(X)\sqrt{p}} \right).$$

Here, $I_0(\frac{\lambda B(X)}{A(X)\sqrt{p}}) = \mathbb{E}[e^{\Re(\frac{\lambda B(X)}{A(X)\sqrt{p}} z_p)}]$ denotes the modified Bessel function (see [27] for example); z_p is a random variable equidistributed on the unit circle[19]. Thus one can determine the refined large deviations of R_X (and thus of $P_X(\frac{1}{2} + it)$) by studying the behaviour of the modified Bessel function as $X \to \infty$.

[19]Note that we have used X and R_X here instead of N and X_N to avoid confusion with prior notation.

The following results for the large-argument asymptotics (for fixed ν) will be essential here:

$$I_\nu(z) \sim \frac{e^z}{(2\pi z)^{1/2}} \sum_{k=0}^{\infty} (-1)^k \frac{a_k(\nu)}{z^k}; \quad a_k(\nu) = \frac{(4\nu^2 - 1^2)(4\nu^2 - 3^2)\cdots(4\nu^2 - (2k-1)^2)}{k! 8^k}, \tag{50}$$

as $z \to \infty$. Here $a_0(0) = 1$.

This asymptotic result is valid provided $|\arg(z)| < \pi/2$. This is indeed the case for us as we are working in the case where our argument $\lambda B(X)/A(X)\sqrt{p}$ is real and positive, so has argument 0 in the complex plane. Further, the function $I_0(z)$ is real and positive here, so taking the logarithm of I_0 is well-defined.

We also require knowledge of the small-argument asymptotics (again for fixed ν):

$$I_\nu(z) \sim \left(\frac{z}{2}\right)^\nu \left[\frac{1}{\Gamma(\nu+1)} + \frac{(\frac{1}{4}z^2)}{\Gamma(\nu+2)}\right] \tag{51}$$

as $z \to 0$.

To formally compute the refined large deviations results for R_X we split into cases subject to the size of p, as well as the size of X; if for example $\sqrt{p}$ is large the argument $\lambda B(X)/A(X)\sqrt{p}$ will be small and different asymptotics must be considered. Taking logarithms and splitting into cases gives

$$\frac{1}{B(X)} \log \mathbb{E}\left[e^{\frac{\lambda B(X)}{A(X)} R_X}\right] = \sum_{\substack{p \le X \\ \sqrt{p} \ll \lambda B(X)/A(X)}} \frac{1}{B(X)} \log I_0\left(\frac{\lambda B(X)}{A(X)\sqrt{p}}\right) \tag{52}$$

$$+ \sum_{\substack{p \le X \\ \sqrt{p} \asymp \lambda B(X)/A(X)}} \frac{1}{B(X)} \log I_0\left(\frac{\lambda B(X)}{A(X)\sqrt{p}}\right) \tag{53}$$

$$+ \sum_{\substack{p \le X \\ \sqrt{p} \gg \lambda B(X)/A(X)}} \frac{1}{B(X)} \log I_0\left(\frac{\lambda B(X)}{A(X)\sqrt{p}}\right). \tag{54}$$

It should be clear above that in the first summation, the argument inside the modified Bessel function will be large and so large argument asymptotics will be necessary, whilst in the third summation the argument is small and there small-argument asymptotics will be essential here.

Having split this according to the size of p subject to X (resp. $\lambda B(X)/A(X)\sqrt{p}$) we must now consider the resulting summations subject to the size of X. This idea will become clear when we consider the following three cases.

4.3.3 Case $X \ll (\lambda B(X)/A(X))^2$

The summation in (52) gives, using the appropriate asymptotics:

$$\sum_{\substack{p\le X\\ \sqrt{p}\ll\lambda B(X)/A(X)}} \frac{1}{B(X)}\log I_0\left(\frac{\lambda B(X)}{A(X)\sqrt{p}}\right) \sim \sum_{\substack{p\le X\\ \sqrt{p}\ll\lambda B(X)/A(X)}} \frac{\lambda}{A(X)\sqrt{p}} \tag{55}$$
$$-\frac{1}{2B(X)}\log\left(\frac{2\pi\lambda B(X)}{A(X)\sqrt{p}}\right)+\frac{1}{B(X)}\log\left(\sum_{k=0}^{\infty}(-1)^k\frac{a_k(0)\sqrt{p}^k A(X)^k}{\lambda^k B(X)^k}\right)$$
.

Applying a strong form of the prime number theorem and isolating the leading order term inside the logarithm, we see that (55) can be written as

$$\sum_{\substack{p\le X\\ \sqrt{p}\ll\lambda B(X)/A(X)}} \frac{1}{B(X)}\log I_0\left(\frac{\lambda B(X)}{A(X)\sqrt{p}}\right) = \frac{\lambda}{A(X)}\frac{2\sqrt{X}}{\log X}-\frac{\log 2\pi\lambda}{2B(X)}\frac{X}{\log X}$$
$$-\frac{\log B(X)}{2B(X)}\frac{X}{\log X}+\frac{\log A(X)}{2B(X)}\frac{X}{\log X}+\frac{X}{4B(X)}+\frac{A(X)}{12\lambda B(X)^2}\frac{X^{3/2}}{\log X}$$
$$+O\left(\frac{A(X)^2X^2}{B(X)^3\log X}\right).$$

The O-term here comes from examining the subleading term inside the final logarithmic term in (55).

For $X \ll (\lambda B(X)/A(X))^2$, the second and third sums (52) and (53) both equal zero, since $X \ll (\lambda B(X)/A(X))^2$ implies there are no primes p such that $p \le X$ and $\sqrt{p} \gg \lambda B(X)/A(X)$.

4.3.4 Case $X \asymp (\lambda B(X)/A(X))^2$

Since the magnitude of X is at least that of the first case, the first sum (52) gives the same results as before.

In (53), the argument within the modified Bessel function is of constant order, thus when we divide through by $B(X)$ and take the limit as $X \to \infty$ we expect the limit here to be zero. Hence we write this term as $O(1/B(X))$ or $o(1)$.

The third sum (54) can be shown to equal zero by applying the same argument to (53) and (54) as seen in the first case.

4.3.5 Case $X \gg (\lambda B(X)/A(X))^2$

For the sum in (52), when using the same methods as above all terms in the expression are of the same order, so we simply write this as

$$\sum_{\substack{p \le X \\ \sqrt{p} \ll \lambda B(X)/A(X)}} \frac{1}{B(X)} \log\left(I_0 \left(\frac{\lambda B(X)}{A(X)\sqrt{p}} \right) \right) = O\left(\frac{B(X)}{A(X)^2 \log(B(X)/A(X))} \right).$$

As before, the sum in (53) is equal to $O(1/B(X))$.

For (54), taking the Taylor expansion of the logarithm and again utilising a strong form of the prime number theorem, as well as the asymptotic result in (51), gives

$$\sum_{p \le X} \frac{1}{B(X)} \log\left(1 + \frac{\lambda^2 B(X)^2}{4A(X)^2 p}\right) = \frac{\lambda^2 B(X)}{4A(X)^2} \log\log X - \frac{\lambda^2 B(X)}{4A(X)^2} \log\log\left(\frac{\lambda^2 B(X)^2}{A(X)^2}\right) + O\left(\frac{B(X)}{A(X)^2} e^{-c\sqrt{\log X}}\right).$$

We therefore conclude in this case that

$$\frac{1}{B(X)} \log \mathbb{E}\left[e^{\frac{\lambda B(X)}{A(X)} R_X}\right] = \frac{\lambda^2 B(X)}{4A(X)^2} \log\log X - \frac{\lambda^2 B(X)}{4A(X)^2} \log\log\left(\frac{\lambda^2 B(X)^2}{A(X)^2}\right) + O\left(\frac{B(X)}{A(X)^2 \log(B(X)/A(X))}\right).$$

We are working in the case $X \gg (\lambda B(X)/A(X))^2$; in other words, we have $(\lambda B(X)/A(X))^2 = \lambda^2 X^\mu$, $0 < \mu < 1$. With this in mind, we see to balance the leading order term in the above equation.

Taking $(\lambda B(X)/A(X))^2 = \lambda^2 X^\mu$ gives in the two leading terms

$$\begin{aligned}
&\frac{\lambda^2 B(X)}{4A(X)^2} \log\log X - \frac{\lambda^2 B(X)}{4A(X)^2} \log\log\left(\frac{\lambda^2 B(X)^2}{A(X)^2}\right) \\
=&\frac{\lambda^2 B(X)}{4A(X)^2} \log\log X - \frac{\lambda^2 B(X)}{4A(X)^2} \log\log(\lambda^2 X^\mu) \\
=&\frac{\lambda^2 B(X)}{4A(X)^2} \log\log X - \frac{\lambda^2 B(X)}{4A(X)^2} \log(\mu \log X) - \frac{\lambda^2 B(X)}{4A(X)^2} \log\left(1 + \frac{\log \lambda^2}{\mu \log X}\right) \\
=&\frac{\lambda^2 B(X)}{4A(X)^2} \log\log X - \frac{\lambda^2 B(X)}{4A(X)^2} \log \mu - \frac{\lambda^2 B(X)}{4A(X)^2} \log\log X - \frac{\lambda^2 B(X)}{4A(X)^2} \log\left(1 + \frac{\log \lambda^2}{\mu \log X}\right) \\
=&\frac{\lambda^2 B(X)}{4A(X)^2} \log(1/\mu) - \frac{\lambda^2 B(X)}{4A(X)^2} \log\left(1 + \frac{\log \lambda^2}{\mu \log X}\right).
\end{aligned}$$

So we seek to balance the leading term, which is simply a case of ensuring that $A(X)^2 = B(X)\log(1/\mu)$. We have two equations to solve:

$$
\begin{aligned}
(B(X)/A(X))^2 &= X^\mu \\
A(X)^2 &= B(X)\log(1/\mu).
\end{aligned}
$$

Solving these yields the solution $A(X) = X^{\mu/2}\log(1/\mu)$, $B(X) = X^\mu \log(1/\mu)$. With these choices of $A(X)$ and $B(X)$, the sequence of random variables R_X satisfies the large deviation principle with rate function $\Lambda(\lambda) = \frac{\lambda^2}{4}$.

Putting it all together, we have

$$
\begin{aligned}
\log \mathbb{E}\left[e^{\frac{\lambda B(X)}{A(X)}R_X}\right] = \frac{\lambda^2 B(X)^2}{4A(X)^2}\log\log X - \frac{\lambda^2 B(X)^2}{4A(X)^2}\log\log\left(\frac{\lambda^2 B(X)^2}{A(X)^2}\right) \\
+ O\left(\frac{B(X)^2}{A(X)^2\log(B(X)/A(X))}\right)
\end{aligned}
$$

and this is the result in Theorem 4.3. $\square$

The proof of Corollary 4.2 follows directly from this.

Proof of Corollary 4.2. Using results from large deviations theory, we have

$$
\begin{aligned}
\log \mathbb{P}(R_X \geq A(X)x) &= \log \mathbb{E}\left[e^{\frac{\lambda B(X)}{A(X)}R_X}\right] - B(X)\bar{\lambda}x - \frac{1}{2}\log B(X) + o(1) \\
&= \frac{\bar{\lambda}^2 B(X)^2}{4A(X)^2}\log\log X - \frac{\bar{\lambda}^2 B(X)^2}{4A(X)^2}\log\log\left(\frac{\bar{\lambda}^2 B(X)^2}{A(X)^2}\right) \\
&\quad + O\left(\frac{B(X)^2}{A(X)^2\log(B(X)/A(X))}\right) - B(X)\bar{\lambda}x - \frac{1}{2}\log B(X) + o(1).
\end{aligned}
$$

Here $\bar{\lambda}$ is the solution to the equation

$$
\begin{aligned}
-\Lambda^*(x) = \sup_\lambda(\lambda x - \Lambda(\lambda)) &= \sup_\lambda\left(\lambda x - \frac{\lambda^2}{4}\right) \\
&= \bar{\lambda} - \frac{\bar{\lambda}^2}{4}.
\end{aligned}
$$

As we see in Chapter 4.1 this gives the result $\bar{\lambda} = 2x$. Plugging this back into the above equation gives the results

$$
\begin{aligned}
\log \mathbb{P}\left(R_X \geq X^{\mu/2}\log(1/\mu)\right) = -x^2 X^\mu \log(1/\mu) - x^2 X^\mu \log\left(1 + \frac{2\log 2x}{\mu \log X}\right) + O\left(\frac{X^\mu}{\mu\log X}\right) \\
- \frac{\mu}{2}\log X - \frac{1}{2}\log\log(1/\mu) + o(1).
\end{aligned}
$$

Expanding the logarithmic term and simplifying gives the following:

$$\log \mathbb{P}\left(R_X \geq x X^{\mu/2}\log(1/\mu)\right) = -x^2 X^{\mu}\log(1/\mu) + O\left(\frac{X^{\mu}}{\mu \log X}\right).$$

and this is the result in Corollary 4.2.

□

These results are in agreement with those in [21] since the central limit regime gives a limiting Gaussian, hence we should expect a quadratic rate function. Further, as with the characteristic polynomial the above result is invariant under a change of variable, and the details of this are included in Appendix A.

If we consider the case where the main contribution comes from P_X, taking $X = \exp(\sqrt{\log T})$ and choosing $T\log^c T$ values of t, we want the expression

$$\left(1 - \mathbb{P}\left(\exp\left(\operatorname{Re}\sum_{p\leq \exp(\sqrt{\log T})}\frac{z_p}{\sqrt{p}}\right) > K\right)\right)^{T\log^c T}$$

to approach 1 as $T \to \infty$. Taking logarithms and making a change of variable in μ such that $X^{\mu/2}\log(1/\mu) \to \log K$ gives the following:

$$\begin{aligned}
X^{\mu/2}\log(1/\mu) = \log K \implies X^{\mu/2} &= \frac{\log K}{\log(1/\mu)} \\
\implies \frac{\mu}{2}\log X &= \log\log K - \log\log(1/\mu) \\
\implies \mu &= \frac{2\log\log K - 2\log\log(1/\mu)}{\log X} \\
&= \frac{2\log\log K}{\log X} - \frac{2\log\log\log X}{\log X} + \cdots .
\end{aligned}$$

With μ as above and $X = \exp(\sqrt{\log T})$, plugging these into our equation gives the following:

$$\begin{aligned}
&T\log^c T \cdot \mathbb{P}\left(\exp\left(\operatorname{Re}\sum_{p\leq \exp(\sqrt{\log T})}\frac{z_p}{\sqrt{p}}\right) > K\right) \\
&= T\log^c T \cdot \exp\left(-\frac{2\log^2 K}{\log\log T}\left[1 + \frac{2}{\log\log T}\log(2(\log\log K - \log\log(1/\mu)) + \cdots\right]\right. \\
&\qquad \left. + O\left(\frac{\log^2 K}{(\log\log T)^2}\right)\right).
\end{aligned}$$

We seek to balance leading order terms, i.e. to find K which satisfies

$$T\log^c T = \exp\left(\frac{2\log^2 K}{\log\log T}\right).$$

Balancing this expression gives the leading order solution

$$K = \max_{\theta\in[0,2\pi)} \exp(R_X) = \exp\left(\sqrt{\frac{1}{2}\log T\log\log T(1+o(1))}\right).$$

as T tends to infinity. However, with these refined results for the large deviations of R_X (and respectively P_X) we can compute more terms in this expression for the maximum, which leads to the following:

Theorem 4.4. *For $X = \exp(\sqrt{\log T})$, we have*

$$\max_{\theta\in[0,2\pi)} \exp(R_X) = \exp\left(\sqrt{\frac{1}{2}\log T\log\log T - \log T\log\log\log T + o(\log T\log\log\log T)}\right).$$

Proof of Theorem 4.4. The method is identical to that displayed in the proof of Theorem 3.3.

With the values of μ and X mentioned above,

$$\begin{aligned}\mu &= \frac{2\log\log K}{\log X} - \frac{2\log\log X}{\log X} + \cdots\\ X &= \exp(\sqrt{\log T}),\end{aligned}$$

we seek to balance the following expression:

$$\begin{aligned}1 &= T\log^c T\cdot\exp\left(-X^\mu\log(1/\mu) + O\left(\frac{X^\mu}{\mu\log X}\right)\right)\\ &= T\log^c T\cdot\exp\left(-\frac{2\log^2 K}{\log\log T}\left[1+\frac{2\log\log\log K}{\log\log T}+\cdots\right] + O\left(\frac{\log^2 K}{(\log\log T)^3}\right)\right).\end{aligned}$$

Balancing leading order terms gives the leading order result

$$K = \exp\left(\sqrt{\frac{1}{2}\log T\log\log T + \frac{c}{2}(\log\log T)^2}\right).$$

As in the proof of Theorem 3.3 we set

$$K = \exp\left(\sqrt{\frac{1}{2}\log T\log\log T + \frac{c}{2}(\log\log T)^2 + \epsilon}\right)$$

and plug this expression into our large deviations result, which gives

$$
\begin{aligned}
1 &= T\log^c T\cdot\exp\left(-\frac{2\log^2 K}{\log\log T}\left[1+\frac{2\log\log\log K}{\log\log T}\right]+O\left(\frac{\log^2 K}{(\log\log T)^3}\right)\right)\\
&= T\log^c T\cdot\exp\Bigg(-\log T-\frac{c}{2}\log\log T-\frac{2\epsilon}{\log\log T}-\frac{2\log T\log\log\log T}{\log\log T}\\
&\quad+O\left(\frac{\log T}{(\log\log T)^2}\right)\Bigg).
\end{aligned}
$$

Balancing leading order terms gives $\epsilon = -\log T\log\log\log T$, so we have the refined expression for K

$$
K=\exp\left(\sqrt{\frac{1}{2}\log T\log\log T-\log T\log\log\log T+o(\log T\log\log\log T)}\right)
$$

which is the result in Theorem 4.4.

Note that we are able to go further as we have not yet exceeded the the value of the O-term. However, additional terms will not be of much use to us in computing refined large deviations results for the Riemann zeta function as we require the same number of terms in our large deviations results for both P_X and Z_X.

□

In this regime where the dominant contribution comes from the product over primes P_X, the contribution from the product over zeros is given by

$$
O\left(\exp\left(\sqrt{\log T}\right)\right).
$$

Combining these results together leads to the following:

Conjecture 4.3. *Provided that R_X is a suitable model for P_X, we have*

$$
\max_{t\in[0,T]}\left|\zeta\left(\frac{1}{2}+it\right)\right| = \exp\left(\sqrt{\frac{1}{2}\log T\log\log T-\log T\log\log\log T+o(\log T\log\log\log T)}\right).
$$

With refined large deviations results computed for both Z_X and P_X (resp. R_X) we now proceed by applying these results directly to the zeta function.

4.4 Refined Large Deviations Results for Zeta

We now apply our results to the Riemann zeta function, and to do this we look back to the theory surrounding the hybrid Euler-Hadamard product. We discussed in Chapter 1.3.7 how in the regime where X is large, the prime contribution is increased and only the zeros away from s affect the product Z_X, so in this regime the prime contribution is the dominant contribution; on the other hand, for small X the prime contribution is diminished and the dominant contribution comes from Z_X.

We now look at the intermediate values where both Z_X and P_X contribute. Evidence that both Z_X and P_X contribute in this regime is given by Gonek, Hughes and Keating (see again [26]). We summarise the method utilised here, which again follow those in [21].

First and foremost, we consider $X = \exp(\log^{\alpha} T)$ with $0 < \alpha < 1/2$; here the value of α ensures that as we vary α we transition between the two previous regimes, and by considering now the intermediate regime we should be able to demonstrate that the results attained above are indeed consistent. Note that if we were to take $N = \log T/\log X$ (so that we can model Z_X by the characteristic polynomial $\Lambda_U(\theta)$) and $M = T \log X$ (so that we sample enough characteristic polynomials to cover the critical line up to height T), the analysis of the previous Sections tells us that at leading order, $|Z_X(\frac{1}{2} + it)|$ gets as large as

$$\exp\left(\frac{1}{\sqrt{2}}\sqrt{(1-2\alpha)\log T \log\log T}\right)$$

while $|P_X(\frac{1}{2} + it)|$ gets as large as

$$\exp\left(\sqrt{\alpha \log T \log\log T}\right).$$

What is the significance of these results? If one combines them to look at the results for zeta, the resulting product exceeds that of the conjectured maximum for zeta. This is reasonable; given that the two dominant regimes (P_X and Z_X) occur for differing values of X (see Chapter 1.3.7) we should not expect that Z_X and P_X attain their maxima simultaneously, as we discussed in our calculations in Chapters 4.1 and 4.2.

Our approach comes from studying the distribution of the large values of $|Z_X P_X|$. Assuming statistical independence of the tails of the distributions of Z_X and P_X, we expect that the distribution of $\log|Z_X| + \log|P_X|$ can be evaluated by the convolution of the two distributions. With this idea, and with the additional terms attained from our analysis, we should be able to compute refined large deviations results for zeta.

For large K, we expect that

$$
\begin{aligned}
&\frac{1}{T}\operatorname{meas}\left\{0<t<T : \log\left|P_X\left(\frac{1}{2}+it\right)\right| + \log\left|Z_X\left(\frac{1}{2}+it\right)\right| \geq \log K\right\} \\
&\quad = \int_{-\infty}^{\infty} \exp\Bigg(- \frac{x^2}{(1-\alpha)\log\log T - \log x} \\
&\quad + \frac{x^2}{((1-\alpha)\log\log T - \log x)^2}\log\log\log T + \frac{x^2\left(\frac{3}{2}+\log(1-\alpha)-2\log 2\right)}{((1-\alpha)\log\log T-\log x)^2} \\
&\quad + \frac{x^2}{((1-\alpha)\log\log T - \log x)^2}\log\left(1-\frac{\log x}{(1-\alpha)\log\log T}\right) + O\left(\frac{x^2}{(\log\log T)^3}\right)\Bigg) \\
&\quad \times \exp\Bigg(- \frac{(\log K - x)^2}{\alpha\log\log T} - \frac{(\log K - x)^2}{(\alpha\log\log T)^2}\log(2\log(\log K - x)) \\
&\quad + O\left(\frac{(\log K - x)^2}{(\log\log T)^2\log(\log K - x)}\right)\Bigg)\ \mathrm{d}x \\
&\quad = \int_{-\infty}^{\infty} \exp(-f_K(x))\ \mathrm{d}x,
\end{aligned}
$$

where $f_K(x)$ is the negative function inside the exponent (the minus sign is included here for convention). Here we utilise our refined large deviations results for both Z_X and P_X (resp. $\exp(R_X)$).

By the saddle point method this result equals

$$\exp(-f_K(x_0))$$

where x_0 is the solution to the equation $f_K'(x_0) = 0$. That is, x_0 satisfies the equation

$$
\begin{aligned}
0 = f_K'(x_0) &= \frac{2x}{(1-\alpha)\log\log T - \log x} + \frac{x}{((1-\alpha)\log\log T - \log x)^2} \\
&- \frac{2x}{((1-\alpha)\log\log T - \log x)^2}\log\log\log T - \frac{2x}{((1-\alpha)\log\log T - \log x)^3}\log\log\log T \\
&+ \frac{2x\left(\frac{3}{2}+\log(1-\alpha)-2\log 2\right)}{((1-\alpha)\log\log T-\log x)^2} + \frac{2x\left(\frac{3}{2}+\log(1-\alpha)-2\log 2\right)}{((1-\alpha)\log\log T-\log x)^3} + O\left(\frac{x}{(\log\log T)^3}\right) \\
&- \frac{2(\log K - x)}{\alpha\log\log T} - \frac{2(\log K - x)}{(\alpha\log\log T)^2}\log(2\log(\log K - x)) \\
&+ O\left(\frac{\log K - x}{(\log\log T)^2\log(\log K - x)}\right).
\end{aligned}
$$

Taking $K = \exp\left(d\sqrt{\log T\log\log T}\right)$ (as this is the conjectured order of the maximum) and balancing the leading order contributions from P_X and Z_X gives

$$0 = f_K(x_0) = \frac{2x_0}{(1-\alpha)\log\log T - \log x_0} - \frac{2(d\sqrt{\log T\log\log T} - x_0)}{\alpha\log\log T}.$$

Solving this equation gives the solution

$$
\begin{aligned}
x_0 = d(1-2\alpha)\sqrt{\log T \log\log T} &- \frac{2\alpha d\sqrt{\log T\log\log T}\log\log\log T}{\log\log T} \\
&- \frac{4\alpha d\log(d(1-2\alpha))\sqrt{\log T\log\log T}}{\log\log T} - \frac{2\alpha d\sqrt{\log T\log\log T}(\log\log\log T)^2}{(\log\log T)^2} \\
&+ O\left(\frac{\sqrt{\log T\log\log T}\log\log\log T}{(\log\log T)^2}\right).
\end{aligned}
$$

Here the O-term comes from evaluating the order of the sub-leading contributions from both P_X and Z_X. Substituting this back into our expression for $f_K(x_0)$ and making use of the fact that

$$
\begin{aligned}
\log\left(1-\frac{x_0}{(1-\alpha)\log\log T}\right) &= \log\left(1-\frac{1}{2(1-\alpha)}-\frac{\log\log\log T}{2(1-\alpha)\log\log T}+\cdots\right) \\
&= \log\left(\frac{1-2\alpha}{2(1-\alpha)}\right)+\log\left(1-\frac{\log\log\log T}{(1-2\alpha)\log\log T}+\cdots\right).
\end{aligned}
$$

When multiplied by the corresponding term in f_K this second logarithmic term is absorbed by the O-term, while the constant terms in the logarithm combine with the other terms in our expression. Upon simplifying, this gives

$$
\begin{aligned}
f_K(x_0) = 2d^2\log T &+ \frac{2d^2\log T\log\log\log T}{\log\log T} + \frac{4d^2(\log d - \frac{3}{2} + 3\log 2)\log T}{\log\log T} \\
&+ \frac{2d^2\log T(\log\log\log T)^2}{(\log\log T)^2} + O\left(\frac{\log T\log\log\log T}{(\log\log T)^2}\right).
\end{aligned}
$$

As we see, there is no α present in the final expression for $f_K(x_0)$. This is what we should expect, as the maximum value attained by the zeta function should be independent of the value of α.

This expression for $f_K(x_0)$ enables us to conjecture the following:

Conjecture 4.4. *For $d>0$ fixed and $T\to\infty$, we have*

$$
\begin{aligned}
&\frac{1}{T}\operatorname{meas}\left\{0<t<T : \left|\zeta\left(\frac{1}{2}+it\right)\right| > \exp\left(d\sqrt{\log T\log\log T}\right)\right\} \\
&\quad = \exp\Bigg(-2d^2\log T - \frac{2d^2\log T\log\log\log T}{\log\log T} - \frac{4d^2(\log d-\frac{3}{2}+3\log 2)\log T}{\log\log T} \\
&\qquad - \frac{2d^2\log T(\log\log\log T)^2}{(\log\log T)^2} + O\left(\frac{\log T\log\log\log T}{(\log\log T)^2}\right)\Bigg).
\end{aligned}
$$

This conjecture allows us to compute refined large deviations results for zeta. Provided our results for P_X (resp. R_X) and Z_X are valid we can balance both sides of this expression as we have done throughout this thesis: at a height T up the critical line, we know that there are (neglecting constants) $T\log T$ zeros; multiplying both sides of the above expression by $T\log T$ gives

$$\mathbb{P}\left(\left|\zeta\left(\frac{1}{2}+it\right)\right| > \exp\left(d\sqrt{\log T\log\log T}\right)\right) = T\log T\exp\Bigg(-2d^2\log T - \frac{2d^2\log T\log\log\log T}{\log\log T} - \frac{4d^2(\log d-\frac{3}{2}+3\log 2)\log T}{\log\log T} - \frac{2d^2\log T(\log\log\log T)^2}{(\log\log T)^2} + O\left(\frac{\log T\log\log\log T}{(\log\log T)^2}\right)\Bigg).$$

If we then balance this equation and solve for d, we obtain (at leading order)

$$d = \sqrt{\frac{1}{2}+\frac{\log\log T}{\log T}}$$

and so we recover the original conjecture at leading order for the maximum of $\zeta(\frac{1}{2}+it)$ along the critical line. Once again, the presence of these additional terms enables us to go further and obtain a refined solution for d:

$$d = \sqrt{\frac{1}{2}-\frac{1}{2}\frac{\log\log\log T}{\log\log T}+\frac{\log\log T}{\log T}}.$$

Balancing the above probability and solving for the maximum, we can then plug our solution for d into this expression to obtain the following final result:

Consequence 4.2. *If our models for P_X and Z_X are both suitable, then*

$$\max_{t\in[0,T]}|\zeta(\tfrac{1}{2}+it)| = \exp(\sqrt{\tfrac{1}{2}\log T\log\log T - \tfrac{1}{2}\log T\log\log\log T + o(\log T\log\log\log T)}).$$

A reasonable question here is to ask whether this result is feasible given the results we have for P_X and Z_X. The subleading term here in the result for the maximum of zeta is the sum of the subleading parts from both the P_X and Z_X case. While both are contributing here to the results for the zeta function, we should be anticipating some consistency between results.

This implies that in the three regimes, the maximum of the zeta function varies. This is unreasonable, since the maximum of zeta is not dependent on the parameter X. We therefore must ask ourselves the question of what is going on here, and is there some way we can still obtain a more precise answer as to the subleading behaviour in the maximum of the zeta function if the result above is to not be believed?

We propose an answer to this final question in Chapter 6, while Chapter 5 is dedicated to showing that R_X is a suitable model for the product over primes P_X, and therefore we can believe the results we attain for this are a good representation for the maximum of the zeta function.

5 Mathematical Justification for the use of R_X as a Model for P_X^*

We've introduced and implemented an alternative model, R_X, for computing refined large deviations results for P_X^*; but is this model valid, and for what values of X is this the case?

As we saw in Section 4.3.2 we model P_X^* by the summand R_X, and it is there we assert that R_X is a good model for P_X^*. The question we seek to answer in this Chapter is: can we formally verify that this is indeed the case?

To do this, we utilise a notion from the theory of probability, and we discuss this in detail in the following Section.

5.1 The Hamburger Moment Problem

The Hamburger Moment problem asks the following question: when is a distribution completely determined by its moments? That is, if we know all the moments of a distribution, when does this allow us to completely determine what that distribution must be?

To help us answer this question, we rely on the following result of Carlemann, which is stated in [50] as follows:

Theorem 5.1 (Theorem 1.10, Carlemann)**.** *A sufficient condition that the Hamburger moment problem be determined is that*

$$\sum_{n=1}^{\infty} \mu_{2n}^{-1/2n} = \infty.$$

More generally, it is sufficient that

$$\sum_{n=1}^{\infty} \gamma_{2n}^{-1/2n} = \infty$$

where

$$\gamma_{2n} = \inf_{\nu \geq n} (\mu_{2\nu})^{1/2\nu}.$$

Here the μ_{2n} denotes the $2n^{\text{th}}$ moment. How can we make use of this result? If we were able to determine that both P_X and R_X were determined by their moments, showing that P_X aligns with R_X in the limit as T tends to infinity amounts to showing that the moments of these two distributions agree with one another in the limit.

Formally, we have the following expressions for P_X^* and R_X:

$$\Re\left(\sum_{p\le X}\frac{1}{p^{1/2+it}}\right)\quad\text{and}\quad\Re\left(\sum_{p\le X}\frac{z_p}{\sqrt{p}}\right),$$

where z_p is an i.i.d random variable uniformly distributed on the unit circle. If we are able to show for all natural numbers k that

$$\mathbb{E}\left[\left(\Re\sum_{p\le X}\frac{1}{p^{1/2+it}}\right)^k\right]\to\mathbb{E}\left[\left(\Re\sum_{p\le X}\frac{z_p}{\sqrt{p}}\right)^k\right]$$

as the variable T tends to infinity (where $t\in[0,T]$ and $X=\exp(\sqrt{\log T})$ as before), and that the result in Theorem 5.1 holds, then we can say that P_X^* converges to R_X in distribution.

Some care is needed when evaluating these expectations, however; the first of these is evaluated with respect to t, while the second is evaluated with respect to the variable z_p. As integrals then, we aim to prove that for $k\in\mathbb{N}$ we have

$$\frac{1}{T}\int_0^T\left(\Re\sum_{p\le X}\frac{p^{-it}}{\sqrt{p}}\right)^k\,\mathrm{d}t\;\to\;\frac{1}{2\pi}\int_0^{2\pi}\left(\Re\sum_{p\le X}\frac{e^{i\theta_p}}{\sqrt{p}}\right)^k\,\mathrm{d}\theta_p \tag{56}$$

as T approaches infinity.

For this, we proceed by evaluating the moments of P_X^* and R_X for values of k on a case-by-case basis and determining whether these expressions agree in the limit.

5.1.1 Case $k=1$

We begin by looking at the first moment calculation for P_X^*. For this integral expression, this evaluates to

$$\frac{1}{T}\sum_{p\le X}\int_0^T\frac{\cos(t\log p)}{\sqrt{p}}\,\mathrm{d}t=\frac{1}{T}\sum_{p\le X}\left[\frac{\sin(t\log p)}{\sqrt{p}\log p}\right]_0^T=\frac{1}{T}\sum_{p\le X}\frac{\sin(T\log p)}{\sqrt{p}\log p}$$

Since both sum and integral are finite we can interchange integral and summation at will; we then see that this expression vanishes in the limit $T\to\infty$ as

$$\frac{1}{T}\sum_{p\le X}\frac{\sin(T\log p)}{\sqrt{p}\log p}\ll\frac{1}{T}\sum_{p\le X}\frac{1}{\sqrt{p}}\sim\frac{1}{T}\cdot\frac{2\sqrt{X}}{\log X}$$

and for $X=\exp(\sqrt{\log T})$ this expression approaches zero in the limit. Therefore in the $k=1$ case, this expression vanishes.

If we consider now the first moment calculation for R_X, we find that this evaluates to give

$$\mathbb{E}\left[\Re\sum_{p\le X}\frac{z_p}{\sqrt{p}}\right]=\frac{1}{2\pi}\sum_{p\le X}\int_0^{2\pi}\frac{\cos(\theta_p)}{\sqrt{p}}\,\mathrm{d}\theta_p=\frac{1}{2\pi}\sum_{p\le X}\left[\frac{\sin(\theta_p)}{\sqrt{p}}\right]_0^{2\pi}=0,$$

and so we see clear agreement between the models in the large T limit for the first moment.

5.1.2 Case $k = 2$

The steps for the second moment calculation are the same as above, only a little more work is required in the $k = 2$ case. Our reasons for considering the second moment separately will become clear when we later look at higher moments, as there are some subtleties that present themselves for higher order moments which are not on display here.

Beginning with the second moment for P_X^* we find that this is equal to

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{p\le X}\frac{\cos(t\log p)}{\sqrt{p}}\right)^2\right]&=\mathbb{E}\left[\sum_{p\le X}\frac{\cos^2(t\log p)}{p}+\sum_{\substack{p,p'\le X\\ p\ne p'}}\frac{\cos(t\log p)\cos(t\log p')}{\sqrt{pp'}}\right]\\&=\mathbb{E}\left[\sum_{p\le X}\frac{\cos^2(t\log p)}{p}\right]+\mathbb{E}\left[\sum_{\substack{p,p'\le X\\ p\ne p'}}\frac{\cos(t\log p)\cos(t\log p')}{\sqrt{pp'}}\right].\end{aligned}$$

For this, we evaluate both terms individually; the first of these gives

$$\begin{aligned}\mathbb{E}\left[\sum_{p\le X}\frac{\cos^2(t\log p)}{p}\right]&=\frac{1}{T}\sum_{p\le X}\int_0^T\frac{\cos^2(t\log p)}{p}\,\mathrm{d}t=\frac{1}{T}\sum_{p\le X}\int_0^T\frac{1+\cos(2t\log p)}{2p}\,\mathrm{d}t\\&=\frac{1}{T}\sum_{p\le X}\left[\frac{t}{2p}+\frac{\sin(2t\log p)}{4p\log p}\right]_0^T\\&=\sum_{p\le X}\frac{1}{2p}+\frac{1}{T}\sum_{p\le X}\frac{\sin(2T\log p)}{4p\log p}.\end{aligned}$$

The first of these terms does not vanish in the limit, instead growing like $\log\log X$. The second term vanishes as the sum is bounded above by another of order equal to that of the first, $\sum_{p\le X}1/p$, which we previously stated is order $\log\log X\sim\frac{1}{2}\log\log T$ as $X=\exp(\sqrt{\log T})$; therefore the $1/T$ scaling present ensures that this vanishes in the limit.

If we look at the second term in the second moment, we find that

$$
\begin{aligned}
\mathbb{E}\left[\sum_{\substack{p,p'\le X\\ p\neq p'}}\frac{\cos(t\log p)\cos(t\log p')}{\sqrt{pp'}}\right] &= \frac{1}{T}\sum_{\substack{p,p'\le X\\ p\neq p'}}\int_0^T\frac{\cos(t\log p)\cos(t\log p')}{\sqrt{pp'}}\,\mathrm{d}t\\
&= \frac{1}{T}\sum_{\substack{p,p'\le X\\ p\neq p'}}\int_0^T\frac{\cos(t(\log p+\log p'))+\cos(t(\log p-\log p'))}{2\sqrt{pp'}}\,\mathrm{d}t\\
&= \frac{1}{T}\sum_{\substack{p,p'\le X\\ p\neq p'}}\frac{\sin(T(\log p+\log p'))}{2\sqrt{pp'}(\log p+\log p')}+\frac{\sin(T(\log p-\log p'))}{2\sqrt{pp'}(\log p-\log p')}.
\end{aligned}
$$

Numerically, these two terms appear to vanish in the large T limit, but our aim is to formally prove this.

Our approach is to bound each term by other terms which vanish in the large T limit. In fact the argument implemented here can be applied to either term; looking at the first term, without loss of generality we evaluate the sum first in terms of p by utilising an upper bound which allows us to split the double sum into two sums; following which we then apply an Euler-Maclaurin argument with respect to the variable p:

$$
\begin{aligned}
\frac{1}{T}\sum_{\substack{p,p'\le X\\ p\neq p'}}\frac{\sin(T(\log p+\log p))}{2\sqrt{pp'}(\log p+\log p')} &\ll \frac{1}{T}\sum_{p'\le X}\sum_{p\le X}\frac{1}{2\sqrt{pp'}\log p}\ll\frac{1}{T}\sum_{p'\le X}\sum_{p\le X}\frac{1}{2\sqrt{pp'}}\\
&= \frac{1}{T}\sum_{p'\le X}\left[\frac{1}{2\sqrt{2p'}}+\frac{1}{2\sqrt{p'}}\int_2^X\frac{1}{\sqrt{t}\log t}\,\mathrm{d}t\right]\\
&= \frac{1}{T}\sum_{p'\le X}\frac{1}{2\sqrt{2p'}}+\frac{1}{2\sqrt{p'}}\cdot\frac{2\sqrt{X}}{\log X}+O\left(\frac{\sqrt{X}}{\sqrt{p}(\log X)^2}\right)\\
&= \frac{1}{T}\left(\frac{2X}{(\log X)^2}+O\left(\frac{X}{(\log X)^3}\right)\right),
\end{aligned}
$$

Looking at the first of these terms we see that this vanishes in the limit as $X = \exp(\sqrt{\log T})$ is sublinear. We therefore see from this expression that this double sum vanishes in the large T limit.

We now look at the second of these terms, and for this we cannot utilise the same arguments. Since p and p' are not equal we split this term into two dominant sums:

$$
\begin{aligned}
\frac{1}{T}\sum_{\substack{p,p'\le X\\ p\ne p'}}\frac{\sin(T(\log p-\log p'))}{2\sqrt{pp'}(\log p-\log p')} &= \frac{1}{T}\sum_{p'<p\le X}\frac{\sin(T(\log p-\log p'))}{2\sqrt{pp'}(\log p-\log p')}\\
&\quad+\frac{1}{T}\sum_{p<p'\le X}\frac{\sin(T(\log p-\log p'))}{2\sqrt{pp'}(\log p-\log p')}\\
&=\frac{1}{T}\sum_{p'\le X}\frac{\sin(T(\log(p'+2)-\log p'))}{2\sqrt{p'(p'+2)}(\log(p'+2)-\log p')}\\
&\quad+\frac{1}{T}\sum_{p'\le X}\frac{\sin(T(\log(p'-2)-\log p'))}{2\sqrt{p'(p'-2)}(\log(p'-2)-\log p')}\\
&\quad+\text{(Subleading Terms)}.
\end{aligned}
$$

Here, our calculations stem from the fact that we can split the sum in terms of p both exceeding and not exceeding p', since we have the added restriction that $p \neq p$. We then observe that the dominant contributions come from making $\log p - \log p'$ as small as possible (since this expression with the $1/T$-scaling is the well-known sinc function and $\operatorname{sinc}(x)$ gets larger as the variable x gets smaller). We therefore want $\log p - \log p'$ as small as possible, and this occurs when the primes are closest together (which for general primes is when p and p' are separated by a distance of 2).

From here we upper bound this expression, collect logarithms and expand. The first sum gives

$$
\begin{aligned}
\frac{1}{T}\sum_{p'\le X}\frac{\sin(T(\log(p'+2)-\log p'))}{2\sqrt{p'(p'+2)}(\log(p'+2)-\log p')} &\ll \frac{1}{T}\sum_{p'\le X}\frac{1}{2\sqrt{p'(p'+2)}(\log(p'+2)-\log p')}\\
&=\frac{1}{T}\sum_{p'\le X}\frac{1}{2\sqrt{p'(p'+2)}(2/p'+\cdots)}\\
&=\frac{1}{T}\sum_{p'\le X}\frac{p'}{4\sqrt{p'(p'+2)}}\left[1+\frac{1}{p'}+\cdots\right].
\end{aligned}
$$

The leading term here is constant order, which when summing over all primes $p' \le X$ gives a factor of $X/\log X$, which for $X = \exp(\sqrt{\log T})$ vanishes due to the $1/T$ scaling present. This argument also applies to the second of these terms, albeit with some sign changes present.

Turning our attention now to the second moment for R_X we find that this is significantly easier to evaluate; here we have

$$\mathbb{E}\left[\left(\sum_{p\le X}\frac{\cos(\theta_p)}{\sqrt{p}}\right)^2\right] = \mathbb{E}\left[\sum_{p\le X}\frac{\cos^2(\theta_p)}{p} + \sum_{\substack{p,p'\le X\\ p\ne p'}}\frac{\cos(\theta_p)\cos(\theta_{p'})}{\sqrt{pp'}}\right]$$
$$= \mathbb{E}\left[\sum_{p\le X}\frac{\cos^2(\theta_p)}{p}\right] + \mathbb{E}\left[\sum_{\substack{p,p'\le X\\ p\ne p'}}\frac{\cos(\theta_p)\cos(\theta_{p'})}{\sqrt{pp'}}\right].$$

The first of these expressions evaluates to give

$$\begin{aligned}\mathbb{E}\left[\sum_{p\le X}\frac{\cos^2(\theta_p)}{p}\right] &= \frac{1}{2\pi}\sum_{p\le X}\int_0^{2\pi}\frac{\cos^2(\theta_p)}{p}\,\mathrm{d}\theta_p = \frac{1}{2\pi}\sum_{p\le X}\int_0^{2\pi}\frac{1+\cos(2\theta_p)}{2p}\\ &= \frac{1}{2\pi}\sum_{p\le X}\left[\frac{\theta_p}{2p}+\frac{\sin(2\theta_p)}{4p}\right]_0^{2\pi}\\ &= \sum_{p\le X}\frac{1}{2p}.\end{aligned}$$

So we see agreement in the initial terms for the second moment; we now verify this for the subleading term. Evaluating this we find that

$$\begin{aligned}\mathbb{E}\left[\sum_{\substack{p,p'\le X\\ p\ne p'}}\frac{\cos(\theta_p)\cos(\theta_{p'})}{\sqrt{pp'}}\right] &= \frac{1}{(2\pi)^2}\sum_{\substack{p,p'\le X\\ p\ne p'}}\int_0^{2\pi}\int_0^{2\pi}\frac{\cos(\theta_p)\cos(\theta_{p'})}{\sqrt{pp'}}\,\mathrm{d}\theta_p\mathrm{d}\theta_{p'}\\ &= \frac{1}{(2\pi)^2}\sum_{\substack{p,p'\le X\\ p\ne p'}}\int_0^{2\pi}\frac{\cos(\theta_p)}{\sqrt{p}}\,\mathrm{d}\theta_p\int_0^{2\pi}\frac{\cos(\theta_{p'})}{\sqrt{p'}}\,\mathrm{d}\theta_{p'}\\ &= 0\end{aligned}$$

upon evaluating either integral. Therefore we see agreement also in the case $k = 2$.

5.1.3 Case $k > 2$ for P_X^*

For the k^{th} ($k > 2$) moment a little more care is needed as there may be additional terms which do not vanish in the limit; we consider the k^{th} moments for both P_X^* and R_X separately. Before looking at these terms we go through the main steps for P_X^*.

Starting with P_X^*, we have the following for the higher moments:

$$
\begin{aligned}
\mathbb{E}\left[\left(\sum_{p\le X}\frac{\cos(t\log p)}{\sqrt{p}}\right)^k\right] &= \mathbb{E}\left[\sum_{p\le X}\frac{\cos^k(t\log p)}{\sqrt{p}^k} + \sum_{\substack{p,p'\le X\\ p\neq p'}}\frac{\cos^{k-1}(t\log p)\cos(t\log p')}{\sqrt{p}^{k-1}\sqrt{p'}} + \cdots\right] \\
&= \mathbb{E}\left[\sum_{p\le X}\frac{\cos^k(t\log p)}{\sqrt{p}^k}\right] + \mathbb{E}\left[\sum_{\substack{p,p'\le X\\ p\neq p'}}\frac{\cos^{k-1}(t\log p)\cos(t\log p')}{\sqrt{p}^{k-1}\sqrt{p'}}\right] \\
&\quad + \cdots .
\end{aligned}
$$

For the initial term in the expression, the limit of the evaluation depends on whether or not k is odd or even. What do we mean by this?

For even $k > 2$ we find that $\cos^k(x)$ can be written as a sum of $\cos(lx)$ terms for even $l \le k$, with non-zero constant term. We arrive at this result via the following argument: using Euler's theorem and the Binomial theorem we can rewrite the n^{th} power of cosine as

$$
\begin{aligned}
\cos^n(x) = \left(\frac{e^{ix}+e^{-ix}}{2}\right)^n &= \frac{1}{2^n}\sum_{k=0}^{n}\binom{n}{k}e^{ikx}e^{-i(n-k)x} = \frac{1}{2^n}\sum_{k=0}^{n}\binom{n}{k}e^{i(2k-n)x} \\
&= \frac{1}{2^n}\left[e^{-inx} + \binom{n}{1}e^{i(2-n)x} + \binom{n}{2}e^{i(4-n)x} + \cdots + \binom{n}{n-1}e^{i(n-2)x} + e^{inx}\right].
\end{aligned}
$$

There are $n+1$-terms in this expression, so for odd n the terms combine with one another to give

$$
\cos^n(x) = \frac{1}{2^{n-1}}\sum_{k=\frac{n+1}{2}}^{n}\binom{n}{k}\cos((2k-n)x)
$$

and there is no constant term in this expression. If n is even, there are an odd number of terms in the above expression and the $k = (n/2)^{\text{th}}$ term gives a constant term equal to $\frac{1}{2^{n-1}}\binom{n}{n/2}$.

Looking then at the leading moment for even k, performing this integration over the interval $[0, T]$ means that the cosine terms vanish and the factor of T picked up from integrating the constant term cancels with the $1/T$ scaling. Therefore the leading term in the expectation is equal to

$$
\frac{1}{2^{k-1}}\binom{k}{k/2}\sum_{p\le X}\frac{1}{\sqrt{p}^k} = O(1) \text{ for } k > 2.
$$

In the case where we are looking at odd $k > 2$, using the expression for cosine above we see that there is no constant term in this expression and the other terms in the expression integrate to become sine terms, which vanish in the large T limit due to the $1/T$ scaling present.

We now consider the sub-leading terms in the expression for P_X, and we again split this based on whether k is odd or even.

If k is odd, in each sub-leading term at least one of the cosine terms is odd. This means that, in each of the sub-leading terms, there is no constant present. Using our formula above for the k^{th} power of cosine, all terms can be decomposed into products of cosine terms. Using also the fact that

$$\cos(x)\cos(y) = \frac{1}{2}(\cos(x+y) + \cos(x-y)),$$

we can apply this result repeatedly to decompose these products of cosines into sums of single cosines with exponent 1. After integrating, these all then become sine terms and thus vanish in the limit due to the $1/T$ scaling. We conclude then that all sub-leading terms for the odd k^{th} moment have limit 0, and there are therefore no sub-leading terms for P_X when k is odd.

If k is even we can apply the same ideas from the odd case. If there are any odd powers of cosine in the sub-leading terms, there are no constant terms present and as in the odd case these terms vanish in the large-T limit. If all the cosine terms in any of the sub-leading terms are even, there is a constant in each term which cancels with the $1/T$-scaling. Therefore any terms with all even powers gives a constant, and since the powers add to make k, after integrating and looking at the large T limit each of these terms is equal to

$$\prod_{k_i} \frac{1}{2^{k_i-1}} \binom{k_i}{k_i/2} \sum_{p \le X} \frac{1}{\sqrt{p}^{k_i}}.$$

The k^{th} moment is given by the sum of these terms, and is therefore given by

$$\sum_{i=1}^{k/2} \sum_{j=1}^{i} \left(\prod_{\substack{k_{i_j} \text{ even} \\ \sum k_{i_j} = k/2}} \frac{1}{2^{k_{i_j-1}}} \binom{k_{i_j}}{k_{i_j}/2} \sum_{p \le X} \frac{1}{\sqrt{p}^{k_{i_j}}} \right)$$

Thus we have deduced the limit of each term in the expansion for the k^{th} moment of P_X.

Case $k > 2$ for R_X

We now turn our attention to the k^{th} moment for R_X. As was the case with P_X^* we consider the leading term first for both odd and even k, before looking at the subleading terms.

For the k^{th} moment for R_X, we find the following:

$$
\begin{aligned}
\mathbb{E}\left[\left(\sum_{p\leq X}\frac{\cos(\theta_p)}{\sqrt{p}}\right)^k\right] &= \mathbb{E}\left[\sum_{p\leq X}\frac{\cos^k(\theta_p)}{\sqrt{p}^k} + \sum_{\substack{p,p'\leq X\\ p\neq p'}}\frac{\cos^{k-1}(\theta_p)\cos(\theta_{p'})}{\sqrt{p}^{k-1}\sqrt{p'}} + \cdots\right] \\
&= \mathbb{E}\left[\sum_{p\leq X}\frac{\cos^k(\theta_p)}{\sqrt{p}^k}\right] + \mathbb{E}\left[\sum_{\substack{p,p'\leq X\\ p\neq p'}}\frac{\cos^{k-1}(\theta_p)\cos(\theta_{p'})}{\sqrt{p}^{k-1}\sqrt{p'}}\right] + \cdots .
\end{aligned}
$$

As before the evaluation of this expression depends on the value of k; if k is odd, the expectation evaluates to give

$$
\begin{aligned}
\mathbb{E}\left[\sum_{p\leq X}\frac{\cos^k(\theta_p)}{\sqrt{p}^k}\right] &= \frac{1}{2\pi}\sum_{p\leq X}\int_0^{2\pi}\frac{\cos^k(\theta_p)}{\sqrt{p}^k}\,\mathrm{d}\theta_p \\
&= 0.
\end{aligned}
$$

That this is the case can be seen graphically. If one looks at the plot of $\cos^k(x)$ for $x\in[0,2\pi]$ the area under the curve, $x\in[\pi/2,\frac{3\pi}{2}]$ is equal to the sum of the areas under the curve for $x\in[0,\frac{\pi}{2}]$ and $x\in[\frac{3\pi}{2},2\pi]$, so these contributions cancel to give resulting integral zero.

In the case where $k>2$ is even, there is a constant term in the expression for $\cos^k(x)$, which as we see above is

$$
\frac{1}{2^{k-1}}\binom{k}{k/2}\sum_{p\leq X}\frac{1}{\sqrt{p}^k} = O(1) \text{ as } T \;\to\; \infty,
$$

and thus these initial terms for $k>2$ agree in the large T limit.

What about the sub-leading terms? Again, whether k is odd or even will come into play here.

We consider the case where k is odd first: if k is odd, in each of the sub-leading terms at least one of the powers of cosine which makes up the product must be odd. Otherwise there will be a term where the sums of the powers of the cosine terms will be even, contradicting the fact that we are looking at the odd k^{th} moment.

That is, if we have terms of the form

$$
\cos^{k_1}(\theta_{p_1})\cdots\cos^{k_i}(\theta_{p_i})
$$

where $k_1+\cdots+k_i=n$, at least one of the k_i must be odd otherwise the sum will be even.

For these powers of cosine for which the exponent is odd, we know from looking at the leading term that this will integrate to give zero. Therefore every sub-leading term in the odd k case integrates to zero, and we thus see agreement between P_X^* and R_X for the odd k^{th} moment.

In the case where k is even, we can apply a similar argument. If there are any terms of the form $\cos^l(x)$ where l is odd we know from looking at the leading term that these integrate to zero. Therefore any non-zero terms in the even k case occur when all cosine terms which make up the product in each term all have even exponent. Therefore the value of the k^{th} moment here is the product

$$\sum_{i=1}^{k/2}\sum_{j=1}^{i}\left(\prod_{\substack{k_{i_j}\text{ even}\\ \sum k_{i_j}=k/2}} \frac{1}{2^{k_{i_j}-1}}\binom{k_{i_j}}{k_{i_j}/2}\sum_{p\le X}\frac{1}{\sqrt{p}^{k_{i_j}}}\right)$$

Therefore we see the agreement between the moments of P_X^* and R_X.

5.1.4 Verifying Carlemann's Theorem

Having verified that the moments of P_X^* and R_X agree with one another in the large T limit, all that remains is to verify the condition in Carlemann's theorem; that is, if we can show that

$$\sum_{n=1}^{\infty}\gamma_{2n}^{-1/2n}=\infty$$

for $\gamma_{2n}=\inf_{\nu\ge n}(\mu_{2\nu})^{1/2\nu}$, then we have all the information needed to verify agreement between the two models.

As the moments for P_X^* and R_X are the same, we need not verify this by considering P_X^* and R_X separately. Proving this for one of P_X^* and R_X leads to the same conclusion for the other.

We want to show that this series is divergent, and this follows from showing that the limit

$$\lim_{n\to\infty}\gamma_{2n}^{-1/2n}$$

exists and is 0, which is equivalent to showing that the limit

$$\lim_{n\to\infty}-\frac{\log\gamma_{2n}}{2n}$$

is negative infinity. The infimum in the expression for γ_{2n} occurs when $(\mu_{2n})^{1/2n}$ is at its smallest, and we can determine this by looking at the ratio

$$
\begin{aligned}
\log\left(\frac{(\mu_{2n+2})^{2n+2}}{(\mu_{2n})^{2n}}\right) &= (2n+2)\log\mu_{2n+2} - 2n\log\mu_{2n}\\
&= 2n\log\left(\frac{\mu_{2n+2}}{\mu_{2n}}\right) + 2\log\mu_{2n+2}.
\end{aligned}
$$

We consider this expression since the logarithm is a convex function, and so the limit of this expression will be equal to the logarithm of the limit $\lim_{n\to\infty}(\mu_{2n+2})^{2n+2}/(\mu_{2n})^{2n}$.

We show that this limit is $-\infty$: consider again the expressions for P_X^* and R_X:

$$
\begin{aligned}
P_X^* &= \sum_{p\leq X} \frac{\cos(-t\log p)}{\sqrt{p}},\\
R_X &= \sum_{p\leq X} \frac{\cos(\theta_p)}{\sqrt{p}}.
\end{aligned}
$$

The denominator within the sums is increasing with p and the numerator lies in the interval $[-1,1]$ for fixed t in the case of P_X^*. Taking random θ_p or $t\log p$ generally leads to values of cosine not equal to ± 1, and due to the cosine function fluctuating between positive and negative, we do not expect that these sums will produce values exceeding either ± 1. Using now the result from Probability Theory which states that $x > y$ implies $\mathbb{E}[x] > \mathbb{E}[y]$, we conclude that

$$
\mu_{2n+2} < \mu_{2n},
$$

and as a result this first logarithm term is negative (since the ratio of moments is less than 1) and therefore tends to negative infinity in the limit due to the multiplying factor of $2n$. The second term, $2\log\mu_{2n+2}$ also tends to negative infinity, since the sequence μ_{2n} decreases for $\nu > 2n$ with limit 0.

This is useful as we have demonstrated that μ_{2n} is decreasing, and therefore the infimum for γ_{2n} is given by the limit

$$
\lim_{n\to\infty}(\mu_{2n})^{1/2n}.
$$

Given both μ_{2n} and $1/2n$ have limit zero as $n\to\infty$, we have three possibilities. The first is that the infimum (and hence the above limit) is zero, and this occurs when μ_{2n} approaches zero at a faster rate than the exponent; the second possibility is that the infimum is 1, which occurs when the exponent approaches zero at a faster rate than the $2n^{\text{th}}$ moment; the final case is where the two approach zero at a similar rate and the limit is some constant which lies between 0 and 1.

In each of these cases, we arrive at the result we need. We know that if expression inside the sum $\sum_{n=1}^{\infty}\gamma_{2n}^{-1/2n}$ does not converge to zero then the sum diverges. If the value of the infimum is 1, each term inside the sum is simply 1 and so we have an infinite sum of 1's which diverges and so the result follows; if the value is a constant between 0 and 1, since the exponent is negative we find that each term inside the sum is of the form $1/c^n$ and this again has limit 1. Finally we consider if the expression has infimum 0; because we are looking at the even moments (since we saw above that these are the only ones with non-zero moments) it means that we approach the infimum as n gets large from the positive direction. Hence the reciprocal of this expression would approach positive infinity, and in all three cases the series diverges.

We can therefore conclude that Carlemann's theorem holds for our purposes, and as such P_X and R_X are determined by their respective moments. Since we have demonstrated that the moments of these two distributions are the same, we conclude that these two distributions align with one another in the large T limit.

5.2 Numerical Support

We've presented some mathematical support for why the random model R_X is a good fit for P_X^* in the large X (resp. large T) limit, but we now include some numerics to further support our argument here. There are some subtleties at play, which we outline below.

As the model R_X includes a random variable uniformly distributed on the unit circle, every realisation of R_X generates a different value/output due to the random nature. Therefore attempting to compare the numerics of these two models directly is challenging, as there is not a direct one-to-one correspondence between P_X^* and R_X.

Instead we look to show that the range of values covered by both P_X^* and R_X is comparable in the large T limit. While this does not say anything about the validity of the R_X model in the large T limit, it does supplement the above theory to suggest there is some credibility to our conjecture.

How then should we proceed? We return to the literature of Farmer, Gonek and Hughes for inspiration: in their original paper they utilised a Gaussian model to determine the leading order behaviour for the maximum of P_X, and for their modelling they took $T \log^c T$ values of t.

With this in mind, we use the following code:

```
P[X_,t_] := Re[Sum[1/(Prime[i])^(1/2+I*t), {i,1,PrimePi[X]}]];
R[X_] := Re[Sum[RandomReal[-1,1]/(Sqrt[Prime[i]]), {i,1,PrimePi[X]}]];

T = 100000;
r = T*100;
A = ConstantArray[0,T];
```

```
dX = Range[0,T,T/r];

dP = P[Exp[Sqrt[Log[T]]],dX];
dQ = Sort[dP,Less];

For[i = 1, i <= Length[A], i++, A[[i]] = R[Exp[Sqrt[Log[T]]]]];
B=Sort[A,Less];

ListPlot[dQ, B, PlotLegends->{Subscript[P, X], Subscript[R, X]}]
```

We summarise now what we our code is doing: for our first sum over primes P_X^* we evaluate it for a fixed X, $X = \exp(\sqrt{\log T})$, as t ranges from 0 to T. For R_X we perform T iterations of this model to match the number of points we evaluate P_X^* at, and we then numerically order both sets of points and plot them against one another.

These figures allow us to compare and contrast the range of values between the models in order to determine whether these seem reasonable, as well as determine how the increase in T affects how well matched the two models are. This is especially important due to the fact that we propose that the maximum for R_X aligns with that of P_X, and so a significant difference in the data sets would quickly dispel such an argument.

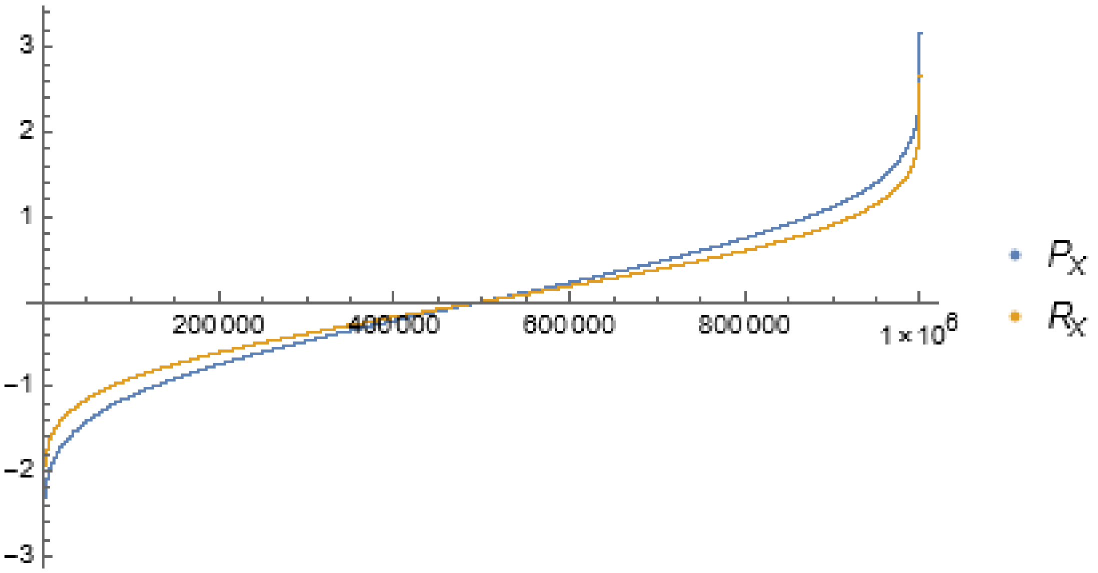


Figure 2: Listplot of $P_X^*(\frac{1}{2}+it)$ against R_X for $T = 10000$.

Comparing the results for P_X^* and R_X in Figure 2 below where we take $T = 10000$, we see that the range of values covered by both functions closely align, although P_X attains marginally higher peaks and troughs. This is reassuring, although this marginal difference may grow with T, and so it is important that we try other values of T to make sure that this is not something that occurs.

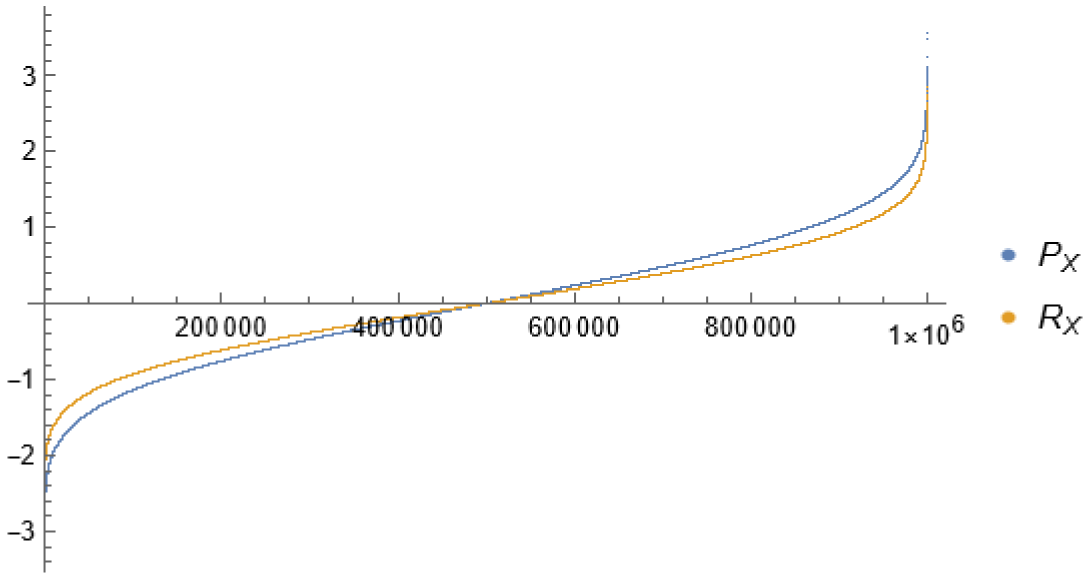


Figure 3: Listplot of $P_X^*(\frac{1}{2}+it)$ against R_X for $T = 100000$.

Comparing the results in Figure 3 and Figure 2 we see that the range of values covered by both functions closely align, and in this case the two ranges more closely align than the previous data set. It is desirable to go beyond this, however we were unable to go beyond $r = 10 \times T$ in Figure 3 due to numerical limitations.

It is worth mentioning however that in both of these graphs we do notice a slight difference between the two models at the tails which does not vanish as we increase the value of T in our model, despite the two curves more closely aligning with one another. This is something that we will discuss in more detail in Chapter 6.

6 P_X vs Z_X

We've computed refined large deviations results for both the product over primes P_X and the product over zeros Z_X, but there remains some ambiguity which we need to address, as well as the reasons for why this ambiguity persists.

If we look at the results for the maximum of the zeta function in these two regimes we see a slight difference between the two models. We recall these results here: in the regime where the dominant contribution comes from the product over zeros Z_X, we find that (up to subleading order)

$$\max_{t\in[0,T]}\left|\zeta\left(\frac{1}{2}+it\right)\right| = \exp\left(\sqrt{\frac{1}{2}\log T\log\log T + \frac{1}{2}\log T\log\log\log T + o(\log T\log\log\log T)}\right)$$

as $T\to\infty$; while our results in Chapter 4.2 are for Z_X these results should also model zeta since we are in the regime where the contribution from the product over primes is diminished, and therefore does not contribute significantly to the maximum. In the regime where the dominant contribution comes from the product over the primes P_X we argue that, provided our alternate model R_X is suitable, that

$$\max_{t\in[0,T]}\left|\zeta\left(\frac{1}{2}+it\right)\right| = \exp\left(\sqrt{\frac{1}{2}\log T\log\log T - \log T\log\log\log T + o(\log T\log\log\log T)}\right)$$

as $T\to\infty$. In the intermediate regime where both P_X and Z_X contribute (the convolution case present in Chapter 4.4), our results suggest that

$$\max_{t\in[0,T]}\left|\zeta\left(\frac{1}{2}+it\right)\right| = \exp\left(\sqrt{\frac{1}{2}\log T\log\log T - \frac{1}{2}\log T\log\log\log T + o(\log T\log\log\log T)}\right)$$

The issue which presents itself arises when looking at the subleading behaviour. With both Z_X and P_X being models for the zeta function in their respective regimes, we should expect that the results these models generate are in agreement with one another. However we see a difference in the subleading term, and this raises a few natural and instinctive questions which we must consider: why do we witness this difference between the two models, and which model (if any) do we favour for providing accurate refined large deviations results for the zeta function?

These are questions we attempt to answer in this section, starting first with the difference between the results of the models utilised in this thesis.

6.1 The Difference in the Models

The first question to consider is why there is a difference between the two models, and whether we determine the mathematical understanding as to why this difference persists. There are a number of possible justifications for the results above, and it is our aim here to determine which of these arguments is the correct one.

The first possibility here comes from studying the Euler-Hybrid Hadamard product, and looking at whether P_X is a good model for the zeta function in the large X regime. A quick look at the work of Gonek, Hughes and Keating [26] quickly dispels this possibility, since the results the authors present in their paper are theoretical (as opposed to conjectural) and unconditional. Furthermore splitting the zeta function as a product of two terms, Z_X and P_X, and agreeing with the results of the former term Z_X but not the latter P_X seems like an unreasonable conclusion. If R_X is indeed a perfectly valid model for P_X, given the theoretical nature of Gonek, Hughes and Keating's work the results should align between the two products, which as we saw in Chapter 4 is clearly not the case for the results we present here.

The second possibility is simply that, while P_X is an effective model for the zeta function, R_X is ill-fitted as a model for P_X (resp. P_X^*), and this possibility is far more feasible than the first.

In Chapter 5.2 we computed some numerics which compared the value distributions of both R_X and P_X^*, and while the two models fit reasonably well we did observe a slight difference between the plots which doesn't seem to disappear as T gets larger (although we were unable to explore this discrepancy further due to numerical limitations). This could lead one to believe there is some deficiency stemming from the alternate model R_X, and it is this line of thinking that we find believe to be the most plausible, and are therefore in agreement with.

Elaborating more on this, in Chapter 5.1 we utilised the method of moments to show that our alternate model R_X converges in distribution to P_X^* in the large T limit. While this is sufficient to determine convergence of distributions, it could be speculated that the two models in question agree only at leading order (and thus in the large T limit since only the leading order contributes anything of significance here) and we then encounter some divergence in the results of the two models at the subleading order.

This would certainly explain why we see alignment between the two models at the leading order but a lack of alignment beyond this. This also aligns with the fact that our numerical results for P_X in Chapter 7.2 do not seem to align well, adding credence to the idea that R_X is not suitably fitted for large (but not extremely large) values of T.

6.2 Choice of Model

Having discussed potential reasons for the differences between the results, we now have the matter of which results we agree with, as discussed in Chapter 6.1.

When it comes to choosing between the two models (Z_X and P_X), first impressions would be that the logical and mathematically reasonable decision would be to favour the results generated from the product over zeros, Z_X. The reasons for why this choice might seem the most rational are as follows.

Both models rely on conjecture here; Z_X relies on the conjecture of Gonek, Hughes and Keating to model Z_X via the characteristic polynomial provided we take the right choice of matrix size, N (see Conjecture 3 of [26]). In the case of P_X, we rely on the conjecture that the alternative model which we introduce, R_X, is a good model for P_X (resp. P_X^*).

Of the two conjectures that are relied upon here, Gonek, Hughes and Keating provide a great deal of support for the conjecture for Z_X (Chapter 4 of [26]), and as we outlined in Chapter 6.1 it is our belief that the difference in results between the two models most likely stems from some deficiency in the R_X model at subleading behaviour.

Further, the refined large deviations results we computed for the CUE in Chapter 3 are theoretic as oppposed to conjectural, providing a more solid foundation for believing in these results.

If we also look at the numerical results which we present in Chapter 7, the results for Z_X align closely with the conjectured maximum which we have computed. On the contrary, the numerical results for R_X do not appear to match well the maximum for P_X for the values of T used, suggesting that R_X and P_X agree at leading order only and so they only see close alignment for extremely large values of T; values of which are beyond our computational ability.

There are however reasons to not ultimately trust in the results of Z_X, which we must also take into account. As we later see in Chapter 7 it is not as clear cut that one of these solutions for the maximum of the zeta function must be the correct one. There is the possibility, although we believe it to be unlikely given the strength of our results in Chapters 3 and 4, that neither the results for Z_X or P_X are reasonable. Without further research, it is unclear which results (if any) are true.

What is clear however, is that despite the differences in the coefficients of the subleading term in all three regimes, the order of magnitude of the subleading term remains the same throughout, namely $\log T \log\log\log T$.

With this in mind, while it is my personal belief that the true answer is closer to that of the results of Z_X (although how much closer is not clear and remains to be seen) we put forward here the following broader conjecture:

Conjecture 6.1.

$$\max_{t\in[0,T]}\left|\zeta\left(\frac{1}{2}+it\right)\right| = \exp\left(\sqrt{\frac{1}{2}\log T\log\log T + c\log T\log\log\log T + o(\log T\log\log\log T)}\right),$$

where c is a constant yet to be determined.

Numerical results in Chapter 7 support this idea, and seem to suggest that the constant c lies between $-1/2$ and $1/2$. However, we do not make a definitive conclusion due to the limitations of our numerical results, and we defer the rest of the details until Chapter 7. This is an improvement over the previous conjecture from Farmer, Gonek and Hughes.

7 Numerical Results

Having computed refined large deviations results for the circular beta ensemble and the zeta function we now compare our theoretical findings against the numerical data in order to determine if the theory aligns with the data; while we can argue this is certainly the case for the circular beta ensemble due to this being a theorem, the results for the product over primes, product over zeros and the zeta function are all without formal proof and are conjectural or are the results of conjectures; having viable numerical data would provide invaluable support for these results.

We summarise the numerical methods on display here, before discussing what the data shows when compared to the theory.

7.1 Numerical Results for the CβE

We begin by looking at the numerics for the circular beta ensemble, all of which are computed and run using Wolfram Mathematica Version 13.1. For these numerics we used the values $M = T$, $N = \log T$, aligning with the theory in Chapter 3. Our results from the theory of large deviations suggest the following maxima:

- The proposed maximum for the characteristic polynomial of a matrix from the COE ($\beta = 1$) is:

$$\exp\left(\sqrt{\log T \log\log T + \log T \log\log\log T - \log T(6\log 2 - 3) + o(\log T)}\right).$$

- For $\beta = 2$ we propose that the maximum of the characteristic polynomial of a matrix from the CUE is:

$$\exp\left(\sqrt{\frac{1}{2}\log T \log\log T + \frac{1}{2}\log T \log\log\log T - \frac{1}{2}\log T(5\log 2 - 3) + o(\log T)}\right).$$

- For $\beta = 4$ we propose that the maximum of the characteristic polynomial of a matrix from the CSE is:

$$\exp\left(\sqrt{\frac{1}{4}\log T \log\log T + \frac{1}{4}\log T \log\log\log T - \frac{1}{4}\log T(4\log 2 - 3) + o(\log T)}\right).$$

We first outline the theory behind our numerics: recall from Section 4.2 that the characteristic polynomial of a random unitary matrix can be used to model Z_X, provided we take N to be $[\log T/e^{\gamma}\log X]$. For our purposes, we take matrices of size $N = [\log T]$. Asymptotically we know there are $T \log T$ zeros along the critical line up to height T, and we need to ensure that we are matching the number of zeros with the number of eigenvalues, $M = [T \log T/N] = T$.

With these values of M and N we should expect that our CUE ($\beta = 2$) results for the characteristic polynomial align with those for the product over zeros Z_X. As for the other circular ensembles ($\beta = 1$ and $\beta = 4$), computing numerics for these ensembles will enable us to compare against the case $\beta = 2$ as well as confirm if these values exceed or do not exceed those of the circular unitary ensemble, harking back to our discussion in Chapter 3.4.1.

To implement these numerically we take M and N as above, with the purpose being to look at the maxima of the characteristic polynomial which here comes from extreme values. Taking an $N \times N$ matrix from the corresponding circular ensemble, we compute its characteristic polynomial and evaluate this at r points equidistributed around the unit circle (r is defined below). From these entries we take the maximum, which gives one point in our data set. We then repeat this procedure until we have M points.

Finally, these M points are numerically ordered and plotted against the proposed maximum for the CβE, where we produce this plot for each of value of β corresponding to one of the circular ensembles.

7.1.1 Circular Orthogonal Ensemble ($\beta = 1$)

For the circular orthogonal ensemble, the following code is implemented:

```
T = 10000;
n = IntegerPart[Log[T]];
r = 50;
m = T;
A = ConstantArray[0,m];
Li = ConstantArray[0,r];
b = 1;

For[i=1, i<=m, i++,
For[j=1, j<=r, j++,
  L = RandomVariate[CircularOrthogonalMatrixDistribution[n]];
  f[x_] := Abs[CharacteristicPolynomial[L,y] /. y -> Exp[I*(2*Pi*x)/r]];
  Li[[j]] = Max[Table[f[k], {k,1,r}]];];
  A[[i]] = Max[Li]]

B = Sort[A, Less];

Show[Plot[Exp[Sqrt[(1/b)*Log[T]*Log[Log[T]] + (1/b)*Log[T]*Log[Log[Log[T]]]
```

```
- (2/b)*Log[T]*((5/2)*Log[2] + (1/2)*Log[2/b] - 3/2)]], {x,0,Length[B]},
PlotLegends -> {"Proposed Result"}], ListPlot[B, PlotLegends -> {"RMT"}],
PlotRange -> Full, AxesOrigin -> True]
```

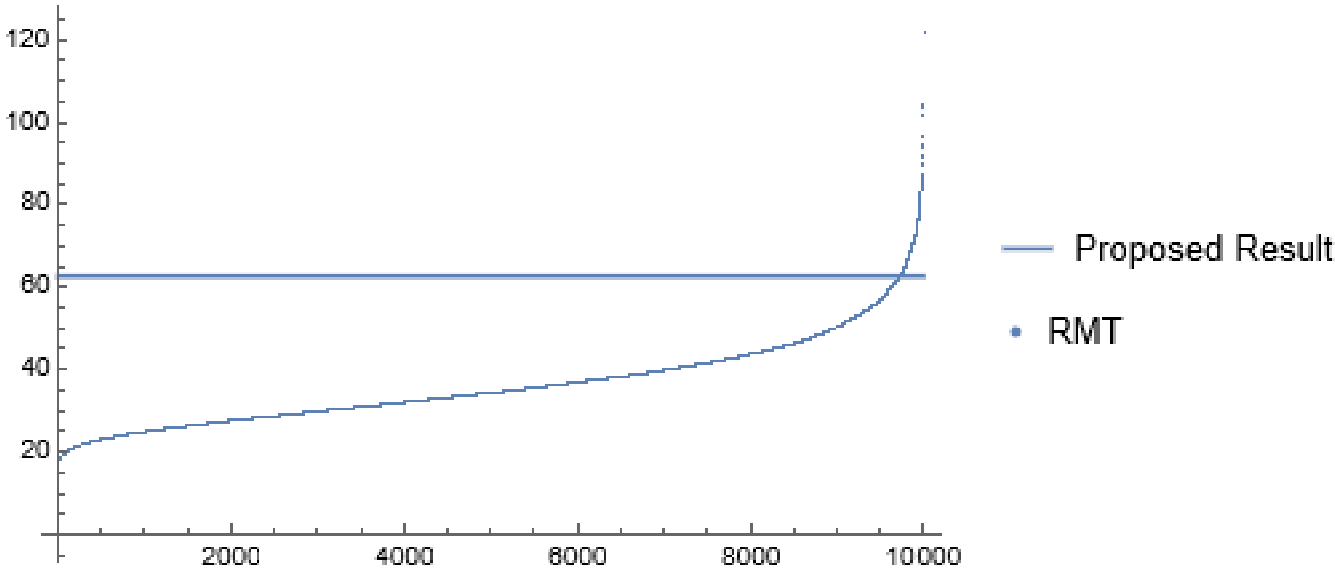


Figure 4: Listplot of T valuations of COE matrices of dimension $[\log T]$, where $T = 10,000$ and the characteristic polynomials are evaluated at $r = 50$ equally spaced points around the unit circle. The horizontal blue line represents our proposed result for the maximum, while the dotted line represents the the Random Matrix data.

Looking at the results produced in Figure 4 for $T = 10,000$, first impressions would be that these results don't appear particularly illuminating. What is clear from the data, however, is that our proposed result for the maximum in the COE case intersects the random matrix theory data somewhere between the $9,500^{\text{th}}$ and $10,000^{\text{th}}$ data points, indicating that more than 95% of our extreme maxima lie below the result proposed by our theorem.

If we compare those results with the results in Figure 5 we see an increased proportion of the maxima lie below the proposed result for the maximum compared to Figure 4, indicating that in the large T limit 100% of the data lies beneath our proposed result, as expected by our theorem.

7.1.2 Circular Unitary Ensemble ($\beta = 2$)

For the circular unitary ensemble, the following code is implemented:

```
T = 10000;
n = IntegerPart[Log[T]];
r = 50;
m = T;
A = ConstantArray[0,m];
Li = ConstantArray[0,r];
```

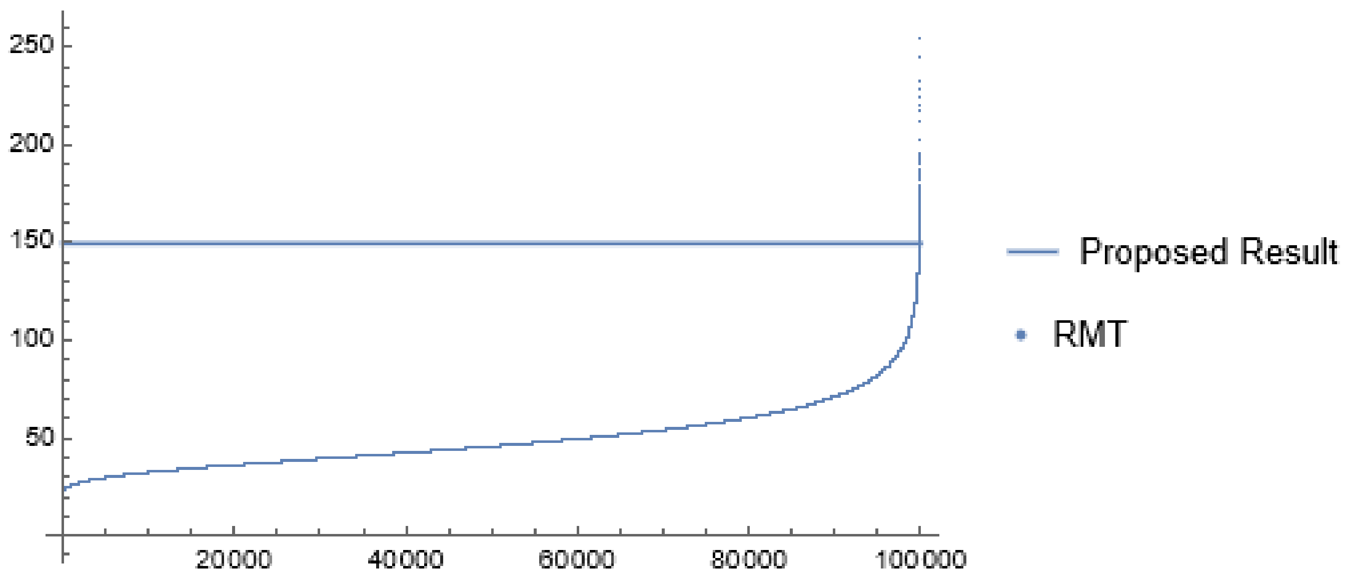


Figure 5: Listplot of T valuations of COE matrices of dimension $[\log T]$, where $T = 100,000$ and the characteristic polynomials are evaluated at $r = 50$ equally spaced points around the unit circle. The horizontal blue line represents our proposed result for the maximum, while the dotted line represents the the Random Matrix data.

```
b = 2;

For[i = 1, i <= m, i++,
 For[j = 1, j <= r, j++,
  L = RandomVariate[CircularUnitaryMatrixDistribution[n]];
  f[x_] := Abs[CharacteristicPolynomial[L,y] /. y -> Exp[I*(2*Pi*x)/r]];
  Li[[j]] = Max[Table[f[k], {k,1,r}]];];
  A[[i]] = Max[Li]]

B = Sort[A, Less];

Show[Plot[Exp[Sqrt[(1/b)*Log[T]*Log[Log[T]]]],
   Exp[Sqrt[(1/b)*Log[T]*Log[Log[T]] + (1/b)*Log[T]*Log[Log[Log[T]]]
   - (2/b)*Log[T]*((5/2)*Log[2] + (1/2)*Log[2/b] - 3/2)]], {x,0,Length[B]},
   PlotLegends -> {"FGH", "Refined Result"}],
   ListPlot[B, PlotLegends -> {"RMT"}], PlotRange -> Full,
   AxesOrigin -> True]
```

Here some slight modifications to the code have been made in comparison to the code for the circular orthogonal ensemble. As we use random unitary matrices to effectively model the zeta function, here we compare not only against our proposed result but the initial result of Farmer, Gonek and Hughes from which our results stem. This enables us to determine not only whether this result is a good fit, but also whether or not our result is a suitable refinement of the result of Farmer, Gonek and Hughes as we initially claim.

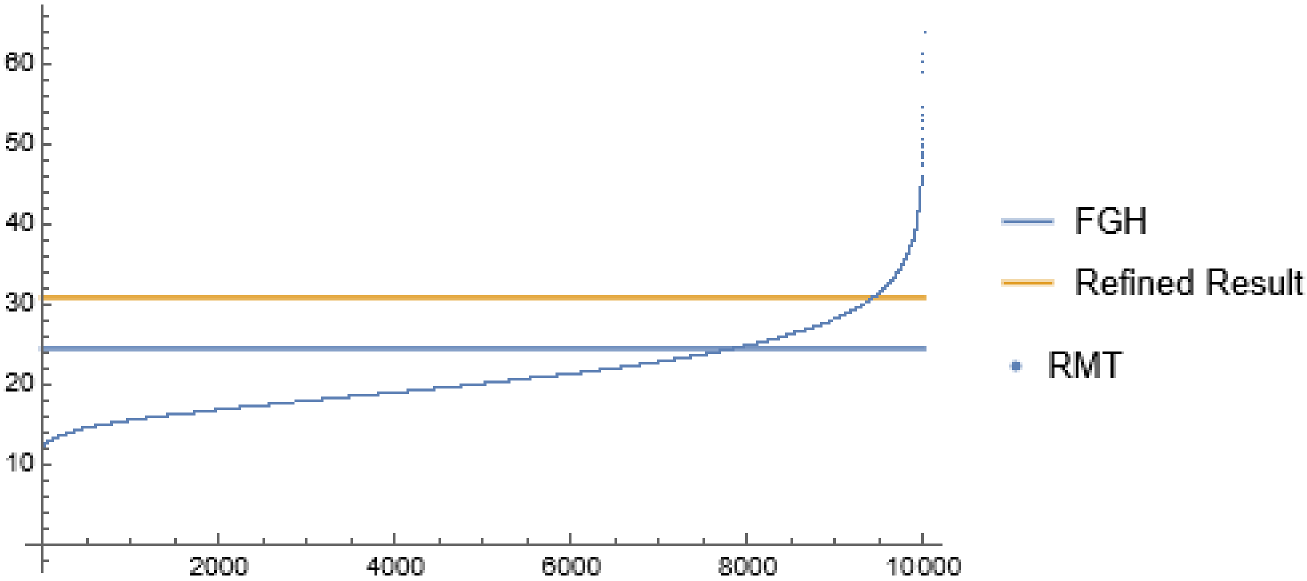


Figure 6: Listplot of T valuations of CUE matrices of dimension $[\log T]$, where $T = 10,000$ and the characteristic polynomials are evaluated at $r = 50$ equally spaced points around the unit circle. The horizontal blue line represents the conjecture of Farmer, Gonek and Hughes, while the orange line represents our refined conjecture.

Looking at the data in Figure 6 we see that our refined Large Deviations result more closely captures the random matrix theory (RMT) data. In order to get a better understanding as to whether this refined result is reasonable we again evaluate this for larger T. As our result is conjectured for T tending to infinity, we should expect fewer points in our RMT data to lie above the horizontal line. The results for this are located in Figure 7.

7.1.3 Circular Symplectic Ensemble ($\beta = 4$)

For the circular symplectic ensemble, the following code is implemented:

```
T = 10000;
n = IntegerPart[Log[T]];
r = 50;
m = T;
A = ConstantArray[0,m];
Li = ConstantArray[0,r];
b = 4;

For[i = 1, i <= m, i++,
 For[j = 1, j <= r, j++,
  L = RandomVariate[CircularUnitaryMatrixDistribution[n]];
  f[x_] := Abs[Sqrt[CharacteristicPolynomial[L,y] /. y -> Exp[I*(2*Pi*x)/r]]];
```

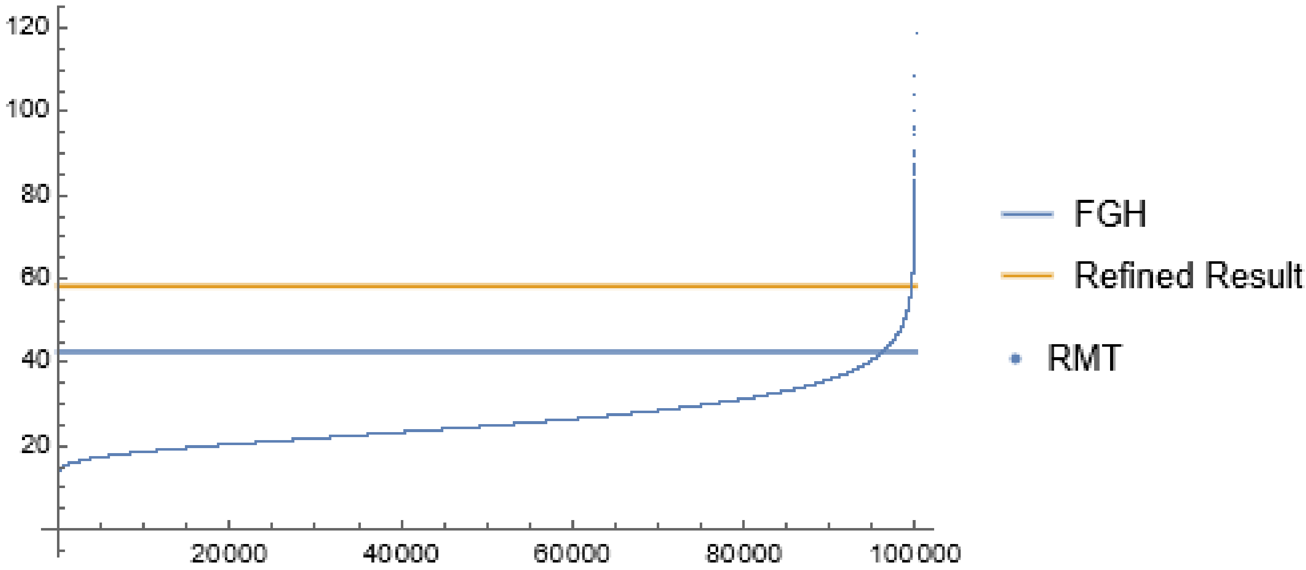


Figure 7: Listplot of T valuations of CUE matrices of dimension $[\log T]$, where $T = 100,000$ and the characteristic polynomials are evaluated at $r = 50$ equally spaced points around the unit circle. The horizontal blue line represents the original conjecture of Farmer, Gonek and Hughes, while the orange line represents our proposed refined result.

```
  Li[[j]] = Max[Table[f[k], {k,1,r}]];];
  A[[i]] = Max[Li]]

B = Sort[A, Less];

Show[Plot[Exp[Sqrt[(1/b)*Log[T]*Log[Log[T]] + (1/b)*Log[T]*Log[Log[Log[T]]]
- (2/b)*Log[T]*((5/2)*Log[2] + (1/2)*Log[2/b] - 3/2)]], {x,0,Length[B]},
PlotLegends -> {"Proposed Result"}], ListPlot[B, PlotLegends -> {"RMT"}],
PlotRange -> Full, AxesOrigin -> True]
```

It should be noted that with the code for the circular symplectic ensemble, a slight adjustment has been made: rather than study the absolute value of the characteristic polynomial, here we study the absolute value of the square root of the characteristic polynomial, and we outline here why we are doing this.

For the results of Keating and Snaith, they studied the characteristic polynomial of an $N \times N$ matrix taken from one of the circular ensembles, giving the result seen earlier:

$$
\begin{aligned}
M_N(\beta, s) &= \mathbb{E}\left[|\Lambda_{U,\beta}(\theta)|^s\right] \\
&= \frac{1}{(2\pi)^N Z_{N,\beta}} \int_0^{2\pi} \cdots \int_0^{2\pi} \mathrm{d}\theta_1 \ldots \mathrm{d}\theta_N \cdot \prod_{1\leq j<m\leq N} \left|e^{i\theta_j} - e^{i\theta_m}\right|^\beta \times \left|\prod_{p=1}^{N} \left(1 - e^{i(\theta_p - \theta)}\right)\right|^s \\
&= \prod_{j=0}^{N-1} \frac{\Gamma(1 + j\beta/2)\Gamma(1 + s + j\beta/2)}{(\Gamma(1 + s/2 + j\beta/2))^2}.
\end{aligned}
$$

However, when one uses the in-built command CircularSymplecticMatrixDistribution[N] in Wolfram Mathematica, this generates a $2N \times 2N$ matrix as opposed to an $N \times N$ matrix. Fortunately, a simple observation allows us to obtain the required results.

For $2N \times 2N$ CSE matrices, each eigenvalue occurs with multiplicity two, meaning that there are N eigenvalues each of which is twice accounted for. If one computes the characteristic polynomial for this matrix, this generates a characteristic polynomial of degree N raised to the power of 2.

Keating and Snaith's work centred on a $N \times N$ matrix with N distinct eigenvalues, and to ensure that our numerics align with the theory discussed in Section 3.4.1 we need to take the square root of the characteristic polynomial in our code. The results for this are given below.

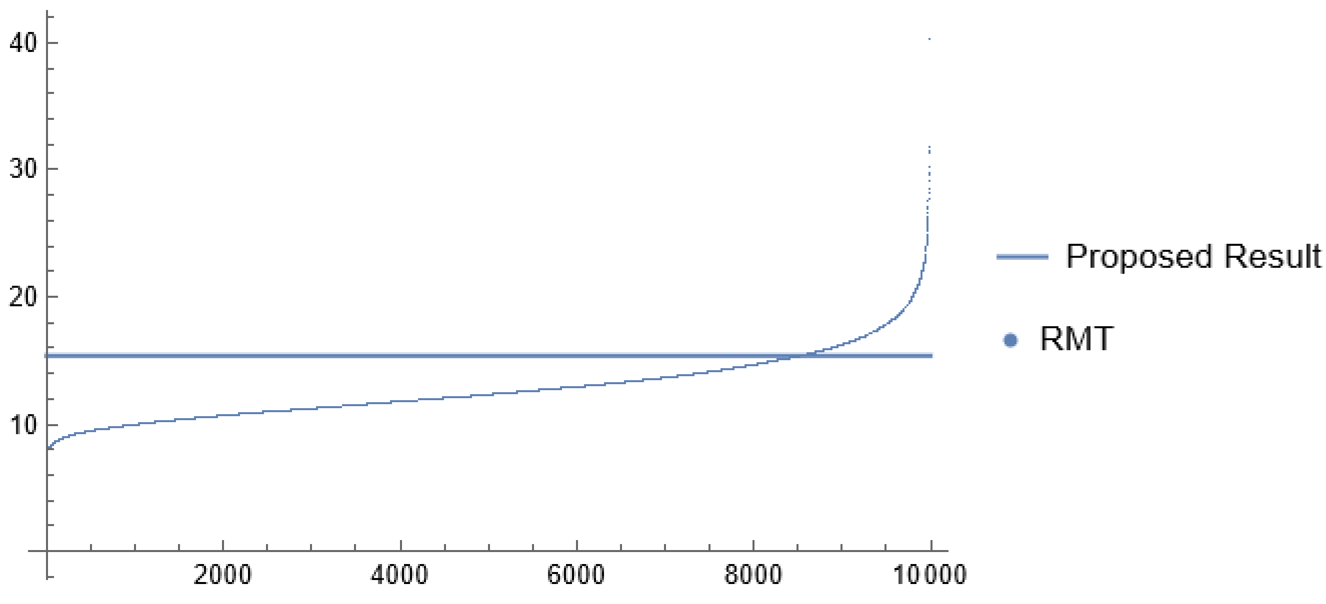


Figure 8: Listplot of T valuations of CSE matrices of dimension $[\log T]$, where $T = 10,000$ and the characteristic polynomials are evaluated at $r = 50$ equally spaced points around the unit circle. The horizontal blue line represents the conjecture of Farmer, Gonek and Hughes, while the dotted line represents the Random Matrix data.

Looking at the result in Figure 8 we find that after taking the square root of the characteristic polynomial in our code, the numerics reasonably match the result in

Theorem 3.2 with M and N as above.

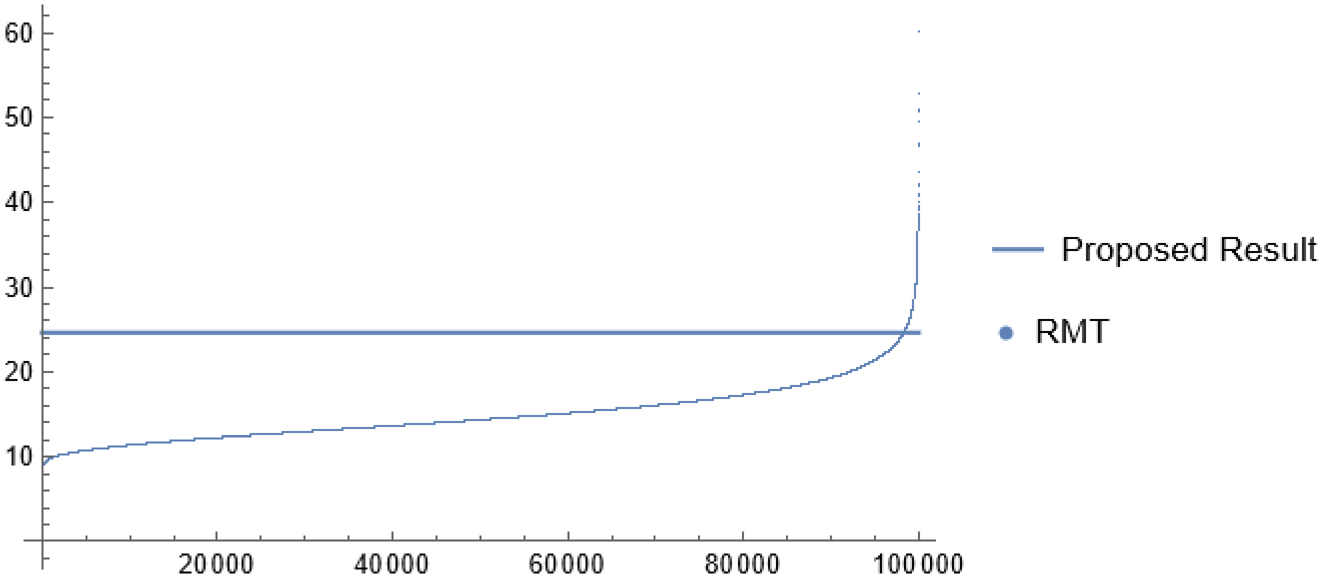


Figure 9: Listplot of T valuations of CSE matrices of dimension $[\log T]$, where $T = 10000$ and the characteristic polynomials are evaluated at $r = 50$ equally spaced points around the unit circle. The horizontal blue line represents the conjecture of Farmer, Gonek and Hughes, while the dotted line represents the Random Matrix data.

Looking now at the data in Figure 9 we arrive at the same conclusion as prior ensembles, in which the increase in T leads to a larger proportion of the data being smaller than the refined maximum, thus aligning with the theory.

7.2 Numerical Results for P_X

We now focus our attention on the numerical results for P_X (resp. R_X). For this, we take a fairly direct approach.

We take a collection of evenly spaced points - of spacing size m - in the interval $[0, T]$ and evalute $|P_X(\frac{1}{2}+it)|$ at these points. We then numerically order these evaluations and plot these against the original estimate of Farmer, Gonek and Hughes, as well as our the consequential result for the maximum in this regime,

$$\max_{t\in[0,T]} \left|P_X\left(\frac{1}{2}+it\right)\right| = \exp\left(\sqrt{\frac{1}{2}\log T\log\log T - \log T\log\log\log T + o(\log T\log\log\log T)}\right).$$

The code which was implemented for this is as follows:

```
T = 10000;
m = 0.01;
PX[t_] := Exp[Sum[1/(Prime[i]^(1/2 + I*t)),
          {i,1,PrimePi[Exp[Sqrt[Log[T]]]]}]];
```

```
V = Range[0,T,m];
A = Abs[PX[V]];
B = Sort[A,Less];

Show[Plot[Exp[Sqrt[(1/2)*Log[T]*Log[Log[T]]]],
 Exp[Sqrt[(1/2)*Log[T]*Log[Log[T]] - Log[T]*Log[Log[Log[T]]]]],
 {x,0,Length[V]}, PlotLegends -> {"FGH", "Refined Result"}],
 ListPlot[B,PlotLegends -> {Subscript[P,"X"]}], PlotRange -> All,
 AxesOrigin -> True]
```

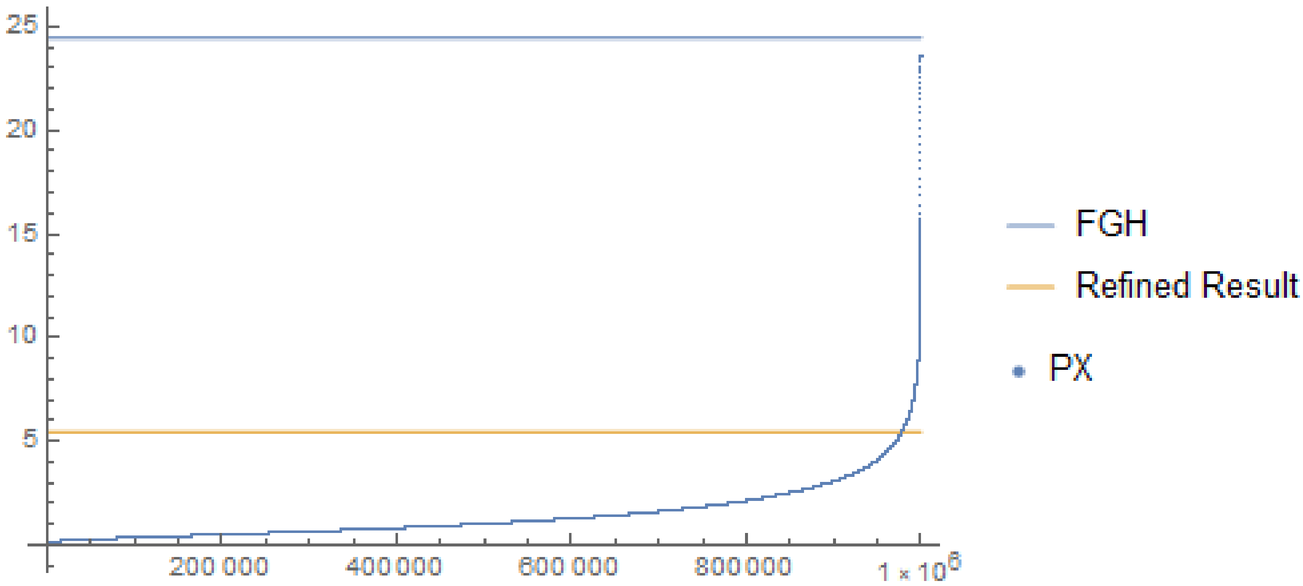


Figure 10: Listplot of $1,000,000$ evaluations of $|P_X(\frac{1}{2}+it)|$ for $T = 10,000$, $m = 0.01$ against the estimate of Farmer, Gonek and Hughes as well as our refined estimate.

In the first of these plots (Figure 10) we see that the refined result exceeds $> 95\%$ of points, and that for the list of points we've included here none of them exceed the original conjecture of Farmer, Gonek and Hughes.

If we then compare the results of Figure 11 against those of Figure 10, we see that the percentage of points exceeding our refined result is smaller, suggesting that as we consider larger and larger T our results are aligning with our conjecture (provided we take enough points in the interval $[0, T]$).

7.3 Numerical Results for Zeta

With the results for both Z_X and P_X computed and seen to be in agreement with our theoretical results, we now turn to computing numerical results for zeta. This will help determine whether our results from Chapter 4.4 are reasonable.

For these numerics, we implement the following code, computed in Wolfram Mathematica Version 13.1:

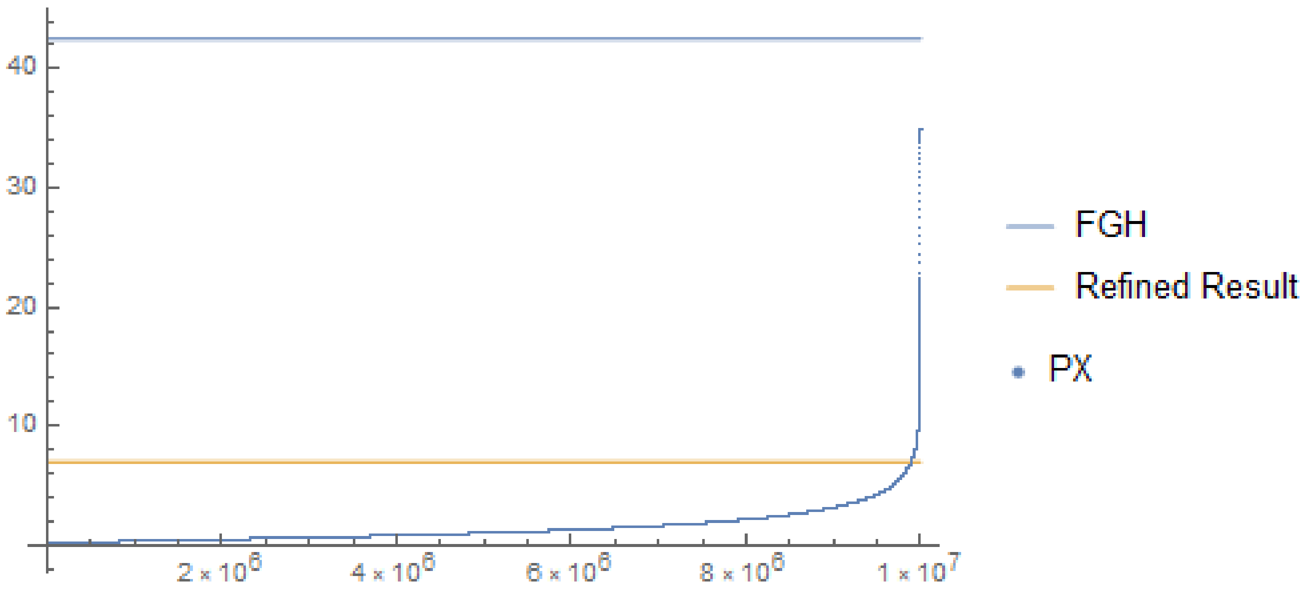


Figure 11: Listplot of 10, 000, 000 evaluations of $|P_X(\frac{1}{2}+it)|$ for $T = 100,000$, $m = 0.01$ against the estimate of Farmer, Gonek and Hughes as well as our refined estimate.

```
T = 10000;
m = 0.01;
v = Range[0,T,m];

w = Abs[Zeta[1/2+I*v]];

z = Sort[w, Less];

Show[Plot[Exp[Sqrt[(1/2)*Log[T]*Log[Log[T]]]],
   Exp[Sqrt[(1/2)*Log[T]*Log[Log[T]] + (1/2)*Log[T]*
       Log[Log[Log[T]]] - (1/2)*Log[T]*(5*Log[2]-3)]],
   Exp[Sqrt[(1/2)*Log[T]*Log[Log[T]] - Log[T]*Log[Log[Log[T]]]]],
   Exp[Sqrt[(1/2)*Log[T]*Log[Log[T]] - (1/2)*Log[T]*
       Log[Log[Log[T]]]]], {x,0,Length[v]},
  PlotLegends -> {"FGH",Subscript[Z,X], Subscript[P,X],"Convolution"}],
  ListPlot[z,PlotLegends -> {Zeta[1/2 + it]}],
  PlotRange -> All, AxesOrigin -> True]
```

These results are straightforward, but we nevertheless summarise our methods. As we are looking for the maximum of $|\zeta(\frac{1}{2}+it)|$ for t in $[0, T]$ we take a collection of equally spaced points in this interval, evaluate these and order them numerically. We then plot these points against the initial conjecture of Farmer, Gonek and Hughes, as well as our refined conjecture.

Looking at these results for $T = 10,000$ (Figure 12) we see that the proposed maximum stemming from the product over zeros Z_X significantly exceeds the max-

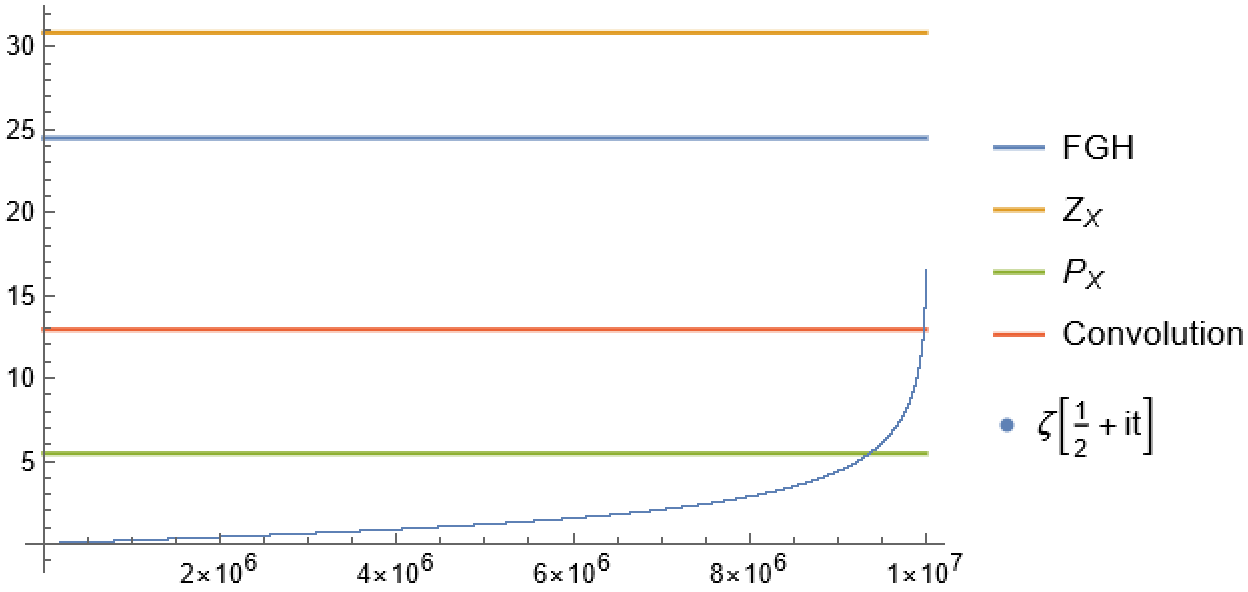


Figure 12: Listplot of $|\zeta(\frac{1}{2}+it)|$ for points $t \in [0, T]$ for $T = 10,000$ with spacing $m = 0.001$ between points. The horizontal lines correspond to the proposed maxima in each of the three regimes, as well as the conjectured maximum of Farmer, Gonek and Hughes.

imum from our data set. The same applies to the result of Farmer, Gonek and Hughes here.

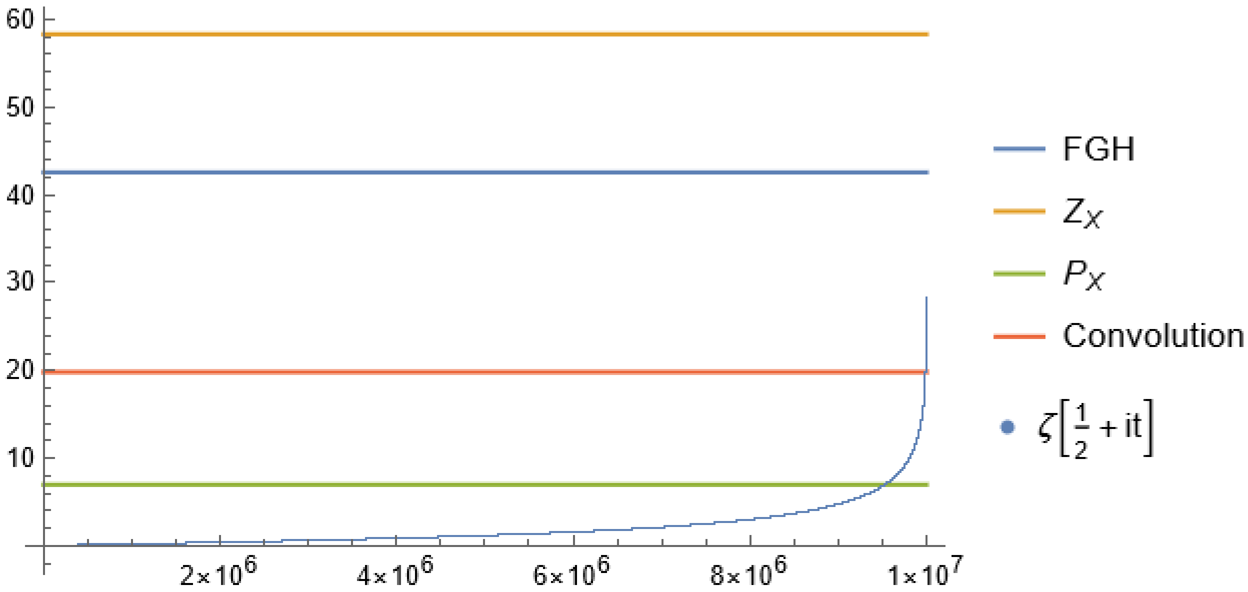


Figure 13: Listplot of $|\zeta(\frac{1}{2}+it)|$ for points $t \in [0, T]$ for $T = 100,000$ with spacing $m = 0.01$ between points. The horizontal lines correspond to the proposed maxima in each of the three regimes, as well as the conjectured maximum of Farmer, Gonek and Hughes.

Turning our attention to Figure 13 we see little difference in what is presented when compared to Figure 12. Given that the maximum of the zeta function is shown in Figures 12 and 13 to lie between the conjectured results for the maximum of Z_X and the zeta function (using the convolution method in Chapter 4.4),

$$\exp\left(\sqrt{\frac{1}{2}\log T\log\log T+\frac{1}{2}\log T\log\log\log T+o(\log T\log\log\log T)}\right),$$

$$\exp\left(\sqrt{\frac{1}{2}\log T\log\log T-\frac{1}{2}\log T\log\log\log T+o(\log T\log\log\log T)}\right),$$

this suggests that the constant c in our conjectured maximum in Chapter 6.2 lies in the range $-\frac{1}{2}<c<\frac{1}{2}$, with the data in Figure 13 placing it closer to $-\frac{1}{2}$.

A Consistency Check

We noted here that the results for P_X and Z_X are consistent, but what do we mean when we discuss consistency for our results?

Consider the following expression first computed in Chapter 3.4:

$$\begin{aligned}\mathbb{P}\left(\log\max_{\theta}|\Lambda_{U,\beta}(\theta)| \geq N^{\mu}x\right) = \exp\Bigg(&-\frac{\beta x^2 N^{2\mu}}{2(1-\mu)\log N} + \frac{\beta x^2 N^{2\mu}}{2(1-\mu)^2\log^2 N}\log\log N \\ &- \frac{\beta x^2 N^{2\mu}}{2(1-\mu)^2\log^2 N}\left(\log x + 2\log 2 - \frac{3}{2}\right) \\ &+ \frac{\beta x^2 N^{2\mu}}{2(1-\mu)^2\log^2 N}\log(1-\mu) + o\left(\frac{N^{\mu}}{(1-\mu)^2\log^2 N}\right)\Bigg).\end{aligned}$$

Making a change of variable $x \to cx$ as well as the change of variable $\mu \to \mu - \frac{\log c}{\log N}$ results in the expression $N^{\mu}x$ on the left-hand side of this expression being unchanged. With the above equation being an equality we should therefore expect that the right-hand side is unchanged also. So when talking about equations being consistent, what we mean is that the refined Large Deviations results we have computed are consistent under a (suitable) change of variable.

If we make the above changes of variable, we find that $N^{2\mu} \to N^{2\mu}/c^2$, and $x^2 \to c^2x^2$. Thus the factors of c cancel in the numerator. For the denominator of terms in this expression, we have that

$$\begin{aligned}\frac{1}{(1-\mu)\log N} \to \frac{1}{(1-\mu+\frac{\log c}{\log N})\log N} &= \frac{1}{(1-\mu)\log N[1+\frac{\log c}{(1-\mu)\log N}]} \\ &= \frac{1}{(1-\mu)\log N}\left[1 - \frac{\log c}{(1-\mu)\log N} + \cdots\right].\end{aligned}$$

Plugging these into the above expression gives

$$
\begin{aligned}
\mathbb{P}\left(\log\max_\theta|\Lambda_{U,\beta}(\theta)|\geq N^\mu x\right) = \exp\Bigg(& -\frac{\beta x^2N^{2\mu}}{2(1-\mu)\log N}\left[1-\frac{\log c}{(1-\mu)\log N}+\cdots\right] \\
& + \frac{\beta x^2N^{2\mu}}{2(1-\mu)^2\log^2 N}\log\log N\left[1-\frac{\log c}{(1-\mu)\log N}+\cdots\right]^2 \\
& + \frac{3\beta x^2N^{2\mu}}{4(1-\mu)^2\log^2 N}\left[1-\frac{\log c}{(1-\mu)\log N}+\cdots\right]^2 \\
& - \frac{\beta x^2N^{2\mu}}{(1-\mu)^2\log^2 N}\log 2\left[1-\frac{\log c}{(1-\mu)\log N}+\cdots\right]^2 \\
& - \frac{\beta x^2N^{2\nu}}{(1-\mu)^2\log^2 N}\log cx\left[1-\frac{\log c}{(1-\mu)\log N}+\cdots\right]^2 \\
& + \frac{\beta x^2N^{2\mu}}{2(1-\mu)\log^2 N}\left[1-\frac{\log c}{(1-\mu)\log N}+\cdots\right]^2\log\left(1-\mu+\frac{\log c}{\log N}\right) \\
& + o\left(\frac{N^{2\mu}}{(1-\mu)^2\log^2 N}\right)\Bigg).
\end{aligned}
$$

Expanding and collecting terms, we find that this is equal to

$$
\begin{aligned}
\mathbb{P}\left(\log\max_\theta|\Lambda_{U,\beta}(\theta)|\geq N^\mu x\right) = \exp\Bigg(& -\frac{\beta x^2N^{2\mu}}{2(1-\mu)\log N}+\frac{\beta x^2N^{2\mu}}{2(1-\mu)^2\log^2 N}\log\log N \\
& + \frac{3\beta x^2N^{2\mu}}{4(1-\mu)^2\log^2 N}-\frac{\beta x^2N^{2\mu}}{(1-\mu)^2\log^2 N}\log 2 \\
& - \frac{\beta x^2N^{2\mu}}{(1-\mu)^2\log^2 N}\log x+\frac{\beta x^2N^{2\mu}}{(1-\mu)^2\log^2 N}\log\left(1-\mu+\frac{\log c}{\log N}\right) \\
& + o\left(\frac{N^{2\mu}}{(1-\mu)^2\log^2 N}\right)\Bigg).
\end{aligned}
$$

For the final term, we split the logarithm as $\log(1-\mu+\frac{\log c}{\log N}) = \log(1-\mu)+\log(1+\frac{\log c}{(1-\mu)\log N})$, at which point we can use the series expansion of the logarithm to give

$$
\begin{aligned}
\frac{x^2N^{2\mu}}{(1-\mu)^2\log^2 N}\log\left(1-\mu+\frac{\log c}{\log N}\right) &= \frac{x^2N^{2\mu}}{(1-\mu)^2\log^2 N}\log(1-\mu) \\
&\quad + \frac{x^2N^{2\mu}}{(1-\mu)^2\log^2 N}\left[-\frac{\log c}{(1-\mu)\log N}+\cdots\right] \\
&= \frac{x^2N^{2\mu}}{(1-\mu)^2\log^2 N}\log(1-\mu)+\frac{x^2N^{2\mu}}{(1-\mu)^3\log^3 N}\log c \\
&\quad + \cdots
\end{aligned}
$$

From this, we see that the sub-leading term and all subsequent lower order terms are absorbed by the o-term in our large deviations result. Therefore we see that the formula is consistent under a change of variable.

We now verify that the formula for R_X computed in Chapter 4.3 is consistent; Recall the result from Corollary 4.2: for $0 < \mu < 1$,

$$\log \mathbb{P}\left((R_X \geq xX^{\mu/2}\log(1/\mu)\right) = -x^2 X^{\mu}\log(1/\mu) + O\left(\frac{X^{\mu}}{\mu \log X}\right).$$

Suppose we make the change of variable $x \to cx$. To ensure that the left-hand side remains unchanged, we need to find the change of variable μ' such that

$$X^{\mu'/2}\log(1/\mu') = \frac{1}{c}X^{\mu/2}\log(1/\mu).$$

Solving this gives the solution

$$\mu' = \mu - \frac{2\log c}{\log X} - \frac{2}{\log X}\left[\frac{1}{\log(1/\mu)}\log\left(1 + \frac{2\log c}{\mu \log X} + \cdots\right) + \cdots\right].$$

With this choice the left-hand side of the result is unchanged. We therefore expect that the right-hand side is unchanged also when these changes of variable are made. If we consider then the right-hand side of Theorem 4.2 we find that

$$\begin{aligned}
-x^2X^{\mu}\log(1/\mu) \to -c^2x^2X^{\mu-2\log c/\log X+\cdots}\left(\log(1/\mu) - \log\left(1 - \frac{2\log c}{\mu\log X} + \cdots\right)\right) \\
= -c^2x^2X^{\mu}\frac{1}{c^2}\exp\left(-\frac{2}{\log(1/\mu)}\log\left(1 + \frac{2\log c}{\mu\log X} + \cdots\right)\right) \\
\times\left[\log(1/\mu) - \log\left(1 - \frac{2\log c}{\mu\log X} + \cdots\right)\right]
\end{aligned}$$

If one simplifies the above expression and performs a Taylor expansion of the exponential term, it can be seen that any subleading terms are absorbed by the O-term. We therefore conclude that our calculations for P_X are consistent under a change of variable.

With these being the key large deviations results, we conclude that our large deviations results are consistent.

B Barnes G-Function

The Barnes G-function was introduced in [8] as the function which satisfies the difference equation

$$G(N+1) = \Gamma(N)G(N).$$

Here, $\Gamma(z)$ is the Euler-gamma function. The Barnes G-function can be expressed via the formula

$$G(N+1) = (2\pi)^{N/2} \exp\left(-\frac{N+(\gamma+1)N^2}{2}\right) \prod_{k=1}^{\infty} \left[\left(1+\frac{N}{k}\right)^k \exp\left(-N+\frac{N^2}{2k}\right)\right].$$

In particular when N is real, large and positive, we have the following asymptotic expansion for the logarithm:

$$\begin{aligned}\log G(N+k+1) = &\frac{N+k}{2}\log 2\pi + \zeta'(-1) - \frac{3N^2}{4} - kN + \left(\frac{N^2}{2} - \frac{1}{12} + \frac{k^2}{2} + kN\right)\log N \\ &+ O\left(\frac{1}{N}\right).\end{aligned}$$

In this expression, $|k|$ is small. For the purpose of our findings we consider the case where $|k|$ is large and dependent on N(albeit smaller than N).

Looking then at $|k| < N$ and treating $N+k$ as N in the above asymptotic expression, we have

$$\begin{aligned}\log G(N+k+1) = &\frac{N+k}{2}\log 2\pi + \zeta'(-1) - \frac{3(N+k)^2}{4} + \left(\frac{(N+k)^2}{2} - \frac{1}{12}\right)\log(N+k) \\ &+ O\left(\frac{1}{N}\right) \\ = &\frac{N+k}{2}\log 2\pi + \zeta'(-1) - \frac{3N^2}{4} - \frac{3kN}{2} - \frac{3k^2}{4} \\ &+ \left(\frac{N^2}{2} - \frac{1}{12} + \frac{k^2}{2} + kN\right)\log(N+k) + O\left(\frac{1}{N}\right).\end{aligned}$$

For the logarithm term, we know that N is the dominant term, so rewriting this term and using the Taylor expansion of $\log(1+x)$ gives

$$
\begin{aligned}
\log G(N+k+1) &= \frac{N+k}{2}\log 2\pi + \zeta'(-1) - \frac{3N^2}{4} - \frac{3kN}{2} - \frac{3k^2}{4} \\
&\quad + \left(\frac{N^2}{2} - \frac{1}{12} + \frac{k^2}{2} + kN\right)\log N + \left(\frac{N^2}{2} - \frac{1}{12} + \frac{k^2}{2} + kN\right)\left(\frac{k}{N}\right. \\
&\quad \left. - \frac{k^2}{2N^2} + \frac{k^3}{3N^3} + \cdots\right) + O\left(\frac{1}{N}\right) \\
&= \frac{N+k}{2}\log 2\pi + \zeta'(-1) - \frac{3N^2}{4} - \frac{3kN}{2} - \frac{3k^2}{4} \\
&\quad + \left(\frac{N^2}{2} - \frac{1}{12} + \frac{k^2}{2} + kN\right)\log N + \left(\frac{kN}{2} - \frac{k}{12N} + \frac{k^3}{2N} + k^2 - \frac{k^2}{4}\right. \\
&\quad \left. - \frac{k^2}{24N^2} + \frac{k^3}{6N} - \frac{k^3}{36N^3} + \frac{k^5}{6N^3} + \frac{k^4}{3N^2} + \cdots\right) \\
&\quad + O\left(\frac{1}{N}\right).
\end{aligned}
$$

and this simplifies to give the expression

$$
\begin{aligned}
\log G(N+k+1) &= \frac{N+k}{2}\log 2\pi + \zeta'(-1) - \frac{3N^2}{4} - kN + \left(\frac{N^2}{2} - \frac{1}{12} + \frac{k^2}{2} + kN\right)\log N \\
&\quad + O\left(\frac{k^3}{N}\right)
\end{aligned}
$$

where the adjusted O-term is to account for the largest order term stemming from the Taylor expansion of the logarithm. We utilise this result to compute refined large deviations results for the characteristic polynomial (and subsequently Z_X) in Chapter 4.1.

References

[1] L.-P. Arguin, D. Belius, and P. Bourgade. Maximum of the characteristic polynomial of random unitary matrices. *Comm. Math. Phys.*, 349(2):703–751, 2017.

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

[3] L.-P. Arguin, D. Belius, and A. J. Harper. Maxima of a randomized Riemann zeta function, and branching random walks. *Ann. Appl. Probab.*, 27(1):178–215, 2017.

[4] L.-P. Arguin, P. Bourgade, and M. Radziwiłł. The Fyodorov-Hiary-Keating conjecture. I., 2020.

[5] L.-P. Arguin, G. Dubach, and L. Hartung. Maxima of a random model of the riemann zeta function over intervals of varying length, 2021.

[6] L.-P. Arguin, F. Ouimet, and M. Radziwiłł. Moments of the Riemann zeta function on short intervals of the critical line, 2021.

[7] R. Balasubramanian and K. Ramachandra. On the frequency of Titchmarsh's phenomenon for $\zeta(s)$. VI. *Acta Arith.*, 53(4):325–331, 1990.

[8] E. W. Barnes. The theory of the G–function. *Quart. J. Pure Appl. Math.*, 31:264–314, 1899.

[9] A. Bondarenko and K. Seip. Large greatest common divisor sums and extreme values of the Riemann zeta function. *Duke Math. J.*, 166(9):1685–1701, 2017.

[10] M. D. Bramson. Maximal displacement of branching Brownian motion. *Comm. Pure Appl. Math.*, 31(5):531–581, 1978.

[11] H. M. Bui, B. Conrey, and M. P. Young. More than 41% of the zeros of the zeta function are on the critical line. *Acta Arith.*, 150(1):35–64, 2011.

[12] R. Chhaibi, T. Madaule, and J. Najnudel. On the maximum of the CβE field. *Duke Math. J.*, 167(12):2243–2345, 2018.

[13] J. B. Conrey. More than two fifths of the zeros of the Riemann zeta function are on the critical line. *J. Reine Angew. Math.*, 399:1–26, 1989.

[14] J. B. Conrey and A. Ghosh. A conjecture for the sixth power moment of the Riemann zeta-function. *Internat. Math. Res. Notices*, (15):775–780, 1998.

[15] J. B. Conrey and S. M. Gonek. High moments of the Riemann zeta-function. *Duke Math. J.*, 107(3):577–604, 2001.

[16] M. Coram and P. Diaconis. New tests of the correspondence between unitary eigenvalues and the zeros of Riemann's zeta function. *J. Phys. A*, 36(12):2883–2906, 2003. Random matrix theory.

[17] C.-J. De la Vallée Poussin. *Recherches analytiques sur la théorie des nombres premiers*. Hayez, Imprimeur de l'Académie royale de Belgique, 1897.

[18] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.

[19] F. J. Dyson. Statistical theory of the energy levels of complex systems. I. *J. Mathematical Phys.*, 3:140–156, 1962.

[20] H. M. Edwards. *Riemann's zeta function*. Dover Publications, Inc., Mineola, NY, 2001. Reprint of the 1974 original [Academic Press, New York; MR0466039 (57 #5922)].

[21] D. W. Farmer, S. M. Gonek, and C. P. Hughes. The maximum size of L-functions. *Journal für die reine und angewandte Mathematik (Crelle's Journal)*, 609:215–236, 2007. math.NT/0506218.

[22] S. Feng. Zeros of the Riemann zeta function on the critical line. *J. Number Theory*, 132(4):511–542, 2012.

[23] Y. V. Fyodorov, G. A. Hiary, and J. P. Keating. Freezing transition, characteristic polynomials of random matrices, and the Riemann zeta function. *Physical review letters*, 108(17):170601, 2012.

[24] Y. V. Fyodorov and J. P. Keating. Freezing transitions and extreme values: random matrix theory, and disordered landscapes. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 372(2007):20120503, 32, 2014.

[25] D. A. Goldston. On the function $S(T)$ in the theory of the Riemann zeta-function. *J. Number Theory*, 27(2):149–177, 1987.

[26] S. M. Gonek, C. P. Hughes, and J. P. Keating. A hybrid Euler-Hadamard product for the Riemann zeta function. *Duke Math. J.*, 136:507–549, 2007. math.NT/0511182.

[27] A. Granville and K. Soundararajan. Extreme values of $|\zeta(1+it)|$. In *The Riemann zeta function and related themes: papers in honour of Professor K. Ramachandra*, volume 2 of *Ramanujan Math. Soc. Lect. Notes Ser.*, pages 65–80. Ramanujan Math. Soc., Mysore, 2006.

[28] J. Hadamard. Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques. *Bull. Soc. Math. France*, 24:199–220, 1896.

[29] G. H. Hardy and J. E. Littlewood. Contributions to the theory of the Riemann zeta-function and the theory of the distribution of primes. *Acta Math.*, 41(1):119–196, 1916.

[30] G. H. Hardy and J. E. Littlewood. The zeros of Riemann's zeta-function on the critical line. *Math. Z.*, 10(3-4):283–317, 1921.

[31] A. J. Harper. A note on the maximum of the Riemann zeta function, and log-correlated random variables, 2013.

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

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

[34] A. E. Ingham. Mean-value theorems in the theory of the Riemann zeta-function. *Proceedings of the London Mathematical Society*, 2(1):273–300, 1928.

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

[36] T. Kotnik. Computational estimation of the order of $\zeta(\frac{1}{2}+it)$. *Math. Comp.*, 73(246):949–956, 2004.

[37] N. Levinson. More than one third of zeros of Riemann's zeta-function are on $\sigma = 1/2$. *Advances in Math.*, 13:383–436, 1974.

[38] E. S. Meckes. *The random matrix theory of the classical compact groups*, volume 218 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2019.

[39] M. L. Mehta. *Random matrices and the statistical theory of energy levels*. Academic Press, New York-London, 1967.

[40] H. L. Montgomery. The pair correlation of zeros of the zeta function. pages 181–193, 1973.

[41] H. L. Montgomery. Extreme values of the Riemann zeta function. *Comment. Math. Helv.*, 52(4):511–518, 1977.

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

[43] C. Niculescu and L.-E. Persson. *Convex functions and their applications*, volume 23. Springer, 2006.

[44] A. M. Odlyzko. The 10^{22}-nd zero of the Riemann zeta function. In *Dynamical, spectral, and arithmetic zeta functions (San Antonio, TX, 1999)*, volume 290 of *Contemp. Math.*, pages 139–144. Amer. Math. Soc., Providence, RI, 2001.

[45] E. Paquette and O. Zeitouni. The maximum of the CUE field. *Int. Math. Res. Not. IMRN*, (16):5028–5119, 2018.

[46] K. Pratt, N. Robles, A. Zaharescu, and D. Zeindler. More than five-twelfths of the zeros of ζ are on the critical line. *Res. Math. Sci.*, 7(2):Paper No. 2, 74, 2020.

[47] N. Robles, A. Roy, and A. Zaharescu. Twisted second moments of the Riemann zeta-function and applications. *J. Math. Anal. Appl.*, 434(1):271–314, 2016.

[48] A. Selberg. On the zeros of Riemann's zeta-function. *Skr. Norske Vid.-Akad. Oslo I*, 1942(10):59, 1942.

[49] S. Serfaty. *Coulomb gases and Ginzburg-Landau vortices*. Zurich Lectures in Advanced Mathematics. European Mathematical Society (EMS), Zürich, 2015.

[50] J. A. Shohat and J. D. Tamarkin. *The problem of moments*, volume 1. American Mathematical Society (RI), 1950.

[51] K. Soundararajan. Extreme values of zeta and L-functions. *Math. Ann.*, 342(2):467–486, 2008.

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

[53] E. C. Titchmarsh. *The theory of the Riemann zeta-function*. The Clarendon Press, Oxford University Press, New York, second edition, 1986. Edited and with a preface by D. R. Heath-Brown.

[54] E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)*, 62:548–564, 1955.

[55] E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. II. *Ann. of Math. (2)*, 65:203–207, 1957.