Chapter 1
From the Law of Large Numbers to Large Deviation Theory in Statistical Physics: An Introduction

Fabio Cecconi, Massimo Cencini, Andrea Puglisi, Davide Vergni, and Angelo Vulpiani

Abstract This contribution aims at introducing the topics of this book. We start with a brief historical excursion on the developments from the law of large numbers to the central limit theorem and large deviations theory. The same topics are then presented using the language of probability theory. Finally, some applications of large deviations theory in physics are briefly discussed through examples taken from statistical mechanics, dynamical and disordered systems.

1.1 Introduction

Describing the physical properties of macroscopic bodies via the computation of (ensemble) averages was the main focus of statistical mechanics at its early stage. In fact, as macroscopic bodies are made of a huge number of particles, fluctuations were expected to be too small to be actually observable. Broadly speaking, we can say that the theoretical basis of statistical descriptions was guaranteed by the law of large numbers. Boltzmann wrote: *In the molecular theory we assume that the laws of the phenomena found in nature do not essentially deviate from the limits that they*

F. Cecconi • M. Cencini
Istituto dei Sistemi Complessi, Consiglio Nazionale delle Ricerche, Via dei Taurini 19, Rome, I-00185, Italy
e-mail: fabio.cecconi@roma1.infn.it; massimo.cencini@cnr.it

A. Puglisi (✉) • A. Vulpiani
Istituto dei Sistemi Complessi, Consiglio Nazionale delle Ricerche, and Dipartimento di Fisica, Università "Sapienza", Piazzale Aldo Moro 5, Rome, I-00185, Italy
e-mail: andrea.puglisi@roma1.infn.it; angelo.vulpiani@roma1.infn.it

D. Vergni
Istituto per le Applicazioni del Calcolo, Consiglio Nazionale delle Ricerche, Via dei Taurini 19, Rome, I-00185, Italy
e-mail: davide.vergni@cnr.it

A. Vulpiani et al. (eds.), *Large Deviations in Physics*, Lecture Notes in Physics 885,
DOI 10.1007/978-3-642-54251-0_1,

would approach in the case of an infinite number of infinitely small molecules, while Gibbs[1] remarked ... *[the fluctuations] would be in general vanishing quantities, since such experience would not be wide enough to embrace the more considerable divergences from the mean values* [1].

Although very small, the importance of fluctuations was recognized quite early to find conclusive evidence for the atomistic hypothesis. At the end of the nineteenth century, atomic theory was still considered, by influential scientists as Ostwald and Mach, useful but non real for the building of a consistent description of nature: *The atomic theory plays a part in physics similar to that of certain auxiliary concepts in mathematics; it is a mathematical model for facilitating the mental reproduction of facts* [1].

The situation changed at the beginning of the twentieth century, when Einstein realized the central role played by the fluctuations and wrote: *The equation* [NA: for the energy fluctuations $\langle E^2\rangle - \langle E\rangle^2 = kT^2C_V$, $C_V = \partial\langle E\rangle/\partial T$ being the specific heat] *we finally obtained would yield an exact determination of the universal constant* [NA: the Avogadro number], *if it were possible to determine the average of the square of the energy fluctuations of the system; this is however not possible according to our present knowledge*. For macroscopic objects the equation for the energy fluctuations cannot actually be used for the determination of the Avogadro number. However, Einstein's intuition was correct as he understood how to relate the Avogadro number to a macroscopic quantity—the diffusion coefficient D—obtained by observing the fluctuations of a Brownian particle—D indeed describes the long time ($t \to \infty$) behavior of particle displacement $\langle (x(t) - x(0))^2\rangle \simeq 2Dt$, which is experimentally accessible. The theoretical work by Einstein and the experiments by Perrin gave a conclusive evidence of atomism: the celebrated relationship between the diffusion coefficient (measurable at the macroscopic level) and the Avogadro number N_A (related to the atomistic description) is

$$D = \frac{RT}{6N_A\pi\eta a},$$

where T and η are the temperature and dynamic viscosity of the fluid respectively, a the radius of the colloidal particle, $R = N_Ak$ is the perfect gas constant and k is the Boltzmann constant.

Einstein's seminal paper on Brownian motion contains another very important result, namely the first example of Fluctuation-Dissipation Theorem (FDT): a relation between the fluctuations (given by correlation functions) of an unperturbed system and the mean response to a perturbation. In the specific case of Brownian motion, FDT appears as a link between the diffusion coefficient (a property of the unperturbed system) and the mobility, which measures how the system reacts to a small perturbation.

[1] Who, by the way, already knew the expression for the mean square energy fluctuations.

Beyond their conceptual relevance and the link with response functions, fluctuations in macroscopic systems are quantitatively extremely small and hard to detect (but for the case of second order phase transition in equilibrium systems). However, in recent years statistical mechanics of small systems is becoming more and more important due to the theoretical and technological challenges of micro- and nano-physics. In such small systems[2] since large excursions from averages values become increasingly important, it is mandatory to go beyond the Gaussian approximation (i.e. beyond the realm of validity of the central limit theorem) by means of the large deviation theory.

Next section presents a non-exhaustive historical survey from the law of large numbers to large deviation theory. Then, in Sect. 1.3 we illustrate with two examples how large deviation theory works. Section 1.4 illustrates some applications of large deviations in statistical physics.

1.2 An Informal Historical Note

Perhaps the most straightforward way to understand the connection between Law of Large Numbers (LLN), the Central Limit Theorem (CLT) and the Large Deviation Theory (LDT) is to consider a classical topic of probability theory, namely the properties of the empirical mean

$$y_N = \frac{1}{N}\sum_{j=1}^{N} x_j \tag{1.1}$$

of a sequence $\{x_1, \ldots, x_N\}$ of N random variables. Three basic questions naturally arise when N is very large:

(a) The behavior of the empirical mean y_N, the possible convergence to an asymptotic value and its dependence on the sequence;
(b) The statistics of small fluctuations of y_N around $\langle y_N \rangle$, i.e., of $\delta y_N = y_N - \langle y_N \rangle$ when $|\delta y_N|$ is "small";
(c) The statistical properties of rare events when such fluctuations are "large".

In the simplest case of sequences $\{x_1, \ldots, x_N\}$ of independent and identically distributed (i.i.d.) random variables with expected value $\langle x \rangle$ and with finite variance, the law of large numbers answers point (a): the empirical average gets close and closer to the expected value $\langle x \rangle$ when N is large:

$$\lim_{N\to\infty} P\left(|y_N - \langle x \rangle| < \epsilon\right) \to 1\,. \tag{1.2}$$

[2]We note that sometimes even in macroscopic systems (e.g. granular materials) the number of effective elementary constituents (e.g. the seeds) is not astonishingly large as in gases or liquids.

In the more general case of dependent variables, in principle, the empirical mean may depend on the specific sequence of random variables. This is the essence of the ergodic problem which generalizes the LLN and had a crucial role for the development of statistical mechanics.

Issue (b) is addressed by the central limit theorem. For instance, in the simple case of i.i.d. variables with expected value $\langle x \rangle$ and finite variance σ^2, the CLT describes the statistics of small fluctuations, $|\delta y_N| \lesssim O(\sigma/\sqrt{N})$, around the mean value when N is very large. Roughly speaking, the CLT proves that, in the limit $N \gg 1$, the quantity

$$z_N = \frac{1}{\sigma\sqrt{N}} \sum_{j=1}^{N} (x_j - \langle x \rangle) \tag{1.3}$$

is normally distributed, meaning that

$$p(z_N = z) \simeq \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} , \tag{1.4}$$

independently of the distribution of the random variables. Under suitable hypothesis the theorem can be extended to dependent (weakly correlated) variables.

Finally, the last point (c) is the subject of large deviation theory which, roughly, states that in the limit $N \gg 1$

$$p(y_N = y) \sim e^{-N\mathscr{C}(y)} . \tag{1.5}$$

Unlike the central limit theorem result with the "universal" limit probability density (1.4), the detailed functional dependence of $\mathscr{C}(y)$—the Cramér or rate function—depends on the probability distribution of $\{x_1, \ldots, x_N\}$. However, $\mathscr{C}(y)$ possesses some general properties: it is zero for $y = \langle y_N \rangle$ and positive otherwise, moreover—when the variables are independent (or weakly correlated)—it is a convex function. Clearly, whenever the CLT applies, $\mathscr{C}(y)$ can be approximated by a parabola around it minimum in $\langle y \rangle$.

As frequently occurring in the development of science, the actual historical path did not follow the simplest trajectory: (a) then (b) and at the end (c). Just to mention an example, Boltzmann introduced the ergodic problem and developed—ante litteram—some aspects of large deviations well before the precise mathematical formulation of the central limit theorem.

1.2.1 Law of Large Numbers and Ergodicity

In the origins, the calculus of probabilities was, to a large extent, a collection of specific rules for specific problems, mainly a matter for rolling dice and card games. For instance, the works by Pascal and Fermat originated by practical questions

in gambling raised by the chevalier de Méré (a French nobleman in love with gambling) [2].

1.2.1.1 J. Bernoulli

J. Bernoulli gave the first important contribution moving the theory of probability away from gambling context with the posthumous book *Ars Conjectandi* (The art of conjecturing), published in 1713 and containing the LLN.[3] In modern terms, if $\{x_1, \ldots, x_N\}$ are i.i.d. with finite variance and expected value $\langle x \rangle$ then for each $\epsilon > 0$ and if $N \to \infty$

$$P\left(\left|\frac{1}{N}\sum_{j=1}^{N} x_j - \langle x \rangle\right| > \epsilon\right) \to 0 \,. \tag{1.6}$$

A particularly important case of the above result is

$$P(|f_N - p| > \epsilon) \to 0 \,, \tag{1.7}$$

where f_N is the frequency of an event over N independent trials, and p is its occurrence probability in a single trial. The result (1.7) stands at the basis of the interpretation of probability in terms of frequencies.

1.2.1.2 Boltzmann

Boltzmann introduced the ergodic hypothesis while developing statistical mechanics [3]. In modern language, we can state the ergodic problem as follows. Consider a deterministic evolution law U^t in the phase space Ω,

$$\mathbf{X}(0) \to \mathbf{X}(t) = U^t \mathbf{X}(0) \,,$$

and a probability measure $d\mu(\mathbf{X})$ invariant under the evolution U^t, meaning that $d\mu(\mathbf{X}) = d\mu(U^{-t}\mathbf{X})$. The dynamical system $\{\Omega, U^t, d\mu(\mathbf{X})\}$ is ergodic, with respect to the measure $d\mu(\mathbf{X})$, if, for every integrable function $A(\mathbf{X})$ and for almost all initial conditions $\mathbf{X}(t_0)$, time and phase average coincide:

$$\overline{A} \equiv \lim_{\mathscr{T}\to\infty} \frac{1}{\mathscr{T}} \int_{t_0}^{t_0+\mathscr{T}} A(\mathbf{X}(t))\mathrm{d}t = \int A(\mathbf{X}) d\mu(\mathbf{X}) \equiv \langle A \rangle \,, \tag{1.8}$$

where $\mathbf{X}(t) = U^{t-t_0}\mathbf{X}(t_0)$.

[3] The most rudimentary form of the LLN seems to be credited to Cardano.

It is worth recalling why the ergodic hypothesis was so important for the development of statistical mechanics. Simplifying, Boltzmann's program was to derive thermodynamics for macroscopic bodies—composed by, say, $N \gg 1$ particles—from the microscopic laws of the dynamics. Thermodynamics consists in passing from the $6N$ degrees of freedom to a few macroscopic, experimentally accessible quantities such as, e.g., the temperature and pressure. An experimental measurement is actually the result of a single observation during which the system passes through a very large number of microscopic states. Denoting with $\mathbf{q}_i$ and $\mathbf{p}_i$ the position and momentum vectors of the i-th particle, the microscopic state of the N-particles system at time t is described by the $6N$-dimensional vector $\mathbf{X}(t) \equiv (\mathbf{q}_1(t), \ldots, \mathbf{q}_N(t); \mathbf{p}_1(t), \ldots, \mathbf{p}_N(t))$, which evolves according to the Hamilton equations. The measurement of an observable $A(\mathbf{x})$ effectively corresponds to an average performed over a very long time (from the microscopic point of view): $\overline{A}_{\mathscr{T}} = (1/\mathscr{T}) \int_{t_0}^{t_0+\mathscr{T}} A(\mathbf{X}(t))\mathrm{d}t$. The theoretical calculation of the time average $\overline{A}_{\mathscr{T}}$, in principle, requires both the knowledge of the microscopic state at time t_0 and the determination of its evolution. The ergodic hypothesis eliminates both these necessities, provided we know the invariant measure. In statistical mechanics of, e.g., isolated systems a natural candidate for the invariant measure $d\mu(\mathbf{X})$ is the microcanonical measure on the constant energy surface $H = E$.

To the best of our knowledge the first precise result on ergodicity, i.e. the validity of (1.6) for non independent stochastic processes has been obtained by A.A. Markov, for a wide class of stochastic processes (now called Markov Chains). Consider an aperiodic and irreducible Markov Chain with M states, transition probabilities $\{P_{i\to j}\}$, and invariant probabilities $\pi_1, \pi_2, \ldots, \pi_M$, and an observable A which takes value A_j on the state j, then for almost all the realizations $\{j_t\}$ we have

$$\overline{A} \equiv \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} A_{j_t} = \sum_{j=1}^{N} A_j \pi_j = \langle A \rangle \,, \tag{1.9}$$

where j_t indicates the state of the chain at time t of a "walker" performing a trajectory according to the transition probabilities $\{P_{i\to j}\}$. There is a curious story at the origin of the above result [4]. Markov, who was an atheist and a strong *critic* of both the tsarist government and the Orthodox Church, at the beginning of the twentieth century had a rather hot diatribe with the mathematician Nekrasov, who had opposite political and religious opinions. The subject of the debate was about the statistical regularities and their role for the problem of free will. Nekrasov noted that the LLN of Bernoulli is based on the independence of successive experiments, while, among human beings, the situation is rather different, hence the LLN cannot, in any way, explain the statistical regularities observed in social life. Such a remark led Markov to find an example of non independent variables for which a generalized LLN holds; in a letter to a colleague he wrote:

> I considered variables connected in a simple chain, and from this came the idea of the possibility of extending the limit theorems of the calculus of probability also to a complex chain.

The ergodic problem in deterministic systems is much more difficult than its analogous for Markov chains. It is rather natural, both from a mathematical and a physical point of view, to wonder under which conditions a dynamical system is ergodic. At an abstract level for a dynamical system $(\Omega, U^t, d\mu(\mathbf{X}))$, the problem has been tackled by Birkhoff and von Neumann who proved the following fundamental theorems:

Theorem 1.1. *For almost every initial condition* $\mathbf{X}_0$ *the time average*

$$\overline{A}(\mathbf{X}_0) \equiv \lim_{\mathscr{T}\to\infty} \frac{1}{\mathscr{T}} \int_0^{\mathscr{T}} A(U^t \mathbf{X}_0) \mathrm{d}t \tag{1.10}$$

exists.

Theorem 1.2. *A necessary and sufficient condition for the system to be ergodic, is that the phase space* Ω *be metrically indecomposable. The latter property means that* Ω *can not be subdivided into two invariant (under the dynamics* U^t*) parts each of positive measure.*

Sometimes instead of metrically indecomposable the equivalent term "metrically transitive" is used. Theorem 1.1 is rather general and not very stringent, in fact time average $\overline{A}(\mathbf{X}_0)$ can depend on the initial condition. The result of Theorem 1.2, while concerning the foundations of statistical mechanics, remains of poor practical utility, since, in general, it is almost impossible to decide whether a given system satisfies the condition of metrical indecomposability. So that, at a practical level, Theorem 1.2 only shifts the problem.

1.2.1.3 Ergodicity and Law of Large Numbers in Statistical Mechanics

Strictly speaking, the ergodicity is a too demanding property to be verified and proved in systems of practical interest. Khinchin in his celebrated book *Mathematical Foundation of the Statistical Mechanics* [5] presents some important results on the ergodic problem which overcome the formal mathematical issues.

The general idea of his approach is based on the following facts:

(a) In the systems which are of interest to statistical mechanics the number of degrees of freedom is very large;
(b) In statistical mechanics, the important observables are not generic (in mathematical sense) functions, so it is enough to restrict the validity of the ergodic hypothesis (1.8) just to the relevant observables;
(c) One can accept that Eq. (1.8) does not hold for initial conditions $\mathbf{X}_0$ in a region of small measure (which goes to zero as $N \to \infty$).

Khinchin considers a separable Hamiltonian system i.e.:

$$H = \sum_{n=1}^{N} H_n(\mathbf{q}_n, \mathbf{p}_n) \tag{1.11}$$

and a special class of observables (called *sum functions*) of the form

$$f(\mathbf{X}) = \sum_{n=1}^{N} f_n(\mathbf{q}_n, \mathbf{p}_n) \tag{1.12}$$

where $f_n = O(1)$. Interesting examples of sum functions are the pressure, the kinetic energy, the total energy and the single-particle distribution function. Notice that a change $O(1)$ in a single f_n results in a relative variation $O(1/N)$ in $f(\mathbf{X})$: the sum functions are "good" macroscopic functions, since they are not so sensitive to microscopic details.

The main result, obtained using the LLN, is:

$$\text{Prob}\left(\frac{|\overline{f} - \langle f \rangle|}{|\langle f \rangle|} \geq K_1 N^{-1/4}\right) \leq K_2 N^{-1/4} \tag{1.13}$$

where K_1 and K_2 are $O(1)$.

The restriction to the separable structure of the Hamiltonian, i.e. (1.11), had been removed by Mazur and van der Linden [6]. They extended the result to systems of particles interacting through a short range potential. Let us stress that in the Khinchin result, as well as in the generalization of Mazur and van der Linden, basically the dynamics has no role and the existence of good statistical properties follows from the LLN, i.e. using the fact that $N \gg 1$.

1.2.1.4 Ergodicity at Work in Statistical Mechanics

We conclude this short excursus on LLN and ergodicity mentioning some important uses of such topics in statistical physics.

The Boltzmann ergodic hypothesis and the result (1.9) for Markov chains are the conceptual starting point for two powerful computational methods in statistical mechanics: molecular dynamics and Monte Carlo method, respectively. In the first approach one assumes (without a mathematical proof) ergodicity[4] and computes time averages from the numerical integration of the "true" Hamilton's equations. In the Monte Carlo approach one selects an ergodic Markov chain[5] with the correct equilibrium probability. Of course, in practical computations, one has to

[4]It is now well known, e.g. from KAM theorem and FPU simulations, that surely in some limit ergodicity fails, however it is fair to assume that the ergodic hypothesis holds for liquids or interacting gases.

[5]Note that, at variance with the molecular dynamics, the Monte Carlo dynamics is somehow artificial (and not unique), therefore the dynamical properties, e.g. correlation functions, are not necessarily related to physical features.

face nontrivial problems, firstly how to estimate the typical time necessary to have a good average and how to control the errors.

Another interesting application is the following: consider a simple multiplicative process: $x_N = a_N x_{N-1}$ where $\{a_j\}$ are i.i.d. positive. Using the LLN[6] it is simple to show that for almost all the realizations one has $x_N \sim e^{\lambda N}$, where $\lambda = \langle \ln a \rangle$ or more formally $P(|(1/N)\ln(x_N/x_0) - \lambda| > \epsilon) \to 0$ as $N \to \infty$. Let us now repeat the problem for non commutative random matrices $\{\mathbb{A}_j\}$, i.e. the multiplicative process $\mathbb{X}_N = \mathbb{A}_N \mathbb{X}_{N-1}$, we can wonder about the limit for $N \to \infty$ of $(1/N)\ln||\mathbb{X}_N||/||\mathbb{X}_o||$, where $||(.)||$ indicates a norm. At first glance the above problem can sound rather artificial, on the contrary it is important for disordered systems[7] and chaotic dynamics. In the 1960s Furstenberg and Kester [7] have proven, under suitable general conditions, the existence of the limit $(1/N)\ln||\mathbb{X}_N||/||\mathbb{X}_o||$ for almost all the realizations: assume that $\langle \ln^+ ||\mathbb{A}_j|| \rangle < \infty$ (where $\ln^+ x = 0$ if $x \leq 1$ and $\ln^+ x = \ln x$ otherwise) then the limit $\lambda_1 = \lim_{N\to\infty}(1/N)\ln||\mathbb{X}_N||/||\mathbb{X}_o||$ exists with probability 1. This result had been extended to deterministic ergodic system by Oseledec [8] in the case the $\{\mathbb{A}_j\}$ are obtained linearizing the dynamics along the trajectory.

1.2.2 Central Limit Theorems

1.2.2.1 The Beginning

The first version of the CLT is due to A. de Moivre who studied the asymptotic behavior of the sum

$$S_N = x_1 + \ldots + x_N$$

in the specific case of binomial random variables with $P(x_j = 1) = p$ and $P(x_j = 0) = 1 - p$. Starting from the binomial distribution and the Stirling approximation de Moivre discovered that

$$\lim_{N\to\infty} P\left(a \leq \frac{S_N - Np}{\sqrt{Np(1-p)}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx \,. \tag{1.14}$$

The history of the CLT as universal law, i.e. not only for dichotomic variables, began with Laplace who was able to prove a generalization of the de Moivre's result. With the use of the characteristic functions and asymptotic methods of approximating

[6] It is enough to consider the variables $t_j = \ln x_j$ and $q_j = \ln a_j$, and then, noting that $t_N = q_1 + q_2 + \ldots + q_N$, one can use the LLN and obtain the result.

[7] For instance the discrete one-dimensional Schrödinger equation with a random potential can be written in terms of a product of 2×2 random matrices.

integrals, Laplace proved that, for the case where $\{x_i\}$ are i.i.d. discrete variables with mean value $\langle x \rangle$ and variance $\sigma^2 < \infty$:

$$\lim_{N\to\infty} P\left(a \le \frac{S_N - N\langle x\rangle}{\sqrt{N\sigma^2}} \le b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx \,. \tag{1.15}$$

1.2.2.2 The Russian School and Lindeberg

The first mathematical detailed treatment of the CLT, i.e. the validity of (1.15) for generic i.i.d. (even non discrete) with finite variance, is due to the Russian school with Chebyshev and Markov who used in a rigorous way the method of the characteristic functions and moments [9].

A generalization of the CLT for independent variables with different distribution is due to Lindeberg (around 1920) who proved that, if $\langle x_j \rangle = 0$ (this is not a real limitation) and $0 < \sigma_j^2 < \infty$, under the hypothesis that, for any τ,

$$\lim_{N\to\infty} \frac{1}{D_N^2} \sum_{n=1}^{N} \int_{|x|>\tau D_N} x^2 p_{x_n}(x) dx = 0 \,, \quad \text{where} \quad D_N^2 = \sum_{n=1}^{N} \sigma_n^2 \,, \tag{1.16}$$

one has

$$\lim_{N\to\infty} P\left(a \le \frac{S_N}{\sqrt{D_N^2}} \le b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx \,. \tag{1.17}$$

Intuitively the Lindeberg condition means that each variance σ_n^2 must be small respect to D_N^2: for any τ and for N large enough one has $\sigma_n < \tau D_N$ for all $n \le N$.

Feller and Lévy found that the Lindeberg condition is not only sufficient but also necessary for the validity of the CLT [9].

1.2.2.3 Modern Times

The case of independent variables is quite restrictive, so it is interesting to wonder about the possibility of extending the validity of CLT to non independent variables. Intuition suggests that if the correlation among variables is weak enough a CLT is expected to hold. Such an argument is supported by precise rigorous results [10]. We just mention the basic one. Consider a stationary process with zero mean and correlation function $c(k) = \langle x_{n+k} x_n \rangle$. If the correlation is summable,

$$\sum_{k=1}^{\infty} c(k) < \infty \,, \tag{1.18}$$

it is possible to prove that

$$\lim_{N\to\infty} P\Big(a \le \frac{S_N}{\sqrt{N\sigma_{\text{eff}}^2}} \le b\Big) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx \,, \tag{1.19}$$

where $\sigma_{\text{eff}}^2 = \sigma^2 + 2\sum_{k=1}^{\infty} c(k)$. In other words σ^2 is replaced by σ_{eff}^2; note that for $N \gg 1$

$$\langle S_N \rangle \simeq 2\sigma_{\text{eff}}^2 N \ . \tag{1.20}$$

The previous result is rather important in the context of diffusion. Interpreting N as a discrete time, Eq. (1.20) simply expresses the diffusive behavior of S_N with diffusion coefficient σ_{eff}^2. Violation of the condition (1.18) are thus at the origin of anomalous diffusive behaviors, for instance to observe $\langle S_N \rangle \sim N^\alpha$ with $\alpha > 1$ and $\sigma < \infty$ it is necessary to have $\sum_{k=1}^{\infty} c(k) = \infty$, i.e. strongly correlated random variables. Another possible violation is when $\sigma = \infty$, in such case CLT can be generalized and this is the subject of the infinitely divisible and Lévy stable distributions.

1.2.3 Large Deviation Theory

The large deviation theory studies the rare events and can be seen as a generalization of the CLT, as it describes not only the "typical" fluctuations but also the very large excursions.

The first general mathematical formulation of LDT has been introduced in the 1930s mainly by Cramér for i.i.d. random variables $x_1, x_2, \ldots$ with mean value $\langle x \rangle$. Under the rather general assumption of existence of the moment generating function $\langle \mathrm{e}^{qx} \rangle$ in some neighborhood of $q = 0$, it is possible to prove that for the "empirical mean" $y_N = (x_1 + \ldots + x_N)/N$

$$\lim_{N\to\infty} \frac{1}{N} \ln P(y_N > y) = -\mathscr{C}(y) \tag{1.21}$$

provided $y > \langle x \rangle$ and $P(x > y) > 0$. Of course, by repeating the previous reasoning for the variable reflected with respect to the mean (i.e. $x \to 2\langle x \rangle - x$), one proves the complementary result for $y_N < y < \langle x \rangle$. The Cramér function $\mathscr{C}(y)$ depends on the probability distribution of x, is positive everywhere but for $y = \langle x \rangle$ where it vanishes. In addition, it is possible to prove that is convex, i.e. $\mathscr{C}'' > 0$.

Roughly speaking, the essence of the above result is that for very large N the probability distribution function of the empirical mean takes the form

$$p(y_N = y) \sim e^{-N\mathscr{C}(y)} \,. \tag{1.22}$$

It is interesting to remind that the first LDT calculation has been carried out by Boltzmann. He was able to express the asymptotic behavior of the multinomial probabilities in terms of relative entropy, see Sect. 1.3.1. In his approach a crucial physical aspect is the statistical interpretation of the entropy as a bridge between microscopic and macroscopic levels.

Let us note that, in general, the Cramér function (for independent or weakly correlated variables) must obey some constraints:

(i) $\mathscr{C}(y) > 0$ for $y \neq \langle y \rangle = \langle x \rangle$;
(ii) $\mathscr{C}(y) = 0$ for $y = \langle y \rangle$;
(iii) $\mathscr{C}(y) \simeq (y - \langle y \rangle)^2/(2\sigma^2)$, where $\sigma^2 = \langle (x - \langle x \rangle)^2 \rangle$, if y is close to $\langle y \rangle$.

Properties (i) and (ii) are consequences of the law of large numbers, and (iii) is nothing but the central limit theorem.

Moreover, the Cramér function $\mathscr{C}(y)$ is linked via a Legendre transform

$$\mathscr{C}(y) = \sup_q \{qy - L(q)\} \,, \tag{1.23}$$

to the cumulant generating function of the variable x

$$L(q) = \ln \langle e^{qx} \rangle \,. \tag{1.24}$$

The result (1.23) is easily understood by noticing that the average $\langle e^{qNy_N} \rangle$ can be written in two equivalent ways,

$$\langle e^{qNy_N} \rangle = \langle e^{qx} \rangle^N = e^{NL(q)}$$

$$\langle e^{qNy_N} \rangle = \int e^{qNy_N} \, p(y_N) dy_N \sim \int e^{[qy - \mathscr{C}(y)]N} dy \,,$$

yielding

$$\int e^{[qy - \mathscr{C}(y)]N} dy \sim e^{NL(q)} \,. \tag{1.25}$$

In the limit of large N, a steepest descent evaluation of the above integral provides

$$L(q) = \sup_y \{qy - \mathscr{C}(y)\} \,, \tag{1.26}$$

which is the inverse of (1.23). Due to the convexity of $\mathscr{C}(y)$, Eqs. (1.23) and (1.26) are fully equivalent. For a nice general discussion on large deviations see the booklet by Varadhan [11].

For dependent variables, in analogy with the central limit theorem, we expect that if the dependence is weak enough a large deviations description such as (1.22) holds, where the Cramér function depends on the specific features of the correlations.

We sketch in the following the case of ergodic Markov chains with a finite number of states.

Consider a sequence $\mathscr{S}^{(N)} = (s_1, s_2, \ldots, s_N)$ where s_t denotes the state of the chain at time t. Given a function of state, $f(s_t)$, the Cramér function of the sum

$$F_N = \frac{1}{N} \sum_{t=1}^{N} f(s_t) \tag{1.27}$$

can be explicitly computed [12]. From the transition probabilities $P_{i\to j}$ for any real q we can define the matrix

$$\mathbb{P}_{ij}^{(q)} = P_{i\to j} \mathrm{e}^{q f_i} \tag{1.28}$$

where f_i is the value of the function $f(s_t)$ when $s_t = i$. Denoting with $\zeta(q)$ the largest eigenvalue of $\mathbb{P}^{(q)}$, whose uniqueness is ensured by the Perron-Frobenius theorem for positive-entry matrices [13], the Cramér function is given by the formula

$$\mathscr{C}(F) = \sup_q \{qF - \ln \zeta(q)\}\,, \tag{1.29}$$

which generalizes (1.23) to Markov chains.

For general non independent variables, $L(q)$ is defined as

$$L(q) = \lim_{N\to\infty} \frac{1}{N} \ln\langle e^{q\sum_{n=1}^{N} x_n}\rangle\ ,$$

and (1.23) is exact whenever $\mathscr{C}(y)$ is convex, otherwise Eq. (1.23) just gives the convex envelop of the correct $\mathscr{C}(y)$.

1.3 LDT for the Sum and Product of Random Independent Variables

1.3.1 A Combinatorial Example

A natural way to introduce the large deviation theory and show its deep relation with the concept of entropy is to perform a combinatorial computation. Consider the simple example of a sequence of independent unfair-coin tosses. The possible outcomes are head $(+1)$ or tail (-1). Denote the possible result of the n-th toss by x_n, where head has probability π, and tail has probability $1-\pi$. Let y_N be the mean value after $N \gg 1$ tosses,

$$y_N = \frac{1}{N}\sum_{n=1}^{N} x_n. \tag{1.30}$$

The number of ways in which K heads occur in N tosses is $N!/[K!(N-K)!]$, therefore, the exact binomial distribution yields

$$P\left(y_N = \frac{2K}{N} - 1\right) = \frac{N!}{K!(N-K)!}\pi^K(1-\pi)^{N-K} \ . \tag{1.31}$$

Using Stirling's approximation and writing $K = pN$ and $N - K = (1-p)N$ one obtains

$$P(y_N = 2p-1) \sim e^{-NI(\pi,p)} \ , \tag{1.32}$$

where

$$I(\pi, p) = p\ln\frac{p}{\pi} + (1-p)\ln\frac{1-p}{1-\pi} \ . \tag{1.33}$$

$I(\pi, p)$ is called "relative entropy" (or Kullback-Leibler divergence), and $I(\pi, p) = 0$ for $\pi = p$, while $I(\pi, p) > 0$ for $\pi \neq p$. It is easy to repeat the argument for the multinomial case, where $x_1, \ldots, x_N$ are independent variables that take m possible different values $a_1, a_2, \ldots, a_m$ with probabilities $\{\pi\} = \pi_1, \pi_2, \ldots, \pi_m$.[8] In the limit $N \gg 1$, the probability of observing the frequencies $\{f\} = f_1, f_2, \ldots, f_m$ is

$$P(\{f\} = \{p\}) \sim e^{-NI(\{\pi\},\{p\})}$$

where

$$I(\{\pi\},\{p\}) = \sum_{j=1}^{m} p_j \ln\frac{p_j}{\pi_j}$$

is called "relative entropy" of the probability $\{p\}$, with respect to the probability $\{\pi\}$. Such a quantity measures the discrepancy between $\{p\}$ and $\{\pi\}$ in the sense that $I(\{\pi\},\{p\}) = 0$ if and only if $\{p\} = \{\pi\}$, and $I(\{\pi\},\{p\}) > 0$ if $\{p\} \neq \{\pi\}$.

From the above computation one understands that it is possible to go beyond the central limit theory, and to estimate the statistical features of extreme (or tail) events, as the number of observations grows without bounds. Writing $I(\pi, p)$ in terms of $y = 2p - 1$, we have the asymptotic behavior of the probability density:

$$p(y) \sim e^{-N\mathscr{C}(y)} \ , \tag{1.34}$$

[8] Such a result has been obtained by Boltzmann, who firstly noted the basic role of the entropy [14].

with

$$\mathscr{C}(y) = \frac{1+y}{2}\ln\frac{1+y}{2\pi} + \frac{1-y}{2}\ln\frac{1-y}{2(1-\pi)} \,. \tag{1.35}$$

For p close to π, i.e. $y \simeq \langle y \rangle$, a Taylor expansion of $\mathscr{C}(y)$ reproduces the central limit theorem.

1.3.2 Product of Random Variables

Large deviation theory accounts for rare events pertaining to the tails of the probability density function (pdf) of the *sum* of random variables. Ironically, one of the best examples to appreciate its importance is the *product* of random variables such as

$$M_N = \prod_{k=1}^{N} \beta_k \,, \tag{1.36}$$

where $\{\beta_k\}$ are real and positive random variables. The statistical properties of the product M_N can be straightforwardly related to those of the sum of random numbers by noticing that

$$M_N = \prod_{k=1}^{N} \beta_k = e^{N\left(\frac{1}{N}\sum_{k=1}^{N} x_k\right)} = e^{N y_N} \qquad \text{with} \qquad x_k = \ln \beta_k \,, \tag{1.37}$$

where again $y_N = (1/N)\sum_{k=1}^{N} \ln \beta_k$ denotes the empirical mean. Below, we illustrate the importance of LDT product of random numbers partially following Ref. [15], using a simple example which allows us to use the results of the previous section. In particular, we can take $\beta_k = \mathrm{e}$ and e^{-1} (i.e. $x_k = \ln \beta_k = \pm 1$) with probability π and $1-\pi$, respectively, so that we can write $P(M_N = \mathrm{e}^K \mathrm{e}^{-(N-K)}) = P(y_N = 2K/N - 1)$ as given by Eq. (1.31). Therefore, we can directly compute the moments of order q

$$\langle M_N^q \rangle = (\pi \mathrm{e}^q + (1-\pi)\mathrm{e}^{-q})^N \,. \tag{1.38}$$

Using Eq. (1.34) we can write the moments as

$$\langle M_N^q \rangle = \langle \mathrm{e}^{N q\, y_N} \rangle \approx \int \mathrm{d}y\, \mathrm{e}^{-N(\mathscr{C}(y) - q\, y)} \approx \mathrm{e}^{-N \inf_y \{\mathscr{C}(y) - q\, y\}} \,, \tag{1.39}$$

where in the second equality we used the LDT result with the Cramér function $\mathscr{C}$ given by Eq. (1.35) and, in the third, we made a steepest descent estimate of the

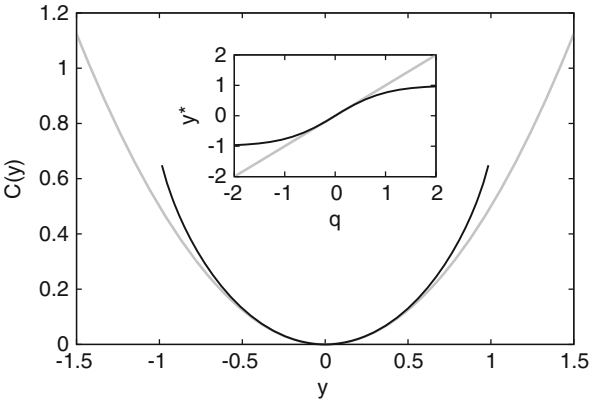

Fig. 1.1 Comparison between the Cramér function (1.35) (*black*) and its parabolic approximation $\tilde{\mathscr{C}}$ (*grey*), for $\pi = 1/2$. *Inset*: $y^*(q)$ vs q, as obtained using LDT or CLT

integral; both steps require N to be large. Using Eq. (1.35), a rather straightforward computation shows that the minimum in Eq. (1.39) is realized at $y^*(q) = (\pi e^q - (1-\pi)e^{-q})/(\pi e^q + (1-\pi)e^{-q})$, with

$$\inf_y\{\mathscr{C}(y) - q\, y\} = \mathscr{C}(y^*) - q\, y^* = -\ln\left(\pi e^q + (1-\pi)e^{-q}\right)\,, \tag{1.40}$$

so that we recover the result (1.38).

Now to appreciate the importance of rare events, we can disregard them by repeating the estimate of the moments using the CLT. In practice, this amounts to Taylor expanding $\mathscr{C}$ in (1.35) around its minimum $y = 2\pi - 1$, i.e. to approximate $\mathscr{C}$ with the parabola

$$\tilde{\mathscr{C}}(y) = \frac{(y + 1 - 2\pi)^2}{8\pi(1-\pi)}\,. \tag{1.41}$$

This approximation corresponds to assume a lognormal distribution for the product M_N [16]. The moments can be computed by finding the minimum in (1.39) with $\mathscr{C}$ replaced by its parabolic approximation $\tilde{\mathscr{C}}$. A straightforward computation gives $y^* = 2\pi - 1 + 4\pi(1-\pi)q$ and $\inf_y\{\tilde{\mathscr{C}}(y) - q\, y\} = \tilde{\mathscr{C}}(y^*) - q\, y^* = -q(2\pi - 1 + (1-\pi)q)$, which leads to moments very different from the correct ones (1.38) also for moderate values of q. Moreover, the fast growth of the moments ($\sim \exp(\text{const.}Nq^2)$) makes the lognormal distribution not uniquely determined by the values of its moments [17]. Figure 1.1 shows the Cramér function (1.39) and its parabolic approximation. The minimum position $y^*(q)$ obtained with the lognormal deviates from the correct value also for moderate values of q (see inset).

In the above example, the CLT (and thus the lognormal approximation) does not take into account the fact that y_N cannot exceed 1, which is the value corresponding to a sequence consisting of N consecutive $\beta_k = \mathrm{e}$. Such a sequence has an exponentially small probability to appear, but it carries an exponentially large contribution compared to the events described by CLT.

For an introductory discussion of LDT in multiplicative processes see Ref. [15].

1.4 Large Deviation Theory: Examples From Physics

1.4.1 Energy Fluctuations in the Canonical Ensemble

The large deviation theory finds a rather natural application in statistical mechanics, e.g. for the fluctuations of the energy e per particle in a system of N particles at temperature T:

$$p(e) \simeq \frac{1}{\mathscr{Z}_N} \exp\{-N\beta[e - Ts(e)]\} \ ,$$

where $s(e)$ is the microcanonical entropy per particle. Since $\int p(e)de = 1$, the constant $\mathscr{Z}_N$ (partition function) turns out to be

$$\mathscr{Z}_N \sim \exp\{-N\beta f(T)\} \ ,$$

where $f(T)$ is the free energy per particle

$$f(T) = \min_e \{e - Ts(e)\} \ .$$

The value e^* for which the function $e - Ts(e)$ reaches its minimum is determined by

$$\frac{1}{T} = \frac{\partial s(e)}{\partial e} \ , \tag{1.42}$$

i.e. it is the value such that the corresponding microcanonical ensemble has temperature T. It is rather obvious what is the Cramér function and its physical meaning:

$$\mathscr{C}(e) = \beta[e - Ts(e) - f(T)] \ .$$

Let us note that the value of e such that $\mathscr{C}(e)$ is minimum (zero) is nothing but $e^* = \langle e \rangle$ given by (1.42). The Gaussian approximation around e^* is

$$\mathscr{C}(e) \simeq \frac{1}{2}\mathscr{C}''(e^*)(e - e^*)^2 \ ,$$

and therefore $\langle (e - e^*)^2 \rangle = 1/[N\mathscr{C}''(e^*)]$, since

$$\langle (e - e^*)^2 \rangle = \frac{k_B}{N} T^2 c_V \ ,$$

where $c_V = \partial \langle e \rangle / \partial T$ is the specific heat per particle. The convexity of the Cramér function has a clear physical meaning: $c_V(T)$ must be positive. The case of non-convex Cramér function corresponds to phase transitions, i.e. non-analytic $f(T)$.

1.4.2 Multiplicative Cascade in Turbulence

Turbulent flows are characterized by fluctuations over a wide range of scales, with a disordered alternation of quiescent regions and sparse bursting events—*intermittency* [18]. Intermittency of this kind is well captured by multiplicative processes which, in turbulence, find their justification in the phenomenology of the energy cascade [18]: the nonlinear process by which velocity fluctuations flow from the large scales (of injection) to the small ones, where they are dissipated by molecular diffusion. The book in Ref. [18] provides a detailed discussion of turbulence within the framework of LDT (and the multifractal model). Here we just illustrate a simple d-dimensional multiplicative process, inspired to turbulence, able to generate an intermittent signal similar to those experimentally observed.

At step $N = 0$ consider a (mother) hypercube of side ℓ_0 (the forcing scale) where energy dissipation is nonrandom and equal to ϵ_0. The $N = 1$ step is obtained subdividing the hypercube in 2^d (daughter) hypercubes of side $\ell_0/2$ (powers of 2 are just for simplicity). In each daughter hypercube the energy dissipation is obtained by multiplying ϵ_0 by independent random variables $w \geq 0$ (such that $\langle w \rangle = 1$ and $\langle w^q \rangle < \infty$ for any $q > 0$). At the n-th step we thus have 2^{Nd} hypercubes of side $\ell_N = \ell_0 2^{-N}$, with energy dissipation

$$\epsilon_N = w_N \epsilon_{N-1} = \prod_{k=1}^{N} w_k \epsilon_0 \,. \tag{1.43}$$

Although the prescription $\langle w \rangle = 1$ ensures that $\langle \epsilon_N \rangle = \epsilon_0$, the multiplicative process is non-conservative, i.e. the value of the energy dissipation of a specific hypercube of side ℓ_N is not equal to the sum of the energy dissipation in the daughters hypercube at scale $\ell_N/2$. Moreover, as discussed in Sect. 1.3.2, large fluctuations are typical of product of random variables, so that we can expect that for N large intermittency shows up. For instance, the choice

$$w = \begin{cases} \beta^{-1} & \text{with prob.} \quad \beta \\ 0 & \text{with prob.} \quad 1-\beta \end{cases} \qquad 0 < \beta \leq 1 \,, \tag{1.44}$$

corresponds to a popular model known as β-model for turbulence [19]. Clearly, with (1.44) at the N-th step energy dissipation will be different from zero only in a fraction $\beta^N = 2^{N \log_2 \beta} = (\ell_0/\ell_N)^{\log_2 \beta}$ of the $2^{Nd} = (\ell_0/\ell_N)^d$ hypercubes, in other terms energy dissipation will distribute on a fractal of dimension $D_F = d - \log_2(1/\beta)$. This qualitatively explains the sparseness of bursting events. However, whenever different from zero energy dissipation will be equal to $\beta^{-N}\epsilon_0$. Therefore, to account for the unevenness of energy dissipative values in each hypercube where it is different from zero, one possibility is to generalize (1.44) by assuming that β is not a fixed value but a realization of i.i.d. random variables with a given pdf $p(\beta)$ [20]. Essentially this leads the energy dissipation to be a multifractal measure [18], which can be characterized in terms of the moments

$$\langle \epsilon_N^q \rangle = \int \prod_{k=1}^{N} \beta_k \, p(\beta_k) d\beta_k \epsilon_0^q \beta_k^{-q} \,. \tag{1.45}$$

To compute the moments in the limit $N \to \infty$ we can proceed similarly to (1.39). In particular, we have $\epsilon_N = \epsilon_0 (\ell_N/\ell_0)^{-y_N}$ with $y_N = \sum_{k=1}^{N} \log_2 \beta_k / N$. LDT implies that $p(y_N = y) \sim (\ell_N/\ell_0)^{\mathscr{C}(y)/\log 2}$, so that estimating the integral in (1.45) with the saddle point method we obtain

$$\langle \epsilon_N^q \rangle = \epsilon_0^q \left(\frac{\ell_N}{\ell_0} \right)^{\tau_q} \quad \text{with} \quad \tau_q = \inf_y \left\{ \frac{\mathscr{C}(y)}{\log 2} - y(q-1) \right\} \,. \tag{1.46}$$

In general, τ_q will be a nonlinear function of q: the signature of multifractality and intermittency. Conversely, in the model (1.44) with β non-random, $\tau^q = (q-1)(d - D_F)$ is a linear function. The exponents τ_q is linked to the scaling behavior of moments of the difference of velocities, the so-called structure functions, which are directly accessible experimentally. As shown in Ref. [20] a careful choice of $p(\beta)$ allows for reproducing the behavior of the structure functions' exponents which display a seemingly universal nonlinear dependence on q.

1.4.3 Chaotic Systems

The most characterizing feature of chaotic systems is the sensitive dependence on initial conditions: starting from nearby initial conditions, trajectories exponentially diverges. The classical indicators of the degree of instability of trajectories are the Lyapunov Exponents (LE), that quantify the mean rate of divergence of trajectories which start infinitesimally close. For the sake of simplicity we consider a $1d$ discrete time dynamical system

$$x(t+1) = f(x(t)) \tag{1.47}$$

and given an initial condition $x(0)$, we look at two trajectories, $x(t)$ and $\tilde{x}(t)$ starting from $x(0)$ and $\tilde{x}(0) = x(0) + \delta x(0)$, respectively, where $|\delta x(0)| \ll 1$. Denoting with $\delta x(t) = |x(t) - \tilde{x}(t)|$ the distance between the two trajectories, we expect that for non-chaotic systems $|\delta x(t)|$ remains bounded or increases algebraically in time, while for chaotic systems it grows exponentially

$$|\delta x(t)| = |\delta x(0)| e^{\gamma t} \,, \tag{1.48}$$

where

$$\gamma = \frac{1}{t} \ln \frac{|\delta x(t)|}{|\delta x(0)|} \,, \tag{1.49}$$

is the local exponential rate of divergence between trajectory.

The Maximum Lyapunov Exponent, characterizing the sensitivity to initial conditions, is defined by the limit

$$\lambda_{\max} = \lim_{t\to\infty} \lim_{|\delta x(0)|\to 0} \frac{1}{t} \ln \frac{|\delta x(t)|}{|\delta x(0)|} . \tag{1.50}$$

Note that γ is fluctuating while $\lambda_{\max}$ is a non-fluctuating quantity, but it can depend on $x(0)$. It is easy to understand that the existence of the limit in Eq. (1.50) is a generalization of LLN for dependent variables. In order to obtain $\delta x(t)$ from $\delta x(t-1)$, in the case of an infinitesimal $|\delta x(t-1)|$ one can use a simple Taylor expansion of the first order and the local exponent γ can be computed as

$$\gamma = \frac{1}{t} \ln \frac{|\delta x(t)|}{|\delta x(0)|} = \frac{1}{t} \sum_{k=1}^{t} \ln |f'(x(k-1))| . \tag{1.51}$$

The Maximum Lyapunov Exponent is nothing but

$$\lambda_{\max} = \lim_{t\to\infty} \frac{1}{t} \sum_{k=1}^{t} \ln |f'(x(k-1))| ,$$

and, if the system is ergodic, it does not depend on $x(0)$. Moreover, in simple cases, it is possible to obtain also the Cramér function of γ. Let us consider the tent map

$$x(t+1) = f(x(t)) = \begin{cases} \dfrac{x(t)}{p} & 0 \le x(t) < p \\ \dfrac{1-x(t)}{1-p} & p \le x(t) \le 1, \end{cases} \tag{1.52}$$

with $p \in (0,1)$. The derivative of the map takes only two values, $1/p$ and $1/(p-1)$, moreover the map can be shown to generate a memory-less process so that the sum (1.51) can be interpreted as the sum of Bernoullian random variables

$$\xi_j = \begin{cases} -\ln p & \text{with prob. } p \\ -\ln(1-p) & \text{with prob. } 1-p. \end{cases}$$

Therefore the effective Lyapunov exponent on a time interval t is

$$\gamma(t) = -\frac{k \ln p + (t-k)\ln(1-p)}{t} \qquad \text{with prob.} \qquad \binom{t}{k} p^k (1-p)^{t-k} ,$$

where k is the number of occurrences of $\xi_j = -\ln p$. Using the Stirling approximation, with some algebra it is possible to obtain the probability of the occurrence

of γ in a time interval t as $P_t(\gamma) \simeq \exp(-t\mathscr{C}(\gamma))$ where the Cramér function is given by

$$\mathscr{C}(\gamma) = \left[\frac{\gamma + \ln(1-p)}{\ln \frac{1-p}{p}} \ln \left(\frac{\gamma + \ln(1-p)}{p \ln \frac{1-p}{p}} \right) - \frac{\gamma + \ln p}{\ln \frac{1-p}{p}} \ln \left(- \frac{\gamma + \ln p}{(1-p) \ln \frac{1-p}{p}} \right) \right]. \tag{1.53}$$

The Cramér function has its minimal value in $\gamma = -p \ln p - (1-p)\ln(1-p)$ (where it also vanishes) which is the Maximum Lyapunov Exponent, and the Taylor expansion of Eq. (1.53) around this minimum provides the Central Limit Theorem for the sum (1.51). Unfortunately this computation can be performed almost only for piecewise linear maps.

For generic dynamical systems

$$\mathbf{x}(t+1) = \mathbf{f}(\mathbf{x}(t))$$

there exists a theorem due to Oseledec that under very general hypothesis, states the existence of the Lyapunov exponents. But a major difficulty arises, i.e., the product of Eq. (1.51) cannot be factorized because of the non commutativity of the Jacobian matrix with entries $A_{ij} = \partial f_i / \partial x_j$.

1.4.4 Disordered Systems

Products of matrices and Oseledec's limit theorem find a natural application to the study of statistical mechanics of disordered systems. Indeed, their thermodynamical properties can be recast, via transfer matrix formalism, as the evaluation of the asymptotic properties of products of matrices. The presence of randomness induced by disorder introduces sample to sample fluctuations of observables which require proper averaging procedures over different disorder realizations. In this case the transfer-matrix approach involves products of *random matrices*.

As an example, which already includes all the difficulties, consider an array of N binary variables $\sigma_i = \pm 1$ (spins) whose interaction is defined by the Hamiltonian,

$$H(\boldsymbol{\sigma}) = -J \sum_{i=1}^{N} \sigma_i \sigma_{i+1} + \sum_{i=1}^{N} h_i \sigma_i, \tag{1.54}$$

where J determines a ferromagnetic internal coupling between nearest neighbor sites and $\{h_i\}_{i=1}^N = \mathbf{h}$ represent a set of local magnetic fields acting on site, each independently extracted from a distribution $\rho(h)$. Typically, $\rho(h)$ is chosen to be a Gaussian or a bimodal distribution and usually periodic boundary conditions are assumed, $\sigma_{i+N} = \sigma_i$. For finite N, the system has 2^N possible configurations, however we are interested in the thermodynamic limit $N \to \infty$, where extensive

thermodynamics quantities becomes independent of N and the choice of boundary conditions is irrelevant.

Once a given realization of the disorder is assigned, $\mathbf{h} = \{h_i\}_{i=1}^{N}$, the equilibrium thermodynamics of the N-spin chain is determined by the free-energy:

$$f_N(\beta, \mathbf{h}) = -\frac{1}{\beta N} \ln Z_N(\beta, \mathbf{h}), \tag{1.55}$$

where

$$Z_N(\beta, \mathbf{h}) = \sum_{\sigma_1} \dots \sum_{\sigma_N} e^{-\beta(J\sigma_1\sigma_2 - h_1\sigma_1)} \cdots e^{-\beta(J\sigma_N\sigma_1 - h_N\sigma_N)} \tag{1.56}$$

is the partition function of the system, the summation covers all the 2^N spin configurations and $\beta = 1/(k_B T)$. In principle, the free-energy (1.55) for every finite N is a random variable, because it depends on the disorder realizations, however, as we shall see in the transfer matrix formalism, a straightforward application of Oseledec's limit theorem implies that

$$\lim_{N\to\infty} f_N(\beta, \mathbf{h}) = \lim_{N\to\infty} -\frac{1}{\beta N} \langle \ln Z_N(\beta, \mathbf{h}) \rangle_h, \tag{1.57}$$

where the average $\langle \dots \rangle_{\mathbf{h}}$ is meant over the random field distribution. Result (1.57) can be interpreted as follows, in the thermodynamic limit $f_N(\beta, \mathbf{h})$ is a non-random quantity as it converges to its limit average over the disorder, for almost all disorder configurations. In an equivalent physical language, when $N \to \infty$, f_N is practically independent of $\{h_i\}_{i=1}^{N}$, and it is a self-averaging observable with respect to sample to sample disorder fluctuations.

The transfer matrix approach amounts to re-writing the partition function

$$Z_N = \sum_{\{\sigma_i\}} \prod_{i=1}^{N} e^{-\beta(J\sigma_i\sigma_{i+1} - h_i\sigma_i)} = Tr\left(\prod_{i=1}^{N} \mathbf{T}[i] \right) \tag{1.58}$$

as an iterated matrix product in indexes $\sigma_2, \sigma_3, \dots, \sigma_N$ and the summation over σ_1 as a trace operation, where the 2×2 fundamental matrix $\mathbf{T}[i]$ has entries: $T(\sigma_i, \sigma_{i+1}) = \exp[\beta(J\sigma_i\sigma_{i+1} - h_i\sigma_i)]$, more explicitly:

$$\mathbf{T}[i] = \begin{bmatrix} e^{\beta(J-h_i)} & e^{-\beta(J+h_i)} \\ e^{-\beta(J-h_i)} & e^{\beta(J+h_i)} \end{bmatrix} . \tag{1.59}$$

In the thermodynamic limit, the free energy per spin is given by the maximum Lyapunov exponent λ_1 of the product of matrices in Eq. (1.58):

$$-\beta f(\beta, \mathbf{h}) = \lim_{N\to\infty} \frac{1}{N} \ln\left[Tr\left(\prod_{i=1}^{N} \mathbf{T}[i] \right) \right] = \lambda_1. \tag{1.60}$$

When the field $h_i = H$ is the same on every site (no disorder), the computation of free-energy is particularly simple because the product involves identical symmetric matrices: $Z_N = Tr(\mathbf{T}^N) = \mu_+^N + \mu_-^N$ being $\mu_\pm$ the eigenvalues of $\mathbf{T}$. Therefore the free energy coincides with the logarithm of the maximum eigenvalue of $\mathbf{T}$: $f(\beta, H) = -\beta^{-1} \ln(\mu_+)$, where

$$\mu_\pm = e^{\beta J} \cosh(\beta H) \pm \sqrt{\cosh^2(\beta H) 2e^{2\beta J} \sinh(2\beta J)}.$$

When h_i is not constant, the matrices (1.59) are not commuting, and the asymptotic behavior of the random matrix product has to be numerically evaluated. Practically, one resorts to compute the exponential growth rate of an arbitrary initial vector $\mathbf{z}_0 = (u_0, v_0)$, with positive components, under the effect of the iterated matrix multiplication $\mathbf{z}_{n+1} = \mathbf{T}[n]\mathbf{z}_{n-1}$,

$$\lambda_1 = \lim_{N\to\infty} \frac{1}{N} \ln\left(\frac{|\mathbf{z}_N|}{|\mathbf{z}_0|} \right) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \ln\left(\frac{|\mathbf{z}_{n+1}|}{|\mathbf{z}_n|} \right).$$

Oseledec's theorem grants that, under rather general conditions, the above limit exists and it is a non-random quantity (self averaging property). Then, the computation of free-energy of a one-dimensional random field Ising model to some extent constitutes a physical example of the application of the law of large numbers. Moreover the self-averaging property of the free-energy in the context of disordered systems corresponds to the ergodicity condition for dynamical systems.

A large deviation approach can be formulated also for the fluctuations of the free-energy of a random field Ising model at finite N around its thermodynamic limit value. The transfer random-matrix formalism makes the characterization of large deviations an application of the generalized Lyapunov exponents. It is easy to compute the asymptotic behavior of $\langle |\mathbf{z}_n|^q \rangle$ for $q = 1, 2, 3, \ldots$ and therefore compute the generalized Lyapunov exponents

$$L(q) = \lim_{n\to\infty} \frac{1}{n} \ln \langle |\mathbf{z}_n|^q \rangle. \tag{1.61}$$

It is possible to show that $L(1)$ is the logarithm of the largest eigenvalue of $\langle \mathbf{T} \rangle$ while $L(2)$ is the logarithm of the largest eigenvalue of $\langle \mathbf{T}^{\otimes 2} \rangle$ where $\langle \mathbf{T}^{\otimes 2} \rangle$ is the tensorial product $\langle \mathbf{T} \otimes \mathbf{T} \rangle$ and so on for $\langle \mathbf{T}^{\otimes 3} \rangle$, etc. In such a way we have an exact bound $\lambda_1 \leq L(q)/q$ for $q = 1, 2, \ldots$. To consider $L(1)$ instead of λ_1 corresponds, in physical terms, to consider an annealed average, i.e. $\ln \langle Z_N \rangle$ instead of $\langle \ln Z_N \rangle$. The knowledge of $L(q)$, for all q, is equivalent to the knowledge of the Cramér function $C(\gamma)$.

1.4.5 Entropy Production in Markov Processes

A recent application of the theory of large deviations concerns the dynamical behaviour of deterministic and stochastic systems at large times.

Consider a continuous time Markov process with a finite number of states, whose evolution is such that: if the system is in state x, it remains in such a state for a random time $t \geq 0$ extracted with a probability density $p(t) = \omega(x)e^{-\omega(x)t}$ and then jumps to a new state x' with transition probability $\frac{w(x \to x')}{\omega(x)}$. The functions $w(x \to x')$ are said to be the transition rates of the Markov process and $\omega(x) = \sum_{x'} w(x \to x')$ is the total exit rate from x. It is useful to introduce also a notion of *time-reversed state* $\overline{x}$ for a given state x: for the so-called "even" variables, such as positions or forces, one has $\overline{x} \equiv x$, while for "odd" variables, such as velocities, one has $\overline{x} \equiv -x$. For what follows, a further assumption is crucial: if $w(x \to x') > 0$ then $w(\overline{x'} \to \overline{x}) > 0$.

From the above definitions, a trajectory of time-length t can be written as $\Omega_0^t = \{(x_0, t_0), (x_1, t_1), (x_2, t_2), \ldots, (x_n, t_n)\}$, where the system undergoes n jumps visiting states x_i in temporal order from $i = 0$ to $i = n$ and stays in each of them for a waiting time t_i, with $\sum_i t_i = t$. Its time-reversal reads $\overline{\Omega_0^t} = \{(\overline{x}_n, t_n), (\overline{x}_{n-1}, t_{n-1}), \ldots, (\overline{x}_2, t_2), (\overline{x}_1, t_1), (\overline{x}_0, t_0)\}$.

The probability $P_x(t)$ of finding the system in state x at time t evolves according to the master equation:

$$\frac{dP_x(t)}{dt} = \sum_{x'} P_{x'}(t) w(x' \to x) - \omega(x) P_x(t). \tag{1.62}$$

We denote by P_x^{inv} the steady state solution of (1.62). The particular steady state where $P_{x'}^{\text{inv}} w(x' \to x) = P_{\overline{x}}^{\text{inv}} w(\overline{x} \to \overline{x'})$ is a steady state which is said to satisfy *detailed balance*. The detailed balance conditions imply that the probability of occurrence of any trajectory is invariant under time-reversal $P(\Omega_0^t) = P(\overline{\Omega_0^t})$: in short, a *movie* of the system of any time-length cannot be discriminated to be played in the forward or backward direction. Markov processes describing physical systems at thermal equilibrium (or isolated), satisfy the detailed balance conditions. On the contrary, the presence of external forces and/or internal dissipation leads to steady states with physical currents, with the consequent breakdown of the detailed balance conditions.

Following a series of studies [21–24], a "fluctuating entropy production functional" has been proposed in [25] for the general case of Markov processes. The functional, for a trajectory which in the time $[0, t]$ includes n jumps, reads

$$W_t(\Omega_0^t) = \sum_{i=1}^{n} \ln \frac{w(x_{i-1} \to x_i)}{w(\overline{x_i} \to \overline{x_{i-1}})}. \tag{1.63}$$

It is immediate to verify that in a steady state satisfying detailed balance $W_t = -\ln[P^{\text{inv}}_{x_0}/P^{\text{inv}}_{\overline{x}_n}]$ and therefore—given the finiteness of the space of states—one has $\lim_{t\to\infty}\frac{W_t}{t} = 0$. Otherwise, as discussed below, $\lim_{t\to\infty}\frac{\langle W_t\rangle}{t} > 0$.

More precisely, a large deviation principle for the stochastic variable W_t can be obtained, such that its associated Cramér function satisfies a particular relation, called "Fluctuation-Relation". An instructive way to derive it is the following [25]. Let us define the joint probability $p_x(W_t, t)$ of finding the system at time t in state x with a value of the entropy production (measured starting from time 0) W_t; we also define the vector $\mathbf{p}(W_t, t) = \{p_{x_1} \dots p_{x_M}\}$ where M is the number of possible states for the system. It is not difficult to realize that its evolution is governed by a modified master equation that reads

$$\frac{dp_x(W_t,t)}{dt} = \sum_{x'} p_{x'}\left[W_t - \Delta W(x'\to x), t\right] w(x'\to x) - \omega(x)\, p_x(W_t,t). \tag{1.64}$$

With $\Delta W(x'\to x) = \ln\frac{w(x'\to x)}{w(\overline{x}\to\overline{x'})}$. If we consider the generating function for W_t conditioned to state x, i.e.

$$g_x(s,t) = \int d\, W_t e^{-sW_t}\, p_x(W_t,t), \tag{1.65}$$

we find for its time evolution, immediately descending from Eq. (1.64):

$$\frac{dg_x}{dt} = \sum_{x'} w(x'\to x) e^{-s\Delta W(x'\to x)} g_{x'}(s,t) - \omega(x) g_x(s,t) =$$

$$\sum_{x'} w(x'\to x)^{1-s} w(\overline{x}\to\overline{x'})^s g_{x'}(s,t) - \omega(x) g_x(s,t) = [L(s)\mathbf{g}(s,t)]_x \tag{1.66}$$

where we have used the definition of $\Delta W(x'\to x)$. The initial conditions for Eq. (1.66) is $g_x(s,0) = \int d\, W_t e^{-sW_t} P_x(0)\delta(W_t) = P_x(0)$, so that

$$g_x(s,t) = \sum_y [e^{Lt}(s)]_{xy} P_y(0). \tag{1.67}$$

Finally, summing over all possible states x, weighted with their probability, we get the unconditioned generating function, that reads

$$g(s,t) = \sum_{x,y} P_x(t)[e^{Lt}(s)]_{xy} P_y(0). \tag{1.68}$$

The Perron-Frobenius theorem guarantees that $L(s)$ has a unique maximal eigenvector $\tilde{g}(s) > 0$ with real eigenvalue $-\mu(s)$. This allows one to define the limit

$$\lim_{t\to\infty} -\frac{1}{t}\ln g(s,t) = \mu(s). \tag{1.69}$$

It is immediate to verify that $\mu(s)$ is the time-rescaled cumulant generating function for the steady state of the variable W_t. Its Legendre transform is the Cramér function for the large deviations of the same variable, i.e.

$$p(W_t) \approx \exp\left[t \sup_s \left(s\frac{W_t}{t} + \mu(s)\right)\right] \approx \exp[t\mathscr{C}(W_t/t)]. \tag{1.70}$$

From its definition in Eq. (1.66), it is straightforward to realize that $L^*(s) = L(1-s)$ and therefore $\mu(s) = \mu(1-s)$. This immediately reflects into the following relation for the Cramér function of $w_t = W_t/t$:

$$\mathscr{C}(w_t) - \mathscr{C}(-w_t) = w_t, \tag{1.71}$$

which is known as Steady State Fluctuation Relation (SSFR).

In the limit of an infinite space of states ($M \to \infty$) problems may arise in the derivation sketched above, when the inverse transform is operated to retrieve the large deviation rate function $\mathscr{C}(w_t)$. In some cases a modified SSFR holds true instead of Eq. (1.71): to recover the validity of formula (1.71) one has to measure a different entropy production, modified by adding so-called "boundary terms", as discussed in [26–28].

Notwithstanding the problems for unbounded spaces, the result (1.70) together with (1.71) is remarkable: the "entropy production" measured on very long trajectories tends to be sharply peaked around its average value, which is positive for non-equilibrium systems and zero otherwise. Moreover, if the trajectories have finite time-length, one can observe also *negative* fluctuations, representing a sort of "finite size violation" of the second principle of thermodynamics, but with exponentially small probability.

References

1. J. Mehra, *The Golden Age of Theoretical Physics* (World Scientific, Singapore, 2001)
2. I. Todhunter, *History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace* (BiblioBazaar, Charleston, 2009)
3. C. Cercignani, *Ludwig Boltzmann: The Man Who Trusted Atoms* (Oxford University Press, Oxford, 2007)
4. L. Graham, J.M. Kantor, *Naming Infinity: A True Story of Religious Mysticism and Mathematical Creativity* (Harvard University Press, Cambridge, 2009)
5. A. Khinchin, *Mathematical Foundations of Statistical Mechanics* (Dover, New York, 1949)
6. P. Mazur, J. van der Linden, J. Math. Phys. **4**, 271 (1963)
7. H. Furstenberg, H. Kesten, Ann. Math. Stat. **31**, 457 (1960)
8. V.I. Oseledec, Trans. Moscow Math. Soc. **19**, 197 (1968)
9. H. Fischer, *A History of the Central Limit Theorem* (Springer, Berlin, 2010)

10. W. Hoeffding, H. Robbins, Duke Math. J. **15**, 773 (1948)
11. S.R.S. Varadhan, *Large Deviations and Applications* (Society for Industrial and Applied Mathematics, Philadelphia, 1984)
12. H. Touchette, Phys. Rep. **478**, 1 (2009)
13. G. Grimmett, D. Stirzaker, *Probability and Random Processes* (Oxford University Press, Oxford, 2001)
14. R.S. Ellis, Physica D **133**, 106 (1999)
15. S. Redner, Am. J. Phys. **58**, 267 (1990)
16. J. Aitchison, J.A.C Brown, *The Lognormal Distribution* (Cambridge University Press, Cambridge, 1963)
17. S.A. Orszag, Phys. Fluids **13**, 2211 (1970)
18. U. Frisch, *Turbulence: The Legacy of A. N. Kolmogorov* (Cambridge University Press, Cambridge, 1995)
19. E.A. Novikov and R.W. Stewart, Isv. Akad. Nauk. USSR Ser. Geophys. **3**, 408 (1964)
20. R. Benzi, G. Paladin, G. Parisi, A. Vulpiani, J. Phys. A **17**, 3521 (1984)
21. D.J. Evans, E.G.D. Cohen, G.P. Morriss, Phys. Rev. Lett. **71**, 2401 (1993)
22. G. Gallavotti, E.G.D. Cohen, Phys. Rev. Lett. **74**, 2694 (1995)
23. D.J. Evans, D.J. Searles, Adv. Phys. **51**, 1529 (2002)
24. J. Kurchan, J. Phys. A.: Math. Gen. **31**, 3719 (1998)
25. J.L. Lebowitz, H. Spohn, J. Stat. Phys. **95**, 333 (1999)
26. R. van Zon, E.G.D. Cohen, Phys. Rev. Lett. **91**, 110601 (2003)
27. F. Bonetto, G. Gallavotti, A. Giuliani, F. Zamponi, J. Stat. Phys. **123**, 39 (2006)
28. A. Puglisi, L. Rondoni, A. Vulpiani, J. Stat. Mech. P08010 (2006)

http://www.springer.com/978-3-642-54250-3

Large Deviations in Physics
The Legacy of the Law of Large Numbers
Vulpiani, A.; Cecconi, F.; Cencini, M.; Puglisi, A.; Vergni, D. (Eds.)
2014, XIV, 314 p. 58 illus., 18 illus. in color., Softcover
ISBN: 978-3-642-54250-3