Large deviation theory applied to study rare and extreme events in turbulence, atmosphere, and climate dynamics - Lecture I

F. BOUCHET – CNRS and ENS de Lyon

Boulder summer school 2022 – Hydrodynamics Across Scales – 07/2022

ENS DE LYON

SIMONS FOUNDATION

Statistical Mechanics and Large Deviation Theory

The mathematical tools of statistical mechanics.

1. Probability theory.
2. Large deviation theory.
3. Averaging and stochastic averaging (reduction methods).
4. Dynamical system theory and ergodic theory.
5. Linear response theory.

Large Deviation Theory

- Large deviation theory is a general framework to describe probability distributions in asymptotic limits

$$P(x,\varepsilon) = \mathbb{P}[X_\varepsilon = x] \underset{\varepsilon \ll 1}{\asymp} e^{-\frac{\mathscr{F}[x]}{\varepsilon}}.$$

For equilibrium statistical mechanics, $\mathscr{F}$ is the free energy, and $\varepsilon = k_B T/N$.

Maths: Cramer 30', Sanov 50', Lanford 70', Freidlin–Wentzell 70' and 80', Varadhan, ... In parallel with theoretical physicists.

Outline

I) Introduction to large deviation theory and its applications to dynamical problems (Wednesday)
II) Large deviation theory for kinetic theories, geostrophic turbulence, and atmosphere dynamics (Thursday)
III) Rare and extreme events in climate dynamics: sampling using rare event algorithms and machine learning (Friday)

Outline

I) Introduction to large deviation theory and its applications to dynamical problems (Wednesday)
II) Large deviation theory for kinetic theories, geostrophic turbulence, and atmosphere dynamics (Thursday)
III) Rare and extreme events in climate dynamics: sampling using rare event algorithms and machine learning (Friday)

Outline for Lecture I.

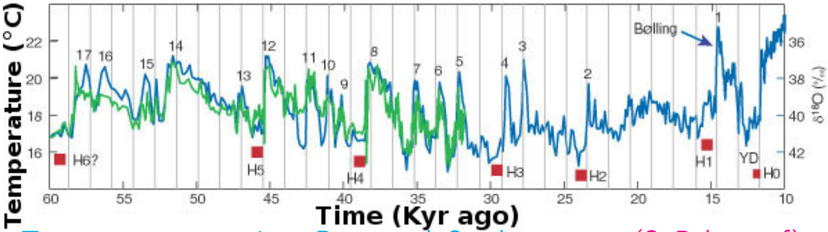
1. Rare events in complex dynamical systems
   - Abrupt climate changes and transitions between turbulent attractors
   - Rare events with a huge impact: extreme heat waves
   - Rare and extreme events in astronomy
2. Introduction to large deviation theory
   - What are large deviations?
   - Large deviations for the sum of N i.i.d. random variables
   - Large deviations for the empirical distribution
3. Dynamical large deviations (large deviations for path measures)
   - Large deviations for stochastic differential equations with weak noises
   - General dynamical large deviations
   - Quasipotential, Hamilton–Jacobi equation, relaxation and fluctuation paths

Outline

1. Rare events in complex dynamical systems
   - Abrupt climate changes and transitions between turbulent attractors
   - Rare events with a huge impact: extreme heat waves
   - Rare and extreme events in astronomy
2. Introduction to large deviation theory
   - What are large deviations?
   - Large deviations for the sum of N i.i.d. random variables
   - Large deviations for the empirical distribution
3. Dynamical large deviations (large deviations for path measures)
   - Large deviations for stochastic differential equations with weak noises
   - General dynamical large deviations
   - Quasipotential, Hamilton–Jacobi equation, relaxation and fluctuation paths

Abrupt Climate Changes (Last Glacial Period)

Long times matter

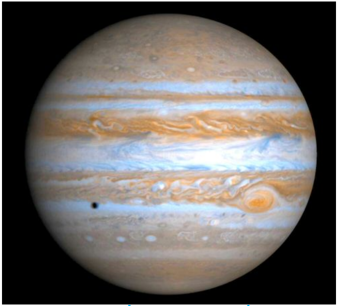


Temperature versus time: Dansgaard–Oeschger events (S. Rahmstorf)

- What is the dynamics and probability of abrupt climate changes?

Jupiter's Zonal Jets

An example of a geophysical turbulent flow (Coriolis force, huge Reynolds number, ...)

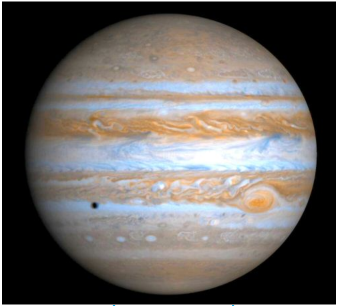


Jupiter's troposphere

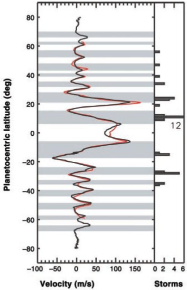
Jupiter's motions (Voyager)

Jupiter's Zonal Jets

We look for a theoretical description of zonal jets



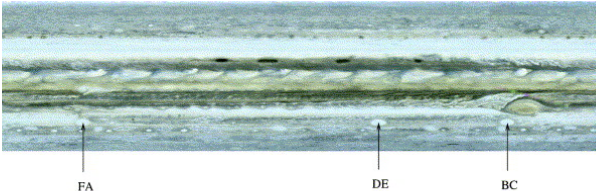
Jupiter's troposphere

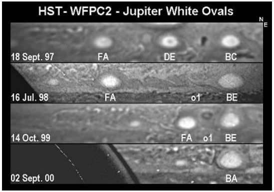


Jupiter's zonal winds (Voyager and Cassini, from Porco et al 2003)

Jupiter's Abrupt Climate Change

Have we lost one of Jupiter's jets ?



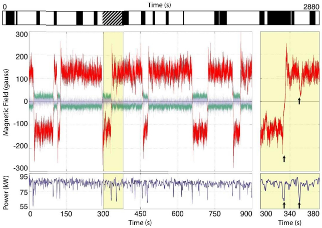


Jupiter's white ovals (see Youssef and Marcus 2005)

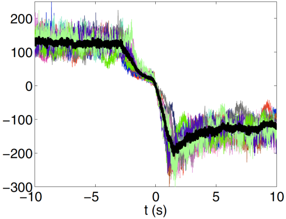
The white ovals appeared in 1939-1940 (Rogers 1995). Following an instability of one of the zonal jets?

Random Transitions in Turbulence Problems

Magnetic Field Reversal (Turbulent Dynamo, MHD Dynamics)



Magnetic field timeseries

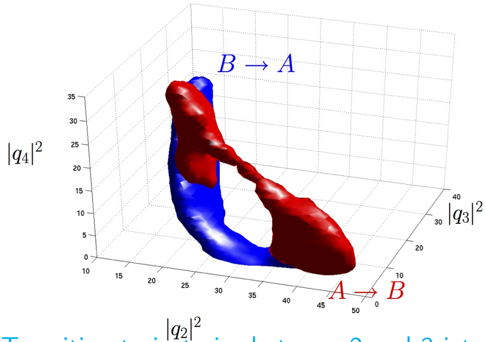


Zoom on transition paths

(VKS experiment)

In turbulent flows, transitions from one attractor to another (reactive paths) often occur through a predictable path called instanton.

Atmosphere Jet "Instantons"



Transition trajectories between 2 and 3 jet states

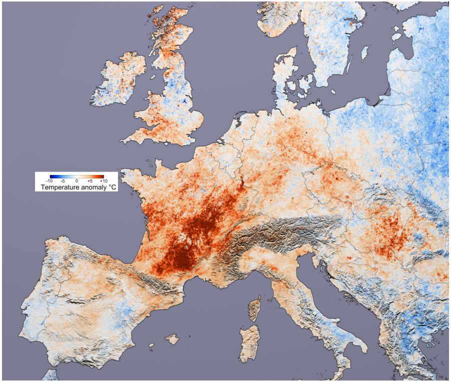
- The dynamics of turbulent transitions is predictable.
- Asymmetry between forward and backward transitions.

F. Bouchet, J. Rolland and E. Simonnet, 2019, PRL.

Outline

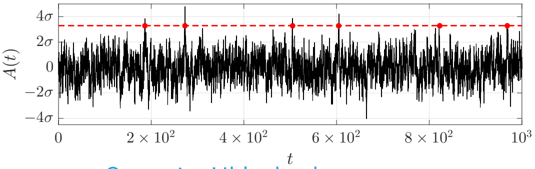
Extreme Heat Waves

Example: the 2003 heat wave over western Europe

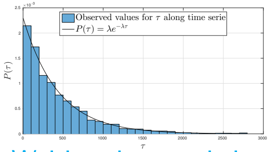


July 20 2003-August 20 2003 land surface temperature minus the average for the same period for years 2001, 2002 and 2004 (TERRA MODIS).

Extreme Events, Poisson Statistics, and Return Times



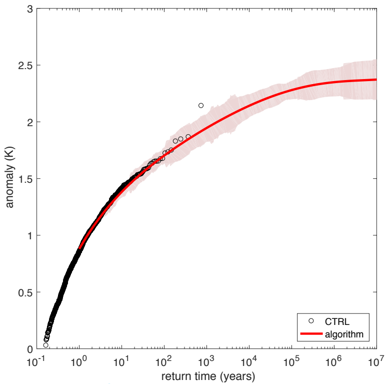
Ornstein–Uhlenbeck extremes



Waiting time statistics

For systems with a single state, rare enough events are uncorrelated and have a Poisson statistics

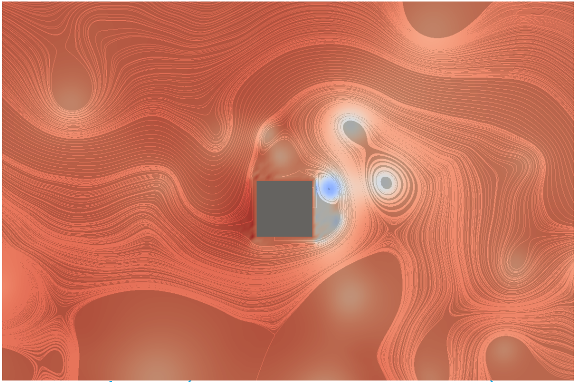
The Return Time of Extreme Heat Waves



Return time of 90 day European heat waves

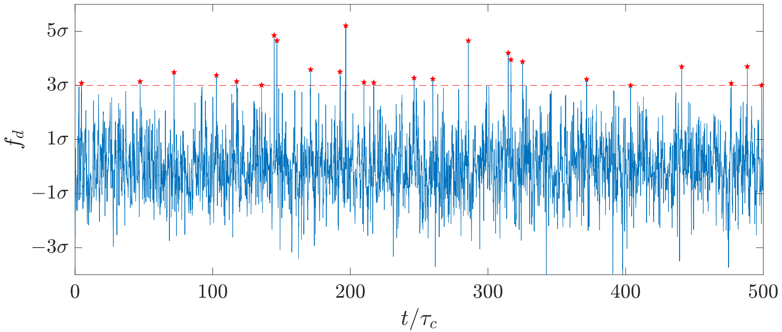
F. Ragone, J. Wouters, and F. Bouchet, PNAS, 2018

Lattice Boltzmann Simulations of 2D Flows



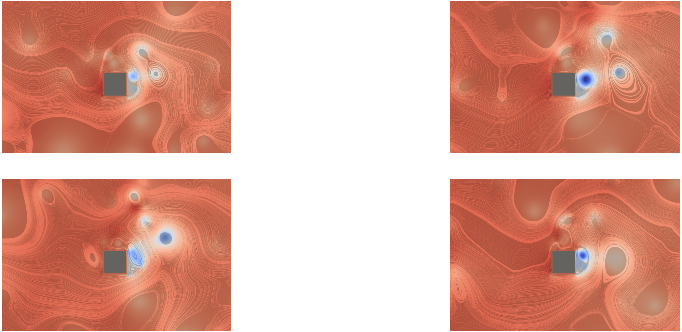
Velocity ($Re_G = 1\,200,\ Re_O \simeq 500$)

Extreme Drags for 2D Grid Turbulence Flows



Velocity ($Re_G = 1\,200$, $Re_O \simeq 500$) (with T. Lestang)

Predictability of Extreme Patterns for Turbulent Flows



Streamlines and pressure for the four events with the most extreme drag

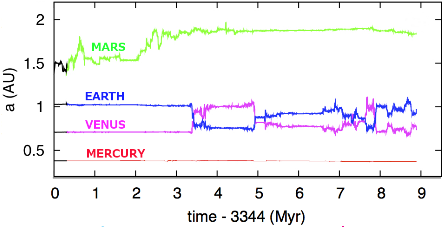
- Extreme event patterns are often predictable. Instanton?

Outline

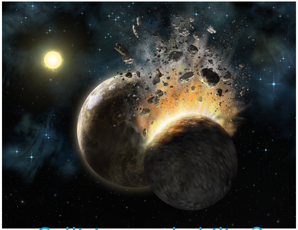
1. Rare events in complex dynamical systems
   - Abrupt climate changes and transitions between turbulent attractors
   - Rare events with a huge impact: extreme heat waves
   - Rare and extreme events in astronomy
2. Introduction to large deviation theory
   - What are large deviations?
   - Large deviations for the sum of N i.i.d. random variables
   - Large deviations for the empirical distribution
3. Dynamical large deviations (large deviations for path measures)
   - Large deviations for stochastic differential equations with weak noises
   - General dynamical large deviations
   - Quasipotential, Hamilton–Jacobi equation, relaxation and fluctuation paths

Collisional Trajectories in the Solar System

In collaboration with J. Laskar.



Distance from the sun vs time (J. Laskar) (7.10^6 hours of CPU, p=1/100 000)

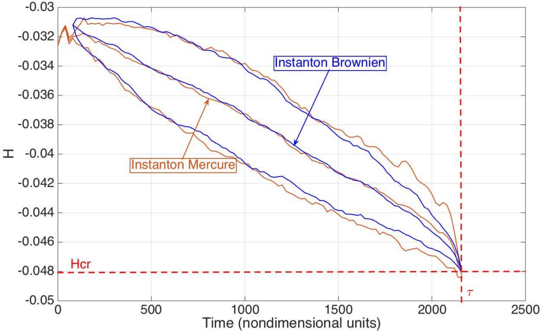


Collision probability?

- What are the probabilities of past and future qualitative changes of the solar system?

Instanton for Mercury–Jupiter Resonance

For the Batygin–Morbidelli–Holman model



Instanton for sample paths of Mercury–Jupiter resonance

E. Woillez and F. Bouchet, 2020, PRL and Nature Reviews Physics

Rare Events in Complex Dynamics

The scientific questions:

- What is the probability and the dynamics of those rare events?
- Is the dynamics leading to such rare events predictable?
- How to sample rare events, their probability, and their dynamics?
- Are direct numerical simulations a reasonable approach?
- Can we devise new theoretical and numerical tools to tackle these issues?

Outline

Large Deviation Theory

- Large deviation theory is a general framework to describe probability distributions in asymptotic limits

$$P(x,\varepsilon) = \mathbb{P}\left[X_\varepsilon = x\right] \underset{\varepsilon \ll 1}{\asymp} \mathrm{e}^{-\frac{\mathscr{F}[x]}{\varepsilon}}.$$

For equilibrium statistical mechanics, $\mathscr{F}$ is the free energy, and $\varepsilon = k_B T/N$.

Maths: Cramer 30', Sanov 50', Lanford 70', Freidlin–Wentzell 70' and 80', Varadhan, ... In parallel with theoretical physicists.

Log-Equivalence

- Log-equivalence:

$$f(x,\varepsilon) \underset{\varepsilon \ll 1}{\asymp} e^{-\frac{\mathscr{F}[x]}{\varepsilon}} \Leftrightarrow \lim_{\varepsilon \to 0} \varepsilon \log f(x,\varepsilon) = -\mathscr{F}[x]$$

- Example $f(x,\varepsilon) = C\varepsilon^{\alpha} e^{-\frac{\mathscr{F}[x]}{\varepsilon}}$. $C\varepsilon^{\alpha}$ is called a prefactor. Then $f(x,\varepsilon) \underset{\varepsilon \ll 1}{\asymp} e^{-\frac{\mathscr{F}[x]}{\varepsilon}}$. The prefactor is not considered.
- While large deviations rate functions are usually rather simple and have a generic structure, prefactors can be tricky to compute and depend on plenty of non-universal details.

Example: Laplace Integrals in dimension 1

- Asymptotics of Laplace integrals:

$$\int_a^b g(x)e^{-\frac{f(x)}{\varepsilon}}\mathrm{d}x \underset{\varepsilon\to 0}{\sim} g(x^*)\sqrt{\frac{2\pi\varepsilon}{f''(x^*)}}e^{-\frac{1}{\varepsilon}f(x^*)} \text{ with } f(x*) = \inf_{x\in(a,b)} f(x)$$

and $x* \in (a,b)$ or

$$\int_a^b g(x)e^{-\frac{f(x)}{\varepsilon}}\mathrm{d}x \underset{\varepsilon\to 0}{\sim} \varepsilon\frac{g(a)}{f'(a)}e^{-\frac{1}{\varepsilon}f(a)} \text{ if } f(a) = \inf_{x\in[a,b]} f(x)$$

- Log-equivalence:

$$\int_a^b g(x)e^{-\frac{1}{\varepsilon}f(x)}\mathrm{d}x \underset{\varepsilon\ll 1}{\asymp} e^{-\frac{f(x*)}{\varepsilon}} \text{ with } f(x*) = \inf_{x\in[a,b]} f(x)$$

Laplace Principle

- If $h(x)$ is bounded and continuous

$$\int g(x)e^{-\frac{1}{\varepsilon}h(x)}e^{-\frac{1}{\varepsilon}I(x)}\mathrm{d}x \underset{\varepsilon\ll 1}{\asymp} \mathrm{e}^{-\frac{\inf_x\{h(x)+I(x)\}}{\varepsilon}}$$

- Can we generalize this type of computations to families of probability measures, with sets of hypothesis which are natural for probability measures?

Large Deviation Principle

1. Let (X_ε) be a family of real random variables with probability measures μ_ε, then we will say that the sequence of measures μ_ε (or the family of variables (X_ε) with $\mathrm{d}\mu_\varepsilon = \mathbb{P}_\varepsilon(x)\mathrm{d}x$) satisfies the large deviation principle with rate function $I(x)$ if those two inequalities are satisfied:

$$\limsup_{\varepsilon\to 0} \quad \varepsilon \ln \mu_\varepsilon(C) \leq -\inf\{I(x)|x\in C\} \quad \textit{for every close set } C$$

$$\liminf_{\varepsilon\to 0} \quad \varepsilon \ln \mu_\varepsilon(O) \geq -\inf\{I(x)|x\in O\} \quad \textit{for every open set } O$$

2. The large deviation principle is equivalent to the statement that for every bounded continuous function h we have

$$\lim_{\varepsilon\to 0} \varepsilon \ln \int e^{-\frac{h(x)}{\varepsilon}} \mathbb{P}_\varepsilon(x)\mathrm{d}x = -\inf_{x\in\mathbb{R}} \{h(x)+I(x)\}.$$

Outline

1. Rare events in complex dynamical systems
   - Abrupt climate changes and transitions between turbulent attractors
   - Rare events with a huge impact: extreme heat waves
   - Rare and extreme events in astronomy
2. Introduction to large deviation theory
   - What are large deviations?
   - Large deviations for the sum of N i.i.d. random variables
   - Large deviations for the empirical distribution
3. Dynamical large deviations (large deviations for path measures)
   - Large deviations for stochastic differential equations with weak noises
   - General dynamical large deviations
   - Quasipotential, Hamilton–Jacobi equation, relaxation and fluctuation paths

Sum of Independent and Identically Distributed Random Variables and the Law of Large Numbers

- Let $(X_1,...,X_N)$ be independent and identically distributed random variables, with probability distribution function P_0.
- We assume that, for P_0, the average μ and the standard deviation σ exist.
- We consider the sum

$$S_N = \frac{1}{N}\sum_{n=1}^{N} X_n.$$

- Law of large numbers: $\lim_{N\to\infty} S_N = \mu$.

The Central Limit Theorem

- Let $(X_1,...,X_N)$ be independent and identically distributed random variables, with probability distribution function P_0.
- We assume that, for P_0, the average μ and the standard deviation σ exist.
- We consider the sum

$$Z_N = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (X_n - \mu).$$

- Central limit theorem: the probability measure of Z_N converges to a centered Gaussian distribution of variance σ.

Beyond Gaussian Fluctuations, Large Deviations: Cramer's Theorem

- Let $(X_1,...,X_N)$ be independent and identically distributed random variables, with probability distribution function P_0.
- We consider the sum

$$S_N = \frac{1}{N}\sum_{n=1}^{N} X_n.$$

- Large deviation principle (Cramer's theorem)

$$P_N(s) = \mathbb{P}(S_N = s) \underset{N\to\infty}{\asymp} \mathrm{e}^{-NI(s)},$$

$$\text{with } I(s) = \sup_k \left\{ks - \lambda(k)\right\},$$

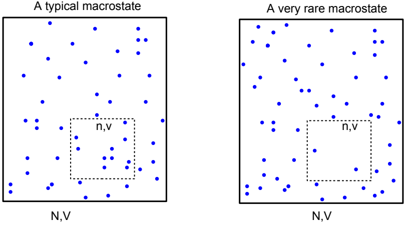
$$\text{and } \lambda(k) = \log \mathbb{E}\left(e^{kX_1}\right) = \log\left(\int P_0(x)\mathrm{e}^{kx}\,\mathrm{d}x\right).$$

Asymptotic Results for Random Variables

- In many problems in physics and mathematics, we often deal with the triptic:

1. Law of large numbers.
2. Central limit theorem.
3. Large deviation principle.

Example 1: Density Fluctuations for a Perfect Gas



Two possible microscopic states.

- The particles are statistically independent. The probability to be inside the box is $p = v/V$, the probability to be outside is $(1-p)$.
- The number of particle inside the box follows a binomial distribution with parameter p.

Example 1: Density Fluctuations for a Perfect Gas

- $x_N = n/N$ ratio of particles in the small box.
- Enumerating all possible cases, we get the binomial distribution

$$P_N(n) = \begin{pmatrix} N \\ n \end{pmatrix} p^n (1-p)^{N-n} \text{ or } P_N(x) = \begin{pmatrix} N \\ xN \end{pmatrix} p^{xN} (1-p)^{N(1-x)}.$$

- Law of large numbers: $\lim_{N\to\infty} x_N = p = v/V$.
- Central limit theorem: the distribution converge to a Gaussian distribution with a variance that scale like $1/\sqrt{N}$.
- We use the Stirling approximation

$$\log P_N(x) = -N \left[x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{1-p} \right] + O(\ln N).$$

Large Deviations for the Binomial Distribution

Convergence of the distribution to the rate function

$$\log P_N(x) = -N\left[x\ln\frac{x}{p} + (1-x)\ln\frac{1-x}{1-p}\right] + O(\ln N).$$

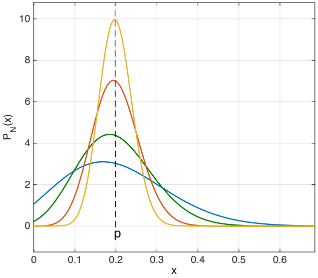
$$I_N(x) = -\frac{1}{N}\ln \mathbb{P}_N(x) \underset{N\to\infty}{\longrightarrow} I_p(x)$$

$$\text{with } \; I_p(x) = x\ln\frac{x}{p} + (1-x)\ln\frac{1-x}{1-p}.$$

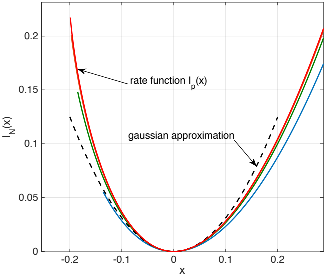
- We recognize an expression similar to the volume part of the entropy in thermodynamics.

Large Deviations for the Binomial Distribution

Convergence of the distribution of the sum of N i.i.d. random variables, distributed according to the binomial distribution ($p = 0.2$), to the rate function



Distribution of the sum of N samples ($N = 5, 20, 50, 100$)



Convergence to the large deviation rate

$$I_p(x) = x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p}.$$

The Scaled Cumulant Generating Function

- Definition of a cumulant generating function:

$$I(k) = \log\left(\int e^{kx} P(x)\,dx\right) = \log \mathbb{E}\left(e^{kx}\right).$$

- If you expect a large deviation principle for the variable X_ε, $P_\varepsilon(x) \underset{\varepsilon \ll 1}{\asymp} e^{-\frac{\mathscr{F}[x]}{\varepsilon}}$, you should adopt a scaled version of the cumulant generating function:

$$\lambda(k) = \lim_{\varepsilon\to 0} \varepsilon \log \int e^{\frac{kx}{\varepsilon}} P_\varepsilon(x)\,dx = \lim_{\varepsilon\to 0} \varepsilon \log \mathbb{E}\left(e^{\frac{kx}{\varepsilon}}\right),$$

or equivalently

$$\mathbb{E}\left(e^{\frac{kX_\varepsilon}{\varepsilon}}\right) \underset{\varepsilon\to 0}{\asymp} e^{\frac{\lambda(k)}{\varepsilon}}.$$

- λ is the scaled cumulant generating function.

Scaled Cumulant Generating Function and Legendre–Fenchel Transform

- Using Laplace's principle

$$\mathbb{E}\left(e^{\frac{kx}{\varepsilon}}\right) = \int e^{\frac{kx}{\varepsilon}} P_\varepsilon(x)\,dx \underset{\varepsilon\to 0}{\asymp} \int e^{\frac{1}{\varepsilon}\{kx - I(x)\}}\,dx \underset{\varepsilon\to 0}{\asymp} e^{\frac{\sup_x\{kx-I(x)\}}{\varepsilon}}.$$

- As λ is defined by $\mathbb{E}\left(e^{\frac{kX_\varepsilon}{\varepsilon}}\right) \underset{\varepsilon\to 0}{\asymp} e^{\frac{\lambda(k)}{\varepsilon}}$, we have

$$\lambda(k) = \sup_x \{kx - I(x)\}.$$

- The scaled cumulant generating function is the Legendre–Fenchel transform of the large deviation rate function.

Computing the Large Deviation Rate Function from the Scaled Cumulant Generating Function

- Basics in convex analysis: If I is convex then the Legendre–Fenchel transform is invertible and

$$I(x) = \sup_k \{kx - \lambda(k)\}.$$

- Gärtner–Ellis theorem: If λ is differentiable, then I exists and

$$I(x) = \sup_k \{kx - \lambda(k)\}.$$

- Whenever λ is differentiable, the large deviation rate function is the Legendre–Fenchel transform of the scaled cumulant generating function.

Proof of Cramer's Theorem

- Let $(X_1,...,X_N)$ be independent and identically distributed random variables, with probability distribution function P_0.
- We consider the sum

$$S_N = \frac{1}{N}\sum_{n=1}^{N} X_n$$

- Computing the scaled cumulant generating function:

$$\frac{1}{N}\log \mathbb{E}\left(e^{NkS_N}\right) = \frac{1}{N}\log \mathbb{E}\left(e^{k\sum_{n=1}^{N} X_n}\right) = \frac{1}{N}\log\left[\mathbb{E}\left(e^{kX_1}\right)\right]^N = \log \mathbb{E}\left(e^{kX_1}\right)$$

- Then $\lambda(k) = \log \mathbb{E}\left(e^{kX_1}\right) = \log\left(\int P_0(x)e^{kx}\,dx\right)$. It is easily proven that λ is differentiable and convex.
- Then we obtain the large deviation principle (Cramer's theorem):

$$P_N(s) = \mathbb{P}(S_N = s) \underset{N\to\infty}{\asymp} e^{-NI(s)} \text{ with } I(s) = \sup_k \left\{ks - \lambda(k)\right\}.$$

Outline

Sanov's Theorem: Large Deviations for the Empirical Distribution

- We now look at the empirical distribution

$$\rho_N(x) = \frac{1}{N}\sum_{n=1}^{N} \delta\left(x - X_n\right).$$

- We consider the probability distribution functional of ρ_n:

$$P_N\left[\rho\right] \equiv \mathbb{E}\left[\delta\left(\rho - \rho_N\right)\right].$$

- Sanov's theorem is a statement about the asymptotic behavior of $P_N\left[\rho\right]$. It states that

$$\ln P_N\left[\rho\right] \underset{N\to\infty}{\sim} -N\int \rho(x)\ln\left(\frac{\rho(x)}{P_0(x)}\right)\mathrm{d}x,$$

if $\int \rho\,\mathrm{d}x = 1$ and $-\infty$ otherwise.

Sanov's Theorem for Variables with Discrete Values

- When the variable x takes only K possible values $\{\sigma_k\}_{1\leq k\leq K}$ with probability $\{\pi_k\}_{1\leq k\leq K}$

$$P_0(x)=\sum_{k=1}^{K}\pi_k\delta(x-\sigma_k),$$

with $\sum_{k=1}^{K}\pi_k=1$, then $\rho_N(z)=\sum_{k=1}^{K}\rho_{k,N}\delta(z-\sigma_k)$.

- PDF of $(\rho_{1,N},...,\rho_{K,N})$:

$$P_N(p_1,...,p_K)\equiv\mathbb{E}[\delta(\rho_{1,N}-p_1,...,\rho_{K,N}-p_K)].$$

- Then Sanov's theorem states that

$$\ln P_N(p_1,...,p_K)\underset{N\to\infty}{\sim}-N\sum_{k=1}^{K}p_k\ln\left(\frac{p_k}{\pi_k}\right),$$

if $\sum_{k=1}^{K}p_k=1$ and $-\infty$ otherwise.

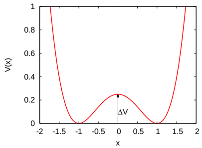
Outline

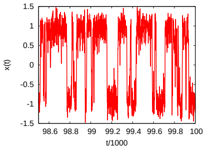
1. Rare events in complex dynamical systems
   - Abrupt climate changes and transitions between turbulent attractors
   - Rare events with a huge impact: extreme heat waves
   - Rare and extreme events in astronomy
2. Introduction to large deviation theory
   - What are large deviations?
   - Large deviations for the sum of N i.i.d. random variables
   - Large deviations for the empirical distribution
3. Dynamical large deviations (large deviations for path measures)
   - Large deviations for stochastic differential equations with weak noises
   - General dynamical large deviations
   - Quasipotential, Hamilton–Jacobi equation, relaxation and fluctuation paths

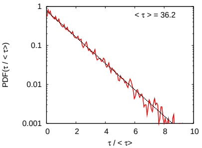
Kramers' Problem: a Pedagogical Example for Bistability

Historical example: Computation by Kramer of Arrhenius' law for a bistable mechanical system with stochastic noise

$$\frac{dx}{dt} = -\frac{dV}{dx}(x) + \sqrt{2k_B T_e}\eta(t) \ \text{Rate}: \ \lambda = \frac{1}{\tau}\exp\left(-\frac{\Delta V}{k_B T_e}\right).$$







The problem was solved by Kramer (30'). Modern approach: path integral formulation (instanton theory, physicists) or large deviation theory (Freidlin–Wentzell, mathematicians).

Kramers' Problem: Kinetic Approach

$$\frac{dx}{dt} = -\frac{dV}{dx}(x) + \sqrt{2k_B T_e}\eta(t). \text{ Rate}: \ \lambda = \frac{1}{\tau}\exp\left(-\frac{\Delta V}{k_B T_e}\right).$$

- P_{-1} is the probability for the particle to be in the basin of attraction of x_{-1}.
- Time scales $\tau_i = \left(\frac{d^2V}{dx^2}(x_i)\right)^{-1}$. If $\lambda \ll 1/\max_i(\tau_i)$, we expect a sequence of uncorrelated jumps (Markovian).
- Then if $k_B T_e \ll \Delta V$, we have for $t \gg \max_i(\tau_i)$, the kinetic eq.

$$\frac{dP_{-1}}{dt} = \lambda(1-P_{-1}) - \lambda P_{-1} = 1 - 2\lambda P_{-1}.$$

- The transition probabilities. $P(x_{-1},T;x_1,0)$ is the solution $P_{-1}(T)$ with initial conditions $P_1(0)=1$.

$$P(x_{-1},T;x_1,0) = \frac{1}{2}\left(1-e^{-2\lambda T}\right).$$

$$P(x_{-1},T;x_1,0) \underset{\max_i(\tau_i)\ll T \ll 1/\lambda}{\simeq} \lambda T = \frac{T}{\tau}\exp\left(-\frac{\Delta V}{k_B T_e}\right).$$

Gaussian White Noise

- What is a Gaussian white noise?
- We consider a Gaussian vector $\eta = \{\eta_i\}_{0\le i\le I}$ with zero mean $\mathbb{E}(\eta_i) = 0$ and covariance $\mathbb{E}(\eta_i\eta_j) = \delta_{ij}$. Its PDF is

$$P(\eta) = \frac{1}{(2\pi)^{I/2}} \mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{I}\eta_i^2}$$

- A Gaussian stochastic process $\eta(t)$ with correlation function

$$\langle \eta(t)\eta(t')\rangle = \delta(t-t').$$

has a Probability Density Functional

$$P_{WN}[\eta] \propto \mathrm{e}^{-\frac{1}{2}\int_0^T \eta^2(t)\,\mathrm{d}t}.$$

Probability Measure Over Paths

- What is the probability for the path $\{x(t)\}_{0\leq t\leq T}$, solution of

$$\frac{dx}{dt}=-\frac{dV}{dx}(x)+\sqrt{2k_BT_e}\eta(t).$$

- We start from the white noise probability

$$P_{WN}[\eta]=\mathrm{e}^{-\frac{1}{2}\int_0^T\eta^2(t)\mathrm{d}t}\mathscr{D}[\eta].$$

- We make a change of variables in order to get the probability for a path $\{x(t)\}_{0\leq t\leq T}$. It is

$$P_P[x]=e^{-\frac{1}{4k_BT_e}\int_0^T\left[\dot{x}+\frac{dV}{dx}(x)\right]^2(t)\mathrm{d}t}J[\eta\,|x]\mathscr{D}[x],$$

where $J[\eta\,|x]$ is the Jacobian of the change of variables.
- If we assume Ito convention, then $J[\eta\,|x]=1$, and

$$P_P[x]=e^{-\frac{1}{4k_BT_e}\int_0^T\left[\dot{x}+\frac{dV}{dx}(x)\right]^2(t)\mathrm{d}t}\mathscr{D}[x].$$

Path Integrals for ODE – Onsager Machlup (50')

- Path integral representation of transition probabilities:

$$P(x_T, T; x_0, 0) = \int_{x(0)=x_0}^{x(T)=x_T} e^{-\frac{\mathscr{A}_T[x]}{2k_B T_e}} \mathscr{D}[x]$$

$$\text{with } \mathscr{A}_T[x] = \int_0^T \mathscr{L}[x, \dot{x}]\ \mathrm{d}t \text{ and } \mathscr{L}[x, \dot{x}] = \frac{1}{2}\left[\dot{x} + \frac{\mathrm{d}V}{\mathrm{d}x}(x)\right]^2.$$

- The most probable path from x_0 to x_T is the minimizer of

$$A_T(x_0, x_T) = \min_{\{x(t)\}} \left\{ \mathscr{A}_T[x] \,|\, x(0) = x_0 \text{ and } x(T) = x_T \right\}.$$

- We may consider the low temperature limit, using a saddle point approximation (WKB), Then we obtain the large deviation result

$$\log P(x_T, T; x_0, 0) \underset{\frac{k_B T}{\Delta V} \to 0}{\asymp} -\frac{A_T(x_0, x_T)}{2k_B T_e}.$$

Relaxation Paths Minimize the Action

$$\mathscr{A}_T[x] = \int_0^T \mathscr{L}[x,\dot{x}]\,\mathrm{d}t \text{ and } \mathscr{L}[x,\dot{x}] = \frac{1}{2}\left[\dot{x} + \frac{\mathrm{d}V}{\mathrm{d}x}(x)\right]^2.$$

- A relaxation path $\{x_r(t)\}_{0\leq t\leq T}$ is a solution of

$$\dot{x} = -\frac{\mathrm{d}V}{\mathrm{d}x}.$$

  Then we see that

$$\mathscr{A}_T[x_r] = 0.$$

- Interpretation: if one follows the deterministic dynamics, no noise is needed and the cost is zero.
- Because for any path $\mathscr{A}_T[x_r] \geq 0$, any relaxation path minimizes the action.

Most Transition Paths Follow the Instanton

- In the weak noise limit, most transition paths follow the most probable path (instanton)

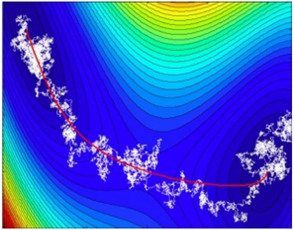


Figure by Eric Van den Eijnden

- For gradient dynamics, instantons are time reversed relaxation paths from a saddle to an attractor. Arrhenius law then follows

$$P(x_1, T; x_{-1}, 0) \underset{k_B T_e \to 0}{\asymp} \mathrm{e}^{-\frac{\Delta V}{k_B T_e}}.$$

Outline

1. Rare events in complex dynamical systems
   - Abrupt climate changes and transitions between turbulent attractors
   - Rare events with a huge impact: extreme heat waves
   - Rare and extreme events in astronomy
2. Introduction to large deviation theory
   - What are large deviations?
   - Large deviations for the sum of N i.i.d. random variables
   - Large deviations for the empirical distribution
3. Dynamical large deviations (large deviations for path measures)
   - Large deviations for stochastic differential equations with weak noises
   - General dynamical large deviations
   - Quasipotential, Hamilton–Jacobi equation, relaxation and fluctuation paths

Large Deviations for the Probability of the Newton Ratio and Path Large Deviations

$$\frac{dX_\varepsilon}{dt} = b(X_\varepsilon) + \sqrt{2\varepsilon}\eta(t).$$

- We compute the probability of the Newton ratio

$$\mathbb{P}\left(\frac{X_\varepsilon(t+\Delta t)-x}{\Delta t} = \dot{x} \middle| X_\varepsilon(t) = x\right) = \mathbb{E}_x\left[\delta\left(\frac{X_\varepsilon(\Delta t)-x}{\Delta t} - \dot{x}\right)\right].$$

- It is the effect of the deterministic part plus the sum of an infinite number of small amplitude Gaussian contributions

$$\mathbb{E}_x\left[\delta\left(\frac{X_\varepsilon(\Delta t)-x}{\Delta t} - \dot{x}\right)\right] \underset{\varepsilon\downarrow 0}{\asymp} \exp\left(-\frac{\Delta t L(x,\dot{x})}{\varepsilon}\right) \text{ with } L(x,\dot{x}) = \frac{1}{4}(\dot{x}-b(x))^2.$$

- Then using Bayes rule, markovianity and a continuous time limit

$$\mathbb{P}\left[\{X_\varepsilon(t)\}_{0\le t<T} = \{x(t)\}_{0\le t<T}\right] \underset{\varepsilon\downarrow 0}{\asymp} \exp\left[-\frac{1}{\varepsilon}\int_0^T \mathrm{d}t\, L(x,\dot{x})\right].$$

Large Deviations Produced by a Large Number of Small Amplitude Independent Moves

- In many processes in physics, chemistry, economy, or science in general, dynamical increments are produced by the sum of a large number of small amplitude independent contributions. However, those do not need to be Gaussian. We need a generalization of dynamical large deviations for those cases.
- We will have to compute the probability of the Newton ratio and prove the result

$$\mathbb{E}_x\left[\delta\left(\frac{X_\varepsilon(\Delta t)-x}{\Delta t}-\dot{x}\right)\right] \underset{\varepsilon\downarrow 0}{\asymp} \exp\left(-\frac{\Delta t L(x,\dot{x})}{\varepsilon}\right).$$

- Then using Bayes rule, markovianity and a continuous time limit

$$\mathbb{P}\left[\{X_\varepsilon(t)\}_{0\leq t<T}=\{x(t)\}_{0\leq t<T}\right] \underset{\varepsilon\downarrow 0}{\asymp} \exp\left[-\frac{1}{\varepsilon}\int_0^T \mathrm{d}t\, L(x,\dot{x})\right].$$

Infinitesimal Generator of a Continuous Time Markov Process

- We consider continuous time Markov processes $\{X(t)\}_{0\leq t<\infty}$.
- Definition of the infinitesimal generator. ϕ is a test function:

$$G[\phi](x) = \lim_{\Delta t\downarrow 0} \frac{\mathbb{E}_x[\phi(X(\Delta t))] - \phi(x)}{\Delta t}.$$

- The infinitesimal generator quantifies the Newton ratio for any observable. It thus generalize the notion of a time derivative.
- It fully characterizes the Markov process.
- For example, for a diffusion $dX = b(X)dt + \sqrt{2}dW_t$, the infinitesimal generator is $G[\phi](x) = b(x)\nabla\phi + \Delta\phi$, the adjoint of the Fokker–Planck equation (sometimes called the backward Fokker-Planck equation).

Large Deviations Produced by a Large Number of Small Amplitude Independent Moves

- Continuous time Markov processes $\{X_\varepsilon(t)\}_{0\le t<\infty}$, with infinitesimal generator G_ε.
- We assume that for all $p\in\mathbb{R}^n$ the limit

$$H(x,p)=\lim_{\varepsilon\downarrow 0}\varepsilon G_\varepsilon\left[\mathrm{e}^{\frac{p.x}{\varepsilon}}\right]\mathrm{e}^{-\frac{p.x}{\varepsilon}}.$$

- Then the family X_ε verifies a large deviation principle

$$P\left[\{X_\varepsilon(t)\}_{0\le t<T}=\{X(t)\}_{0\le t<T}\right]\underset{\varepsilon\downarrow 0}{\asymp}\exp\left(-\frac{\int_0^T\mathrm{d}t\,L\left(X,\dot{X}\right)}{\varepsilon}\right),$$

$$\text{with } L(x,\dot{x})=\sup_p\{p\dot{x}-H(x,p)\}.$$

(see for instance physics literature, or J. Feng papers, or Freidlin-Wentzell, 3rd edition, 2012)

A Formal Proof

- We have to compute the probability of the Newton ratio and prove the result

$$\mathbb{E}_x\left[\delta\left(\frac{X_\varepsilon(\Delta t)-x}{\Delta t}-\dot{x}\right)\right]\underset{\varepsilon\downarrow 0}{\asymp}\exp\left(-\frac{\Delta t L(x,\dot{x})}{\varepsilon}\right).$$

- We can rather compute the cumulant generating function and use the Gärtner–Ellis theorem. If for all p, the limit

$$H(x,p)=\lim_{\varepsilon\downarrow 0}\lim_{\Delta t\downarrow 0}\frac{\varepsilon}{\Delta t}\log\mathbb{E}_x\left[\exp\left(\frac{\Delta t}{\varepsilon}p.\frac{X_\varepsilon(\Delta t)-x}{\Delta t}\right)\right]$$

exists and H is everywhere differentiable then

$$L(x,\dot{x})=\sup_p\left\{p\dot{x}-H(x,p)\right\}.$$

A Formal Proof

$$H(x,p)=\lim_{\varepsilon\downarrow 0}\lim_{\Delta t\downarrow 0}\frac{\varepsilon}{\Delta t}\log\mathbb{E}_x\left[\exp\left(\frac{\Delta t}{\varepsilon}p.\frac{X_\varepsilon(\Delta t)-x}{\Delta t}\right)\right]$$

- We connect this to the definition of the infinitesimal generator $\mathbb{E}_x\left[\phi(X(\Delta t))\right]=\phi(x)+G\left[\phi\right](x)\Delta t+o(\Delta t)$:

$$\frac{1}{\Delta t}\log\mathbb{E}_x\left[\exp\left(\frac{\Delta t}{\varepsilon}p.\frac{X_\varepsilon(\Delta t)-x}{\Delta t}\right)\right]$$

$$=\frac{1}{\Delta t}\log\left\{\mathbb{E}_x\left[\exp\left(\frac{p.X_\varepsilon(\Delta t)}{\varepsilon}\right)\right]\exp\left(-\frac{p.x}{\varepsilon}\right)\right\}$$

$$=\frac{1}{\Delta t}\log\left\{1+\Delta t G_\varepsilon\left[\exp\left(\frac{px}{\varepsilon}\right)\right]\exp\left(-\frac{px}{\varepsilon}\right)+o(\Delta t)\right\}$$

$$=G_\varepsilon\left[\exp\left(\frac{px}{\varepsilon}\right)\right]\exp\left(-\frac{px}{\varepsilon}\right)+o(\Delta t).$$

Large Deviation Rate Function from the Infinitesimal Generator of a Continuous Time Markov Process

- Continuous time Markov processes $\{X_\varepsilon(t)\}_{0\leq t<\infty}$.
- We assume that for all $p\in\mathbb{R}^n$ the limit

$$H(x,p)=\lim_{\varepsilon\downarrow 0}\varepsilon G_\varepsilon\left[\mathrm{e}^{\frac{p.x}{\varepsilon}}\right]\mathrm{e}^{-\frac{p.x}{\varepsilon}}\ \text{exists.}$$

- Then the family X_ε verifies a large deviation principle

$$P\left[\{X_\varepsilon(t)\}_{0\leq t<T}=\{x(t)\}_{0\leq t<T}\right]\underset{\varepsilon\downarrow 0}{\asymp}\exp\left(-\frac{\int_0^T \mathrm{d}t\,L(x,\dot{x})}{\varepsilon}\right).$$

with rate ε and rate function

$$L(x,\dot{x})=\sup_p\{p\dot{x}-H(x,p)\}.$$

(see for instance physics literature, or J. Feng papers, or Freidlin-Wentzell, 3rd edition, 2012)

Example 1: Radioactive Decay

- We consider the decay from a state $x = 1$ to a state $x = 0$ at rate λ (the event $1 \to 0$ occurs after a time τ_d distributed according to $\mathbb{P}(\tau_d = \tau) = \lambda \exp(-\lambda\tau)$). Pure death process.
- N independent particles x_n, with $1 \leq n \leq N$, each of which undergo a radioactive decay.

$$X_N(t) = \frac{1}{N}\sum_{n=1}^{N} x_n(t).$$

- Law of large numbers: $X_N \underset{N\to\infty}{\to} \bar{X}$ with

$$\frac{\mathrm{d}\bar{X}}{\mathrm{d}t} = -\lambda\bar{X} \text{ thus } \bar{X}(t) = \exp(-\lambda t).$$

- We expect

$$P\left[\{X_N(t)\}_{0\leq t\leq T} = \{X(t)\}\right] \underset{N\uparrow\infty}{\asymp} \exp\left(-N\int_0^T \mathrm{d}t\, L(x,\dot{x})\right).$$

Large Deviations for the Radioactive Decay of N Particles

- Definition of the infinitesimal generator. ϕ is a test function.

$$G\left[\phi\right](x) = \lim_{\Delta t \downarrow 0} \frac{\mathbb{E}_x\left[\phi(X(\Delta t))\right] - \phi(x)}{\Delta t}.$$

- For the radioactive decay of N independent particle, $X_N = \frac{1}{N}\sum_{n=1}^{N} x_n(t)$ is a Markov chain. *If* $X_N = x$, during an infinitesimal increment Δt, it has a probability $N\lambda\Delta t x$ to decrease to the value $x - 1/N$

$$G_N\left[\phi\right](x) = N\lambda x\left[\phi\left(x - \frac{1}{N}\right) - \phi\left(x\right)\right].$$

$$\lim_{N\to\infty} \frac{1}{N} G_N\left[e^{Npx}\right] e^{-Npx} = \lambda x\left(e^{-p} - 1\right) \equiv H(x,p).$$

Radioactive Decay: Lagrangian and Relaxation Path

- Then

$$P\left[\{X_N(t)\}_{0\leq t\leq T}=\{X(t)\}\right]\underset{N\uparrow\infty}{\asymp}\exp\left(-N\int_0^T \mathrm{d}t\,L(x,\dot{x})\right).$$

- Lagrangian $L(x,\dot{x})=\sup_p\{p\dot{x}-H(x,p)\}$:

$$L(x,\dot{x})=\dot{x}+\lambda x-\dot{x}\log\left(-\frac{\dot{x}}{\lambda x}\right)\text{ if }\dot{x}<0\text{ and }-\infty\text{ otherwise.}$$

- Relaxation path:

$$\dot{x}=R(x)=\frac{\partial H}{\partial p}(x,0)=-\lambda x.$$

Example 2: Network of Reacting Particles

- N particles X_n, with $1 \leq n \leq N$. Each of them can be in K different states: $X_n = k$ with $1 \leq k \leq K$
- The number of particles which are in the state k is NF_k:

$$\text{Empirical distribution :}\mathbf{F}^N = \{F_k\}_{1 \leq k \leq K}$$

- We assume that the particles have reactions:

$$k_1 + k_2 \to k_3 + k_4 \text{ at a rate } \frac{\lambda_{k_1 k_2 k_3 k_4}}{N}.$$

- Number of reactions per unit of time: of order N. Number of reactions per particle per unit of time: of order 1.
- During a time of order 1, the empirical distribution changes N times, each time with a modification of order $1/N$.
- We expect

$$P\left[\left\{\mathbf{F}^N(t)\right\}_{0 \leq t \leq T} = \{\mathbf{f}(t)\}\right] \underset{N\uparrow\infty}{\asymp} \exp\left(-N\int_0^T \mathrm{d}t\, L\left(\mathbf{f},\dot{\mathbf{f}}\right)\right).$$

Larges Deviations for the Network of Reacting Particles

- The infinitesimal generator for the empirical distribution $\mathbf{F}^N = \{F_k\}_{1\leq k\leq K}$ is

$$G(\mathbf{f}) = N \sum_{k_1,k_2,k_3,k_4} \lambda_{k_1 k_2 k_3 k_4} f_{k_1} f_{k_2} \left[\phi\left(f_{k_1} - \frac{1}{N}, f_{k_2} - \frac{1}{N}, f_{k_3} + \frac{1}{N}, f_{k_4} + \frac{1}{N}, \tilde{\mathbf{f}}\right) - \phi(\mathbf{f})\right].$$

- Using the formula for the Hamiltonian $H(x,p) = \lim_{\varepsilon\downarrow 0} \varepsilon G_\varepsilon \left[e^{\frac{p.x}{\varepsilon}}\right] e^{-\frac{p.x}{\varepsilon}}$, we get

$$H(\mathbf{f},\mathbf{p}) = \sum_{k_1,k_2,k_3,k_4} \lambda_{k_1 k_2 k_3 k_4} f_{k_1} f_{k_2} \left(e^{p_3+p_4-p_2-p_1} - 1\right).$$

Outline

Abstract Properties of Dynamical Large Deviations

- We assume dynamical large deviations

$$P\left[\{X_{\varepsilon}(t)\}_{0\leq t<T}=\{x(t)\}_{0\leq t<T}\right]\underset{\varepsilon\downarrow 0}{\asymp}\exp\left(-\frac{\int_0^T \mathrm{d}t\, L(x,\dot{x})}{\varepsilon}\right),$$

with

$$L(x,\dot{x})=\sup_p\left[p\dot{x}-H(x,p)\right].$$

- Why is the large deviation rate function useful?
- What are the properties which are common to any dynamical large deviations, and thus common to all kinetic theories?

Quasipotential

- We assume dynamical large deviations

$$P\left[\{X_\varepsilon(t)\}_{0\le t<T} = \{x(t)\}_{0\le t<T}\right] \underset{\varepsilon\downarrow 0}{\asymp} \exp\left(-\frac{\int_0^T \mathrm{d}t\, L(x,\dot{x})}{\varepsilon}\right),$$

  with

$$L(x,\dot{x}) = \sup_p \left[p\dot{x} - H(x,p)\right].$$

- We assume the process is stationary, and the invariant measure has a large deviation rate

$$P_{s,\varepsilon}(x) \equiv \mathbb{E}\left[\delta\left(X_\varepsilon - x\right)\right] \underset{\varepsilon\downarrow 0}{\asymp} \exp\left(-\frac{U(x)}{\varepsilon}\right),$$

- U is the quasipotential. We assume that U has a single minimizer x_0, then $U(x_0) = 0$. Then

$$U(x) = \inf_{\left\{\{X(t)\}_{-\infty\le t\le 0} \middle| X(-\infty)=x_0 \text{ and } X(0)=x\right\}} \int_{-\infty}^{0} \mathrm{d}t\, L\left(X,\dot{X}\right).$$

Hamilton–Jacobi eq., Relaxation and Fluctuation Paths

1. H is a convex function of the variable p and $H(x,0)=0. L\geq 0.$
2. **The relaxation paths** (most probable paths) solve

$$\dot{X}_r = R(X_r) \text{ with } R(x) = \arg\inf_{\dot{x}} L(x,\dot{x}) = \frac{\partial H}{\partial p}(x,0).$$

3. The quasipotential solves the **Hamilton–Jacobi equation**

$$H(x,\nabla U)=0.$$

4. **The fluctuation paths** solve

$$\dot{x} = F(x) \equiv \frac{\partial H}{\partial p}(x,\nabla U(x)).$$

Monotonicity of U Along Relaxation and Fluctuation Paths

- As H is convex, the quasipotential decreases along the relaxation paths

$$\frac{\mathrm{d}U}{\mathrm{d}t}(X_r) = H(X_r,0) - H(X_r,\nabla U(X_r)) + \frac{\partial H}{\partial p}(X_r,0).\nabla U(X_r) \leq 0.$$

- As H is convex, the quasipotential increases along the fluctuation paths

$$\frac{\mathrm{d}U}{\mathrm{d}t}(X_f) = H(X_f,0) - H(X_f,\nabla U(X_f)) + \frac{\partial H}{\partial p}(X_f,\nabla U(X_f)).\nabla U(X_f) \geq 0.$$

Conservation laws

- $C(x)$ is a conserved quantity iff

$$\text{for any } x \text{ and } p,\;\; L(x,\dot{x}) = +\infty \text{ if } \dot{x}.\frac{\partial C}{\partial x}(x) \neq 0,$$

or equivalently

$$\text{for any } x \text{ and } p,\;\; \frac{\partial H}{\partial p}(x,p).\frac{\partial C}{\partial x}(x) = 0.$$

A Sufficient Condition for U to be the Quasipotential

1. If U solves the Hamilton–Jacobi equation,
2. if U has a single minimum x_0 with $U(x_0) = 0$,
3. if for any x the solution of the reverse fluctuation path dynamics $\dot{X} = -F(X) = -\frac{\partial H}{\partial p}(X, \nabla U(X))$ with $X(0) = x$, converges to x_0 for large times

then U is the quasipotential.

Abstract Properties of Dynamical Large Deviations

- We assume dynamical large deviations

$$P\left[\{X_{\varepsilon}(t)\}_{0\leq t<T}=\{x(t)\}_{0\leq t<T}\right]\underset{\varepsilon\downarrow 0}{\asymp}\exp\left(-\frac{\int_0^T \mathrm{d}t\, L(x,\dot{x})}{\varepsilon}\right),$$

with

$$L(x,\dot{x})=\sup_p\left[p\dot{x}-H(x,p)\right].$$

There are more abstract general properties that are important and that I will discuss later. Those include

- Characterization of time-reversibility at the level of the large deviations.
- Relations between the large deviation rate function and the geometry of the associated partial differential equations or kinetic equations.

Suggested reading about today lecture

- Lecture notes for a basic introduction to large deviation theory

1a) For the sum of independent random variables
1b) For dynamical systems with weak noises

- Section 3 of the paper "Is the Boltzmann Equation Reversible? A Large Deviation Perspective on the Irreversibility Paradox", F. Bouchet, J. Stat. Phys., 2020.