HAL open science

Revisiting classical and quantum disordered systems from the unifying perspective of large deviations

Cecile Monthus

▶ To cite this version:

Cecile Monthus. Revisiting classical and quantum disordered systems from the unifying perspective of large deviations. The European Physical Journal B: Condensed Matter and Complex Systems, 2019, 92 (7), pp.149. 10.1140/epjb/e2019-100151-9 . hal-02322039

HAL Id: hal-02322039

https://hal.science/hal-02322039v1

Submitted on 11 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

arXiv:1903.05899v2 [cond-mat.dis-nn] 13 May 2019

Revisiting classical and quantum disordered systems from the unifying perspective of large deviations

Cécile Monthus
Institut de Physique Théorique, Université Paris Saclay, CNRS, CEA, 91191 Gif-sur-Yvette, France

The theory of large deviations is already the natural language for the statistical physics of equilibrium and non-equilibrium. In the field of disordered systems, the analysis via large deviations is even more useful to describe within a unified perspective the typical events and the rare events that occur on various scales. In the present pedagogical introduction, we revisit various emblematic classical and quantum disordered systems in order to highlight the common underlying mechanisms from the point of view of large deviations.

I. INTRODUCTION

Just like Mr Jourdain discovering that he has been speaking in prose all his life without knowing it, physicists working in statistical physics become aware at some point that they have been using the theory of large deviations without realizing it since their very first acquaintance with the Boltzmann notion of entropy and the Gibbs theory of ensembles. This language of large deviations has turned out to be very powerful to unify the statistical physics of equilibrium, non-equilibrium and dynamical systems (see the reviews [1–3] and references therein) and to formulate an appropriate statistical physics approach of dynamical trajectories for various Markovian processes (see the reviews [4–10] and the PhD Theses [11–14] and the HDR Thesis [15]).

In the field of disordered systems, the presence of random disorder variables induce a lot of subtle effects for the probabilities of interesting observables. Physicists have understood from the very beginning that some observables are non-self-averaging, i.e. their disorder-averaged value is completely different from their typical value (see the books [16, 17] and references therein). It was also realized very early that in each large typical sample, there will nevertheless occur rare anomalous regions of a certain size that may dominate some observables : famous examples are the Lifshitz essential singularities of the density of states near spectrum edges in Anderson localization models [16, 18–21], the Griffiths singularities for the statics [22, 23] and the dynamics [24–26] of random classical models, and the Griffiths phases in random quantum models (see the reviews [27, 28] and references therein). Finally at critical points, it was found that multifractal properties appear, for instance for the inverse participation ratios of eigenfunctions at Anderson localization transitions (see the reviews [29, 30] and references therein) or for correlation functions in random classical spin models [31–38], while at Infinite Disorder fixed points, many observables are even more broadly distributed [27, 28]. These few examples indicate that that the language of large deviations is even more useful in the presence of disorder in order to describe within a unified perspective all these phenomena involving typical and rare events on various scales.

The aim of the present pedagogical introduction is thus to explain to physicists how the general theory of large deviations is the natural language to analyze the properties of various well-known classical and quantum random models. It is of course not meant for mathematicians who have been using the large deviation framework for a very long time (see the books [39–44] and references therein), in particular in the area of disordered systems (see the the books [45–47], the review [48] and references therein). This pedagogical introduction is thus intended only for physicists who are disheartened by the technical vocabulary used in the mathematical literature on large deviations (like Polish space, Borel sigma-field, cadlag function, ...).

The following sections are organized as follows. In section II, we introduce the generic notations for one-dimensional random models and describe how observables can be classified according to the order of the empirical property of the disorder configuration that determine them. We then analyze the various levels of this hierarchy : the 'Level 1' of large deviations allows to study the properties of observables given by products of random variables (section III); the 'Level 2' of large deviations corresponds to the fluctuations of the empirical 1-point histogram of the disorder configuration (section IV); the 'Level 2.5' of large deviations corresponds to the fluctuations of the empirical 2-point histogram of the disorder configuration (section V); finally the 'Level 3' of large deviations corresponds to the whole series of empirical histograms of arbitrary order (section VI). In Section VII, we turn to random models defined on Cayley trees to analyze their properties in terms of large deviations of branches. Our conclusions are summarized in section VIII. In Appendix A, we describe an alternative classification of one-dimensional disorder configurations in terms of empirical intervals where the disorder remains the same.

II. CLASSIFICATION OF OBSERVABLES IN ONE-DIMENSIONAL RANDOM MODELS

A. Transfer-matrix formulation of one-dimensional random models

Many classical and quantum disordered models in one dimension can be reformulated in terms of the product of random matrices (see the books [16, 17] and references therein). To have generic notations, it will be convenient to denote by $v(x)$ the disorder variable at point x that is drawn independently with some probability distribution p_v normalized to unity

$$\sum_v p_v = 1 \tag{1}$$

that should be translated into $\int dv p_v = 1$ whenever the disorder v is a continuous random variable. In this paper, we have chosen to write the general equations for the case of discrete disorder v (Eq 1) without the constant translation into the case of continuous disorder, but some examples of application will involve continuous disorder.

A disorder configuration $[v(.)] \equiv [v(x)]_{x=1,2,..,L}$ on a sample of L sites occurs with the factorized probability

$$\mathbb{P}_L[v(.)] \equiv p_{v(1)} p_{v(2)} ... p_{v(L)} = \prod_{x=1}^{L} p_{v(x)} \tag{2}$$

In this disordered sample, various physical observables can be then obtained by considering the product of the L corresponding transfer matrices $T_{v(x)}$ [16, 17]. One of the most important observable is the trace of this product

$$\mathcal{T}_L[v(.)] \equiv Trace \left[T_{v(L)} T_{v(L-1)} ... T_{v(2)} T_{v(1)} \right] \tag{3}$$

The exponential growth with L of its modulus $|\mathcal{T}_L[v(.)]|$ can be then measured by the finite-size Lyapunov exponent

$$\lambda[v(.)] \equiv \frac{\ln |\mathcal{T}_L[v(.)]|}{L} \tag{4}$$

Of course a more complete analysis would involve the whole Lyapunov spectrum [17] of the product of matrices but will not be discussed here.

B. Statistics of the Lyapunov exponent λ over the disorder configurations

For large L, the probability distribution $\mathcal{P}_L(\lambda)$ of the finite-size Lyapunov exponent λ of Eq. 4 over the disorder configurations $[v(.)]$ drawn with the probabilities of Eq. 2 is expected to follow the large deviation form [16, 17]

$$\mathcal{P}_L(\lambda) \equiv \sum_{[v(.)]} \mathbb{P}_L[v(.)] \delta \left(\lambda - \frac{\ln |\mathcal{T}_L[v(.)]|}{L} \right) \underset{L \to +\infty}{\simeq} e^{-L I(\lambda)} \tag{5}$$

where $I(\lambda)$ is called the 'rate function' in the field of large deviations : it is positive $I(\lambda) \geq 0$ and vanishes only at its minimum corresponding to the typical value λ^{typ} that will be realized with probability one in the thermodynamic limit $L \to +\infty$.

$$I(\lambda^{typ}) = 0 = I'(\lambda^{typ}) \tag{6}$$

All other values $\lambda \neq \lambda^{typ}$ appear with a probability $\mathcal{P}_L(\lambda)$ that is exponentially small in L in Eq. 5, but they are nevertheless important to understand the behavior of the moments of non-integer order k of the trace of Eq. 3, as a consequence of their evaluation via the Laplace saddle-point method of the following integral over λ

$$\overline{|\mathcal{T}_L[v(.)]|^k} \equiv \sum_{[v(.)]} P_L[v(.)] \, |\mathcal{T}_L[v(.)]|^k = \int d\lambda \, \mathcal{P}_L(\lambda) \, e^{L\lambda k} \underset{L \to +\infty}{\simeq} \int d\lambda \, e^{L[\lambda k - I(\lambda)]} \underset{L \to +\infty}{\simeq} e^{L\phi(k)} \tag{7}$$

The function $\phi(k)$ governing their exponential growth in L is called the 'scaled cumulant generating function' in the field of large deviations. It corresponds to the Legendre transform of the rate function $I(\lambda)$ of Eq. 5 as a consequence of the saddle-point evaluation of Eq. 7

$$\begin{aligned} \phi(k) &= \lambda k - I(\lambda) \\ 0 &= k - I'(\lambda) \end{aligned} \tag{8}$$

with the reciprocal Legendre transform

$$\begin{aligned} I(\lambda) &= \lambda k - \phi(k) \\ 0 &= \lambda - \phi'(k) \end{aligned} \tag{9}$$

For $k=0$ where $\phi(k=0)=0$ as a consequence of the normalization in Eq. 7, one obtains that the typical value λ^{typ} where the rate function vanishes (Eq. 6) corresponds to the derivative

$$\lambda^{typ} = \phi'(k=0) \tag{10}$$

while all moments of order $k \neq 0$ are dominated by non-typical values of the Lyapunov exponent in the saddle-point calculation of Eq. 7.

Since the typical Lyapunov exponent λ^{typ} appear with probability one in the thermodynamical limit $L \to +\infty$, one of the main goal in the field of products of random matrices has been to compute it in various models via the Dyson-Schmidt invariant measure method [16, 17, 49]. In the present paper, our goal will be instead to focus on the simplest cases where the whole large deviations rate function $I(\lambda)$ can be explicitly obtained.

C. Examples of observables corresponding to products of random variables

It is clear that the simplest case is of Eq. 3 is when the transfer matrices T_v are replaced by numbers t_v

$$\tau_L[v(.)] = t_{v(L)} t_{v(L-1)} ... t_{v(2)} t_{v(1)} = \prod_{x=1}^{L} t_{v(x)} \tag{11}$$

This case occurs in various disordered models, either exactly or approximately in some region of parameters, as displayed by the following examples.

1. *Examples of observables that are exactly given by products of random variables*

(1-a) In the classical Ising chain with random couplings $J(x)$, the two-spin correlation function reads [16, 17, 50]

$$C(x_0, x_0+r) = \prod_{x=x_0}^{x_0+r-1} \tanh(\beta J(x)) \tag{12}$$

(1-b) In the random quantum spin chains corresponding to free majorana fermions, the possible edge Majorana zero modes that characterize the topological phases are given in terms of product of random variables in the simplest cases (see [51] and references therein for various examples).

2. *Observables that can be approximated by products of random variables in certain regions of parameters*

(2-a) For the Anderson Localization tight-binding model with hopping V and random on-site-energy $\epsilon(x)$, the eigenfunction ψ_{x_0} localized on site x_0 for $V=0$ can be approximated at lowest order in the hopping V in the so-called Forward Approximation [52–55] by the product

$$\psi_{x_0}(x_0+r) \simeq \prod_{x=x_0+1}^{x_0+r} \left(\frac{V}{\epsilon(x_0) - \epsilon(x)} \right) \tag{13}$$

(2-b) For the quantum Ising chain with random couplings $J(x)$ and random transverse fields $h(x)$, the two-spin correlation function is given at lowest order in perturbation in the couplings by the product

$$C(x_0, x_0+r) = J(x_0) \prod_{x=x_0+1}^{x_0+r-1} \frac{J(x)}{h(x)} \tag{14}$$

This form can also be understood from the Strong Disorder RG approach [27, 28] when only sites are decimated, or from the Cavity approach [56–58].

D. Classification of observables in terms of empirical histograms of the disorder configuration

For each disorder configuration $[v(x)]_{x=1,2,..,L}$ with periodic boundary conditions $v(L+x)=v(x)$, the empirical 1-point histogram

$$Q_{v_1}[v(.)] \equiv \frac{1}{L}\sum_{x=1}^{L}\delta_{v_1,v(x+1)} \tag{15}$$

measures the frequencies of the possible values v_1 of the disorder variable. More generally, the empirical r-point histogram

$$Q_{v_r...v_2v_1}[v(.)] \equiv \frac{1}{L}\sum_{x=1}^{L}\delta_{v_r,v(x+r)}...\delta_{v_2,v(x+2)}\delta_{v_1,v(x+1)} \tag{16}$$

measures the frequencies of the occurrence of the r consecutive values $(v_r,...v_2,v_1)$ in the disordered sample. This hierarchy can be constructed up to the maximal value $r_{max}=L$ that corresponds to the total length L of the disorder configuration

$$Q_{v_Lv_{L-1}...v_2v_1}[v(.)] \equiv \frac{1}{L}\sum_{x=1}^{L}\delta_{v_L,v(x+L)}\delta_{v_{L-1},v(x+L-1)}...\delta_{v_2,v(x+2)}\delta_{v_1,v(x+1)} \tag{17}$$

i.e. this represents the average over the L translations via $x=1,2,..,L$ of the initial disorder configuration.

The observables of the disordered models can be then classified according to the order r of the empirical r-point histogram that allows to reconstruct them. For instance, the product of Eq. 11 can be rewritten in terms of the empirical 1-point histogram $Q_{v_1}[v(.)]$ of Eq. 15 as

$$\tau_L[v(.)] = \prod_{x=1}^{L} t_{v(x)} = \prod_{v_1}(t_{v_1})^{LQ_{v_1}[v(.)]} \tag{18}$$

The physical interpretation is that the product of random variables is not sensitive to the order of appearance of the disorder variables $v(x)$, but depends only on the global frequencies of the possible values v_1 that are summarized in the empirical 1-point histogram $Q_{v_1}[v(.)]$.

An example of observable that depends only on the empirical r-point histogram of Eq. 16 is the Spatial-Average within a given sample of the 2-point correlation function at distance r of Eq. 12 in a given sample

$$C^{SpAv}(r) \equiv \frac{1}{L}\sum_{x=1}^{L}C(x,x+r) = \frac{1}{L}\sum_{x=1}^{L}\prod_{y=x}^{x+r-1}\tanh(\beta J(y)) = \sum_{J_r}...\sum_{J_1}\left(\prod_{j=1}^{r}\tanh(\beta J_j)\right)Q_{J_r...J_2J_1}[J(.)] \tag{19}$$

whose statistics is discussed in [50] to stress that it will coincide with the disorder-averaged correlation function only for the small sizes $r \leq (cst)\ln L$. Finally, the most general observables depend on the empirical L-point histogram of Eq. 17 that contains the complete information on the disorder configuration.

The usefulness of this classification is that once one has identified that an observable $A[v(.)]$ depends on the disorder configuration $[v(.)]$ only via its empirical r-point histogram $Q_{(rpoints)}[v(.)]$ of Eq. 16

$$A[v(.)] = \mathcal{A}\left(Q_{(rpoints)}[v(.)]\right) \tag{20}$$

then its probability distribution over the disorder configurations drawn with Eq. 2 depends only on the probability distribution $P_L[Q_{(rpoints)}]$ of the empirical r-point histogram

$$\mathcal{P}_L(A) \equiv \sum_{[v(.)]}\mathbb{P}_L[v(.)]\delta\left(A-\mathcal{A}\left(Q_{v_r...v_2v_1}[v(.)]\right)\right) = \sum_{[Q_{(rpoints)}]}P_L[Q_{(rpoints)}]\delta\left(A-\mathcal{A}\left(Q_{(rpoints)}\right)\right) \tag{21}$$

In the theory of large deviations, it turned that the probability distributions $P_L[Q_{(rpoints)}]$ of the empirical r-point histograms of various order r have been labelled by levels as follows [1, 3] : the Level 2 corresponds to the empirical 1-point histogram $Q_.$, the Level 2.5 corresponds to the empirical 2-point histogram $Q_{..}$, the Level 3 corresponds to the full hierarchy of arbitrary r up to the limit $r \to +\infty$. In the following sections, we will thus describe this hierarchy, starting with the Level 1 that corresponds to the large deviations properties of sums of random variables, that are important to fully characterize the statistics of products of random variables.

III. PRODUCT OF RANDOM VARIABLES AS THE LEVEL-1 OF LARGE DEVIATIONS

In this section, we focus on the product of random variables corresponding to the modulus of Eq 11

$$|\tau_L[v(.)]| = \prod_{x=1}^{L} |t_{v(x)}| \tag{22}$$

and on the corresponding finite-size Lyapunov exponent of Eq. 4

$$\lambda[v(.)] \equiv \frac{\ln |\tau_L[v(.)]|}{L} = \frac{1}{L}\sum_{x=1}^{L} \ln |t_{v(x)}| \tag{23}$$

As explained in detail in the previous section, this is the simplest problem that occur in the field of disordered systems. In the language of large deviations, the properties of the sum of random variables of Eq. 23 is also the simplest example corresponding to the so-called 'Level-1' description [1, 3].

A. Moments of non-integer order k

The moments of non-integer order k of the product in Eq. 22 can be directly computed as a consequence of the independence of the disorder variables $v(x)$ on the L sites (Eq. 2)

$$\overline{|\tau_L[v(.)]|^k} = \overline{\prod_{x=1}^{L} |t_{v(x)}|^k} = \left[\overline{|t_v|^k}\right]^L \tag{24}$$

So the scaled cumulant generating function $\phi(k)$ governing their exponential growth in L (Eq 7) is given, actually even for any finite L, by the simple expression

$$\phi(k) = \frac{\ln\left(\overline{|\tau_L[v(.)]|^k}\right)}{L} = \ln\left[\overline{|t_v|^k}\right] \tag{25}$$

in terms of the moments $\overline{|t_v|^k}$ of the elementary variable $|t_v|$.

B. Rate function $I(\lambda)$ governing the large deviations of the Lyapunov exponent λ

The rate function $I(\lambda)$ governing the large deviations (Eq 5) of the Lyapunov exponent λ of Eq. 23 can be computed either directly if the probability distribution of the sum of Eq. 23 is known or it can be obtained via the reciprocal Legendre transform (Eq. 9) from the knowledge of the function $\phi(k)$ of Eq. 25. Let us now recall some simple examples that will be useful later (in section VII).

C. Examples for the equilibrium of disordered classical models

In the field of disordered classical models, the simplest example is when the variable $t_{v(x)}$ corresponds to the Boltzmann weight at inverse temperature β of the random potential $v(x)$

$$t_{v(x)} = e^{\beta v(x)} \tag{26}$$

Then Eq. 22 represents the Boltzmann weight of the L sites

$$\tau_L[v(.)] = \prod_{x=1}^{L} e^{\beta v(x)} \tag{27}$$

and Eq. 23 corresponds to the energy per site (up to the factor β)

$$\lambda[v(.)] = \beta\frac{1}{L}\sum_{x=1}^{L} v(x) \tag{28}$$

For instance if the distribution of the potential v is Gaussian of zero mean

$$p^{Gauss}(v) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{v^2}{2\sigma^2}} \tag{29}$$

then both the rate function $I(\lambda)$ and the scaled cumulant generating function $\phi(k)$ are simply quadratic

$$\begin{aligned} I^{Gauss}(\lambda) &= \frac{\lambda^2}{2\beta^2\sigma^2} \\ \phi^{Gauss}(k) &= \frac{k^2\beta^2\sigma^2}{2} \end{aligned} \tag{30}$$

Another example is when the distribution of the potential v is the Bernoulli distribution

$$p_v^{Bernoulli} = p\delta(v - v_0) + (1-p)\delta(v) \tag{31}$$

then the rate function $I(\lambda)$ and the scaled cumulant generating function $\phi(k)$ read

$$\begin{aligned} I^{Bernoulli}(\lambda) &= \frac{\lambda}{\beta v_0} \ln\left(\frac{\lambda}{p\beta v_0}\right) + \left(1 - \frac{\lambda}{\beta v_0}\right) \ln\left(\frac{1 - \frac{\lambda}{\beta v_0}}{1-p}\right) \\ \phi^{Bernoulli}(k) &= \ln\left[p e^{k\beta v_0} + (1-p)\right] \end{aligned} \tag{32}$$

So it is important to stress here that the large deviations properties depend on all the details of the disorder distribution p_v, in contrast to the small deviations region described by the Central-Limit-Theorem that corresponds to the expansion at lowest order of the rate function $I(\lambda)$ around its vanishing minimum at the typical value λ_{typ} of Eq. 6

$$I^{CLT}(\lambda) = \frac{I''(\lambda_{typ})}{2}(\lambda - \lambda_{typ})^2 + o((\lambda - \lambda_{typ})^2) \tag{33}$$

D. Examples for disordered quantum models

For the Anderson Localization model in the Forward approximation of Eq. 13, it is usual to consider the box distribution of width $(2W)$ for the random on-site energy $\epsilon(x)$

$$p^{Box}(\epsilon) = \frac{\theta(-W \leq \epsilon \leq W)}{2W} \tag{34}$$

The elementary variable $t_{\epsilon(x)}$ in the product in Eq. 13 at the center of the band $\epsilon(x_0) = 0$

$$t_{\epsilon(x)} = \frac{V}{|\epsilon(x)|} \tag{35}$$

has then moments only in the region $k < 1$

$$\overline{|t_\epsilon|^k} = V^k \int_0^W \frac{d\epsilon}{W} \epsilon^{-k} = \left(\frac{V}{W}\right)^k \frac{1}{1-k} \tag{36}$$

So the scaled cumulant generating function $\phi(k)$ of Eq. 25 reads

$$\phi^{Anderson}(k) = \ln\left[\overline{|t_\epsilon|^k}\right] = k \ln\left(\frac{V}{W}\right) - \ln(1-k) \tag{37}$$

with the corresponding rate function

$$I^{Anderson}(\lambda) = \lambda - \ln\left(\frac{V}{W}\right) - 1 - \ln\left[\lambda - \ln\left(\frac{V}{W}\right)\right] \tag{38}$$

IV. EMPIRICAL 1-POINT HISTOGRAM AS THE LEVEL-2 OF LARGE DEVIATIONS

In this section, we focus on the probability of the empirical 1-point histogram of Eq. 15 over the disorder configurations $v(.)$ drawn with Eq. 2

$$P_L[Q_.] \equiv \sum_{[v(.)]} \mathbb{P}_L[v(.)] \prod_{v_1} \delta\left(Q_{v_1} - \frac{1}{L}\sum_{x=1}^{L} \delta_{v_1,v(x+1)}\right) \tag{39}$$

Of course the typical value of this histogram is the 'true' probability distribution p_v of the disorder (Eq. 1)

$$Q_v^{typ} = p_v \tag{40}$$

but here the goal is to describe its fluctuations for large L. In the language of large deviations [1–3], this is known as the 'Level-2 description of the empirical measure'. The essential result is the large deviation form for large L

$$P_L[Q_.] \underset{L\to+\infty}{\simeq} C_1[Q_.] e^{-LS^{rel}(Q_.|p_.)} \tag{41}$$

where

$$C_1[Q_.] = \delta\left(1 - \sum_v Q_v\right) \tag{42}$$

represents the normalization constraint of the empirical histogram (the notation $\delta(Y)$ represents the discrete Kronecker symbol $\delta_{0,Y}$ but has been chosen here for better readability of the argument Y), while the rate function is the relative entropy of the empirical 1-point histogram Q_v with respect to the true probability distribution p_v of the disorder

$$S^{rel}(Q_.|p_.) \equiv \sum_v Q_v \ln \frac{Q_v}{p_v} \tag{43}$$

This result is known as the Sanov theorem in the field of large deviations [1–3] and can be considered as the true cornerstone of the whole theory, with many further generalizations for the higher levels. It is thus important to fully understand its origin and its physical meaning, via the three following different derivations.

A. First approach via the multinomial distribution

Since each disorder value $v(x)$ is drawn with probability $p_{v(x)}$ independently on each of the L sites $x = 1, 2, .., L$ (Eq 2), the probability of the empirical 1-point histogram $Q_.$ of Eq. 15 amounts to analyze the integer numbers (LQ_v) of the occurrences of each value v and is thus given by the multinomial distribution

$$P_L[Q_.] = \delta\left(1 - \sum_v Q_v\right) \frac{L!}{\prod_v (LQ_v)!} \prod_v [p_v]^{LQ_v} \tag{44}$$

The Stirling's approximation for the factorials $m! \simeq \sqrt{2\pi m}\; m^m e^{-m}$ then yields the large deviation form of Eq. 41 with the relative entropy of Eq. 43. This derivation based on the application of the Stirling's approximation to the multinomial distribution of Eq. 44 goes back to Boltzmann [2] and appears in all statistical physics lectures.

B. Second approach via the generating function

Another derivation is based on the generating function of the empirical 1-point histogram of Eq. 39

$$\begin{aligned} G_L[\nu_.] &\equiv \sum_{Q_.} P_L[Q_.]\; e^{L\sum_v \nu_v Q_v} = \sum_{v(1)} ... \sum_{v(L)} p_{v(1)} p_{v(2)} ... p_{v(L)}\; e^{\sum_v \nu_v \sum_{x=1}^{L} \delta_{v,v(x)}} \\ &= \prod_{x=1}^{L} \left(\sum_{v(x)} p_{v(x)} e^{\nu_{v(x)}}\right) = \left(\sum_v p_v e^{\nu_v}\right)^L \end{aligned} \tag{45}$$

This factorized form is valid already for any finite L and the corresponding scaled cumulant generating function $\Phi[\nu_.]$ governing the exponential growth with L

$$G_L[\nu_.] \quad = e^{L\Phi[\nu_.]} \tag{46}$$

is given in terms of the generating function of the disorder distribution p_v

$$\Phi[\nu_.] \quad = \ln\left(\sum_v p_v e^{\nu_v}\right) \tag{47}$$

where the analogy with Eq. 25 is clear. It is now useful to show the link with the the relative entropy of Eq. 43 via the Legendre transform and the reciprocal Legendre transform respecify.

1. *Link with the relative entropy via the Legendre transform*

The generating function of Eq 45 can be rewritten in terms of Eq. 41 as

$$G_L[\nu_.] \quad \equiv \sum_{Q_.} P_L[Q_.] e^{L\sum_v \nu_v Q_v} \underset{L\to+\infty}{\simeq} \sum_{Q_.} \delta\left(1-\sum_v Q_v\right) e^{L\left[\sum_v \nu_v Q_v - S^{rel}(Q_.|p_.)\right]} \tag{48}$$

The Laplace's saddle point method for large L yields that one should optimize over $Q_.$ the function in the exponential in the presence of the normalization constraint $(1-\sum_v Q_v)$ in order to obtain the function $\Phi[\nu_.]$ of Eq. 46

$$\Phi[\nu_.] \quad = \max_{Q_.:1=\sum_v Q_v} \left[\sum_v \nu_v Q_v - S^{rel}(Q_.|p_.)\right] = \max_{Q_.:1=\sum_v Q_v} \left[\sum_v \nu_v Q_v - \sum_v Q_v \ln\frac{Q_v}{p_v}\right] \tag{49}$$

Taking into account the constraint via some Lagrange multiplier η, one needs to optimize the functional

$$\mathcal{L}(Q_.) = \sum_v Q_v\left[\nu_v - \ln\frac{Q_v}{p_v}\right] + \eta\left(1-\sum_v Q_v\right) \tag{50}$$

over the values Q_v

$$0 = \frac{\partial \mathcal{L}(Q_.)}{\partial Q_v} = \nu_v - \ln\frac{Q_v}{p_v} - 1 - \eta \tag{51}$$

One obtains the optimal solution

$$Q_v^* = p_v e^{\nu_v - 1 - \eta} \tag{52}$$

where the Lagrange multiplier η is fixed by the constraint

$$1 = \sum_v Q_v^* = e^{-1-\eta}\sum_v p_v e^{\nu_v} \tag{53}$$

The optimal value of the functional of Eq. 50

$$\mathcal{L}(Q_.^*) = \sum_v Q_v^*\left[\nu_v - \ln\frac{Q_v^*}{p_v}\right] = 1+\eta = \ln\left(\sum_v p_v e^{\nu_v}\right) = \Phi[\nu_.] \tag{54}$$

indeed coincides with the result of Eq. 47.

2. Link with the relative entropy via the reciprocal Legendre transform

The reciprocal Legendre transform of Eq. 49 reads

$$S^{rel}(Q_.|p_.) \quad = \max_{\nu_.} \left[\sum_v \nu_v Q_v - \Phi[\nu_.] \right] = \max_{\nu_.} \left[\sum_v \nu_v Q_v - \ln \left(\sum_v p_v e^{\nu_v} \right) \right] \tag{55}$$

The optimization over ν_v

$$0 \quad = \frac{\partial}{\partial \nu_v} \left[\sum_{v'} \nu_{v'} Q_{v'} - \ln \left(\sum_{v'} p_{v'} e^{\nu_{v'}} \right) \right] = Q_v - \frac{p_v e^{\nu_v}}{\sum_{v'} p_{v'} e^{\nu_{v'}}} \tag{56}$$

yields the optimal solution

$$\nu_v^* = \ln \left[\frac{Q_v \left(\sum_{v'} p_{v'} e^{\nu_{v'}} \right)}{p_v} \right] \tag{57}$$

and the optimal value of the functional of Eq. 55

$$\max_{\nu_.} \left[\sum_v \nu_v Q_v - \ln \left(\sum_v p_v e^{\nu_v} \right) \right] = \sum_v \nu_v^* Q_v - \ln \left(\sum_v p_v e^{\nu_v^*} \right) = \sum_v Q_v \ln \left[\frac{Q_v}{p_v} \right] = S^{rel}(Q_.|p_.) \tag{58}$$

coincides with the relative entropy as it should.

These calculations based on generating functions, Laplace's saddle-point method with constraints taken into account via Lagrange multipliers, and Legendre transforms are very standard both in statistical physics and in the theory of large deviations.

C. Third approach via some appropriate change of measure

The third approach via some appropriate change of measure is very common in the whole field of large deviations, but appears to be less well known among physicists. It seems thus useful to explain it here in more physical terms than usual. The starting point is that the probability of the disorder configuration $[v(x)]_{x=1,2,..,L}$ of Eq. 2 can be rewritten only in terms of the empirical 1-point histogram of Eq. 15

$$\mathbb{P}_L[v(.)] = e^{\sum_{x=1}^{L} \ln(p_{v(x)})} = e^{L \sum_v Q_v \ln(p_v)} \tag{59}$$

So all the disorder configurations that have the same empirical 1-point histogram $Q_.$ have the same probability in Eq. 59. As a consequence, the normalization of Eq. 59 over all disorder configurations $[v(.)]$ can be rewritten as a sum over the possible empirical 1-point histogram $Q_.$

$$1 = \sum_{[v(.)]} \mathbb{P}_L[v(.)] = \sum_{Q_.} C_1[Q_.] \Omega_L[Q_.] e^{L \sum_v Q_v \ln(p_v)} \tag{60}$$

where

$$\Omega_L[Q_.] \equiv \sum_{v(.)} \prod_v \left(Q_v - \frac{1}{L} \sum_{x=1}^{L} \delta_{v,v(x)} \right) \tag{61}$$

counts the number of disorder configurations that are associated to the same value $Q_.$ of the empirical histogram. So the probability $P_L[Q_.]$ of Eq 39 to observe the empirical histogram $Q_.$ reads

$$P_L[Q_.] \quad = C_1[Q_.] \Omega_L[Q_.] e^{L \sum_v Q_v \ln(p_v)} \tag{62}$$

When the empirical 1-point histogram takes its typical value p_v of Eq. 40, the probability of Eq. 62

$$P_L[Q_.^{typ}] \quad = P_L[p_.] = \Omega_L[p_.] e^{L \sum_v p_v \ln(p_v)} \tag{63}$$

should not decay exponentially in L, so that $\Omega_L[p_.]$ should grow exponentially in L in order to compensate exactly the other exponential factor

$$\Omega_L[p_.] \underset{L \to +\infty}{\simeq} e^{-L \sum_v p_v \ln(p_v)} \tag{64}$$

To obtain the behavior of $\Omega_L[Q_.]$ when the empirical 1-point histogram $Q_.$ is different from its typical value $Q_.^{typ} = p_.$, we may consider a modified model where the disorder is drawn with the modified probability $\tilde{p}_v = Q_v$ that will make Q_v typical for this modified model, and one obtains

$$\Omega_L[Q_.] \underset{L \to +\infty}{\simeq} e^{L S_1[Q_.]} \tag{65}$$

where

$$S_1[Q_.] \equiv - \sum_v Q_v \ln(Q_v) \tag{66}$$

represents the entropy of the empirical 1-point histogram $Q_.$. Plugging this result into Eq. 62 yields that the large deviation behavior of the probability of the empirical 1-point histogram

$$P_L[Q_.] \quad = C_1[Q_.] \Omega_L[Q_.] e^{L \sum_v Q_v \ln(p_v)} \underset{L \to +\infty}{\simeq} C_1[Q_.] \; e^{-L \sum_v Q_v \ln\left(\frac{Q_v}{p_v}\right)} = C_1[Q_.] \; e^{L S^{rel}(Q_.|p_.)} \tag{67}$$

involves again the relative entropy $S^{rel}(Q_.|p_.)$ as it should to recover Eq 41 and Eq. 43.

This idea to evaluate the large deviations properties of the untypical values of the empirical observable via the introduction of a modified model that make this empirical observable typical is used extensively in the field of large deviation for the two following reasons. From the conceptual point of view, this way of thinking is very illuminating because it shows very clearly why the entropy $S_1[Q_.]$ appears in Eq. 65 and why the relative entropy $S^{rel}(Q_.|p_.)$ appears in Eq 67. From the technical point of view, it is extremely powerful, since it allows to obtain directly the results without any actual computations : indeed, one does not need to use combinatorics to enumerate the appropriate configurations in finite size as in Eq. 44, and one does not need either to compute the generating function of Eq. 45 and to perform the reciprocal Legendre transform, but one obtains directly the rate function from simple considerations. In the following sections concerning the more complicated cases of empirical histograms of higher orders, as well as in the Appendix, we will see how this approach can be adapted to each purpose in order to obtain directly the appropriate rate functions without any calculation.

V. EMPIRICAL 2-POINT HISTOGRAM AS THE LEVEL 2.5 OF LARGE DEVIATIONS

In this section, we focus on the probability of the empirical 2-point histogram of Eq. 16 for $r = 2$ over the disorder configurations $[v(.)]$ drawn with Eq. 2

$$P_L[Q_{..}] \equiv \sum_{[v(.)]} \mathbb{P}_L[v(.)] \prod_{v_2} \prod_{v_1} \delta \left(Q_{v_2 v_1} - \frac{1}{L} \sum_{x=1}^{L} \delta_{v_2, v(x+2)} \delta_{v_1, v(x+1)} \right) \tag{68}$$

Its large deviations properties have been analyzed in the context of Markov chains [3, 11, 59]. Together with its analog formulations for Markov jump processes in continuous time [11, 15, 60–69] and for diffusion processes [15, 63, 64, 70, 71], it is nowadays called the 'Level 2.5' in the field of large deviations.

A. Constraints on the empirical 2-point histogram $Q_{..}$

Since the empirical 1-point histogram $Q_{.}$ can be reconstructed by summing over the last or the first value of the empirical 2-point histogram $Q_{..}$, it is convenient to introduce the following notation to summarize these constraints

$$C_2[Q_{..}, Q_{.}] = \prod_{v_1} \left[\delta \left(Q_{v_1} - \sum_{v_2} Q_{v_2 v_1} \right) \delta \left(Q_{v_1} - \sum_{v_2} Q_{v_1 v_2} \right) \right] \tag{69}$$

while the empirical 1-point histogram $Q_{.}$ should of course still satisfy the normalization constraint $C_1[Q_{.}]$ of Eq. 42.

B. Generalized Markovian model for the disorder

In order to analyze the statistical properties of the empirical 2-point histogram, it is useful to introduce a generalized model where the disorder configurations are generated by a Markov chain where the transition probability matrix $W_{v'v}$ to go from v to v' is normalized to unity

$$\sum_{v'} W_{v'v} \quad = 1 \tag{70}$$

The probability of Eq. 2 for a disorder configuration $[v(x)]_{x=1,2,..,L}$ is thus replaced by the product of the transition probabilities along the configuration (up to boundary terms that become negligible for large $L \to +\infty$)

$$\mathbb{P}_L^{Markov}[v(.)] \simeq \prod_{x=1}^{L} W_{v(x+1)v(x)} = e^{L \sum_{x=1}^{L} \ln \left(W_{v(x+1)v(x)} \right)} \tag{71}$$

It is also useful to introduce the stationary state ρ_v of this Markov chain satisfying

$$\rho_{v'} \quad = \sum_{v} W_{v'v} \rho_v \tag{72}$$

with the normalization

$$1 = \sum_{v} \rho_v \tag{73}$$

For this generalized Markovian model, the typical value of the empirical 1-point histogram of Eq. 15 is simply the stationary state ρ_v introduced in Eq. 72

$$Q_v^{typ} \quad = \rho_v \tag{74}$$

while the typical value of the empirical 2-point histogram is given by the corresponding flow appearing in Eq. 72

$$Q_{v'v}^{typ} = W_{v'v} \rho_v \tag{75}$$

that satisfy the constraints of Eqs 69 and Eq 42.

Since the probability of Eq. 71 can be rewritten only in terms of the empirical 2-point histogram $Q_{v'v}$ as

$$P_L^{Markov}[v(.)] \simeq e^{L \sum_{v'} \sum_{v} Q_{v'v} \ln \left(W_{v'v} \right)} \tag{76}$$

the normalization over disorder configurations can be rewritten as a sum over the empirical 1-point and 2-point histograms with their constraints of Eq. 42 and Eq 69 as

$$1 = \sum_{v(.)} P_L^{Markov}[v(.)] = \sum_{Q_{.}} C_1[Q_{.}] \sum_{Q_{..}} C_2[Q_{..}, Q_{.}] \; \Omega_L[Q_{..}, Q_{.}] e^{L \sum_{v'} \sum_{v} Q_{v'v} \ln \left(W_{v'v} \right)} \tag{77}$$

where $\Omega_L[Q_{..},Q_{.}]$ counts the number of disorder configurations that have the empirical observables $[Q_{..},Q_{.}]$ and is thus the direct generalization of Eq. 61, while the probability to observe these empirical observables reads

$$P_L[Q_{..},Q_{.}] = C_1[Q_{.}]C_2[Q_{..},Q_{.}]\ \Omega_L[Q_{..},Q_{.}]e^{L\sum_{v'}\sum_{v} Q_{v'v}\ln\left(W_{v'v}\right)} \tag{78}$$

For the typical values of Eq. 74 and Eq 75 of the empirical observables, this probability should not be exponentially small in L so that $\Omega_L[Q_{..}^{typ},Q_{.}^{typ}]$ should exactly compensate the other exponential factor in Eq. 78

$$\Omega_L[Q_{..}^{typ},Q_{.}^{typ}] \underset{L\to+\infty}{\simeq} e^{-L\sum_{v'}\sum_{v} Q_{v'v}^{typ}\ln\left(W_{v'v}\right)} \tag{79}$$

For other values of the empirical observables, one may consider a modified Markov transition matrix $\tilde{W}_{v'v}$ that would make these empirical histograms typical : Eqs 74 and 75 yields that the appropriate choice is

$$\tilde{W}_{v'v} = \frac{Q_{v'v}}{Q_v} \tag{80}$$

so that Eq 79 becomes

$$\Omega_L[Q_{..},Q_{.}] \underset{L\to+\infty}{\simeq} e^{-L\sum_{v'}\sum_{v} Q_{v'v}\ln\left(\tilde{W}_{v'v}\right)} = e^{-L\sum_{v'}\sum_{v} Q_{v'v}\ln\left(\frac{Q_{v'v}}{Q_v}\right)} \underset{L\to+\infty}{\simeq} e^{L\left(S_2[Q_{..}]-S_1[Q_{.}]\right)} \tag{81}$$

where $S_2[Q_{..}]$ represents the entropy of the empirical 2-point histogram $Q_{..}$

$$S_2[Q_{..}] \equiv -\sum_{v'}\sum_{v} Q_{v'v}\ln\left(Q_{v'v}\right) \tag{82}$$

while $S_1[Q_{.}]$ is the entropy of the empirical 1-point histogram $Q_{.}$ introduced in Eq 66.

Plugging Eq. 81 into Eq 78 yields the large deviation form [3, 11, 59]

$$P_L[Q_{..},Q_{.}] \underset{L\to+\infty}{\simeq} C_1[Q_{.}]C_2[Q_{..},Q_{.}]e^{-L\sum_{v'}\sum_{v} Q_{v'v}\ln\left(\frac{Q_{v'v}}{W_{v'v}Q_v}\right)} \tag{83}$$

that is called nowadays the 'Level 2.5' for Markov chains. The rate function can be interpreted as the relative entropy for Markov chains [3, 11, 59]. The analog results have been much studied for Markov jump processes in continuous time [11, 15, 60–69] and for diffusion processes [15, 63, 64, 70, 71],

C. Return to the initial disorder of Eq. 2

The initial disorder model of Eq. 2 corresponds to the special case where the Markov matrix of Eq. 70 reduces to

$$W_{v'v} = p_{v'} \tag{84}$$

Then Eq 83 simplifies into

$$P_L[Q_{..},Q_{.}] \underset{L\to+\infty}{\simeq} C_1[Q_{.}]C_2[Q_{..},Q_{.}]e^{-L\sum_{v'}\sum_{v} Q_{v'v}\ln\left(\frac{Q_{v'v}}{p_{v'}Q_v}\right)} = C_1[Q_{.}]\ e^{LS^{rel}(Q_{.}|p_{.})}C_2[Q_{..},Q_{.}]\ e^{T(S_2[Q_{..}]-2S_1[Q_{.}])} \tag{85}$$

In the last expression, one recognizes the probability $P_L[Q_{.}]$ of the empirical 1-point histogram $Q_{.}$ of Eq. 67. This yields that the conditional probability to observe the empirical 2-point histogram $Q_{..}$ once the empirical 1-point histogram $Q_{.}$ is given reads

$$P_L[Q_{..}|Q_{.}] = \frac{P_L[Q_{..},Q_{.}]}{P_L[Q_{.}]} \underset{L\to+\infty}{\simeq} C_2[Q_{..},Q_{.}]e^{L(S_2[Q_{..}]-2S_1[Q_{.}])} \tag{86}$$

In particular, once the empirical 1-point histogram $Q_{.}$ is given, the typical value of the empirical 2-point histogram $Q_{..}$ is simply the product

$$Q_{v_2v_1}^{typ} = Q_{v_2}Q_{v_1} \tag{87}$$

as it should, while Eq 86 described the large deviations away from this typical value.

VI. EMPIRICAL HIGHER ORDER HISTOGRAMS AS THE LEVEL 3 OF LARGE DEVIATIONS

In the language of large deviations, the Level 3 actually denotes the empirical process that can be constructed from the knowledge of the empirical r-point histogram in the limit $r \to +\infty$ [1, 3]. In this section, we will not be interested into taking this limit, but we wish to analyze the hierarchy of the empirical r-point histograms of arbitrary order r up to the maximal value $r_{max} = L$ (Eq 17), in order to characterize the sample-to-sample fluctuations for a disordered ring of large size L. So strictly speaking, this section is between the Level 2.5 of the previous section and the Level 3 concerning the limit $r \to +\infty$.

A. Large deviations properties of the empirical r-point histograms of arbitrary order r

In the two previous sections, we have described in detail the large deviations properties of the empirical 1-point histogram $Q_.$ and 2-point histogram $Q_{..}$. Via iteration, one may analyze similarly the properties of the empirical r-point histogram $Q_{(rpoints)}$ of Eq. 16 of arbitrary order r. Since the empirical $(r-1)$-point histogram $Q_{((r-1)points)}$ can be reconstructed by summing over the last or the first value of the r-point histogram $Q_{(rpoints)}$, it is convenient to introduce the following notation analogous to Eq 69 to summarize them

$$C_r[Q_{(rpoints)}, Q_{((r-1)points)}] = \prod_{v_{r-1}} ... \prod_{v_1} \delta\left(Q_{v_{r-1}...v_1} - \sum_{v_r} Q_{v_r...v_1}\right) \delta\left(Q_{v_{r-1}...v_1} - \sum_{v_r} Q_{v_{r-1}...v_1 v_r}\right) \tag{88}$$

The final result is that the probability $P_L(Q_{(rpoints)}, ..., Q_{..}, Q_.)$ to observe the empirical histograms up to the r-point histogram $Q_{(rpoints)}$ normalized to unity

$$1 = \sum_{Q_{(rpoints)}} ... \sum_{Q_{..}} \sum_{Q_.} P_L(Q_{(rpoints)}, ..., Q_{..}, Q_.) \tag{89}$$

follows the large deviation form

$$P_L(Q_{(rpoints)}, ..., Q_.) \underset{L\to+\infty}{\simeq} \left(C_1[Q_.] e^{L \sum_v Q_v \ln(p_v)}\right) C_2[Q_{..}, Q_.] C_3[Q_{...}, Q_{..}] ... C_{r-1}[Q_{((r-1)points)}, Q_{((r-2)points)}]$$
$$C_r[Q_{(rpoints)}, Q_{((r-1)points)}] e^{L\left(S_r[Q_{(rpoints)}] - S_{r-1}[Q_{((r-1)points)}]\right)} \tag{90}$$

that generalizes Eq. 85. Besides the consistency constraints $(C_1, .., C_r)$ up to order r (Eq 88) and besides the disorder configuration weight $e^{L\sum_v Q_v \ln(p_v)}$ of Eq. 59 that only involves the empirical 1-point histogram Q_v, the remaining factor corresponds to the exponential growth of the number of configurations that have some empirical r-point histogram $Q_{(rpoints)}$

$$\Omega_L[Q_{(rpoints)}; Q_{((r-1)points)}] \underset{L\to+\infty}{\simeq} e^{L\left(S_r[Q_{(rpoints)}] - S_{r-1}[Q_{((r-1)points)}]\right)} \tag{91}$$

in terms of the entropy of the empirical r-point histogram $Q_{(rpoints)}$

$$S_r[Q_{(rpoints)}] \equiv -\sum_{v_r} ... \sum_{v_2} \sum_{v_1} Q_{v_r...v_2 v_1} \ln\left(Q_{v_r...v_2 v_1}\right) \tag{92}$$

Equivalently, Eq. 90 means that the conditional probability to observe the empirical r-point histogram $Q_{(rpoints)}$ once the empirical $(r-1)$-point histogram $Q_{((r-1)points)}$ is given reads

$$P_L[Q_{(rpoints)} | Q_{((r-1)points)}] \equiv \frac{P_L(Q_{(rpoints)}, ..., Q_{..}, Q_.)}{P_L(Q_{((r-1)points)}, ..., Q_{..}, Q_.)}$$
$$\underset{L\to+\infty}{\simeq} C_r[Q_{(rpoints)}, Q_{((r-1)points)}] e^{L\left(S_r[Q_{(rpoints)}] - 2S_{r-1}[Q_{((r-1)points)}] + S_{r-2}[Q_{((r-2)points)}]\right)} \tag{93}$$

which is the generalization of Eq. 86.

B. Analysis of the hierarchy in the backward direction via contraction

Up to now we have described the hierarchy of empirical histograms by considering successively higher and higher order r. But it is also useful to see now how one goes backwards in this hierarchy, via the notion of 'contraction' which is the generic name in the field of large deviations for the operation needed to go from a higher to a lower level of description. In our present case, the contraction consists in finding the optimal empirical r-point histogram that maximizes the conditional probability Eq 93 when all the lower-order empirical histograms are given. One needs to maximize the exponential factor in Eq 93 in the presence of the constraints $C_r[Q_{(rpoints)}, Q_{(r-1)points}]$ of Eq. 88 that can be taken into account via Lagrange multipliers. So one considers the following functional of $Q_{(rpoints)}$

$$\mathcal{L}[Q_{(rpoints)}] = -\sum_{v_r} ... \sum_{v_2} \sum_{v_1} Q_{v_r...v_2v_1} \ln\left(Q_{v_r...v_2v_1}\right) - 2S_{r-1}[Q_{((r-1)points)}] + S_{r-2}[Q_{((r-2)points)}]$$
$$+ \sum_{v_{r-1}} ... \sum_{v_2} \sum_{v_1} g_{v_{r-1}...v_1} \left(Q_{v_{r-1}...v_1} - \sum_{v_r} Q_{v_r...v_1}\right) + \sum_{v_r} ... \sum_{v_3} \sum_{v_2} f_{v_r...v_2} \left(Q_{v_r...v_2} - \sum_{v_1} Q_{v_r...v_2v_1}\right) \quad (94)$$

The optimization with respect to $Q_{v_r...v_2v_1}$

$$0 = \frac{\partial \mathcal{L}[Q_{(rpoints)}]}{\partial Q_{v_r...v_2v_1}} = -\ln\left(Q_{v_r...v_2v_1}\right) - 1 - g_{v_{r-1}...v_1} - f_{v_r...v_2} \quad (95)$$

yields the optimal solution

$$Q^*_{v_r...v_2v_1} = e^{-1-f_{v_r...v_2}-g_{v_{r-1}...v_1}} \quad (96)$$

where the Lagrange multipliers $f_{v_r...v_2}$ and $g_{v_{r-1}...v_1}$ have to be chosen to satisfy the constraints

$$Q_{v_{r-1}...v_1} = \sum_{v_r} Q^*_{v_r...v_1} = e^{-1}\left(\sum_{v_r} e^{-f_{v_r...v_2}}\right) e^{-g_{v_{r-1}...v_1}}$$
$$Q_{v_r...v_2} = \sum_{v_1} Q^*_{v_r...v_2v_1} = e^{-1} e^{-f_{v_r...v_2}} \left(\sum_{v_1} e^{-g_{v_{r-1}...v_1}}\right) \quad (97)$$

A further consequence is thus the following constraint involving the empirical histogram of order $(k-2)$

$$Q_{v_{r-1}...v_2} = \sum_{v_r} \sum_{v_1} Q^*_{v_r...v_1} = e^{-1}\left(\sum_{v_r} e^{-f_{v_r...v_2}}\right)\left(\sum_{v_1} e^{-g_{v_{r-1}...v_1}}\right) \quad (98)$$

These four last equations yield that the optimal solution of Eq. 96 can be simply rewritten as the product of the two empirical observables of order $(k-1)$ of Eq 97 divided by the empirical observable of order $(k-2)$ of Eq 98

$$Q^*_{v_r...v_2v_1} = \frac{Q_{v_r...v_2} Q_{v_{r-1}...v_1}}{Q_{v_{r-1}...v_2}} \quad (99)$$

One then needs to evaluate the entropy of Eq. 92 of this optimal solution $Q^*_{v_r...v_2v_1}$

$$S_r[Q^*_{(rpoints)}] = -\sum_{v_r} ... \sum_{v_2} \sum_{v_1} Q^*_{v_r...v_2v_1} \ln\left(Q^*_{v_r...v_2v_1}\right)$$
$$= -\sum_{v_r} ... \sum_{v_2} \sum_{v_1} \frac{Q_{v_r...v_2} Q_{v_{r-1}...v_1}}{Q_{v_{r-1}...v_2}} \left[\ln(Q_{v_r...v_2}) + \ln(Q_{v_{r-1}...v_1}) - \ln(Q_{v_{r-1}...v_2})\right]$$
$$= -\sum_{v_r} ... \sum_{v_2} Q_{v_r...v_2} \ln(Q_{v_r...v_2}) - \sum_{v_{r-1}} ... \sum_{v_2} \sum_{v_1} Q_{v_{r-1}...v_1} \ln(Q_{v_{r-1}...v_1}) + \sum_{v_{r-1}} ... \sum_{v_2} Q_{v_{r-1}...v_2} \ln(Q_{v_{r-1}...v_2})$$
$$= 2S_{r-1)}[Q_{((r-1)points)}] - S_{r-2)}[Q_{((r-2)points)}] \quad (100)$$

So the functional of Eq. 94 vanishes for this optimal solution $Q^*_{(rpoints)}$

$$\mathcal{L}[Q^*_{(rpoints)}] = 0 \quad (101)$$

i.e. the conditional probability of Eq. 93 does not decay exponentially in L for this optimal solution $Q^*_{(rpoints)}$, that represents the typical value of $Q_{v_r...v_2v_1}$ once all the empirical histograms of lower order are given

$$Q^{typ}_{v_r...v_2v_1} = Q^*_{v_r...v_2v_1} = \frac{Q_{v_r...v_2}Q_{v_{r-1}...v_1}}{Q_{v_{r-1}...v_2}} \tag{102}$$

The probability of all other values is described by the large deviation form of Eq. 93.

VII. RANDOM MODELS ON THE CAYLEY TREE FROM LARGE DEVIATIONS OF BRANCHES

Many random models have been studied on the geometry of the Cayley tree, where the absence of loops allows to write exact recurrences on probability distributions : two famous examples are the Directed Polymer on the Cayley tree [72, 73] and the Anderson Localization on the Cayley tree [74–77]. In the Cayley tree of branching ratio K around the central root O, the number of sites at distance r

$$N^{tree}(r) = (K+1)K^{r-1} = \frac{K+1}{K}K^r \tag{103}$$

grows exponentially with the distance r, in contrast to the power-law growth as r^{d-1} in any finite dimension d. The Cayley tree is thus considered as an appropriate way to define the mean-field version of random models in infinite dimensionality $d = \infty$.

It is interesting to compare the properties of the same random model defined in the two following geometries :

(i) in the finite Cayley tree of branching ratio K with L generations around the central root O, where the number of leaves is given by Eq. 103 for $r = L$

$$N^{tree}(L) = \frac{K+1}{K}K^L \tag{104}$$

(ii) in the star geometry, where the central root O is linked to K^L independent one-dimensional lattices of L sites, so that the number of sites at distance r is actually independent of r

$$N^{star}_{K^L}(r) = K^L \tag{105}$$

but the number of leaves at $r = L$ displays the same exponential behavior in L as Eq. 104.

Although (ii) may look as an extremely crude approximation of (i), the properties of some random models defined on (i) and (ii) have turned out to be very close, as exemplified by the exact solutions of (i) the Directed Polymer on the Cayley tree [72, 73] and of (ii) the Directed Polymer in the star geometry that coincides with the Random Energy Model [78] (a model that had been introduced before with completely different motivations coming from mean-field spin-glasses). The differences between the two only appear in the finite-size scaling properties of the freezing transition [73].

In the star geometry (ii), it is clear that the random model will be governed by the large deviations properties of the corresponding one-dimensional model of length L that appear on the K^L independent branches. In this section, the goal is thus to describe how the large deviations properties of one-dimensional models that have been discussed in the previous sections can be used to analyze the properties of the same model on this star geometry (ii).

A. Model on the star geometry where each branch corresponds to a product of random variables

We wish the analyze the star geometry (ii) above, where each of the independent K^L branches labelled by $b = 1, 2, .., K^L$ can be described by a product of L random variables as Eq. 11

$$\tau_L[v_b(.)] = \prod_{x=1}^{L} t_{v_b(x)} \tag{106}$$

with its corresponding finite-size Lyapunov exponent of Eq. 23

$$\lambda_b \equiv \lambda[v_b(.)] \equiv \frac{\ln|\tau_L[v(.)]|}{L} = \frac{1}{L}\sum_{x=1}^{L} \ln|t_{v_b(x)}| \tag{107}$$

whose large deviations properties for large L are described by some rate function $I(\lambda)$

$$\mathcal{P}_L(\lambda) \underset{L\to+\infty}{\simeq} e^{-LI(\lambda)} \tag{108}$$

Each disordered configuration on the star geometry can be then characterized by the empirical histogram of the Lyapunov exponent λ_b of Eq. 107 for the K^L independent branches

$$\mathcal{Q}_L(\lambda) \equiv \frac{1}{K^L}\sum_{b=1}^{K^L}\delta(\lambda-\lambda_b) \tag{109}$$

while the empirical number of branches having the Lyapunov exponent λ reads

$$\mathcal{N}_L(\lambda) \equiv \sum_{b=1}^{K^L}\delta(\lambda-\lambda_b) = K^L\mathcal{Q}_L(\lambda) \tag{110}$$

In various models, an interesting class of observables are given by the sums over the K^L independent branches of the powers of non-integer k of the products $\tau_L[v_b(.)]$ of Eq. 106

$$\mathcal{S}_L(k) \equiv \sum_{b=1}^{K^L}|\tau_L[v_b(.)]|^k \tag{111}$$

that can be rewritten in terms of the Lyapunov exponents λ_b (Eq 107) of the K^L branches or in terms of the empirical observables of Eqs 109 and 110 as

$$\mathcal{S}_L(k) = \sum_{b=1}^{K^L} e^{kL\lambda_b} = \int d\lambda e^{kL\lambda}\mathcal{N}_L(\lambda) = K^L\int d\lambda e^{kL\lambda}\mathcal{Q}_L(\lambda) \tag{112}$$

B. Statistical properties of the empirical histogram $\mathcal{Q}_L(\lambda)$ of the Lyapunov exponent

The typical value of the empirical histogram of Eq. 109 is given by the true probability of the Lyapunov exponent of Eq. 108

$$\mathcal{Q}_L^{typ}(\lambda) = \mathcal{P}_L(\lambda) \underset{L\to+\infty}{\simeq} e^{-LI(\lambda)} \tag{113}$$

so that in a given sample, the empirical number of branches of Eq. 110 has for typical value

$$\mathcal{N}_L^{typ}(\lambda) = K^L\mathcal{Q}_L(\lambda) \underset{L\to+\infty}{\simeq} e^{L[\ln K - I(\lambda)]} \tag{114}$$

The typical value λ^{typ} of the one-dimensional model corresponding to the vanishing of the rate function $I(\lambda^{typ}) = 0$ will thus appear in an extensive number of the branches

$$\mathcal{N}_L^{typ}(\lambda^{typ}) \underset{L\to+\infty}{\simeq} e^{L\ln K} = K^L \tag{115}$$

while all the other values in the interval $\lambda^- < \lambda < \lambda^+$ where

$$I(\lambda) < \ln K = I(\lambda^-) = I(\lambda^+) \tag{116}$$

will appear in a sub-extensive number $e^{L[\ln K - I(\lambda)]}$ of branches. Finally, the values of the Lyapunov exponent outside this interval, i.e. in the two regions $\lambda < \lambda^-$ and $\lambda > \lambda^+$ where the rate function satisfies $I(\lambda) > \ln K$ are too rare to appear in a typical sample of the star geometry, so that Eq. 114 should be rewritten more precisely for a typical sample as

$$\mathcal{N}_L^{TypicalSample}(\lambda) \underset{L\to+\infty}{\simeq} e^{L[\ln K - I(\lambda)]}\theta(I(\lambda) \leq \ln K) = e^{L[\ln K - I(\lambda)]}\theta(\lambda^- \leq \lambda \leq \lambda^+) \tag{117}$$

However the values $\lambda < \lambda^-$ and $\lambda > \lambda^+$ that do not appear in a typical sample may appear in atypical samples, and it is thus interesting to consider the large deviations of the empirical histogram $\mathcal{Q}_L(.)$ of Eq. 109 with respect to its typical value $\mathcal{Q}_L^{typ}(.) = \mathcal{P}_L(.)$ of Eq. 113 : since the K^L branches are independent, one may directly adapt the Sanov result of Eq. 41 to our present notations : the probability to observe the empirical histogram $\mathcal{Q}_L(.)$ follows the large deviation form with respect to the size K^L

$$P_L[\mathcal{Q}_L(.)] \underset{L\to+\infty}{\simeq} \delta\left(1-\int d\lambda \mathcal{Q}_L(\lambda)\right) e^{-K^L S^{rel}(\mathcal{Q}_L(.)|\mathcal{P}_L(.))} \tag{118}$$

where the rate function corresponds to the relative entropy

$$S^{rel}(\mathcal{Q}_L(.)|\mathcal{P}_L(.)) = \int d\lambda \mathcal{Q}_L(\lambda) \ln\left(\frac{\mathcal{Q}_L(\lambda)}{\mathcal{P}_L(\lambda)}\right) \tag{119}$$

of the empirical histogram $\mathcal{Q}_L(.)$ with respect to the true probability distribution $\mathcal{P}_L(.)$ of the Lyapunov exponent (Eq. 108). As explained in detail in section IV B, the Sanov result of Eq. 118 is equivalent to the following expression of the generating function that is valid for any finite L (Eq. 45 as adapted to our present context)

$$\begin{aligned}\mathcal{G}_L[\nu(.)] &\equiv \sum_{\mathcal{Q}_L(.)} P_L[\mathcal{Q}_L(.)]\ e^{K^L \int d\lambda \nu(\lambda)\mathcal{Q}_L(\lambda)} = \int d\lambda_1 ... \int d\lambda_{K^L} \mathcal{P}_L(\lambda_1)...\mathcal{P}_L(\lambda_{K^L})\ e^{\sum_{b=1}^{K^L}\nu(\lambda_b)} \\ &= \prod_{b=1}^{K^L}\left(\int d\lambda_b \mathcal{P}_L(\lambda_b) e^{\nu(\lambda_b)}\right) = \left(\int d\lambda \mathcal{P}_L(\lambda) e^{\nu(\lambda)}\right)^{K^L}\end{aligned} \tag{120}$$

In particular, the successive derivatives with respect to $\nu(\lambda)$

$$\begin{aligned}\frac{\partial \mathcal{G}_L[\nu(.)]}{\partial \nu(\lambda)} &= \sum_{\mathcal{Q}_L(.)} P_L[\mathcal{Q}_L(.)]\ K^L \mathcal{Q}_L(\lambda)\ e^{K^L \int d\lambda' \nu(\lambda')\mathcal{Q}_L(\lambda')} = K^L \mathcal{P}_L(\lambda) e^{\nu(\lambda)} \left(\int d\lambda' \mathcal{P}_L(\lambda') e^{\nu(\lambda')}\right)^{K^L-1} \\ \frac{\partial^2 \mathcal{G}_L[\nu(.)]}{\partial^2 \nu(\lambda)} &= \sum_{\mathcal{Q}_L(.)} P_L[\mathcal{Q}_L(.)]\ \left(K^L \mathcal{Q}_L(\lambda)\right)^2\ e^{K^L \int d\lambda' \nu(\lambda')\mathcal{Q}_L(\lambda')} = K^L \mathcal{P}_L(\lambda) e^{\nu(\lambda)} \left(\int d\lambda' \mathcal{P}_L(\lambda') e^{\nu(\lambda')}\right)^{K^L-1} \\ &+ K^L(K^L-1)\left(\mathcal{P}_L(\lambda) e^{\nu(\lambda)}\right)^2 \left(\int d\lambda' \mathcal{P}_L(\lambda') e^{\nu(\lambda')}\right)^{K^L-2}\end{aligned} \tag{121}$$

gives the integer moments of the number $\mathcal{N}_L(\lambda) = K^L \mathcal{Q}_L(\lambda)$ of branches with some Lyapunov exponent λ (Eq 110) by taking $\nu(.) = 0$. The first moment

$$\overline{\mathcal{N}_L(\lambda)} = \sum_{\mathcal{Q}_L(.)} P_L[\mathcal{Q}_L(.)]\ K^L \mathcal{Q}_L(\lambda) = \left.\frac{\partial \mathcal{G}_L[\nu(.)]}{\partial \nu(\lambda)}\right|_{\nu(.)=0} = K^L \mathcal{P}_L(\lambda) = \mathcal{N}_L^{typ}(\lambda) \tag{122}$$

coincides with the typical value $\mathcal{N}_L^{typ}(\lambda)$ of Eq. 114. The second moment

$$\overline{(\mathcal{N}_L(\lambda))^2} = \sum_{\mathcal{Q}_L(.)} P_L[\mathcal{Q}_L(.)]\ \left(K^L \mathcal{Q}_L(\lambda)\right)^2 = \left.\frac{\partial^2 \mathcal{G}_L[\nu(.)]}{\partial^2 \nu(\lambda)}\right|_{\nu(.)=0} = K^L \mathcal{P}_L(\lambda) + K^L(K^L-1)\left(\mathcal{P}_L(\lambda)\right)^2 \tag{123}$$

can be rewritten in terms of the typical value $\mathcal{N}_L^{typ}(\lambda)$ of Eq. 114 as

$$\overline{(\mathcal{N}_L(\lambda))^2} \simeq \mathcal{N}_L^{typ}(\lambda) + \left(\mathcal{N}_L^{typ}(\lambda)\right)^2 \tag{124}$$

and will thus change of behavior at the values $\lambda^\pm$ introduced in Eq. 116. In the region $\lambda^- < \lambda < \lambda^+$ where $\mathcal{N}_L^{typ}(\lambda)$ is exponentially large, the second term dominates over the first term that corresponds to a small fluctuation. In the other regions where $\mathcal{N}_L^{typ}(\lambda)$ is exponentially small, the first term dominates and actually represents the very small probability to have a single rare event

$$\begin{aligned}\overline{(\mathcal{N}_L(\lambda))^2} &\simeq \left(\mathcal{N}_L^{typ}(\lambda)\right)^2 \quad \text{for} \quad \lambda^- < \lambda < \lambda^+ \\ &\simeq \mathcal{N}_L^{typ}(\lambda) \quad \text{for} \quad \lambda < \lambda^- \quad \text{and} \quad \lambda^+ < \lambda\end{aligned} \tag{125}$$

This result can be generalized to arbitrary moments, as described in the context of the Random Energy Model [78].

C. Statistical properties of the empirical sums $\mathcal{S}_L(k)$ of Eq. 112

The disorder-averaged value of the empirical sum $\mathcal{S}_L(k)$ of Eq. 111 reads

$$\overline{\mathcal{S}_L(k)} = \sum_{b=1}^{K^L} \overline{|\tau_L[v_b(.)]|^k} = K^L \overline{|\tau_L[v(.)]|^k} \tag{126}$$

where the moments $\overline{|\tau_L[v(.)]|^k}$ of non-integer order k for the product of random variables have been already discussed in Eq. 24

$$\overline{|\tau_L[v(.)]|^k} = \int d\lambda e^{kL\lambda} \mathcal{P}_L(\lambda) = \int d\lambda e^{L[k\lambda - I(\lambda)]} = e^{L\phi(k)} \tag{127}$$

in terms of the scaled cumulant generating function $\phi(k)$

$$\phi(k) = \ln \left[\overline{|t_v|^k} \right] \tag{128}$$

that corresponds to the Legendre transform of the rate function $I(\lambda)$.

On the other hand, Eq. 112 yields that the sum $\mathcal{S}_L(k)$ in a typical sample can be computed from the empirical histogram in a typical sample (Eq. 117)

$$\mathcal{S}_L^{TypicalSample}(k) = \int d\lambda e^{kL\lambda} \mathcal{N}_L^{TypicalSample}(\lambda) = K^L \int d\lambda e^{L[k\lambda - I(\lambda)]} \theta(\lambda^- \leq \lambda \leq \lambda^+) \tag{129}$$

So the only difference with the averaged value (Eqs 126 and 127)

$$\overline{\mathcal{S}_L(k)} = K^L \int d\lambda e^{L[k\lambda - I(\lambda)]} \tag{130}$$

lies in the boundaries $\lambda^- \leq \lambda \leq \lambda^+$ for the integration over the Lyapunov exponent that appear for the value in a typical sample (Eq 129) but that are absent in the averaged value of Eq. 130. As a consequence, one needs to discuss the position of the saddle-point value λ_k that governs the integral governing the averaged value of Eq. 130

$$k = I'(\lambda_k) \tag{131}$$

with respect to the two boundaries $\lambda^\pm$ of the integral governing the typical-sample value of Eq. 129. It is thus useful to introduce the two values $k^\pm$ satisfying $\lambda_{k^\pm} = \lambda^\pm$ i.e.

$$k^\pm = I'(\lambda^\pm) \tag{132}$$

and to distinguish the three following cases :

(a) In the region $k^- < k < k^+$, the saddle-point value λ_k of Eq. 131 is in the interval

$$\lambda^- < \lambda_k < \lambda^+ \tag{133}$$

The typical-sample value $\mathcal{S}_L^{TypicalSample}(k)$ of Eq. 129 has then the same exponential behavior in L as the averaged value $\overline{\mathcal{S}_L(k)}$ involving the Legendre transform $\phi(k)$ (Eqs 127 and 25) of $I(\lambda)$

$$\mathcal{S}_L^{TypicalSample}(k) \underset{L\to+\infty}{\propto} K^L e^{L\phi(k)} = e^{L[\ln K + \phi(k)]} \tag{134}$$

(b) In the region $k > k^+$, the saddle-point value λ_k of Eq. 131 is bigger than λ^+

$$\lambda_k > \lambda^+ \tag{135}$$

The typical-sample value $\mathcal{S}_L^{TypicalSample}(k)$ of Eq. 129 is then governed by the saddle-point evaluation frozen at the boundary λ_+ satisfying $I(\lambda_+) = \ln K$ (Eq 116)

$$\mathcal{S}_L^{TypicalSample}(k) \underset{L\to+\infty}{\propto} K^L e^{L\left[k\lambda^+ - I(\lambda^+)\right]} = e^{Lk\lambda^+} \tag{136}$$

(c) In the region $k < k^-$, the saddle-point value λ_k of Eq. 131 is smaller than λ^-

$$\lambda_k < \lambda^- \tag{137}$$

The typical-sample value $\mathcal{S}_L^{TypicalSample}(k)$ of Eq. 129 is then governed by the saddle-point evaluation frozen at the boundary λ_- satisfying $I(\lambda_-) = \ln K$ (Eq 116)

$$\mathcal{S}_L^{TypicalSample}(k) \underset{L\to+\infty}{\propto} K^L e^{L\left[k\lambda^- - I(\lambda^-)\right]} = e^{Lk\lambda^-} \tag{138}$$

D. Sample-to-sample fluctuations in the frozen phase $k > k^+$

In the frozen phase $k > k^+$, the sample-dependent version of Eq. 136 is that the sum $\mathcal{S}_L(k)$ in a given sample will be actually governed by the biggest Lyapunov exponents available among the K^L branches. It is thus convenient to relabel in each sample the Lyaponov exponents according to their magnitudes

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 ... \geq \lambda_{K^L} \tag{139}$$

and to analyze the statistics of the first biggest terms in the sum of Eq. 112

$$\mathcal{S}_L(k) = \sum_{b=1}^{K^L} e^{kL\lambda_b} = e^{kL\lambda_1} + e^{kL\lambda_2} + e^{kL\lambda_3} + ... \tag{140}$$

and in particular the first one that involves the maximal Lyapunov exponent λ_1

$$\mathcal{S}_L^{first}(k) \equiv e^{Lk\lambda_1} \tag{141}$$

1. *Probability distribution of the maximal Lyapunov exponent λ_1 in each sample*

The maximal Lyapunov exponent λ_1 is typically of order λ^+, but here we wish to analyze its probability distribution $R(\lambda_1)$ over the samples. The corresponding cumulative distribution reads in terms of $\mathcal{P}_L(\lambda)$ of Eq. 108

$$\int^{\lambda} d\lambda_1 R(\lambda_1) \quad = \left[1 - \int_{\lambda}^{+\infty} d\lambda' \mathcal{P}_L(\lambda')\right]^{K^L} \underset{L\to+\infty}{\simeq} e^{-K^L \int_{\lambda}^{+\infty} d\lambda' e^{-LI(\lambda')}} \underset{L\to+\infty}{\simeq} e^{-e^{L(\ln K - I(\lambda))}} \tag{142}$$

The change of variables

$$\lambda = \lambda^+ + \frac{u}{Lk^+} \tag{143}$$

centered around the value λ^+ where $I(\lambda^+) = \ln K$ and $I(\lambda^+) = k^+$ (Eq. 132) leads to the Taylor expansion of the rate function

$$I(\lambda) = I(\lambda^+) + \frac{u}{Lk^+} I'(\lambda^+) + O\left(\frac{1}{L^2}\right) = \ln K + \frac{u}{L} + O\left(\frac{1}{L^2}\right) \tag{144}$$

Plugging this expansion into Eq 142

$$\int^{\lambda^+ + \frac{u}{Lk^+}} d\lambda_1 R(\lambda_1) \underset{L\to+\infty}{\simeq} e^{-e^{-u}} \equiv \int^{u} du' G(u') \tag{145}$$

yields the convergence towards the Gumbel distribution (well-known as one of the three universality classes for the extreme-value statistics of independent random variables [82, 83])

$$G(u) = e^{-u} e^{-e^{-u}} \tag{146}$$

for the $O(1)$ random variable u introduced in Eq. 143.

2. *Probability distribution of $\mathcal{S}_L^{first}(k) = e^{Lk\lambda_1}$ over the samples*

Eq 141 yields that its logarithm reads with the change of variables of Eq. 143

$$\ln \mathcal{S}_L^{first}(k) = Lk\lambda_1 = Lk\lambda^+ + \frac{k}{k^+} u \tag{147}$$

where u is distributed with the Gumbel distribution of Eq. 146. This means that the probability distribution of $\left(\ln \mathcal{S}_L^{first}(k)\right)$ propagates as a traveling wave as L grows : the first term $Lk\lambda^+$ corresponds to a motion with the

non-random velocity ($k\lambda^+$) with respect to L, while the second term $\frac{k}{k^+}u$ is random and independent of L, i.e. its probability distribution corresponds to the fixed shape of the traveling wave. This notion of traveling wave has been stressed here because it plays a major role in the analysis of random models defined on Cayley trees, as first discovered with the exact solution of the Directed Polymer on the Cayley tree [72].

Eq. 147 translates into

$$\mathcal{S}_L^{first}(k) = e^{Lk\lambda^+} e^{\frac{k}{k^+}u} \equiv \left(\mathcal{S}_L^{TypicalSample}(k)\right) X \tag{148}$$

where $\mathcal{S}_L^{TypicalSample}(k) = e^{Lk\lambda^+}$ is the value in a typical sample introduced in 136, while

$$X \equiv e^{\frac{k}{k^+}u} \tag{149}$$

is an $O(1)$ positive random variable, whose distribution reads in terms of the Gumbel distribution $G(u)$ of Eq. 146

$$L_k(X) = \frac{k^+}{kX} G\left(\frac{k^+}{k}\ln X\right) = \frac{\mu_k}{X^{1+\mu_k}} e^{-\frac{1}{X^{\mu_k}}} \tag{150}$$

where the exponent

$$\mu_k \equiv \frac{k^+}{k} \tag{151}$$

governs the power-law decay of Eq. 150 for large X

$$L_k(X) \underset{X\to+\infty}{\simeq} \frac{\mu_k}{X^{1+\mu_k}} \tag{152}$$

The exponent μ_k decays continuously in the frozen phase $k \geq k^+$ from the value $\mu_{k=k^+} = 1$ to vanishing values $\mu_{(k\to+\infty)} \to 0$. Since it remains smaller than one in the whole frozen phase $k \geq k^+$

$$\mu_{(k\geq k^+)} = \frac{k^+}{k} \leq 1 \tag{153}$$

the averaged value of the variable X is infinite

$$\int dX X L_k(X) = +\infty \tag{154}$$

i.e. the averaged value $\overline{\mathcal{S}_L^{first}(k)}$ in Eq. 148 has a different exponential behavior in L than the typical value $\mathcal{S}_L^{TypicalSample}(k)$, in consistency with the discussion around Eq. 130.

E. Application to the Directed Polymer and the Random Energy Model

With respect to the generic notations of section VII A, the Random Energy Model [78] corresponds to $K = 2$ and to the case where λ is an energy distributed with a Gaussian distribution of Eq. 29 so that the rate function $I(\lambda)$ and the scaled cumulant generating function $\phi(k)$ are quadratic

$$\begin{aligned} I^{Gauss}(\lambda) &= \frac{\lambda^2}{2\sigma^2} \\ \phi^{Gauss}(k) &= \frac{k^2\sigma^2}{2} \end{aligned} \tag{155}$$

The empirical number $\mathcal{N}_L(\lambda)$ of Eq. 110 corresponds to the number of accessible states in the microcanonical ensemble where the energy density λ is fixed, and its value $\mathcal{N}_L(\lambda)$ in a typical sample (Eq. 117) yields that the function in the exponential corresponds to the entropy as a function of the energy density λ in the microcanonical ensemble [78]

$$S^{Microcanonical}(\lambda) = \left[\ln K - I^{Gauss}(\lambda)\right] \theta(\ln K - I^{Gauss}(\lambda) \geq 0) = \left[\ln 2 - \frac{\lambda^2}{2\sigma^2}\right] \theta(\lambda^- \leq \lambda \leq \lambda^+) \tag{156}$$

with the boundaries

$$\lambda^{\pm} = \pm\sigma\sqrt{2\ln 2} \tag{157}$$

With the change of notation $k \to \beta$, the empirical sum $\mathcal{S}_L(k)$ of Eq. 111 and 112 corresponds to the partition function $Z_L(\beta)$ in the canonical ensemble at inverse temperature β

$$Z_L(\beta) = \sum_{b=1}^{K^L} e^{\beta L \lambda_b} = \int d\lambda e^{kL\lambda} \mathcal{N}_L(\lambda) \tag{158}$$

with its disordered-averaged value (Eqs 126 and 127)

$$\overline{Z_L(\beta)} = K^L e^{L\phi(\beta)} = e^{L\left[\ln 2 + \frac{\beta^2\sigma^2}{2}\right]} \tag{159}$$

while its value in a typical sample (Eq 129) involves the microcanonical entropy of Eq. 156

$$Z_L^{TypicalSample}(\beta) = K^L \int d\lambda e^{L\left[\ln 2 - \frac{\lambda^2}{2\sigma^2}\right]} \theta(\lambda^- \leq \lambda \leq \lambda^+) \tag{160}$$

Since the inverse temperature β is positive $\beta > 0$ (instead of k of arbitrary sign above), the critical temperature β_c of the freezing transition corresponds to the solution k^+ of Eq 132

$$\beta_c = I'(\lambda^+) = \frac{\lambda^+}{\sigma^2} = \frac{\sqrt{2\ln 2}}{\sigma} \tag{161}$$

The two phases are [78]

(a) the high-temperature phase $\beta < \beta_c$ where the partition function in a typical sample (Eq 129) coincides with the averaged value of Eq. 159.

(b) the low-temperature frozen phase $\beta < \beta_c$ where the partition function in a typical sample is different from the averaged value of Eq. 159 because it is governed by the boundary λ_+ (Eq. 136)

$$Z_L^{TypicalSample}(\beta) \underset{L\to+\infty}{\propto} e^{L\beta\lambda^+} = e^{L\beta\sigma\sqrt{2\ln 2}} \tag{162}$$

In this frozen phase, the exponent of Eq. 151

$$\mu_\beta \equiv \frac{\beta_c}{\beta} = \frac{T}{T_c} \tag{163}$$

of the heavy-tail distribution of Eq. 152 allows to analyze further the statistics of overlaps in terms of the weights of individual terms within in a Lévy sum of random variables distributed with heavy tails [79–81].

F. Application to Anderson Localization

The notations for the Anderson Localization model have been explained in the subsection III D with the rate function $I(\lambda)$ and the scaled cumulant generating function $\phi(k)$ given by Eqs 37 and 38

$$\begin{aligned} I^{Anderson}(\lambda) &= \lambda - \ln\left(\frac{V}{W}\right) - 1 - \ln\left[\lambda - \ln\left(\frac{V}{W}\right)\right] \\ \phi^{Anderson}(k) &= k\ln\left(\frac{V}{W}\right) - \ln(1-k) \end{aligned} \tag{164}$$

Here the analysis concerns the localized phase in the regime of small hopping V where the forward perturbation formula of Eq. 13 is valid, so it will be possible to use this approach up to the critical hopping V_c of the delocalization transition only if the branching ratio K is large $K \gg 1$.

The empirical number of Eq 110 counts the number of leaves (among the K^L branches) where the wave-function $|\psi_b(L)|$ is of order $e^{L\lambda}$ with respect to the finite wave-function at the center. The empirical number in a typical sample (Eq. 117) reads

$$\mathcal{N}_L^{TypicalSample}(\lambda) \underset{L\to+\infty}{\simeq} e^{L\left[\ln K - I^{Anderson}(\lambda)\right]} \theta(\lambda^- \leq \lambda \leq \lambda^+) = e^{L\left[\ln\left(\frac{KVe}{W}\right) - \lambda + \ln\left[\lambda - \ln\left(\frac{V}{W}\right)\right]\right]} \theta(\lambda^- \leq \lambda \leq \lambda^+) \tag{165}$$

where the boundaries $\lambda^{\pm}$ are given by Eq. 116

$$0 = \ln K - I^{Anderson}(\lambda^{\pm}) = \ln\left(\frac{KVe}{W}\right) - \lambda^{\pm} + \ln\left[\lambda^{\pm} - \ln\left(\frac{V}{W}\right)\right] \tag{166}$$

For large $K \gg 1$, the upper boundary is given by

$$\lambda^{+} \simeq \ln\left(\frac{KVe}{W}\right) + \ln\left[\ln\left(\frac{KVe}{W}\right) - \ln\left(\frac{V}{W}\right)\right] = \ln\left[\frac{KVe}{W}\ln(Ke)\right] \tag{167}$$

The localized phase correspond to the region $\lambda_{+} < 0$, where the wave-function decays exponentially on all the K^L branches, while the delocalization transition occurs when λ^{+} vanishes

$$(\lambda^{+})_{criti} = 0 \tag{168}$$

so the critical hopping V_c for the delocalization transition is given for large $K \gg 1$ by

$$V_c \simeq \frac{W}{Ke\ln(Ke)} \tag{169}$$

At this critical point $V = V_c$, the inverse participation ratios

$$Y_L(q) \equiv \sum_{b=1}^{K^L} |\psi_b(L)|^{2q} \tag{170}$$

correspond to the empirical sums of Eq. 111 with the change of notation $k = 2q$, so that their disordered-averaged values (Eqs 126 and 127) read for $q < \frac{1}{2}$

$$\overline{Y_L(q)} = K^L e^{L\phi^{Anderson}(2q)} = e^{L\left[\ln K + 2q\ln\left(\frac{V_c}{W}\right) - \ln(1-2q)\right]} = e^{L[(1-2q)\ln K - 2q\ln(e\ln(Ke)) - \ln(1-2q)]} \underset{L\to+\infty}{\simeq} (K^L)^{-\tau_q^{av}} \tag{171}$$

where the exponents τ_q^{av} defined with respect to the number K^L of sites read for large $K \gg 1$

$$\tau_q^{av} \simeq (2q-1)\theta\left(q < \frac{1}{2}\right) \tag{172}$$

Eq 132 yields that the boundary value $k^{+} = 2q^{+}$ using Eq. 167 and Eq. 169

$$2q^{+} = I'(\lambda^{+}) = 1 + \frac{1}{\lambda^{+} - \ln\left(\frac{V_c}{W}\right)} = 1 - \frac{1}{\ln\left[Ke\ln(Ke)\right]} \tag{173}$$

is close to unity for large K, so that the inverse participation ratios in a typical sample

$$Y_L^{TypicalSample}(q) \underset{L\to+\infty}{\simeq} (K^L)^{-\tau_q^{typ}} \tag{174}$$

involve essentially the same exponents as the averaged values of Eq. 171

$$\tau_q^{typ} \simeq \tau_q^{av} \simeq (2q-1)\theta\left(q < \frac{1}{2}\right) \tag{175}$$

These exponents are known as the 'Strong Multifractality spectrum' in the field of Anderson transitions [30], where they appears either in the limit of infinite dimensionality $d \to +\infty$ or in related long-ranged power-law hoppings in one-dimension [84–98], or more recently in toy models of Many-Body-Localization [99, 100]. Although the freezing transitions at the values $q^{\pm}$ is not very important for this 'Strong Multifractality spectrum', they have been much discussed in the general theory of multifractality at Anderson transitions in finite dimension d [29, 30, 88]

G. Application to the Quantum Ising Model

As a third and final example, let us mention the case of the random transverse field spin-glass model on the Cayley tree that has been studied recently via real-space renormalization and where the large deviations properties of the one-dimensional model play a major role [101]. Here the difference with the two previous examples of the Random Energy Model and of Anderson Localization is that the one-dimensional model has already its phase transition between the spin-glass phase and the paramagnetic phase, where the exact critical properties have been obtained by the Strong Disorder renormalization approach [27, 28]. As a consequence, one obtains three phases that can be explained as follows in the star-geometry (ii) of Eq. 105, where one considers that the center O is linked to K^L independent chains of length L [101], i.e. each branch $b = 1, .., K^L$ is characterized by the Lyapunov exponent (Eq 14)

$$\lambda_b \equiv \frac{1}{L} \sum_{x=1}^{L} \ln \left| \frac{J_b(x)}{h_b(x)} \right| \tag{176}$$

(a) the star is in its paramagnetic phase if all the K^L chains are in their paramagnetic state $\lambda_b < 0$, i.e. the boundary value λ_+ of Eq. 116 should be negative $\lambda_+ < 0$.

(b) the star is in its spin-glass phase with an extensive spin-glass order if an extensive number of the K^L chains are in their spin-glass phase $\lambda_b > 0$ i.e. the typical Lyapunov exponent should be positive $\lambda_{typ} > 0$.

(c) in between, i.e. in the region $\lambda_{typ} < 0 < \lambda_+$, the star is a spin-glass phase with an sub-extensive spin-glass order, because only the subextensive number of chains are in their spin-glass phase $\lambda_b > 0$, while an extensive number of chains are in their paramagnetic state $\lambda_b < 0$.

VIII. CONCLUSION

In this pedagogical introduction, we have explained why the general theory of large deviations is the natural language to analyze the properties of disordered systems in order to offer a unified perspective on the typical events and on the rare events that occur on various scales. We have first focused on one-dimensional random models in order to emphasize the various levels of description. We have first recalled how the Level 1 allows to analyze the properties of observables given by products of random variables that occur in many classical or quantum models. We have then described how a finer analysis in terms of the whole hierarchy of empirical histograms allows to classify the set of disorder configurations into subsets that have the same empirical properties up to a certain order. We have then turned our attention to random models defined on Cayley trees, in order to analyze their properties in terms of the large deviations of branches. We have taken as examples various emblematic classical and quantum disordered systems in order to highlight the common underlying mechanisms from the point of view of large deviations.

The large deviation analysis of disordered systems in finite dimension $1 < d < +\infty$ clearly goes beyond the scope of the present introduction. Although some notions can be directly applied, like the Sanov theorem for the empirical 1-point histogram, or the multifractal analysis at Anderson transitions [29, 30] or at phase transitions of random classical models [31–38], one should be aware that qualitatively new phenomena may also occur. For instance the large deviations properties that have been exactly computed [102–104] for the Directed Polymer in dimension $d = 2$ display an asymmetry between values bigger or smaller than the typical value, with two different scalings with respect to the length L of the polymer : an 'anomalously good' ground state energy requires only L anomalously good on-site energies along the polymer, while an 'anomalously bad' ground state energy requires L^2 bad on-site energies in the two-dimensional sample. So this single example already shows that some properties of random systems in finite dimensions d call for a much broader large deviation theory with two different scalings for values bigger or smaller than the typical value, as discussed in more details in the recent preprint [105].

Appendix A: Alternative classification of disorder configurations in terms of empirical intervals

In the text, we have described the classification of one-dimensional disorder configurations in terms of the hierarchy of the empirical r-point histograms. In this Appendix, we discuss an alternative classification in terms of the empirical intervals during which the disorder keeps a constant value, since this framework is more appropriate to analyze the Lifshitz and the Griffiths singularities as we now recall.

1. Observables corresponding to products of contributions from intervals of random lengths

After the product of random variables discussed in section II C, the next simpler case of Eq. 3 concerns the case where the disorder variable can take only two values that will be labelled by $v = \pm$. It is then useful to replace the disorder configuration $[v(x)]_{x=1,2,..,L}$ by its decomposition into intervals during which the disorder keeps the same value. For a model defined on a ring of L sites (i.e. with periodic boundary conditions $L + x = x$) there will be an empirical even number $(2N)$ of intervals, where the N odd intervals $(2i-1)$ of lengths l_{2i-1} are associated to the value $v = -$, while the N even intervals $(2i)$ of lengths l_{2i} are associated to the value $v = +$. The lengths l_i satisfy the sum rule

$$L = \sum_{i=1}^{2N} l_i \tag{A1}$$

When the disorder configuration $[v(x)]_{x=1,2,..,L}$ is replaced by the list $[l_i]_{i=1,2,..,2N}$ of the lengths of the intervals, the trace of Eq. 3 becomes

$$\Theta_L\left([l_i]_{i=1,2,..,2N}\right) \equiv Tr\left[T_+^{l_{2N}} T_-^{l_{2N-1}} ... T_-^{l_3} T_+^{l_2} T_-^{l_1}\right] \tag{A2}$$

To analyze the Lifshitz and the Griffiths singularities mentioned in the Introduction, various models have been studied in the regime where the value $v = -$ corresponds to a very strong disorder value where the associated transfer matrix T_- can be approximated by a projector on some state $|0>$ with some eigenvalue t_- [16]

$$T_- \simeq t_- |0><0| \tag{A3}$$

Then Eq. A2 simplifies into the product of the contributions of the intervals

$$\Theta_L\left([l_i]_{i=1,2,..,2N}\right) = \prod_{i=1}^{N} \left(\theta_+(l_{2i})\ \theta_-(l_{2i-1})\right) \tag{A4}$$

where the contribution of an interval $v = -$ of length l is simply

$$\theta_-(l) \equiv (t_-)^l \tag{A5}$$

while the contribution of an interval $v = +$ of length l corresponds to the the pure model $v = +$ with the boundary conditions $|0>$ fixed by the projector form of Eq. A3

$$\theta_+(l) \equiv <0|T_+^l|0> \tag{A6}$$

Various examples concerning Anderson Localization models and classical spin chains are described in the book [16], while an example concerning random DNA is analyzed in [23].

2. Empirical 1-interval observables with their constraints

The observables of the form of Eq. A4 suggests that it is appropriate to analyze the disorder configurations in terms the empirical 1-interval observables

$$\begin{aligned} n_+(l) &\equiv \frac{1}{L}\sum_{i=1}^{N} \delta_{l_{2i},l} \\ n_-(l) &\equiv \frac{1}{L}\sum_{i=1}^{N} \delta_{l_{2i-1},l} \end{aligned} \tag{A7}$$

The summation over the length l corresponds to the density $\frac{N}{L}$ of intervals $v = +$ or $v = -$

$$\sum_l n_+(l) = \frac{N}{L} = \sum_l n_-(l) \tag{A8}$$

while the total length L of the disorder configurations fixes the normalization (Eq. A1)

$$1 = \sum_{l=1}^{+\infty} l \left[n_+ (l) + n_- (l) \right] \tag{A9}$$

It is thus useful to introduce the following notation to summarize these constraints on the empirical 1-interval observables $n^{\pm}(.)$

$$c_1[n_+(.); n_-(.)] \equiv \quad \delta \left(\sum_l n_+ (l) - \sum_l n_- (l) \right) \delta \left(1 - \sum_l l \left[n_+ (l) + n_- (l) \right] \right) \tag{A10}$$

where again the notation $\delta(X)$ is introduced for better readability of the arguments X but actually represents the Kronecker symbol $\delta_{0,X}$.

3. Typical values of the empirical 1-interval observables

Since the probability of a disorder configuration is given by Eq. 2 with $p_+ + p_- = 1$, the probability distributions of the lengths l of the intervals $v = \pm$ are given by the geometrical distributions

$$p_{\pm}^{geo}(l) = (1 - p_{\pm})(p_{\pm})^{l-1} \tag{A11}$$

with the normalization

$$\sum_{l=1}^{+\infty} p_{\pm}^{geo}(l) = 1 \tag{A12}$$

and the averaged lengths

$$\sum_{l=1}^{+\infty} l p_{\pm}^{geo}(l) = \frac{1}{1 - p_{\pm}} = \frac{1}{p_{\mp}} \tag{A13}$$

As a consequence, the typical density $\frac{N^{typ}}{L}$ of the intervals reads

$$\frac{N^{typ}}{L} = \frac{1}{\sum_{l=1}^{+\infty} l \left[p_+^{geo}(l) + p_-^{geo}(l) \right]} = p_+ p_- \tag{A14}$$

and the typical values of the empirical 1-interval observables are

$$\begin{aligned} n_+^{typ}(l) \quad &= \frac{N^{typ}}{L} p_+^{geo}(l) = p_-^2 p_+^l \\ n_-^{typ}(l) \quad &= \frac{N^{typ}}{L} p_-^{geo}(l) = p_+^2 p_-^l \end{aligned} \tag{A15}$$

4. Large deviations of empirical 1-interval observables

In order to analyze the large deviations of empirical 1-interval observables, one needs to introduce a generalized semi-Markovian model for the disorder, where the lengths l_i of the intervals are drawn with some general distributions $p_{\pm}(l)$ (instead of the geometric distributions of Eq. A11). The probability of some configuration of the intervals then reads (up to boundary terms that can be neglected for $L \to +\infty$

$$\begin{aligned} \mathbb{P}_L^{SemiMarkov} \left[(l_{2N} ... l_1) \right] \quad &\simeq \delta \left(L - \sum_{i=1}^{2N} l_i \right) \prod_{i=1}^{N} p_+(l_{2i}) p_-(l_{2i-1}) = \delta \left(L - \sum_{i=1}^{2N} l_i \right) e^{\sum_{i=1}^{N} [\ln(p_+(l_{2i})) + \ln(p_-(l_{2i-1}))]} \\ &\simeq c_1[n_+(.); n_-(.)] e^{-La[n_+(.); n_-(.)]} \end{aligned} \tag{A16}$$

where the action in the exponential is a function of the empirical 1-interval observables introduced in Eq. A7

$$a[n_+(.);n_-(.)] = -\sum_l \left[n_+(l)\ln\left(p_+(l)\right) + n_-(l)\ln\left(p_-(l)\right)\right] \tag{A17}$$

while $c_1[n_+(.);n_-(.)]$ has been introduced in Eq. A10 to summarize the constraints. In this semi-Markovian model, all the disorder configurations that have the same empirical 1-interval observables $n^{\pm}(.)$ have the same probability. As a consequence, the probability $P_L[n_+(.);n_-(.)]$ to see these empirical observables is given by

$$P_L[n_+(.);n_-(.)] = c_1[n_+(.);n_-(.)]\omega_L[n_+(.);n_-(.)]e^{-La[n_+(.);n_-(.)]} \tag{A18}$$

where $\omega_L[n_+(.);n_-(.)]$ counts the number of disorder configurations that correspond to these empirical observables, while the normalization reads

$$1 \quad = \left(\prod_l \int dn_+(l) \int dn_-(l)\right) P_L[n_+(.);n_-(.)] \tag{A19}$$

When the empirical 1-interval observables take their typical values for this semi-Markovian generalized model (adapted from Eqs A14 and A15)

$$n_{\pm}^{typ}(l) \quad = \frac{p_{\pm}(l)}{\sum_{l'=1}^{+\infty} l' \left[p_+(l') + p_-(l')\right]} \tag{A20}$$

the probability $P_L[n_+^{typ}(.);n_-^{typ}(.)]$ should remain finite as $L \to +\infty$. So the factor $\omega_L[n_+^{typ}(.);n_-^{typ}(.)]$ should compensate exactly the exponential factor of Eq. A18, i.e. it should display the exponential growth

$$\omega_L[n_+^{typ}(.);n_-^{typ}(.)] \quad \underset{L\to+\infty}{\simeq} e^{+La[n_+^{typ}(.);n_-^{typ}(.)]} = e^{-L\sum_l \left[n_+^{typ}(l)\ln\left(p_+(l)\right) + n_-^{typ}(l)\ln\left(p_-(l)\right)\right]} \tag{A21}$$

When the empirical observables $[n_+(.);n_-(.)]$ are different from their typical values $[n_+^{typ}(.);n_-^{typ}(.)]$, we may consider a modified semi-Markovian model with modified probability distributions $\tilde{p}_{\pm}(l)$ for the lengths of the intervals that would make the empirical observables $[n_+(.);n_-(.)]$ typical for this modified model. Equations A20 yield that the modified probability distributions $\tilde{p}_{\pm}(l)$ should be chosen as

$$\tilde{p}_{\pm}(l) \quad = \frac{n_{\pm}(l)}{\sum_{l'} n_{\pm}(l')} \tag{A22}$$

where the two denominators coincide as a consequence of the constraints of Eq. A10

Then Eq. A21 translates for this modified model into

$$\begin{aligned}\omega_L[n_+(.);n_-(.)] \quad &\underset{L\to+\infty}{\simeq} e^{-L\sum_l \left[n_+(l)\ln\left(\tilde{p}_+(l)\right) + n_-(l)\ln\left(\tilde{p}_-(l)\right)\right]} \\ &= e^{-L\sum_l \left[n_+(l)\ln\left(\frac{n_+(l)}{\sum_{l'} n_+(l')}\right) + n_-(l)\ln\left(\frac{n_-(l)}{\sum_{l'} n_-(l')}\right)\right]}\end{aligned} \tag{A23}$$

Plugging this result into Eq. A18 yields the large deviation form

$$P_L[n_+(.);n_-(.)] \underset{L\to+\infty}{\simeq} c_1[n_+(.);n_-(.)]e^{-LJ[n_+(.);n_-(.)]} \tag{A24}$$

with the rate function

$$J[n_+(.);n_-(.)] = \sum_l \left[n_+(l)\ln\left(\frac{n_+(l)}{p_+(l)\sum_{l'} n_+(l')}\right) + n_-(l)\ln\left(\frac{n_-(l)}{p_-(l)\sum_{l'} n_-(l')}\right)\right] \tag{A25}$$

Related studies on large deviations properties of various semi-Markov processes in continuous time can be found in [11, 106–109].

Here we wish to return to the initial disorder model corresponding to the geometric distributions $p_{\pm}^{geo}(l)$ of Eq. A11, where the result of Eq. A25, concerning the generalized semi-Markov model of disorder configurations with arbitrary distributions $p_{\pm}(l)$ for the lengths of the intervals, becomes

$$J^{geo}[n_+(.);n_-(.)] = \sum_l \left[n_+(l)\ln\left(\frac{n_+(l)}{p_+^{geo}(l)\sum_{l'} n_+(l')}\right) + n_-(l)\ln\left(\frac{n_-(l)}{p_-^{geo}(l)\sum_{l'} n_-(l')}\right)\right] \tag{A26}$$

5. Large deviations for observables given by the product of the intervals contributions

The modulus of Eq. A4 can be rewritten in terms of the empirical 1-interval observables of Eq. A7 as

$$\left|\Theta_L\left([l_i]_{i=1,2,..,2N}\right)\right| = e^{\sum_{i=1}^{N}\left[\ln|\theta_+(l_{2i})|+\ln|\theta_-(l_{2i-1})|\right]} = e^{L\sum_l\left[n_+(l)\ln|\theta_+(l)|+n_-(l)\ln|\theta_-(l)|\right]} \tag{A27}$$

So the corresponding finite-size Lyapunov exponent of Eq. 4 is a linear function of the empirical 1-interval observables

$$\lambda[n_+(.);n_-(.)] = \frac{\ln|\Theta_L|}{L} = \sum_l\left[n_+(l)\ln|\theta_+(l)|+n_-(l)\ln|\theta_-(l)|\right] \tag{A28}$$

Its typical value can be obtained from the typical values of the empirical 1-interval observables of Eq. A15

$$\lambda^{typ} = \sum_l\left[n_+^{typ}(l)\ln|\theta_+(l)|+n_-^{typ}(l)\ln|\theta_-(l)|\right] = \sum_l\left[p_-^2p_+^l\ln|\theta_+(l)|+p_+^2p_-^l\ln|\theta_-(l)|\right] \tag{A29}$$

The moments of non-integer order k of Eq A27 read in terms of the probability $P_L[n_+(.);n_-(.)]$ of Eq A24

$$\begin{aligned}\overline{|\Theta_L|^k} &= \left(\prod_l\int dn_+(l)\int dn_-(l)\right)P_L[n_+(.);n_-(.)]e^{Lk\lambda[n_+(.);n_-(.)]}\\ &\underset{L\to+\infty}{\simeq}\left(\prod_l\int dn_+(l)\int dn_-(l)\right)c_1[n_+(.);n_-(.)]e^{L[k\lambda[n_+(.);n_-(.)]-J^{geo}[n_+(.);n_-(.)]]}\end{aligned} \tag{A30}$$

One thus needs to optimize the function $[k\lambda[n_+(.);n_-(.)]-J^{geo}[n_+(.);n_-(.)]]$ in the exponential in the presence of the constraints $c_1[n_+(.);n_-(.)]$ of Eq. A10 that can be taken into account via Lagrange multipliers. It is technically more convenient to introduce the empirical density of intervals $\pm$ that appear in the constraints $c_1[n_+(.);n_-(.)]$ and in the rate function $J^{geo}[n_+(.);n_-(.)]$ of Eq. A26

$$n \equiv \sum_l n_+(l) = \sum_l n_-(l) \tag{A31}$$

via another constraint.

So we will consider the functional

$$\begin{aligned}\mathcal{L}_k[n_+(.),n_-(.),n] &= k\lambda[n_+(.);n_-(.)]-J^{geo}[n_+(.);n_-(.)]\\ &+\chi_+\left(n-\sum_l n_+(l)\right)+\chi_-\left(n-\sum_l n_-(l)\right)+\varphi\left(1-\sum_l l\left[n_+(l)+n_-(l)\right]\right)\\ &=\sum_l n_+(l)\left[k\ln|\theta_+(l)|-\ln\left(\frac{n_+(l)}{p_+^{geo}(l)}\right)\right]+\sum_l n_-(l)\left[k\ln|\theta_-(l)|-\ln\left(\frac{n_-(l)}{p_-^{geo}(l)}\right)\right]+2n\ln(n)\\ &+\chi_+\left(n-\sum_l n_+(l)\right)+\chi_-\left(n-\sum_l n_-(l)\right)+\varphi\left(1-\sum_l l\left[n_+(l)+n_-(l)\right]\right)\end{aligned} \tag{A32}$$

The optimization with respect to the empirical 1-interval observable $n_\pm(l)$

$$0 = \frac{\partial\mathcal{L}_k[n_+(.),n_-(.),n]}{\partial n_\pm(l)} = -1+k\ln|\theta_\pm(l)|-\ln\left(\frac{n_\pm(l)}{p_\pm^{geo}(l)}\right)-\chi_\pm-\varphi l \tag{A33}$$

yields the forms

$$n_\pm(l) = e^{-1-\chi_\pm}\left|\theta_\pm(l)\right|^k p_\pm^{geo}(l)e^{-\varphi l} \tag{A34}$$

The constraints

$$n = \sum_l n_\pm(l) = e^{-1-\chi_\pm}\sum_l\left|\theta_\pm(l)\right|^k p_\pm^{geo}(l)e^{-\varphi l} \tag{A35}$$

determine the Lagrange multipliers $\chi_\pm$ as a function of the other parameters

$$e^{-1-\chi_\pm} = \frac{n}{\sum_l |\theta_\pm(l)|^k \, p_\pm^{geo}(l) e^{-\varphi l}} \tag{A36}$$

The optimization with respect to the interval density n

$$0 = \frac{\partial \mathcal{L}_k[n_+(.), n_-(.), n]}{\partial n} \quad = 2 + 2\ln(n) + \chi_+ + \chi_- \tag{A37}$$

yields together with Eq. A36 that the value of the Lagrange multiplier φ is fixed by the condition

$$1 = \left[\sum_l |\theta_+(l)|^k \, p_+^{geo}(l) e^{-\varphi l}\right] \left[\sum_{l'} |\theta_-(l')|^k \, p_-^{geo}(l') e^{-\varphi l'}\right] \tag{A38}$$

while the remaining constraint

$$1 = \sum_l l \left[n_+(l) + n_-(l)\right] \tag{A39}$$

determines the value of the density n.

The value of the functional of Eq. A32 for the optimal solution satisfying the constraints

$$\begin{aligned}
\mathcal{L}_k^{opt} &= \sum_l n_+(l) \left[k \ln|\theta_+(l)| - \ln\left(\frac{n_+(l)}{p_+^{geo}(l)}\right)\right] + \sum_l n_-(l) \left[k \ln|\theta_-(l)| - \ln\left(\frac{n_-(l)}{p_-^{geo}(l)}\right)\right] + 2n\ln(n) \\
&= \sum_l n_+(l) \left[1 + \chi_+ + \varphi l\right] + \sum_l n_-(l) \left[1 + \chi_- + \varphi l\right] + 2n\ln(n) \\
&= n(2 + \chi_+ + \chi_- + 2\ln(n)) + \varphi \sum_l l \left[n_+(l) + n_-(l)\right] = \varphi
\end{aligned} \tag{A40}$$

actually reduces to the Lagrange multiplier φ.

In summary, the scaled cumulant generating function $\varphi(k)$ governing the exponential growth of the moments of Eq. A30

$$\overline{|\Theta_L|^k} \quad = \underset{L\to+\infty}{\simeq} e^{L\varphi(k)} \tag{A41}$$

is the solution of Eq A38

$$1 = \left[\sum_{l=1}^{+\infty} |\theta_+(l)|^k \, p_+^{geo}(l) e^{-l\varphi(k)}\right] \left[\sum_{l'=1}^{+\infty} |\theta_-(l')|^k \, p_-^{geo}(l') e^{-l'\varphi(k)}\right] \tag{A42}$$

that involves the distribution $p_\pm^{geo}(l)$ of the lengths of the intervals of the disorder configurations (Eq A11) and the functions $\theta_\pm(l)$ of Eq. A4 of the observable under study. One can check that the expansion at first order in k around $k = 0$ with Eq. 10

$$\varphi(k) \underset{k\to 0}{\simeq} k\lambda^{typ} \tag{A43}$$

allows to recover the typical value λ^{typ} de Eq A29, while the special case $\theta^\pm(l) = (t_\pm)^l$ allows to recover the scaled cumulant generating function $\phi(k)$ of Eq 25 concerning the simpler case of products of random variables.

[1] Y. Oono, Progress of Theoretical Physics Supplement 99, 165 (1989).
[2] R.S. Ellis, Physica D 133, 106 (1999).
[3] H. Touchette, Phys. Rep. 478, 1 (2009).
[4] B. Derrida, J. Stat. Mech. P07023 (2007).
[5] R J Harris and G M Schütz, J. Stat. Mech. P07020 (2007).
[6] E.M. Sevick, R. Prabhakar, S. R. Williams, D. J. Searles, Ann. Rev. of Phys. Chem. Vol 59, 603 (2008).

[7] H. Touchette and R.J. Harris, chapter "Large deviation approach to nonequilibrium systems" of the book "Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond", Wiley 2013.
[8] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim Rev. Mod. Phys. 87, 593 (2015).
[9] A. Lazarescu, J. Phys. A: Math. Theor. 48 503001 (2015).
[10] A. Lazarescu, J. Phys. A: Math. Theor. 50 254004 (2017).
[11] A. de La Fortelle, PhD (2000).
[12] V. Lecomte, PhD Thesis (2007).
[13] R. Chétrite, PhD Thesis (2008).
[14] B. Wynants, PhD Thesis (2010), arXiv:1011.4210.
[15] R. Chétrite, HDR Thesis (2018).
[16] J.M. Luck, Aléa Saclay (1992) "Systèmes désordonnés unidimensionnels".
[17] A. Crisanti, G. Paladin, A. Vulpiani, "Products of random matrices in statistical physics", Springer Verlag (1993).
[18] I.M. Lifshitz, Adv. Phys. 13 (1964) 483.
[19] I.M. Lifshitz, Sov. Phys. Usp. 7 (1965) 549.
[20] I.M. Lifshitz, S.A. Gredeskul and L.A. Pastur, John Wiley and Sons (1987), "Introduction to the theory of disordered systems".
[21] B.I. Shklovskii and A.L. Efros, "Electronic properties of doped semiconductors", Springer Verlag Belin (1984).
[22] R. B. Griffiths, Phys. Rev. Lett. 23, 17 (1969).
[23] Y. Kafri and D. Mukamel, Phys. Rev. Lett. 91, 055502 (2003).
[24] M. Randeria, J. P. Sethna and R. G. Palmer Phys. Rev. Lett. 54, 1321 (1985)
[25] A.J. Bray, Phys. Rev. Lett. 59, 586 (1987).
[26] A.J. Bray, Phys. Rev. Lett. 60, 720 (1988).
[27] F. Iglói and C. Monthus, Phys. Rep. 412, 277 (2005).
[28] F. Iglói and C. Monthus, Eur. Phys. J. B 91, 290 (2018).
[29] M. Janssen, Phys. Rep. 295, 1 (1998).
[30] F. Evers and A.D. Mirlin, Rev. Mod. Phys. 80, 1355 (2008).
[31] A.W.W. Ludwig, Nucl. Phys. B 330, 639 (1990).
[32] J.L. Jacobsen and J.L. Cardy, Nucl. Phys., B515, 701 (1998).
[33] C. Chatelain and B. Berche, Nucl. Phys., B572, 626 (2000).
[34] C. Chatelain, B. Berche and L.N. Shchur, J. Phys. A Math. Gen. 34, 9593 (2001).
[35] G. Palágyi, C. Chatelain, B, Berche and F. Iglói, Eur. Phys. J. B13, 357 (2000).
[36] B. Berche and C. Chatelain, in *Order, disorder, and criticality*, ed. by Yu. Holovatch, World Scientific, Singapore 2004, p.146.
[37] M.J. Thill and H.J. Hilhorst, J. Phys. I 6, 67 (1996).
[38] C. Monthus, B. Berche and C. Chatelain, J Stat. Mech. P12002 (2009)
[39] R.S. Ellis, "Entropy, Large Deviations, and Statistical Mechanics", Springer Verlag New-York (1985).
[40] J.D. Deuschel and D.W. Stroock, "Large Deviations", Academic Press Boston (1989).
[41] A. Dembo and O. Zeitouni, "Large Deviations Techniques and Applications", Springer Verlag Berlin (1998).
[42] F. den Hollander, "Large Deviations", Fields Institute Monographs, American Mathematical Society, Providence (2008).
[43] R. Azencott, M.I. Freidlin and S.R.S. Varadhan, "Large Deviations at Saint-Flour ", Probability at Saint-Flour, Springer Heidelberg (2012)
[44] F. Rassoul-Agha and T. Seppalainen, "A Course on Large Deviations with an Introduction to Gibbs Measures", Graduate Studies in Mathematics volume 162, American Mathematical Society (2015)
[45] M. Talagrand, "Spin-glasses : a challenge for mathematicians", springer New-York (2003)
[46] A. Bovier, "Statistical mechanics of disordered systems : a mathematical perspective", Cambridge University Press, Cambridge (2006).
[47] F. Comets, "Directed Polymers in Random Environments", Probability in Saint-Flour, Lecture Notes in Mathematics 2175, Springer International Publishing (2017).
[48] O. Zeitouni, J. Phys. A Math. Gen. 39, R433 (2006).
[49] A. Comtet and Y. Tourigny, arxiv: 1601.01822
[50] B. Derrida and H. Hilhorst, J. Phys. C Solid State Phys. 14, L539 (1981).
[51] C. Monthus, J. Phys. A: Math. Theor. 51, 465301 (2018).
[52] B.L. Altshuler, Y. Gefen, A. Kamenev and L.S. Levitov, Phys. Rev. Lett. 78, 2803 (1997).
[53] A. De Luca, B.L. Altshuler, V.E. Kravtsov and A. Scardicchio, Phys. Rev. Lett. 113, 046806 (2014).
[54] F. Pietracaprina, V. Ros, A. Scardicchio, Phys. Rev. B 93, 054201 (2016).
[55] C. Monthus, J. Stat. Mech. (2016) 123303
[56] L.B. Ioffe and M. Mézard, Phys. Rev. Lett. 105, 037001 (2010).
[57] M.V. Feigelman, L.B. Ioffe and M. Mézard, Phys. Rev. B 82, 184534 (2010).
[58] O. Dimitrova and M. Mézard, J. Stat. Mech. P01020 (2011).
[59] G. Fayolle and A. de La Fortelle, Problems of Information Transmission 38, 354 (2002).
[60] A. de La Fortelle, Problems of Information Transmission 37 , 120 (2001).
[61] C. Maes and K. Netocny, Europhys. Lett. 82, 30003 (2008)
[62] C. Maes, K. Netocny and B. Wynants, Markov Proc. Rel. Fields. 14, 445 (2008).
[63] B. Wynants, PhD Thesis (2010), arXiv:1011.4210.

[64] A. C. Barato and R. Chetrite, J. Stat. Phys. 160, 1154 (2015).
[65] L. Bertini, A. Faggionato and D. Gabrielli, Ann. Inst. Henri Poincare Prob. and Stat. 51, 867 (2015).
[66] L. Bertini, A. Faggionato and D. Gabrielli, Stoch. Process. Appli. 125, 2786 (2015).
[67] C. Monthus, J. Stat. Mech. 023206 (2019).
[68] C. Monthus, J. Phys. A: Math. Theor. 52, 025001 (2019).
[69] C. Monthus, J. Phys. A: Math. Theor. 52 135003(2019).
[70] C. Maes, K. Netocny and B. Wynants Physica A 387, 2675 (2008).
[71] J. Hoppenau, D. Nickelsen and A. Engel, New J. Phys. 18 083010 (2016).
[72] B. Derrida and H. Spohn, J. Stat. Phys., **51**, 817 (1988).
[73] J. Cook and B. Derrida, J. Stat. Phys. **63**, 505 (1991).
[74] R. Abou-Chacra, P.W. Anderson and D.J. Thouless, J. Phys. C : Solid State Physics 6, 1734 (1973).
[75] R. Abou-Chacra and D. J. Thouless, J. Phys. C: Solid State Phys. 7, 65 (1974).
[76] B. Derrida and G.J. Rodgers, J. Phys. A : Math. Gen. 26, L457 (1993).
[77] J.D. Miller and B. Derrida, J. Stat. Phys. 75, 357 (1994).
[78] B. Derrida, Phys. Rev. B 24, 2613 (1981).
[79] B. Derrida and G. Toulouse, J. Phys. Lett. (France), **46**, L223 (1985).
[80] B. Derrida and H. Flyvbjerg, J. Phys. A Math. Gen. 20, 5273 (1987).
[81] B. Derrida, "Non-self-averaging effects in sums of random variables, spin glasses, random maps and random walks", in "On three levels" Eds M. Fannes et al (1994) New-York Plenum Press.
[82] E.J. Gumbel, " Statistics of extreme" (Columbia University Press, NY 1958).
[83] J. Galambos, " The asymptotic theory of extreme order statistics" (Krieger , Malabar, FL 1987).
[84] L.S. Levitov, Europhys. Lett. 9, 83 (1989).
[85] L.S. Levitov, Phys. Rev. Lett. 64, 547 (1990).
[86] B.L. Altshuler and L.S. Levitov, Phys. Rep. 288, 487 (1997).
[87] L.S. Levitov, Ann. Phys. (Leipzig) 8, 5, 507 (1999).
[88] F. Evers and A. D. Mirlin Phys. Rev. Lett. 84, 3690 (2000);
A.D. Mirlin and F. Evers, Phys. Rev. B 62, 7920 (2000).
[89] Y.V. Fyodorov, A. Ossipov and A. Rodriguez, J. Stat. Mech. L12001 (2009).
[90] Y.V. Fyodorov, A. Kupiainen and C. Webb. arXiv:1509.01366.
[91] O. Yevtushenko and V. E. Kratsov, J. Phys. A 36, 8265 (2003).
[92] O. Yevtushenko and A. Ossipov, J. Phys. A 40, 4691 (2007).
[93] S. Kronmüller, O. M. Yevtushenko and E. Cuevas, J. Phys. A 43, 075001 (2010).
[94] V. E. Kratsov, A. Ossipov, O. M. Yevtushenko and E. Cuevas, Phys. Rev. B 82, 161102(R) (2010).
[95] E. Bogomolny and O. Giraud, Phys. Rev. E 84, 036212 (2012).
[96] E. Bogomolny and O. Giraud, Phys. Rev. E 84, 046208 (2012).
[97] E. Bogomolny and O. Giraud, Phys. Rev. Lett. 106, 044101 (2011).
[98] C. Monthus and T. Garel, J. Stat. Mech. (2010) P09015.
[99] C. Monthus, Entropy 18, 122 (2016).
[100] C. Monthus, J. Stat. Mech. 073301 (2016).
[101] C. Monthus, J. Stat. Mech. 123304 (2017).
[102] B. Derrida and J.L. Lebowitz, Phys. Rev. Lett. 80, 209 (1998)
[103] D.S. Dean and S.N. Majumdar, Phys. Rev. Lett. 97, 160201 (2006) and Phys. Rev. E 77, 041108 (2008).
[104] S. N. Majumdar and M. Vergassola, Phys. Rev. Lett. 102, 060601 (2009).
[105] C. Monthus, arXiv:1904.02448
[106] D. Andrieux and P. Gaspard, JSTAT P11007 (2008).
[107] C. Maes, K.l Netocny and B. Wynants, J. Phys. A: Math. Theor. 42 (2009) 365002
[108] M. Mariani and L. Zambotti, Adv. Appl. Prob. 48, 648 (2016).
[109] A. Faggionato, arXiv:1709.05653