Inference of Markov models from trajectories via large deviations at level 2.5 with applications to random walks in disordered media

Cecile Monthus

Inference of Markov models from trajectories via Large Deviations at Level 2.5 with applications to random walks in disordered media

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

The inference of Markov models from data on stochastic dynamical trajectories over the large time-window T is revisited via the Large Deviations at Level 2.5 for the time-empirical density and the time-empirical flows. The goal is to obtain the large deviations properties for the probability distribution of the inferred Markov parameters in order to characterize their possible fluctuations around the true Markov parameters for large T. The explicit rate functions are given for several settings, namely discrete-time Markov chains, continuous-time Markov jump processes, and diffusion processes in dimension d. Applications to various models of random walks in disordered media are described, where the goal is to infer the quenched disordered variables defining a given disordered sample.

I. INTRODUCTION

Inference has always played a major role in probability and statistics [1–3]. However the recent availability of big data in many fields has triggered an enormous increase of applications of inference methods. In particular, the inference of Markov models from data on stochastic dynamical trajectories has been implemented within various settings [4] including discrete-time Markov chains [5–7], continuous-time Markov jump processes [8–12], and Langevin dynamics [13–31], as well as for several active matter models [32–34].

Besides the development of efficient numerical inference procedures, it is thus important to characterize theoretically the statistical fluctuations of these inferred parameters of Markov models with respect to the 'true' values of parameters. The goal of the present paper is to analyze their large deviations properties with respect to the time-window size T of the observed dynamical trajectories. Within the recent progresses in the field of non-equilibrium stochastic processes (see the reviews with different scopes [35–43], the PhD theses [44–47] and the HDR thesis [48]), the theory of large deviations (see the reviews [49–51] and references therein) plays the role of the unifying language. However the usual classification into three levels, namely Level 1 for empirical observables, Level 2 for the empirical measure, and Level 3 for the empirical process has turned out to be insufficient. So the new intermediate 'Level 2.5' concerning the joint distribution of the empirical measure and of the empirical flows has emerged as the appropriate level for non-equilibrium steady-states, where one can write explicit expressions for the rate functions. This large deviation analysis at Level 2.5 has been applied to various settings, including discrete-time Markov chains [44, 51–54], continuous-time Markov jump processes [44, 47, 48, 54–65] and diffusion processes [47, 48, 54, 58, 67–69]. In the present paper, these large deviations properties at Level 2.5 for the empirical observables of dynamical trajectories are translated into the large deviations for the inferred parameters of Markov models. Applications to some models of random walks in disordered media (see the reviews [70–74]) are described, where the goal is to infer the quenched disordered variables defining a given disordered sample.

The paper is organized as follows. Section II explains the general framework to analyze the large deviations properties for the inference of Markov models from dynamical trajectories. The case of Markov chains in discrete time is discussed in section III, with the application to the random walk on the disordered ring in section IV. The case of Markov jump processes in continuous time is presented in section V, with the application to the directed trap model on a disordered ring in section VI. Finally, the case of diffusion processes in dimension d is described in section VII with the application to the diffusion in a disordered potential in section VIII. Our conclusions are summarized in section IX. Appendix A contains more technical details on the path-integral analysis for diffusion processes.

II. GENERAL PRINCIPLES TO INFER MARKOV MODELS FROM DYNAMICAL TRAJECTORIES

Before the analysis of various specific Markov models (discrete or continuous time, discrete or continuous space) in the next sections, it seems useful in the present section to outline the general principles.

A. Goals and Notations

The guideline of the present paper can be summarized in terms of the following unifying notations :

(i) M is the 'true' Markov model with steady state. Its parameters are unknown and need to be inferred from data.

(ii) D are the data concerning a single dynamical trajectory $x(0 \leq t \leq T)$ of the model M over the long time T.

(iii) E are the relevant time-empirical observables : they can be computed as the averages of simple time-local operators over the trajectory $x(0 \leq t \leq T)$ and they are sufficient to evaluate the probability of the trajectory $x(0 \leq t \leq T)$ for the model M. The large deviation properties of these relevant time-empirical observables are explicitly known for many Markov models M and are called 'Large deviations at Level 2.5', as already mentioned in the introduction.

(iv) $\hat{M}(E)$ is the best Markov model that can be inferred via the maximum-likelihood method from the values E of the relevant time-empirical observables computed from the data D concerning the single dynamical trajectory $x(0 \leq t \leq T)$. The large deviation properties at Level 2.5 of the empirical observables E mentioned in (iii) can be translated into the large deviation properties of the inferred model $\hat{M}$ in order to characterize the fluctuations of inferred model $\hat{M}$ around the true model M for large T.

Let us now describe in more details the various steps of such an analysis.

B. Identification of the relevant time-empirical observables that determine the trajectories probabilities

For the 'true' Markov model M, the first step consists in rewriting the probability of a long dynamical trajectory $x(0 \leq t \leq T)$

$$\mathcal{P}[x(0 \leq t \leq T)] \underset{T \to +\infty}{\simeq} e^{-T A_{[M]}\left(E[x(0 \leq t \leq T)]\right)} \tag{1}$$

in terms of an intensive action $A_{[M]}\left(E[x(0 \leq t \leq T)]\right)$ that depends on the model parameters M, and that only involves a few relevant time-empirical observables $E[x(0 \leq t \leq T)]$ of the dynamical trajectory $x(0 \leq t \leq T)$.

C. Number of dynamical trajectories of length T with the same value of the time-empirical observables

Since all the individual dynamical trajectories $x(0 \leq t \leq T)$ that have the same empirical observables $E = E[x(0 \leq t \leq T)]$ have the same probability given by Eq. 1, one can rewrite the normalization over all possible trajectories as a sum over these empirical observables

$$1 = \sum_{x(0 \leq t \leq T)} \mathcal{P}[x(0 \leq t \leq T)] \underset{T \to +\infty}{\simeq} \sum_{E} \Omega_T(E) e^{-T A_{[M]}(E)} \tag{2}$$

where the number of dynamical trajectories of length T associated to given values E of these empirical observables

$$\Omega_T(E) \equiv \sum_{x(0 \leq t \leq T)} \delta\left(E[x(0 \leq t \leq T)] - E\right) \tag{3}$$

is expected to grow exponentially with respect to the length T of the trajectories

$$\Omega_T(E) \underset{T \to +\infty}{\simeq} C(E)\ e^{T S(E)} \tag{4}$$

while the prefactor $C(E)$ denotes the appropriate constitutive constraints for the empirical observables. The factor $S(E) = \frac{\ln \Omega_T(E)}{T}$ represents the Boltzmann intensive entropy of the set of trajectories of length T with given intensive empirical observables E. Let us now recall how it can be evaluated without any actual computation (i.e. one does not need to use combinatorial methods to count the appropriate configurations).

The normalization of Eq. 2 becomes for large T

$$1 \underset{T \to +\infty}{\simeq} \sum_{E} C(E)\ e^{T\left[S(E) - A_{[M]}(E)\right]} \tag{5}$$

When the empirical variables E take their typical values E_M^{typ} for the model M, the exponential behavior in T of Eq. 5 should exactly vanish, i.e. the entropy $S(E_M^{typ})$ should exactly compensate the action $A_{[M]}\left(E_M^{typ}\right)$

$$S(E_M^{typ}) = A_{[M]}\left(E_M^{typ}\right) \tag{6}$$

To obtain the intensive entropy $S(E)$ for any other given value E of the empirical observables, one just needs to introduce the modified model $\hat{M}(E)$ that would make the empirical values E typical for this modified model

$$E = E_{\hat{M}(E)}^{typ} \tag{7}$$

and to use Eq. 6 for this modified model to obtain

$$S(E) = S(E_{\hat{M}(E)}^{typ}) = A_{[\hat{M}(E)]}\left(E_{\hat{M}(E)}^{typ}\right) = A_{[\hat{M}(E)]}\left(E\right) \tag{8}$$

D. Large deviations for the relevant time-empirical observables E

Eq 2 means that, for the model M, the probability $P_T(E)$ of the empirical observables E over the set of dynamical trajectories of length T, with the normalization

$$1 = \sum_E P_T(E) \tag{9}$$

follows the large deviation form

$$P_T(E) \underset{T\to+\infty}{\simeq} C(E)\ e^{-TI_{[M]}(E)} \tag{10}$$

where the rate function

$$I_{[M]}(E) = A_{[M]}\left(E\right) - S(E) = A_{[M]}(E) - A_{[\hat{M}(E)]}(E) \tag{11}$$

is simply given by the difference between the intensive action $A_{[M]}(E)$ associated to the true model M and the intensive action $A_{[\hat{M}(E)]}(E)$ associated to the modified model $\hat{M}(E)$ that would make the empirical value E typical (see Eq. 7). It is positive $I_{[M]}(E) \geq 0$ and vanishes only for the typical value E_M^{typ}

$$0 = I_{[M]}(E_M^{typ}) \tag{12}$$

i.e. only when the modified model $\hat{M}(E)$ coincides with the true model M.

E. Large deviations for the inferred model $\hat{M}$ obtained via the principle of maximum likelihood

Now we wish to infer the true model M from the empirical observables E computed from the data D. The likelihood $\mathcal{L}_T(M|E)$ of the model M given E is defined as the probability $P_T(E|M)$ to obtain the empirical variables E if the true model is M, i.e. by the probability of Eq. 10 displaying the large deviation behavior

$$\mathcal{L}_T(M|E) = P_T(E|M) \underset{T\to+\infty}{\simeq} C(E)\ e^{-TI_{[M]}(E)} \tag{13}$$

The maximum of this likelihood corresponds to the vanishing of the positive rate function $I_{[M]}(E)$ discussed in Eq. 12, i.e. the best inferred model is the modified model $\hat{M}(E)$ introduced in Eq. 7 that makes the empirical observables E typical. Via the bijective change of variables $\hat{M}(E)$ of Eq. 7 between the empirical data E and the best inferred model $\hat{M}(E)$, Eq. 10 can be translated into the large deviation form for the probability to infer the model $\hat{M}$

$$\begin{aligned} P_T^{Infer}(\hat{M}) &= \sum_E P_T(E)\delta\left(\hat{M} - \hat{M}(E)\right) \underset{T\to+\infty}{\simeq} \sum_E C(E)\ e^{-TI_{[M]}(E)}\delta\left(\hat{M} - \hat{M}(E)\right) \\ &\underset{T\to+\infty}{\simeq} \mathcal{C}(\hat{M}) e^{-T\ \mathcal{I}_{[M]}(\hat{M})} \end{aligned} \tag{14}$$

where the rate function $\mathcal{I}_M(\hat{M})$ can be explicitly obtained via the translation of the rate function $I_{[M]}(E)$ at Level 2.5 of Eq. 11, while the prefactor $\mathcal{C}(\hat{M})$ represents the translation of the constraints $C(E)$ introduced in Eq. 4.

F. Large deviations for inferred parameters if the true model M is parametrized by a few parameters θ

In the previous subsection, we have described the 'full inference' problem where one considers the best inferred model $\hat{M}$ that can be reconstructed from the full information on the relevant empirical observables E. However sometimes one prefers to assume that the true model M belongs to some subspace M_θ parametrized by a few parameters θ that one wishes to infer. The probability to infer the parameters $\hat{\theta}$ is then obtained by applying Eq. 14 to the special case $M = M_\theta$ and $\hat{M} = M_{\hat{\theta}}$ on the right handside

$$P_T^{Infer}(\hat{\theta}) \underset{T \to +\infty}{\simeq} \mathcal{C}(\hat{M}_{\hat{\theta}}) e^{-T \, \mathcal{I}_{[M_\theta]}(M_{\hat{\theta}})} \tag{15}$$

G. Application to the simplest example concerning the drawing of T independent variables

In order to see more concretely how the general formalism described above works in practice, it is useful to revisit now the trivial example of independent variables before focusing on Markov models in the other sections. In this subsection, we thus consider the problem of drawing T independent random variables $x(t)$ where $t = 1, .., T$ with the discrete probability distribution P_x normalized to unity

$$\sum_x P_x = 1 \tag{16}$$

so that the distribution $P_.$ represents the model M that one wishes to infer.

1. *Identification of the relevant empirical observables E*

The probability to draw the sequence $x(1 \leq t \leq T)$ can be rewritten in the form of Eq. 1

$$\mathcal{P}[x(0 \leq t \leq T)] = \prod_{t=1}^{T} P_{x(t)} = e^{\sum_{t=1}^{T} \ln P_{x(t)}} = e^{T \sum_x \rho_x \ln P_x} \tag{17}$$

The only relevant empirical observable E is thus the empirical density (or the empirical histogram)

$$\rho_x \equiv \frac{1}{T} \sum_{t=1}^{T} \delta_{x(t),x} \tag{18}$$

normalized to unity

$$\sum_x \rho_x = 1 \tag{19}$$

The intensive action defined in Eq. 1 is simply

$$A_{[P]}(\rho_.) = -\sum_x \rho_x \ln P_x \tag{20}$$

2. *Large deviations for the empirical observable E, i.e. for the empirical density $\rho_.$*

The typical value of the empirical density $\rho_.$ of Eq. 18 is the true probability $P_.$

$$\rho_x^{typ} = P_x \tag{21}$$

Reciprocally, the modified probability $\hat{P}_.$ that would make the empirical density $\rho_.$ typical is simply

$$\hat{P}_x = \rho_x \tag{22}$$

The corresponding action of Eq. 20 for this modified model

$$A_{[\hat{P}]}\left(\rho_{.}\right) = -\sum_{x} \rho_x \ln \hat{P}_x = -\sum_{x} \rho_x \ln \rho_x \tag{23}$$

is simply the Shannon entropy of the empirical density $\rho_.$. The probability to see the empirical density $\rho_.$ given the true probability $P_.$ follows the Large Deviation form for large T

$$P_T(\rho_.) \underset{T\to+\infty}{\simeq} \delta\left(\sum_x \rho_x - 1\right) e^{-T I_{[P_.]}(\rho_.)} \tag{24}$$

where the normalization constraint ($\sum_x \rho_x = 1$) represents the constitutive constraint denoted by $C(E)$ in Eq. 10, while the rate function $I(\rho_.)$ corresponds to the difference of Eq. 11 between the actions of Eqs 20 and 23

$$I_{[P_.]}(\rho_.) = A_{[P]}(\rho_.) - A_{[\hat{P}]}(\rho_.) = \sum_x \rho_x \ln \frac{\rho_x}{P_x} \tag{25}$$

This standard result is known as Sanov's Theorem : the rate function of Eq. 25 involves the relative entropy of the empirical density $\rho_.$ with respect to the true distribution $P_.$. It is positive $I(\rho_.) \geq 0$ and vanishes only for the typical value of the empirical density of Eq. 21.

3. *Large deviations for the inferred model $\hat{M}$, i.e. for the inferred distribution $\hat{P}_.$*

Here the best inferred distribution $\hat{P}$ simply coincides with the empirical distribution $\rho_.$ (Eq. 22). As a consequence, Eq. 24 can be directly rephrased as the probability to infer the probability distribution $\hat{P}_.$ from the T variables drawn with the true probability $P_.$

$$P_T^{Infer}(\hat{P}_.|P_.) \underset{T\to+\infty}{\simeq} \delta\left(\sum_x \hat{P}_x - 1\right) e^{-T\sum_x \hat{P}_x \ln \frac{\hat{P}_x}{P_x}} \tag{26}$$

4. *Translation for the case of continuous distribution $P(.)$*

Up to now we have considered the case of a discrete distribution $P_.$. However if one wishes to infer a continuous distribution $P(x)$, one just needs to replace discrete sums by integrals in the final result of Eq. 26

$$P_T^{Infer}(\hat{P}(.)|P(.)) \underset{T\to+\infty}{\simeq} \delta\left(\int dx \hat{P}(x) - 1\right) e^{-T\int dx \hat{P}(x) \ln \frac{\hat{P}(x)}{P(x)}} \tag{27}$$

5. *Example of inference of two parameters only*

As an example of the parameters inference described in subsection II F, let us assume that the true distribution is the gamma distribution normalized on $x \in [0, +\infty[$

$$P_{\alpha,\theta}(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)\theta^\alpha} e^{-\frac{x}{\theta}} \tag{28}$$

with the properties

$$\begin{aligned} \int_0^{+\infty} dx x P_{\alpha,\theta}(x) &= \alpha\theta \\ \int_0^{+\infty} dx (\ln x) P_{\alpha,\theta}(x) &= \ln(\theta) + \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \end{aligned} \tag{29}$$

One considers that both the shape parameter α and the scale parameter θ are unknown and need to be inferred from data. The probability to infer the two parameters $(\hat{\alpha}, \hat{\theta})$ follows the large deviation of Eq. 15 that can be evaluated using Eq. 27

$$
\begin{aligned}
P_T^{Infer}(\hat{\alpha}, \hat{\theta} | \alpha, \theta) & \underset{T \to +\infty}{\simeq} e^{-T \int_0^{+\infty} dx P_{\hat{\alpha},\hat{\theta}}(x) \ln \frac{P_{\hat{\alpha},\hat{\theta}}(x)}{P_{\alpha,\theta}(x)}} \\
& \underset{T \to +\infty}{\simeq} e^{-T \left[\ln \left(\frac{\Gamma(\alpha) \theta^{\alpha}}{\Gamma(\hat{\alpha}) \hat{\theta}^{\hat{\alpha}}} \right) + (\hat{\alpha} - \alpha) \int_0^{+\infty} dx (\ln x) P_{\hat{\alpha},\hat{\theta}}(x) + \left(\frac{1}{\theta} - \frac{1}{\hat{\theta}} \right) \int_0^{+\infty} dx x P_{\hat{\alpha},\hat{\theta}}(x) \right]} \\
& \underset{T \to +\infty}{\simeq} e^{-T \left[\ln \left(\frac{\Gamma(\alpha) \theta^{\alpha}}{\Gamma(\hat{\alpha}) \hat{\theta}^{\hat{\alpha}}} \right) + (\hat{\alpha} - \alpha) \left(\ln(\hat{\theta}) + \frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} \right) + \left(\frac{\hat{\theta}}{\theta} - 1 \right) \hat{\alpha} \right]}
\end{aligned}
\tag{30}
$$

III. INFERENCE FOR MARKOV CHAIN IN DISCRETE TIME WITH STEADY STATE

The inference for discrete-time Markov chains has a long history in the mathematical literature [5–7]. In this section, the goal is to revisit this problem via the large deviations at Level 2.5 that have emerged more recently [44, 51–54].

A. Markov chain in discrete time and discrete space parametrized by the Markov matrix $W_{.,.}$

In this section, we consider the Markov chain dynamics for the probability $P_y(t)$ to be at position y at time t

$$
P_x(t+1) = \sum_y W_{x,y} P_y(t) \tag{31}
$$

where the Markov Matrix elements are positive $W_{x,y} \geq 0$ and satisfy the normalization

$$
\sum_x W_{x,y} \quad = 1 \tag{32}
$$

So here the Markov Matrix $W_{.,.}$ represents the model M that one wishes to infer.

We will assume that the steady-state solution $P_x^* \geq 0$ of Eq. 31

$$
P_x^* = \sum_y W_{x,y} P_y^* \tag{33}
$$

exists. From the point of view of the Perron–Frobenius theorem, Eqs 32 and 33 mean that unity is the highest eigenvalue of the positive Markov Matrix $W_{.,.}$, where the positive left eigenvector l_x is constant

$$
l_x = 1 \tag{34}
$$

while the right eigenvector r_x is the steady state

$$
r_x = P_x^* \tag{35}
$$

B. Identification of the relevant time-empirical observables that determine the trajectories probabilities

The probability of the whole trajectory $x(0 \leq t \leq T)$ starting at the fixed position x_0 at time $t = 0$

$$
\mathcal{P}[x(0 \leq t \leq T)] = \delta_{x(0),0} \left[\prod_{t=1}^{T} W_{x(t),x(t-1)} \right] = \delta_{x(0),0} e^{\sum_{t=1}^{T} \ln \left(W_{x(t),x(t-1)} \right)} \tag{36}
$$

can be rewritten in terms of the time-empirical 2-point density that characterizes the flows between two consecutive positions within this trajectory $x(0 \leq t \leq T)$

$$\rho^{(2)}_{x,y} \equiv \frac{1}{T}\sum_{t=1}^{T} \delta_{x(t),x}\delta_{x(t-1),y} \tag{37}$$

as

$$\mathcal{P}[x(0 \leq t \leq T)] = \delta_{x(0),0}\ e^{T\sum\limits_{x,y} \rho^{(2)}_{x,y} \ln\left(W_{x,y}\right)} \tag{38}$$

With respect to the general formalism summarized in Section II, this means that the relevant empirical observable E is the 2-point density $\rho^{(2)}_{.,.}$ of Eq. 37, and that the intensive action introduced in Eq. 1 reads

$$A_{[W]}\left(\rho^{(2)}_{.,.}\right) = -\sum_{x,y} \rho^{(2)}_{x,y} \ln\left(W_{x,y}\right) \tag{39}$$

Note that the 2-point density of Eq. 37 contains the information on the empirical 1-point density that can be obtained via the sum over the first or the second position (up to a boundary term of order $1/T$ that is negligible for large duration $T \to +\infty$)

$$\rho_x \equiv \frac{1}{T}\sum_{t=1}^{T} \delta_{x(t),x} = \sum_y \rho^{(2)}_{x,y} = \sum_y \rho^{(2)}_{y,x} \tag{40}$$

with the normalization

$$\sum_x \rho_x = 1 \tag{41}$$

C. Typical values of the empirical observables

The typical value of the empirical 1-point density is the steady state of Eq. 33

$$\rho_x^{typ} = P^*_x \tag{42}$$

while the typical value of the empirical 2-point density is given by the steady-state flows

$$\rho^{(2)typ}_{x,y} = W_{x,y}P^*_y \tag{43}$$

Reciprocally, the modified Markov matrix elements $\hat{W}_{.,.}$ that would make the empirical densities $\rho_.$ and $\rho^{(2)}_{.,.}$ typical are given by the following ratios

$$\hat{W}_{x,y} \equiv \frac{\rho^{(2)}_{x,y}}{\rho_y} \tag{44}$$

With respect to the general formalism summarized in Section II, this means that the intensive action of Eq. 39 reads for the modified Markov matrix $\hat{W}_{.,.}$ of Eq. 44

$$A_{[\hat{W}]}\left(\rho^{(2)}_{.,.}\right) = -\sum_{x,y} \rho^{(2)}_{x,y} \ln\left(\frac{\rho^{(2)}_{x,y}}{\rho_y}\right) \tag{45}$$

D. Large deviations at level 2.5 for the relevant time-empirical observables

For large T, the joint probability to see the empirical 2-point and 1-point densities follows the large deviation form at Level 2.5 [44, 51–54]

$$P_T(\rho^{(2)}_{.,.};\rho_.) \underset{T\to+\infty}{\simeq} C(\rho^{(2)}_{.,.};\rho_.)e^{-TI_{2.5}(\rho^{(2)}_{.,.};\rho_.)} \tag{46}$$

with the constraints discussed in Eqs 40 and 41

$$C(\rho^{(2)}_{.,.};\rho_.) = \delta\left(\sum_x \rho_x - 1\right)\left[\prod_x \delta\left(\sum_y \rho^{(2)}_{x,y} - \rho_x\right)\right]\left[\prod_y \delta\left(\sum_x \rho^{(2)}_{x,y} - \rho_y\right)\right] \tag{47}$$

while the rate function

$$I_{2.5}(\rho^{(2)}_{.,.};\rho_.) = \sum_x \sum_y \rho^{(2)}_{x,y} \ln\left(\frac{\rho^{(2)}_{x,y}}{W_{x,y}\rho_y}\right) \tag{48}$$

is positive and vanishes only for the typical values of Eqs 42 and 43. With respect to the general formalism summarized in Section II, the rate function of Eq. 48 indeed corresponds to the difference of Eq. 11 between the actions of Eqs 39 and 45

$$I_{2.5}(\rho^{(2)}_{.,.};\rho_.) = A_{[W]}\left(\rho^{(2)}_{.,.}\right) - A_{[\hat{W}]}\left(\rho^{(2)}_{.,.}\right) \tag{49}$$

E. Probability to infer the Markov matrix $\hat{W}_{.,.}$ with its associated steady state $\hat{P}^*_.$

From the empirical 1-point and 2-point densities measured over a very long trajectory $x(0 \leq t \leq T)$ (see Eqs 37 and 40), the maximum likelihood inference yields that the best inferred steady state $P^*_.$ corresponds to the 1-point empirical density (see Eq. 42)

$$\hat{P}^*_x = \rho_x \tag{50}$$

while the best inferred Markov matrix $\hat{W}_{.,.}$ corresponds to the modified matrix of Eq. 44.

Via this bijective change of variables, Eq. 46 yields that the joint probability to infer the Markov matrix $\hat{W}_{.,.}$ and its associated steady state $\hat{P}^*_.$ reads

$$\begin{aligned} P_T^{Infer}(\hat{W}_{.,.};\hat{P}^*_.) \quad & \underset{T\to+\infty}{\simeq} \delta\left(\sum_x \hat{P}^*_x - 1\right)\left[\prod_x \delta\left(\sum_y \hat{W}_{x,y}\hat{P}^*_y - \hat{P}^*_x\right)\right]\left[\prod_y \delta\left(\sum_x \hat{W}_{x,y} - 1\right)\right] \\ & e^{-T\sum_y \hat{P}^*_y \sum_x \hat{W}_{x,y} \ln\left(\frac{\hat{W}_{x,y}}{W_{x,y}}\right)} \end{aligned} \tag{51}$$

The first constraint corresponds to the normalization of the inferred steady state $\hat{P}^*$. The two other constraints mean that the inferred Markov Matrix $\hat{W}_{.,.}$ has unity as highest eigenvalue, with the inferred steady state $\hat{r}_. = \hat{P}^*_.$ as right eigenvector, and the trivial left eigenvector $\hat{l}_x = 1$. So these constraints are in direct correspondence with the properties of the true steady state and the true Markov Matrix (see Eqs 32 and 33).

The formula of Eq. 51 will be applied in section IV to the random walk on a disordered ring.

F. Translation for discrete-time Markov chains in continuous space $\vec{x}$ in dimension d

For discrete-time Markov chains in continuous space $\vec{x}$ in dimension d with kernel $W(\vec{x},\vec{y})$

$$P_{t+1}(\vec{x}) = \int d^d\vec{y}\; W(\vec{x},\vec{y}) P_t(\vec{y}) \tag{52}$$

one just needs to replace discrete sums by integrals in the final result of Eq. 51 to obtain that the joint probability to infer the Markov kernel $\hat{W}(.,.)$ with its associated steady state $\hat{P}^*(.)$ is given by

$$\begin{aligned} & P_T^{Infer}(\hat{W}(.,.);\hat{P}^*(.)) \underset{T\to+\infty}{\simeq} \delta\left(\int d^d\vec{x}\; \hat{P}^*(\vec{x}) - 1\right)\left[\prod_{\vec{x}} \delta\left(\int d^d\vec{y}\; \hat{W}(\vec{x},\vec{y})\hat{P}^*(\vec{y}) - \hat{P}^*(\vec{x})\right)\right]\left[\prod_y \delta\left(\int d^d\vec{x}\; \hat{W}(\vec{x},\vec{y}) - 1\right)\right] \\ & e^{-T\int d^d\vec{y}\; \hat{P}^*(\vec{y}) \int d^d\vec{x}\; \hat{W}(\vec{x},\vec{y}) \ln\left(\frac{\hat{W}(\vec{x},\vec{y})}{W(\vec{x},\vec{y})}\right)} \end{aligned} \tag{53}$$

This formula will be useful in subsection VII G.

IV. APPLICATION TO THE RANDOM WALK ON THE DISORDERED RING OF L SITES

In this section, the large deviations analysis of inference for discrete-time Markov chains described in the previous section is applied to the example of the random walk on a disordered ring [75, 76].

A. Model parametrization and non-equilibrium steady state

The Derrida-Pomeau model [75, 76] is defined on a ring of L sites with periodic boundary conditions $x + L \equiv x$, and corresponds to the dynamics of Eq. 31 where the Markov Matrix

$$W_{x,y} = \delta_{x,y+1} R_y + \delta_{x,y-1}(1 - R_y) \tag{54}$$

is parametrized by the L parameters $R_y \in]0,1[$ for $y = 1, .., L$. So when the particle is on site y at time t, the new position at time $(t+1)$ can be either the right neighbor $(y+1)$ with probability $R_y \in]0,1[$ or the left neighbor $(y-1)$ with the complementary probability $(1 - R_y) \in]0,1[$.

The steady state of Eq. 33

$$P^*_x = R_{x-1} P^*_{x-1} + (1 - R_{x+1}) P^*_{x+1} \tag{55}$$

reads [75, 76]

$$P^*_x = \frac{K}{R_x} \left[1 + \sum_{z=1}^{L-1} \prod_{y=1}^{z} s_{x+y} \right] = \frac{K}{R_x} \left[1 + s_{x+1} + s_{x+1} s_{x+2} + ... + s_{x+1} s_{x+2} ... s_{x+L-1} \right] \tag{56}$$

in terms of the ratios

$$s_x \equiv \frac{1 - R_x}{R_x} \tag{57}$$

while the constant K is fixed by the normalization

$$1 = \sum_{x=1}^{L} P^*_x = K \sum_{x=1}^{L} \frac{1}{R_x} \left[1 + \sum_{z=1}^{L-1} \prod_{y=1}^{z} s_{x+y} \right] \tag{58}$$

When the probabilities R_y are random, the characteristic structure of Eq. 56 is known as Kesten random variables and appears in many disordered systems [54, 71, 77–85]. The generalization to the matrix framework is discussed in the recent work [86] and in references therein.

B. Inference of the L parameters R_y of the model

Here the trajectory data are the positions $x(t)$ for the discrete times $t = 0, 1, 2, .., T$. For each site $y = 1, .., L$ on the ring, one computes the 2-point density $\rho^{(2)}_{x,y}$ of Eq. 37 for the only two possible values $x = y \pm 1$ for the model of Eq. 54

$$\begin{aligned} \rho^{(2)}_{y+1,y} &= \frac{1}{T} \sum_{t=1}^{T} \delta_{x(t),y+1} \delta_{x(t-1),y} \\ \rho^{(2)}_{y-1,y} &= \frac{1}{T} \sum_{t=1}^{T} \delta_{x(t),y-1} \delta_{x(t-1),y} \end{aligned} \tag{59}$$

The 1-point density ρ_y of Eq 40 corresponds to their sum

$$\rho_y = \frac{1}{T} \sum_{t=1}^{T} \delta_{x(t-1),y} = \rho^{(2)}_{y+1,y} + \rho^{(2)}_{y-1,y} \tag{60}$$

So the non-vanishing matrix elements of the best inferred Markov matrix $\hat{W}_{x,y}$ of Eq. 44 for $x = y \pm 1$ are computed via the ratios

$$\begin{aligned} \hat{W}_{y+1,y} &\equiv \frac{\rho^{(2)}_{y+1,y}}{\rho_y} = \frac{\rho^{(2)}_{y+1,y}}{\rho^{(2)}_{y+1,y} + \rho^{(2)}_{y-1,y}} \equiv \hat{R}_y \\ \hat{W}_{y-1,y} &\equiv \frac{\rho^{(2)}_{y-1,y}}{\rho_y} = \frac{\rho^{(2)}_{y-1,y}}{\rho^{(2)}_{y+1,y} + \rho^{(2)}_{y-1,y}} \equiv 1 - \hat{R}_y \end{aligned} \tag{61}$$

In summary, the trajectory data have been used to compute the L parameters $\hat{R}_y \in]0,1[$ that parametrize the best inferred Markov matrix $\hat{W}_{.,.}$ of Eq. 44

$$\hat{W}_{x,y} = \delta_{x,y+1}\hat{R}_y + \delta_{x,y-1}(1 - \hat{R}_y) \tag{62}$$

C. Large deviations for the L inferred parameters $\hat{R}_y$

Now one wishes to know how the L inferred parameters $\hat{R}_y \in]0,1[$ computed from the data can fluctuate with respect to the 'true' values R_y of the 'true' model of Eq. 54. The inferred steady state $\hat{P}^*_x$ of Eq. 50 corresponds to the steady state associated to the model with the inferred parameters $\hat{R}_y$ and is thus given by the analog of Eqs 56 and 57

$$\hat{P}^*_x = \frac{\hat{K}}{\hat{R}_x}\left[1 + \sum_{z=1}^{L-1}\prod_{y=1}^{z}\left(\frac{1-\hat{R}_{x+y}}{\hat{R}_{x+y}}\right)\right] \tag{63}$$

where the constant $\hat{K}$ is fixed by the normalization as in Eq. 58

$$1 = \sum_{x=1}^{L}\hat{P}^*_x = \hat{K}\sum_{x'=1}^{L}\frac{1}{\hat{R}_{x'}}\left[1 + \sum_{z=1}^{L-1}\prod_{y=1}^{z}\left(\frac{1-\hat{R}_{x'+y}}{\hat{R}_{x'+y}}\right)\right] \tag{64}$$

As a consequence, the joint probability to infer the L parameters $\hat{R}_.$ of a given disordered ring follows the large deviation form of Eq. 51

$$P_T^{Infer}(\hat{R}_.) \underset{T\to+\infty}{\simeq} e^{-T\mathcal{I}_{R_.}(\hat{R}_.)} \tag{65}$$

with the explicit rate function

$$\begin{aligned} \mathcal{I}_{R_.}(\hat{R}_.) &= \sum_{x=1}^{L}\left[\hat{R}_x \ln\left(\frac{\hat{R}_x}{R_x}\right) + (1-\hat{R}_x)\ln\left(\frac{1-\hat{R}_x}{1-R_x}\right)\right]\hat{P}^*_x \\ &= \sum_{x=1}^{L}\left[\hat{R}_x \ln\left(\frac{\hat{R}_x}{R_x}\right) + (1-\hat{R}_x)\ln\left(\frac{1-\hat{R}_x}{1-R_x}\right)\right]\frac{\frac{1}{\hat{R}_x}\left[1 + \sum_{z=1}^{L-1}\prod_{y=1}^{z}\left(\frac{1-\hat{R}_{x+y}}{\hat{R}_{x+y}}\right)\right]}{\sum_{x'=1}^{L}\frac{1}{\hat{R}_{x'}}\left[1 + \sum_{z=1}^{L-1}\prod_{y=1}^{z}\left(\frac{1-\hat{R}_{x'+y}}{\hat{R}_{x'+y}}\right)\right]} \end{aligned} \tag{66}$$

The main qualitative conclusion is thus that the L inferred parameters $\hat{R}_x$ are coupled via the inferred steady state $\hat{P}^*_.$ that they produce together.

V. INFERENCE FOR MARKOV JUMP PROCESS IN CONTINUOUS TIME WITH STEADY STATE

The inference for continuous-time jump processes has been analyzed in various contexts including evolution models [8], DNA unzipping [9, 10] and the continuous-time random walk in a one-dimensional disordered medium [11, 12]. In this section, the goal is to revisit this problem via the large deviations at Level 2.5 that have emerged more recently [44, 47, 48, 54–65].

A. Markov jump process in continuous time and discrete space

In this section, we consider the continuous-time dynamics in discrete space defined by the Master Equation

$$\frac{\partial P_x(t)}{\partial t} = \sum_y w_{x,y} P_y(t) \tag{67}$$

where the off-diagonal $x \neq y$ positive matrix elements $w_{x,y} \geq 0$ represent the transitions rates per unit time from y to x, while the corresponding diagonal elements are negative and fixed by the conservation of probability to be

$$w_{y,y} \quad = -\sum_{x \neq y} w_{x,y} \tag{68}$$

As in Eq. 33, we will assume that the steady-state P^*_x of Eq. 67

$$0 = \sum_y w_{x,y} P^*_y \tag{69}$$

exists. Eqs 68 and 69 mean that zero is the highest eigenvalue of the Markov Matrix $w_{.,.}$, with the positive left eigenvector

$$l_x = 1 \tag{70}$$

and the positive right eigenvector r_x given by the steady state

$$r_x = P^*_x \tag{71}$$

B. Identification of the relevant time-empirical observables that determine the trajectories probabilities

The probability of the whole trajectory $x(0 \leq t \leq T)$

$$\mathcal{P}[x(0 \leq t \leq T)] = e^{\left[\sum_{t: x(t^+) \neq x(t^-)} \ln(w_{x(t^+),x(t^-)}) + \int_0^T dt w_{x(t),x(t)} \right]} \tag{72}$$

can be rewritten as

$$\mathcal{P}[x(0 \leq t \leq T)] = e^{T \left[\sum_y \sum_{x \neq y} q_{x,y} \ln(w_{x,y}) + \sum_x \rho_x w_{x,x} \right]} \tag{73}$$

in terms of the empirical density

$$\rho_x \quad \equiv \frac{1}{T} \int_0^T dt \ \delta_{x(t),x} \tag{74}$$

normalized to unity

$$\sum_x \rho_x \quad = 1 \tag{75}$$

and in terms of the jump densities for $x \neq y$

$$q_{x,y} \equiv \frac{1}{T} \sum_{t: x(t^+) \neq x(t^-)} \delta_{x(t^+),x} \delta_{x(t^-),y} \tag{76}$$

that should satisfy the stationarity constraint (for any x, the total incoming flow should be equal to the total outgoing flow)

$$\sum_{y \neq x} q_{x,y} = \sum_{y \neq x} q_{y,x} \tag{77}$$

With respect to the general formalism summarized in Section II, this means that the relevant empirical observables E are the empirical density $\rho_.$ and the flows $q_{.,.}$. The corresponding intensive action introduced in Eq. 1 reads using Eq. 68

$$A_{[w]}\left(\rho_.;q_{.,.}\right) \quad = -\sum_y \rho_y w_{y,y} - \sum_y \sum_{x\neq y} q_{x,y}\ln(w_{x,y}) = \sum_y \sum_{x\neq y} \left[w_{x,y}\rho_y - q_{x,y}\ln(w_{x,y})\right] \tag{78}$$

C. Typical values of the empirical observables

The typical value of the empirical density is the steady state of Eq. 69

$$\rho_x^{typ} = P_x^* \tag{79}$$

while the typical value of the jump densities read

$$q_{x,y}^{typ} = w_{x,y}P_y^* \tag{80}$$

Reciprocally, the modified off-diagonal $x \neq y$ matrix elements $\hat{w}_{x,y}$ that would make typical the empirical density ρ_y and the empirical jump-density $q_{x,y}$ correspond to the following ratios

$$\hat{w}_{x,y} \equiv \frac{q_{x,y}}{\rho_y} \tag{81}$$

The corresponding modified diagonal matrix elements are fixed by the conservation of probabilities as in Eq. 68

$$\hat{w}_{y,y} \quad = -\sum_{x\neq y} \hat{w}_{x,y} = -\sum_{x\neq y} \frac{q_{x,y}}{\rho_y} \tag{82}$$

With respect to the general formalism summarized in section II, this means that the intensive action of Eq. 78 reads for the modified Markov matrix $\hat{w}_{.,.}$ of Eqs 81 and 82

$$A_{[\hat{w}]}\left(\rho_.;q_{.,.}\right) = \sum_y \sum_{x\neq y} \left[q_{x,y} - q_{x,y}\ln\left(\frac{q_{x,y}}{\rho_y}\right)\right] \tag{83}$$

D. Large deviations at level 2.5 for the empirical density and the empirical flows

The joint probability distribution of the empirical density $\rho_.$ and flows $q_{.,.}$ satisfy the following large deviation form at level 2.5 [44, 47, 48, 54–66]

$$P_T[\rho_.;q_{.,.}] \underset{T\to+\infty}{\propto} C[\rho_.;q_{.,.}]e^{-TI_{2.5}[\rho_.;q_{.,.}]} \tag{84}$$

with the constraints discussed in Eqs 75 and 77

$$C[\rho_.;q_{.,.}] = \delta\left(\sum_x \rho_x - 1\right)\prod_x \delta\left(\sum_{y\neq x} q_{x,y} - \sum_{y\neq x} q_{y,x}\right) \tag{85}$$

while the rate function reads

$$I_{2.5}[\rho_.;q_{.,.}] = \sum_y \sum_{x\neq y} \left[q_{x,y}\ln\left(\frac{q_{x,y}}{w_{x,y}\rho_y}\right) - q_{x,y} + w_{x,y}\rho_y\right] \tag{86}$$

and vanishes only for the typical values of Eqs 79 and 80. With respect to the general formalism summarized in section II, the rate function of Eq. 86 indeed corresponds to the difference of Eq. 11 between the actions of Eqs 78 and 83

$$I_{2.5}(\rho_.;q_{.,.}) = A_{[w]}\left(\rho_.;q_{.,.}\right) - A_{[\hat{w}]}\left(\rho_.;q_{.,.}\right) \tag{87}$$

E. Probability to infer the Markov matrix $\hat{w}_{.,.}$ with its associated steady state $\hat{P}^*_.$

From the empirical density and the empirical flows measured over a very long trajectory $x(0 \leq t \leq T)$ (see Eqs 74 and 76), the maximum likelihood inference yields that the best inferred steady state $P^*_.$ corresponds to the empirical density (see Eq. 79)

$$\hat{P}^*_x = \rho_x \tag{88}$$

while the best inferred Markov matrix $\hat{w}_{.,.}$ is given by Eqs 81 and 82. Via this bijective change of variables, Eq. 84 yields that the joint probability to infer the Markov matrix $\hat{w}_{.,.}$ and its associated steady state $\hat{P}^*_.$ follows the large deviation form

$$\begin{aligned} P_T^{Infer}[\hat{w}_{.,.};\hat{P}^*_.]_{\underset{T\to+\infty}{\propto}} \quad & \delta\left(\sum_x \hat{P}^*_x - 1\right)\left[\prod_x \delta\left(\sum_y \hat{w}_{x,y}\hat{P}^*_y\right)\right]\left[\prod_y \delta\left(\sum_x \hat{w}_{x,y}\right)\right] \\ & e^{-T\sum_y \hat{P}^*_y \sum_{x\neq y}\left[\hat{w}_{x,y}\ln\left(\frac{\hat{w}_{x,y}}{w_{x,y}}\right) - \hat{w}_{x,y} + w_{x,y}\right]} \end{aligned} \tag{89}$$

The first constraint corresponds to the normalization of the inferred steady state $\hat{P}^*$. The two other constraints mean that the inferred Markov Matrix has zero as highest eigenvalue, with the inferred steady state $\hat{r}_. = \hat{P}^*_.$ as right eigenvector, and the trivial left eigenvector $\hat{l}_x = 1$. Again these constraints are in direct correspondance with the properties of the true steady state and the true Markov Matrix (see Eqs 68 and 69).

The formula of Eq. 89 will be applied in section VI to the directed trap model on a disordered ring.

F. Translation for continuous-time Markov jump processes in continuous space $\vec{x}$ in dimension d

For continuous-time Markov jump processes in continuous space $\vec{x}$ in dimension d with kernel $w(\vec{x},\vec{y})$

$$\frac{\partial P_t(\vec{x})}{\partial t} = \int d^d y\ w(\vec{x},\vec{y})P_t(\vec{y}) - \left(\int d^d y\ w(\vec{y},\vec{x})\right)P_t(\vec{x}) \tag{90}$$

Eq. 89 translates into the following probability to infer the Markov kernel $\hat{w}(.,.)$ with its associated steady state $\hat{P}^*(.)$

$$\begin{aligned} P_T[\hat{w}(.,.);\hat{P}^*(.)]_{\underset{T\to+\infty}{\propto}} \quad & \delta\left(\int d^d y \hat{P}^*(\vec{x}) - 1\right)\prod_x \delta\left(\int d^d y\ w(\vec{x},\vec{y})P^*(\vec{y}) - \left(\int d^d y\ w(\vec{y},\vec{x})\right)P^*(\vec{x})\right) \\ & e^{-T\int d^d y \hat{P}^*(\vec{y})\int d^d x\left[\hat{w}(\vec{x},\vec{y})\ln\left(\frac{\hat{w}(\vec{x},\vec{y})}{w(\vec{x},\vec{y})}\right) - \hat{w}(\vec{x},\vec{y}) + w(\vec{x},\vec{y})\right]} \end{aligned} \tag{91}$$

VI. APPLICATION TO THE DIRECTED TRAP MODEL ON A DISORDERED RING OF L SITES

In this section, the large deviations analysis of inference for continuous-time Markov jump processes described in the previous section is applied to the directed disordered trap model [87–90], whose large deviations properties have been studied recently [61, 91, 92]. Note that besides this directed disordered trap model, many other trap models have been also under study in the context of anomalously slow glassy behaviors [93–101].

A. Model parametrization and steady state

In this section, we consider the directed disordered trap model [61, 87–92] on a ring of L sites with periodic boundary conditions $x + L \equiv x$. The dynamics is defined by the master equation 67 where the Markov matrix

$$w_{x,y} = \delta_{x,y+1}\frac{1}{\tau_y} - \delta_{x,y}\frac{1}{\tau_y} \tag{92}$$

is parametrized by the L trapping times $\tau_y \in]0, +\infty[$. So when the particle is on site y at time t, the only possible move is to jump to the right neighbor $(y+1)$ with the rate $\frac{1}{\tau_y}$ per unit time.

The solution for the steady state of Eq. 69

$$0 = \frac{1}{\tau_{x-1}} P^*_{x-1} - \frac{1}{\tau_x} P^*_x \tag{93}$$

is simply given by

$$P^*_x = \frac{\tau_x}{\sum_{x'=1}^{L} \tau_{x'}} \tag{94}$$

B. Inference of the L trapping times τ_y of the model

Here the trajectory data are the positions $x(t)$ for the continuous time $t \in [0, T]$. For each site $y = 1, .., L$ on the ring, one computes the empirical density ρ_y of Eq. 74, and the empirical density of jumps $q_{x,y}$ of Eq 76 towards the only possible site $x = y + 1$ in the present directed trap model of Eq. 92

$$q_{y+1,y} = \frac{1}{T} \sum_{t: x(t^+) \neq x(t^-)} \delta_{x(t^+),y+1} \delta_{x(t^-),y} \tag{95}$$

So the non-vanishing off-diagonal matrix elements of the best inferred Markov matrix $\hat{w}_{x,y}$ for $x = y+1$ are computed via the ratios of Eq. 81

$$\hat{w}_{y+1,y} = \frac{q_{y+1,y}}{\rho_y} = \frac{\sum_{t: x(t^+) \neq x(t^-)} \delta_{x(t^+),y+1} \delta_{x(t^-),y}}{\int_0^T dt \ \delta_{x(t),y}} \equiv \frac{1}{\hat{\tau}_y} \tag{96}$$

and provide the L inferred trapping times $\hat{\tau}_y$. The corresponding diagonal matrix elements are given by Eq. 82

$$\hat{w}_{y,y} = -\hat{w}_{y+1,y} = -\frac{1}{\hat{\tau}_y} \tag{97}$$

In summary, the trajectory data have been used to compute the L parameters $\hat{\tau}_y$ that parametrize the best inferred Markov matrix

$$\hat{w}_{x,y} = \delta_{x,y+1} \frac{1}{\hat{\tau}_y} - \delta_{x,y} \frac{1}{\hat{\tau}_y} \tag{98}$$

C. Large deviations for the L inferred parameters $\hat{\tau}_y$

Now one wishes to know how the L inferred parameters $\hat{\tau}_y$ computed from the data can fluctuate with respect to the 'true' values τ_y of the 'true' model of Eq. 92. The inferred steady state $\hat{P}^*_x$ corresponds to the steady state associated to the model with the inferred parameters $\hat{\tau}_y$ and is thus given by the analog of Eq. 93

$$\hat{P}^*_x = \frac{\hat{\tau}_x}{\sum_{x'=1}^{L} \hat{\tau}_{x'}} \tag{99}$$

As a consequence, the joint probability to infer the L parameters $\hat{\tau}_.$ of a given disordered ring follows the large deviation form of Eq. 89

$$P_T^{Infer}(\hat{\tau}_.) \underset{T \to +\infty}{\simeq} e^{-T \mathcal{I}_{\tau_.}(\hat{\tau}_.)} \tag{100}$$

with the explicit rate function

$$\mathcal{I}_{\tau_.}(\hat{\tau}_.) = \sum_{x=1}^{L} \hat{P}^*_x \left[\frac{1}{\hat{\tau}_x} \ln\left(\frac{\tau_x}{\hat{\tau}_x}\right) - \frac{1}{\hat{\tau}_x} + \frac{1}{\tau_x} \right] = \frac{\sum_{x=1}^{L} \left[-\ln\left(\frac{\hat{\tau}_x}{\tau_x}\right) - 1 + \frac{\hat{\tau}_x}{\tau_x} \right]}{\sum_{x'=1}^{L} \hat{\tau}_{x'}} \tag{101}$$

As in Eq. 66, the main qualitative conclusion is that the L inferred parameters $\hat{\tau}_x$ are coupled via the inferred steady state $\hat{P}^*_.$ that they produce together.

VII. INFERENCE FOR DIFFUSION PROCESSES IN DIMENSION d WITH STEADY STATE

The inference for Langevin dynamics has been applied to many contexts [13–31]. In this section, the goal is to revisit this problem via the large deviations at Level 2.5 for diffusion processes [47, 48, 54, 58, 67–69].

A. Fokker-Planck generator parametrized by the force $\vec{F}(\vec{x})$ and the diffusion coefficient $D(\vec{x})$

The Fokker-Planck dynamics in the force field $\vec{F}(\vec{x})$ with the diffusion coefficient $D(\vec{x})$ in dimension d

$$\frac{\partial P_t(\vec{x})}{\partial t} = -\vec{\nabla}.\left[P_t(\vec{x})\vec{F}(\vec{x}) - D(\vec{x})\vec{\nabla} P_t(\vec{x}) \right] \equiv \mathcal{F} P_t(.) \tag{102}$$

corresponds to a conserved continuity equation involving the probability density $P_t(\vec{x})$ and the current

$$\vec{J}_t(\vec{x}) \equiv P_t(\vec{x})\vec{F}(\vec{x}) - D(\vec{x})\vec{\nabla} P_t(\vec{x}) \tag{103}$$

So here the Markov model M that one wishes to infer is the Fokker-Planck generator $\mathcal{F}$ parametrized by the force $\vec{F}(\vec{x})$ and the diffusion coefficient $D(\vec{x})$.

As in Eqs 33 and 69, we will assume that the steady-state solution $P^*(\vec{x})$ of Eq. 102

$$0 = \mathcal{F} P^*(.) = -\vec{\nabla}.\left[P^*(\vec{x})\vec{F}(\vec{x}) - D(\vec{x})\vec{\nabla} P^*(\vec{x}) \right] \tag{104}$$

exists. Again the steady state P^* corresponds to the right eigenvector of the Fokker-Planck generator $\mathcal{F}$ associated to the eigenvalue zero, while the corresponding left eigenvector is constant.

B. Empirical observables and their constraints

For the Fokker-Planck dynamics of Eq. 102, the relevant time-empirical observables are :

(i) the empirical density

$$\rho(\vec{x}) \quad \equiv \frac{1}{T} \int_0^T dt\ \delta^{(d)}(\vec{x}(t) - \vec{x}) \tag{105}$$

normalized to unity

$$\int d^d\vec{x}\ \rho(\vec{x}) \quad = 1 \tag{106}$$

(ii) the empirical current

$$\vec{j}(\vec{x}) \equiv \frac{1}{T} \int_0^T dt\ \frac{d\vec{x}(t)}{dt} \delta^{(d)}(\vec{x}(t) - \vec{x}) \tag{107}$$

that measures the time-average of the velocity $\frac{d\vec{x}(t)}{dt}$ when the position $\vec{x}(t)$ at the same time t is $\vec{x}$. This empirical current should be divergence-free

$$\vec{\nabla}.\vec{j}(\vec{x}) = 0 \tag{108}$$

in order to be consistent with stationarity.

C. Large deviations at Level 2.5 for the empirical density $\rho(.)$ and the empirical current $\vec{j}(.)$

The joint probability distribution of the normalized empirical density $\rho(.)$ and of the empirical divergence-less current $\vec{j}(.)$ satisfies the large deviation form [47, 48, 54, 58, 67–69]

$$P_T[\rho(.),\vec{j}(.)] \underset{T\to+\infty}{\simeq} \delta\left(\int d^d\vec{x}\rho(\vec{x})-1\right)\left[\prod_{\vec{x}}\delta\left(\vec{\nabla}.\vec{j}(\vec{x})\right)\right] e^{-TI_{2.5}\left[\rho(.);\vec{j}(.)\right]} \tag{109}$$

where the rate function

$$I_{2.5}\left[\rho(.);\vec{j}(.)\right] = \int \frac{d^d\vec{x}}{4D(\vec{x})\rho(\vec{x})}\left[\vec{j}(\vec{x})-\rho(\vec{x})\vec{F}(\vec{x})+D(\vec{x})\vec{\nabla}\rho(\vec{x})\right]^2 \tag{110}$$

vanishes only for the typical values corresponding to the steady-state of Eq. 104

$$\begin{aligned}\rho^{typ}(\vec{x}) &= P^*(\vec{x})\\ \vec{j}^{typ}(\vec{x}) &= P^*(\vec{x})\vec{F}(\vec{x})-D(\vec{x})\vec{\nabla}P^*(\vec{x})\end{aligned} \tag{111}$$

D. Inferred Fokker-Planck dynamics in the strict continuous-time limit

Reciprocally, from the empirical density $\rho(.)$ and the empirical current $\vec{j}(.)$ measured from the trajectory data, the best inferred steady state $\hat{P}^*(.)$ and the best inferred force $\hat{\vec{F}}(.)$ are given by

$$\begin{aligned}\hat{P}^*(\vec{x}) &= \rho(\vec{x})\\ \hat{\vec{F}}(\vec{x}) &= \frac{\vec{j}(\vec{x})+D(\vec{x})\vec{\nabla}\rho(\vec{x})}{\rho(\vec{x})}\end{aligned} \tag{112}$$

while the inferred diffusion coefficient $\hat{D}(\vec{x})$ has to coincide with the true diffusion coefficient $D(\vec{x})$

$$\hat{D}(\vec{x}) = D(\vec{x}) \tag{113}$$

in the strict continuous-time limit, as explained in detail in Appendix A from the path-integral point of view.

E. Large deviations for the inferred Fokker-Planck parameters in the strict continuous-time limit

Via the change of variables of Eq 112, Eqs 109 and 110 yields that the probability to infer the force $\hat{\vec{F}}(\vec{x})$ and the steady state $\hat{P}^*(.)$ of the corresponding Fokker-Planck generator $\hat{\mathcal{F}}$

$$0 = \hat{\mathcal{F}}\hat{P}^*(\vec{x}) = -\vec{\nabla}.\left[\hat{\vec{F}}(\vec{x})\hat{P}^*(\vec{x})-D(\vec{x})\vec{\nabla}\hat{P}^*(\vec{x})\right] \tag{114}$$

follows the large deviation form

$$P_T^{Infer}[\hat{P}^*(.),\hat{\vec{F}}(.)] \underset{T\to+\infty}{\simeq} \delta\left(\int d^d\vec{x}\hat{P}^*(\vec{x})-1\right)\left[\prod_{\vec{x}}\delta\left(\vec{\nabla}.\left[\hat{\vec{F}}(\vec{x})\hat{P}^*(\vec{x})-D(\vec{x})\vec{\nabla}\hat{P}^*(\vec{x})\right]\right)\right] e^{-T\mathcal{I}\left[\hat{P}^*(.),\hat{\vec{F}}(.)\right]} \tag{115}$$

with the rate function

$$\mathcal{I}\left[\hat{P}^*(.),\hat{\vec{F}}(.)\right] = \int d^d\vec{x}\hat{P}^*(\vec{x})\left[\frac{\left(\hat{\vec{F}}(\vec{x})-\vec{F}(\vec{x})\right)^2}{4D(\vec{x})}\right] \tag{116}$$

As in Eqs 66 and 101, the main qualitative conclusion is that the values $\hat{\vec{F}}(\vec{x})$ of the inferred force field are coupled via the inferred steady state $\hat{P}^*(.)$ that they produce together.

The impossibility to consider the fluctuations of the inferred diffusion coefficient (Eq. 113) might be somewhat surprising, but it is due to the strict continuous-time limit as explained in detail in Appendix A from the path-integral point of view. However in practice, the numerical inference of diffusion processes is based on discretized data. It is thus useful in the following subsections to re-analyze in detail the inference problem for diffusion processes from the point of view of discretized Langevin equations.

F. Equivalent Langevin dynamics with their discretized interpretations

When the diffusion coefficient $D(\vec{x})$ depends on the position $\vec{x}$, the Fokker-Planck dynamics of Eq. 102 corresponds to various Langevin stochastic differential equations involving d independent Gaussian white noise components $\mu = 1, .., d$

$$\begin{aligned} \langle \eta_\mu(t) \rangle &= 0 \\ \langle \eta_\mu(t) \eta_\nu(t') \rangle &= \delta_{\mu,\nu} \delta(t-t') \end{aligned} \tag{117}$$

as follows.

1. Equivalent Langevin dynamics within the Stratonovich interpretation

The Fokker-Planck dynamics of Eq. 102 is equivalent to the Stratonovich Langevin dynamics

$$\frac{d\vec{x}(t)}{dt} = \vec{f}_S(\vec{x}(t)) + \sqrt{2D(\vec{x}(t))}\ \vec{\eta}(t) \tag{118}$$

with the effective Stratonovich force

$$\vec{f}_S(\vec{x}) \equiv \vec{F}(\vec{x}) + \frac{1}{2}\vec{\nabla} D(\vec{x}) \tag{119}$$

Eq. 118 should be interpreted with the mid-point discretization scheme for the multiplicative factor of the noise

$$\vec{x}(t+\Delta t) = \vec{x}(t) + \vec{f}_I(\vec{x}(t))\Delta t + \sqrt{D(\vec{x}(t+\Delta t)) + D(\vec{x}(t))} \int_t^{t+\Delta t} dt' \vec{\eta}(t') \tag{120}$$

2. Equivalent Langevin dynamics within the Ito interpretation

The Fokker-Planck dynamics of Eq. 102 is equivalent to the Ito Langevin dynamics

$$\frac{d\vec{x}(t)}{dt} = \vec{f}_I(\vec{x}(t)) + \sqrt{2D(\vec{x}(t))}\ \vec{\eta}(t) \tag{121}$$

with the effective Ito force

$$\vec{f}_I(\vec{x}) \equiv \vec{F}(\vec{x}) + \vec{\nabla} D(\vec{x}) \tag{122}$$

Eq. 121 should be interpreted with the causal discretization scheme

$$\vec{x}(t+\Delta t) = \vec{x}(t) + \vec{f}_I(\vec{x}(t))\Delta t + \sqrt{2D(\vec{x}(t))} \int_t^{t+\Delta t} dt' \vec{\eta}(t') \tag{123}$$

i.e. the corresponding Gaussian propagator for the time-interval Δt reads

$$\begin{aligned} W^{[\Delta t]}(\vec{x}(t+\Delta t)|\vec{x}(t)) &= \left(\frac{1}{4\pi D(\vec{x}(t))\Delta t}\right)^{\frac{d}{2}} e^{-\frac{\left(\vec{x}(t+\Delta t) - \vec{x}(t) - \vec{f}_I(\vec{x}(t))\Delta t\right)^2}{4D(\vec{x}(t))\Delta t}} \\ &= \left(\frac{1}{4\pi D(\vec{x}(t))\Delta t}\right)^{\frac{d}{2}} e^{-\Delta t \frac{\left(\frac{\vec{x}(t+\Delta t) - \vec{x}(t)}{\Delta t} - \vec{f}_I(\vec{x}(t))\right)^2}{4D(\vec{x}(t))}} \end{aligned} \tag{124}$$

G. Inference for the Markov chain kernel corresponding to the Ito discretization of the Langevin equation

Eq. 124 corresponds to the discrete-time continuous-space Markov chain Gaussian kernel of parameters $[\vec{f}_I(.); D(.)]$

$$W^{[\Delta t]}_{[\vec{f}_I(.);D(.)]}(\vec{x}|\vec{y}) = \left(\frac{1}{4\pi D(\vec{y})\Delta t}\right)^{\frac{d}{2}} e^{-\frac{\left(\vec{x}-\vec{y}-\vec{f}_I(\vec{y})\Delta t\right)^2}{4D(\vec{y})\Delta t}} \tag{125}$$

while the time T corresponds to $N = \frac{T}{\Delta t}$ time steps. As a consequence, Eq. 53 yields that the probability to infer the generator $W^{[\Delta t]}_{[\hat{\vec{f}}_I(.),\hat{D}(.)]}$ together with its steady state $\hat{P}^*(.)$ follows the large deviation form

$$P_T^{Infer}(W_{[\hat{\vec{f}}_I(.),\hat{D}(.)]}(.,.);\hat{P}^*(.)) \underset{T\to+\infty}{\simeq} \delta\left(\int d^d\vec{x}\ \hat{P}^*(\vec{x}) - 1\right) \prod_{\vec{x}} \delta\left(\int d^d\vec{y}\ W_{[\hat{\vec{f}}_I(.),\hat{D}(.)]}(\vec{x},\vec{y})\hat{P}^*(\vec{y}) - \hat{P}^*(\vec{x})\right)$$
$$e^{-\frac{T}{\Delta t}\int d^d\vec{y}\ \hat{P}^*(\vec{y})\int d^d\vec{x}\ W_{[\hat{\vec{f}}_I(.),\hat{D}(.)]}(\vec{x},\vec{y})\ln\left(\frac{W_{[\hat{\vec{f}}_I(\ .),\hat{D}(.)]}(\vec{x},\vec{y})}{W_{[\vec{f}_I(.),D(.)]}(\vec{x},\vec{y})}\right)} \tag{126}$$

where the last constraint on the first line of Eq. 53 does not appear, since the conservation of probability is always satisfied by the generator $W^{[\Delta t]}_{[\hat{\vec{f}}_I(.),\hat{D}(.)]}$ (see the form of Eq. 125).

The last integral of the rate function in the exponential reads using the Gaussian kernel of Eq. 125

$$\int d^d\vec{x}\ W_{[\hat{\vec{f}}_I(.),\hat{D}(.)]}(\vec{x},\vec{y})\ln\left(\frac{W_{[\hat{\vec{f}}_I(\ .),\hat{D}(.)]}(\vec{x},\vec{y})}{W_{[\vec{f}_I(.),D(.)]}(\vec{x},\vec{y})}\right)$$
$$= \int d^d\vec{x}\ \left(\frac{1}{4\pi\hat{D}(\vec{y})\Delta t}\right)^{\frac{d}{2}} e^{-\frac{\left(\vec{x}-\vec{y}-\hat{\vec{f}}_I(\vec{y})\Delta t\right)^2}{4\hat{D}(\vec{y})\Delta t}} \left[\frac{d}{2}\ln\left(\frac{D(\vec{y})}{\hat{D}(\vec{y})}\right) + \frac{\left(\vec{x}-\vec{y}-\vec{f}_I(\vec{y})\Delta t\right)^2}{4D(\vec{y})\Delta t} - \frac{\left(\vec{x}-\vec{y}-\hat{\vec{f}}_I(\vec{y})\Delta t\right)^2}{4\hat{D}(\vec{y})\Delta t}\right]$$
$$= \frac{\left[\hat{\vec{f}}_I(\vec{y}) - \vec{f}_I(\vec{y})\right]^2}{4D(\vec{y})}\Delta t + \frac{d}{2}\left[-\ln\left(\frac{\hat{D}(\vec{y})}{D(\vec{y})}\right) + \frac{\hat{D}(\vec{y})}{D(\vec{y})} - 1\right] \tag{127}$$

so that Eq. 126 reduces to

$$P_T^{Infer}(W_{[\hat{\vec{f}}_I(.),\hat{D}(.)]}(.,.);\hat{P}^*(.)) \underset{T\to+\infty}{\simeq}$$
$$\delta\left(\int d^d\vec{x}\ \hat{P}^*(\vec{x}) - 1\right) \prod_{\vec{x}} \delta\left(\int d^d\vec{y}\ \left(\frac{1}{4\pi\hat{D}(\vec{y})\Delta t}\right)^{\frac{d}{2}} e^{-\frac{\left(\vec{x}-\vec{y}-\hat{\vec{f}}_I(\vec{y})\Delta t\right)^2}{4\hat{D}(\vec{y})\Delta t}} \hat{P}^*(\vec{y}) - \hat{P}^*(\vec{x})\right)$$
$$e^{-T\int d^d\vec{y}\ \hat{P}^*(\vec{y})\left(\frac{\left[\hat{\vec{f}}_I(\vec{y}) - \vec{f}_I(\vec{y})\right]^2}{4D(\vec{y})} + \frac{d}{2\Delta t}\left[-\ln\left(\frac{\hat{D}(\vec{y})}{D(\vec{y})}\right) + \frac{\hat{D}(\vec{y})}{D(\vec{y})} - 1\right]\right)} \tag{128}$$

In conclusion, it is important to distinguish whether the time step Δt remains finite or tends towards zero :

(i) if the time-step Δt of the Ito discretization scheme of Eq. 123 remains finite, then the probability to infer the Ito force $\vec{f}_I(.)$ and the diffusion coefficient $\hat{D}(.)$ follows the large deviation form of Eq. 128.

(ii) in the limit $\Delta t \to 0$, the last term of the rate function of Eq. 128 diverges if $\hat{D}(\vec{y}) \neq D(\vec{y})$. One recovers that the inferred diffusion coefficient $\hat{D}(\vec{x})$ has to coincide with the true diffusion coefficient $D(\vec{x})$ as in Eq. 113. Within the present Ito discretization computation, the origin of this property is the divergence of the number $N = \frac{T}{\Delta t} \to +\infty$ of time-steps in the limit $\Delta t \to 0$, so that the inferred diffusion $\hat{D}(\vec{x})$ coefficient cannot fluctuate any more but is fixed to its typical value given by the true diffusion coefficient.

VIII. APPLICATION TO THE DIFFUSION IN A DISORDERED POTENTIAL IN DIMENSION d

In this section, the large deviations analysis of inference for diffusion processes of the previous section is applied to the thermal equilibrium diffusion in a disordered potential $U(\vec{x})$ in dimension d.

A. Model and steady state

In this section, we consider the Fokker-Planck dynamics of Eq. 102 for the special case where the diffusion coefficient is uniform and fixed by the inverse temperature β (note that in the present paper, the notation T represents the time window of the trajectory and not the temperature, so that the temperature will only appear via its inverse β)

$$D(\vec{x}) = \frac{1}{\beta} \tag{129}$$

while the force $\vec{F}(\vec{x})$ derives from some disordered potential $U(\vec{x})$

$$\vec{F}(\vec{x}) = -\vec{\nabla} U(\vec{x}) \tag{130}$$

So the Fokker-Planck dynamics of Eq. 102 becomes

$$\frac{\partial P_t(\vec{x})}{\partial t} = -\vec{\nabla}.\left[P_t(\vec{x}) \left(-\vec{\nabla} U(\vec{x}) \right) - \frac{1}{\beta} \vec{\nabla} P_t(\vec{x}) \right] \tag{131}$$

The steady state of Eq. 104 corresponds to the Boltzmann equilibrium at inverse temperature β in the potential $U(\vec{x})$ on the appropriate domain $\vec{x} \in V$ for the model under study

$$P^*(\vec{x}) = \frac{e^{-\beta U(\vec{x})}}{\int_V d^d\vec{x}' e^{-\beta U(\vec{x}')}} \tag{132}$$

where the partition function of the denominator ensures the normalization of the steady state over the domain V.

B. Large deviations for the inference of the disordered potential $U(\vec{x})$

Eqs 115 and 116 yields that the probability to infer the potential $\hat{U}(\vec{x})$ instead of the true potential $U(\vec{x})$ follows the large deviation form with respect to the time-window T of the trajectory

$$P_T^{Infer}[\hat{U}(.)] \underset{T \to +\infty}{\simeq} e^{-T\mathcal{I}\left[\hat{U}(.)\right]} \tag{133}$$

with the explicit rate function

$$\mathcal{I}\left[\hat{U}(.)\right] = \left(\frac{\beta}{4}\right) \frac{\int_V d^d\vec{x} \left(\vec{\nabla}\hat{U}(\vec{x}) - \vec{\nabla} U(\vec{x}) \right)^2 e^{-\beta \hat{U}(\vec{x})}}{\int_V d^d\vec{x}' e^{-\beta \hat{U}(\vec{x}')}} \tag{134}$$

As in the other previous examples (see Eqs 66,101 and 116), the values $\hat{U}(\vec{x})$ of the inferred potential are coupled via the inferred steady state $\hat{P}^*(.)$ that they produce together, given here by the corresponding Boltzmann equilibrium

$$\hat{P}^*(\vec{x}) = \frac{e^{-\beta \hat{U}(\vec{x})}}{\int_V d^d\vec{x}' e^{-\beta \hat{U}(\vec{x}')}} \tag{135}$$

IX. CONCLUSION

In this paper, we have revisited the inference of Markov models via the large deviations properties at Level 2.5 within the following point of view :

(1) The input is a single dynamical trajectory $x(0 \leq t \leq T)$ over a very long time T

(2) From the trajectory data of (1), one computes the following empirical time-averaged observables (that should not depend too much on the time window T, otherwise this means that T is not large enough to see the convergence towards some steady state) :

(2a) the empirical time-averaged density ρ_x, given by Eq 40 in discrete-time, or by Eq. 74 in continuous-time

(2b) the empirical density of jumps, given by $\rho^{(2)}_{x,y}$ of Eq 37 in discrete-time, or by $q_{x,y}$ of Eq 76 in continuous-time

(3) The best Markov model that can be inferred from the empirical time-averaged observables of (2) via the principle of maximum likelihood is as follows :

(3a) the best inferred steady-state $\hat{P}^*$ is directly the empirical density ρ_x. Note that the possible states in the inferred model are given by the different states that are actually seen in the empirical density of the data in (2a). If some states of the true model have a too small probability in the steady state to be visited during the trajectory $x(0 \leq t \leq T)$, these states will not appear in the inferred model.

(3b) the best inferred Markov matrix, given by $\hat{W}_{x,y} = \frac{\rho^{(2)}_{x,y}}{\rho_y}$ of Eq. 44 in discrete-time, or by $\hat{w}_{x,y} \equiv \frac{q_{x,y}}{\rho_y}$ of Eq. 81 in continuous-time. Note that the possible jumps between states in the inferred model are given by the different jumps that are actually seen in the empirical jump-density of the data in (2b). If some possible jumps of the true model have a too small probability in the steady state to be visited during the trajectory $x(0 \leq t \leq T)$, these jumps will not appear in the inferred model.

4) The fluctuations of the best inferred Markov matrix of (3b) and of its corresponding steady state of (3a) with respect to the 'true' Markov matrix and the true steady state is given by the large deviations of Eq. 51 in discrete-time, or of Eq. 89 in continuous-time.

Applications to various models of random walks in disordered media have been described, where the goal was to infer the quenched disordered variables defining a given disordered sample. Given the recent availability of big data in many fields and the corresponding extensive use of inference methods, we hope that the present analysis can be useful to characterize the statistical fluctuations of the inferred parameters of Markov models with respect to the 'true' values of these parameters.

As a final remark, we should stress that for simplicity, we have only considered the case of time-independent Markov models with steady-states. However the Level 2.5 has been also formulated for other frameworks, so let us mention two possible natural extensions of the present framework :

(i) when the dynamics does not converge towards some steady state, or when the Markov matrix is time-dependent, one should use instead the formulation of Ref. [57], where the Level 2.5 concerns the ensemble-averaged observables at fixed time, i.e. one should consider that the input in (1) is not a single trajectory $x(0 \leq t \leq T)$ over a very long time T anymore, but instead a large number N of independent trajectories $x_n(0 \leq t \leq T)$ labelled by the index $n = 1, 2, .., N$ over the finite time T that one wishes to analyze for the inference problem.

(ii) when the Markov matrix is time-periodic, one should instead use the formulation of Ref. [64, 65], where the Level 2.5 concerns the empirical period-averaged observables for any given position inside the period.

Appendix A: Inference for diffusion processes via the path-integral approach

Since the property of Eq. 113 concerning the inference of the diffusion coefficient for the Fokker-Planck dynamics can be very surprising at first, it is useful in this Appendix to discuss in details the origin of this property from the point of view of the general principles of section II using path-integral methods.

1. Identification of the relevant time-empirical observables that determine the trajectories probabilities

For the Fokker-Planck dynamics of Eq. 102, the probability of the trajectory $x(0 \leq t \leq T)$

$$\mathcal{P}[\vec{x}(0 \leq t \leq T)] = e^{-\int_0^T dt \left[\frac{[\frac{d\vec{x}(t)}{dt} - \vec{F}(\vec{x}(t))]^2}{4D(\vec{x}(t))} - \frac{[\vec{\nabla} D(\vec{x}(t))]^2}{16 D(\vec{x}(t))} + \frac{\Delta D(\vec{x}(t))}{4} + \frac{\vec{\nabla}.\vec{F}(\vec{x}(t))}{2} \right]} \prod_t (D(\vec{x}(t)))^{-\frac{d}{2}} \quad \text{(A1)}$$

involves both the action in the exponential and the non-trivial measure-factor that depends on the diffusion coefficient $D(\vec{x}(t))$ along the trajectory. If one wishes to include this measure-factor in the action, one needs to introduce some regularization with some time-step Δt and the $N = \frac{T}{\Delta t}$ times $t_k = k\Delta t$ with $k = 1, 2, .., N$ to obtain

$$\left[\prod_t (D(\vec{x}(t)))^{-\frac{d}{2}} \right]^{Reg} = \prod_{k=1}^{\frac{T}{\Delta t}} [D(\vec{x}(t = k\Delta t))]^{-\frac{d}{2}} = e^{-\frac{d}{2} \sum_{k=1}^{\frac{T}{\Delta t}} \ln [D(\vec{x}(t = k\Delta t))]} \simeq e^{-\frac{d}{2\Delta t} \int_0^T dt \ln [D(\vec{x}(t))]} \quad \text{(A2)}$$

In the trajectory probability of Eq. A1, the relevant time-empirical observables that appear are thus not only the empirical density of Eq. 105 and the empirical current of Eq. 107 discussed in the main text, but also the empirical kinetic energy

$$k(\vec{x}) \equiv \frac{1}{T} \int_0^T dt \ \frac{1}{2} \left(\frac{d\vec{x}(t)}{dt} \right)^2 \delta^{(d)}(\vec{x}(t) - \vec{x}) \quad \text{(A3)}$$

that measures the time-average of the kinetic energy $\frac{1}{2} \left(\frac{d\vec{x}(t)}{dt} \right)^2$ when the position $\vec{x}(t)$ at the same time t is $\vec{x}$.

If one introduces the following notation for the normalized average of an arbitrary observable $\mathcal{O}$ at position $\vec{x}$

$$\langle \mathcal{O} \rangle_{\vec{x}} \equiv \frac{\frac{1}{T} \int_0^T dt \ \mathcal{O} \delta^{(d)}(\vec{x}(t) - \vec{x})}{\frac{1}{T} \int_0^T dt \ \delta^{(d)}(\vec{x}(t) - \vec{x})} = \frac{\frac{1}{T} \int_0^T dt \ \mathcal{O} \delta^{(d)}(\vec{x}(t) - \vec{x})}{\rho(\vec{x})} \quad \text{(A4)}$$

one obtains in terms of the empirical observables

$$\begin{aligned} \langle 1 \rangle_{\vec{x}} &= 1 \\ \langle \frac{d\vec{x}(t)}{dt} \rangle_{\vec{x}} &= \frac{\vec{j}(\vec{x})}{\rho(\vec{x})} \\ \langle \frac{1}{2} \left(\frac{d\vec{x}(t)}{dt} \right)^2 \rangle_{\vec{x}} &= \frac{k(\vec{x})}{\rho(\vec{x})} \end{aligned} \quad \text{(A5)}$$

As a consequence, the positivity of the variance of the velocity gives the following constraint for the empirical kinetic energy $k(\vec{x})$

$$\langle \left(\frac{d\vec{x}(t)}{dt} - \langle \frac{d\vec{x}(t)}{dt} \rangle_{\vec{x}} \right)^2 \rangle_{\vec{x}} = \langle \left(\frac{d\vec{x}(t)}{dt} \right)^2 \rangle_{\vec{x}} - \left(\langle \frac{d\vec{x}(t)}{dt} \rangle_{\vec{x}} \right)^2 = \frac{2k(\vec{x})}{\rho(\vec{x})} - \left(\frac{\vec{j}(\vec{x})}{\rho(\vec{x})} \right)^2 = \frac{2}{\rho(\vec{x})} \left[k(\vec{x}) - \frac{\vec{j}^2(\vec{x})}{2\rho(\vec{x})} \right] \geq 0 \quad \text{(A6)}$$

It is thus more convenient to replace the empirical kinetic energy $k(\vec{x})$ by the empirical positive excess of kinetic energy

$$e(\vec{x}) \equiv \frac{k(\vec{x})}{\rho(\vec{x})} - \frac{1}{2} \left(\frac{\vec{j}(\vec{x})}{\rho(\vec{x})} \right)^2 \geq 0 \quad \text{(A7)}$$

i.e. the empirical kinetic energy $k(\vec{x})$ is decomposed into the two contributions

$$k(\vec{x}) = \frac{\vec{j}^2(\vec{x})}{2\rho(\vec{x})} + \rho(\vec{x}) e(\vec{x}) \quad \text{(A8)}$$

In terms of these empirical observables, the probability of the trajectory $x(0 \leq t \leq T)$ of Eq. A1 with the regularized form of Eq. A2 can be rewritten as

$$\begin{aligned}
&\mathcal{P}[\vec{x}(0 \leq t \leq T)] \\
&= e^{-T \int d^d\vec{x} \left(\frac{k(\vec{x})}{2D(\vec{x})} - \frac{\vec{j}(\vec{x}).\vec{F}(\vec{x})}{2D(\vec{x})} + \rho(\vec{x}) \left[\frac{[\vec{F}(\vec{x})]^2}{4D(\vec{x})} - \frac{[\vec{\nabla} D(\vec{x})]^2}{16 D(\vec{x})} + \frac{\Delta D(\vec{x})}{4} + \frac{\vec{\nabla}.\vec{F}(\vec{x})}{2} \right] \right)}_{e} - \frac{dT}{2\Delta t} \int d^d\vec{x} \rho(\vec{x}) \ln[D(\vec{x})] \\
&= e^{-T \int d^d\vec{x} \rho(\vec{x}) \left[\frac{\left(\frac{\vec{j}(\vec{x})}{\rho(\vec{x})} - \vec{F}(\vec{x}) \right)^2}{4D(\vec{x})} + \frac{e(\vec{x})}{2D(\vec{x})} - \frac{[\vec{\nabla} D(\vec{x})]^2}{16 D(\vec{x})} + \frac{\Delta D(\vec{x})}{4} + \frac{\vec{\nabla}.\vec{F}(\vec{x})}{2} + \frac{d}{2\Delta t} \ln[D(\vec{x})] \right]}
\end{aligned} \tag{A9}$$

With respect to the general formalism summarized in Section II, this means that the relevant empirical observables E are the empirical density $\rho(.)$, the empirical current $\vec{j}(.)$ and the empirical excess of kinetic energy $e(.)$, while the intensive action introduced in Eq. 1 reads

$$\begin{aligned}
&A_{\vec{F}(.);D(.)} \left([\rho(.); \vec{j}(.); e(.)] \right) = \\
&\int d^d\vec{x} \rho(\vec{x}) \left[\frac{\left(\frac{\vec{j}(\vec{x})}{\rho(\vec{x})} - \vec{F}(\vec{x}) \right)^2}{4D(\vec{x})} + \frac{e(\vec{x})}{2D(\vec{x})} - \frac{[\vec{\nabla} D(\vec{x})]^2}{16 D(\vec{x})} + \frac{\Delta D(\vec{x})}{4} - \frac{\vec{F}(\vec{x}).\vec{\nabla}\rho(\vec{x})}{2\rho(\vec{x})} + \frac{d}{2\Delta t} \ln[D(\vec{x})] \right]
\end{aligned} \tag{A10}$$

The penultimate term has been rewritten via an integration by parts with respect to Eq. A9.

2. Typical values of the empirical observables

The typical values for the empirical density and for the current have been given in Eq. 111 of the text, while the typical value of the empirical excess of kinetic energy $e(.)$ is given by the diffusion coefficient

$$e^{typ}(\vec{x}) = D(x) \tag{A11}$$

Reciprocally, the modified Fokker-Planck operator $\hat{\mathcal{F}}$ that would make these three empirical observables $[\rho(.); \vec{j}(.); e(.)]$ typical is parametrized by the modified diffusion coefficient

$$\hat{D}(\vec{x}) = e(\vec{x}) \tag{A12}$$

and by the modified force

$$\hat{\vec{F}}(\vec{x}) = \frac{\vec{j}(\vec{x}) + e(\vec{x}) \vec{\nabla}\rho(\vec{x})}{\rho(\vec{x})} \tag{A13}$$

With respect to the general formalism summarized in Section II, this means that the intensive action of Eq. A10 reads for the modified Fokker-Planck operator $\hat{\mathcal{F}}$ with the parameters given by Eqs A12 and A13

$$\begin{aligned}
&A_{\hat{\vec{F}}(.);\hat{D}(.)} \left([\rho(.); \vec{j}(.); e(.)] \right) \\
&= \int d^d\vec{x} \rho(\vec{x}) \left[\frac{\left(\frac{\vec{j}(\vec{x})}{\rho(\vec{x})} - \hat{\vec{F}}(\vec{x}) \right)^2}{4\hat{D}(\vec{x})} + \frac{e(\vec{x})}{2\hat{D}(\vec{x})} - \frac{[\vec{\nabla}\hat{D}(\vec{x})]^2}{16\hat{D}(\vec{x})} + \frac{\Delta\hat{D}(\vec{x})}{4} - \frac{\hat{\vec{F}}(\vec{x}).\vec{\nabla}\rho(\vec{x})}{2\rho(\vec{x})} + \frac{d}{2\Delta t} \ln\left[\hat{D}(\vec{x})\right] \right] \\
&= \int d^d\vec{x} \rho(\vec{x}) \left[\frac{e(\vec{x}) \left(\frac{\vec{\nabla}\rho(\vec{x})}{\rho(\vec{x})} \right)^2}{4} + \frac{1}{2} - \frac{[\vec{\nabla} e(\vec{x})]^2}{16 e(\vec{x})} + \frac{\Delta e(\vec{x})}{4} - \frac{\left(\frac{\vec{j}(\vec{x}) + e(\vec{x})\vec{\nabla}\rho(\vec{x})}{\rho(\vec{x})} \right).\vec{\nabla}\rho(\vec{x})}{2\rho(\vec{x})} + \frac{d}{2\Delta t} \ln[e(\vec{x})] \right]
\end{aligned} \tag{A14}$$

3. Large deviations for the relevant time-empirical observables

So the joint distribution of the empirical density $\rho(.)$,the empirical current $\vec{j}(.)$ and the empirical excess of kinetic energy $e(.)$ satisfy the large deviation form

$$P_T[\rho(.),\vec{j}(.);e(.)]\underset{T\to+\infty}{\simeq}\delta\left(\int d^d\vec{x}\rho(\vec{x})-1\right)\left[\prod_{\vec{x}}\delta\left(\vec{\nabla}.\vec{j}(\vec{x})\right)\right]e^{-TI_{2.75}\left[\rho(.);\vec{j}(.);e(.)\right]} \tag{A15}$$

where the rate function $I_{2.75}\left[\rho(.);\vec{j}(.);e(.)\right]$ of Level 2.75 (this name just means that it is higher than the usual 2.5 Level discussed below in Eq. A18) corresponds to the difference (Eq. 11) between the actions of Eqs A10 and A14

$$\begin{aligned}
&I_{2.75}\left[\rho(.);\vec{j}(.);e(.)\right]=A_{\vec{F}(.);D(.)}\left([\rho(.);\vec{j}(.);e(.)]\right)-A_{\hat{\vec{F}}(.);\hat{D}(.)}\left([\rho(.);\vec{j}(.);e(.)]\right)\\
&=\int d^d\vec{x}\rho(\vec{x})\left[\frac{\left(\frac{\vec{j}(\vec{x})}{\rho(\vec{x})}-\vec{F}(\vec{x})\right)^2}{4D(\vec{x})}+\frac{e(\vec{x})}{2D(\vec{x})}-\frac{[\vec{\nabla}D(\vec{x})]^2}{16D(\vec{x})}+\frac{\Delta D(\vec{x})}{4}-\frac{\vec{F}(\vec{x}).\vec{\nabla}\rho(\vec{x})}{2\rho(\vec{x})}\right]\\
&-\int d^d\vec{x}\rho(\vec{x})\left[\frac{e(\vec{x})\left(\frac{\vec{\nabla}\rho(\vec{x})}{\rho(\vec{x})}\right)^2}{4}+\frac{1}{2}-\frac{[\vec{\nabla}e(\vec{x})]^2}{16e(\vec{x})}+\frac{\Delta e(\vec{x})}{4}-\frac{\left(\frac{\vec{j}(\vec{x})+e(\vec{x})\vec{\nabla}\rho(\vec{x})}{\rho(\vec{x})}\right).\vec{\nabla}\rho(\vec{x})}{2\rho(\vec{x})}\right]\\
&+\frac{d}{2\Delta t}\int d^d\vec{x}\rho(\vec{x})\ln\left[\frac{D(\vec{x})}{e(\vec{x})}\right]
\end{aligned} \tag{A16}$$

where we have written separately on the last line the regularized contribution involving the time-step Δt (see Eq. A2). In the limit $\Delta t\to 0$, this contribution becomes singular unless the empirical excess of kinetic energy $e(.)$ coincides with the diffusion coefficient $D(.)$

$$e(\vec{x})=D(\vec{x}) \tag{A17}$$

This means that the empirical excess of kinetic energy $e(.)$ is actually not allowed to fluctuate but is fixed by Eq. A17 to its typical value.

As a consequence, in the strict continuous-time limit $\Delta t\to 0$, the only empirical observables that can fluctuate are the empirical density $\rho(.)$ and the empirical current $\vec{j}(.)$ as mentioned in the text, and their large deviations properties are described by Eq. 109, where the rate function of Eq. 110 is given by Eq. A15 for the case $e(\vec{x})=D(\vec{x})$ of Eq. A17

$$\begin{aligned}
&I_{2.5}\left[\rho(.);\vec{j}(.)\right]=I_{2.75}\left[\rho(.);\vec{j}(.);e(.)=D(.)\right]\\
&=\int d^d\vec{x}\rho(\vec{x})\left[\frac{\left(\vec{F}(\vec{x})-\frac{\vec{j}(\vec{x})}{\rho(\vec{x})}\right)^2-D^2(\vec{x})\left(\frac{\vec{\nabla}\rho(\vec{x})}{\rho(\vec{x})}\right)^2}{4D(\vec{x})}-\frac{\left(\vec{F}(\vec{x})-\frac{\vec{j}(\vec{x})+D(\vec{x})\vec{\nabla}\rho(\vec{x})}{\rho(\vec{x})}\right).\vec{\nabla}\rho(\vec{x})}{2\rho(\vec{x})}\right]\\
&=\int d^d\vec{x}\rho(\vec{x})\left[\frac{\left(\vec{F}(\vec{x})-\frac{\vec{j}(\vec{x})+D(\vec{x})\vec{\nabla}\rho(\vec{x})}{\rho(\vec{x})}\right)^2}{4D(\vec{x})}\right]
\end{aligned} \tag{A18}$$

4. Consequence for the inferred diffusion coefficient

So the inferred diffusion coefficient $\hat{D}(\vec{x})$ has to coincide with the true diffusion coefficient $D(\vec{x})$ in the strict continuous-time limit $\Delta t\to 0$ (see Eqs A12 and A17)

$$\hat{D}(\vec{x})=e(\vec{x})=D(\vec{x}) \tag{A19}$$

Within the present path-integral approach, the origin of this property is the non-trivial measure-factor of Eq. A2 in the trajectory probability of Eq. A1 that would give an infinite result in the last contribution of the rate function of

Eq. A16 for any other diffusion coefficient $\hat{D}(\vec{x}) \neq D(\vec{x})$ than the true diffusion coefficient $D(\vec{x})$. Another perspective of the property of Eq. A19 is given in the text after Eq. 128.

---

[1] T. M. Cover and J. A. Thomas, Elements of Information Theory, J. Wiley, New Jersey (2006).
[2] L. Wasserman, All of statistics: a concise course in statistical inference, Springer (2020)
[3] D. J. C. MacKay, Information theory, inference and learning algorithms, Cambridge University Press (2003)
[4] A. Gomez-Corral, D. Rios Insua, F. Ruggeri and M. Wiper, Wiley Statistics Reference Online 2015 https://doi.org/10.1002/9781118445112.stat07837
[5] T. W. Anderson and L. A. Goodman, The Annals of Mathematical Statistics, 28, 89 (1957).
[6] P. Billingsley, Ann. Math. Statist. 32, 12 (1961).
[7] A. Dembo and O. Zeitouni, Large Deviations Techniques and Applications, Springer Berlin (2010).
[8] M. A. Suchard, R. E. Weiss, and J. S. Sinsheimer, Mol. Biol. Evol. 18, 1001 (2001).
[9] V. Baldazzi, S. Cocco, E. Marinari, and R. Monasson, Phys. Rev. Lett. 96, 128102 (2006).
[10] V. Baldazzi, S. Bradde, S. Cocco, E. Marinari, and R. Monasson, Phys. Rev. E 75, 011904 (2007).
[11] S. Cocco and R. Monasson, Euro. Phys. Lett, 81 20002 (2008).
[12] S. Cocco and R. Monasson, Journal of Physics: Conference Series 197, 012005 (2009).
[13] F. Boettcher, J. Peinke, D. Kleinhans, R. Friedrich, P. G. Lind, M. Haase, Phys. Rev. Lett. 97, 090603 (2006).
[14] J.B. Masson, D. Casanova, S. Turkcan, G. Voisinne, M. Popoff, M. Vergassola and A. Alexandrou, Phys. Rev. Lett 102, 048103 (2009).
[15] G. Voisinne, A. Alexandrou and J. B. Masson, Biophysical Journal 98, 596 (2010).
[16] S. Turkcan, A. Alexandrou and J.B. Masson, Biophysical Journal 102, 2288 (2012).
[17] S. Turkcan, J. B. Masson, D. Casanova, G. Mialon, T. Gacoin, J. P. Boilot, M. R. Popoff and A. Alexandrou, Biophysical Journal 102, 2299 (2012).
[18] S. Turkcan and J. B. Masson, PLOS ONE 8(12): e82799 (2013).
[19] M. U. Richly, S. Turkcan, A. Le Gall, N. Fiszman, J. B. Masson, N. Westbrook, K. Perronet and A. Alexandrou, Optics Express 21, 31578 (2013).
[20] S. Turkcan S, M. U. Richly, A. Alexandrou, J. B. Masson , PLOS ONE 8(1): e53073 (2013).
[21] J. B. Masson, P. Dionne, C. Salvatico, M. Renner, C.G. Specht, A. Triller, and M. Dahan, Biophysical Journal 106, 74 (2014).
[22] C. L. Vestergaard, P. C. Blainey and H. Flyvbjerg, Phys. Rev E 89, 022726 (2014).
[23] M. El Beheiry, S. Turkcan, M. U. Richly, A.Triller, A. Alexandrou, M. Dahan and J. B. Masson, Biophysical Journal 110, 1209 (2016).
[24] S. Thapa, M. A. Lomholt, J. Krog, A. G. Cherstvy and R. Metzler, Phys. Chem. Chem. Phys. 20, 29018 (2018).
[25] A. G. Cherstvy, S. Thapa, C. E. Wagner and R. Metzler, Soft Matter 15(12), 2526 (2019).
[26] F. S. Gnesotto, G. Gradziuk, P. Ronceray, C. P. Broedersz, Nature Communications 11, 5378 (2020).
[27] A. Frishman and P. Ronceray, Phys. Rev. X 10, 021009 (2020).
[28] D. B Bruckner, P. Ronceray, C. P Broedersz Phys. Rev. Lett. 125, 058103 (2020).
[29] F. Ferretti, V. Chardes, T. Mora, A. M. Walczak, I. Giardina, Phys. Rev. X 10, 031018 (2020).
[30] G. Munoz-Gil, M. A. Garcia-March, C. Manzo, J. D. Martin-Guerrero and M. Lewenstein, New J. Phys. 22, 013010 (2020).
[31] D. B. Bruckner, N. Arlt, A. Fink, P. Ronceray, J. O. Radler, C. P. Broedersz, arXiv:2008.03978.
[32] A. Cavagna, I. Giardina, F. Ginelli, T. Mora, D. Piovani, R. Tavarone, A. M. Walczak, Phys. Rev. E 89 042707 (2014).
[33] G. Junot, E. Clement, H. Auradou, R. Garcia-Garcia, arXiv:2012.04528.
[34] R. Supekar, B. Song, A. Hastewell, A. Mietke, J. Dunkel , arXiv:2101.06568.
[35] B. Derrida, J. Stat. Mech. P07023 (2007).
[36] R. J. Harris and G. M. Schütz, J. Stat. Mech. P07020 (2007).
[37] E. M. Sevick, R. Prabhakar, S. R. Williams, D. J. Searles, Ann. Rev. of Phys. Chem. Vol 59, 603 (2008).
[38] 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.
[39] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio and C. Landim Rev. Mod. Phys. 87, 593 (2015).
[40] R. L. Jack, P. Sollich, The European Physical Journal Special Topics 224, 2351 (2015).
[41] A. Lazarescu, J. Phys. A: Math. Theor. 48 503001 (2015).
[42] A. Lazarescu, J. Phys. A: Math. Theor. 50 254004 (2017).
[43] R. L. Jack, Eur. Phy. J. B 93, 74 (2020).
[44] A. de La Fortelle, PhD (2000) "Contributions to the theory of large deviations and applications" INRIA Rocquencourt.
[45] V. Lecomte, PhD Thesis (2007) "Thermodynamique des histoires et fluctuations hors d'équilibre" Université Paris 7.
[46] R. Chétrite, PhD Thesis 2008 "Grandes déviations et relations de fluctuation dans certains modèles de systèmes hors d'équilibre", ENS Lyon.
[47] B. Wynants, arXiv:1011.4210, PhD Thesis (2010), "Structures of Nonequilibrium Fluctuations", Catholic University of Leuven.

[48] R. Chétrite, HDR Thesis (2018) "Pérégrinations sur les phénomènes aléatoires dans la nature", Laboratoire J. A. Dieudonné, Université de Nice.
[49] Y. Oono, Progress of Theoretical Physics Supplement 99, 165 (1989).
[50] R.S. Ellis, Physica D 133, 106 (1999).
[51] H. Touchette, Phys. Rep. 478, 1 (2009);
H. Touchette, Modern Computational Science 11: Lecture Notes from the 3rd International Oldenburg Summer School, BIS-Verlag der Carl von Ossietzky Universitat Oldenburg, 2011.
[52] G. Fayolle and A. de La Fortelle, Problems of Information Transmission 38, 354 (2002).
[53] C. Monthus, Eur. Phys. J. B 92, 149 (2019) in the topical issue " Recent Advances in the Theory of Disordered Systems" edited by F. Igloi and H. Rieger.
[54] C. Monthus, J. Stat. Mech. (2021) 033201.
[55] A. de La Fortelle, Problems of Information Transmission 37 , 120 (2001).
[56] C. Maes and K. Netocny, Europhys. Lett. 82, 30003 (2008).
[57] C. Maes, K. Netocny and B. Wynants, Markov Proc. Rel. Fields. 14, 445 (2008).
[58] A. C. Barato and R. Chétrite, J. Stat. Phys. 160, 1154 (2015).
[59] L. Bertini, A. Faggionato and D. Gabrielli, Ann. Inst. Henri Poincare Prob. and Stat. 51, 867 (2015).
[60] L. Bertini, A. Faggionato and D. Gabrielli, Stoch. Process. Appli. 125, 2786 (2015).
[61] C. Monthus, J. Stat. Mech. (2019) 023206.
[62] C. Monthus, J. Phys. A: Math. Theor. 52, 135003 (2019).
[63] C. Monthus, J. Phys. A: Math. Theor. 52, 025001 (2019).
[64] A. C. Barato, R. Chétrite, J. Stat. Mech. (2018) 053207.
[65] L. Chabane, R. Chétrite, G. Verley, J. Stat. Mech. (2020) 033208.
[66] C. Monthus, arXiv:2103.08885 and arXiv:2104.10392.
[67] C. Maes, K. Netocny and B. Wynants, Physica A 387, 2675 (2008).
[68] J. Hoppenau, D. Nickelsen and A. Engel, New J. Phys. 18 083010 (2016).
[69] C. Monthus, J. Stat. Mech. (2021) 033303.
[70] J. W. Haus et K. W. Kehr, Phys. Rep. 150, 263 (1987).
[71] J. P. Bouchaud and A. Georges, Phys. Rep. 195, 127 (1990).
[72] J. P. Bouchaud, A. Comtet, A. Georges and P. Le Doussal, Ann. Phys. 201, 285 (1990).
[73] D. Ben-Avraham and S. Havlin, "Diffusion and reactions in fractals and disordered systems" Cambridge University Press (2000) .
[74] C. Monthus, Lett. Math. Phys. 78, 207 (2006).
[75] B. Derrida and Y. Pomeau, Phys. Rev. Lett. 48 , 627 (1982).
[76] B. Derrida, J. Stat. Phys. 31, 433 (1983).
[77] H. Kesten, Acta Math. 131, 208 (1973); H. Kesten et al. , Compositio Math 30, 145 (1975).
[78] F. Solomon, Ann. Proba. 1,31 (1975).
[79] Y. G. Sinai, Theo. Prob. and Appl. 27, 256 (1982).
[80] B. Derrida and H. Hilhorst, J. Phys. A 16, 2641 (1983).
[81] C. de Callan, J. M. Luck, Th. Nieuwenhuizen and D. Petritis, J. Phys. A 18, 501 (1985).
[82] F. Igloi and C. Monthus, Phys. Rep. 412, 277 (2005).
[83] C. Monthus, Phys. Rev. B 69, 054431 (2004).
[84] C. Monthus, J. Stat. Mech. P06036 (2015).
[85] C. Monthus, J. Stat. Mech. 123304 (2017).
[86] T. Gautié, J.P. Bouchaud and P. Le Doussal, arXiv:2101.08082.
[87] J. P. Bouchaud, A. Georges and P. Le Doussal, J. Physique 48, 1855 (1987).
[88] C. Aslangul, M. Barthelemy, N. Pottier and D. Saint-James, J. Stat. Phys. 59, 11 (1990).
[89] A. Compte and J. P. Bouchaud, J. Phys. A. 31, 6113 (1998).
[90] C. Monthus, Phys. Rev. E 67, 046109 (2003)
[91] K. Van Duijvendijk, G. Schehr and F. Van Wijland, Phys. Rev. E 78, 011120 (2008).
[92] C. Monthus, arXiv:2102.10834.
[93] J.P. Bouchaud, J. Phys. I (France) 2 , 1705 (1992).
[94] J. P. Bouchaud and D. Dean, J. Phys. I (France) 5, 265 (1995).
[95] C. Monthus and J. P. Bouchaud, J. Phys. A 29, 3847 (1996).
[96] E. M. Bertin and J. P. Bouchaud, Phys. Rev. E 67, 026128 (2003).
[97] E. M. Bertin and J. P. Bouchaud, Phys. Rev. E 67, 065105(R) (2003).
[98] C. Monthus, Phys. Rev. E 68, 036114 (2003).
[99] C. Monthus, J. Phys. A 36, 11605 (2003).
[100] C. Monthus, Phys. Rev. E 69, 026103 (2004).
[101] M. Ueda and S. Sasa, J. Phys. A: Math. Theor. 50, 125001 (2017).