

EXTREMA OF LOG-CORRELATED RANDOM VARIABLES: PRINCIPLES AND EXAMPLES

LOUIS-PIERRE ARGUIN

ABSTRACT. These notes were written for the mini-course *Extrema of log-correlated random variables: Principles and Examples* at the Introductory School held in January 2015 at the Centre International de Rencontres Mathématiques in Marseille. There have been many advances in the understanding of the high values of log-correlated random fields from the physics and mathematics perspectives in recent years. These fields admit correlations that decay approximately like the logarithm of the inverse of the distance between index points. Examples include branching random walks and the two-dimensional Gaussian free field. In this paper, we review the properties of such fields and survey the progress in describing the statistics of their extremes. The branching random walk is used as a guiding example to prove the correct leading and subleading order of the maximum following the multiscale refinement of the second moment method of Kistler. The approach sheds light on a conjecture of Fyodorov, Hiary & Keating on the maximum of the Riemann zeta function on an interval of the critical line and of the characteristic polynomial of random unitary matrices.

CONTENTS

1. Introduction	1
1.1. Statistics of extremes	3
1.2. Log-correlated fields	6
1.3. Relations to statistical physics	8
2. Examples and general properties	10
2.1. Branching random walk	10
2.2. 2D Gaussian free field	12
3. Order of the maximum	15
3.1. Leading order of the maximum	15
3.2. Subleading order of the maximum	19
4. Universality classes of log-correlated fields	23
4.1. Maximum of the Riemann zeta function on an interval	23
4.2. Maximum of the characteristic polynomial of random unitary matrices	26
References	29

1. INTRODUCTION

There has been tremendous progress recently in the understanding of the extreme value statistics of stochastic processes whose variables exhibit strong correlations. The purpose

2000 *Mathematics Subject Classification.* 60G70, 60G60.

Key words and phrases. Extreme Value Theory, Gaussian Fields, Branching Random Walk.

L.-P. A. is supported by the NSF grant DMS 1513441, the PSC-CUNY Research Award 68784-00 46, and partially by a NSERC Discovery grant and FQRNT *Nouveaux chercheurs* grant.

of this paper is to survey the recent progress in this field for processes with logarithmically decaying correlations. The material has been developed for the mini-course *Extrema of log-correlated random variables: Principles and Examples* at the Introductory School held in January 2015 at the Centre International de Rencontres Mathématiques in Marseille for the trimester *Disordered systems, Random spatial processes and Applications* of the Institut Henri-Poincaré.

The study of extreme values of stochastic processes goes back to the early twentieth century. The theory was developed in the context of independent or weakly correlated random variables. The first book regrouping the early advances in the field is the still relevant *Statistics of Extremes* by E.J. Gumbel [43] published in 1958. Gumbel credits the first work on extreme values to Bortkiewicz in 1898, which is one of the first paper emphasizing the importance of Poisson statistics. After progress by Fisher and Tippett in understanding the limit law of the maximum of a collection of IID random variables, the theory culminated in 1943 with the classification theorem of Gnedenko [42] which showed that the distribution of the maximum of N independent and identically distributed (IID) random variables when properly recentered and rescaled can only be of three types: Fréchet, Weibull or Gumbel. Prior to this, von Mises had shown sufficient conditions for convergence to the three types. All in all, it took the theory sixty years between the first rigorous results and an essentially complete theory as it appears in Gumbel's review. There are now many excellent textbooks on extreme value statistics, see e.g. [19, 28, 49, 54].

The classical theory however does not apply to collections of random variables that exhibit strong correlations, i.e. correlations of the order of the variance. As Gumbel points out at the very beginning of *Statistics of Extremes*:

Another limitation of the theory is the condition that the observations from which the extremes are taken should be independent. This assumption, made in most statistical work, is hardly ever realized.

In the last thirty years, there have been important breakthroughs in mathematics and in physics to identify *universality classes* for the distributions of extreme values of strongly correlated stochastic processes. One such class that has attracted much interest in recent years is the class of *log-correlated fields*. Essentially, these are stochastic processes for which the correlations decay logarithmically with the distance. Even though the theory has not reached a complete status as in the IID case, there has been tremendous progress in identifying the distribution of extremes in this case and to develop rich heuristics that describe accurately the statistics of large values. There are already excellent survey papers on the subject by Bovier [20] and Zeitouni [57], all offering a different perspective on log-correlated fields. The present paper modestly aims at complementing the growing literature on the subject by focusing on the techniques to obtain precise estimates on the order of the maximum of log-correlated fields and their potential applications to seemingly unrelated problems: the maxima of the characteristic polynomials of random matrices and the maxima of the Riemann zeta function on an interval of the critical line. The approach we take is based on Kistler's multiscale refinement of the second moment method introduced in [46].

The presentation is organized as follows. In Section 1, we review the basic theory in the case of IID random variables and will introduce log-correlated fields in generality. Section 2 focuses on two important examples of log-correlated fields: *branching random walks* and *the two-dimensional Gaussian free field*. In particular, three properties of these

fields (and of log-correlated fields in general) are singled out as they play a crucial role in the analysis of extremes. In Section 3, we describe the general method of Kistler to prove fine asymptotics for the maximum of log-correlated fields [46]. Finally, in Section 4, we will explain how the method is expected to be applicable to study the high values of the characteristic polynomials of random matrix ensemble as well as the large values of the Riemann zeta function on an interval of the critical line.

Acknowledgements

I am grateful to Jean-Philippe Bouchaud, Pierluigi Contucci, Cristian Giardinà, Pierre Nolin, Vincent Vargas, and Vladas Sidoravicius for the organization of the trimester *Disordered systems, Random spatial processes and Applications* at the Institut Henri-Poincaré in the spring 2015. I also thank the *Centre International de Rencontres Mathématiques* and its staff for the hospitality and the organization during the Introductory School. This work would not have been possible without the numerous discussions with my collaborators on the subject: David Belius, Paul Bourgade, Anton Bovier, Adam J. Harper, Nicola Kistler, Olivier Zeitouni, and my students Samuel April, Frédéric Ouimet, Roberto Persechini, and Jean-Sébastien Turcotte.

1.1. Statistics of extremes. We are interested in the problem of describing the maxima of a stochastic process in the limit where there are a large number of random variables. For example, consider the process consisting of IID centered Gaussian random variables. One realization of this process for 2^{10} variables is depicted in Figure 1 with a particular choice of the variance. The maximum of the rugged landscape of this realization lies around 6. If we repeat this exercise for several realizations, we would observe that this is the case for every realization: the maximum has fluctuations roughly of order one around a deterministic value which is close to 6.

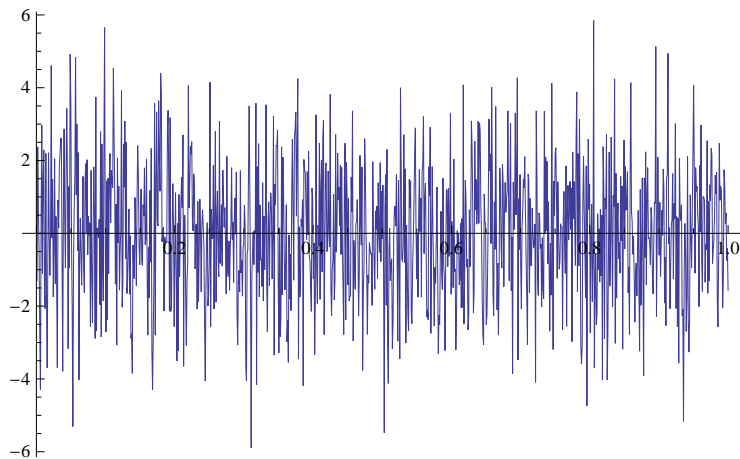


FIGURE 1. A realization of 2^{10} centered Gaussian random variables of variance $\frac{1}{2} \log 2^{10}$.

Now consider a drastically different stochastic process: let U be a random $2^{10} \times 2^{10}$ unitary matrix sampled uniformly from the unitary group and consider the modulus of the characteristic polynomial on the unit circle

$$P_U(\theta) = |\det(e^{i\theta} - U)| .$$

A realization of the logarithm of this quantity on the interval $[0, 2\pi]$ is given in Figure 2. This random process is of course qualitatively very different from the IID process above.

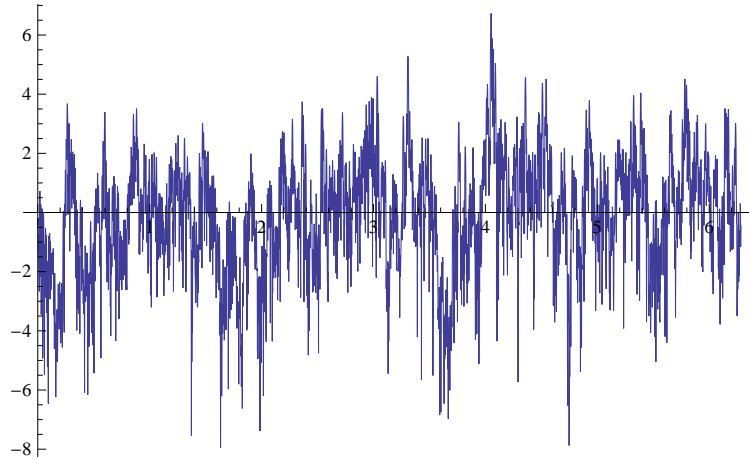


FIGURE 2. A realization of $P_U(\theta)$ on $[0, 2\pi]$ of a random unitary matrix U of size $2^{10} \times 2^{10}$.

It is not Gaussian and it is a continuous process with singularities at the eigenvalues of U . However, perhaps by coincidence, the maximum of the landscape lies also around 6. In fact, if we repeat the experiment, we would again observe that the maximum fluctuates around a deterministic value close to 6. If we were precise enough, we would see that the deterministic value is close to the one for IID variables but there is a slight discrepancy. The conjecture of Fyodorov-Hiary-Keating [39, 40] makes a detailed prediction for the maximum of $P_U(\theta)$. The distribution of the maximum is different in the limit of a large number of variables from the one of IID, though it is close. However, it is a perfect match with the distribution of log-correlated Gaussian field. This conjecture is still open (even to first order). However, the mechanism that would be responsible for this conjecture to hold is now quite clear. This is the goal of this review to illustrate the techniques for log-correlated fields in the forthcoming sections and their relations to $P_U(\theta)$ in Section 4.

Throughout the paper, the framework will be as follows. The random field will be indexed by a discrete space V_n with 2^n points. One important example is $V_n = \{-1, 1\}^n$ representing the configurations of n spins. In most examples, V_n will be equipped with a metric that we denote $d(v, v')$, $v, v' \in V_n$. We consider the random field

$$X(n) = \{X_v(n), v \in V_n\} .$$

defined on a probability space (Ω, P) . We will suppose without loss of generality that $E[X_v(n)]$ for all $v \in V_n$. In most examples, the variance will be the same for every $v \in V_n$. This is not true for the two-dimensional Gaussian free field because of the boundary effect. However, the variance is almost uniform for most points which ultimately does not affect the results, but complicate the analysis. With this in mind, we choose to parametrize the variance as

$$E[X_v(n)] = \sigma^2 n ,$$

for some σ^2 . This turns out to be the correct physical parametrization: it gives a a non-trivial limit for the free energy per spin in the limit $n \rightarrow \infty$, cf. Section 1.3.

Even though the analysis we present are not limited to Gaussian fields, it is convenient to present the ideas in this setting. There are many advantages to do so. First, the distribution of the field $X(n)$ is then entirely determined by the covariances

$$E[X_v(n)X_{v'}(n)] .$$

Second, estimates for the probability of large values for Gaussians are very precise: it is not hard to prove the Gaussian estimate for $X_v(n) \sim \mathcal{N}(0, \sigma^2 n)$,

$$(1) \quad P(X_v(n) > a) = \frac{(1 + o(1))}{\sqrt{2\pi}} \frac{\sqrt{\sigma^2 n}}{a} \exp\left(-\frac{a^2}{2\sigma^2 n}\right) \quad \text{for } a > \sqrt{\sigma^2 n}.$$

Finally, comparison results for Gaussian processes can be used in some instances to bound functionals of the processes above and below with functionals of processes with simpler correlation structures.

The central question of extreme value theory is to describe the functional

$$\max_{v \in V_n} X_v(n) \text{ in the limit } n \rightarrow \infty.$$

As for other global functionals of a stochastic process such as the sum, it is reasonable to expect universality results for the limit, i.e. results that depend mildly on the fine details of the distributions of $X(n)$. From a physics perspective, this question corresponds to describing the ground state energy of a system with state space V_n and energy $-X_v(n)$ for the state v . The system is said to be *disordered* because the energies depend on the realization of the sample space Ω . A realization of the process as depicted in Figures 1 and 2 from this point of view represents the energy landscape of the system for a given disorder.

With a perhaps naive analogy with the limit theorems for sum of IID random variables, one might expect a corresponding *law of large numbers* and *central limit theorems* for the maximum. More precisely, the problem is to:

(2)

$$\text{Find a recentering } a_n \text{ and rescaling } b_n \text{ such that } \max_{v \in V_n} \frac{X_v(n) - a_n}{b_n} \text{ converges in law.}$$

If these sequences exist, it suggests that there is a non-trivial random variable \mathcal{M} independent of n such that the approximate equality holds:

$$\max_{v \in V_n} X_v(n) \approx a_n + b_n \mathcal{M}.$$

In this regard, the recentering a_n is the term analogous to the law of large numbers whereas b_n sizes the magnitude of the fluctuations with distribution \mathcal{M} , thus belongs to the central limit theorem regime. Of course, a_n should be unique up to an additive constant and b_n , up to a multiplicative constant. We refer to a_n as the *order of the maximum*. We shall focus on general techniques to rigorously obtain the order of the maximum a_n for log-correlated fields in Section 3. The proofs of fluctuations are technically more involved, but the heuristics developed for the order of the maximum are crucial to the analysis at a more refined level.

The starting point of our analysis is the result when $X_v(n)$ are IID Gaussians with variance $\sigma^2 n$. In this case, the problem (2) is a simple exercise and has a complete answer.

Proposition 1. *Let $X(n) = (X_v(n), v = 1, \dots, 2^n)$ be IID centered Gaussians with variance $\sigma^2 n$. Then for $c = \sqrt{2 \log 2} \sigma$ and*

$$(3) \quad a_n = cn - \frac{1}{2} \frac{\sigma^2}{c} \log n, \quad b_n = 1,$$

we have

$$\lim_{n \rightarrow \infty} P \left(\max_{v \in V_n} X_v(n) \leq a_n + b_n x \right) = \exp(-C e^{-cx}) \quad , \quad \text{with } C = \frac{\sigma/c}{\sqrt{2\pi}}.$$

In other words, the maximum converges in distribution to a *Gumbel random variable* also known as *double-exponential*. The order of the maximum is linear in n at first order with velocity $c = \sqrt{2 \log 2} \sigma$. The subleading order is logarithmic in n . The proof of this is simple but is instructive for the case of log-correlated fields. The same type of arguments can be used for the *order statistics*, i.e. the joint distribution of first, second, third, etc maxima. It converges to a Poisson point process with exponential density.

Proof of Proposition 1. Clearly, since the random variables are IID, we have

$$P \left(\max_{v \in V_n} X_v(n) \leq a_n + b_n x \right) = \left(1 - \frac{2^n P(X_v(n) > a_n + b_n x)}{2^n} \right)^{2^n}.$$

This implies that the convergence of the distribution reduces to show that

$$2^n P(X_v(n) > a_n + b_n x) \text{ converges for an appropriate choice of } a_n \text{ and } b_n.$$

The right side is exactly the expected *number of exceedances*, that is

$$E[\mathcal{N}(y)], \quad \text{where } \mathcal{N}(y) = \#\{v \in V_n : X_v(n) > y\}$$

for $y = a_n + b_n x$. It is not hard to prove using the Gaussian estimate (1) that the choice $a_n = cn - \frac{1}{2} \frac{\sigma^2}{c} \log n$ and b_n ensures the convergence. In fact, the leading order cn lies in the large deviation regime for the distribution $X_v(n)$. It compensates exactly the *entropy* term 2^n . As for the subleading order $-\frac{\sigma^2}{2c}$, it balances out the fine asymptotics

$$\frac{\sigma^2 n}{a_n} = \frac{\sqrt{\sigma^2 n}}{cn(1 + o(1))} = O(n^{-1/2})$$

in the Gaussian estimate. □

It is a real challenge to achieve precise results of the kind of Proposition 1 for a stochastic $X(n)$ with correlations of the order of the variance. Correlations should affect the order of the maximum and its fluctuations, but to which extent? It is reasonable to expect the statistics of extremes of such processes to retain features of the IID statistics if we look at a coarse-enough scale. When correlations are present the expected number of exceedances is no longer a good proxy for the correct level of the maximum, since by linearity of the expectations, it is blind to correlations and is the same as IID. The expected number of exceedances turns out to be inflated by the rare events of exceedances far beyond the right level of the maximum. As we will see in Section 3, the number of exceedances need to be modified to reflect the correct behavior of the maxima.

1.2. Log-correlated fields. In the framework where $X(n) = (X_v(n), v \in V_n)$ is a stochastic process on a metric space V_n with distance d , a *log-correlated* field has correlations

$$(4) \quad E[X_v(n)X_{v'}(n)] \approx -\log d(v, v').$$

In particular, the correlation decays very slowly with the distance! Slower than any power. We emphasize that the covariance might not be exactly logarithmic for the field to exhibit a log-correlated behavior. For one, the covariance matrix must be positive definite, hence the covariance might be exactly the logarithm of the distance. Moreover, the logarithmic behavior might only be exhibited in the bulk due to some boundary effect. As we will see in Section 2, this is the case for the two-dimensional Gaussian free

field. Other important examples of processes in this class are branching random walks and models of Gaussian multiplicative chaos. It is good to re-express (4) in terms of the size of neighborhoods with a strong correlations with a fixed point v . More precisely, for $r > 0$, consider the number of points v' whose covariance is a fraction r of the variance. The logarithmic nature implies that

$$(5) \quad \frac{1}{2^n} \#\{v' \in V_n : \frac{E[X_v(n)X_{v'}(n)]}{E[X_v(n)^2]} > r\} \approx 2^{-rn}.$$

In other words, it takes approximately 2^{rn} balls to cover the space V_n with neighborhoods where correlations between points is greater than r times the variance. In particular, the size of these neighborhoods lies at the mesoscopic scales compared to the size of the systems. This will play an important role in the analysis of Sections 2 and 3.

The motivations for the study of log-correlated fields are now plenty. On one hand, in the physics literature, the work of Carpentier & Ledoussal [26] spurred a lot of interest in the study of such processes as energy landscapes of disordered systems. We mention in particular the works of Fyodorov & Bouchaud [38] and Fyodorov *et al.* [41]. These papers develop a statistical mechanics approach to the problem of describing the extreme value statistics of the systems that are applicable to a wide range of systems. Second, these processes play an essential role in Liouville quantum gravity as well as models of three-dimensional turbulence. We refer to the excellent review on Gaussian Multiplicative Chaos of Rhodes & Vargas [55] and references therein for details in these directions. Third, many models of volatility in finance are now built on the assumption that the time-correlations of the returns are log-correlated, see also [55] for more details on this. Finally, as we shall see in Section 4, log-correlated fields seem to provide the right structure to study the local maxima of the Riemann zeta function on the critical line as suggested in [39, 40].

It is argued in the physics literature that the distribution of the maximum of log-correlated fields is a borderline case where the features of IID statistics should still be apparent. In fact, it is expected that the following should hold for log-correlated fields that are close to Gaussians in some suitable sense.

Conjecture 1 (Informal). *Let $X(n) = \{X_v(n), v \in V_n\}$ be a log-correlated field in the sense of (4) with $E[X_v(n)] = 0$ and $E[X_v(n)^2] \approx \sigma^2 n$. Then for $c = \sqrt{2 \log 2} \sigma$ and*

$$(6) \quad a_n = cn - \frac{3\sigma^2}{2c} \log n, \quad b_n = 1,$$

we have

$$\lim_{n \rightarrow \infty} P \left(\max_{v \in V_n} X_v(n) \leq a_n + b_n x \right) = E[e^{-CZe^{-cx}}].$$

for some constant C and random variable Z called the derivative martingale.

This is to be compared with Proposition 1 for IID random variables. Perhaps surprisingly, the first order of the maximum is the same as for IID with the same velocity. The correlations start having an effect in the subleading where $\frac{1}{2}$ for IID is changed to $\frac{3}{2}$. This seemingly small change hides an important mechanism of the extreme value statistics that will be explained in Section 3.2. For now, let us just observe that under the recentering $a_n = cn - \frac{3\sigma^2}{2c} \log n$, the expected number of exceedances diverges like n . To get the correct order, it will be necessary to modify the exceedances with a description of the values at each scale of the field. Finally, the fluctuations are not exactly Gumbel as for IID Gaussians. The distribution is a *mixture of Gaussians*, the mixture being on

the different realizations of the random variables. This in fact changes qualitatively the nature of the distribution as it is expected that the right tail behaves like xe^{-cx} for x large whereas as it is e^{-cx} for double exponential distribution.

Conjecture 1 was proved in many instances of log-correlated fields. The most important contribution is without a doubt the seminal work of Bramson [25] who proved the result for branching Brownian motion (see Lalley & Sellke [47] for the expression in terms of the derivative martingale). For branching random walks, it was proved in great generality by Aïdekon [3], see also [1, 14, 24]; for the two-dimensional Gaussian free field, by Bramson *et al.* [23] (see Biskup & Louidor [17] for the expression in terms of the derivative martingale), for a type of Gaussian Multiplicative Chaos by Madaule [50], and finally in great generality for log-correlated Gaussian fields by Ding *et al.* [35].

We will not have time in this review paper to touch the subject of the order statistics or *extremal process* of log-correlated fields. Again, it turns out that the statistics retain features of the IID case. In fact, it is expected that the extremal process is a Poisson cluster process or Poisson decorated point process. This was proved for branching Brownian motion in [4, 8–10], for branching random walks in [51], and partially for the two-dimensional Gaussian free field in [17].

1.3. Relations to statistical physics. Questions of extreme value statistics such as Conjecture 1 can be addressed from a statistical physics point of view. In that case the object of interest is the *partition function*

$$(7) \quad Z_n(\beta) = \sum_{v \in V_n} \exp(\beta X_v(n)), \quad \beta > 0.$$

(Observe that we use β as opposed to the customary $-\beta$ since we are interested in the maximum and not the minimum of $X(n)$.) By design, the points v with a high value of the field have a greater contribution to the partition function, and the continuous parameter β (the inverse temperature) adjusts the magnitude of the contribution. The *free energy per particle* or *normalized log-partition function* is also of importance:

$$(8) \quad f_n(\beta) = \frac{1}{\beta n} \log Z_n(\beta).$$

In particular, in the limit $n \rightarrow \infty$, the free energy contains the information on the first order of the maximum. Indeed, we have the elementary inequalities

$$\frac{\max_{v \in V_n} X_v(n)}{n} \leq f_n(\beta) \leq \frac{\log 2}{\beta} + \frac{\max_{v \in V_n} X_v(n)}{n},$$

therefore

$$\lim_{n \rightarrow \infty} \frac{\max_{v \in V_n} X_v(n)}{n} = \lim_{\beta \rightarrow \infty} \lim_{n \rightarrow \infty} f_n(\beta), \quad \text{whenever the limits exist.}$$

To obtain finer information on the maximum such as subleading orders or fluctuations, one could study directly the partition functions $Z_n(\beta)$. For example, one could compute the moments of $Z_n(\beta)$ and try to infer the distribution of $Z_n(\beta)$ from it. This is the approach taken in [38, 41] for example, but there are major obstacles to overcome to make this type of arguments fully rigorous.

For log-correlated fields, it is possible to compute rigorously the free energy by the Laplace method. We have by rearranging the sum over v that

$$(9) \quad f_n(\beta) \approx \frac{1}{n} \log \left(\sum_{E \in [0, c]} \exp(\beta n E + n S_n(E)) \right),$$

where

$$(10) \quad S_n(E) = \frac{1}{n} \log \mathcal{N}_n(En) = \frac{1}{n} \log \#\{v \in V_n : X_v(n) > En\}.$$

is the *entropy* or *log-number of high points*, i.e. the log-number of exceedances at level En . In the limit, it is possible to prove in some cases, see e.g. [27] and [12, 13], that for log-correlated fields

$$(11) \quad S(E) = \lim_{n \rightarrow \infty} S_n(E) = \begin{cases} 0 & \text{if } E \geq c \\ \log 2 - \frac{E^2}{2\sigma^2} & \text{if } E \in [0, c] \end{cases} \text{ in probability,}$$

where $c = \sqrt{2 \log 2} \sigma$. We will see a general method to prove this in Section 3.1. Note that this is exactly the result one would obtain for IID random variables by applying (1)! The free energy is then easily calculated like a Gibbs variational principle

$$(12) \quad \lim_{n \rightarrow \infty} f_n(\beta) = \max_{E \in [0, c]} \left\{ \beta E + \log 2 - \frac{E^2}{2\sigma^2} \right\} = \begin{cases} \log 2 + \frac{\beta^2 \sigma^2}{2} & \text{if } \beta < \frac{c}{\sigma^2} \\ c\beta & \text{if } \beta \geq \frac{c}{\sigma^2}. \end{cases}$$

Again, not suprisingly, the free energy of log-correlated fields corresponds to the one for IID variables also known as the *Random Energy Model* (REM) as introduced by Derrida [30]. In particular, the free energy exhibits a *freezing* phase transition, i.e. for $\beta > \frac{c}{\sigma^2}$, the free energy divided by β is constant reflecting the fact the partition function is supported on the maximal values of the field. Equation (12) was first proved for branching random walks by Derrida & Spohn [32] who pioneered the study of log-correlated random fields as disordered systems. Equation (12) was also proved for other log-correlated Gaussian fields in [12, 13].

Another important object in statistical physics is the *Gibbs measure* which in our context is a random probability measure on V_n defined as

$$(13) \quad G_{n, \beta}(v) = \frac{\exp \beta X_v(n)}{Z_n(\beta)}, v \in V_n.$$

This contains refined information on the order statistics of the field $X(n)$. We only mention here that the limiting distribution of the Gibbs measure depends on the parameter β . In particular, the *annealed Gibbs measure*

$$G_{n, \beta}^{ann}(v) = \frac{\exp \beta X_v(n)}{E[Z_n(\beta)]}$$

should have a non-trivial continuous limit for high temperature $\beta < \frac{c}{\sigma^2}$ [55], whereas for low temperature $\beta \geq \beta_c$, the quenched Gibbs measure as in (13) is the relevant limiting object and should be singular [12, 13].

2. EXAMPLES AND GENERAL PROPERTIES

Branching random walk serves as a guiding example to prove rigorous results for log-correlated fields. In particular, we will use it to illustrate three important properties of log-correlated fields. We will refer to these as the *multi-scale decomposition*, the *dichotomy of scales* and the *self-similarity of scales*. As we will see these properties do not hold exactly for general log-correlated fields, like the 2D Gaussian free field. However, they hold approximately enough to reproduce the extremal behavior.

2.1. Branching random walk. We will focus on a simple example of branching random walk (BRW), where the branching is binary and the increments are Gaussian. This process is sometimes called *hierarchical Gaussian field*. More specifically, let V_n be the leafs of a binary rooted tree with n generations. The field is indexed by V_n and thus has 2^n random variables. As before, we denote the field by

$$X(n) = \{X_v(n) : v \in V_n\} .$$

The random variables $X_v(n)$ are constructed as follows. We consider $\{Y_e\}$ IID centered Gaussian random variables of variance σ^2 indexed by the edges of the binary tree. The parameter σ^2 will be adjusted when we compare log-correlated fields to approximate branching random walks. For a given leaf $v \in V_n$, the random variable $X_v(n)$ is given by the sum of the Y_e 's for the edges e on the path from the root \emptyset to the leaf v , i.e. ,

$$X_v(n) = \sum_{e: \emptyset \rightarrow v} Y_e .$$

We will often abuse notation and write $Y_v(l)$ for the variable Y_e for the leaf v at level l in the tree. In other words,

$$(14) \quad X_v(n) = \sum_{l=1}^n Y_v(l) .$$

We refer to $Y_l(v)$ as the *increment at scale l* of $X_v(n)$. With this definition, it is easy to see that the variance of each variable is

$$E[X_v(n)^2] = \sum_{l=1}^n E[Y_v(l)^2] = \sigma^2 n .$$

As for the covariance for $v, v' \in V_n$, we need to define

$v \wedge v'$, the level of the tree where the paths of v and v' to the root split.

We call $v \wedge v'$ the *branching time* or *branching scale* of v and v' . Note that $0 \leq v \wedge v' \leq n$. See Figure 3 for an illustration. It then follows directly from the definition (14) that

$$E[X_v(n)X_{v'}(n)] = \sum_{l=1}^{v \wedge v'} E[Y_v^2(l)] = \sigma^2 (v \wedge v') .$$

The covariance matrix completely determines the distribution of the centered Gaussian field $X(n)$.

Why is the field $X(n)$ log-correlated ? The best way to see this is to think in terms of neighborhood with a certain covariance as in equation (5). Indeed, let $0 \leq r \leq 1$, and

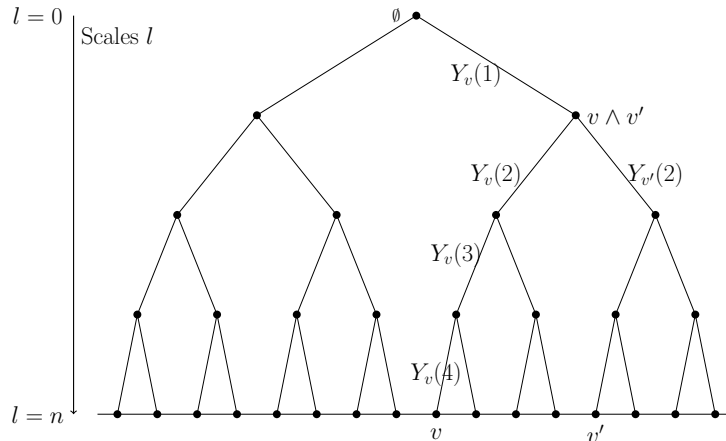


FIGURE 3. An illustration of a binary branching random walk with increments $Y_v(l)$ at each scale $0 \leq l \leq n$.

suppose for simplicity that rn is an integer. Then the size of the neighborhood of a given v with covariance $r\sigma^2n$ with v is exactly

$$\#\{v' \in V_n : \frac{E[X_v(n)X_{v'}(n)]}{\sigma^2n} \geq r\} = 2^{n-rn} .$$

It corresponds to those v' which branched out from v at level rn ! It is a ball of size $2^{n(1-r)}$ and is of mesoscopic scale compared to the size of the systems.

There are three elementary properties of BRW which will have their counterpart for any log-correlated field. These properties will be key to guide our analysis in more complicated models.

- *Multiscale decomposition*: The first property simply refers to the decomposition (14) into independent increments.

$$X_v(n) = \sum_{l=1}^n Y_v(l) .$$

- *Dichotomy of scales*: Another simple consequence of the definition of BRW is the fact that for scales before the branching points of v and v' , the increments are perfectly correlated whereas after the branching point, they are exactly independent. In other words:

$$(15) \quad E[Y_l(v)Y_l(v')] = \begin{cases} \sigma^2 & \text{if } l \leq v \wedge v', \\ 0 & \text{if } l > v \wedge v' . \end{cases}$$

We refer to this abrupt change in correlations for the increments as the *dichotomy of scales*.

- *Self-similarity of scales*: In the case of BRW, the increments at each scale are identically distributed. In particular, together with the property of dichotomy, this property implies that for a given $0 < l < n$, the variables $X_v(l) = \sum_{k=1}^l Y_v(k)$ define a process $X(l)$ with 2^l distinct values. Moreover, this process is exactly a BRW with l levels. Similarly, for a given v , we can define the variables

$$\{X_{v'}(n) - X_{v'}(l) : v' \wedge v \geq l\} .$$

This defines a BRW on 2^{n-l} points corresponding to the subtree with the common ancestor at level l acting as the root.

We point out that the above construction leads to a natural generalization of fields by modifying the variance of the increments in the multiscale decomposition at each scale. The self-similarity is lost in these models. The study of such types of models go back to Derrida who introduced the *Generalized Random Energy Models* [31]. In the case of BRW, these are called time-inhomogeneous BRW. They were studied in [21, 22, 36, 37, 52]. A similar *scale-inhomogeneous* field for the Gaussian free field was introduced and studied in [11, 13].

2.2. 2D Gaussian free field. The discrete two-dimensional Gaussian free field (2DGFF) is an important model of random surfaces. Let V_n be a finite square box of \mathbb{Z}^2 . To emphasize the similarities with BRW, we suppose that $\#V_n = 2^n$. The 2DGFF is a centered Gaussian field $X(n) = \{X_v(n), v \in V_n\}$ defined as follows.

Let \mathbb{P}_v be the law of a simple random walk $(S_k, k \geq 0)$ starting at $v \in V_n$ and \mathbb{E}_v , the corresponding expectation. Denote by τ_n the first exit time of the random walk from V_n . We consider the covariance matrix given by

$$(16) \quad E[X_v(n)X_{v'}(n)] = \mathbb{E}_v \left[\sum_{k=0}^{\tau_n} 1_{S_k=v'} \right] = G_n(v, v') .$$

The right-hand side is the expected number of visits to v' of a random walk starting at v before exiting V_n . It is not hard to show, using the Markov property of the random walk, that the right-hand side is the Green's function of the discrete Laplacian on V_n with Dirichlet boundary conditions. More precisely, for f a function on the vertices of V_n , the discrete Laplacian is

$$-\Delta f(v) = \frac{1}{4} \sum_{\omega \sim v} f(\omega) - f(v) = \mathbb{E}_v[f(S_1) - f(v)] ,$$

where $\omega \sim v$ means that ω and v share an edge in \mathbb{Z}^2 . The Green's function can be seen as the inverse of the discrete Laplacian in the sense that it satisfies

$$-\Delta G_n(v, v') = \delta_{v'}(v) = \begin{cases} 1 & \text{if } v = v', \\ 0 & \text{if } v \neq v'. \end{cases}$$

In particular, the Green's function is symmetric and positive definite, since the Laplacian is. The density of the 2DGFF is easily obtained from the covariance (16). Indeed, by rearranging the sum, we have

$$\sum_{v \sim v'} (x_v - x_{v'})^2 = \frac{1}{4} \sum_v x_v (-\Delta) x_v .$$

Again, since the Green's function is the inverse of $-\Delta$, we have that

$$(17) \quad P(X_v(n) \in dx_v, v \in V_n) = \frac{1}{Z} \exp \left(-\frac{1}{8} \sum_{v \sim v'} (x_v - x_{v'})^2 \right) ,$$

for the appropriate normalization Z .

Why is this field log-correlated? It is reasonable to expect this since the Green's function of the Laplacian in the continuous setting decays logarithmically with the distance.

This analogy can be made rigorous using random walk estimates. In fact, it is possible to show, see e.g. [48], that

$$(18) \quad G_n(v, v') = E_v \left[\sum_{k=0}^{\tau_n} 1_{S_k=v'} \right] = E_v[a(v', S_{\tau_n})] - a(v, v')$$

where $a(v, v) = 0$ and if $v \neq v'$

$$a(v, v') = \frac{1}{\pi} \log d(v, v')^2 + O(1) + O(d(v, v')^{-2}) ,$$

where $d(v, v')$ is the Euclidean distance in \mathbb{Z}^2 . In particular, this implies that for points not too close to the boundary, the field is log-correlated: $G_n(v, v') = \frac{1}{\pi} \log \frac{2^n}{d(v, v')^2} + O(d(v, v')^{-2})$. Moreover, we get the following upper bound on the variance:

$$G_n(v, v) \leq \frac{1}{\pi} \log 2^n + O(1) .$$

Again, a matching lower bound follows from (18), but only for points far from the boundary. As first step towards a connection to BRW, these estimates already suggest that the BRW parameter σ^2 should be taken to be

$$\sigma^2 \longrightarrow \frac{\log 2}{\pi} .$$

To make the connection with an approximate BRW more precise, we shall need three fundamental facts. First, it is not hard to prove from the density (17) that the field satisfies the *Markov property*, that is for any finite box B in V_n ,

$$(19) \quad \{X_v(n), v \in B\} \text{ is independent of the field in } B^c \text{ given the field on the boundary } \partial B .$$

Here the boundary ∂B refers to the vertices in B^c that share an edge with a vertex in B . This property essentially follows from (17) since the correlations are expressed in terms of nearest neighbors. Second, it follows from the Markov property that the conditional expectation of the field inside v given the boundary is simply a linear combination of the field on the boundary. More precisely, we have

$$(20) \quad E[X_v(n) | \{X_{v'}(n), v' \in B^c\}] = \sum_{u \in \partial B} p_u(v) X_u(n)$$

where $p_u(v)$ is the probability of a simple random walk starting at v to exit B at u . The specific form of the coefficient of the linear combination is proved using the strong Markov property of the random walk. A similar argument shows that the process is self-similar. Namely, if we write $X_v(B) = E[X_v(n) | \{X_u(n), u \in \partial B\}]$ for the harmonic average of $X_v(n)$ on the boundary of B , we have

$$(21) \quad \{X_v(n) - X_v(B), v \in B\} \quad \text{is a 2DGFF on } B,$$

in the sense that it is a Gaussian free field with covariance (16) restricted to B .

The third property shows that if v and v' are in B , then $X_v(B)$ and $X_{v'}(B)$ should be close, if they are not too close to the boundary of B . More precisely, say B is a square box or a disc containing 2^l points for some l . Estimates of the Green's function as in (18) can be used to show that, see e.g. Lemma 12 in [18],

$$(22) \quad E[(X_v(B) - X_{v'}(B))^2] \leq O(d(v, v')2^{-l/2})$$

whenever the distance of v and v' to the center of B is of the order of $2^{(l-1)/2}$ or less, to avoid boundary effects.

The connection between the 2D Gaussian free field and an approximate BRW is made through Equations (19), (21), and (22). For each $v \in V_n$ and $0 \leq l \leq n$, we define $[v]_l$, a neighborhood of v containing 2^{n-l} points. For simplicity, suppose that $[v]_l$ is a square box. If such a neighborhood is not entirely contained in V_n , we define $[v]_l$ to be the intersection of the neighborhood with V_n . Note that these neighborhoods shrink as l increases. By convention, we take $[v]_0 = V_n$ and $[v]_n = \{v\}$. The boundary is denoted $\partial[v]_l$ and the union of $\partial[v]_l$ with its boundary will be denoted by $\overline{[v]_l}$. The connection of 2DGFF with BRW is illustrated in Figure 4 and is as follows. For simplicity, we assume that all vertices v are not too close to the boundary of V_n .

Multiscale decomposition: The decomposition here is simply a *martingale decomposition*. Namely, we define the *field of $v \in V_n$ at scale l* as

$$X_v(l) = E[X_v(n) | \{X_{v'}(n) : v' \in [v]_l^c\}] = E[X_v(n) | \{X_u(n) : u \in \partial[v]_l\}] ,$$

where the second equality is by the Markov property. Note that for each $v \in V_n$, $(X_v(l), 0 \leq l \leq n)$ is a martingale by construction. Define the *increments at scale l* as the martingale difference

$$Y_v(l) = X_v(l) - X_v(l-1) , \quad 1 \leq l \leq n .$$

These increments are orthogonal by construction. Since the field is Gaussian, so are the conditional expectations. In particular, the increments are independent of each other for a given v . Thus, the decomposition

$$(23) \quad X_v(n) = \sum_{l=1}^n Y_v(l) , \quad v \in V_n ,$$

seems so far to be the right analogue of the multiscale decomposition of BRW. It remains to check the other two properties.

Self-similarity of scales: First, note that the increments are Gaussian by definition. The fact that the increments have approximately the same variance follows from (21). Indeed, for $0 \leq l \leq n$, the variance of $X_v(n) - X_v(l)$ is by (18)

$$E[(X_v(n) - X_v(l))^2] = (n-l) \frac{\log 2}{\pi} + O(1) ,$$

Because $X_l(v)$ is orthogonal to $X_v(n) - X_v(l)$, we deduce that

$$E[X_v(l)^2] = l \frac{\log 2}{\pi} + O(1) .$$

In particular, since $X_v(l+1)$ admits the orthogonal decomposition $X_v(l) + Y_v(l+1)$, this implies

$$E[Y_v(l)^2] = \frac{\log 2}{\pi} + o(1)$$

where $o(1)$ is summable, so the contribution of the sum over l of the error is $O(1)$. This justifies the choice $\sigma^2 = \log 2/\pi$ for the comparison with BRW. Of course, here, the increments are not exactly identically distributed, but since the error is summable, it will not affect the behavior of the extremes.

Dichotomy of scales: It remains to establish the correlations of the increments between two distinct v, v' . For this purpose, it is useful to define the *branching scale $v \wedge v'$* as

follows

$$(24) \quad v \wedge v' = \min\{0 \leq l \leq n : \overline{[v]}_l \cap \overline{[v']}_l = \emptyset\},$$

that is the smallest scale for which the neighborhoods of v and v' (including their boundaries) are disjoint. Note that because the neighborhoods have size 2^{n-l} , the branching scale $v \wedge v'$ is related to the distance by

$$d(v, v')^2 \approx 2^{n-v \wedge v'}.$$

For small scales, a direct application of the Markov property (19) yields

$$E[Y_v(l)Y_{v'}(l)] = 0 \quad \text{if } l > v \wedge v'.$$

This is because if $l > v \wedge v'$, then the neighborhoods of v and v' do not intersect and are contained in each other's complement. For scales larger than the branching scale, we expect the increments to be almost perfectly correlated. To see this, note that by definition, for $l = v \wedge v' - 1$, then the neighborhoods intersect but for larger neighborhoods, that is $l < v \wedge v' - 1$, v and v' must be contained in each other's neighborhood. In particular, an application of (22) gives

$$E[(X_v(l) - X_l(v'))^2] = O(1).$$

In other words, the vectors $X_v(l)$ and $X_l(v')$ are close in L^2 . Since these vectors are sums of orthogonal increments, these increments must be also close. In particular, we get

$$E[Y_v(l)Y_{v'}(l)] = \frac{\log 2}{\pi} + o(1) \quad \text{if } l < v \wedge v' - 1,$$

where $o(1)$ is summable in l .

This completes the construction of an approximate BRW embedded in 2DGFF. We remark that the dichotomy of scales is not as clean cut as the one of BRW in (15). In particular, nothing precise can be said for the correlation between exactly at the branching scale $l = v \wedge v'$. It turns out that the coupling and decoupling of the increments do not have to be perfect as long as they occur fast enough. The same will be true for non-Gaussian models in Section 4.

3. ORDER OF THE MAXIMUM

In this section, we explain a general method to show the leading and subleading orders of the maximum for log-correlated fields in Conjecture 1. We will prove them in the case of Gaussian branching random walk on the binary tree, following the treatment in [46]. We will also prove the convergence of the log-number of high points or entropy, cf. (11). The heuristic is (or expected to be) the same for more complicated log-correlated models, such as the two-dimensional Gaussian free field. The multiscale decomposition of the field plays a fundamental role in the proof. This is also true when proving more refined results such as the convergence of the maximum and of the extremal process. In a nutshell, to prove such results on the maximum, one needs to understand the contribution of the increments at each scale to a large value of the field.

3.1. Leading order of the maximum. Recall the definition of the Gaussian BRW $X(n)$ on a binary tree defined in Section 2.1. The following result yields the leading order of the maximum

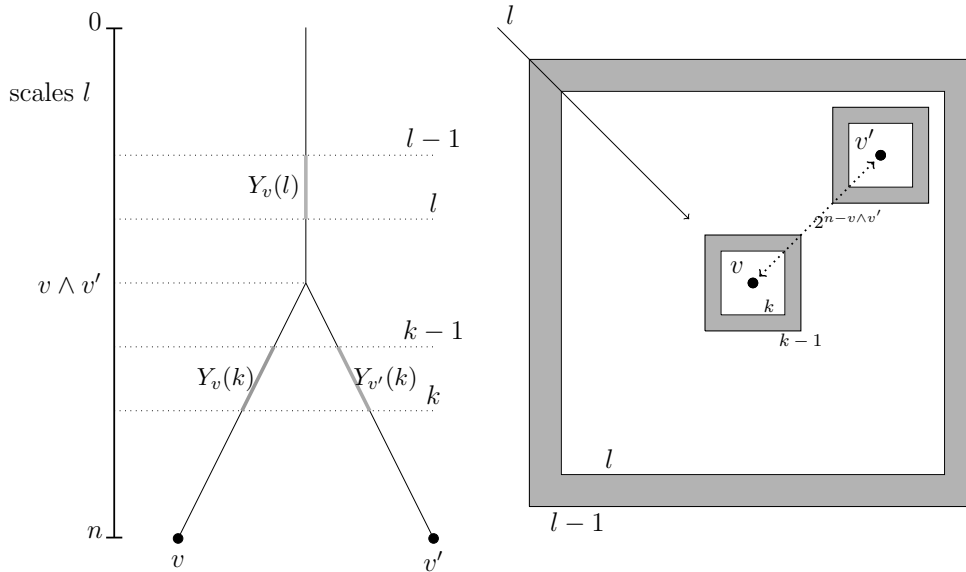


FIGURE 4. An illustration of the approximate branching structure defined by the 2DGFF. The grey area represents the increments at each scale.

Theorem 2. Let $X(n) = \{X_v(n) : v \in V_n\}$ be a branching random walk on a binary tree with n levels as in (23) with Gaussian increments of variance σ^2 . Then

$$\lim_{n \rightarrow \infty} \frac{\max_{v \in V_n} X_v(n)}{n} = \sqrt{2 \log 2} \sigma \quad \text{in probability.}$$

The theorem goes back to Biggins for general BRW [16]. The analogue for the 2DGFF was proved by Bolthausen, Deuschel & Giacomin [18]. We point out that Kistler greatly simplified the proof in both cases [46] using a *multiscale refinement of the second moment method*. This approach also makes the analysis amenable to non-Gaussian models. We refer the reader directly to [46] for a proof of the theorem. We shall instead use Kistler's method to prove a very similar result on the *log-number of high points or exceedances*.

Theorem 3. Let $X(n) = \{X_v(n) : v \in V_n\}$ be a branching random walk on a binary tree with n levels as in (23) with Gaussian increments of variance σ^2 . Then for $0 \leq E < c = \sqrt{2 \log 2} \sigma$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \#\{v \in V_n : X_v(n) > En\} = \log 2 \left(1 - \frac{E^2}{c^2}\right) \quad \text{in probability.}$$

In other words, the number of points whose field value is larger than En is approximately $2^{n(1-E^2/c^2)}$ with large probability for E smaller than the maximal level c . We stress that this is the same result for 2^n IID Gaussian random variables of variance σ^2 . The analogue of this result for 2DGFF was shown by Daviaud in [27].

Proof. Upper bound: To bound the number of points beyond a given level, it suffices to use Markov's inequality. Let

$$\mathcal{N}_n(En) = \#\{v \in V_n : X_v(n) > En\}.$$

Then for a given $\varepsilon > 0$, we have

$$(25) \quad P\left(\mathcal{N}_n(En) > 2^{n(1+\varepsilon)(1-\frac{E^2}{c^2})}\right) \leq 2^{-n(1+\varepsilon)(1-\frac{E^2}{c^2})} E[\mathcal{N}_n(En)].$$

By linearity, the expectation is simply

$$E[\mathcal{N}_n(En)] = 2^n P(X_v(n) > En) \leq 2^n \exp\left(-\frac{E^2 n}{2\sigma^2}\right) = 2^{n(1-\frac{E^2}{2c^2})}$$

where we use the Gaussian estimate (1) with $X_v(n) \sim \mathcal{N}(0, \sigma^2 n)$. It follows that the probability in (25) goes to 0. This proves the upper bound.

Lower bound: It remains to show that

$$P\left(\mathcal{N}_n(En) > 2^{n(1-\varepsilon)\left(1-\frac{E^2}{c^2}\right)}\right) \rightarrow 1 .$$

The lower bound is an application of the *Paley-Zygmund inequality*, which states that for any random variable $\mathcal{N} > 0$ and $0 \leq \delta \leq 1$,

$$(26) \quad P(\mathcal{N} > \delta E[\mathcal{N}]) \geq (1 - \delta)^2 \frac{E[\mathcal{N}]^2}{E[\mathcal{N}^2]}$$

So ultimately, one needs to prove that in the limit the second moment matches the first moment square. This would certainly be the case if the variables were independent. In the case of BRW, this is not so. However, we can create enough independence and not losing too much precision by dropping the increments at lower scales. In addition, as observed in [46], the matching of the moments is greatly simplified if one considers increments rather than the sum of increments. More precisely, since the leading of the maximum is linear in the scales, it is reasonable to expect that each scale contributes equally to the maximal value to leading order (since they have the same variance). It turns out this heuristic is correct and extend to the high points.

With this in mind, we introduce a parameter K which divides the n scales coarser K levels. We define accordingly, for $m = 1, \dots, K$, the increments

$$W_m(v) = \sum_{\frac{m-1}{K} < l \leq \frac{m}{K}} Y_l(v) .$$

The idea is that for a high point at level En , each variable $W_m(v)$ should contribute $E\frac{n}{K}$. Without loss of generality, we can suppose ε is small enough so that $(1 + \varepsilon)E < c$. We consider the event

$$\mathcal{E}_m(v) = \{W_m(v) > (1 + \varepsilon)\frac{n}{K} E\} , m = 1, \dots, K,$$

and the *modified number of exceedances*,

$$\tilde{\mathcal{N}}_n = \sum_{v \in V_n} \prod_{m=2}^K 1_{\mathcal{E}_m(v)} .$$

Note that if v is counted in $\tilde{\mathcal{N}}_n$, then each $W_m(v)$ for $m \geq 2$ exceeds $(1 + \varepsilon)K^{-1}nE$. Moreover, it is easy to check using the Gaussian estimate (1) that $|W_1(v)|$ is less than $K^{-1}n$ with large probability. In particular, $X_v(n) > (1 + \varepsilon)(1 - K^{-1})En$ on the events, so that we can take K large enough depending on ε so that $X_v(n) > En$. We conclude that

$$\tilde{\mathcal{N}}_n \leq \mathcal{N}_n(En) ,$$

hence it suffices to show

$$P\left(\tilde{\mathcal{N}}_n > 2^{n(1-\varepsilon)\left(1-\frac{E^2}{c^2}\right)}\right) \rightarrow 1 .$$

We apply the inequality (26) with $\delta = 2^{-n\varepsilon(1-\frac{E^2}{c^2})}$. We have by the Gaussian estimate (1)

$$E[\tilde{\mathcal{N}}_n] = 2^n \frac{(1+o(1))}{\sqrt{2\pi}} \exp\left(-\frac{(1-K^{-1})(1+\varepsilon)^2 E^2 n}{2\sigma^2}\right) \geq 2^{n(1-\frac{E^2}{c^2})},$$

for n large enough. Thus

$$P\left(\tilde{\mathcal{N}}_n > 2^{n(1-\varepsilon)(1-\frac{E^2}{c^2})}\right) \geq P\left(\tilde{\mathcal{N}}_n > \delta E[\tilde{\mathcal{N}}_n]\right)$$

It remains to show

$$(27) \quad \frac{(E[\tilde{\mathcal{N}}_n])^2}{E[(\tilde{\mathcal{N}}_n)^2]} \rightarrow 1 \text{ as } n \rightarrow \infty.$$

The idea is to split the second moment according to the branching scale. More precisely, we have

$$(28) \quad \begin{aligned} E[(\tilde{\mathcal{N}}_n)^2] &= \sum_{\substack{v, v' \in V_n \\ v \wedge v' \leq \frac{n}{K}}} P\left(\bigcap_{m=2}^K \mathcal{E}_m(v) \cap \mathcal{E}_m(v')\right) + \sum_{r=2}^K \sum_{\substack{v, v' \in V_n \\ \frac{(r-1)n}{K} < v \wedge v' \leq \frac{rn}{K}}} P\left(\bigcap_{m=2}^K \mathcal{E}_m(v) \cap \mathcal{E}_m(v')\right) \\ &= \sum_{\substack{v, v' \in V_n \\ v \wedge v' \leq \frac{n}{K}}} \prod_{m=2}^K P(\mathcal{E}_m(v))^2 + \sum_{r=2}^K \sum_{\substack{v, v' \in V_n \\ \frac{(r-1)n}{K} < v \wedge v' \leq \frac{rn}{K}}} \prod_{m=2}^K P(\mathcal{E}_m(v) \cap \mathcal{E}_m(v')) \end{aligned}$$

where the second line holds by the independence of the increments between the scale and the fact that $W_m(v) = W_m(v')$ for all $m \geq 2$ if $v \wedge v' \leq n/K$ by the dichotomy of scales. Since there are at least $2^n(2^n - 2^{n-n/K}) = 2^{2n}(1 - o(1))$ pairs such that $v \wedge v' \leq n/K$, the first term of (28) is $(1 - o(1))(E[\tilde{\mathcal{N}}_n])^2$. The proof will be concluded once we prove that the second term is $o(1)(E[\tilde{\mathcal{N}}_n])^2$.

Guided by the dichotomy of scales, we do not lose by dropping the constraints on m smaller than the branching scales for one of the v 's since they are identical. The second term is smaller than

$$\sum_{r=2}^K \sum_{\substack{v, v' \in V_n \\ \frac{(r-1)n}{K} < v \wedge v' \leq \frac{rn}{K}}} \prod_{m=2}^K P(\mathcal{E}_m(v)) \prod_{m \geq r+1} P(\mathcal{E}_m(v')),$$

where we use the independence of the increments after the branching scales. Since there are at most $2^n(2^{n-\frac{(r-1)n}{K}})$ pairs $v \wedge v'$ such that $\frac{(r-1)n}{K} < v \wedge v' \leq \frac{rn}{K}$, we have that the above is smaller than

$$(E[\tilde{\mathcal{N}}_n])^2 \sum_{r=2}^K \left(2^{\frac{(r-1)n}{K}} \prod_{m=2}^r P(\mathcal{E}_m(v))\right)^{-1} \leq (E[\tilde{\mathcal{N}}_n])^2 \sum_{r=2}^{\infty} \left(2^{(r-1)nK^{-1}(1-(1+\varepsilon)^2 E^2/c^2)}\right)^{-1}$$

where we used the Gaussian estimate with variance n/K for each variable W_m . The sum is geometric. It converges because of the choice $(1+\varepsilon)E < c$. Thus the second term of (28) is smaller than

$$2(E[\tilde{\mathcal{N}}_n])^2 2^{-nK^{-1}(1-(1+\varepsilon)^2 E^2/c^2)} = o(1)(E[\tilde{\mathcal{N}}_n])^2,$$

for n large enough. This proves (27) and concludes the proof of the theorem. \square

3.2. Subleading order of the maximum. We now turn to the proof of the subleading order of the maximum for a Gaussian branching random walk. The basic idea goes back to the seminal work of Bramson who showed the convergence of the law of the maximum of branching Brownian motion. It consists in observing that the increments of the maximum must satisfy a linear constraint at each scale. This is consistent with the fact the first order is linear in the scales and that at each scale l , the variables $X_l(v)$ form themselves a BRW with 2^l points. To prove convergence, more precise estimates are needed than for the subleading order. We follow here the treatment of [46] with the additional use of the *ballot theorem* as in [5], cf. Theorem 5 below.

Theorem 4. *Let $X(n) = \{X_v(n) : v \in V_n\}$ be a branching random walk on a binary tree with n levels as in (23) with Gaussian increments of variance σ^2 . Then*

$$\lim_{n \rightarrow \infty} \frac{\max_{v \in V_n} X_v(n) - cn}{\log n} = -\frac{3}{2} \quad \text{in probability.}$$

where $c = \sqrt{2 \log 2} \sigma$.

For simplicity, we define the deterministic displacement of the maximum as

$$(29) \quad m_n = cn - \frac{3\sigma^2}{2c} \log n, \text{ and more generally, } m_n(\varepsilon) = m_n + \varepsilon \log n .$$

The first observation is to realize that the expected of exceedances at the level m_n diverges! Indeed, using the Gaussian estimate (1) and linearity

$$E[\mathcal{N}_n(m_n)] = E[\#\{v \in V_n : X_v(n) > m_n\}] = O(n) .$$

This is the first sign that the branching structure matters for the subleading order. The divergence of the expectation comes from atypical events that inflate the expectation. Therefore, it is necessary to restrict the exceedances to the typical behavior values. Since the first order is linear in the scales and the variables $\{X_l(v) : v \in V_n\}$ form a BRW on 2^l values, it is reasonable to expect that $X_l(v) = \sum_{k=1}^l Y_k(v) \leq cl + B$ for some appropriate choice of *barrier* B . Keeping this in mind, we introduce a *modified number of exceedances*

$$\tilde{\mathcal{N}}_n = \#\{v \in V_n : X_v(n) > m_n, X_l(v) \leq cl + B \forall l \leq n\} .$$

It turns out that $E[\tilde{\mathcal{N}}_n] = O(1)$, because the probability of a random walk (in this instance $(X_l(v), l \leq n)$) to stay below a barrier is exactly of the order $1/n$. This is the content of the ballot theorem that we now state precisely. The reader is referred to [1, 2, 24] for more details. The result quoted here is the version appearing in [7].

Theorem 5. *Let $(S_n)_{n \geq 0}$ be a Gaussian random walk with increments of mean 0 and variance $\sigma^2 > 0$, with $S_0 = 0$. Let $\delta > 0$. There is a constant $C = C(\sigma, \delta)$ such that for all $B > 0$, $b \leq B - \delta$ and $n \geq 1$*

$$(30) \quad P[X_n \in (b, b + \delta) \text{ and } X_k \leq B \text{ for } 0 < k < n] \leq C \frac{(1+B)(1+B-b)}{n^{3/2}} .$$

Moreover, if $\delta < 1$,

$$(31) \quad P[X_n \in (0, \delta) \text{ and } X_k \leq 1 \text{ for } 0 < k < n] \geq \frac{1}{Cn^{3/2}} .$$

With this new tool in hand, we are ready to prove the subleading order.

Proof of Theorem 4. Upper bound: We want to show that for $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(\max_{v \in V_n} X_v(n) \geq m_n(\varepsilon)) = 0 .$$

As mentioned previously, the idea is to first show that we can constrain the random walks to stay below a linear barrier. More precisely, we prove that for $B = \log^2 n$,

$$(32) \quad \lim_{n \rightarrow \infty} P(\exists v \in V_n : X_v(n) \geq m_n(\varepsilon), X_l(v) > cl + B \text{ for some } l \leq n) = 0 .$$

The proof is by a Markov's inequality at each scale. Indeed, we can bound the above by

$$\sum_{l=1}^n P(\exists v : X_l(v) > cl + B) \leq \sum_{l=1}^n 2^l \exp\left(-\frac{(cl + B)^2}{2\sigma^2 l}\right) = n \exp(-Bc/\sigma^2) = o(1) ,$$

where the first inequality is by the Gaussian estimate and the last by the choice of $B = \log^2 n$.

To conclude the proof of the upper bound, consider the event

$$\mathcal{E}^+(v) = \{X_v(n) \geq m_n(\varepsilon), X_l(v) \leq cl + B \forall l \leq n\} .$$

By (32), it follows that

$$P(\max_{v \in V_n} X_v(n) \geq m_n(\varepsilon)) = P\left(\sum_{v \in V_n} \mathbf{1}_{\mathcal{E}^+(v)} \geq 1\right) + o(1) .$$

Therefore, it remains to show that the first probability goes to 0. By Markov's inequality, this is simply

$$P\left(\sum_{v \in V_n} \mathbf{1}_{\mathcal{E}^+(v)} \geq 1\right) \leq 2^n P(\mathcal{E}^+(v)) .$$

To prove that this goes to 0, it is useful to make a change of measure to make the end value $m_n(\varepsilon)$ typical. To this aim, consider the measure Q defined from P through the density

$$\frac{dQ}{dP} = \prod_e \frac{e^{\lambda Y_e}}{e^{\varphi(\lambda)}}, \quad \varphi(\lambda) = \log E[e^{\lambda Y_e}] = \frac{\lambda^2 \sigma^2}{2} ,$$

where the product is over all the edges in the binary tree. Note that by definition

$$\varphi'(\lambda) = E_Q[Y_e] = \lambda \sigma^2 .$$

In particular, to make $m_n(\varepsilon)$ typical, we choose

$$\lambda = \frac{m_n(\varepsilon)}{\sigma^2 n} .$$

Under this change of measure, we can write the probability as

$$(33) \quad \begin{aligned} P(\mathcal{E}^+(v)) &= e^{-(\lambda m_n(\varepsilon) - \varphi(\lambda))} E_Q[e^{-\lambda(X_v(n) - m_n(\varepsilon))} \mathbf{1}_{\mathcal{E}^+(v)}] \\ &\leq e^{-\frac{(m_n(\varepsilon))^2}{2\sigma^2 n}} Q(\mathcal{E}^+(v)) , \end{aligned}$$

where we use the choice of λ and the fact that $X_v(n) - m_n(\varepsilon) \geq 0$ on the event $\mathcal{E}^+(v)$. Since

$$2^n \exp\left(-\frac{(m_n(\varepsilon))^2}{2\sigma^2 n}\right) = O(n^{-\varepsilon c/\sigma^2}) n^{3/2} ,$$

the upper bound will be proved if

$$n^{3/2} Q(\mathcal{E}^+(v)) = o(n^{\varepsilon c/\sigma^2}) .$$

But this follows from the ballot theorem (Theorem 5) with $B = (\log n)^2$, since $(Y_l(v) - m_n(\varepsilon)/n, l \leq n)$ under Q is a random walk of finite variance.

Lower bound: As for the leading order, to get a matching lower bound, it is necessary to create independence by cutting off the increments at low scales. Of course, we cannot truncate as many scales. We choose

$$r = \log \log n ,$$

and consider the truncated walks $X_v(n) - X_v(r)$. We do not lose much by dropping these since, for $\varepsilon > 0$,

$$(34) \quad P(\exists v : X_v(n) < m_n(-2\varepsilon), X_v(n) - X_r(v) \geq m_{n-r}(-\varepsilon)) \leq P(\exists v : X_v(r) < -10cr) ,$$

where the inequality holds for n large enough depending on ε . By a union bound and the Gaussian estimate (1) (note that the variables are symmetric!),

$$P(\exists v : X_v(r) < -10cr) \leq 2^r \exp(-100c^2r^2/(2\sigma^2)) = 2^{-99r} \rightarrow 0 .$$

In view of (34) and by redefining $\varepsilon > 0$, we reduced the proof to showing

$$(35) \quad P(\exists v : X_v(n) - X_r(v) \geq m_{n-r}(-\varepsilon)) \rightarrow 1, n \rightarrow \infty .$$

For this purpose, consider

$$(36) \quad \mu = \frac{m_{n-r}(-\varepsilon)}{n-r} \quad \frac{\mu^2}{2\sigma^2} = \log 2 - \left(\frac{3}{2} - \frac{c}{\sigma^2} \varepsilon \right) \frac{\log(n-r)}{n-r} + o(n^{-1}) .$$

Denote the truncated walk and its recentering by

$$X_v(r, n) = X_v(n) - X_r(v) , \quad \bar{X}_v(r, n) = X_v(n) - X_r(v) - \mu(n-r) .$$

With this notation, the event of exceedance with the barrier $B = 1$ is

$$\mathcal{E}^-(v) = \{v \in V_n : \bar{X}_v(r, n) \in [0, \delta], \bar{X}_v(l, n) \leq 1 \forall l \geq r+1\} ,$$

where $\delta > 0$ is arbitrary. The relevant number of exceedances is

$$\mathcal{N}^-(v) = \sum_{v \in V_n} 1_{\mathcal{E}^-(v)} .$$

By the Paley-Zygmund inequality, we have

$$P(\mathcal{N}^-(v) \geq 1) \geq \frac{(E[\mathcal{N}^-(v)])^2}{E[(\mathcal{N}^-(v))^2]} .$$

To prove (35), it remains to show that

$$(37) \quad E[(\mathcal{N}^-(v))^2] = (1 - o(1))(E[\mathcal{N}^-(v)])^2 .$$

The second moment can be split as a sum over $v \wedge v'$. We get

$$(38) \quad \begin{aligned} E[(\mathcal{N}^-(v))^2] &= \sum_{v, v': v \wedge v' \leq r} P(\mathcal{E}^-(v) \cap \mathcal{E}^-(v')) \\ &+ \sum_{m=r+1}^n \sum_{v, v': v \wedge v' = m} P(\mathcal{E}^-(v) \cap \mathcal{E}^-(v')) . \end{aligned}$$

For the first term, note that $\mathcal{E}^-(v)$ is independent of $\mathcal{E}^-(v')$ because $v \wedge v' \leq r$ so the walks share no increments. Thus, the first term is

$$(39) \quad \sum_{v, v': v \wedge v' \leq r} P(\mathcal{E}^-(v) \cap \mathcal{E}^-(v')) = (1 - o(1)) (E[\mathcal{N}^-(v)])^2$$

since there are $2^{2n}(1 - o(1))$ pairs v, v' with $v \wedge v' \leq r$. We show that the second term is $o(1)(E[\mathcal{N}^-(v)])^2$ to conclude the proof.

Observe that, by proceeding as (33) with the measure Q (for $\lambda = \mu/\sigma^2$) and the Gaussian estimate (1), the expected number of exceedances satisfy

$$(40) \quad \begin{aligned} E[\mathcal{N}^-(v)] &\geq C2^n e^{-\lambda\delta} e^{-\frac{\mu^2(n-r)}{2\sigma^2}} Q(\mathcal{E}^-(v)) \\ &\geq C2^r e^{-\lambda\delta} (n-r)^{\frac{\varepsilon}{\sigma^2}} (n-r)^{3/2} Q(\mathcal{E}^-(v)) \\ &\geq C2^r (n-r)^{\frac{\varepsilon}{\sigma^2}} e^{-\lambda\delta} \end{aligned}$$

for some constant $C > 0$. Here we used the estimate on $\frac{\mu^2}{2\sigma^2}$ in (36). The last inequality is a consequence of the ballot theorem 5. For $v \wedge v' = m$, we decompose the probability $P(\mathcal{E}^-(v) \cap \mathcal{E}^-(v'))$ on the events $\{\bar{X}_v(r, m) \in (-q, -q+1]\}$ for $q \geq 0$. Note that $X_v(m) = X_{v'}(m)!$ We thus get

$$\begin{aligned} &P(\mathcal{E}^-(v) \cap \mathcal{E}^-(v')) \\ &= \sum_{q=0}^{\infty} P(\mathcal{E}^-(v) \cap \mathcal{E}^-(v') \cap \{\bar{X}_v(m, n) \in (-q, -q+1]\}) \\ &\leq \sum_{q=0}^{\infty} P(\bar{X}_v(r, m) \in (-q, -q+1], \bar{X}_v(r, l) \leq 1 \forall l = r+1, \dots, m) \\ &\quad \times \left(P(\bar{X}_v(m, n) \geq q-1, \bar{X}_v(m, l) \leq 1+q \forall l = m+1, \dots, n) \right)^2 \end{aligned}$$

where we used the bound on $X_v(m)$, and the square of the probability comes from the independence of increments after the branching scale m . We evaluate these two probabilities. Using the change of measure with $\lambda = \mu/\sigma^2$ and the bounds on $\bar{X}_v(m, n)$, the first one is smaller than

$$e^{\lambda q} e^{-\frac{\mu^2(n-r)}{2\sigma^2}} Q(\bar{X}_v(r, m) \in (-q, -q+1], \bar{X}_v(r, l) \leq 1 \forall l = r+1, \dots, m) .$$

By the ballot theorem, this is smaller than

$$e^{\lambda q} e^{-\frac{\mu^2(n-r)}{2\sigma^2}} \frac{q+1}{(m-r)^{3/2}} .$$

An identical treatment of the square of the second probability yields that it is smaller than

$$e^{-2\lambda q} e^{-\frac{\mu^2(m-r)}{\sigma^2}} \frac{q+1}{(n-m)^3} .$$

Putting all the above together, we finally obtained that

$$\sum_{m=r+1}^n \sum_{v, v': v \wedge v' = m} P(\mathcal{E}^-(v) \cap \mathcal{E}^-(v')) \leq \sum_{m=r+1}^n 2^{2n-m} \frac{e^{-\frac{\mu^2}{2\sigma^2}(2n-m-r)}}{(m-r)^{3/2}(n-m)^3} \sum_{q=0}^{\infty} (q+1)^2 e^{-\lambda q} ,$$

since there are at most 2^{2n-m} pairs with $v \wedge v' = m$. The last sum is finite. Moreover, using the estimate in (36), the first sum is smaller than

$$C2^r(n-r)^{2\varepsilon\frac{c}{\sigma^2}} \sum_{m=r+1}^n \frac{(n-r)^{\frac{3}{2}(2-\frac{m-r}{n-r})}}{(m-r)^{3/2}(n-m)^3} .$$

It is not hard to show by an integral test that the series is bounded for any n . Thus

$$(41) \quad \sum_{m=r+1}^n \sum_{v,v':v \wedge v'=m} P(\mathcal{E}^-(v) \cap \mathcal{E}^-(v')) \leq C2^r(n-r)^{2\varepsilon\frac{c}{\sigma^2}} .$$

Using (41), (40), and (39) in (38), we conclude that

$$P(\mathcal{N}^-(v) \geq 1) \geq \frac{1}{1 + C2^r e^{2\lambda\delta}}$$

which goes to 0 by taking $n \rightarrow \infty$ ($r = \log \log n$) than $\delta \rightarrow 0$. This proves the theorem. \square

4. UNIVERSALITY CLASSES OF LOG-CORRELATED FIELDS

The heuristics developed in Section 3 from branching random walks turns out to be applicable to a wide class of stochastic processes. It is a research topic of current interest to extend these techniques to other non-Gaussian log-correlated models with the purpose of proving results similar to Conjecture 1. One interesting problem that we will not have time to allude to is the cover time of the discrete random walk on the two-dimensional torus, which perhaps surprisingly, can be estimated using such methods. We refer the reader to [15, 29] for results on this question. We focus our attention here to two open problems: the large values of the Riemann zeta function and of the characteristic polynomials of random matrices. The behavior of the large values of these two models was conjectured by Fyodorov, Hiary & Keating [39, 40] to mimic that of a Gaussian log-correlated model. We shall explain here this connection in terms of branching random walks.

4.1. Maximum of the Riemann zeta function on an interval. Let $s \in \mathbb{C}$. If $\operatorname{Re} s > 1/2$, the Riemann zeta function is defined by

$$(42) \quad \zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ primes}} (1 - p^{-s})^{-1} .$$

This definition can be analytically continued using the functional equation

$$\zeta(s) = \chi(s)\zeta(1-s) , \quad \chi(s) = 2^s \pi^{s-1} \sin\left(\frac{\pi}{2}s\right) \Gamma(1-s) ,$$

to the whole complex plane with a pole at $s = 1$. The distribution of the primes is famously linked to information on the zeros of the function. There are trivial zeros (coming from the functional equation) at the negative even integers. The Riemann Hypothesis states that all other zeros lie on the critical line $\operatorname{Re} s = 1/2$ suggested by the symmetry of the functional equation. See Figure 5 for the plot of the modulus of ζ on the critical line.

We shall be interested instead in the large values of the modulus of ζ in an interval. In other words, what can be said about the local maxima of zeta? The global behavior is much more intricate, see the Lindelöf's hypothesis.

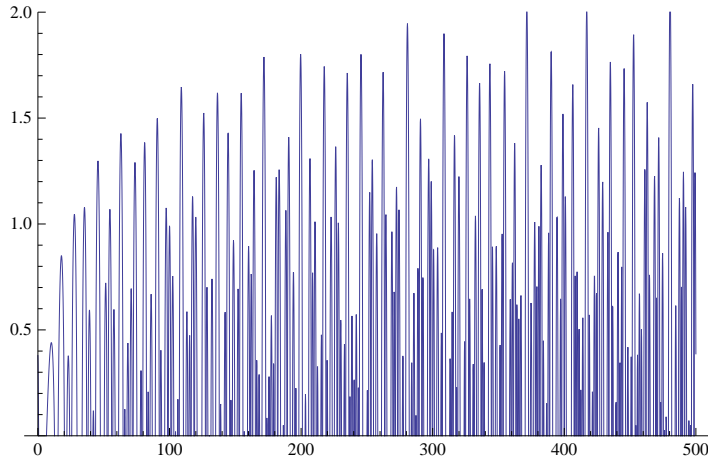


FIGURE 5. The function $\log |\zeta(s)|$ for $s = 1/2 + it$ for $t \in [0, 500]$.

Fyodorov, Hiary and Keating based on numerics and connection with random matrices made an astonishing prediction for the local maximum of ζ on an interval, say $[0, 1]$, of the critical line.

Conjecture 2 (Fyodorov–Hiary–Keating [39, 40]). *For τ sampled uniformly from $[0, T]$,*

$$(43) \quad \lim_{T \rightarrow \infty} \frac{\max_{h \in [0, 1]} \log |\zeta(1/2 + i(\tau + h))| - \log \log T}{\log \log \log T} = -\frac{3}{4} \text{ in probability.}$$

In fact, their conjecture is even more precise, suggesting an explicit distribution for the fluctuation. Roughly speaking, the above means that the maximum in a typical interval of the critical line is

$$\max_{h \in [0, 1]} \log |\zeta(1/2 + i(\tau + h))| \approx \log \log T - \frac{3}{4} \log \log \log T + O(1) .$$

The connection with Conjecture 1 is made by taking $\log T = 2^n$ and $\sigma^2 = \frac{\log 2}{2}$. Observe that it is a perfect match! As we will see, the relation with log-correlated models goes beyond the numerological curiosity. It also proposes a starting point to prove the conjecture. In fact, the conjecture can be proved to the subleading order for a random model of zeta, cf. Theorem 6.

Suppose first the Euler product of the Riemann zeta function extended to the critical line. By expanding the logarithm, we would have

$$(44) \quad \log |\zeta(1/2 + it)| = -\operatorname{Re} \sum_p \log(1 - p^{-(1/2 + it)}) = \sum_{k=1}^{\infty} \frac{1}{k} \sum_p \operatorname{Re} \frac{1}{p^{k(1/2 + it)}} = \sum_p \frac{\operatorname{Re} p^{-it}}{p^{1/2}} + O(1) .$$

This expansion can be made partially rigorous if the sum over primes is cut off at T , see Proposition 1 of [44] based on [56]. Now replace t by a random $\tau + h$ where $h \in [0, 1]$ and τ is sampled uniformly from $[0, T]$. The randomness entirely comes from the random variables $(p^{-i\tau}, p \text{ primes})$. It is an easy exercise by moment computations to verify that, since the values of $\log p$ are linearly independent for distinct primes, the process $(p^{-i\tau}, p \text{ primes})$ converges in the sense of finite-dimensional distribution as $T \rightarrow \infty$ to independent random variables distributed uniformly on the unit circle. This suggests the following model for zeta: let $(U_p, p \text{ primes})$ be IID uniform random variables on the unit

circle index by the primes and take

$$(45) \quad \left(\sum_{p \leq T} \frac{\operatorname{Re}(U_p p^{-ih})}{p^{1/2}}, h \in [0, 1] \right).$$

For this model, we can prove the conjecture up to subleading order.

Theorem 6 ([7]). *Let $(U_p, p \text{ primes})$ be independent random variables distributed uniformly on the unit circle. Then*

$$(46) \quad \lim_{T \rightarrow \infty} \frac{\max_{h \in [0, 1]} \sum_{p \leq T} \frac{\operatorname{Re}(U_p p^{-ih})}{p^{1/2}} - \log \log T}{\log \log T} = -\frac{3}{4} \text{ in probability.}$$

The proof is based on finding an approximate branching random walk embedded in zeta. What is this approximate branching random walk? In particular, what is the multiscale decomposition and does it exhibit the usual self-similarity and dichotomy of scales of log-correlated models? It is tempting to think of the decomposition $\sum_p \frac{\operatorname{Re} p^{-it}}{p^{1/2}}$ as the right multiscale decomposition since the variables $\operatorname{Re} p^{-it}$ decorrelate in the limit. However, for the scales to be self-similar, it is necessary to consider a coarse-graining of the sum. For the sake of clarity, we take $\log T = 2^n$. For the random model (45), the increment at scale l for $l = 1, \dots, n$ is

$$(47) \quad Y_h(l) = \sum_{2^{l-1} < \log p \leq 2^l} \frac{\operatorname{Re}(U_p p^{-ih})}{p^{1/2}}, h \in [0, 1].$$

The multiscale decomposition is therefore

$$\sum_{p \leq T} \frac{\operatorname{Re}(U_p p^{-ih})}{p^{1/2}} = \sum_{l=1}^n Y_h(l).$$

For the random model, the increments are independent by construction. For zeta, they are only independent in the limit which is a substantial difficulty. We verify that the $Y(l)$'s are self-similar and exhibit a dichotomy. The variance is easily calculated for the random model

$$E[(Y_h(k))^2] = \sum_{2^{k-1} < \log p \leq 2^k} \frac{1}{p} E[(\operatorname{Re} U_p)^2] = \frac{1}{2} \sum_{2^{k-1} < \log p \leq 2^k} \frac{1}{p}$$

where we used the fact that $E[U_p^2] = 0$ and $E[U_p \bar{U}_p] = 0$ where \bar{z} is the complex conjugate of z . The sum over primes can be evaluated using the *Prime Number Theorem*, see e.g. [53], which states that

$$(48) \quad \#\{p \leq x : p \text{ prime}\} = \int_2^x \frac{1}{\log u} du + O(xe^{-c\sqrt{\log x}}).$$

We get

$$E[(Y_h(l))^2] = \frac{\log 2}{2} + O(e^{-c\sqrt{2^l}}).$$

Thus the variances of the increments are approximately equal. Moreover, the parameter σ^2 should be taken to be $\log 2/2$. This motivates the above coarse-graining to get *self-similarity*. For the *dichotomy*, we need the equivalent of the branching scale. For $h, h' \in [0, 1]$, take

$$(49) \quad h \wedge h' = \log_2 |h - h'|^{-1}.$$

The covariance between increments at scale l is

$$E[Y_h(k)Y_{h'}(k)] = \frac{1}{2} \sum_{2^{k-1} < \log p \leq 2^k} \frac{\cos(|h - h'| \log p)}{p}.$$

Again, the sum can be evaluated using (48). This time, the oscillating nature of cosine will produce a dichotomy depending on the branching scale between h and h' . In fact, it is not hard to show that, see Lemma 2.1 in [7],

$$(50) \quad E[Y_k(h)Y_k(h')] = \begin{cases} \frac{\log 2}{2} + O(2^{-2(h \wedge h' - k)}) & \text{if } k \leq h \wedge h', \\ O(2^{-(k - h \wedge h')}) & \text{if } k > h \wedge h'. \end{cases}$$

The dichotomy is not as clean as for BRW. On the other hand, the increments couple and decouple exponentially fast in the scales. This turns out to be sufficient for the purpose of the subleading order.

To prove Theorem 6, there are additional difficulties. On one hand, the process is continuous and not discrete like BRW. Therefore, an argument is needed to show that a discrete set of 2^n points capture the order of the maximum. This is based on a chaining argument. Finally, the process here is not Gaussian so the estimate (1) cannot be used. It is replaced by precise large deviation principles obtained from estimates on the exponential moments of the sum.

Of course, there are tremendous technical obstacles to extend the method of proof of Theorem 6 to prove Conjecture 2 on the actual zeta function. For one, large deviation estimates are much harder to get. Moreover, the contribution to ζ of primes larger than T needs to be addressed. Despite these hurdles, the branching random walk heuristic gives a promising path to prove at least the leading order of the conjecture.

4.2. Maximum of the characteristic polynomial of random unitary matrices.

In this final section, we return to the problem we alluded to in Section 1.1. Let U_N be a random $N \times N$ unitary matrix sampled uniformly from the unitary group. This is often referred to as the *Circular Unitary Ensemble* (CUE). We consider the modulus of the characteristic polynomial on the unit circle

$$|P_{U_N}(\theta)| = |\det(e^{i\theta} - U_N)| = \prod_{j=1}^{2^{10}} |e^{i\theta} - e^{i\lambda_j}|.$$

where the eigenvalues of U_N are denoted by $(\lambda_j, j \leq N)$ and lie on the unit circle. A realization of this process was given for $N = 2^{10}$ in Figure 2. The prediction for the order of the maximum of this process is

Conjecture 3 (Fyodorov–Hiary–Keating [39, 40]). *For $N \in \mathbb{N}$, let U_N be a random matrix sampled uniformly from the group of $N \times N$ unitary matrices. Then*

$$(51) \quad \lim_{N \rightarrow \infty} \frac{\max_{\theta \in [0, 2\pi]} \log |P_{U_N}(\theta)| - \log N}{\log \log N} = -\frac{3}{4} \text{ in probability.}$$

In other words, it is expected that

$$\max_{\theta \in [0, 2\pi]} \log |P_{U_N}(\theta)| \approx \log N - \frac{3}{4} \log \log N + O(1).$$

The conjecture is well motivated by numerics and precise computations by Fyodorov & Keating of moments of the partition function of the models. They infer from the expression for the moments the order of magnitude of the moments as well as a prediction

for the fluctuations of the maximum assuming that the system undergoes a *freezing transition* similar to the random energy model. Since there is strong empirical evidence that the characteristic polynomial of CUE is a good model for the Riemann zeta function locally, the authors use in part Conjecture 3 to motivate Conjecture 2.

As for the Riemann zeta function, the conjecture is exactly what would be expected for the maximum of a log-correlated model with $N = 2^n$ random variables with variance $\sigma^2 n = \frac{\log 2}{2} n$, cf. Conjecture 1. It turns out that for a given θ , it was shown in [45] that $\log |P_{U_N}(\theta)|$ normalized by $\sqrt{\frac{1}{2} \log N}$ converges in distribution to a standard Gaussian. Therefore, the choice of σ^2 is already consistent with this result. Kistler’s multiscale refinement of the second moment method as described in Section 3 can be adapted to prove the leading order of the conjecture.

Theorem 7 ([6]). *For $N \in \mathbb{N}$, let U_N be a random matrix sampled uniformly from the group of $N \times N$ unitary matrices. Write $P_N(\theta)$, $\theta \in [0, 2\pi]$, for its characteristic polynomial on the unit circle. Then*

$$(52) \quad \lim_{N \rightarrow \infty} \frac{\max_{\theta \in [0, 2\pi]} \log |P_N(\theta)|}{\log N} = 1 \quad \text{in probability.}$$

In this section, we explain the connection to an approximate branching random walk which illustrates the approach to prove the theorem and the general conjecture.

To see what plays the role of the *multiscale decomposition*, write the characteristic polynomial as

$$\log |P_{U_N}(\theta)| = \sum_{j=1}^N \log |1 - e^{i(\lambda_j - \theta)}|.$$

By expanding the logarithm, we get

$$(53) \quad \log |P_{U_N}(\theta)| = \sum_{j=1}^N \sum_{k=1}^{\infty} -\frac{\operatorname{Re}(e^{ik(\lambda_j - \theta)})}{j} = \sum_{k=1}^{\infty} -\frac{\operatorname{Re}(e^{-ik\theta} \operatorname{Tr} U_N^k)}{k}$$

where Tr stands for the trace. It turns out that an integrable periodic function always has a pointwise convergent Fourier series wherever it is differentiable, so the above series makes sense for θ away from the eigenvalues. This already seems a good candidate for the *multiscale decomposition*. However, as it was the case for zeta, it will be necessary to group the traces of powers to get the equivalent of increments at each scale. In other words, the traces of the powers play the role of the primes.

To see this, we will need the seminal result of Diaconis & Shahshahani [34], see also [33], who proved that

$$(54) \quad \mathbb{E} \left[\operatorname{Tr} U_N^j \overline{\operatorname{Tr} U_N^k} \right] = \delta_{kj} \min(k, N).$$

where \mathbb{E} stands for the expectation under the uniform measure on the unitary group. It is also easy to see by rotation invariance of the uniform measure that $\mathbb{E} [\operatorname{Tr} U_N^j \operatorname{Tr} U_N^k] = 0$. This shows that traces of powers are *uncorrelated*. However, and this is one of the main issues, they are not independent for fixed N . The formula (54) is useful to our heuristic in many ways. First, an easy calculation shows that the variance of the powers

less than N is

$$E \left[\left(\sum_{k=1}^N -\frac{\operatorname{Re}(e^{-ik\theta} \operatorname{Tr} U_N^k)}{k} \right)^2 \right] = \frac{1}{2} \sum_{k \leq N} \frac{1}{k} = \frac{1}{2} \log N + O(1).$$

In particular, if we take $N = 2^n$ to emphasize the analogy with BRW, it leads us to define the *increment at scale l* for $l = 1, \dots, n$ as

$$Y_\theta(l) = \sum_{2^{l-1} < k \leq 2^l} -\frac{\operatorname{Re}(e^{-ik\theta} \operatorname{Tr} U_N^k)}{k}.$$

This gives

$$\mathbb{E}[(Y_\theta(l))^2] = \frac{\log 2}{2} + o(1),$$

where $o(1)$ is summable in l . From (53), the candidate for the multiscale decomposition is

$$(55) \quad \log |\mathbb{P}_{U_N}(\theta)| = \sum_{l=1}^n Y_\theta(l) + Z_\theta(n),$$

where $Z_\theta(n)$ stands for the sum of powers greater than n . Moreover, equation (54) suggests that the contribution of $Z_\theta(n)$ to $\log |\mathbb{P}_N(\theta)|$ in (53) should be of order 1 since $\sum_{j>N} \frac{N}{j^2} = O(1)$. We stress that, in this decomposition, the increments are uncorrelated though not independent. They are however asymptotically independent Gaussians, cf. [34]. Finally, they are almost *self-similar* in that regard since their variances are almost identical.

It remains to show the *dichotomy of scales* to complete the connection with BRW. A simple calculation using (54) gives that the covariance between increments for $\theta, \theta' \in [0, 2\pi]$ is

$$(56) \quad \mathbb{E}[Y_\theta(l)Y_{\theta'}(l)] = \sum_{2^{l-1} < k \leq 2^l} \frac{\cos(j\|\theta - \theta'\|)}{2k},$$

where $\|\theta - \theta'\|$ stands for the periodic distance on $[0, 2\pi]$. As it was the case for the Riemann zeta function, the presence of the cosine is responsible for the dichotomy. To see this, define the *branching scale*

$$(57) \quad \theta \wedge \theta' = -\log_2 \|\theta - \theta'\|.$$

For j such that $j\|h - h'\|$ is small the cosine is essentially 1, and for j such that $j\|\theta - \theta'\|$ is large the oscillation of the cosine leads to cancellation. This can be proved by Taylor expansion of \cos in the first case and summation by parts in the other. This argument shows

$$(58) \quad \mathbb{E}[Y_\theta(l)Y_{\theta'}(l)] = \begin{cases} \frac{1}{2} + O(2^{l-\theta \wedge \theta'}) & \text{if } l \leq \theta \wedge \theta', \\ O(2^{-2(l-\theta \wedge \theta')}) & \text{if } l > \theta \wedge \theta'. \end{cases}$$

In particular, this shows that the sum of traces of powers less than N are *log-correlated*:

$$\mathbb{E} \left[\left(\sum_{l=1}^n Y_\theta(l) \right) \left(\sum_{l=1}^n Y_{\theta'}(l) \right) \right] = \frac{\theta \wedge \theta'}{2} + O(1) = -\frac{1}{2} \log \|\theta - \theta'\| + O(1).$$

There are substantial difficulties to implement the method of Section 3 to prove Theorem 7 and more generally Conjecture 3. First, it is necessary to control the contribution

of the high powers, that is $Z_\theta(n)$ in (55). This turns out to be technically very challenging from a random matrix standpoint. Second, as opposed to BRW, the increments are not independent. This makes it much harder to get good large deviation estimates on the sum of the increments. Finally, since we are dealing with a continuous process on $[0, 2\pi]$ with singularities at the eigenvalues, one needs to justify that taking the maximum on $N = 2^n$ discrete points is enough to capture the order of the maximum.

REFERENCES

- [1] L. Addario-Berry and B. Reed, *Minima in branching random walks*, Ann. Probab. **37** (2009), no. 3, 1044–1079. MR2537549 (2011b:60338)
- [2] L. Addario-Berry and B. A. Reed, *Ballot theorems, old and new*, Horizons of combinatorics, 2008, pp. 9–35.
- [3] E. Aïdékon, *Convergence in law of the minimum of a branching random walk*, Ann. Probab. **41** (2013), no. 3A, 1362–1426. MR3098680
- [4] E. Aïdékon, J. Berestycki, É. Brunet, and Z. Shi, *Branching Brownian motion seen from its tip*, Probab. Theory Related Fields **157** (2013), no. 1-2, 405–451. MR3101852
- [5] E. Aïdékon and Z. Shi, *Weak convergence for the minimal position in a branching random walk: a simple proof*, Period. Math. Hungar. **61** (2010), no. 1-2, 43–54. MR2728431 (2011g:60153)
- [6] L.-P. Arguin, D. Belius, and P. Bourgade, *Maximum of the characteristic polynomial of random unitary matrices*, Preprint arXiv:1511.07399 (2015).
- [7] L.-P. Arguin, D. Belius, and A. J. Harper, *Maxima of a randomized riemann zeta function, and branching random walks*, Preprint arXiv:1506.00629 (2015).
- [8] L.-P. Arguin, A. Bovier, and N. Kistler, *Genealogy of extremal particles of branching Brownian motion*, Comm. Pure Appl. Math. **64** (2011), no. 12, 1647–1676. MR2838339 (2012g:60269)
- [9] ———, *Poissonian statistics in the extremal process of branching Brownian motion*, Ann. Appl. Probab. **22** (2012), no. 4, 1693–1711. MR2985174
- [10] ———, *The extremal process of branching Brownian motion*, Probab. Theory Related Fields **157** (2013), no. 3-4, 535–574. MR3129797
- [11] L.-P. Arguin and F. Ouimet, *Extremes of the two-dimensional gaussian free field with scale-dependent variance*, Preprint arXiv:1508.06253 (2015).
- [12] L.-P. Arguin and O. Zindy, *Poisson-Dirichlet statistics for the extremes of a log-correlated Gaussian field*, Ann. Appl. Probab. **24** (2014), no. 4, 1446–1481. MR3211001
- [13] ———, *Poisson-Dirichlet statistics for the extremes of the two-dimensional discrete Gaussian free field*, Electron. J. Probab. **20** (2015), no. 59, 19. MR3354619
- [14] M. Bachmann, *Limit theorems for the minimal position in a branching random walk with independent logconcave displacements*, Adv. in Appl. Probab. **32** (2000), no. 1, 159–176. MR1765165 (2001m:60189)
- [15] D. Belius and N. Kistler, *The subleading order of two dimensional cover times*, Preprint arXiv:1405.0888 (2014).
- [16] J. D. Biggins, *The first- and last-birth problems for a multitype age-dependent branching process*, Advances in Appl. Probability **8** (1976), no. 3, 446–459.
- [17] M. Biskup and O. Louidor, *Extreme local extrema of two-dimensional discrete gaussian free field*, Preprint arXiv:1306.2602 (2013).
- [18] E. Bolthausen, J.-D. Deuschel, and G. Giacomin, *Entropic repulsion and the maximum of the two-dimensional harmonic crystal*, Ann. Probab. **29** (2001), no. 4, 1670–1692. MR1880237 (2003a:82028)
- [19] A. Bovier, *Extremes*, Online lecture notes (2006). available at http://wt.iam.uni-bonn.de/fileadmin/WT/Inhalt/people/Anton_Bovier/lecture-notes/extreme.pdf.
- [20] ———, *From spin glasses to branching brownian motion – and back?*, Random walks, random fields, and disordered systems, 2015, pp. 1–64.
- [21] A. Bovier and L. Hartung, *The extremal process of two-speed branching brownian motion*, Elect. J. Probab. **19** (2014), no. 18, 1–28.
- [22] ———, *Variable speed branching Brownian motion 1. Extremal processes in the weak correlation regime*, ALEA Lat. Am. J. Probab. Math. Stat. **12** (2015), no. 1, 261–291. MR3351476

- [23] M. Bramson, J. Ding, and O. Zeitouni, *Convergence in law of the maximum of the two-dimensional discrete gaussian free field*, Preprint arXiv:1301.6669 (2013).
- [24] ———, *Convergence in law of the maximum of nonlattice branching random walk*, Preprint arXiv:1404.3423 (2014).
- [25] M. D. Bramson, *Maximal displacement of branching Brownian motion*, Comm. Pure Appl. Math. **31** (1978), no. 5, 531–581. MR0494541 (58 #13382)
- [26] D. Carpentier and P. Le Doussal, *Glass transition of a particle in a random potential, front selection in nonlinear renormalization group, and entropic phenomena in liouville and sinh-gordon models*, Phys. Rev. E **63** (2001Jan), 026110.
- [27] O. Daviaud, *Extremes of the discrete two-dimensional Gaussian free field*, Ann. Probab. **34** (2006), no. 3, 962–986.
- [28] L. de Haan and A. Ferreira, *Extreme value theory*, Springer Series in Operations Research and Financial Engineering, Springer, New York, 2006. An introduction. MR2234156 (2007g:62008)
- [29] A. Dembo, Y. Peres, J. Rosen, and O. Zeitouni, *Cover times for Brownian motion and random walks in two dimensions*, Ann. of Math. (2) **160** (2004), no. 2, 433–464. MR2123929 (2005k:60261)
- [30] B. Derrida, *Random-energy model: an exactly solvable model of disordered systems*, Phys. Rev. B (3) **24** (1981), no. 5, 2613–2626. MR627810 (83a:82018)
- [31] ———, *A generalisation of the random energy model that includes correlations between the energies*, J. Phys. Lett. **46** (1985), no. 3, 401–407.
- [32] B. Derrida and H. Spohn, *Polymers on disordered trees, spin glasses, and traveling waves*, J. Statist. Phys. **51** (1988), no. 5-6, 817–840. New directions in statistical mechanics (Santa Barbara, CA, 1987). MR971033 (90i:82045)
- [33] P. Diaconis and S. N. Evans, *Linear functionals of eigenvalues of random matrices*, Trans. Amer. Math. Soc. **353** (2001), no. 7, 2615–2633. MR1828463 (2002d:60003)
- [34] P. Diaconis and M. Shahshahani, *On the eigenvalues of random matrices*, J. Appl. Probab. **31A** (1994), 49–62. Studies in applied probability. MR1274717 (95m:60011)
- [35] J. Ding and O. Roy R. and Zeitouni, *Convergence of the centered maximum of log-correlated gaussian fields*, Preprint arXiv:1503.04588 (2015).
- [36] M. Fang and O. Zeitouni, *Branching random walks in time-inhomogeneous environments*, Electron. J. Probab. **17** (2012), no. 67, 18.
- [37] ———, *Slowdown for time inhomogeneous branching Brownian motion*, J. Stat. Phys. **149** (2012), no. 1, 1–9. MR2981635
- [38] Y. V. Fyodorov and J.-P. Bouchaud, *Freezing and extreme-value statistics in a random energy model with logarithmically correlated potential*, J. Phys. A **41** (2008), no. 37, 372001, 12. MR2430565 (2010g:82036)
- [39] Y. V. Fyodorov, G. A. Hiary, and J. P. Keating, *Freezing transition, characteristic polynomials of random matrices, and the riemann zeta function*, Phys. Rev. Lett. **108** (2012Apr), 170601.
- [40] 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** (2014), no. 2007, 20120503, 32. MR3151088
- [41] Y. V. Fyodorov, P. Le Doussal, and A. Rosso, *Statistical mechanics of logarithmic REM: duality, freezing and extreme value statistics of $1/f$ noises generated by Gaussian free fields*, J. Stat. Mech. Theory Exp. **10** (2009), P10005, 32. MR2882779
- [42] B. Gnedenko, *Sur la distribution limite du terme maximum d’une série aléatoire*, Ann. of Math. (2) **44** (1943), 423–453. MR0008655 (5,41b)
- [43] E. J. Gumbel, *Statistics of extremes*, Columbia University Press, New York, 1958. MR0096342 (20 #2826)
- [44] A. J. Harper, *A note on the maximum of the riemann zeta function, and log-correlated random variables*, Preprint arXiv:1304.0677 (2013).
- [45] J. P. Keating and N. C. Snaith, *Random matrix theory and $\zeta(1/2 + it)$* , Comm. Math. Phys. **214** (2000), no. 1, 57–89. MR1794265 (2002c:11107)
- [46] N. Kistler, *Derrida’s random energy models*, Correlated random systems: Five different methods, 2015, pp. 71–120.
- [47] S. P. Lalley and T. Sellke, *A conditional limit theorem for the frontier of a branching Brownian motion*, Ann. Probab. **15** (1987), no. 3, 1052–1061. MR893913 (88h:60161)

- [48] G. F. Lawler and V. Limic, *Random Walk: a Modern Introduction*, Cambridge Studies in Advanced Mathematics, vol. 123, Cambridge University Press, Cambridge, 2010.
- [49] M. R. Leadbetter, G. Lindgren, and H. Rootzén, *Extremes and related properties of random sequences and processes*, Springer Series in Statistics, Springer-Verlag, New York-Berlin, 1983. MR691492 (84h:60050)
- [50] T. Madaule, *Maximum of a log-correlated gaussian field*, Preprint arXiv:1307.1365 (2014).
- [51] ———, *Convergence in law for the branching random walk seen from its tip*, Journal of Theoretical Probability (2015), 1–37 (English).
- [52] P. Maillard and O. Zeitouni, *Slowdown in branching brownian motion with inhomogeneous variance*, Preprint arXiv:1307.3583 (2015).
- [53] H. L. Montgomery and R. C. Vaughan, *Multiplicative number theory. I. Classical theory*, Cambridge Studies in Advanced Mathematics, vol. 97, Cambridge University Press, Cambridge, 2007. MR2378655 (2009b:11001)
- [54] S. I. Resnick, *Extreme values, regular variation and point processes*, Springer Series in Operations Research and Financial Engineering, Springer, New York, 2008. Reprint of the 1987 original. MR2364939 (2008h:60002)
- [55] R. Rhodes and V. Vargas, *Gaussian multiplicative chaos and applications: A review*, Probab. Surveys **11** (2014), 315–392.
- [56] K. Soundararajan, *Moments of the Riemann zeta function*, Ann. of Math. (2) **170** (2009), no. 2, 981–993. MR2552116 (2010i:11132)
- [57] O. Zeitouni, *Branching random walks and gaussian fields*, Online lecture notes (2013). available at <http://www.wisdom.weizmann.ac.il/~zeitouni/notesGauss.pdf>.

L.-P. ARGUIN, DEPARTMENT OF MATHEMATICS, BARUCH COLLEGE AND GRADUATE CENTER,
CITY UNIVERSITY OF NEW YORK, NEW YORK, NY 10010, USA.

E-mail address: louis-pierre.arguin@baruch.cuny.edu