

STOCHASTIC PROCESSES

Master 2 Physics of Complex Systems



Christophe Texier



Contents

0	Introduction	3
1	Probability : some useful concepts	3
1.1	Events, probability and random variable	3
1.2	Stochastic independence – Joint and conditional probabilities	7
1.3	Mean, variance, moments and generating function	9
1.4	Variance and cumulants	11
1.5	Multivariate Gaussian distribution	14
1.6	The central limit theorem	15
1.7	Large deviations : typical and atypical fluctuations	18
1.8	Generalization of the central limit theorem for power law distributions	22
1.9	Extreme value statistics	27
	Additional exercises	34
2	The Langevin equation for a particle in a fluid	38
2.1	Fluctuations and Langevin force	38
2.2	A fluctuation-dissipation relation	40
2.3	Diffusion	41
2.4	Large scale properties and the overdamped regime	41
2.5	The free Brownian motion (the Wiener process)	43
	Appendix : Microscopic foundations of the Langevin equation	46
	Additional exercises	51
3	Markov processes and the master equation	52
3.1	Generalities : joint probabilities and conditional probabilities	52
3.2	Markov processes	53
3.3	Master equation	56
3.4	Classification of Markov processes	63
3.5	Spectral decomposition - Relaxation	65
3.6	Markov chains	67
3.7	Beyond the master equation : renewal processes	72
3.8	Fourier analysis of stochastic processes – Wiener-Khintchin theorem	73
	Additional exercises	76

4	Stochastic differential equations	79
4.1	SDE with drift and additive noise	79
4.2	SDE with multiplicative noise : Itô or Stratonovich ?	80
	Historical note on Döblin-Itô calculus	90
	Appendix : Stochastic integrals	90
	Additional exercises	92
5	The Fokker-Planck equation	96
5.1	Interpretation of the Fokker-Planck equation	96
5.2	From the master equation to the Fokker-Planck equation	98
5.3	The generator of the diffusion and the stationary state	101
5.4	Relaxation and spectral analysis	104
5.5	Forward and backward FPE	111
5.6	Boundary conditions for the FPE	111
5.7	Random walk on the lattice : persistence and recurrence	114
5.8	First passage and exit problem (in 1D)	121
	Additional exercises	126
6	Functionals	131
6.1	Introduction/motivations	131
6.2	Path integrals	133
6.3	Functionals of the Brownian motion	136
6.4	Functional for an arbitrary diffusion	143
6.5	First passage Brownian functionals (not discussed)	146
	Additional exercises	147
A	Mathematical tools	148
A.1	Linear differential equations with source terms	148
A.2	Fourier transform	148
A.3	Poisson formulae	150
A.4	Laplace transform	151
A.5	Asymptotics of integrals	151
A.6	Special functions	153
B	Solutions of some of the exercises	156
	References	229
	Index	233

Prerequisites :

- Equilibrium statistical physics (ensemble theory)
- Math: Fourier analysis, Laplace transform, residue theorem,...

0 Introduction

This set of lectures is devoted to stochastic processes and out-of-equilibrium statistical physics.

Equilibrium statistical physics provides a well defined procedure to study the thermodynamic properties of systems with complex dynamics. The main idea is to replace the study of the complex dynamics of the system, i.e. how its state $\vec{\Gamma}(t)$ evolves in time (here $\vec{\Gamma}$ represents a point in phase space), by some statistical information, i.e. the probability $\rho(\vec{\Gamma})$ to find the system in a given state. The first approach would require to solve a macroscopic number of differential equations, while the beauty of the second approach lies on the fact that the determination of the probability density relies on very few information, what can be understood as a result of a maximum entropy principle. The choice of the distribution, microcanonical, canonical, grand canonical, etc, is driven by physical considerations or simply by convenience.

Out-of-equilibrium statistical physics requires a statistical treatment of the dynamics, what can be achieved by various approaches, phenomenological or microscopic. On the more phenomenological side : the Langevin equation, the master equation and the Fokker-Planck equation provide different approaches for the analysis of stochastic processes. On the more microscopic side : kinetic equations (BBGKY hierarchy, Boltzmann equation, Vlasov equation, hydrodynamic equations, out-of-equilibrium Green's function –Keldysh technique,...). Note that the frontier between phenomenological and microscopic is not so sharp, as we will see by deriving a Langevin equation from a microscopic model (§ page 46).

In these lectures, I will describe the main concepts allowing for a probabilistic analysis of stochastic processes, without dwelling on the microscopic origin of the dynamics.

1 Probability : some useful concepts

Let us start with an introduction of several useful concepts of probability theory, with a physicist's point of view. For a mathematical monograph, see the standard book of Feller [26].

1.1 Events, probability and random variable

Whenever the dynamics of a system is too complex and its evolution is unpredictable, it becomes more efficient to rely on a probabilistic description. This is for example the case when one plays heads or tails,¹ as the motion of the coin in the fluid is too difficult to predict and can be considered random.

Consider an event ω belonging to the set of events Ω (e.g. $\Omega = \{\text{head}, \text{tail}\}$ when tossing a coin). We denote by $\mathcal{P}(\omega) \geq 0$ the probability of occurrence of the event, which is the frequency of occurrence of ω when a large number of observations is made. Obviously, $\mathcal{P}(\Omega) = 1$.

For a physicist, an event can be identified by making an observation, i.e. a measure of an “observable” whose value depends on the event, $X(\omega)$. Hence several observations would lead to a random sequence of values and $X(\omega)$ is called a “random variable”. Mathematically, a random variable is a function defined over a certain space (the set of events). Rather to manipulate the distribution of events $\mathcal{P}(\omega)$, we will introduce the distribution of the random variable

$$P(x) = \text{Proba}\{X(\omega) = x\} = \sum_{\omega \in \Omega} \mathcal{P}(\omega) \delta_{X(\omega), x} \quad (1.1)$$

that is the sum of probabilities for all events which realize $X(\omega) = x$. Importantly, the distribution satisfies the normalisation condition

$$\sum_x P(x) = \mathcal{P}(\Omega) = 1. \quad (1.2)$$

¹ “pile ou face” ; “testa o croce”.

For a random variable varying continuously, the Kronecker $\delta_{X(\omega),x}$ is replaced by a Dirac $\delta(X(\omega) - x)$ and the sum by an integral, then $P(x)$ is a *density*.

Probability in mathematics (if you want to read the first lines of math's papers) : Mathematicians have axiomatised probability theory.

- They first introduce the concept of σ -algebra. Considering a given set Ω (the set of events), one first define the family $\mathcal{A}$ of all subsets of Ω , such that (i) $\emptyset \in \mathcal{A}$, (ii) if $A \in \mathcal{A}$, the complementary of A is also in $\mathcal{A}$, (iii) if $A_1, A_2, \dots \in \mathcal{A}$, then $\bigcup_n A_n \in \mathcal{A}$. We say that “ $\mathcal{A}$ is a σ -algebra on Ω ”.

- The probability axioms are : (i) $\mathcal{P}(A) \geq 0$ for all $A \in \mathcal{A}$, (ii) $\mathcal{P}(\Omega) = 1$, (iii) if $A_1, A_2, \dots \in \mathcal{A}$ is a countable set of nonoverlapping sets ($A_i \cap A_j = \emptyset, \forall i \neq j$), then

$$\mathcal{P}\left(\bigcup_i A_i\right) = \sum_i \mathcal{P}(A_i)$$

- $(\Omega, \mathcal{A}, \mathcal{P})$ is called a probability space.

If you like this style, you can look at Øksendal's book [57].

Example 1 : Bernoulli distribution.— Consider a random variable $\xi \in \{0, 1\}$ (e.g. when tossing a coin, the two values correspond to head and tail). The Bernoulli distribution is $\text{Proba}\{\xi = 1\} = p$ and $\text{Proba}\{\xi = 0\} = 1 - p$.

Example 2 : Binomial distribution.— Play to head or tail N times and denote H_N the number of heads, p being the probability to get “head” ($p = 1/2$ for an unbiased coin and $p \neq 1/2$ for a biased coin). The distribution of H_N is the binomial distribution

$$\text{Proba}\{H_N = n\} = C_N^n p^n (1 - p)^{N-n} \quad (1.3)$$

Note that the random variable may be decomposed as the sum of N independent Bernoulli variables (the n -th one counts if one gets head at step n) : $H_N = \sum_{n=1}^N \xi_n$.

Example 3 : Geometric distribution (first success for Bernoulli trials).— Consider a sequence of Bernoulli trials (each trial successes with probability p or fails with probability $q = 1 - p$). The probability that the first success arises at the n -th trial is

$$\mathcal{P}_n = q^{n-1} p. \quad (1.4)$$

This is the “*geometric distribution*” ; it is obviously normalised $\sum_{n=1}^{\infty} \mathcal{P}_n = 1$.

✎ **Exercise 1.1 – Lifetime of an apparatus :** Each time an apparatus is used, it has a probability $p \ll 1$ to break down. What is the distribution of its lifetime $T \in \mathbb{N}^*$? What is the average lifetime $\langle T \rangle$ and its variance $\text{Var}(T)$?

Example 4 : Poisson distribution (counting uncorrelated events).— The distribution of the number of occurrences of uncorrelated events occurring with constant rate during a finite time

$$\text{Proba}\{X = n\} = \frac{q^n}{n!} e^{-q} \quad (1.5)$$

For example, this is the probability that n drops of rain fall into a definite area during a specific amount of time.

Example 5 : Gaussian distribution.— It is characterised by two parameters

$$g_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-(x-\mu)^2/(2\sigma^2)} . \quad (1.6)$$

This is the distribution of one component v_x of the velocity of molecules in a classical fluid (then $\mu = 0$ and $\sigma^2 = k_B T/m$).

Example 6 : Cauchy distribution.— A distribution with power law tail

$$C_{\mu,a}(x) = \frac{a/\pi}{(x - \mu)^2 + a^2} . \quad (1.7)$$

Example 7 : Pareto distribution.— A positive random variable with power law distribution

$$P(x) = \mu x^{-\mu-1} \quad \text{for } x \geq 1 \quad (1.8)$$

for exponent $\mu > 0$.

The Italian scientist Vilfredo Pareto (1848-1923) was interested in wealth distribution. Several examples exhibiting Paretian tails are shown in figures 1.1 and 1.2.

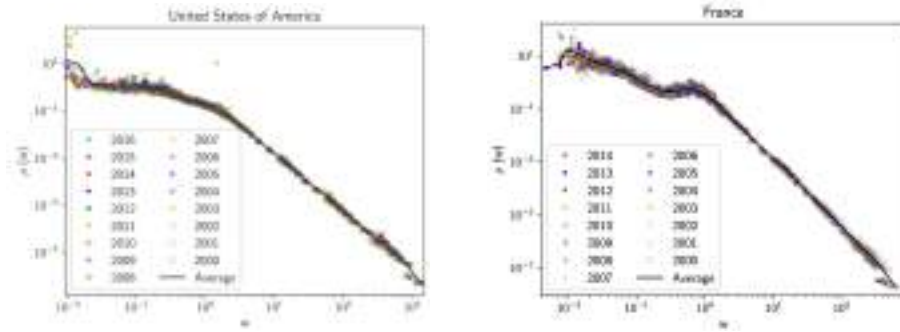


Figure 1.1: *Distribution of normalised wealth ($w = W/\langle W \rangle$) in USA and in France ; tail $\rho(w) \propto w^{-1-\mu}$ characterized by exponents $\mu = 1.4$ and $\mu = 1.68$, respectively. Note that wealth can be negative (debts) ; this part of the distribution is not shown. From I. Gere *et al.*, Physica A **581**, 126194 (2021).*

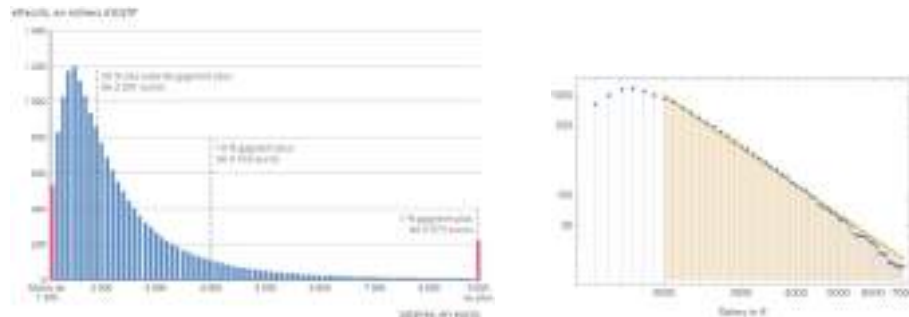


Figure 1.2: *Distribution of French salaries (net) in the private sector in 2022 (source : INSEE). The right figure (in log-log) is an attempt of fit, leading to exponent $\mu \approx 2$.*

Transformation of random variables

Sometimes we are interested in the distribution of a function of a random variable whose distribution is known. For example, the distribution of the velocity v_x of the molecules in a gas is well known (Maxwell distribution), and one could be interested in the distribution of the kinetic energy $\frac{1}{2}mv_x^2$. Let us consider a random variable X with known distribution $p(x)$. Given the monotonous function $\varphi(x)$, what is the distribution $q(y)$ of

$$Y = \varphi(X) ? \quad (1.9)$$

(the case of a non monotonous φ is not much more difficult).

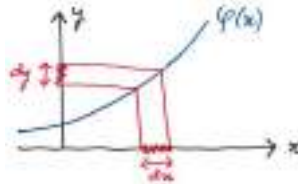


Figure 1.3: $\text{Proba}\{X \in [x, x + dx]\}$ equals $\text{Proba}\{Y \in [y, y + dy] = [\varphi(x), \varphi(x) + d\varphi(x)]\}$.

It is important to remember that *only probabilities can be made equal* (not densities, which may have different physical dimensions) : $p(x) dx = \text{Proba}\{X \in [x, x + dx]\} = \text{Proba}\{Y \in [\varphi(x), \varphi(x) + d\varphi(x)]\} = q(\varphi(x)) \varphi'(x) dx$, i.e. $p(x) = q(\varphi(x)) \varphi'(x)$. To be short, remember

$$\boxed{p(x) dx = q(y) dy \quad \text{with } y = \varphi(x)} \quad (1.10)$$

Note that it can also be convenient to write

$$q(y) = \langle \delta(y - \varphi(X)) \rangle = \int dx p(x) \delta(y - \varphi(x)) \quad (1.11)$$

(this form allows to deal with the case when φ is a non monotonous function).

✎ **Exercise 1.2 – Generate random numbers with a computer :** A computer generates a random number Y with a box distribution $B(y) = 1$ for $y \in [0, 1]$ and $B(y) = 0$ otherwise. Consider a monotonously inscreasing function

$$\varphi : [a, b] \mapsto [0, 1]$$

where the interval $[a, b]$ is arbitrary (boundaries can be sent to $\pm\infty$).

a) What is the distribution of X such that $Y = \varphi(X)$?

b) Deduce a method to generate a random number X with arbitrary distribution $p(x)$ from the computer random number Y .

c) Application : how to generate a positive random variable with exponential distribution $p(x) = \lambda e^{-\lambda x}$?

✎ **Exercise 1.3 – Applications for several distributions with power law tails :** Using the outcome of Exercise 1.2, explain how, using the random generator of a computer (from a box distribution), one can generate random variables with the following distributions :

a) Pareto distribution :

$$p(x) = \mu x^{-1-\mu} \quad \text{for } x \geq 1. \quad (1.12)$$

b) Same question for the symmetric distribution $q(x) = \frac{\mu}{2}(1 + |x|)^{-1-\mu}$ defined on $\mathbb{R}$.

c) Cauchy distribution : same question for $\mathcal{C}(x) = \pi^{-1}(x^2 + 1)^{-1}$ for $x \in \mathbb{R}$.

d) Student distribution : same question for the Student distribution $S^{(2)}(x)$ for $x \in \mathbb{R}$ (cf. Student distribution is defined in Exercise 1.36).

1.2 Stochastic independence – Joint and conditional probabilities

Let us consider A and B , two subsets of Ω . The *joint probability* $P(A \cap B)$ is the probability that both A and B occur, which is possible if the two sets share some events. The *conditional probability* is the probability that A occurs, given that B has occurred :

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (1.13)$$

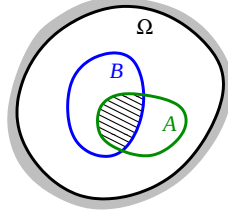


Figure 1.4: It is easy to visualize the meaning of (1.13) : imagine that the probability is proportional to the area of the ensemble of events drawn on the figure. Then $P(A)$ is identified with $\text{Area}(A)/\text{Area}(\Omega)$ and the conditional probability $P(A|B)$ is the ratio $\text{Area}(A \cap B)/\text{Area}(B)$.

As a consequence it is clear that

$$P(A \cap B) = P(A|B) P(B) = P(B|A) P(A) \quad (\text{Bayes theorem}). \quad (1.14)$$

If A and B are **stochastically independent**, the conditional probability is independent of B , hence $P(A|B) = P(A)$, i.e. $P(A \cap B) = P(A) P(B)$.



Figure 1.5: Parisian mice during the 2024 olympic games.... and a cat. $P(n|1)$ is the distribution of the mouse number, when there is one cat. © Eiffel tower photo by B. Rega.

It is probably more intuitive to define the concept for random variables. Let us consider X and Y two random variables. For example the number of mice and cats in an ecological system, which both evolve over time at random. We denote by

$$P(m, c) = \text{Proba}\{\# \text{ of mice} = m \ \& \ \# \text{ of cats} = c\} \quad (1.15)$$

the *joint probability* to have m mice and c cats at a certain time ; we assume for simplicity that the two populations are stable, i.e. that the joint distribution does not depend on time. The *marginal distribution* of the mouse number is

$$p(m) = \sum_c P(m, c) \quad (1.16)$$

(probability to have m mice, irrespectively of the number of cats). We denote $q(c) = \sum_m P(m, c)$ the marginal distribution of the number of cats. We then introduce the conditional probability

$$P(m | c) = \frac{P(m, c)}{q(c)} \quad (1.17)$$

which is the distribution of the population of mice given that the number of cats equals c . Note that it is normalised

$$\sum_m P(m | c) = 1. \quad (1.18)$$

Let us now discuss two different situations :

- If we consider mice in Paris and cats in Torino, we expect the two random variables to be independent,² hence $P(m | c)$ is independent of c (the Turin cats have no influence on the population of Parisian mice), i.e. $P(m | c) = p(m)$. In other terms, in this case we have

$$\boxed{P(m, c) = p(m) q(c) \quad \forall m, c \quad : \text{stochastic independence}} \quad (1.19)$$

- If instead we focus on Parisian cats and mice, we expect that the two populations interact and their numbers are not independent : the more there are cats to eat mice, the smaller is the mouse population.³ Hence the conditional probability $P(m | c)$ does not coincide with $p(m)$ and

$$\exists (m, c) \text{ such that } P(m, c) \neq p(m) q(c) \quad (1.20)$$

In this case we expect that the distribution $P(m | c)$ becomes more concentrated on small values of m when then number of cats increases (they eat more mice). Instead, for small c , the distribution is expected to be broader.

Exercise 1.4 – Conditional probability for 2 Gaussian variables : Consider two real random variables distributed according to the Gaussian distribution $P(x, y) = \mathcal{N} \exp \left[-\frac{1}{2} a x^2 + c x y - \frac{1}{2} b y^2 \right]$.

a) Compute the normalisation constant $\mathcal{N}$. What is the condition on a , b and c ? Give $\langle X^2 \rangle$, $\langle Y^2 \rangle$ and $\langle XY \rangle$.

b) Determine the conditional probability $P(x|y)$. Deduce $\langle X | Y = y \rangle$ the average of X conditioned by $Y = y$. Give also $\langle X^2 | Y = y \rangle$ and the corresponding conditioned variance.

c) Application to the Brownian motion : We consider a mesoscopic particle at equilibrium in a homogeneous fluid. We assume that the joint distribution $P(x_t, x_0)$ of its position at time $t = 0$ and at time t is Gaussian. For simplicity assume $\langle x_0 \rangle = \langle x_t \rangle = 0$. Deduce the conditional probability $P(x_t | x_0)$ and the conditioned mean $\langle X_t | X_0 = x_0 \rangle$.

Hint : you can determine the coefficients a , b and c by noticing that $\langle (x_t - x_0)^2 \rangle = 2Dt$, where D is the diffusion constant.

A remark : the concept of independence should not be confused with that of correlation

independence $\Rightarrow$ uncorrelated
uncorrelated $\nRightarrow$ independence

Correlations are characterized by a specific correlation function. We illustrate this in the following exercise, and allow later in exercise 1.20.

²unless the people of Torino bring their cats in Paris and/or the Parisians their mice in Torino.

³unless all Parisian cats are domestic cats who never meet a mouse in their life!

✎ **Exercise 1.5 – Independence and correlations:** Consider a random angle θ uniformly distributed over $[0, 2\pi]$. We introduce the two coordinates on the circle, $x = \cos \theta$ and $y = \sin \theta$.

a) Compute the correlation $C = \langle x(\theta)y(\theta) \rangle - \langle x(\theta) \rangle \langle y(\theta) \rangle$. Are the coordinates correlated? Are they independent?

b) Propose another correlator which is non zero.

1.3 Mean, variance, moments and generating function

Let us consider a random variable X , now assumed to vary continuously in $\mathbb{R}$, and its probability distribution $P(x)$, i.e. $P(x) dx = \text{Proba}\{X \in [x, x + dx]\}$. The average, or “mean value”⁴ of the variable is $\langle X \rangle = \int dx P(x) x$. For any known function $\phi(x)$, the average of $\phi(X)$ is

$$\langle \phi(X) \rangle = \int dx P(x) \phi(x). \quad (1.21)$$

a) Moments

We denote by

$$\mu_n = \langle X^n \rangle \quad (1.22)$$

the n -th moment. It exists if the integral $\int dx P(x) x^n$ is convergent, which requires that $P(x)$ decays sufficiently rapidly, at least like $|x|^{-1-n-\epsilon}$ for any $\epsilon > 0$.

b) Generating function

If all moments exist, i.e. $P(x)$ decays faster than any power law, we can define the *generating function*

$$G(k) \stackrel{\text{def}}{=} \langle e^{kX} \rangle = \sum_{n=0}^{\infty} \frac{\mu_n}{n!} k^n, \quad (1.23)$$

where k is the “conjugated variable”. Often, the generating function is more simple to obtain than the moments: this is why one computes partition functions in statistical physics. Given $G(k)$, one can deduce the moments by simple derivations

$$\mu_n = G^{(n)}(0), \quad (1.24)$$

the function G “generates” the moments (by derivations).

Important example in physics : the canonical partition function $Z_\beta = \sum_\ell e^{-\beta E_\ell}$ is a generating function for energy levels which are randomly occupied.

✎ **Exercise 1.6 – Gamma distribution:** The Gamma distribution is defined by

$$P^{(\mu)}(x) = c \theta_H(x) x^{\alpha-1} e^{-x} \quad (1.25)$$

where $\alpha > 0$. We introduce the Heaviside function $\theta_H(x)$ and c is a normalisation constant.

a) Compute the normalisation c .

b) Compute the moments μ_n .

c) Compute the generating function $G(k)$. Specify for which values of k the generating function is well defined.

d) Expand $G(k)$ in order to recover the moments.

e) Discuss more specifically the case $\alpha = 1$.

⁴also called “expectation value” by Mathematicians, and usually denoted $\mathbb{E}(X)$ However, unless the distribution is very narrow, there is no reason to “expect” the mean value.

✎ **Exercise 1.7 – :** Compute the moments of the symmetric Gaussian distribution $g_{0,\sigma}(x)$ and the generating function.

Hint : express the moments in terms of the Euler Gamma function $\Gamma(z) = \int_0^\infty dt t^{z-1} e^{-t}$.

✎ **Exercise 1.8 – The domain of definition of $G(k)$ for finite moments:** For finite moments ($\mu_n < \infty \forall n$), the generating function is not always defined on the full real axis. Let us consider three examples to illustrate this point.

a) Gaussian distribution : consider $g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Give the moments μ_n . Discuss the convergence radius of the series $\sum_n \frac{\mu_n}{n!} k^n$. Comment.

b) Exponential distribution : consider now $w(x) = \frac{1}{2} e^{-|x|}$. Same question.

c) Distribution

$$p(x) = \frac{1}{4\sqrt{|x|}} e^{-\sqrt{|x|}} \quad (1.26)$$

Same question.

c) The case of divergent moments

If the distribution does not decay sufficiently fast, highest moments are divergent. However, it is always possible to choose k purely imaginary, which corresponds to consider the *characteristic function*, i.e. the Fourier transform of the distribution

$$\hat{P}(k) = \langle e^{-ikX} \rangle \equiv G(-ik) \quad (1.27)$$

which always exists (the integral $\hat{P}(k) = \int dx P(x) e^{-ikx}$ is absolutely convergent for real k).

For example, consider a positive random variable X with distribution with power law tail

$$P(x) \sim x^{-1-\mu} \quad \text{for } x \rightarrow \infty \quad (1.28)$$

with $\mu > 0$. Clearly

$$\begin{cases} \mu_n = \langle X^n \rangle < \infty & \text{for } n < \mu \\ \mu_n = \langle X^n \rangle = \infty & \text{for } n \geq \mu \end{cases} \quad (1.29)$$

We now show that this has an interesting consequence on the characteristic function.

✎ **Exercise 1.9 – :** Discuss the moments of the Cauchy distribution (1.7). Compute its characteristic function $\hat{C}_{\mu,a}(k)$.

The n -th moment is related to the derivative of the characteristic function :

$$\mu_n = i^n \hat{P}^{(n)}(0). \quad (1.30)$$

If the n -th moment is finite, the characteristic function is differentiable at least n times.

When all moments are finite, we can write the series representation $\hat{P}(k) = \sum_{n=0}^\infty \frac{\mu_n}{n!} (-ik)^n$, i.e. $\hat{P}(k)$ is an analytic function of the variable k . However, when some moments are infinite, such a series representation does not exist. Diverging moments indicate that the characteristic function is *not differentiable*, hence is *non analytic* at $k = 0$. For the tail $P(x) \sim x^{-1-\mu}$, moments up to $n = \lfloor \mu \rfloor$ are finite, thus $\hat{P}(k)$ is differentiable $\lfloor \mu \rfloor$ times but higher derivatives are infinite for $k = 0$. We illustrate this on two examples.

Example : We consider a random variable X with distribution $f(x) = \frac{4/\pi}{4+x^4}$. Only the two first moments are finite : $\langle X \rangle = 0$, $\langle X^2 \rangle = 2$ and $\langle X^{2n} \rangle = \infty$ for $n > 1$ (odd moments being zero by symmetry).

The characteristic function is $\hat{f}(k) = \langle e^{-ikX} \rangle = (\cos k + \sin |k|)e^{-|k|}$. Expansion of the function for $k \rightarrow 0$ reads $\hat{f}(k) = 1 - k^2 + \frac{2}{3}|k|^3 + \mathcal{O}(k^4)$, which shows that $\hat{f}(k) \in \mathcal{C}_2(\mathbb{R})$ (continuous function differentiable twice everywhere).

✎ **Exercise 1.10** – : Compute the Fourier transform of $f(x) = \frac{4/\pi}{x^4+4}$ using residue theorem. Compute also the Fourier transform of $g(x) = \frac{2/\pi}{(x^2+1)^2}$. Compare their $k \rightarrow 0$ behaviours.

✎ **Exercise 1.11** – : Consider a positive random variable X distributed according to the law

$$w(x) = \frac{1}{\Gamma(\mu)} x^{1+\mu} e^{-1/x}, \quad (1.31)$$

where $\mu > 0$. Compute the moments and the generating function $\tilde{w}(\beta) = G(-\beta) = \langle e^{-\beta X} \rangle$. Discuss its expansion for $\beta \rightarrow 0$.

Hint : use the integral representation of the MacDonald function (cf. appendix).

Remark : the law (1.31) for $\mu = 1/2$ is known as the Lévy distribution.

✎ **Exercise 1.12** – “Laplace reciprocity” : In this exercise we show that the $x \rightarrow \infty$ behaviour of the distribution is related to the $k \rightarrow 0$ of its Laplace transform. We consider a positive random variable $X > 0$, with distribution with tail $w(x) \simeq c x^{-1-\mu}$ with $\mu > 0$. Then it is possible to define its Laplace transform (i.e. the generating function).

a) Case $\mu \in]0, 1[$: show that the characteristic function (Laplace transform) presents the behaviour $\tilde{w}(\beta) \stackrel{\text{def}}{=} G(-\beta) \simeq 1 - \tilde{c} \beta^\mu$ for $\beta \rightarrow 0$ (give the constant $\tilde{c}$).

Hint : start from $\tilde{w}(\beta) = \int_0^\infty dx w(x) e^{-\beta x} = 1 - \int_0^\infty dx w(x) (1 - e^{-\beta x})$.

b) Consider now $\mu \in]1, 2[$. How could you adapt the same trick ? Deduce the $\beta \rightarrow 0$ behaviour of $\tilde{w}(\beta)$.

Conclusion : The above discussion and the careful study of the examples of the exercises show that the *large* x power law tail is related to the *short* k behaviour

$$P(x) \underset{x \rightarrow \infty}{\sim} x^{-1-\mu} \quad \leftrightarrow \quad \hat{P}(k) \underset{k \rightarrow 0}{\simeq} \sum_{n=0}^{[\mu]} \frac{\mu_n}{n!} (-ik)^n + \tilde{C} k^\mu \quad (1.32)$$

(cf. Exercise 1.13 for a discussion of the coefficient $\tilde{C}$). When $\mu = N$ is integer, the non analytic term is $|k|^N$ (for a symmetric law with N odd) or $k^N \ln |k|$.

Such property is a manifestation of *Fourier reciprocity*.

✎ **Exercise 1.13** – **Coefficient $\tilde{C}$ for a symmetric distribution** : We consider a real random variable with a symmetric distribution $P(x) = P(-x)$ exhibiting a power law behaviour $P(x) \simeq C |x|^{-1-\mu}$ at infinity. Using the results of Exercise 1.12, deduce the coefficient $\tilde{C}$ of the expansion (1.32).

1.4 Variance and cumulants

a) Variance and standard deviation

The fluctuations of a random variable around its average can be quantified by considering the variance

$$\text{var}(X) \stackrel{\text{def}}{=} \langle [X - \langle X \rangle]^2 \rangle = \langle X^2 \rangle - \langle X \rangle^2 \quad (1.33)$$

We also introduce the *standard deviation*

$$\sigma \stackrel{\text{def}}{=} \sqrt{\text{var}(X)} \quad (1.34)$$

which gives the order of magnitude of the fluctuations.

The average tells us where the distribution of the random variable is centered, while the variance characterizes the width of the distribution. To some extent, the variance is the information about the distribution which remains when we forget the average (if we center the distribution by performing a shift by $-\langle X \rangle$). Can we proceed and remove the information encoded in the variance ? The answer is yes and corresponds to introduce *cumulants*.

b) Generating function of the cumulants

Cumulants are defined from the generating function

$$W(k) = \ln G(k) = \sum_{n=1}^{\infty} \frac{\kappa_n}{n!} k^n \quad (1.35)$$

where κ_n is the cumulant of order n .

✎ **Exercise 1.14** – : Expand $W(k) = \ln G(k) = \ln \left(\sum_{n=0}^{\infty} \frac{\mu_n}{n!} k^n \right)$ and deduce that the expression of the four first cumulants in terms of the moments are

$$\kappa_1 = \mu_1 \quad (1.36)$$

$$\kappa_2 = \mu_2 - \mu_1^2 = \langle [X - \langle X \rangle]^2 \rangle \quad (1.37)$$

$$\kappa_3 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3 = \langle [X - \langle X \rangle]^3 \rangle \quad (1.38)$$

$$\kappa_4 = \mu_4 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4 = \langle [X - \langle X \rangle]^4 \rangle - 3\langle [X - \langle X \rangle]^2 \rangle^2 \quad (1.39)$$

Remark : The expression of the fourth cumulant shows that cumulants should NOT be confused with the centered moments $\langle [X - \langle X \rangle]^n \rangle$

✎ **Exercise 1.15** – : Give all cumulants of the Gaussian distribution $g_{\mu,\sigma}(x)$.

Sketch a symmetric distribution with $\kappa_4 > 0$ and one with $\kappa_4 < 0$.

Convexity of $W(k)$: The cumulant generating function is a convex function. We can deduce this property from the Hölder inequality $\sum_i |x_i y_i| \leq \left(\sum_i |x_i|^{1/\alpha} \right)^\alpha \left(\sum_i |y_i|^{1/\beta} \right)^\beta$, with $\alpha + \beta = 1$ and $\alpha, \beta \in [0, 1]$. The substitution $x_i \rightarrow e^{\alpha k X_i}$ and $y_i \rightarrow e^{\beta k' X_i}$ leads to the form :

$$\ln \left(\sum_i e^{\alpha k X_i + \beta k' X_i} \right) \leq \alpha \ln \left(\sum_i e^{k X_i} \right) + \beta \ln \left(\sum_i e^{k' X_i} \right) \quad (1.40)$$

Because $W(k)$ has the form of a sum $\ln \left(\sum_i e^{k X_i} \right) \rightarrow W(k)$, we can write

$$W(\alpha k + (1 - \alpha)k') \leq \alpha W(k) + (1 - \alpha)W(k') \quad \forall \alpha \in [0, 1] \quad (1.41)$$

which means that $W(k)$ is convex (draw a plot).

The Marcinkiewicz theorem : A theorem proved by Marcinkiewicz in 1939 [49] (see also [11]). If it is polynomial, the function $W(k)$ is necessary of degree 1 or 2. In other terms, either cumulants κ_n all vanish for $n > 2$ (Gaussian variable), or all cumulants are non zero.

Remark : note that the odd cumulants could all vanish for symmetry reason (symmetric distribution). This is trivial.

Meaning of the first cumulants.— The third cumulant κ_3 indicates whether positive ($\kappa_3 > 0$) or negative ($\kappa_3 < 0$) fluctuations are favoured (Fig. 1.6). In order to introduce a dimensionless quantity, it is customary to define the “*skewness*”

$$\text{skewness} = \kappa_3 / \kappa_2^{3/2} . \quad (1.42)$$

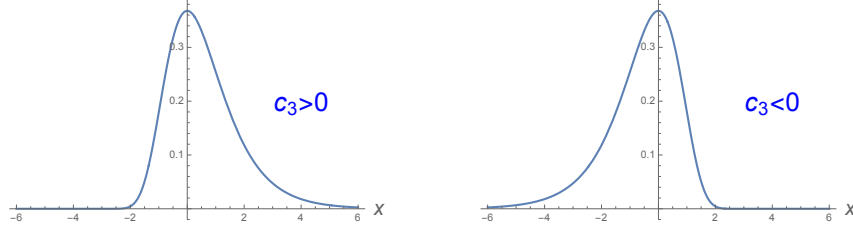


Figure 1.6: À gauche : la loi de Gumbel $f(x) = \exp(-x - e^{-x})$ a pour premiers cumulants : $\kappa_1 = \mathbf{C} = 0.577\dots$ (constante d’Euler), $\kappa_2 = \pi^2/6$, $\kappa_3 = 2\zeta(3)$ (fonction zeta de Riemann), $\kappa_4 = \pi^4/15$, etc. Le signe de κ_3 indique si les fluctuations positives ou négatives (relativement à $\langle X \rangle$) sont favorisées. From [71].

The fourth cumulant κ_4 encodes whether large ($\kappa_4 > 0$) or small ($\kappa_4 < 0$) fluctuations are favoured, compared to the Gaussian case ($\kappa_4 = 0$) : an illustration is shown in Fig. 1.7. One defines the “*kurtosis*”

$$\text{kurtosis} = \kappa_4 / \kappa_2^2 . \quad (1.43)$$

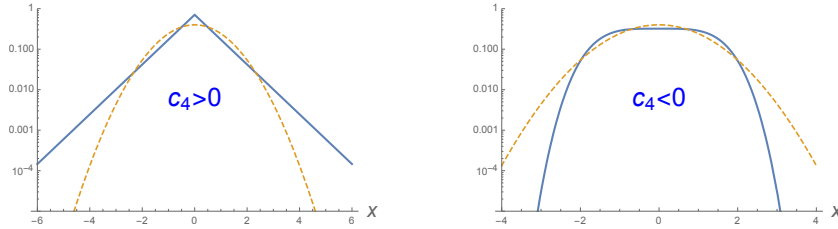


Figure 1.7: La distribution $g_a(x) = \frac{a\sqrt{\Gamma(3/a)}}{2\Gamma(1/a)^{3/2}} \exp\left(-\left|\sqrt{\frac{\Gamma(3/a)}{\Gamma(1/a)}}x\right|^a\right)$, ici tracée pour $a = 1$ (à gauche) et $a = 4$ (à droite), a pour premiers cumulants (non nuls) : $\kappa_2 = 1$, $\kappa_4 = \frac{\Gamma(1/a)\Gamma(5/a)}{\Gamma(3/a)^2} - 3$, etc. Le quatrième cumulant change de signe pour $a = 2$ (les grandes fluctuations sont favorisées si $a < 2$). La courbe en tirets est la distribution gaussienne ($a = 2$). From [71].

c) Addition of random variables

The study of statistical properties of sums of random variables is a standard and important problem which has many occurrences in mathematics, physics, etc. For example, given the Maxwell distribution of the velocity of molecules, what is the distribution of the kinetic energy of a gas ?

If variables are correlated, the problem can be rapidly tricky. On the other hand, for independent variables, there exist a lot of powerful results (like the central limit theorem for identical variables). Let us start with the simplest case of two independent random variables X and Y and denote by $S = X + Y$ their sum. Moments of S are not simply related to those of X and Y , obviously $\mu_n^S \neq \mu_n^X + \mu_n^Y$ (for example, $\langle (X + Y)^2 \rangle \neq \langle X^2 \rangle + \langle Y^2 \rangle$ in general, the difference

between the two sides being $2\langle X\rangle\langle Y\rangle$ for independent variables). Cumulants are the specific combinations of the moments such that additivity is preserved for independent variables. Denote by κ_n^X , κ_n^Y and κ_n^S the cumulants of the three variables. We have

$$\boxed{\kappa_n^{X+Y} = \kappa_n^X + \kappa_n^Y \quad \text{for } X \text{ and } Y \text{ independent.}} \quad (1.44)$$

✎ **Exercise 1.16** – : (IMPORTANT) Prove the equality.

✎ **Exercise 1.17 – Divisibility**: We have shown above that cumulants are additive when independent variables are added. This is a crucial property leading in particular to the central limit theorem discussed below. Can we now discuss the opposite : given a random variable S with distribution $w(s)$, does it exist a couple of i.i.d. random variables X_1 and X_2 such that $S = X_1 + X_2$? If the answer is yes, the variable S is said to be **divisible**.

Denote $\hat{w}(k) = \int ds w(s) e^{-iks} = \langle e^{-ikS} \rangle$ the characteristic function. Divisibility requires that $\hat{p}(k) = \sqrt{\hat{w}(k)}$ can be interpreted as a characteristic function, i.e. $\int \frac{dk}{2\pi} \sqrt{\hat{w}(k)} e^{ikx} = p(x)$ is positive and can be interpreted as a distribution. We discuss two examples.

a) Consider a random variable S with characteristic function $\hat{w}(k) = e^{-|k|^\mu}$ with $\mu \in]0, 2]$. Is S divisible ? If yes, what is the distribution $p(x)$ of X_i ?

b) Now, we consider the box distribution $w(s) = \frac{1}{2} \theta_H(1 - |s|)$. In this case, we use a weaker criterion based on Exercise 1.38. Compute the moments μ_n 's of S . Deduce the first four cumulants κ_n and the kurtosis. Discuss the divisibility.

1.5 Multivariate Gaussian distribution

We consider N Gaussian random variables $x_1, \dots, x_N$. The most general Gaussian distribution has the form

$$P(X) = \sqrt{\det \left(\frac{A}{2\pi} \right)} e^{-\frac{1}{2}(X-X_0)^T A (X-X_0)} \quad (1.45)$$

where X and X_0 are column vectors $\in \mathbb{R}^N$ and A is a real and strictly positive ⁵ symmetric matrix.

The first interesting question is to characterize the *correlations* $\langle x_i x_j \rangle$ between the random variables. The most simple way is to compute the generating function

$$G(K) = \langle e^{K^T X} \rangle = e^{K^T X_0 + \frac{1}{2} K^T A^{-1} K} \quad (1.46)$$

Remark that this does not even require to compute the multiple integral over X as it is sufficient to manipulate the quadratic form :

$$\begin{aligned} & (X - X_0)^T A (X - X_0) - 2K^T X \\ &= (X - X_0)^T A (X - X_0) - (X - X_0)^T K - K^T (X - X_0) - 2K^T X_0 \\ &= [X - X_0 - A^{-1} K]^T A [X - X_0 - A^{-1} K] - K^T A^{-1} K - 2K^T X_0 \end{aligned}$$

From the knowledge of the generating function $G(K) = \langle e^{K^T X} \rangle$ we can deduce any correlation function from simple derivation operations : $\langle x_i \rangle = \frac{\partial G(K)}{\partial k_i} \Big|_{K=0} = \langle x_i e^{K^T X} \rangle \Big|_{K=0}$, $\langle x_i x_j \rangle = \frac{\partial^2 G(K)}{\partial k_i \partial k_j} \Big|_{K=0}$, etc. We can also compute the connex correlation function (cumulant) $\langle x_i x_j \rangle_c \stackrel{\text{def}}{=} \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$

$$\langle x_i x_j \rangle_c = \frac{\partial^2}{\partial k_i \partial k_j} \ln G(K) \Big|_{K=0} \quad (1.47)$$

⁵with positive e.v.

For Gaussian variables, from the above expression, we deduce the important relation

$$\boxed{\langle x_i x_j \rangle_c = (A^{-1})_{ij}} \quad (1.48)$$

The result is remarkable : **it is sufficient to identify the matrix A in the Gaussian measure (and inverse it) to get the correlation function** (no need to compute a multiple integral). This is also true for any correlation function, as we show below (Wick theorem).

✎ **Exercise 1.18 – “Discrete Furutsu-Novikov theorem”** : We consider $f(X)$, a function of $X = (x_1, \dots, x_N)^T \in \mathbb{R}^N$, differentiable and such that $\langle X f(X) \rangle < \infty$. Show that for **Gaussian** random variables with $\langle x_i \rangle = 0$ one has

$$\langle x_i f(X) \rangle = \sum_j \langle x_i x_j \rangle \left\langle \frac{\partial f}{\partial x_j} \right\rangle. \quad (1.49)$$

Furutsu-Novikov theorem.— The Furutsu-Novikov theorem was proven in 1963 by Furutsu in the context of fluctuating electrodynamics and in 1964 by Novikov in the context of turbulence.⁶ Consider $\varphi(x)$ a Gaussian field (i.e. with Gaussian measure), and $F[\varphi]$ a functional of this field, then

$$\langle \varphi(x) F[\varphi] \rangle = \int dx' \langle \varphi(x) \varphi(x') \rangle \left\langle \frac{\delta F[\varphi]}{\delta \varphi(x')} \right\rangle \quad \text{for a Gaussian field } \varphi(x). \quad (1.50)$$

where $\frac{\delta F[\varphi]}{\delta \varphi(x)}$ is a “functional derivative”. This is the continuous version of (1.49).

✎ **Exercise 1.19 – Wick theorem** : We consider N **Gaussian** random variables with distribution $P(X) \propto e^{-\frac{1}{2} X^T A X}$.

- Compute the four point correlation function $\langle x_i x_j x_k x_l \rangle$.
- Generalize to the $2n$ -point correlation function $\langle x_1 x_2 \dots x_{2n} \rangle$. Show that

$$\langle x_1 \dots x_{2n} \rangle = \sum_{\text{contractions}} \langle x_{i_1} x_{j_1} \rangle \langle x_{i_2} x_{j_2} \rangle \dots \langle x_{i_n} x_{j_n} \rangle \quad (1.51)$$

What is the number of terms in the sum (“contractions” of the $2n$ indices) ?

✎ **Exercise 1.20 – Independence and correlations (2)** : Consider a set of Gaussian correlated random variables y_i of zero mean, such that $\langle y_i y_j \rangle = C_y(|i - j|)$, and an independent set of i.i.d. random variables $\epsilon_i = \pm 1$ with $\langle \epsilon_i \rangle = 0$.

- We define $x_i = \epsilon_i y_i$. Are these variables correlated ? (consider $\langle x_i x_j \rangle$)
- Are the variables x_i independent ? What can you study to answer the question ?

1.6 The central limit theorem

A fundamental theorem of probability theory is the “central limit theorem”, which provides statistical informations for the sum of N identically and independently distributed (i.i.d.) random variables $X_1, \dots, X_N$, with common distribution $p(x)$. We denote their sum

$$S_N = \sum_{n=1}^N X_n. \quad (1.52)$$

⁶K. Furutsu, *On the Statistical Theory of Electromagnetic Waves in a Fluctuating Medium* (I), Journal of Research of the National Bureau of Standards **67D**(3), 303–323 (1963).

E. A. Novikov, *Functionals and the random-force method in turbulence theory*, Sov. Phys. JETP **20**(5), 1290–1294 (1965), [Zh. Eksp. Teor. Fiz. 47, 1919 (1964)].

- Assuming that the first moment exists, $\langle X_i \rangle < \infty$, the **law of large numbers** holds : the limit

$$\frac{S_N}{N} \xrightarrow{N \rightarrow \infty} \langle X_i \rangle \quad (1.53)$$

is finite. S_N/N is a random quantity, however the limit is deterministic (non random).

- Assuming that the second moment exists, $\langle X_i^2 \rangle < \infty$, the **central limit theorem** holds : the distribution of S_N is a universal Gaussian law centered on $N \langle X_i \rangle$ and with variance $\text{var}(S_N) = N \text{var}(X_i)$.

It is worth stressing that, while the additivity of the mean is trivial and simply follows from the linearity of the average, the additivity of the variance is quite non trivial and relies on the independence of the variables, as previously emphasized.

a) Proof of the theorem

It is useful to recall the proof of the theorem. The distribution of the sum is ⁷

$$P_N(s) = \left\langle \delta \left(s - \sum_{i=1}^N X_i \right) \right\rangle \stackrel{\text{indep.}}{=} \underbrace{(p * \dots * p)}_{N \text{ times}}(s) = \int_{\mathbb{R}} \frac{dk}{2\pi} e^{iks} \hat{p}(k)^N \quad (1.54)$$

where we have found convenient to write the convolution in terms of the Fourier transform. We now want to analyze the integral in the limit $N \rightarrow \infty$. For this reason we use the expansion in terms of the cumulants

$$\hat{p}(k) = e^{W(-ik)} = e^{-i\kappa_1 k + \frac{1}{2}\kappa_2 (-ik)^2 + \mathcal{O}(k^3)} \quad (1.55)$$

This leads to

$$P_N(s) = \int_{\mathbb{R}} \frac{dk}{2\pi} e^{iks - iN\kappa_1 k - \frac{N}{2}\kappa_2 k^2 + \mathcal{O}(Nk^3)} \quad (1.56)$$

Then we rescale the argument of the distribution as

$$s = N\kappa_1 + \sqrt{N\kappa_2} y \quad (1.57)$$

and change the variable of integration as $k = t/\sqrt{N\kappa_2}$ leading to

$$P_N(N\kappa_1 + \sqrt{N\kappa_2} y) = \frac{1}{2\pi\sqrt{N\kappa_2}} \int_{\mathbb{R}} dt e^{ity - \frac{1}{2}t^2 + \mathcal{O}(N^{-1/2}t^3)} \quad (1.58)$$

Finally we use

$$\lim_{N \rightarrow \infty} \int_{\mathbb{R}} dt e^{ity - \frac{1}{2}t^2 + \mathcal{O}(N^{-1/2}t^3)} = \sqrt{2\pi} e^{-\frac{1}{2}y^2} \quad (1.59)$$

leading to

$$\sqrt{N\kappa_2} P_N(N\kappa_1 + \sqrt{N\kappa_2} y) \underset{N \rightarrow \infty}{\simeq} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \quad (1.60)$$

which completes the proof.

In conclusion, the central limit theorem states that

$$P_N(s) \underset{N \gg 1}{\simeq} \frac{1}{\sqrt{2\pi N \text{var}(X_i)}} \exp \left\{ -\frac{(s - N \langle X_i \rangle)^2}{2N \text{var}(X_i)} \right\}. \quad (1.61)$$

⁷Start with $P_2(s) = \int dx_1 p(x_1) \int dx_2 p(x_2) \delta(s - x_1 - x_2) = \int dx_1 p(x_1) p(s - x_1) = (p * p)(s)$.

b) Convergence towards the Gaussian : Edgeworth expansion

At this stage, a natural question is : for a finite (large) N , how close is $P_N(s)$ to the Gaussian distribution ? The convergence towards the Gaussian distribution as N grows can be characterised precisely. For this purpose, it is convenient to use the rescaled variable $y = (s - N\kappa_1)/\sqrt{N\kappa_2}$ (which is $\sim \mathcal{O}(1)$ for *typical* fluctuations) and consider the rescaled distribution

$$\psi_N(y) \stackrel{\text{def}}{=} \sqrt{N\kappa_2} P_N\left(N\kappa_1 + \sqrt{N\kappa_2} y\right) \quad (1.62)$$

In the limit $N \rightarrow \infty$, we have shown that $\lim_{N \rightarrow \infty} \psi_N(y) = g(y) = (2\pi)^{-1/2} e^{-(1/2)y^2}$. We have (Edgeworth expansion)

$$\psi_N(y) - g(y) = \sum_{k=1}^{\infty} \frac{Q_k(y)}{N^{k/2}} \quad (1.63)$$

where the function $Q_k(y)$ involved in the correction of order $\mathcal{O}(N^{-k/2})$ combines derivatives of the Gaussian weighted the dimensionless parameters $\lambda_k \stackrel{\text{def}}{=} \kappa_k/\kappa_2^{k/2}$ (this is the k -th cumulant of $X_i/\sqrt{\kappa_2}$). In particular

$$Q_1(y) = -\frac{\lambda_3}{6} g'''(y) = \frac{\lambda_3}{6} (y^3 - 3y) g(y), \quad Q_2(y) = \frac{\lambda_4}{24} g^{(4)}(y) + \frac{\lambda_3^2}{72} g^{(6)}(y), \text{ etc.} \quad (1.64)$$

(more terms can be found on the [wikipedia's](#) page). *Stricto sensu* (1.63) is an asymptotic series. It is also known as the Edgeworth-Chebyshev expansion.

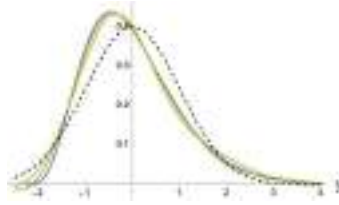


Figure 1.8: *Truncated Edgeworth expansion for an exponential law* ($p(x) = \theta_H(x) e^{-x}$), for $N = 5$. Dashed line is the Gaussian distribution ; the blue line is the exact distribution $P_N(s) = \theta_H(s) \frac{1}{(N-1)!} s^{N-1} e^{-s}$ rescaled ; the orange and green lines show approximations from the Edgeworth expansion (1.63) truncated after $k = 1$ and $k = 2$ terms, respectively.

c) Beyond the central limit theorem : several directions

What are the limitations of the central limit theorem ? Of course, we could ask for generalizations when one relaxes various hypothesis :

- What about the case where $\langle X_i^2 \rangle = \infty$? This might occur when $p(x)$ presents a power law. We will discuss this below in § 1.8.
- What about the case of correlated variables ? The story is then more complicated as there are many ways to introduce correlations. At the end of the lectures (chapter 6) we will explain a method to study the distribution of $\sum_{t=1}^N U(x_t)$, where $U(x)$ is a known function and x_t is a stochastic process correlated in time (like a Brownian motion).
- A more subtle point, somehow hidden in the previous demonstration : the permutation of the limit $\lim_{N \rightarrow \infty}$ and the integral $\int_{\mathbb{R}} dt$ is in fact licit only when the argument s of $P_N(s)$ is *not too large*, being the origin of the theorem's name. The theorem only characterizes the *central* universal Gaussian part of the distribution, of width $\delta S \sim \mathcal{O}(\sqrt{N})$. Where does

the Gaussian approximation breaks down ? Can we say something about the tails of the distribution $P_N(s)$, for $s \rightarrow \infty$? To shed light on this subtle point, we will introduce the concept of *large deviations* in the next §.

- A last direction, which will not be addressed in these lecture notes : generalized central limit theorem involving non commuting objects is an important development in mathematics and physics. The study of random matrix products enters this direction and has many applications. References are [17, 14, 70] (or [9] in the Mathematical literature).

At this point it is important to check that one is comfortable with the steepest descent method : cf. § in the Appendix page 152.

1.7 Large deviations : typical and atypical fluctuations

What is central in the Central Limit Theorem ?

Let us now study more precisely the distribution of S_N . This time, we find more convenient to use the Laplace transform. We introduce the cumulant generating function

$$W(k) = \ln \langle e^{kX} \rangle \quad (1.65)$$

The distribution of S_N can be written as an inverse Laplace transform (cf. Appendix A.4) :

$$P_N(s) = \underbrace{(p * \dots * p)}_{N \text{ times}}(s) = \int_{i\mathbb{R}} \frac{dk}{2i\pi} e^{-ks + N W(k)}. \quad (1.66)$$

In the exponential, we write the argument $s = N y$ with $y = s/N \sim \mathcal{O}(1)$. Then the integral has the form $\int dk e^{-N[ky - W(k)]}$ which suggests to use the steepest descent method (cf. Appendix page 152). Assuming that the integral is dominated by a single (real) saddle point at k_* , we have

$$P_N(s = N y) \underset{N \rightarrow \infty}{\simeq} \frac{1}{\sqrt{2\pi N W''(k_*)}} e^{-N[k_* y - W(k_*)]} \quad (1.67)$$

where the saddle point is given by $\frac{\partial}{\partial k}[ky - W(k)] = 0$, i.e.

$$W'(k_*) = y \quad (1.68)$$

(i.e. k_* is function of y). It is important to note that $W(k)$ being a **convex function** (general property of the cumulant generating function), $W'(k)$ is monotonous and the equation $W'(k_*) = y$ has a unique solution (the function is invertible). We now introduce the “*large deviation function*” $\Phi(y) \stackrel{\text{def}}{=} k_* y - W(k_*)$. This shows that the cumulant generating function $W(k)$ and the large deviation function $\Phi(y)$ are related through a *Legendre transform*, i.e. we can write

$$P_N(s) \underset{N \rightarrow \infty}{\simeq} \exp \left\{ -N \Phi \left(\frac{s}{N} \right) \right\} \quad \text{where} \quad \Phi(y) = \max_k \{k y - W(k)\} \quad (1.69)$$

This result is known as the *Cramér theorem* (1938) [16] and the large deviation function $\Phi(y)$ is sometimes called the Cramér function [8].

Recover the central limit theorem.— Let us stress that the central limit theorem corresponds to a certain limit of (1.69) : consider the limit $y \rightarrow 0$ (i.e. consider that the argument of the distribution $P_N(s)$ is $s = N y \ll N$), we expect that $k_* \rightarrow 0$ as well, so that we can

use the expansion in terms of the cumulants $W(k) \simeq \kappa_1 k + (1/2)\kappa_2 k^2$. The saddle point is $k_* \simeq (y - \kappa_1)/\kappa_2$ and we get

$$\Phi(y) = k_* y - W(k_*) \underset{y \rightarrow \kappa_1}{\simeq} k_* (y - \kappa_1) - \frac{1}{2} \kappa_2 k_*^2 = \frac{(y - \kappa_1)^2}{2\kappa_2}. \quad (1.70)$$

This means that the distribution of the sum has a Gaussian form

$$P_N(s) \underset{\delta s \ll N}{\sim} \exp \left[- (s - N\kappa_1)^2 / (2N\kappa_2) \right] \quad (\text{central limit theorem (1.61)}). \quad (1.71)$$

where $\delta s = s - N\kappa_1$. This now makes clear over which range the Gaussian form is expected : the quadratic behaviour is obtained for $|y - \kappa_1| \ll 1$, i.e. $|\delta S_N| \ll \mathcal{O}(N)$. I.e. (1.69) contains the Central limit theorem. Above this scale, for $\delta S_N \gtrsim \mathcal{O}(N)$, the LDF is not quadratic in general and the distribution deviates from the universal Gaussian form obtained around $N\kappa_1$.

Gärtner-Ellis theorem.— The Gärtner-Ellis theorem relates more generally the scaled cumulant generating function to the large deviation function, through the Legendre transform (1.69), for more general observables (not only the sum of i.i.d. variables). The key point is the existence of a unique scaling with N for all cumulants, i.e. the existence of the limit $\lim_{N \rightarrow \infty} (1/N^\alpha) \ln \langle e^{k S_N} \rangle = W(k)$, where α is some exponent. For the sum of independent random variables (above) we have $\alpha = 1$. Another example where $\alpha = 2$: in random matrix theory, several interesting questions rely on the analysis of linear statistics of eigenvalues, that is, observables of the form $S_N = \sum_{i=1}^N f(\lambda_i)$, where the eigenvalues $\{\lambda_1, \dots, \lambda_N\}$ are *strongly* correlated (in this case, the scaled cumulant generating function is obtained from $W(k) = \lim_{N \rightarrow \infty} (1/N^2) \ln \langle e^{k S_N} \rangle$).

The Cramér theorem is a specific case of the Gärtner-Ellis theorem applying to the sum of i.i.d. random variables.

Universal typical fluctuations and *non* universal large deviations.— The sum $S_N = \sum_{i=1}^N X_i$ is a random variable with *typical* fluctuations scaling like $\delta S_N \sim \sqrt{N}$ (width of the distribution). The typical fluctuations are encoded by the central limit theorem and have a *universal Gaussian character*. However rare events such that $\delta S_N \gtrsim \mathcal{O}(N)$ are not forbidden : probability of such extreme events are described by the large deviation tails, controlled by $\Phi(y)$ for $y \gg 1$. The behaviour of $\Phi(y)$ for $y \gtrsim 1$ is *not* universal and depends strongly on $W(k)$ i.e. on the details of the law $p(x)$: large deviation tails for $|s - N \langle X \rangle| \gtrsim \mathcal{O}(N)$ characterize *atypical* fluctuations (rare events), which are *non universal*.

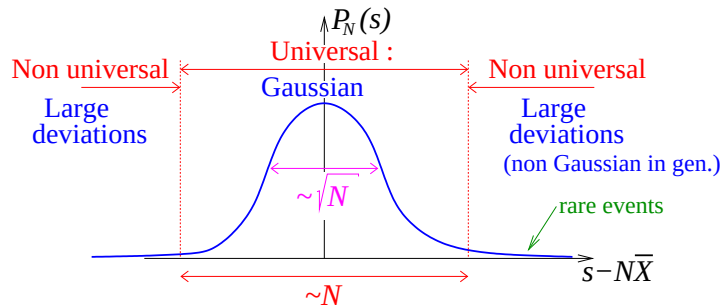


Figure 1.9: The distribution of the sum of i.i.d. variables presents a Gaussian universal central part of width $\sim \sqrt{N}$ (central limit theorem) and large deviation tails which are in general non Gaussian.

✎ **Exercise 1.21 – Large deviation function of the binomial distribution :** Using the Stirling formula, show that the binomial distribution (1.3) obeys the large deviation principle, i.e. has the form $\text{Proba}\{H_N = n\} \underset{N \rightarrow \infty}{\sim} \exp\{-N \Phi(\frac{n}{N})\}$. Give the large deviation function $\Phi(y)$.

Legendre transform and duality.— The Cramér theorem allows to deduce the LDF $\Phi(y)$ from the cumulant generating function $W(k)$. Another Legendre transform allows to go in the other way

$$W(k) = \max_y \{k y - \Phi(y)\} \quad (1.72)$$

which simply follows from the steepest descent approximation of the Laplace transform

$$e^{NW(k)} = \langle e^{kS_N} \rangle = \int ds P_N(s) e^{ks} \sim N \int dy e^{N[ky - \Phi(y)]} \sim e^{N \max_y \{k y - \Phi(y)\}} \quad (1.73)$$

The two Legendre transforms (1.69, 1.72) introduced above are

$$\begin{cases} \Phi(y) = ky - W(k) & \text{where } y = W'(k) \\ W(k) = ky - \Phi(y) & \text{where } k = \Phi'(y) \end{cases} \quad (1.74)$$

It is convenient to write

$$\begin{cases} \Phi(W'(k)) = kW'(k) - W(k) \\ W(\Phi'(y)) = y\Phi'(y) - \Phi(y) \end{cases} \quad (1.75)$$

Setting $k = 0$ in the first equation gives $\Phi(\kappa_1) = -W(0) = 0$, which is the minimum of the function Φ . Differentiations of the two equations give $\Phi'(W'(k)) = k$ and $W'(\Phi'(y)) = y$ as it should (the two functions W' and Φ' are inverse). Differentiating once more gives

$$\begin{cases} \Phi''(W'(k)) = 1/W''(k) \\ W''(\Phi'(y)) = 1/\Phi''(y) \end{cases} \quad (1.76)$$

or in short $\Phi''(y) = 1/W''(k)$ with $y = W'(k)$, i.e. the curvatures are inverse.

✎ **Exercise 1.22 – Duality between W and Φ :** Consider the cumulant generating function with limiting behaviour

$$W(k) \simeq \begin{cases} \kappa_1 k + (1/2)\kappa_2 k^2 & \text{for } k \rightarrow 0 \\ c|k|^{1+\alpha} & \text{for } k \rightarrow \pm\infty \end{cases} \quad (1.77)$$

Performing the Legendre transform, deduce the limiting behaviours of the large deviation function $\Phi(y)$ close to its minimum and at infinity.

✎ **Exercise 1.23 – Large deviations and statistical physics :** In statistical physics, the fundamental function is the entropy $S(E)$ characterizing the number of microstates at a given energy E : $\Omega(E) = e^{S(E)}$ (here $k_B = 1$). Argue that the canonical partition function $Z(\beta)$ is related to $\Omega(E)$ through a Laplace transform.

Deduce that the free energy $F(\beta) = -(1/\beta) \ln Z(\beta)$ and the entropy $S(E)$ are related through a Legendre transform (use extensivity). Interpret $S(E)$ and $F(\beta)$ within the large deviation frame.

For a simple discussion : see chapter 7 of [71]. For a broader perspective on the role of the large deviations in statistical physics, see the review article by Touchette [73].

✎ **Exercise 1.24** – : Consider the symmetric distribution

$$p(x) = \frac{1}{2} e^{-|x|}. \quad (1.78)$$

a) Compute the generating function $W(k) = \ln \langle e^{kX} \rangle$ and discuss its domain of definition. Deduce the two first cumulants.

b) Get the large deviation function from the Legendre transform $\Phi(y) = \max_k \{k y - W(k)\}$ which controls the distribution $P_N(s)$ of the sum of N such i.i.d. random variables. Analyze the $k \rightarrow 0$ result to check your result (compare to the central limit theorem).

c) The distribution $P_N(s)$ can be computed exactly : it is expressed in terms of the MacDonald function

$$P_{n+1}(s) = \frac{1}{\sqrt{\pi} n!} \left| \frac{s}{2} \right|^{n+1/2} K_{n+1/2}(s) = \frac{e^{-|s|}}{2^{2n+1} n!} \sum_{m=0}^n \frac{(n+m)!}{m!(n-m)!} (2|s|)^{n-m}. \quad (1.79)$$

Using the asymptotic of the MacDonald function (Exercise A.3), compare to the behaviour of previous question.

Does the large deviation form $P_N(s) \sim \exp \{-N \Phi(s/N)\}$ always holds ? (More advanced)

The previous exercise reveals some interesting observation : for a distribution with purely exponential tail $p(x) \sim e^{-|x|}$, the large deviation is also linear $\Phi(y) \sim |y|$ at large y , hence N is absent in the exponential tail of the distribution of the sum : $P_N(s) \sim \exp[-|s|]$. On the other hand, the domain of definition of $G(k) = \langle e^{kX} \rangle$ is a bounded domain in this case. From these two remarks we can guess that such a tail is a kind of threshold in the space of distributions. Indeed, if we consider distribution with stretched exponential tail

$$p(x) \underset{x \rightarrow \infty}{\sim} e^{-|x|^\alpha} \quad \text{for } \alpha \in]0, 1[\quad (1.80)$$

all the moments exist $\mu_n = \langle X^n \rangle < \infty$ however the generating function does not exist $G(k) = \infty$ for $k \in \mathbb{R}$. In this case, it is possible to show that the distribution of the sum presents a different large deviation form

$$P_N(s) \underset{N \rightarrow \infty}{\sim} \exp \left\{ -N^{\frac{\alpha}{2-\alpha}} \Phi \left(N^{-\frac{1}{2-\alpha}} s \right) \right\} \quad (1.81)$$

We check that it matches with the usual form when $\alpha = 1$.

In statistical physics, there are several occurrences of such large deviation form $P_N(X) \sim e^{-N^a \Phi(X/N^b)}$: the large deviation function $\Phi(y)$ is also called the *rate function*, N^a is the *speed* and N^b the *scale*.

✎ **Exercise 1.25** – : Consider i.i.d. random variables with distribution $p(x) = (4\sqrt{|x|})^{-1} e^{-\sqrt{|x|}}$ (studied in exercise 1.8). Using the above remark, give the corresponding large deviation form (1.81).

Summary : if one considers i.i.d. random variables with distribution presenting an exponential tail $p(x) \sim e^{-|x|^\alpha}$ for $x \rightarrow \infty$, the situation is as follows :

- For $\alpha \geq 1$: the distribution of the sum presents the usual large deviation form $P_N(s) \sim \exp \{-N \Phi(s/N)\}$.
- For $0 < \alpha \leq 1$: the distribution of the sum presents the form (1.81), involving non trivial powers of N .

✎ **Exercise 1.26 – Large deviation function for power law distributions :** Let us discuss the possible forms of the large deviation function when the distribution exhibits some power law tail. For simplicity we consider the distribution of the sum of $N \gg 1$ positive i.i.d. random variables, with support $\mathbb{R}^+$. Assume the large deviation form $P_N(s) \sim \exp\{-N\Phi(s/N)\}$. Imagine that it presents the behaviour $P_N(s) \sim \exp\{-N^2/s\}$ for $s \rightarrow 0$.

- a) Sketch the large deviation function $\Phi(y)$ if $P_N(s) \sim s^{-N-1}$ for $s \rightarrow \infty$.
b) Same question if $P_N(s) \sim s^{-3}$ for $s \rightarrow \infty$.

1.8 Generalization of the central limit theorem for power law distributions

The central limit theorem applies when the second moment is finite, $\langle X^2 \rangle < \infty$, meaning that the characteristic function behaves as ⁸ $\hat{p}(k) = 1 - i\mu_1 k - \frac{1}{2}\mu_2 k^2 + o(k^2)$. When $\langle X^2 \rangle = \infty$ (or $\langle X \rangle = \infty$), we have seen that the characteristic function is non analytic for $k \rightarrow 0$ and the central limit theorem does not apply. Because the characteristic function is well defined, we can as well characterize the distribution of the sum $P_N(s)$ with the help of a large deviation function.

a) Case of symmetric distributions $p(x)$: symmetric Lévy laws

Let us first consider the specific case of a symmetric distribution $p(x) = p(-x)$, then $\hat{p}(k)$ is real symmetric : this will simplify the discussion. There are three situations in general :

- (i) $\hat{p}(k)$ is analytic for $k \rightarrow 0$, all moments are finite, the central limit theorem holds.
- (ii) $\hat{p}(k)$ is *not* analytic for $k \rightarrow 0$, however the second moment is finite, meaning that $\hat{p}(k) \simeq 1 - \frac{1}{2}\kappa_2 k^2 + \dots + \tilde{c}|k|^\mu$ where $\mu > 2$ is non integer. The central limit theorem still holds.
- (iii) $\hat{p}(k)$ is *not* analytic for $k \rightarrow 0$, and behaves as

$$\hat{p}(k) \simeq 1 - \tilde{c}|k|^\mu \text{ for } \mu \in]0, 2[\quad (1.82)$$

meaning that $\langle X^2 \rangle = \infty$ and the central limit theorem does not hold.

We now examine further this last situation. We analyze $P_N(s) = \int_{\mathbb{R}} \frac{dk}{2\pi} e^{iks + N \ln \hat{p}(k)}$ and proceed as for the central limit theorem. The behaviour (1.82) however suggests to introduce a different scaling variable. We now write

$$(\tilde{c}N)^{1/\mu} P_N(s = (\tilde{c}N)^{1/\mu} y) \underset{N \rightarrow \infty}{\simeq} (\tilde{c}N)^{1/\mu} \int_{\mathbb{R}} \frac{dk}{2\pi} e^{(\tilde{c}N)^{1/\mu} iky - N\tilde{c}|k|^\mu} \simeq \int_{\mathbb{R}} \frac{dq}{2\pi} e^{iqy - |q|^\mu} \quad (1.83)$$

In other terms

$$P_N(s) \underset{N \rightarrow \infty}{\simeq} \frac{1}{(\tilde{c}N)^{1/\mu}} \mathcal{L}_{\mu,0} \left(\frac{s}{(\tilde{c}N)^{1/\mu}} \right) \quad \text{where} \quad \mathcal{L}_{\mu,0}(x) \stackrel{\text{def}}{=} \int_{\mathbb{R}} \frac{dk}{2\pi} e^{ikx - |k|^\mu} \quad (1.84)$$

is a new universal function called a *Lévy law*, where the second index “0” refers to the symmetric case (see below). The Lévy law is defined from its Fourier transform $\hat{\mathcal{L}}_{\mu,0}(k) = e^{-|k|^\mu} \simeq 1 - |k|^\mu$ for $k \rightarrow 0$, which shows that $\mathcal{L}_{\mu,0}(x)$ presents the same power law tail as the initial distribution $p(x)$:

$$\mathcal{L}_{\mu,0}(x) \underset{x \rightarrow \infty}{\sim} |x|^{-1-\mu}. \quad (1.85)$$

Thus $P_N(s) \sim N/|s|^{1+\mu}$ for $s \rightarrow \infty$. The precise coefficient of the power law is given below (see below and Exercise 1.28).

⁸I recall that $o(x)$ is something small compared to $x \rightarrow 0$. Precisely $o(x)/x \rightarrow 0$ while $\mathcal{O}(x)/x \rightarrow \text{cst}$ for $x \rightarrow 0$.

Eq. (1.84) can be understood as an *extension of the central limit theorem*, which now involves a family of universal functions $\mathcal{L}_{\mu,0}(x)$ for $0 < \mu < 2$, replacing the Gaussian distribution.

Let us discuss more into detail the scaling of the sum S_N with N . The present discussion holds for symmetric power law distribution for $\mu \in]0, 2[$. The fluctuations of S_N now scale with N as

$$S_N \sim N^{1/\mu}. \quad (1.86)$$

(for the marginal case $\mu = 2$, there is a logarithmic correction, $S_N \sim \sqrt{N \ln N}$). This has some important consequences :

- For $1 < \mu < 2$ the relative fluctuations go to zero, $S_N/N \sim N^{1/\mu-1} \rightarrow 0$ for $N \rightarrow \infty$, hence the distribution of S_N/N is sharply peaked. The law of large numbers holds.
- For $0 < \mu < 1$, the relative fluctuations *grow* with N as $S_N/N \sim N^{1/\mu-1} \rightarrow \infty$ for $N \rightarrow \infty$, and the law of large numbers does not hold. Indeed, we will argue below that the sum is dominated by few extreme values (the largest X_i 's).

The mechanism which explains the scaling is different in the two cases : in the first case, in general (for non symmetric $p(x)$), all terms equally contribute to the sum $S_N \sim N$ (+ fluct. of order $N^{1/\mu} \ll N$), hence the main linear behaviour is explained by the fact that the sum has N terms. In the second case, when $0 < \mu < 1$, the sum is dominated by the largest contribution ; as N grows it becomes more and more probable to pick a larger X_i , which explains the scaling as $N^{1/\mu} \gg N$.

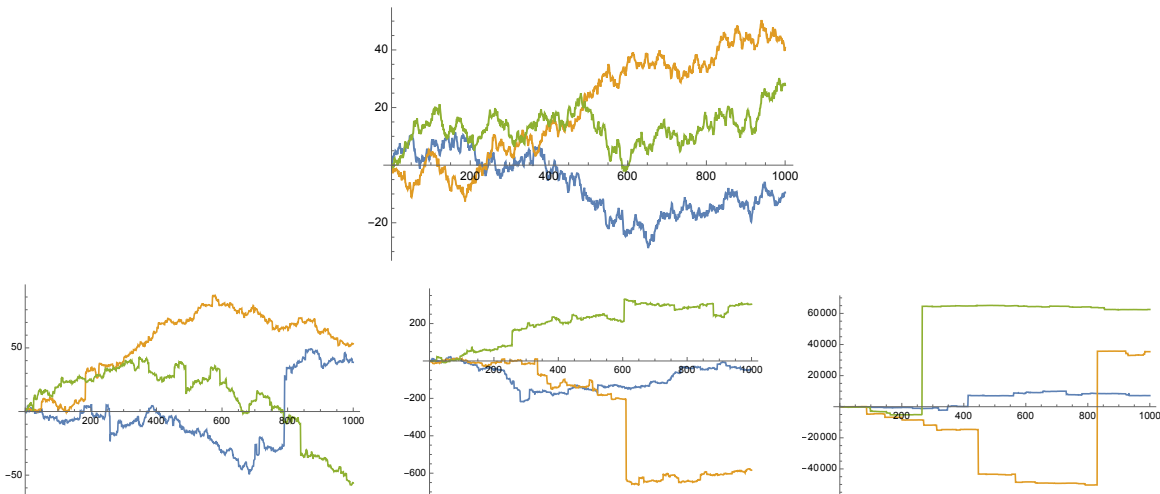


Figure 1.10: **Different random walks.**— We plot $S_N = \sum_{i=1}^N X_i$ as a function N for two different distributions $p(x)$ of the X_i 's : i is the “time”, N the final time and X_i 's are jumps. Top : Gaussian distribution (three realizations). Bottom : Power law distribution with decreasing exponent from left to right : $\mu = 2.5$ (left), $\mu = 1.5$ (center) and $\mu = 0.75$ (right). As μ decreases, large jumps become more important and few large jumps dominate the sum when $\mu \leq 1$. (the random numbers with power law distribution are generated thanks to the method of Exercise 1.3).

To emphasize this point, I have performed some simulations (Fig. 1.10) : generating i.i.d. random numbers, the sum $S_N = \sum_{i=1}^N X_i$ is plotted as a function of N . This is similar to the random walk problem. For Gaussian random numbers, looking at the process at large scale, the random walk seems continuous. For power law distribution with $\mu > 2$, the random walk exhibits some moderate jumps, which do not seem to contribute much to the final position. For $1 < \mu < 2$, we can see that the contributions of the big jumps become more important (look at the range). For $\mu < 1$, very few large jumps are prominent and strongly dominate the sum.

Universal large deviations.— As we have seen above with the central limit theorem when $\langle X^2 \rangle < \infty$, typical fluctuations are described by the universal Gaussian form, while non universality manifests in the atypical fluctuations (tails for large x). For the new universality classes corresponding to Lévy laws, one obtains distributions characterized by power law tails, hence universality extends to large deviations and, in this sense, is stronger than in the central limit theorem.



Figure 1.11: *Paul Lévy (1886-1971).*

b) Stable Lévy laws : general case (More advanced)

In the central limit theorem, the origin of the universal Gaussian law lies in the stability of the law against addition, which is clear from the expression of the characteristic function $\hat{g}_{\mu,\sigma}(k) = e^{-i\mu k - \frac{1}{2}\sigma^2 k^2}$, leading to

$$\hat{g}_{\mu_1,\sigma_1}(k)\hat{g}_{\mu_2,\sigma_2}(k) = \hat{g}_{\mu_1+\mu_2,\sqrt{\sigma_1^2+\sigma_2^2}}(k).$$

Stable laws against addition have been classified by Paul Lévy and are known as α -stable Lévy laws (because the exponent μ is often denoted α). They provide the correct extension of the central limit theorem to the case where the distribution exhibits power law tails with $\mu < 2$, being non symmetric. The characteristic function is [4]

$$\hat{\mathcal{L}}_{\mu,\beta}(k) = \begin{cases} e^{-|k|^\mu [1+i\beta \operatorname{sign}(k) \tan(\pi\mu/2)]} & \text{for } \mu \in]0, 2] \text{ \& } \mu \neq 1 \\ e^{-|k| [1-\frac{2i\beta}{\pi} \operatorname{sign}(k) \ln |k|]} & \text{for } \mu = 1 \end{cases} \quad (1.87)$$

The index μ controls the exponent of the tail $\mathcal{L}_{\mu,\beta}(x) \sim |x|^{-1-\mu}$ and the parameter $\beta \in [-1, +1]$ controls the asymmetry, cf. Eq. (1.90). For $\beta = 0$, the distribution is symmetric. The Lévy distribution $\mathcal{L}_{\mu,\beta}(x)$ does not have a simple analytic form in general, and is defined by its integral representation (inverse Fourier transform of $\hat{\mathcal{L}}_{\mu,\beta}(k)$). In order to get some intuition, let us make few concrete remarks.

- For $\beta = 0$, there exist expansions for $x \rightarrow 0$:

$$\mathcal{L}_{\mu,0}(x) = \frac{1}{\pi\mu} \sum_{n=0}^{\infty} \frac{(-1)^n \Gamma((2n+1)/\mu)}{(2n)!} x^{2n} \simeq \frac{1}{\pi\mu} \left(\Gamma(1/\mu) - \frac{1}{2} \Gamma(3/\mu) x^2 \right) \quad (1.88)$$

and for $x \rightarrow \infty$

$$\mathcal{L}_{\mu,0}(x) = \frac{1}{\pi} \sum_{n=1}^{\infty} (-1)^{n-1} \frac{\Gamma(1+n\mu) \sin(\pi n\mu/2)}{n!} |x|^{-1-\mu n} \simeq \frac{\Gamma(1+\mu) \sin(\pi\mu/2)}{\pi |x|^{1+\mu}} \quad (1.89)$$

✎ **Exercise 1.27** – : Get the convergence radius of the two series representations (1.88) and (1.89).

- $\hat{\mathcal{L}}_{1,0}(k) = e^{-|k|}$ corresponds to the Cauchy distribution $\mathcal{L}_{1,0}(x) = \frac{1/\pi}{x^2+1}$.

- For $\beta \neq 0$, the asymptotic of the Lévy distribution is (wikipedia)

$$\mathcal{L}_{\mu,\beta}(x) \underset{x \rightarrow \pm\infty}{\sim} \frac{\sin(\frac{\pi\mu}{2}) \Gamma(\mu+1)}{\pi} \frac{1 + \beta \operatorname{sign}(x)}{|x|^{\mu+1}} \quad (1.90)$$

In particular, this shows that the power law tail is present only on one side for $\beta = \pm 1$.

- For $\beta = +1$, the characteristic function is $\widehat{\mathcal{L}}_{\mu,1}(k) = e^{-c_\mu (ik)^\mu}$, where $c_\mu = 1/\cos(\pi\mu/2)$. For $\mu \in]0, 1[$, the support of the distribution is $\mathbb{R}_+$. For $\mu \in]0, 1[$ and $\beta = -1$, the support is $\mathbb{R}_-$.
- For $0 < \mu < 1$, the distribution $\mathcal{L}_{\mu,+1}(x)$ presents the essential singularity

$$\mathcal{L}_{\mu,1}(x) \sim x^{-\frac{2-\mu}{2(1-\mu)}} \exp \left\{ -a_\mu x^{-\frac{\mu}{1-\mu}} \right\} \quad \text{for } x \rightarrow 0^+, \quad (1.91)$$

where $a_\mu = \frac{1-\mu}{\mu} [\mu/\cos(\pi\mu/2)]^{1/(1-\mu)}$ (cf. appendix of Ref. [6]).

- $\widehat{\mathcal{L}}_{1/2,1}(k) = e^{-\sqrt{2ik}}$ corresponds to the Lévy distribution ⁹

$$\mathcal{L}_{\frac{1}{2},1}(x) = \frac{1}{\sqrt{2\pi} x^{3/2}} e^{-1/2x} \quad \text{for } x > 0. \quad (1.92)$$

- Penson and Górska [58] have recently obtained analytic expressions for the one-sided Lévy law with fractional exponent $\mathcal{L}_{p/q,1}(x)$ with $p, q \in \mathbb{N}^*$ and $\mu = p/q \in]0, 1[$ (few specific cases were known earlier, for $\mu = 1/2, 1/3, 1/4, 3/4$).
- For $\mu \in]1, 2[$, the support of $\mathcal{L}_{\mu,+1}(x)$ is $\mathbb{R}$ (and *not* $\mathbb{R}_+$), although the power law only exists on $\mathbb{R}_+$ (and the decay is faster on $\mathbb{R}_-$). Example :

$$\mathcal{L}_{\frac{3}{2},1}(x) = \frac{\sqrt{3}}{\sqrt{\pi}|x|} e^{\frac{x^3}{27}} \begin{cases} W_{\frac{1}{2},\frac{1}{6}}\left(-\frac{2x^3}{27}\right) & \text{for } x \leq 0 \\ \frac{1}{6} W_{-\frac{1}{2},\frac{1}{6}}\left(\frac{2x^3}{27}\right) & \text{for } x \geq 0 \end{cases} \simeq \begin{cases} \frac{1}{3} \sqrt{\frac{-2x}{\pi}} e^{\frac{2x^3}{27}} & \text{for } x \rightarrow -\infty \\ \frac{3}{2\sqrt{2\pi} x^{5/2}} & \text{for } x \rightarrow +\infty \end{cases} \quad (1.93)$$

where $W_{\mu,\nu}(z)$ is the Whittaker function.

- $\mathcal{L}_{2,\beta}(x) = \mathcal{L}_{2,0}(x)$ is the Gaussian distribution.

Remark : The Lévy law $\mathcal{L}_{2,\beta}(x) = \mathcal{L}_{2,0}(x)$ is the Gaussian distribution. A distribution with power law tail $p(x) \sim |x|^{-3}$ is in the basin of attraction of the Gaussian distribution, although the variance is infinite (the typical fluctuations of the sum S_N scale as $\sqrt{N \ln N}$). The characteristic function presents an additional logarithmic term, i.e. exhibits a behaviour of the type $\hat{p}(k) \simeq 1 - k^2(1 - \ln|k|)$ for $k \rightarrow 0$. The inverse Fourier transform of $e^{-k^2(1 - \ln|k|)}$ is Gaussian in the central part, with non Gaussian large deviation tails.

✎ **Exercise 1.28 – Tail of $P_N(s)$:** We consider a symmetric distribution with power law tail $p(x) \simeq c|x|^{-1-\mu}$ for $x \rightarrow \infty$, for $\mu \in]0, 2[$. Its Fourier transform behaves as $\hat{p}(k) \simeq 1 - \tilde{c}|k|^\mu$ for $k \rightarrow 0$, where $\tilde{c} = -2c\Gamma(-\mu)\cos(\pi\mu/2) > 0$ (Exercise 1.13). Using the asymptotic (1.89), check that the asymptotic of the distribution of the sum is

$$P_N(s) \underset{x \rightarrow \pm\infty}{\sim} N c |s|^{-1-\mu}. \quad (1.94)$$

⁹To check this, it is more easy to compute the Laplace transform of the distribution, which is related to the MacDonald function $K_{1/2}(\sqrt{2\beta})$, leading to $\int_0^\infty dx \mathcal{L}_{1/2,1}(x) e^{-\beta x} = e^{-\sqrt{2\beta}}$.

Stability and universality classes.— Lévy has classified the different universal laws by identifying the functions which are stable against convolution (addition of the i.i.d. random variables). This is clear by inspection of the Fourier transform, for example $\widehat{\mathcal{L}}_{\mu,\beta}(k)^N = \widehat{\mathcal{L}}_{\mu,\beta}(N^{1/\mu}k)$, meaning that the result of N convolutions is the same function, up to a rescaling. The generalised central limit theorem shows that these laws are not only stable, they are also **attractors** under the action of many convolutions. In other terms, in the space of distributions, the exponent μ of the tail and the symmetry index β , which are stable properties for the convolution, define a universality class : the action of many convolutions on any distribution of a class (μ, β) ends to the Lévy law $\mathcal{L}_{\mu,\beta}(x)$, which is the attractive fixed point.

This illustrates the idea of universality classes for critical phenomena. Here the convolution plays the role of the renormalisation group transformation.

✎ **Exercise 1.29 – :** Consider a set of N i.i.d. random variables X_i distributed according to $p(x)$.

a) When $p(x) = A_1 \frac{1}{(x^6 + 4x^4 + x^2 + 1)^{1/3}}$ for $x \in \mathbb{R}$, what is $P_N(s)$ for $N \gg 1$? (A_1 is a normalisation).

b) Same question for $p(x) = A_2 \frac{2x^2 + 3x + \sqrt{x}}{(7x^7 + 5x^5 + 3)^{1/2}}$ for $x \in \mathbb{R}^+$.

c) Same question for $p(x) = A_3 \frac{3 + \tanh x}{(|x|^5 + x^2 + 2)^{1/4}}$ for $x \in \mathbb{R}$.

✎ **Exercise 1.30 – Condition for CLT:** Consider a set of i.i.d. random variables with common distribution $p(x)$. In the book of Gnedenko and Kolmogorov [31] (cf. p. 172), the condition for the CLT to hold is

$$\frac{\Lambda^2 \int_{|x|>\Lambda} dx p(x)}{\int_{|x|<\Lambda} dx x^2 p(x)} \xrightarrow{\Lambda \rightarrow \infty} 0 \quad (1.95)$$

Discuss the case of an exponential tail $p(x) \sim e^{-a|x|}$ and the case of a power law tail $p(x) \sim |x|^{-\mu-1}$ (with a special attention to the marginal case $\mu = 2$).

c) Scaling of S_N in the general case

Let us now discuss the scaling of S_N with N for power law distribution $p(x) \sim |x|^{-1-\mu}$. I denote $\overline{X} \equiv \langle X_i \rangle$ when it exists.

	CLT	μ_1	μ_2	$P_N(s)$	F
$\mu > 2$	yes	$< \infty$	$< \infty$	$P_N(s) \simeq \frac{1}{\sqrt{N}} F\left(\frac{s - N\overline{X}}{\sqrt{N}}\right)$	Gaussian
$\mu = 2$	yes	$< \infty$	∞	$P_N(s) \simeq \frac{1}{\sqrt{N \ln N}} F\left(\frac{s - N\overline{X}}{\sqrt{N \ln N}}\right)$	Gaussian
$1 < \mu < 2$	no	$< \infty$	∞	$P_N(s) \simeq \frac{1}{N^{1/\mu}} F\left(\frac{s - N\overline{X}}{N^{1/\mu}}\right)$	$\mathcal{L}_{\mu,\beta}(x)$
$\mu = 1$	no	∞	∞	$P_N(s) \simeq \frac{1}{N} F\left(\frac{s - c N \ln N}{N}\right)$	$\mathcal{L}_{1,\beta}(x)$
$0 < \mu < 1$	no	∞	∞	$P_N(s) \simeq \frac{1}{N^{1/\mu}} F\left(\frac{s}{N^{1/\mu}}\right)$	$\mathcal{L}_{\mu,\beta}(x)$

To learn more : The review article by Bouchaud and Georges [6] contains a nice appendix (I borrowed few remarks from it).

✎ **Exercise 1.31 – The marginal case $\mu = 1$:** We discuss the distribution of the sum corresponding to the case $\mu = 1$.

a) We first consider the case of random variables with a distribution given by the stable Lévy law $p(x) = \mathcal{L}_{1,\beta}(x)$, with characteristic function (1.87). Show that the distribution of S_N is

$$P_N(s) = \frac{1}{N} \mathcal{L}_{1,\beta} \left(\frac{s - \frac{2\beta}{\pi} N \ln N}{N} \right) \quad (1.96)$$

b) Consider now random variables with an arbitrary distribution with tail $p(x) \sim x^{-2}$ (with different weights for $x \rightarrow +\infty$ and $x \rightarrow -\infty$). Discuss the distribution of the sum $P_N(s)$.

1.9 Extreme value statistics

The study of extreme events has become very important these last decades, in physics (records, avalanches,...) [48], in climate (hurricane, typhoon,...), economy (market cracks,...), earthquakes, etc. In the paragraph 1.7, we have been interested in the rare events corresponding to atypical values of the sum of N i.i.d. random variables. Here, the “*extreme value statistics*” problem corresponds to analyze the distribution of the largest (or smallest) among N variables. More generally we can consider also the distribution of the second largest, etc, hence one sometimes refers to “*ordered statistics*”. The case of independent identically distributed (i.i.d) random variables is well documented since the pioneering work of Fréchet [27] and Gumbel [34]. We will only focus on this case here (see books [35, 36, 19]). For a recent review discussing the case of correlated variables in the context of statistical physics, see Ref. [48].

The question is as follows : consider a set of N random variables $X_1, \dots, X_N$ with the same law $f(x)$ and joint distribution $\mathcal{P}_N(x_1, \dots, x_N)$. What is the distribution of the maximum

$$M_N = \max(X_1, \dots, X_N) ?$$

In general, the cumulative distribution of the maximum involves a multiple integral of the joint distribution

$$\Phi_N(x) = \text{Proba}\{M_N \leq x\} = \text{Proba}\{X_1 \leq x \ \& \ \dots \ \& \ X_N \leq x\} \quad (1.97)$$

$$= \int_{-\infty}^x dx_1 \dots \int_{-\infty}^x dx_N \mathcal{P}_N(x_1, \dots, x_N) \quad (1.98)$$

which can be extremely difficult to analyze for large N .



Figure 1.12: Maurice Fréchet (1878-1973) and Emil Julius Gumbel (1891-1966).

If now the random variables are **independent**, the joint distribution factorizes $\mathcal{P}_N(x_1, \dots, x_N) = \prod_{i=1}^N f(x_i)$ and the multiple integral is separable. Introducing the cumulative distribution of one random variable

$$F(x) = \int_{-\infty}^x dy f(y) = \text{Proba}\{X_n \leq x\} \quad (1.99)$$

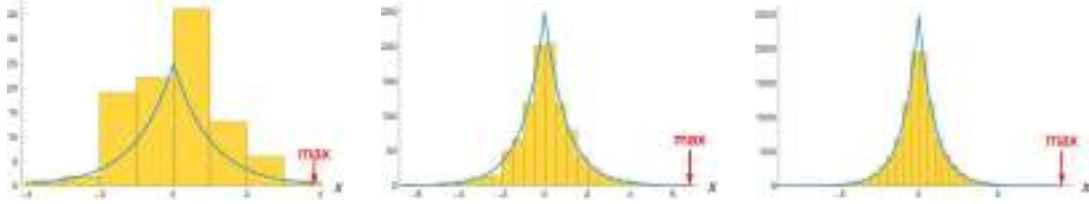
we have the simple form

$$\Phi_N(x) = [F(x)]^N. \quad (1.100)$$

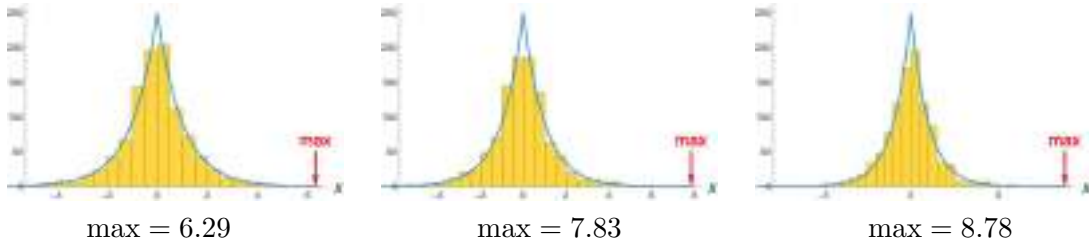
As we will see, the analysis of this distribution for $N \rightarrow \infty$ is interesting and reveals some universal character.

a) Typical value of the maximum

The first question is : if we pick N i.i.d. variables at random, where is typically the maximum ? For N variables, we can draw an histogram (the “empirical distribution”), which will get closer to $N f(x)$ as N grows. This is what is shown on the following plots, where I have picked N variables with distribution $f(x) = \frac{1}{2}e^{-|x|}$. I have produced three histograms for $N = 100$, $N = 1000$ and $N = 10000$:



The red arrow shows the position of the maximum M_N , which clearly grows with N . Here, for the exponential distribution, we show below that it grows like $\ln N$ (this is consistent with the figure since $\ln 100 \simeq 4.6$, $\ln 1000 \simeq 6.9$ and $\ln 10000 \simeq 9.2$). The maximum also fluctuates between different realisations of the N variables, for a given N . Here, several realizations of histograms for $N = 1000$ are



In general, the maximum is typically at a_N given by :

$$F(a_N) = 1 - \frac{1}{N} \quad (1.101)$$

Interpretation of this equation : the histogram is close to the average density for N variables $\bar{\rho}_N(x) = N f(x)$. The average number of variables above x is ($\#$ of variables above x) = $\int_x^\infty dx' \bar{\rho}_N(x') = N [1 - F(x)]$. The position of the largest variable is typically expected to be $M_N \sim a_N$ where a_N is given by $1 = N [1 - F(a_N)]$ (one variable above a_N).

Let us give few examples in the following table :

Distribution	$f(x)$	$F(x)$	a_N	b_N
Exponential	$(1/a)e^{-x/a}$ for $x > 0$	$1 - e^{-x/a}$	$a \ln(N)$	a
Gaussian	$\frac{1}{\sqrt{2\pi}a}e^{-x^2/(2a^2)}$	$\frac{1}{2}[1 + \text{erf}(x/\sqrt{2}a)]$	$a \sqrt{2 \ln(N)}$	$a/\sqrt{2 \ln(N)}$
Power law	$\mu A x^{-\mu-1}$ for $x \rightarrow \infty$	$1 - A x^{-\mu}$	$(A N)^{1/\mu}$	a_N/μ

For the power law tail, the scale of variation is given by a_N , hence there is only one relevant scale (i.e. $b_N \sim a_N$, where b_N is defined below as the scale over which the distribution varies significantly).

✎ **Exercise 1.32** – : Argue that $a_N \sim [\ln(N)]^{1/\alpha}$ for a distribution with tail $f(x) \sim \exp[-x^\alpha]$.

Consequence for the generalised central limit theorem : This table allows us to understand precisely the remark about the law of large numbers. We consider the sum of i.i.d. random variables $S_N = \sum_{n=1}^N X_n$. Let us denote by M_N the maximum of the N variables $\{X_n\}$.

- When the decay of the probability $f(x)$ is of exponential type, $f(x) \sim \exp[-x^\alpha]$, the typical value of the maximum scales as $M_N \sim [\ln(N)]^{1/\alpha} \ll N$, hence it cannot dominate the sum of N terms. All terms equally contribute to the average $\langle S_N \rangle \sim N$.
- When $f(x) \sim x^{-1-\mu}$, we have instead $M_N \sim N^{1/\mu}$. When $\mu > 1$, the scaling of the maximum is $\sim N^{1/\mu} \ll N$, and hence it does not dominate the sum S_N , which again scales as N ; the law of large numbers holds, $S_N/N \xrightarrow{N \rightarrow \infty} \langle X_i \rangle$.
- When $0 < \mu < 1$, the scaling of the maximum is faster than N , hence the sum is dominated by its maximum, not by the N terms. It scales as $S_N \sim M_N \sim N^{1/\mu} \gg N$. The law of large numbers breaks down.

Now we come back to the analysis of the distribution of the maximum and discuss the three universality classes.

b) Case of exponentially decaying distributions : Gumbel class

We consider the case where $f(x)$ presents a decay of exponential type at large x :

$$f(x) \sim e^{-x^\alpha} \quad \text{for } x \rightarrow \infty, \quad (1.102)$$

for $\alpha > 0$. As we have discussed, the maximum M_N of N such i.i.d. random variables is typical close to $a_N \sim (\ln N)^{1/\alpha}$. In the vicinity of a_N , we can approximate the law as

$$f(x = a_N + \epsilon) \underset{\epsilon \rightarrow 0}{\sim} e^{-a_N^\alpha - \alpha a_N^{\alpha-1} \epsilon} \quad \text{i.e.} \quad f(x) \underset{x \sim a_N}{\sim} e^{-a_N^\alpha} e^{-(x-a_N)/b_N} \quad (1.103)$$

where $b_N = a_N^{1-\alpha}/\alpha$ is a second characteristic scale which measures over which scale does the distribution vary significantly.¹⁰ A more general definition of this scale is thus $1/b_N = -f'(a_N)/f(a_N)$, or (almost) equivalently

$$1/b_N \stackrel{\text{def}}{=} \frac{F'(a_N)}{1 - F(a_N)} = N f(a_N). \quad (1.104)$$

This scale also measures the sensitivity of a_N as a function of N : indeed, differentiation of (1.101) with respect to N gives $f(a_N) \frac{da_N}{dN} = 1/N^2$ i.e.

$$b_N = \frac{da_N}{d \ln(N)}. \quad (1.105)$$

Examples are given in the table. An immediate consequence of this relation is that, for exponentially decaying distribution (1.102), we have

$$\boxed{\frac{b_N}{a_N} \simeq \frac{1}{\alpha \ln N}}, \quad (1.106)$$

which is shown below to characterize the relative fluctuations of the maximum.

¹⁰Note that the next term of the $\epsilon \rightarrow 0$ expansion in the exponential is $\frac{1}{2}\alpha(\alpha-1)a_N^{\alpha-2}\epsilon^2$, however the term $a_N^{\alpha-2}$ should be smaller than $a_N^{\alpha-1}$ as a_N grows with N , hence the $\mathcal{O}(\epsilon^2)$ is negligible on the scale b_N where the function decays significantly.

This discussion shows that the distribution behaves locally as an exponential, Eq. (1.103), when $x \sim a_N$. This is also true for the cumulative distribution :

$$F(x) \underset{x \sim a_N}{\simeq} 1 - \frac{1}{N} e^{-(x-a_N)/b_N} \quad (1.107)$$

(from its definition, $1/b_N$ is the decay rate of $F(x)$ around a_N). We stress that *we do not assume that the distribution is exponential*, however it decays *locally* exponentially over the scale b_N (consider the example of the Gaussian distribution, cf. Exercise 1.45). Introducing this form in the expression of the cumulative of the maximum, we have

$$\Phi_N(x) \underset{x \sim a_N}{\simeq} \left[1 - \frac{1}{N} e^{-(x-a_N)/b_N} \right]^N \simeq \exp \left[-e^{-(x-a_N)/b_N} \right] \quad (1.108)$$

where we considered the limit $N \rightarrow \infty$ with $(x-a_N)/b_N \sim \mathcal{O}(1)$. Hence, introducing the scaling variable

$$y = (x - a_N)/b_N \quad (1.109)$$

we deduce the scaling form

$$\Phi_N(x = a_N + b_N y) \underset{N \rightarrow \infty}{\simeq} \exp \left[-e^{-y} \right] \equiv \Psi(y) \quad \text{for } y \sim \mathcal{O}(1), \quad (1.110)$$

characterizing the cumulative distribution in the vicinity of a_N . This also shows that b_N measures the fluctuations of the maximum M_N .

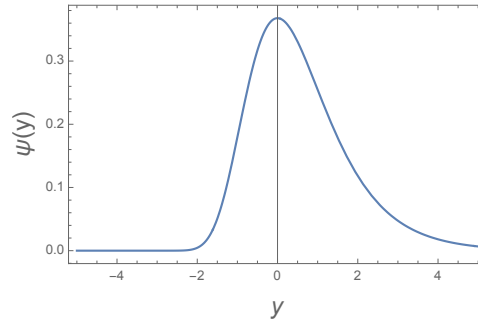


Figure 1.13: The Gumbel distribution (1.111).

The distribution of the maximum $\phi_N(x) = \Phi'_N(x)$ involves the *universal* scaling function $\psi(y) = \Psi'(y)$ known as the *Gumbel distribution*

$$\boxed{\psi(y) = \exp \left[-y - e^{-y} \right]} \quad (1.111)$$

It reaches its maximum for $y = 0$ (i.e. a_N is the value where the distribution of M_N is maximum). The distribution is strongly asymmetric : for $y \rightarrow +\infty$ its decay is reminiscent of the exponential decay of the distribution $f(x)$, while for $y \rightarrow -\infty$ it is strongly suppressed (exponential of exponential). As a result the distribution $\phi_N(x) = \Phi'_N(x)$ of the maximum M_N is

$$\boxed{\phi_N(x) \underset{N \gg 1}{\simeq} \frac{1}{b_N} \psi \left(\frac{x - a_N}{b_N} \right) \quad \text{for } \frac{x - a_N}{b_N} \sim \mathcal{O}(1),} \quad (1.112)$$

✎ **Exercise 1.33 – Mean and variance of M_N :** Compute $\langle M_N \rangle$ and $\text{Var}(M_N)$.

Hint : We give $\int_0^\infty dt (-\ln t) e^{-t} = -\mathbf{C}$ and $\int_0^\infty dt (-\ln t)^2 e^{-t} = \mathbf{C}^2 + \frac{\pi^2}{6}$, where $\mathbf{C} \simeq -0.577216$ is the Euler-Mascheroni constant.

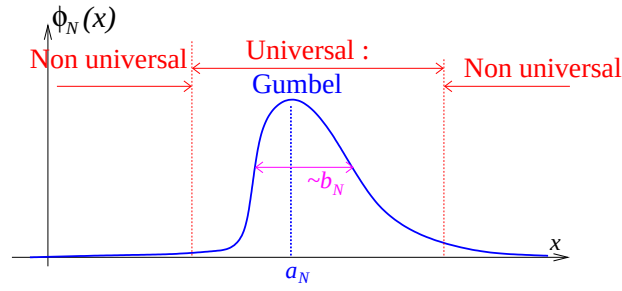


Figure 1.14: Sketch of the distribution $\phi_N(x) = \Phi'_N(x)$ of the maximum in the Gumbel class. The Gumbel distribution describes the universal (central) part of the distribution of width $\sim b_N$. There exists also non universal tails [similar to large deviation tails of $P_N(s)$].

Large deviations for the distribution of the maximum.— Note that the universal form (1.111) has been obtained by keeping $y \sim \mathcal{O}(1)$, while sending $N \rightarrow \infty$, hence it only describes the distribution close to the typical value of the maximum, in the “central part”, corresponding to the typical value of the maximum. Going further away, there exist non universal large deviation tails (cf. Exercise 1.45).

Considering first the limit $x \rightarrow \infty$ for fixed N , the maximum is isolated from the other values and we have $\phi_N(x) \sim f(x) \sim e^{-x^\alpha}$, which indeed depends on the precise distribution. This corresponds to the expansion of the general formula $\Phi_N(x) = [F(x)]^N = [1 - \int_x^\infty du f(u)]^N \simeq 1 - N \int_x^\infty du f(u)$, when x is sent to infinity *first*, i.e. the tail of the distribution is

$$\phi_N(x) \underset{x \rightarrow +\infty}{\simeq} N f(x), \quad (1.113)$$

i.e. is *non universal*, as it depends on the “microscopic details” (the distribution f). The large deviation tail for $y \rightarrow -\infty$ is less straightforward to discuss as this corresponds to rare configurations where the N random values are *all* pushed to the left, below x .

✎ **Exercise 1.34 – Maximum for the exponential distribution:** Consider the distribution $f(x) = e^{-x}$ for $x > 0$. Compute $F(x)$ and deduce explicitly $\Phi_N(x)$. Deduce a_N and b_N .

✎ **Exercise 1.35 – Maximum for the Gaussian distribution:** Now consider a Gaussian distribution $f(x)$? i.e. $F(x) = 1 - \frac{1}{2}\text{erfc}(x/\sqrt{2})$. Find a_N and b_N .

We can also extend the analysis and ask the question of the distribution of the position the of $N - k + 1$ -th largest value, for $k = 1, 2, \dots$ (the maximum corresponding to $k = 1$). We have now to consider

$$\begin{aligned} \Phi_{N,k}(x) &= \text{Proba}\{X_1 \leq x \ \& \ \dots \ \& \ X_{N-k} \leq x \ \& \ X_{N-k+1} \geq x \ \& \ \dots \ \& \ X_N \geq x\} \\ &= [F(x)]^{N-k+1} [1 - F(x)]^{k-1} \end{aligned} \quad (1.114)$$

i.e. $\Phi_N(x) \equiv \Phi_{N,1}(x)$ for the cumulative distribution of the largest value. Introducing again (1.107) we obtain the scaling form

$$\Phi_{N,k}(x = a_N + b_N y) \underset{N \rightarrow \infty}{\simeq} (\dots) \exp[-(k-1)y - k e^{-y}] \equiv \Psi_k(y) \quad (1.115)$$

for $y \sim \mathcal{O}(1)$. The related distribution is $\Psi'_k(y) = \psi_k(y)$ with

$$\psi_k(y) = \frac{k^k}{(k-1)!} \exp[-k y - k e^{-y}]. \quad (1.116)$$

The typical value (max of the distribution) is still for $y = 0$.

c) Case of distributions with power law tails : Fréchet class

The second universality class corresponds to distributions with power law tail

$$f(x) \simeq \mu A x^{-1-\mu} \quad \text{for } x \rightarrow \infty. \quad (1.117)$$

As shown in the table above, the typical position of the maximum among N variable is $a_N = (AN)^{1/\mu}$. Because the distribution is a power law, it exhibits scale invariance properties and is only characterised by *one* scale. We can check that the scale b_N introduced above, Eqs. (1.104, 1.105), is here $b_N = a_N/\mu \sim a_N$ (same N -dependence as a_N). Now, we can rewrite the cumulative distribution $F(x) \simeq 1 - Ax^{-\mu}$ for $x \rightarrow \infty$ as

$$F(x) \underset{x \gtrsim a_N}{\simeq} 1 - \frac{1}{N} \left(\frac{a_N}{x} \right)^\mu. \quad (1.118)$$

Contrary to the Gumbel case, where the exponential form (1.107) only holds for $|x - a_N| \lesssim b_N$, the behaviour (1.118) is valid both in the neighbourhood of a_N , $x \sim a_N$ and in the tail $x \gg a_N$. Inserting this form in $\Phi_N(x) = [F(x)]^N$ we eventually obtain the scaling form

$$\Phi_N(x = a_N y) \underset{N \rightarrow \infty}{\simeq} e^{-y^{-\mu}} \equiv \Psi(y) \quad (1.119)$$

where $y \sim \mathcal{O}(1)$ (i.e. y does not scale as N to some power).

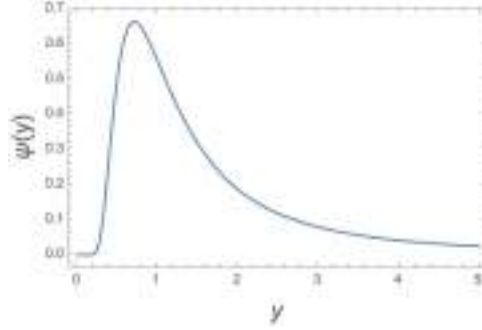


Figure 1.15: The Fréchet distribution (1.120) for $\mu = 3/2$.

The corresponding (Fréchet) distribution is $\psi(y) = \Psi'(y)$

$$\psi(y) = \frac{\mu}{y^{\mu+1}} e^{-y^{-\mu}}. \quad (1.120)$$

I.e. we have obtained another universal function describing the distribution of the maximum for variable with power law distribution. In conclusion, the distribution of the maximum is now

$$\boxed{\phi_N(x) \underset{N \gg 1}{\simeq} \frac{1}{a_N} \psi\left(\frac{x}{a_N}\right) \quad \text{for } x/a_N \gtrsim \mathcal{O}(1),} \quad (1.121)$$

The tail for large y corresponds to (1.113). Interestingly, we see that for power law distribution, the universal behaviour extends to infinity (which was not the case for the Gumbel class).

d) Weibull class

The last universality class corresponds to a distribution defined over a bounded domain. For example, consider positive random variables $X_n > 0$ for a distribution such that $f(x) \sim x^{\mu-1}$ for $x \rightarrow 0$ (with $\mu > 0$). We now discuss the distribution of the *minimum* of N such random

variables. This situation can be analysed straightforwardly by noticing that this is dual to the Fréchet case doing $x \leftrightarrow 1/x$, i.e. $\mu \rightarrow -\mu$. As a result, the distribution of the *minimum* is given by the universal scaling function

$$\psi(y) = \mu y^{\mu-1} e^{-y^\mu} . \quad (1.122)$$

☺ Important points

- *Basic concepts* : random variable, moments, cumulants, generating functions, independence, conditioning,...
- The case of *divergent moments* : distribution with power law tail $p(x) \simeq c|x|^{-1-\mu}$ for $x \rightarrow \pm\infty$ is related to $\hat{p}(k) \simeq 1 - \tilde{c}|k|^\mu$ for $k \rightarrow 0$ when $\mu \in]0, 1[$, $\hat{p}(k) \simeq 1 - ik\bar{x} - \tilde{c}|k|^\mu$ when $\mu \in]1, 2[$, etc.
- *Gaussian random variables* : relate correlations to the measure.
- *Central limit theorem and its generalisations* for power law distributions. For a distribution with power law tail $p(x) \simeq c|x|^{-1-\mu}$, the typical scaling of the fluctuations of N i.i.d. variables is $N^{1/2}$ for $\mu > 2$ and $N^{1/\mu}$ for $\mu < 2$. For $\mu \in]1, 2[$, the sum is dominated by the mean $S_N \simeq N\bar{X} + \text{fluct. of order } N^{1/\mu}$. For $\mu \in]0, 1[$, the average does not exist, $S_N \sim \text{fluct. of order } N^{1/\mu}$.
- *Large deviations*, large deviation function : distinction between typical and atypical fluctuations.
- *Extreme value statistics* : Typical position a_N of the maximum of N i.i.d. variable is given by $F(a_N) = 1 - 1/N$, where $F(x) = \int^x dt p(t)$ is the cumulative.

Additional exercises

✎ **Exercise 1.36 – Student distribution :** The student distribution of index $\mu > 0$ is defined as

$$S^{(\mu)}(x) = c \left(1 + x^2/\mu\right)^{-\frac{1+\mu}{2}} \quad (1.123)$$

It is symmetric and presents the power law tail $S^{(\mu)}(x) \simeq c|x|^{-1-\mu}$.

- Compute the normalisation constant
- We denote by $C^{(\mu)}(x) = \int_{-\infty}^x dt S^{(\mu)}(t)$ the cumulative distribution. Give $C^{(1)}(x)$.
- check that $C^{(2)}(x) = \frac{1}{2} + \frac{x}{2\sqrt{2+x^2}}$.
- What is $\lim_{\mu \rightarrow \infty} S^{(\mu)}(x)$?

✎ **Exercise 1.37 – Generate a Gaussian variable (Box-Muller algorithm) :** We discuss how to generate a Gaussian random number with a computer (preliminary : you should study at least the first questions of Exercise 1.2).

We consider two i.i.d. Gaussian random variables X and Y with zero mean and unit variance.

- What is the distribution of the radius $R = \sqrt{X^2 + Y^2}$? What is the distribution of the angle Θ ?
- What is the distribution of $\xi = \frac{1}{2}(X^2 + Y^2)$?
- Deduce a method to generate a Gaussian random number from a box distribution.

✎ **Exercise 1.38 – Minimal kurtosis :** Let us consider a random variable X with distribution $p(x)$. We demonstrate in this exercise that the kurtosis $\mathcal{K} \stackrel{\text{def}}{=} \kappa_4/(\kappa_2)^2$ is bounded from below.

Without loss of generality, we assume $\langle X \rangle = 0$, hence $\kappa_4 = \mu_4 - 3(\mu_2)^2$, where μ_n are the moments.

- Show that $\mu_4 \geq \mu_2^2$. Deduce an inequality for the kurtosis.
- Consider the symmetric exponential law $p(x) = \frac{1}{2}e^{-|x|}$. Compute the moments μ_n 's and deduce the kurtosis $\mathcal{K}$.
- Consider the Bernoulli distribution $B(x) = \frac{1}{2}[\delta(x-1) + \delta(x+1)]$. Compute the generating function $G(k)$ and the moments μ_n 's.
- Give a distribution for which $\mathcal{K}$ reaches its lower bound.

✎ **Exercise 1.39 – Conditional distribution for i.i.d. random variables :** Consider N i.i.d. random variables $X_1, \dots, X_N$ with distribution $f(x)$. Express formally the joint distribution conditioned on the sum $S = \sum_{i=1}^N X_i$, denoted $P(x_1, \dots, x_N | S)$. Deduce an expression for the marginal distribution conditioned on the sum, $p(x | S)$, in terms of $f(x)$ and $\Omega_N(S) = \int dx_1 \dots dx_N f(x_1) \dots f(x_N) \delta(S - \sum_{i=1}^N x_i)$.

✎ **Exercise 1.40 – Heteroscedasticity :** Let us consider the return X (in finance), assumed to be a centered Gaussian variable with variance $\langle X^2 \rangle = \sigma^2$. During periods of uncertainty, the volatility σ usually increases, compared to quiet periods (the phenomenon of variance inhomogeneity is called heteroscedasticity). Denoting $Q(\sigma)$ the distribution of the volatility, the distribution of return takes the form

$$P(x) = \int_0^\infty d\sigma Q(\sigma) \frac{e^{-x^2/(2\sigma^2)}}{\sqrt{2\pi}\sigma} \quad (1.124)$$

- Analyze the fourth cumulant κ_4 .
- Consider a Gaussian distribution of volatility $Q(\sigma) = \sqrt{2/\pi} e^{-\sigma^2/2}$. Show that $P(x)$ is related

to a MacDonald function (see Appendix A). Analyze the limiting behaviours.

c) Consider now the volatility distribution with tail of the form $Q(\sigma) \sim e^{-\sigma^\alpha/\alpha}$ for $\alpha > 0$. Use a saddle point approximation to deduce the tail of $P(x)$.

✎ **Exercise 1.41 – Discrete Ornstein-Uhlenbeck process :** We consider random Gaussian variables $(\dots, \phi_n, \dots)$ with probability weight $P(\phi) \propto \exp[-S]$ where the action is

$$S = \frac{1}{2} \sum_{t \in \mathbb{Z}} [(\phi_{t+1} - \phi_t)^2 + \mu^2 \phi_t^2] \quad (1.125)$$

Write the action as $S = \frac{1}{2} \phi^T A \phi$ and show that the matrix A involves the discrete Laplace operator $\Delta_{n,m} = \delta_{n,m+1} - 2\delta_{n,m} + \delta_{n,m-1}$.

Give the eigenvalues and (normalised) eigenvectors of Δ on the infinite line ($n \in \mathbb{Z}$). Deduce the correlation function $\langle \phi_t \phi_{t'} \rangle$.

Discuss the limit $\mu \rightarrow 0$.

Hint : we give the integral $\int_0^{2\pi} \frac{d\theta}{2\pi} \frac{\sinh \lambda}{\cosh \lambda + \cos \theta} e^{in\theta} = e^{-\lambda|n|}$.

✎ **Exercise 1.42 – Gaussian random walk and the Wiener measure :** We consider N i.i.d. Gaussian random numbers of zero mean and unit variance, $\xi_1, \dots, \xi_N$.

a) Write the joint PDF $W_N(\xi_N, \dots, \xi_1)$.

We now consider the Gaussian random walk (RW)

$$x_n = x_{n-1} + \xi_n \quad \text{with } x_0 = 0. \quad (1.126)$$

b) Argue that the Jacobian related to the change of variables $\xi_1, \dots, \xi_N \rightarrow x_1, \dots, x_N$ is unity.

c) Give the joint PDF $P_N(x_N, \dots, x_1)$ of the RW. Express the distribution under the form $P_N(X) = C_N e^{-\frac{1}{2} X^T A_N X}$ and give the $N \times N$ matrix A_N and the normalisation constant C_N .

d) Argue that the probability for a configuration of the RW passing through the sequence of positions $(x_1, x_2, \dots, x_N)$ has the form

$$P_N(x_N, \dots, x_1) = \left(\prod_{i=2}^N P(x_i | x_{i-1}) \right) P_1(x_1) \quad (1.127)$$

e) Wiener measure : what is the form of the measure $P_N(X) d^N X \propto d^N X e^{-\frac{1}{2} X^T A_N X}$ in the continuum limit $x_t \rightarrow x(t)$?

✎ **Exercise 1.43 – N identical Gaussian variables :** Consider N Gaussian random variables $x_1, \dots, x_N$ of zero mean, with distribution $P(\mathbf{x}) \propto \exp \left\{ -\frac{1}{2} \mathbf{x}^T A \mathbf{x} \right\}$, where A is a positive definite real symmetric matrix. We assume known the correlation matrix $C_{ij} = \langle x_i x_j \rangle$: the variables are **identical**, $\langle x_1^2 \rangle = \dots = \langle x_N^2 \rangle = v$, but **not independent**, $\langle x_i x_j \rangle = c \ \forall (i, j), i \neq j$.

a) Relate the matrices A and C (in general).

b) Write the correlation matrix C in terms of $\mathbf{1}_N$ and the projector $\Pi = (1/N) X X^T$ on the vector $X = (1 \ 1 \ \dots \ 1)^T$.

c) Show that X is an eigenvector of C . Show that any vector $\perp$ to X is also an eigenvector. Deduce the spectrum of eigenvalues of C .

d) Deduce two constraints on the correlation c . Comment.

e) Compute the variance of the sum $S = \sum_i x_i$. Comment.

f) Give the general form for the measure $P(\mathbf{x})$ for identical (correlated) Gaussian variables.

✎ **Exercise 1.44 – Random trap model:** We consider a line with traps at regular positions. A particle on a site is trapped during a random time τ_α , and eventually jumps to one of the two neighbouring traps with probability $1/2$ (symmetric random walk with waiting times). At a given time t , the particle has done N_t jumps, and is typically at distance $x_t \sim N_t^{1/2}$ (the number of jumps N_t is random). The question is now to determine how the time t scales with the number of jumps N_t . We rather consider a fixed number of jumps N , then the time is random, given by

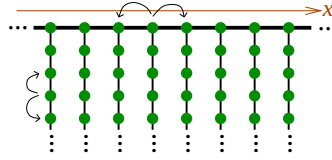
$$t = \sum_{\alpha=1}^N \tau_\alpha \quad (1.128)$$

where the waiting times are i.i.d. random variables with distribution $\psi(\tau)$.

a) Assuming the power law tail $\psi(\tau) \sim \tau^{-1-\mu}$ for $\tau \rightarrow \infty$, discuss how t scales with N depending on μ .

b) Deduce the nature of the random walk.

c) **Diffusion along the skeleton of a comb:** We consider a regular random walk between the nearest neighbour sites of a comb as shown on this figure :



In one dimension, the probability of first return to the starting point is $\psi(\tau) \sim \tau^{-3/2}$ (cf. § 5.8). We denote by x_t the position of the particle along the skeleton. Deduce how x_t scales with time.

✎ **Exercise 1.45 – Typical value and large deviations for the maximum of Gaussian variables:** We consider i.i.d. random Gaussian variables with zero mean and unit variance.

a) Argue that we can write the cumulative of the maximum under the form $\Phi_N(x) \simeq \exp \left[-\frac{N}{2} \operatorname{erfc}(x/\sqrt{2}) \right]$.

b) Compare the behaviour for $y = (x - a_N)/b_N \sim \mathcal{O}(1)$ and $x \sim \mathcal{O}(1)$. We give $a_N \simeq \sqrt{2 \ln N} - \frac{\ln(4\pi \ln N)}{2\sqrt{2 \ln N}}$ and $b_N \simeq 1/a_N$.

✎ **Exercise 1.46 – Concentration of measure for power law random variables – Application for the R.E.M.:** We consider N i.i.d. positive random variables $z_1, \dots, z_N$ with distribution $f(z)$ and denote Z their sum. The relative weights are $w_i = z_i/Z$. Let us introduce the moments $Y_k \stackrel{\text{def}}{=} \sum_{i=1}^N (w_i)^k$, which characterize the relative importance of the variables. The aim of the exercise is to study the average of these moments, $\langle Y_k \rangle$, when the distribution of the z_i 's has a power law tail $f(z) \simeq c z^{-1-\mu}$ for $z \rightarrow \infty$. We show that $\langle Y_k \rangle$ exhibits a phase transition tuned by μ

a) Preliminary.— compare the moments in two situations : (i) all weights are equal, $w_i = 1/N$. (ii) one weight dominates, e.g. $w_1 = 1/2$, with all other weights equal $w_i = w_j$ for $i, j > 1$.

b) now we come back to the case of random variables. Use the identity $Z^{-k} = \frac{1}{\Gamma(k)} \int_0^\infty dt t^{k-1} e^{-tZ}$, in order to express $\langle Y_k \rangle$ in terms of the characteristic function $G(\beta) \stackrel{\text{def}}{=} \langle e^{-\beta z_i} \rangle$.

c) We first consider the case $0 < \mu < 1$: recall the behaviour of $G(\beta)$ for $\beta \rightarrow 0$. Deduce that $\langle Y_k \rangle$ remains finite when $N \rightarrow \infty$.

d) We now consider the case $\mu > 1$: recall the behaviour of $G(\beta)$ for $\beta \rightarrow 0$ then. Deduce that $\langle Y_k \rangle \rightarrow 0$ as $N \rightarrow \infty$. Plot $\langle Y_2 \rangle$ as a function of μ . Plot also $\langle Y_3 \rangle$. Comment.

e) **Random Energy Model.**— The R.E.M. is an important model describing the statistical physics of disordered systems. The model has been introduced in 1980 by Derrida [20, 21].

Despite being the simplest model for a complex system, we will see that it has an interesting phenomenology. The REM is defined as follows : consider a system with $N = 2^M$ energy levels E_i , which are assumed i.i.d. random variables (originally, Derrida assumed a Gaussian distribution for the energies). This assumption was shown to correspond to a model of spin glass with random p -spin interactions between M spins. Here we can consider a generalized version of Derrida's model [7] such that the marginal distribution of energy levels is of exponential type, $p(E) \sim \exp\{-c|E|^\alpha\}$ for $E \rightarrow -\infty$. Argue that, for $N \gg 1$ variables, the distribution can be approximated by an exponential $p(E) \simeq (b_N N)^{-1} \exp\{(E - a_N)/b_N\}$ in the scaling region (i.e. when E is in the neighbourhood of the typical position a_N of the ground state energy). Recall how a_N and b_N behave with N . We consider the partition function $Z = \sum_{i=1}^N e^{-\beta E_i}$. Making the connection with the problem studied previously, show that the R.E.M. exhibits a phase transition tuned by the temperature $T = 1/\beta$. Interpret.

2 The Langevin equation for a particle in a fluid

The aim of this introductory chapter is to start the discussion of stochastic processes with a concrete and simple example, and introduce several general and important ideas. Consider a particle in a fluid, submitted to a friction force. The usual *phenomenological* model is friction proportional to the velocity (Stokes regime) :

$$F_f = -\gamma v, \quad (2.1)$$

where γ is the friction coefficient. For a spherical particle of radius R , fluid mechanics gives $\gamma = 6\pi\eta R$ where η is the viscosity of the fluid (for example, $\eta \simeq 10^{-3} \text{ kg.m}^{-1}.\text{s}^{-1}$ for water at $T = 20^\circ\text{C}$). In the absence of any other external force, the Newton equation of motion takes the form $m\dot{v} = -\gamma v$. The friction coefficient has dimension of a mass divided by a time, hence we can write

$$\gamma = \frac{m}{\tau} \quad (2.2)$$

where τ is the *relaxation time* for the velocity.

2.1 Fluctuations and Langevin force

In 1827, the scottish botanist Robert Brown observed with a microscope that pollen grains at the surface of water move erratically. ¹¹ It was understood later that this observation supports the atomist description of matter as it is the manifestation of the *fluctuations* in the fluid (erratic motion of the molecules). A clear description of the phenomenon was given much later by Albert Einstein in 1905. If the particle (the pollen grain) is small, it is not only submitted to the friction force but it is also sensitive to the *fluctuations in the fluid*, i.e. the collisions with molecules. The typical collision time between molecules in a fluid is $\tau_{\text{coll}} \sim 10^{-15} \text{ s}$, thus we expect that the Brownian particle experiences collisions with the rate $1/\tau_{\text{coll}}$ and which can be considered as *independent*. The friction force is due to the effect of these collisions over a much larger time scale. Additionally to the friction force, we model the frequent collisions by introducing in the Newton equation a force $\xi(t)$ fluctuating in time, called the “**Langevin force**” :

$$m \frac{dv(t)}{dt} = -\gamma v(t) + \xi(t) \quad (2.3)$$

$$\frac{dx(t)}{dt} = v(t) \quad (2.4)$$

Because the collisions are exerted at random along all directions, we expect that

$$\langle \xi(t) \rangle = 0 \quad (2.5)$$

where $\langle \dots \rangle$ denotes *statistical averaging* ¹² (it is also true if we consider averaging over time for a single history). As the Langevin force models the force exerted on the particle by the molecules, it is natural to assume short time correlations $\langle \xi(t)\xi(t') \rangle = \frac{C}{\tau_{\text{coll}}} \varphi((t-t')/\tau_{\text{coll}})$ where φ is normalised function of width ~ 1 centered on the origin (like $(1/2)e^{-|x|}$ or $\pi^{-1/2}e^{-x^2}$) and C

¹¹Brown was not the first one to observe this phenomenon, but he was the first to study it systematically (varying parameters of the experiment). You can find some historical perspectives in the excellent article of Bertrand Duplantier [25].

¹²Statistical averaging corresponds to average over **different histories** of the particle, with same initial conditions but different realisations of the Langevin force. This is the procedure followed by Jean Perrin (Nobel prize in 1926) in his experiments [59] ; cf. Fig. 2.3. The difference between statistical averaging and time averaging is described more concretely in Exercise 2.7.

the strength of the fluctuations. As we are interested in the dynamics of the Brownian particle over time $\gg \tau_{\text{coll}}$ we can simply consider

$$\langle \xi(t)\xi(t') \rangle = C \delta(t - t') \quad (2.6)$$

(which corresponds formally to $\tau_{\text{coll}} \rightarrow 0$). A random function characterised by such local correlations is called a “white noise”.



Figure 2.1: *Robert Brown (1773-1858), Albert Einstein (1879-1955), Paul Langevin (1872-1946) and Jean Perrin (1870-1942).*

This model was introduced by Paul Langevin [42]. Why studying a model for the motion of a pollen grain at the surface of a fluid (or more generally a “colloid” in a fluid) is an important problem ? The reason is that several ideas of the Langevin model have a much broader application in out-of-equilibrium statistical physics.

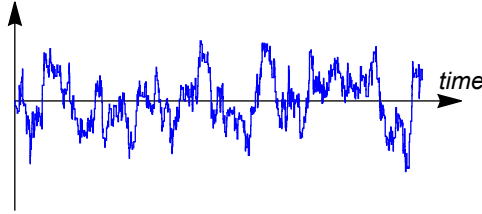


Figure 2.2: *A realization of the Ornstein-Uhlenbeck process $v(t)$ defined by (2.3) when $\xi(t)$ is a Gaussian white noise.*

We now analyse the statistical properties of the particle. Taking advantage that the equation of motion is *linear*, its integration gives

$$v(t) = v(0) e^{-t/\tau} + \frac{1}{m} \int_0^t dt' \xi(t') e^{-(t-t')/\tau} . \quad (2.7)$$

This representation makes easy to deduce the statistical properties of $v(t)$ from those of $\xi(t)$. If the initial velocity is non random, we have

$$\langle v(t) \rangle = v(0) e^{-t/\tau} . \quad (2.8)$$

After a time larger than $\tau = m/\gamma$, the memory of the initial velocity is lost and the velocity is independent of $v(0)$. We also get the correlator $\langle v(t)v(t') \rangle_c \stackrel{\text{def}}{=} \langle v(t)v(t') \rangle - \langle v(t) \rangle \langle v(t') \rangle$:

$$\langle v(t)v(t') \rangle_c = \frac{1}{m^2} \int_0^t dt_1 e^{-(t-t_1)/\tau} \int_0^{t'} dt_2 e^{-(t'-t_2)/\tau} \langle \xi(t_1)\xi(t_2) \rangle = \frac{C e^{-(t+t')/\tau}}{m^2} \int_0^{\min(t,t')} dt_1 e^{2t_1/\tau} \quad (2.9)$$

thus

$$\langle v(t)v(t') \rangle_c = \frac{C \tau}{2m^2} \left(e^{-|t-t'|/\tau} - e^{-(t+t')/\tau} \right) . \quad (2.10)$$

The correlations decay in time over the same time scale τ like the average time. The decorrelation of the velocity can be understood from the fact that the velocity decorrelates from the noise on the same time scale : we check easily that

$$\langle v(t)\xi(t') \rangle = \frac{C}{m} \theta_H(t-t') e^{-(t-t')/\tau}. \quad (2.11)$$

The velocity $v(t)$ is only correlated to the noise in the past, not too far (i.e. for $t-t' \lesssim \tau$).

The stochastic process described by (2.3) where $\xi(t)$ is a Gaussian white noise is known as the “*Ornstein-Uhlenbeck process*”.

2.2 A fluctuation-dissipation relation

After a sufficient long time, we expect that the particle is at thermal equilibrium with the fluid (which should be at thermal equilibrium), hence $\langle v(t)^2 \rangle = k_B T/m$ (equipartition theorem). On the other hand, the outcome of the Langevin analysis is $\langle v(t)^2 \rangle = C/(2m\gamma)$. Consistency of the previous analysis with equilibrium statistical physics imposes a constraint between the strength C of the Langevin noise (the fluctuations in the fluid), the friction coefficient γ and the temperature :

$$C = 2 \overset{\substack{\text{dissipation} \\ \uparrow}}{\gamma} k_B T \underset{\substack{\text{fluctuations} \\ \downarrow}}{\quad} \quad (2.12)$$

this is a “*fluctuation-dissipation relation*” and a first formulation of the **fluctuation-dissipation theorem** (FDT). The mass has disappeared, which emphasizes that this is a property of the thermal bath (the fluid). We could write the correlator of the noise

$$\langle \xi(t)\xi(t') \rangle = 2\gamma k_B T \delta(t-t'). \quad (2.13)$$

The phenomenological coefficients C , the strength of the Langevin force, and γ , the friction coefficient, are *not* two independent parameters (at least when thermal equilibrium holds). This reminds that friction and fluctuations have the same physical origin : the interaction of the particle with the molecules of the fluid. Below (§ p. 46), we will introduce a microscopic model of friction and try to clarify the origin of this relation.

✎ **Exercise 2.1 – Langevin equation for random initial velocity:** The correlator (2.10) corresponds to a fixed initial velocity. Consider now the case where the initial velocity is random, distributed according to $P(v_0) \propto \exp \left\{ -\frac{mv_0^2}{2k_B T} \right\}$.

a) Compute the new correlator, denoted $\langle v(t)v(t') \rangle^{\text{equil}}$.

b) Can we compare the two correlators ?

✎ **Exercise 2.2 – Stationary measure of the Ornstein-Uhlenbeck process:** In the stationary regime, compare the correlator with the one obtained in Exercise 1.41. Deduce what is the measure of the Ornstein-Uhlenbeck process.

In Eq. (2.10), the second term is a transient term coming from the initial condition $v(0) = 0$. Hence for times $t, t' \gg \tau$ we can write

$$\langle v(t)v(t') \rangle = \frac{k_B T}{m} e^{-|t-t'|/\tau} \quad (2.14)$$

The correlations (2.11) between velocity and Langevin force can also be rewritten in terms of the temperature

$$\langle v(t)\xi(t') \rangle = \frac{2k_B T}{\tau} \theta_H(t-t') e^{-(t-t')/\tau}. \quad (2.15)$$

2.3 Diffusion

Now that we have obtained the statistical properties of the velocity, let us analyze the statistical properties of the position. The average position is simply

$$\langle x(t) \rangle = v(0)\tau \left(1 - e^{-t/\tau}\right) \quad (2.16)$$

On average, after a time $t \gg \tau$, the particle covers a distance $v(0)\tau$, which is not surprising as the memory of the initial velocity $v(0)$ is lost after time τ . What about the fluctuations? In the stationary regime the correlator of the speed is a “narrow function” of width τ , with weight

$$\int_{-\infty}^{+\infty} d(t-t') \langle v(t)v(t') \rangle_c^{\text{equil}} = \frac{2k_B T}{\gamma}. \quad (2.17)$$

As a result, neglecting the transient regime at short times, we can write

$$\langle x(t)^2 \rangle_c = \int_0^t dt_1 \int_0^t dt_2 \langle v(t_1)v(t_2) \rangle_c \simeq t \int_{-\infty}^{+\infty} d(t_1 - t_2) \langle v(t_1)v(t_2) \rangle_c^{\text{equil}} = \frac{2k_B T}{\gamma} t \quad (2.18)$$

The linear behaviour characterizes **diffusive motion**. We introduce the *diffusion constant*

$$D \stackrel{\text{def}}{=} \lim_{t \rightarrow \infty} \frac{\langle x(t)^2 \rangle_c}{2t} \quad (2.19)$$

In other terms, we have obtained above a general relation between the velocity correlator and the diffusion constant

$$D \stackrel{\text{def}}{=} \int_0^\infty dt \langle v(t)v(0) \rangle \quad (2.20)$$

Coming back to Langevin model, Eqs. (2.3,2.4), we get the expression

$$D = \frac{k_B T}{\gamma} \quad (2.21)$$

which is known as the “*Einstein relation*”. It was obtained in Einstein 1905’s article on the Brownian motion.¹³ It is also known as the “Einstein-Stokes law” or “Einstein-Stokes-Sutherland law”. This is another formulation of the FDT, relating three different physical quantities, the diffusion constant characterizing the fluctuations of the motion, the friction coefficient characterizing the dissipation and the temperature.

Exercise 2.3 – : Choose initial conditions for a fixed initial velocity $x(0) = 0$ and $v(0) = v_0$. Compute $\langle x(t) \rangle$. Study precisely $\langle x(t)^2 \rangle_c$. Analyze the crossover between the short time and large time $\langle x(t)^2 \rangle_c \propto t$ diffusive behaviour.

2.4 Large scale properties and the overdamped regime

Over large time scales ($\gg \tau$), the correlator (2.14) seems a narrow function which can be replaced by a delta function

$$\langle v(t)v(t') \rangle_c^{\text{equil}} = \frac{C}{\gamma^2} \frac{e^{-|t-t'|/\tau}}{2\tau} \stackrel{\text{large scale}}{\approx} \frac{C}{\gamma^2} \delta(t-t') = \frac{1}{\gamma^2} \langle \xi(t)\xi(t') \rangle. \quad (2.22)$$

¹³A. Einstein, “Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen”, Annalen der Physik **322**(8), 549–560 (1905).

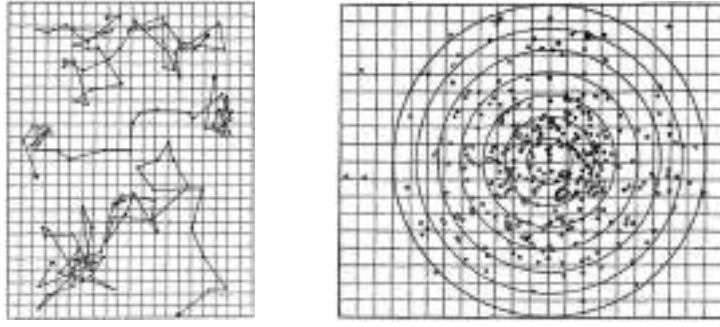


Figure 2.3: *Measurements of Jean Perrin (1908) ; from [59]. Left : few examples of trajectories. Right : final points after the several histories (for a fixed time).*

Furthermore, assuming that $\xi(t)$ is Gaussian,¹⁴ this corresponds to write

$$v(t) \stackrel{\text{large scale}}{\approx} \frac{1}{\gamma} \xi(t) \quad (2.23)$$

i.e. to neglect the acceleration term in Newton's equation :

$$0 \approx -\gamma v(t) + \xi(t) \quad (\text{overdamped regime}) . \quad (2.24)$$

This approximation is called the “*overdamped regime*”, which is achieved either by studying the process over large time scales, $t \gg \tau$, or by formally considering the limit of strong damping, $\gamma \rightarrow \infty$. As a result we obtain that the **velocity equals the force**.

Overdamped regime : “Aristotle equation”.— The pre-Galileo-Newtonian postulate, proposed by Aristotle, asserted that velocity is proportional to force.¹⁵ The above discussion has shown that this makes sense for the motion of a particle in a viscous fluid. If an additional (conservative) force $F(x)$ is introduced in the equation of motion for the overdamped regime (2.24), we have

$$\frac{dx(t)}{dt} \approx \frac{1}{\gamma} [F(x(t)) + \xi(t)] \quad (\text{overdamped regime}) . \quad (2.25)$$

We can call the equation (velocity) = (force)/(friction) the “*Aristotle equation*”, although this is a bit anachronistic (differential calculus was invented only at the end of XVIIIth century by Leibniz and Newton).

A last comment about the physical content of the overdamped regime, which we rewrite

$$\boxed{\frac{dx(t)}{dt} = -\frac{1}{\gamma} \frac{\partial V(x)}{\partial x} + \sqrt{2D} \eta(t)} \quad (\text{overdamped regime}) . \quad (2.26)$$

where $V(x)$ is the potential, such that $F(x) = -V'(x)$, and $\eta(t)$ a normalised Gaussian white noise, i.e. with $\langle \eta(t) \rangle = 0$ and $\langle \eta(t) \eta(t') \rangle = \delta(t - t')$. Contrary to the Newton equation, which describes a *conservative* dynamics, the Aristotle equation describes a *non conservative* dynamics, along which energy is dissipated (cf. Exercise 2.9). In particular, in the $D \rightarrow 0$ limit,

¹⁴Gaussian processes have only two finite cumulants κ_1 and κ_2 , all others being zero. Equality of averages and variances of two Gaussian processes implies that they have the same statistical properties.

¹⁵In the book “Aristotle's Physics”, it is claimed that the velocity of a falling body is proportional to its weight, i.e. the velocity is proportional to the force. In other terms the motion requires a force for Aristotle (384-322 BC). It required almost two thousand years to understand that force induces *acceleration*, not motion : motion can exist without force. This was the achievement of Galileo Galilei (1564-1642 AC).

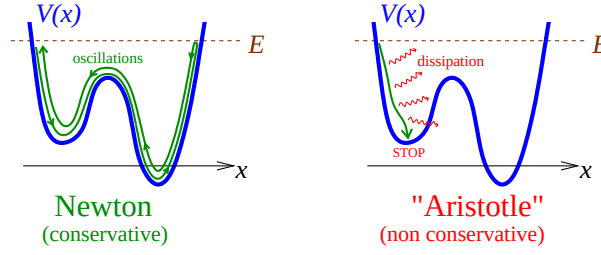


Figure 2.4: *The dynamics described by the Newton equation conserves the energy, whereas the energy is dissipated during the “Aristotle dynamics”.*

the trajectory converges towards the first stationary point (Fig. 2.4). If the first stationary point is a local minimum like on the figure, only a fluctuation (finite D) can untrap the particle in order to visit another neighbouring (local or global) minimum.

We will come back later to a general analysis of this stochastic differential equation.

2.5 The free Brownian motion (the Wiener process)

In the overdamped regime, in the absence of the external force $F(x)$, the position is just the integral of a Gaussian white noise

$$x(t) = \underbrace{x(0)}_{=0} + \frac{1}{\gamma} \int_0^t du \xi(u) \quad (2.27)$$

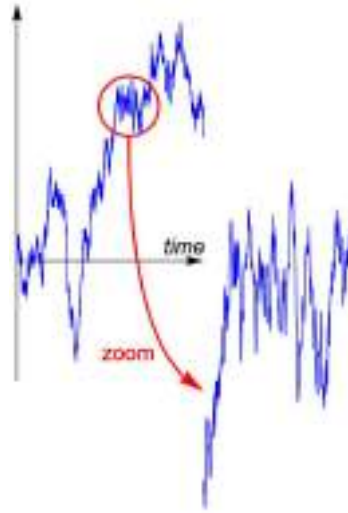


Figure 2.5: *A typical Brownian trajectory : $W(t)$ as a function of t . If one zooms, trajectory looks the same (scale invariance).*

Let us simplify the notations and introduce a normalised Gaussian white noise $\eta(t) = \xi(t)/\sqrt{C}$ so that

$$\langle \eta(t) \rangle = 0 \quad \text{and} \quad \langle \eta(t) \eta(t') \rangle = \delta(t - t'). \quad (2.28)$$

We now consider the normalised free Brownian motion, with no external force (the “Wiener process”) by setting $x(t) = \frac{\sqrt{C}}{\gamma} W(t)$. Obviously

$$\frac{dW(t)}{dt} = \eta(t) \quad \Rightarrow \quad W(t) = \int_0^t du \eta(u) \quad (2.29)$$

thus

$$\langle W(t)W(t') \rangle = \int_0^t du \int_0^{t'} dv \delta(u-v) = \int_0^{\min(t,t')} du \quad (2.30)$$

Finally

$$\boxed{\langle W(t)W(t') \rangle = \min(t, t')} \quad (2.31)$$

Interpretation : consider the case $t < t'$, we have $\langle W(t)W(t') \rangle = \langle [W(t') - W(t)]W(t) \rangle + \langle W(t)^2 \rangle$. The second term is t (diffusion) ; the first term vanishes as the two increments $W(t') - W(t) = \int_t^{t'} du \eta(u)$ and $W(t) = \int_0^t du \eta(u)$ are independent.

✎ **Exercise 2.4 - :** Check that the increment depends only on the time difference

$$\langle [W(t) - W(t')]^2 \rangle = |t - t'| \quad (2.32)$$

Continuity.— In the limit $t \rightarrow t'$ we get $\lim_{t \rightarrow t'} \langle [W(t) - W(t')]^2 \rangle = 0$. This last equation *suggests* that the curve $W(t)$ is **continuous**. As a matter of fact the two statements are disconnected : the fact that this limit is zero does not imply continuity (cf. Exercise 3.5)!

In order to prove continuity, we should consider the probability

$$\text{Proba}\{|W(t) - W(t')| < \epsilon\} = \text{erf}\left(\frac{\epsilon}{\sqrt{2|t - t'|}}\right). \quad (2.33)$$

This expression simply follows from the fact that the increment between times t and t' is Gaussian, with variance $|t - t'|$, hence the cumulative distribution related to the error function $\text{erf}(x)$ (see Appendix A). This expression shows that $\forall \epsilon > 0$, there exists a sufficiently small time difference, choose $|t - t'| \ll \epsilon^2$, so that we are sure that the distance is smaller than ϵ : $\text{Proba}\{|W(t) - W(t')| < \epsilon\} \simeq 1 - \mathcal{O}(e^{-\epsilon^2/(2|t - t'|)})$. Because we can choose ϵ as small as we want, this corresponds to continuity of the process.

Continuity can also be understood in a simpler manner by constructing the Brownian motion from the discrete random walk, $W(t) = \sum_{\tau=1}^{t/\epsilon} \delta W_\tau$ where $\langle \delta W_\tau \rangle = 0$ and $\langle \delta W_\tau \delta W_{\tau'} \rangle = \epsilon \delta_{\tau, \tau'}$, and taking the continuum limit of small jumps $\delta W_\tau \sim \sqrt{\epsilon} \rightarrow 0$ occuring with high rate $1/\epsilon \rightarrow \infty$.

✎ **Exercise 2.5 - :** Prove (2.33).

Non differentiability Another consequence of (2.32) is that

$$\left\langle \left(\frac{W(t) - W(t')}{t - t'} \right)^2 \right\rangle = \frac{1}{|t - t'|} \quad (2.34)$$

which goes to infinity when $t \rightarrow t'$: the derivative of typical curves is infinite, i.e. $W(t)$ is *non differentiable*. We can see on the plot that the curve is indeed extremely irregular on small scale (see Fig. 2.5).

Scale invariance.— This irregularity is related to *scale invariance* and fractal behaviour (with fractal dimension $1/2$). For a scaling factor $\alpha > 0$, it is clear that

$$\langle W(\alpha t)W(\alpha t') \rangle = \alpha \min(t, t') = \alpha \langle W(t)W(t') \rangle \quad (2.35)$$

Because $W(t)$ is Gaussian, all statistical information is encoded in the two point function, hence this equality means that we can identify the statistical properties of $W(\alpha t)$ with those of

$\sqrt{\alpha} W(t)$. Mathematicians express this through and “*equality in law*” (or “*equality in distribution*”)

$$\boxed{W(\alpha t) \stackrel{(\text{law})}{=} \sqrt{\alpha} W(t)} \quad (2.36)$$

(the “laws” of the two sides are equal). The same argument applies to the Gaussian white noise gives

$$\eta(\alpha t) \stackrel{(\text{law})}{=} \frac{1}{\sqrt{\alpha}} \eta(t). \quad (2.37)$$

The property is extended in Exercise [2.10](#).

☺ Important points

- Master the analysis of the linear Langevin equation (2.3) (integrate the eq., averages, etc)
- Fluctuation-dissipation relation (different forms) : Langevin force and damping force have the same origin, hence the relation.
- Wiener process : main properties.

APPENDIX : Microscopic foundations of the Langevin equation

The aim of this paragraph is to go beyond the phenomenological Langevin model and clarify the physical origin of the Langevin equation from a microscopic description. We introduce a model with a purely Hamiltonian (conservative) dynamics, within which will emerge the effective *dissipative* dynamics described by the Langevin equation. This will allow to identify the microscopic origin of the dissipation.

Model.— One could be tempted to model the collisions in the fluid, however the microscopic dynamics would be difficult to analyze. Instead, we consider a particle coupled to a macroscopic number of uncoupled harmonic oscillators modelling the “environment” (also called the “bath”). In this model, the oscillators represent the eigen-modes of the macroscopic system (like the phonon modes in a fluid). We now study the deterministic dynamics governed by the Hamiltonian

$$H = \frac{p^2}{2m} + V(x) + \sum_n \left[\frac{p_n^2}{2} + \frac{1}{2} \omega_n^2 \left(q_n - \frac{c_n x}{\omega_n^2} \right)^2 \right] \quad (2.38)$$

i.e.

$$H_{\text{sys}}(x, p) = \frac{p^2}{2m} + V(x) \quad (2.39)$$

$$H_{\text{env}}(\{q_n, p_n\}) = \sum_n \left[\frac{p_n^2}{2} + \frac{1}{2} \omega_n^2 q_n^2 \right] \quad (2.40)$$

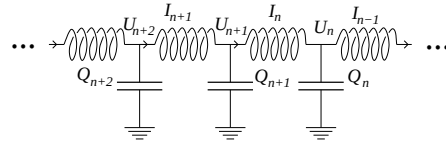
The crucial assumption here is that the coupling between the system and the bath is *linear*

$$H_{\text{int}} = -x \sum_n c_n q_n + \frac{1}{2} x^2 \sum_n \frac{c_n^2}{\omega_n^2} \quad (2.41)$$

where c_n are coupling constants.

Physical realizations are : an electron in an atom, coupled to the electromagnetic modes, or an electric device coupled to a L-C line, etc.

✎ Exercise 2.6 – Dissipation in a transmission line: A perfect transmission line (a coaxial cable) is characterised by an inductance and a capacitance per length. A possible discrete model is a series of discrete capacitive and inductive elements (without resistance), i.e. only non-dissipative elements.



We consider harmonic solutions $I_n(t) = \tilde{I}_n e^{-i\omega t}$. We recall that the impedance of the capacitance is $Z_C = 1/(-i\omega C)$ and that of the inductance $Z_L = -i\omega L$, where ω is the frequency.

1/ We first study the eigenmodes of the infinite line. Using Kirchhoff laws, write the equations satisfied by the currents $\tilde{I}_n$.

2/ Propagative modes.– Show that the modes $I_n(t) = e^{iqn - i\omega(q)t}$ only exist in a finite bandwidth $\omega \in [0, \omega_0]$ where $\omega_0 \stackrel{\text{def}}{=} 2/\sqrt{LC}$. Give the dispersion relation.

3/ Evanescent modes.– Study solutions of the form $I_n(t) = (-1)^n e^{qn - i\omega(q)t}$. Over what distance can propagate such modes ?

4/ Impedance of semi-infinite line.— We denote by Z_n the impedance of a finite line involving n couples of $L - C$ elements. Give the recurrence between Z_n and Z_{n+1} . Deduce the impedance of the semi-infinite line $Z_\infty \equiv Z(\omega)$. Plot $\text{Re } Z(\omega)$ and $\text{Im } Z(\omega)$. Discuss the fact that $\text{Re } Z(\omega) \neq 0$ for a certain interval of frequencies (comment this at the light of question 1).

Integration of the bath equations of motion.— We first derive the equations of motion

$$\begin{cases} m \ddot{x} = F(x) - x \sum_n \frac{c_n^2}{\omega_n^2} + \sum_n c_n q_n \\ \ddot{q}_n = -\omega_n^2 q_n + c_n x \end{cases} \quad (2.42)$$

where $F(x) = -V'(x)$. Let us integrate the equations of motion for the bath, which is possible thanks to the linearity. We can use that the retarded Green's function for the harmonic oscillator, i.e. the causal solution of $\ddot{G}^R(t) + \omega_n^2 G^R(t) = \delta(t)$ is $G^R(t) = \theta_H(t) \frac{\sin(\omega_n t)}{\omega_n}$. Thus we can solve the equation of motion for the oscillators

$$q_n(t) = q_n(0) \cos(\omega_n t) + \dot{q}_n(0) \frac{\sin(\omega_n t)}{\omega_n} + c_n \int_0^t dt' \frac{\sin(\omega_n(t-t'))}{\omega_n} x(t') \quad (2.43)$$

We split the source term in Eq. (2.42) into two parts

$$\sum_n c_n q_n(t) = \overbrace{\sum_n c_n \left(q_n(0) \cos(\omega_n t) + \dot{q}_n(0) \frac{\sin(\omega_n t)}{\omega_n} \right)}^{\stackrel{\text{def}}{=} \xi(t)} + \int_0^t dt' \Gamma(t-t') x(t') \quad (2.44)$$

where

$$\Gamma(t) \stackrel{\text{def}}{=} \sum_n c_n^2 \frac{\sin(\omega_n t)}{\omega_n} \quad (2.45)$$

is a function depending of the microscopic parameters of the model. We denote $\xi(t)$ “the noise”, which is controlled by the dynamical variables of the bath, i.e. a macroscopic number of degrees of freedom. For this reason, it is expected to exhibit a complex dynamics. Using $\int_0^\infty dt e^{i\omega t} = \frac{1}{0+ -i\omega}$ we remark that

$$\int_0^\infty dt \Gamma(t) = \sum_n \frac{c_n^2}{\omega_n^2} \quad (2.46)$$

which appears in the equation of motion above. With these definitions, we can rewrite the effective equation for the particle as

$$m \ddot{x}(t) = F(x(t)) - x(t) \int_0^\infty d\tau \Gamma(\tau) + \int_0^t d\tau \Gamma(\tau) x(t-\tau) + \xi(t) \quad (2.47)$$

Integration over the bath degrees of freedom is responsible for both the integral term and the “noise” term.

Noise and spectral function.— Because the bath involves a macroscopic number of degrees of freedom, it is natural to assume **thermal equilibrium for the bath**, say at $t = 0$, for the bath variables

$$P(\{q_n, p_n\}) \propto e^{-\beta H_{\text{env}}} \quad (2.48)$$

so that

$$\langle q_n(0)q_m(0) \rangle = \delta_{n,m} \frac{k_B T}{\omega_n^2} \quad (2.49)$$

$$\langle \dot{q}_n(0)\dot{q}_m(0) \rangle = \delta_{n,m} k_B T \quad (2.50)$$

Then the noise correlator is

$$C(t-t') = \langle \xi(t)\xi(t') \rangle = k_B T \sum_n \frac{c_n^2 \cos(\omega_n(t-t'))}{\omega_n^2} \quad (2.51)$$

At this stage it is useful to define the spectral function

$$J(\omega) \stackrel{\text{def}}{=} \pi \sum_n \frac{c_n^2}{2\omega_n} \delta(\omega - \omega_n) \quad (2.52)$$

which depends on the distribution of frequencies and coupling constants. We can write the function

$$\Gamma(t) = \frac{2}{\pi} \int_0^\infty d\omega J(\omega) \sin(\omega t) \quad (2.53)$$

and the correlator

$$C(t) = \frac{2k_B T}{\pi} \int_0^\infty d\omega \frac{J(\omega)}{\omega} \cos(\omega t) \quad (2.54)$$

in terms of the spectral function. In particular $C'(t) = -k_B T \Gamma(t)$. Two remarks :

- In practice, we expect a dense spectrum of oscillators for frequencies $\omega \geq 0$ (it is natural to assume that the spectrum of eigenmodes start at $\omega = 0$ since there exist low frequency excitations ususally).

What kind of behaviour can we expect for $J(\omega)$? Imagine that coupling constant is a smooth function of the frequency $c_n^2 = g(\omega_n)$. Then $J(\omega) = \frac{\pi}{2\omega} \sum_n g(\omega_n) \delta(\omega - \omega_n) \simeq \frac{\pi}{2\omega} g(\omega) \rho(\omega)$ for $\omega \rightarrow 0$, where $\rho(\omega)$ is the spectral density. For a linear spectrum (like photons, or phonons) we have $\rho(\omega) \sim \omega^{d-1}$ and thus we expect a power law $J(\omega) \sim g(\omega) \omega^{d-2}$ at low frequency. A simple assumption is $g(0) = \text{cste}$. In dimension $d = 3$, $g(0) = \text{cste}$ gives $J(\omega) \propto \omega$ for $\omega \rightarrow 0$.

- The spectrum of eigen-frequencies is usually cut off at a frequency ω_D related to the microscopic scale (for phonon modes in a crystal, ω_D is the Debye frequency, related to the lattice spacing).

The Ohmic case, $J(\omega) \propto \omega$ for small frequency : a concrete example.— Assuming a broad spectrum of frequencies, of width ω_D of the form

$$J(\omega) = \gamma_0 \omega \frac{\omega_D^2}{\omega^2 + \omega_D^2} \quad (2.55)$$

gives

$$C(t) = k_B T \gamma_0 \omega_D e^{-\omega_D |t|} . \quad (2.56)$$

Its integral is

$$\int_{-\infty}^{+\infty} dt C(t) = 2\gamma_0 k_B T \quad (2.57)$$

which recall us something...

Effective equation of motion.— Let us come back to the analysis of the effective equation of motion (2.47). If the spectral function $J(\omega)$ is broad (width $\sim \omega_D$), we expect the function $\Gamma(t)$ to be narrow in time (width $\sim 1/\omega_D$). For future convenience, we introduce

$$\gamma(t) = \int_t^\infty dt' \Gamma(t') \quad (2.58)$$

which also decays rapidly over the scale $\sim 1/\omega_D$. We introduce a Heaviside function in its definition to make it causal

$$\gamma(t) = \theta_H(t) \sum_n \frac{c_n^2 \cos(\omega_n t)}{\omega_n^2} = \frac{2\theta_H(t)}{\pi} \int_0^\infty d\omega \frac{J(\omega)}{\omega} \cos(\omega t) \quad (2.59)$$

is a “narrow function” of width $1/\omega_D$.

An integration by parts gives

$$\int_0^t d\tau \Gamma(\tau) x(t - \tau) = \gamma(0) x(t) - \gamma(t) x(0) - \int_0^t d\tau \gamma(\tau) \dot{x}(t - \tau) \quad (2.60)$$

Considering times $t \gg 1/\omega_D$, we drop the term $\gamma(t) x(0)$. We end with the effective equation of motion

$$\boxed{m \ddot{x}(t) = F(x(t)) - \int_0^t d\tau \gamma(\tau) \dot{x}(t - \tau) + \xi(t)} \quad (2.61)$$

After “integration over the bath degree of freedom”, we have obtained that the evolution the system is described by a integro-differential equation. This form makes clear the physical interpretation of the integral term as a *friction* term. The nonlocality of the friction term is not a surprise : from a microscopic perspective, friction results from the multiple interactions between the system and the bath, hence damping needs some time to establish.

FDT.— Finally we have the relation between the correlator of the noise and the friction

$$\boxed{C(\tau) = k_B T \gamma(\tau) \quad \text{for } \tau > 0} \quad (\text{FDT}) \quad (2.62)$$

which relates the correlator of the noise to the damping (friction) function.

In the microscopic model, the damping term is an integral term. The relation between the damping and the strength of the noise results from the integration of the microscopic equation of motions, assuming equilibrium for the bath *only* (not for the particle, like in the phenomenological Langevin approach). We can compare the two approaches

- In the § 2 we have introduced two terms in the Langevin equation : the friction controlled by γ_0 and the noise controlled by the strength C . We have then assumed that the *particle* is at canonical equilibrium $P_{\text{sys}}(x, p) \propto \exp[-\frac{\beta}{2}mv^2]$. Comparing with the statistical properties of the solution of the Langevin equation, we have deduced that the two parameters of the model cannot be independent but must be related by $C = 2\gamma_0 k_B T$. To some extent, this relation was *assumed* for consistency.
- Here, we have only assumed that, being macroscopic, the *bath* is at thermal equilibrium $P_{\text{bath}}(x, p) \propto e^{-\beta H_{\text{bath}}}$. As a result of the integration of the conservative dynamics, we have deduced the relation $C = 2\gamma_0 k_B T$. A by-product is that if we study the statistic for the particle, one can show that it is described by canonical equilibrium (the particle reaches equilibrium because it interacts with the bath).

Quantum model : a very similar analysis can be performed within a quantum frame. Mainly, the correlator of the noise (i.e. of the initial bath variables) involves a different function and one is led to a “quantum Langevin equation” (cf. [30] or [67]).

Energetic considerations : We now study the energy of the system

$$\frac{d}{dt}H_{\text{sys}} = \dot{x} [m\ddot{x} - F(x)] = -v(t) \int_0^t d\tau \gamma(\tau) v(t - \tau) + v(t) \xi(t) \quad (2.63)$$

Clearly, the second term corresponds to the work of the Langevin force

$$\frac{dW}{dt} = v(t) \xi(t) \quad (2.64)$$

hence the first term should be interpreted as the heat received by the system

$$\frac{dQ}{dt} = -v(t) \int_0^t d\tau \gamma(\tau) v(t - \tau). \quad (2.65)$$

One can consider the model with $\gamma(\tau) = \gamma_0 \omega_D e^{-\omega_D \tau}$. Assuming $1/\omega_D \ll \tau = m/\gamma_0$, we expect that $v(t)$ is smooth on the scale $1/\omega_D$ so that we can treat $\xi(t)$ as a white noise. Hence $v(t) \simeq \frac{1}{m} \int_0^t dt' \xi(t') e^{-(t-t')/\tau}$. We can estimate the average work of the Langevin force

$$\frac{\langle dW \rangle}{dt} = \langle v(t) \xi(t) \rangle = \frac{2k_B T}{\tau} \theta_H(0) = \frac{k_B T}{\tau} \quad (2.66)$$

where $\langle v(t) \xi(t) \rangle$ was computed in Chapter 2, cf. Eq. (2.15). We have used that $\theta_H(0) = 1/2$ (this is consistent with a symmetric regularised δ function). The particle receives work $\langle dW \rangle > 0$ on average. The averaged heat is

$$\frac{\langle dQ \rangle}{dt} = - \int_0^t dt' \gamma(t - t') \langle v(t) v(t') \rangle \quad (2.67)$$

Because the correlator $\langle v(t) v(t') \rangle$ decays much slower than the damping function (we have assumed $\omega_D \tau \gg 1$) we can write $\gamma(t - t') \langle v(t) v(t') \rangle \simeq \gamma(t - t') \langle v(t)^2 \rangle$, hence

$$\frac{\langle dQ \rangle}{dt} \simeq - \langle v(t)^2 \rangle \int_0^\infty dt'' \gamma(t'') = - \frac{k_B T}{m} \gamma_0 = - \frac{k_B T}{\tau} \quad (2.68)$$

i.e. the bath receives some heat from the particle. As it should the total energy is conserved on average

$$\langle dW \rangle + \langle dQ \rangle = 0 \quad (2.69)$$

On average, the Langevin force gives some work to the particle and the bath receives the heat which is dissipated. Fluctuations allow that the opposite occurs.

The bath receives the entropy $dS_{\text{bath}} = -dQ/T$, thus the dissipation corresponds to the entropy production with rate

$$\boxed{\frac{d \langle S_{\text{bath}} \rangle}{dt} = + \frac{k_B}{\tau}} \quad (2.70)$$

Stochastic thermodynamics.— Here, I have applied some concepts of thermodynamics to a single particle. This type of question has attracted a lot of attention for ~ 25 years and is the subject of the field of “stochastic thermodynamics”. If you are interested you can have a look to the reviews [43, 23, 50, 64, 63, 46] or to the lectures of Bernard Derrida at coll ge de France (2015-2016), <https://www.college-de-france.fr/site/bernard-derrida/>

Additional exercises

✎ **Exercise 2.7 – Comparison between time average and statistical average:** One considers the random “function” given by the sum of impulses $\xi(t) = \sum_{n=1}^N \kappa_n \delta(t - t_n)$ defined over the interval $[0, T]$, where

- the t_n ’s are independent and identically distributed (i.i.d) random times uniformly distributed over $[0, T]$ (i.e. one t_n has distribution $p(t_n) = 1/T$). We denote by $\lambda = N/T$ (for $N \rightarrow \infty$ and $T \rightarrow \infty$) the rate of occurrence of the random times.
- The κ_n ’s are i.i.d random variables with common distribution $w(\kappa)$ with finite $\langle \kappa_n^2 \rangle$.

a) Compute the time average of $\overline{\xi(t)}$, over the time interval $[0, T]$. Compare with the statistical average (over t_n ’s and κ_n ’s).

b) What is the condition on the random function $\xi(t)$ allowing to define a time averaged correlator $\tilde{C}(t - t') = \overline{\xi(t)\xi(t')^c} = \overline{\xi(t)\xi(t')} - \overline{\xi(t)} \overline{\xi(t')}$? Compare to $C(t - t') = \langle \xi(t)\xi(t') \rangle_c = \langle \xi(t)\xi(t') \rangle - \langle \xi(t) \rangle \langle \xi(t') \rangle$.

✎ **Exercise 2.8 – Mean square displacement from the Langevin equation:** Our aim is to compute the mean square displacement $\langle x(t)^2 \rangle$ of a Brownian particle in a fluid described by (2.3,2.4). We assume that $x(0) = 0$ and that the particle is initially at equilibrium with the fluid.

a) Recover the correlator $\langle v(t)\xi(t') \rangle$ and deduce $\langle x(t)\xi(t') \rangle$. What is $\langle x(t)\xi(t) \rangle$?

b) Prove that $\frac{d^2}{dt^2}x(t)^2 + \frac{1}{\tau} \frac{d}{dt}x(t)^2 = 2v(t)^2 + \frac{2}{m}x(t)\xi(t)$.

c) What is $\langle v(t)^2 \rangle$?

d) Argue that $\left. \frac{d}{dt} \langle x(t)^2 \rangle \right|_{t=0} = 0$ and deduce that

$$\langle x(t)^2 \rangle = 2D \left[t - \tau \left(1 - e^{-t/\tau} \right) \right] \quad (2.71)$$

Analyze carefully the limiting behaviours (interpret the $t \rightarrow 0$ behaviour) and plot the function.

✎ **Exercise 2.9 – Dissipation of the energy in the Langevin equation:** We consider the Langevin equation $m \frac{d}{dt}v(t) = -\gamma v(t) + \xi(t)$ for the velocity of a particle in a fluid where $\xi(t)$ is a Gaussian white noise such that $\langle \xi(t) \rangle = 0$ and $\langle \xi(t)\xi(t') \rangle = 2\gamma k_B T \delta(t - t')$.

a) Solve the Langevin equation. Deduce the correlator between the velocity and the Langevin force $\langle v(t)\xi(t') \rangle$.

b) Consider now the kinetic energy $E_{\text{kin}}(t) = \frac{1}{2}mv(t)^2$. Using the Langevin equation, deduce $\frac{dE_{\text{kin}}}{dt}$ and interpret the two terms.

c) Deduce a differential equation for $\langle E_{\text{kin}}(t) \rangle$.

Hint : you should need $\langle v(t)\xi(t) \rangle$ at coinciding times, which requires $\theta_H(0) = 1/2$ (Heaviside function at the origin).

d) Solve the differential equation for $\langle E_{\text{kin}}(t) \rangle$ and discuss the result.

✎ **Exercise 2.10 – From the Wiener process to the Ornstein-Uhlenbeck process:** We consider the Wiener process described by the equation $\frac{dW(u)}{du} = \eta(u)$, where $\eta(u)$ is a normalised Gaussian white noise.

a) Consider $\varphi(u)$ a monotonous function. Argue that

$$\eta(\varphi(u)) \stackrel{(law)}{=} \frac{1}{\sqrt{|\varphi'(u)|}} \eta(u) \quad (2.72)$$

b) Deduce the stochastic differential equation for

$$x(t) = \frac{W(u)}{\sqrt{u}} \quad \text{with} \quad u = u_0 e^{2\gamma t} \quad (2.73)$$

3 Markov processes and the master equation

We have discussed above few simple stochastic processes (the Wiener process and the Ornstein-Uhlenbeck process). Let us now introduce some general ideas and concepts (and the vocabulary) allowing for a general analysis of stochastic processes.



Figure 3.1: *Marian von Smoluchowski (1872-1917) is considered as the father of the theory of stochastic processes.*

3.1 Generalities : joint probabilities and conditional probabilities

The aim of the section is to introduce some useful tools and concepts needed to describe random processes. In general, the characterization of the distribution of a random process $X(t)$ defined on an interval $[0, T]$, i.e. a random function of time, should require to introduce a probability weight taking the form of a *functional* $P[x(t)] \sim \text{Proba}\{X(t) = x(t), \forall t \in [0, T]\}$, which is rather tricky to handle. However if we are interested in an average value, formally written as a functional integral $\langle X(t) \rangle = \int \mathcal{D}x P[x(t)] x(t)$, we do not need all the information about the process at all times, but only at time t . In this chapter and the following two chapters, we will restrict ourselves to simple questions which only require to know the probability that the process $X(t)$ takes some values at a *discrete* set of times (at t for $\langle X(t) \rangle$, or at t_1 and t_2 for $\langle X(t_1)X(t_2) \rangle$, etc). This leads to introduce the joint distribution of the values taken by the process at those times only. At the end of the lectures (chapter 6), we will address more complicated questions requiring the knowledge of the process in the full time interval.

Joint probability : In order to characterize the statistical properties of the random process, we introduce the joint probability or “the n -point function”

$$P_n(\underbrace{x_n, t_n; \dots; x_2, t_2; x_1, t_1}_{\substack{\longleftarrow \\ \text{time}}}) = \langle \delta(x_n - X(t_n)) \dots \delta(x_1 - X(t_1)) \rangle \quad (3.1)$$

corresponding to the probability (density) for the process to be equal to $x_1, \dots, x_n$ at times $t_1, \dots, t_n$. We can also write

$$\begin{aligned} & P_n(x_n, t_n; \dots; x_2, t_2; x_1, t_1) dx_1 \dots dx_n \\ &= \text{Proba}\{X(t_1) \in [x_1, x_1 + dx_1] \& \dots \& X(t_n) \in [x_n, x_n + dx_n]\} \end{aligned} \quad (3.2)$$

From the definition, it is clear that one integration relates the n -point to the $n-1$ -point function

$$\int dx_k P_n(x_n, t_n; \dots; x_{k+1}, t_{k+1}; \underline{x_k, t_k}; x_{k-1}, t_{k-1}; \dots; x_1, t_1) \quad (3.3)$$

$$= P_{n-1}(x_n, t_n; \dots; x_{k+1}, t_{k+1}; x_{k-1}, t_{k-1}; \dots; x_1, t_1) \quad (3.4)$$

The main interest of the joint distribution with n arguments is to allow us to express the average and correlations of the process. Eq. (3.1) makes clear that

$$\langle X(t_1) \rangle = \int dx_1 P_1(x_1, t_1) x_1 \quad (3.5)$$

$$\langle X(t_2)X(t_1) \rangle = \int dx_1 dx_2 P_2(x_2, t_2; x_1, t_1) x_2 x_1 \quad (3.6)$$

etc.

Conditional probability : Another important concept is the one of conditional probability corresponding to the probability for the process to pass through $x_1, \dots, x_n$ at successive times $t_1, \dots, t_n$, *given* that it has previously passed through $y_1, \dots, y_n$ at successive times $\tau_1, \dots, \tau_n$:

$$P_{n|m}(x_n, t_n; \dots; x_1, t_1 | y_m, \tau_m; \dots; y_1, \tau_1) = \frac{P_{n+m}(x_n, t_n; \dots; x_1, t_1; y_m, \tau_m; \dots; y_1, \tau_1)}{P_m(y_m, \tau_m; \dots; y_1, \tau_1)} \quad (3.7)$$

In particular, the conditional probability is useful if we are interested in conditional averages. Examples :

$$\langle X(t_1) | X(0) = x_0 \rangle = \int dx_1 x_1 P_{1|1}(x_1, t_1 | x_0, 0) \quad (3.8)$$

$$\langle X(t_2)X(t_1) | X(0) = x_0 \rangle = \int dx_1 dx_2 x_2 x_1 P_{2|1}(x_2, t_2; x_1, t_1 | x_0, 0), \quad (3.9)$$

etc, which involve averaging over trajectories which all start from the same point $X(0) = x_0$.

Propagator : Integrating the two point-point function $P_2(x, t; x_0, t_0) = P_{1|1}(x, t | x_0, t_0) P_1(x_0, t_0)$ over the initial coordinate, we get

$$P_1(x, t) = \int dx_0 P_{1|1}(x, t | x_0, t_0) P_1(x_0, t_0). \quad (3.10)$$

The equation shows that the conditional probability plays the role of a “*time evolution operator*” for the distribution $P_1(x, t)$: it “propagates” the stochastic evolution. For this reason, it is also called the “*propagator*”.

3.2 Markov processes

A very important class of random processes are *Markov processes*. A Markov process is a random process whose evolution after time t_0 only depends on its value at time t_0 , and *not* on its past history before t_0 . Mathematically, this property takes the form

$$P_{n|m}(x_n, t_n; \dots; x_1, t_1 | y_m, \tau_m; \underbrace{\dots; y_1, \tau_1}_{\text{past history}}) = P_{n|1}(x_n, t_n; \dots; x_1, t_1 | y_m, \tau_m) \quad (3.11)$$

which expresses that history prior to τ_m does not matter : only the *last* position y_m at time τ_m determines the evolution at times $t_1, \dots, t_n$.

Let us examine the consequences of this fundamental assumption. Consider for example the three point function :

$$P_3(x_3, t_3; x_2, t_2; x_1, t_1) = P_{1|2}(x_3, t_3 | x_2, t_2; x_1, t_1) P_2(x_2, t_2; x_1, t_1) \quad (3.12)$$

$$= P_{1|2}(x_3, t_3 | x_2, t_2; x_1, t_1) P_{1|1}(x_2, t_2 | x_1, t_1) P_1(x_1, t_1) \quad (3.13)$$

$$\stackrel{\text{Markov}}{=} P_{1|1}(x_3, t_3 | x_2, t_2) P_{1|1}(x_2, t_2 | x_1, t_1) P_1(x_1, t_1) \quad (3.14)$$

We can generalize this to any joint distribution. We simplify the notation as $P_{1|1}(x, t | y, t_0) \equiv P(x, t | y, t_0)$ and $P_1(x, t) \equiv P(x, t)$ and we conclude that



Figure 3.2: *Andreï Andreïevich Markov (1856-1922).*

A Markov process is fully characterized by $P(x, t | x_0, t_0)$ and $P(x, t)$ only.

In practice, the following property is useful

$$P_{n|1}(x_n, t_n; \dots; x_1, t_1 | x_0, t_0) = \prod_{k=1}^n P(x_k, t_k | x_{k-1}, t_{k-1}) \quad (3.15)$$

as it allows to compute any average $\langle X(t_n) \cdots X(t_1) X(t_0) \rangle$ or $\langle X(t_n) \cdots X(t_1) | X(t_0) = x_0 \rangle$.

Chapman-Kolmogorov equation : Start from the general property

$$\int dx_2 P_3(x_3, t_3; x_2, t_2; x_1, t_1) = P_2(x_3, t_3; x_1, t_1) . \quad (3.16)$$

For a Markov process, using (3.14), one gets the Chapman-Kolmogorov equation

$$\int dx_2 P(x_3, t_3 | x_2, t_2) P(x_2, t_2 | x_1, t_1) = P(x_3, t_3 | x_1, t_1) \quad (3.17)$$

The probability to go from x_1 to x_3 is the sum over x_2 of the probabilities conditioned to passed through x_2 .



Figure 3.3: *Andreï Nikolaïevitch Kolmogorov (1903-1987), well-known by physicists for his major contributions to the theory of dynamical systems and probability.*

Homogeneous Markov processes : In the following, in order to simplify the discussion, we will restrict ourselves to Markov processes such that the transition probability is invariant under time translation

$$P(x_2, t_2 | x_1, t_1) = P(x_2, t_2 - t_1 | x_1, 0) \quad (3.18)$$

Such random processes are denoted “*homogeneous*”. Since it only depends on the time difference, we will also write the propagator as $P_t(x|x_0)$.

a) A first example of Markov process

We can come back to the Langevin equation (2.3) for the velocity. (i) The equation for $v(t)$ is first order, hence the evolution at $t > 0$ is fully determined by $v(0) = v_0$. (ii) The Langevin force is a white noise (ξ uncorrelated in time), thus no time correlations are hidden in the noise. From (i) & (ii) we conclude that the process is Markovian.

Let us now discuss the related fundamental probabilities $P(v, t)$ and $P(v, t|v_0, 0)$, which can be easily obtained if we assume furthermore that the Langevin noise is **Gaussian**.¹⁶ We have obtained above the two first moments, cf. (2.8, 2.10), which are conditioned averages now denoted

$$\langle v(t) | v(0) = v_0 \rangle = v_0 e^{-t/\tau} \quad (3.19)$$

$$\text{Var}[v(t) | v(0) = v_0] = \frac{k_B T}{m} \left(1 - e^{-2t/\tau} \right) \quad (3.20)$$

The process $v(t)$ is a convolution of the Gaussian Langevin force $\xi(t)$, hence it is also Gaussian (a sum of Gaussian variables is also Gaussian). The knowledge of these two moments is sufficient to characterize the full distribution :¹⁷

$$P_t(v|v_0) = \sqrt{\frac{m}{2\pi k_B T (1 - e^{-2t/\tau})}} \exp \left\{ -\frac{m (v - v_0 e^{-t/\tau})^2}{2k_B T (1 - e^{-2t/\tau})} \right\}. \quad (3.21)$$

At large time, the conditional probability converges toward the equilibrium distribution

$$P_t(v|v_0) \xrightarrow{t \rightarrow \infty} P(v) = \sqrt{\frac{m}{2\pi k_B T}} e^{-\frac{m}{2k_B T} v^2} \quad (3.22)$$

the Maxwell distribution (or Gibbs distribution). Here, the “one point distribution” $P(v)$ is independent of the time due to the existence of a stationary state (this is not always the case). This is also why the conditional probability rapidly converges (exponentially fast) toward the equilibrium distribution. The process described by equation (2.3), or the conditional probability (3.21), is known as the **Ornstein-Uhlenbeck process**. It is the subject of a theorem by Doob (1942) : the only homogeneous Gaussian stationary random process is the Ornstein-Uhlenbeck process.¹⁸

Exercise 3.1 – : Express the correlator $\langle v(t)v(t') | v(0) = v_0 \rangle$ as a double integral involving the conditional probability $P_t(v|v_0)$. Check that one recovers Eq. (2.10) by using (3.21). Same question for the correlator $\langle v(t)v(t') \rangle_c$ corresponding to an initially random velocity v_0 .

b) Markovian or non-Markovian ? Few examples

Random walk.— After each step, a random walker forgets his past history (where he is coming from) : this is a Markov process.

Choosing clothes every morning.— Consider choosing clothes from a closet every morning without looking at the weather forecast, as a stochastic process. The selection depends on the weather of the previous day : there is a memory effect hence the process is non Markovian.

¹⁶The distribution of the noise is a Gaussian : $P[\xi] \propto \exp \left\{ -\frac{1}{2C} \int dt \xi(t)^2 \right\}$. One can deduce from this that $\langle \xi(t)\xi(t') \rangle = C \delta(t - t')$ [hint : discretize the time to check this].

¹⁷This argument was used in the historical paper : G. E. Uhlenbeck and L. S. Ornstein, “On the Theory of the Brownian Motion”, Phys. Rev. **36**(5), 823–841 (1930).

¹⁸The Ornstein-Uhlenbeck process is more frequently introduced as a model for a particle attached to a spring in the overdamped regime : $\dot{x}(t) = -k x(t) + \xi(t)$ (I forget the friction coefficient γ).

Velocity $v(t)$: a Markov process.— We have pointed out that the Langevin equation (2.3) describes a Markov process because

- (i) the differential equation is first order, hence the solution for $t > 0$ depends only on some initial value $v(0)$;
- (ii) the noise is δ -correlated, hence there is no memory hidden in the Langevin force.

Position $x(t) = \int_0^t dt' v(t')$: a non Markovian process.— The position $x(t)$ of the particle is a non Markovian process. There are two points of view leading to this conclusion.

(1) it obeys the stochastic differential equation (SDE) $m \ddot{x}(t) = -\gamma \dot{x}(t) + \xi(t)$, which is *second* order. The noise is still δ -correlated, however, the solution of the differential equation is determined both by the initial position $x(0)$ *and* the initial velocity $\dot{x}(0)$, which depends on the history before $t = 0$ (velocity indicates where the particle comes from). We break point (i) while point (ii) holds. The process $x(t)$ is non Markovian.

(2) We could also argue that $x(t)$ obeys a first order SDE, $\dot{x}(t) = v(t)$, whose solution is determined only by $x(0)$, hence point (i) holds. The velocity now plays the role of the noise. This “noise” is now characterized by a finite correlation time (memory time), $\langle v(t)v(t') \rangle = \frac{k_B T}{m} e^{-|t-t'|/\tau}$ (one says that the SDE for $x(t)$ involves a “colored noise”). Within this point of view we break point (ii), which leads to the same conclusion, as it should, that $x(t)$ is non Markovian.

A 2D Markov process : $(x(t), v(t))$.— $x(t)$ alone is a non Markovian process, however it can be considered as the first component of a two-dimensional Markov process $\vec{X}(t) = (x(t), v(t))$. Indeed, the system of differential equations for $(x(t), v(t))$ is *first* order and involves a δ -correlated noise. It presents the more general form of a *multidimensional Langevin equation*

$$\dot{X}_i = A_i(\vec{X}) + B_{ij} \Xi_j(t) \quad (3.23)$$

where here $\vec{A} = (v, F(x) - \gamma v)$ is the drift (we have added a conservative force). The noise $\vec{\Xi}(t) = (\xi_x(t), \xi(t))$ is uncorrelated in time. The matrix is $B_{xx} = B_{xv} = B_{vx} = 0$ and $B_{vv} = 1$. Hence it is a 2D Markovian process. We will see that the joint distribution $P_t(x, v)$ obeys the “Kramers equation” $\partial_t P_t = \left[-\partial_x v - \frac{1}{m} \partial_v (F(x) - \gamma v) + \frac{1}{m^2} \gamma k_B T \partial_v^2 \right] P_t$.

Conclusion : The analysis of this example shows that the identification of a Markov process is sometimes a question of perspective, and also illustrates that Markov processes are elementary building blocks. Many non Markovian processes can be constructed from Markov processes.

3.3 Master equation

a) Master equation for continuous processes

Let us start with the case of continuous processes, which is more general. As we have seen, the time evolution of the distribution $P(x, t)$ of a Markov process can be represented in terms of the conditional probability, Eq. (3.10). However, this equation and (3.17) are not of great help to determine the two fundamental functions $P(x, t)$ and $P(x, t|x_0, t_0)$. The distribution is more conveniently obtained by solving an evolution equation for an infinitesimal time step : such an evolution equation can be related to the above integral equation by considering the evolution during an infinitesimal time $\delta t \rightarrow 0$. In this case we expect

$$P(x, t + \delta t|x_0, t) \simeq \delta(x - x_0) + \delta t W_t(x|x_0) + \mathcal{O}(\delta t^2) \quad (3.24)$$

The linear correction follows from the Markov assumption : at short time, the transition probability is linear with time and involves a transition rate $W_t(x|x_0)dx$ for performing the transition from x_0 to $[x, x + dx]$, which is the elementary information defining the process.

If we restrict ourselves to homogeneous processes (time translation invariant) we have

$$W_t(x|x_0) \rightarrow W(x|x_0) \quad (\text{homogeneous process}) \quad (3.25)$$

independent of time. For homogeneous processes, we find the differential equation (in time)

$$\boxed{\frac{\partial P(x, t)}{\partial t} = \int dx' W(x|x') P(x', t)} \quad (3.26)$$

This is the general form of the Master equation for a continuous process. The fact that the differential equation in time is *first order* means that the future evolution of the probability density for times $t > 0$ depends only on the distribution $P(x, 0)$ and not on the past history (Markov property). This is the (*a posteriori*) justification for the expansion (3.24). Note that the conservation of probability requires that

$$\int dx_f W(x_f|x_i) = 0 \quad \Rightarrow \quad \frac{\partial}{\partial t} \int dx P(x, t) = 0 \quad \forall t \quad (3.27)$$

so that probability $\int dx P(x, t) = 1$ is conserved. This condition follows from the normalization condition of the conditional probability, in the expansion (3.24).

Note that, being related to a transition rate, the transition kernel should be positive $W(x|x_0) \geq 0$ for $x \neq x_0$. As a result, it must contain a singular negative contribution of the form $-\lambda(x) \delta(x - x_0)$ in order to fulfill the condition (3.27), where $\lambda(x)$ is therefore the probability rate for leaving point x (see examples below in this chapter).

Obviously, the conditional probability obeys the same equation

$$\frac{\partial P_t(x|x_0)}{\partial t} = \int dx' W(x|x') P_t(x'|x_0) \quad \text{for initial condition } P_0(x|x_0) = \delta(x - x_0). \quad (3.28)$$

In a specific problem, the transition “rates” $W(x|x_0)$ encode the elementary information about the stochastic process. The aim is to solve the master equation (3.26), or (3.28), given the integral kernel $W(x|x_0)$.

In the most general case, a Markov process can combine

- a diffusion : in this case the integral kernel $W(x|x')$ is replaced by a second order differential operator
- jumps : leading to an integral term in the master equation, like in (3.26).

Below, we will give concrete examples. We will come back on this discussion in chapter 5 (§ 5.2).

b) Master equation for discrete processes

For simplicity, let us first consider a random process which takes discrete values $X(t) \in \{x_1, \dots, x_N\}$ and denote $P_n(t) = \text{Proba}\{X(t) = x_n\}$ (in certain situations the total number of states can be infinite, $N \rightarrow \infty$). The Markovian nature of the process implies that what happens on $[t, t + \delta t]$ is independent to what has occurred on $[0, t]$. Hence the probability $P_n(t + \delta t)$ can be expressed in terms of products of probabilities, one characterizing the evolution up to time t , i.e. the probabilities $P_m(t)$ ’s, and the probabilities for transitions between states during time interval

δt . We denote by $W_{n,m} \geq 0$ the rate for a transition from m to n , with $n \neq m$, meaning that the probability to make a transition $m \rightarrow n$ in the time interval $[t, t + \delta t]$ is $W_{n,m} \delta t$. Therefore

$$P_n(t + \delta t) \simeq \left(1 - \underbrace{\sum_{m(\neq n)} W_{m,n} \delta t}_{\text{proba to leave } n}\right) \overset{\text{Markov}}{\times} P_n(t) + \sum_{m(\neq n)} \underbrace{W_{n,m} \delta t}_{\text{proba for } m \rightarrow n} \overset{\text{Markov}}{\times} P_m(t) \quad (3.29)$$

Letting $\delta t \rightarrow 0$ we obtain the *first order* differential equation

$$\frac{d}{dt} P_n(t) = \sum_{m(\neq n)} [W_{n,m} P_m(t) - W_{m,n} P_n(t)] \quad (3.30)$$

(we can also replace $\sum_{m(\neq n)} \rightarrow \sum_m$). Eq. (3.30) is the *master equation* for a discrete process.

The previous equation does not depend on $W_{n,n}$, which was not defined so far, however it is convenient to introduce $W_{n,n} \stackrel{\text{def}}{=} -\sum_{m(\neq n)} W_{m,n} < 0$, which is the probability rate for leaving state n , up to the sign. We deduce the form

$$\boxed{\frac{dP_n(t)}{dt} = \sum_m W_{n,m} P_m(t)} \quad (3.31)$$

The coefficients $W_{n,m}$ form a $N \times N$ matrix W , which satisfies the condition

$$\boxed{\sum_n W_{n,m} = 0} \quad (3.32)$$

ensuring the conservation of probability $\sum_n P_n(t) = 1 \forall t$. Note that this condition is related to an asymmetry of the matrix ($W_{n,m} \neq W_{m,n}$ in general) : the sum of all elements of a column is zero, however summing elements along a line is non zero in general.

We now discuss several examples of Markov processes (discrete or continuous).

c) The random telegraph process ($N = 2$)

Let us first consider the *most simple Markov process* $X(t)$, taking only two possible values X_1 or X_2 (this is a “two level system” for stochastic processes). A typical realization of the process is represented on Fig. 3.4.

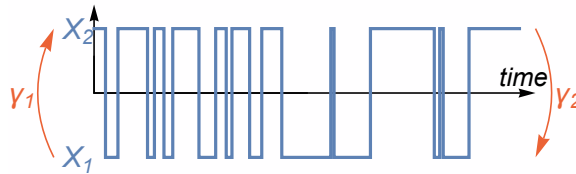


Figure 3.4: *Random telegraph process with $X_1 = -1$ and $X_2 = +1$.*

There are many different realizations of the random telegraph process (according to [wikipedia](#), it is also known as the Kac process, from the name of the famous Mathematician Mark Kac, or random telegraph noise, or burst noise, or popcorn noise). It may occur any time we deal with a two state system with a random switching mechanism. For example, since the 1990s, there has been a huge interest in micrometric structures known as “*quantum dots*” (QD), which can trap a tunable number of eletrons (see for example the review [74]). Such a device is shown in

Fig. 3.5 (see also Fig. 3.6 p. 60) : a small island is patterned in a two dimensional electron gas at an interface GaAs/GaAlAs. Due to Coulomb interaction, the addition of an extra electron in the island has a high energy cost, however, playing with some external gate voltages, the QD can be prepared in a configuration such that the two states, the one with n electrons and the one with $n + 1$ electrons, are degenerate. Then, the QD is switching between the two states. Moreover, the figure 3.5 shows that the switching occurs with a rather low frequency, on a millisecond scale, which is a macroscopic time for the electrons.

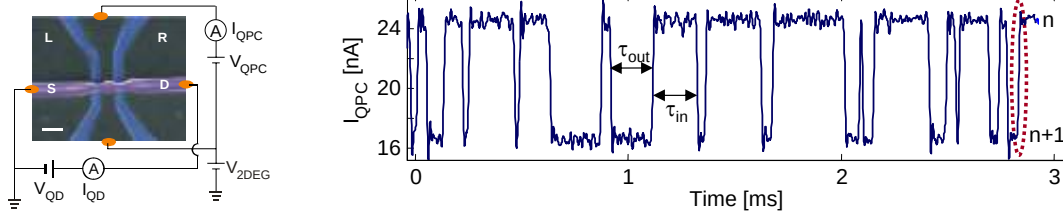


Figure 3.5: A physical realization of the random telegraph process : the charge inside a quantum dot is measured as a function of the time with a neighbouring probe. The current I_{QPC} through the probe is proportional to the number of electrons inside the central island, which fluctuates by one unit (one electron). From : S. Gustavsson, I. Shorubalko, R. Leturcq, S. Schön, and K. Ensslin, “Measuring current by counting electrons in a nanowire quantum dot”, Appl. Phys. Lett. **92**, 152101 (2008)

We introduce two probabilities $P_i(t) = \text{Proba}\{X(t) = X_i\}$ with $i \in \{1, 2\}$ and in general, there are two different transition rates $\gamma_1 = \text{rate}(X_1 \rightarrow X_2)$ and $\gamma_2 = \text{rate}(X_2 \rightarrow X_1)$. The matrix W is 2×2

$$W = \begin{pmatrix} -\gamma_1 & \gamma_2 \\ \gamma_1 & -\gamma_2 \end{pmatrix}. \quad (3.33)$$

The two probabilities obey the set of coupled differential equations

$$\frac{dP_1(t)}{dt} = -\gamma_1 P_1(t) + \gamma_2 P_2(t) \quad (3.34)$$

$$\frac{dP_2(t)}{dt} = +\gamma_1 P_1(t) - \gamma_2 P_2(t) \quad (3.35)$$

in each equation, one term populates the state and the other empties the state.

In the following exercise, we solve this system of equations.

✎ Exercise 3.2 – Random telegraph process :

a) Derive the set of differential equations for $P_1(t)$ and $P_2(t)$ from the definitions of the rates. Deduce a matricial form $\frac{d}{dt}P(t) = W P(t)$, where $P = (P_1 \ P_2)^T$ is the column vector (T denotes transposition).

b) Find the stationary solution, denoted by P_i^* , and give the general solution of the master equation.

c) Determine the conditional probability $P_t(i|j)$. Compute $\sum_j P_t(n|j) P_j^*$ and interpret. Discuss detailed balance.

d) Express $\langle X(t) \rangle$ and $\langle X(t)X(t') \rangle$ in the stationary regime. For simplicity, choose $X_1 = 0$ and $X_2 = 1$. Show that the correlator $C(t - t') = \langle X(t)X(t') \rangle - \langle X(t) \rangle \langle X(t') \rangle$ is

$$C(t - t') = \frac{\gamma_1 \gamma_2}{(\gamma_1 + \gamma_2)^2} e^{-(\gamma_1 + \gamma_2)|t - t'|}. \quad (3.36)$$

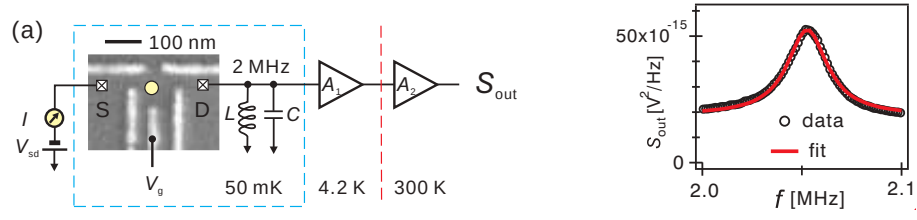


Figure 3.6: Left : *The microstructure studied*. Right : *Power spectrum*. From : Y. Okazaki, S. Sasaki and K. Muraki, *Shot noise spectroscopy on a semiconductor quantum dot in the elastic and inelastic cotunneling regimes*, Phys. Rev. B **87**, 041302(R) (2013).

e) **Electric noise power spectrum** : in the experiments, the fluctuations of the charge in the QD can be characterized by measuring the power spectrum $S(\omega)$ (see the Wiener-Khintchin theorem discussed below, § 3.8, and the relation with the correlation function $C(t)$). Deduce $S(\omega)$ for the random telegraph process. The measurement is reported in another article, Fig. 3.6. Compare your result with the data.

d) Birth and death processes

A subclass of discrete processes are “birth and death processes”. They correspond to the case where the transition matrix is tridiagonal, i.e. allows only transitions between nearest neighbour states. The master equation has the form

$$\frac{dP_n(t)}{dt} = d_{n+1} P_{n+1}(t) + b_{n-1} P_{n-1}(t) - (d_n + b_n) P_n(t) \quad (3.37)$$

where $d_n > 0$ and $b_n > 0$ are death and birth rates, respectively (cf. Fig. 3.7). A simple example is the Poisson process studied below in great detail (in this case the death rates are zero, $d_n = 0$ and birth rates are all equal, $b_n = \lambda \forall n$). Birth and death processes have some applications in queueing theory (Exercise 3.18), in biology, etc. Another example will be studied in Exercise 3.19.

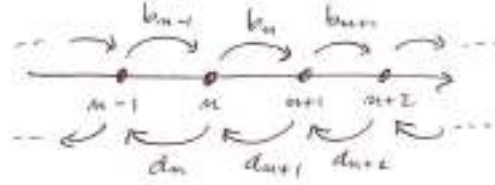


Figure 3.7: *Transition rates in the birth and death process*.

e) The Poisson process (statistics of uncorrelated events)

The Poisson process takes integer values $\mathcal{N}(t) \in \mathbb{N}$ and starts at initial time from $\mathcal{N}(0) = 0$. With probability rate λ , the process is incremented by one, i.e. during an interval of time of duration dt , the process increases by one with probability λdt . We denote $P_n(t) = \text{Proba}\{\mathcal{N}(t) = n\}$ its probability.

The Poisson process (PP) counts the occurrences of *independent events*. For instance the number of drops of rain falling on the floor during a time interval t . Or the number of disintegrations in a radioactive material during a time t .

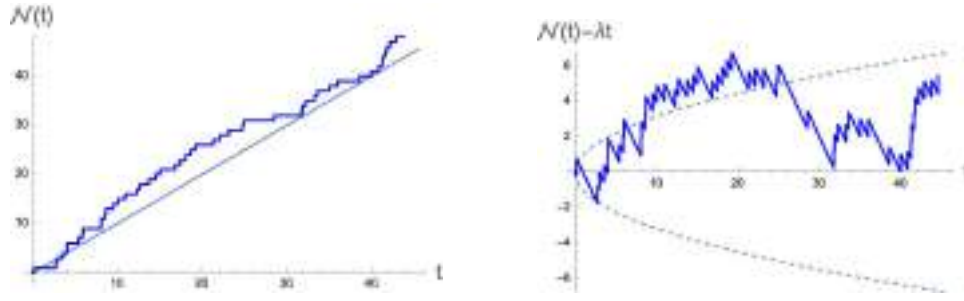


Figure 3.8: A typical realization of the Poisson process : $\mathcal{N}(t)$ as a function of t . On the right plot, the average is removed. Typical fluctuations grow as $\sim \sqrt{t}$.

The master equation of the PP is

$$\begin{cases} \frac{dP_n(t)}{dt} = \lambda P_{n-1}(t) - \lambda P_n(t) & \text{for } n > 0 \\ \frac{dP_0(t)}{dt} = -\lambda P_0(t) \end{cases} \quad (3.38)$$

In other terms, the rate “matrix” has elements on the diagonal and just below the diagonal $W_{n,m} = \lambda (-\delta_{m,n} + \delta_{m,n-1})$.

Exercise 3.3 – Master equation for the PP:

- Show that the master equation for the Poisson process is (3.38).
- Introduce the generating function $G(z; t) \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} z^n P_n(t)$. What is the value of $G(z; 0)$? Get a differential equation for $G(z; t)$ and solve it.
- Deduce that

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad (3.39)$$

- Determine the cumulants $\langle \mathcal{N}(t)^k \rangle_c$ of the Poisson process.
- Give the distribution $q(\tau)$ of the time separating two successive events [indication : relate $q(\tau)$ and $P_0(t)$].

Note that (3.39) was obtained by choosing the initial condition $\mathcal{N}(0) = 0$, hence, (3.39) corresponds to the conditional probability $P_t(n|0)$. Using translation invariance of the Poisson process, we deduce the general expression of the conditional probability

$$P_t(n|m) = \begin{cases} \frac{(\lambda t)^{n-m}}{(n-m)!} e^{-\lambda t} & \text{for } n \geq m \\ 0 & \text{for } n < m \end{cases} \quad (3.40)$$

The Poisson process is both *translation invariant* in time ($W_{n,m}$ does not depend on time) and “space”

$$W_{n,m} = \lambda (-\delta_{m-n,0} + \delta_{n-m,1}) = \text{function}(n-m). \quad (3.41)$$

Exercise 3.4 – Conditional probability for the PP: Consider the Poisson process with $\mathcal{N}(0) = 0$. Check that $\sum_m P_t(n|m) P_{t'}(m|n_0) = P_{t+t'}(n|n_0)$.

Exercise 3.5 – Two-point correlator of the PP:

- For $\mathcal{N}(0) = 0$, express $\langle \mathcal{N}(t) \rangle$ in terms of the probability. Compute it.
- Express $\langle \mathcal{N}(t)\mathcal{N}(t') \rangle$ as a double sum and compute it when $t' < t$. Deduce the correlator $\langle \mathcal{N}(t)\mathcal{N}(t') \rangle_c = \langle \mathcal{N}(t)\mathcal{N}(t') \rangle - \langle \mathcal{N}(t) \rangle \langle \mathcal{N}(t') \rangle$. Compare with the Wiener process.
- Deduce $\left\langle [\mathcal{N}(t) - \mathcal{N}(t')]^2 \right\rangle_c$. Comment on the limit $t \rightarrow t'$.

f) The compound Poisson process (or “Poisson jump process”)

A natural generalization of the Poisson process is the *compound Poisson process* (CPP) ¹⁹ : we consider now that the process $X(t)$ makes random jumps

$$X(t_n^+) = X(t_n^-) + \eta_n \quad (3.42)$$

where the η_n 's are i.i.d., distributed according to a distribution $w(\eta)$. As for the Poisson process, the jumps occur at random times t_n with rate λ (for the PP, we have $\mathcal{N}(t_n^+) = \mathcal{N}(t_n^-) + 1$). After a time t , the number of jumps $\mathcal{N}(t)$ is random (it is a PP), hence we can represent the CPP in terms of a PP as

$$X(t) = \sum_{n=0}^{\mathcal{N}(t)} \eta_n \quad \text{with } X(0) = \eta_0 = 0. \quad (3.43)$$

Numerics : this form shows that it is very simple to simulate the CPP with a computer. One can generate two independent sequences of random numbers : the random time intervals $\tau_n = t_n - t_{n-1} > 0$ with exponential distribution $q(\tau) = \lambda e^{-\lambda\tau}$ (cf. Exercise 3.3) and the random amplitudes η_n 's with distribution $w(\eta)$.

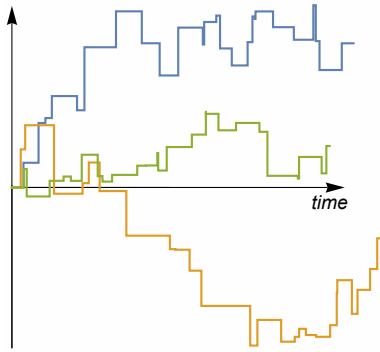


Figure 3.9: Three realizations of the Compound Poisson Process $X(t)$ for Gaussian jumps.

Master equation : The master equation of the CPP is

$$\frac{\partial P(x, t)}{\partial t} = \lambda \int d\eta w(\eta) [P(x - \eta, t) - P(x, t)] \quad (3.44)$$

i.e. of the form (3.26) for

$$W(x|x_0) = \lambda [w(x - x_0) - \delta(x - x_0)] . \quad (3.45)$$

Note that here, the transition kernel is (time and space) translation invariant, as for the PP.

✎ Exercise 3.6 – Master equation for the CPP :

a) Show that the master equation for the CPP is (3.44).

b) *Continuum limit.*— Study the limit $\lambda \rightarrow \infty$ with $\eta_n \rightarrow 0$ such that $a = \lambda \langle \eta_n \rangle$ and $b = \lambda \langle \eta_n^2 \rangle$ are kept finite (argue that, in this limit, $\lambda \langle \eta_n^k \rangle \rightarrow 0$ for $k > 2$).

c) Introducing the Fourier transforms $\hat{P}(k, t) = \int dx e^{-ikx} P(x; t)$ and $\hat{w}(k) = \int d\eta e^{-ik\eta} w(\eta)$, show that the solution is

$$P(x, t) = \int_{-\infty}^{+\infty} \frac{dk}{2\pi} e^{\lambda t [\hat{w}(k) - 1] + ikx} . \quad (3.46)$$

¹⁹in French: “processus de Poisson composé”.

Discuss the continuum limit.

d) When $\langle \eta_n^2 \rangle = \infty$, the process belongs to the class of **Lévy flights**. For example, if $w(\eta) \simeq c/\eta^2$ for $\eta \rightarrow \pm\infty$ we have $\hat{w}(k) \simeq 1 - \tilde{c}|k|$ for $k \rightarrow 0$. Deduce $P(x, t)$ over large scales. Discuss also the more general case where $\hat{w}(k) \simeq 1 - \tilde{c}|k|^\mu$ for $k \rightarrow 0$, with $\mu \in]0, 2[$.

3.4 Classification of Markov processes

We now discuss the different scenarii which might occur when studying the master equation (3.31).

Notations for vectors and matrices : In order to use unified notations, “vectors” in $\mathbb{R}^d$ will be column vectors ($d \times 1$ matrices) :

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \quad \text{and} \quad X^T = (x_1 \quad \cdots \quad x_d) \quad (3.47)$$

is a line vector (a $1 \times d$ matrix). We denote by $(\cdot)^T$ the transposition of a matrix. As a result, if X and Y are two (column) vectors, the scalar product is $X^T Y$, while $X Y^T$ represents a matrix (a projector). Explicitly (for $d = 2$) :

$$(x_1 \quad x_2) \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = x_1 y_1 + x_2 y_2 \quad \text{and} \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} (y_1 \quad y_2) = \begin{pmatrix} x_1 y_1 & x_1 y_2 \\ x_2 y_1 & x_2 y_2 \end{pmatrix}. \quad (3.48)$$

a) Stationary state

The equation (3.32), which guarantees the conservation of the total probability, has a clear linear algebra interpretation : it expresses that the (column) vector with unit components

$$L^{(0)} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \quad (3.49)$$

is a *left* eigenvector with eigenvalue $\lambda_0 = 0$:

$$L^{(0)T} W = 0 \quad \text{or} \quad W^T L^{(0)} = 0. \quad (3.50)$$

Because the $N \times N$ matrix W is non-symmetric in general, the eigenvalue is associated with a pair of left and right eigenvectors which differ. Let us denote $R^{(0)}$ the corresponding right eigenvector, i.e. $W R^{(0)} = 0$, with components $R^{(0)} = (P_1^*, \dots, P_N^*)^T$. Using (3.32), it is convenient for future discussion to rewrite the equation under the form

$$\sum_{m (\neq n)} [W_{nm} P_m^* - W_{mn} P_n^*] = 0 \quad \forall n. \quad (3.51)$$

For matrices W with $W_{n,m} \geq 0$ and constraint $\sum_n W_{n,m} = 0$, there exists a theorem in Mathematics (the Perron-Fröbenius theorem discussed below) which guarantees that (i) the eigenvalue $\lambda_0 = 0$ is the largest one, (ii) it is non degenerate, (iii) it is associated with a right eigenvector with positive components $R^{(0)} = (P_1^*, \dots, P_N^*)^T$, which can therefore be interpreted as the stationary distribution.²⁰

²⁰The Perron-Fröbenius theorem applies to “stochastic matrices”, i.e. matrices with positive entries and such that $\sum_n M_{nm} = 1 \forall m$. Here, the matrix W is a stochastic matrix shifted by the unity matrix : $M = \mathbf{1}_N + \mu W$ where $\mathbf{1}_N$ is the identity matrix and μ is a real number chosen such that all diagonal elements of M are positive.

b) Three types of stochastic processes

The nature of the different stochastic processes can be related to the properties of the stationary state. Depending on the matrix W , which *defines* the stochastic process, one encounters one of the three following situations.

- (i) **Equilibrium.**— Often, the existence of a stationary solution is ensured by a condition *stronger* than (3.51), called the **detailed balance condition**

$$\boxed{W_{nm} P_m^* = W_{mn} P_n^* \quad \forall (n, m)} \quad (\text{detailed balance} \equiv \text{equilibrium}) \quad (3.52)$$

i.e. ratio of occupations coincides with the ratio of transition rates

$$\frac{W_{nm}}{W_{mn}} = \frac{P_n^*}{P_m^*}. \quad (3.53)$$

If detailed balance is fulfilled, one says that the distribution P_n^* describes an *equilibrium state*. This is a *probabilistic definition of equilibrium*.

The detailed balance condition defines a first class of stochastic processes, those eventually reaching an equilibrium. From the point of view of the $N \times N$ matrix W , there are $\sim N^2$ conditions [there are $\sim N^2$ equations (3.52)], which is rather restrictive.

- (ii) **Non-equilibrium stationary state (NESS).**— If the detailed balance condition (3.52) is *not* fulfilled (at least for some couples (n, m)) but the condition

$$\sum_{m(\neq n)} \underbrace{[W_{nm} P_m^* - W_{mn} P_n^*]}_{\neq 0} = 0 \quad (\text{stationarity}) \quad (3.54)$$

holds ($\forall n$) for a normalisable distribution, one says that the stationary state is a *non-equilibrium steady state*.

This property defines a second class of stochastic processes, those reaching a NESS at large time. Note that the existence of the NESS involves only a set of $\sim N$ conditions ²¹

- (iii) **Transient process.**— When the size N of the matrix W is finite, the existence of a stationary state follows from the definition of the matrix (i.e. from the Perron-Fröbenius theorem). However, when $N \rightarrow \infty$, the eigenvector $R^{(0)} = (\dots, P_n^*, \dots)^T$ can be normalisable [cases (i) or (ii)], or not. In this latter case, when $\sum_n P_n^* = \infty$, the system never reaches a stationary state and the process is said to be *transient*. This defines the last family of stochastic processes.

Current interpretation : Detailed balance condition has a simple interpretation in terms of *probability currents* : for $n \neq m$, the matrix element W_{nm} being a rate, $J^*(m \rightarrow n) = W_{nm} P_m^*$ is the probability current from state m to state n . In the stationary state, currents between states should compensate globally, if probability is conserved : (3.54) expresses the vanishing of the sum of all currents arriving and leaving state n .

Detailed balance (3.52) means that probability currents equilibrate, $J^(m \rightarrow n) = J^*(n \rightarrow m)$, for all pairs of states.* There is no circulating probability current between more than two states (Fig. 3.10).

On the other hand, if detailed balance does not hold, the NESS is characterized by the existence of circulating probability currents around loops.

²¹The definition of the matrix W involves N constraints $\sum_n W_{nm} = 0$ for $m = 1, \dots, N$; those conditions are directly related to the N equations (3.54).

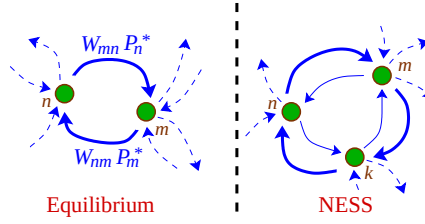


Figure 3.10: **Equilibrium state versus NESS.**— Denote by $J^*(m \rightarrow n) = W_{nm} P_m^*$ the probability current between two states, in the stationary state. At equilibrium, currents equilibrate for all pairs of states $J^*(n \rightarrow m) = J^*(m \rightarrow n)$. In the NESS, there exist loop(s) in which probability current(s) circulate : here, due to an imbalance $J^*(n \rightarrow m) > J^*(m \rightarrow n)$, there is a net circulation $n \rightarrow m \rightarrow k \rightarrow n$.

c) Simple examples :

Molecular vapour at thermal equilibrium : consider a vapour of molecules at thermal equilibrium. Each molecule has energy levels ε_n , expected to be occupied according to canonical weights $P_n^* \propto e^{-\beta \varepsilon_n}$. The molecule in an excited state falls in a state with lower energy by emission. Equilibrium and detailed balance imply that the absorption and emission rates between two levels fulfill the relation

$$\frac{\Gamma_{n \leftarrow m}}{\Gamma_{m \leftarrow n}} = \frac{P_n^*}{P_m^*} = e^{-\beta(\varepsilon_n - \varepsilon_m)} \quad (3.55)$$

i.e. emission is more probable than absorption. This observation (Gibbs equilibrium is compatible with an imbalance between emission and absorption) is the key point of Einstein first theory of spontaneous emission (at the origin of the difference between emission and absorption).²²

Equilibrium statistical physics : more generally, all situations described by equilibrium statistical physics and ensemble theory [71, 72] rely on the existence of transition rates which satisfy detailed balance (dynamical aspects, like transitions between microstates, are usually not deeply described in introductory courses on statistical physics ; however transitions are fundamental to ensure that the system reaches an equilibrium state).

Other simple examples : Exercise 3.10 discusses an extremely simple model of biased random walk in a ring, which illustrates the three situations. Another illustration is given in Exercise 3.19, where a richer model of random walk on $\mathbb{Z}$ for non uniform transition rates modeling the presence of a potential is studied,.

3.5 Spectral decomposition - Relaxation

In the previous paragraph, we have discussed the solution of the master equation (3.31) in the $t \rightarrow \infty$ limit. What can we say about the solution for finite t ? The master equation is a first order differential equation of the form $\frac{d}{dt} \text{Vector}(t) = \text{Matrix} \cdot \text{Vector}(t)$. Obviously such equation is conveniently solved by using the information on the spectrum of eigenvalues and eigenvectors of the matrix. The main difficulty is here related to the fact that the matrix is non symmetric in general.

²²Albert Einstein, “Zur Quantentheorie der Strahlung”, Physikalische Zeitschrift **18**, 121–128 (1917).

Einstein’s article is reproduced (and translated in french) in : A. Einstein, *Œuvres choisies. 1. Quanta*, Seuil (1989), textes choisis et présentés par F. Balibar, O. Darrigol & B. Jech.

🔪 **Exercise 3.7 – Warm up : right and left eigenvectors of a non-symmetric 2×2 matrix :** We consider the real non-symmetric matrix

$$M = \begin{pmatrix} 0 & e^{-h} \\ e^h & 0 \end{pmatrix} \quad (3.56)$$

Show that it is diagonalisable. Find the right and left eigenvectors and discuss the normalisation.

In the master equation, the matrix W is not symmetric in general, $W^T \neq W$. We have seen above that its eigenvalue $\lambda_0 = 0$ is associated with a couple of left and right eigenvectors $L^{(0)}$ and $R^{(0)}$. If W is diagonalisable, its eigenvalues are associated with a biorthogonal set of left and right eigenvectors $L^{(n)}$ and $R^{(n)}$. Let us write for convenience the equation for the eigenvalue as

$$WR^{(n)} = -\lambda_n R^{(n)} \quad \text{and} \quad W^T L^{(n)} = -\lambda_n L^{(n)} \quad (3.57)$$

where $\lambda_0 = 0 < \lambda_1 < \lambda_2 < \dots$ (the Perron-Fröbenius theorem discussed below guarantees this property).

Let us prove now that two left and right eigenvectors for different eigenvalues are orthogonal (but not the right eigenvectors among themselves, nor the left eigenvectors in general). We write the two equations (3.57) for two eigenvalues, $WR^{(n)} = -\lambda_n R^{(n)}$ and $L^{(m)T}W = -\lambda_m L^{(m)T}$ and multiply by the left and right eigenvectors, respectively,

$$\begin{cases} L^{(m)T}WR^{(n)} = -\lambda_n L^{(m)T}R^{(n)} \\ L^{(m)T}WR^{(n)} = -\lambda_m L^{(m)T}R^{(n)} \end{cases} \xrightarrow{\text{difference}} 0 = (\lambda_n - \lambda_m) L^{(m)T}R^{(n)}.$$

Hence, $\lambda_n \neq \lambda_m$ implies orthogonality $L^{(m)T}R^{(n)} = 0$. The choice of normalization is arbitrary. A convenient one is

$$L^{(n)T}R^{(m)} = \delta_{n,m}, \quad (3.58)$$

which ensures that the identity can be written

$$\sum_n R^{(n)} L^{(n)T} = \mathbf{1}_N \quad (3.59)$$

where $R^{(n)} L^{(n)T}$ is the projector on the eigenvector $R^{(n)}$. More importantly for the analysis, this leads to the **spectral representation** of the matrix

$$W = - \sum_n \lambda_n R^{(n)} L^{(n)T}. \quad (3.60)$$

This is very useful in order to solve the master equation. Denoting by $P(0) = (P_1(0), \dots, P_N(0))^T$ the initial state, we can write the solution of (3.31) in a matrix form and then get the probabilities :

$$P(t) = e^{tW} P(0) \quad \text{i.e.} \quad P_n(t) = \sum_j e^{-\lambda_j t} R_n^{(j)} \underbrace{L^{(j)T} P(0)}_{c_j \stackrel{\text{def}}{=}} = \sum_j c_j e^{-\lambda_j t} R_n^{(j)}. \quad (3.61)$$

c_j 's are the coefficients of the initial vector in the basis of eigenvectors $P(0) = \sum_j c_j R^{(j)}$. Note that $c_0 = L^{(0)T} P(0) = \sum_j P_j(0) = 1$ carries all the normalisation (orthonormalisation indeed implies that $L^{(0)T} R^{(j)} = \sum_n R_n^{(j)} = 0$ for $j \neq 0$).

Relaxation towards the stationary state.— Now let us discuss the large time behaviour. Using that $\lambda_0 = 0 < \lambda_1 < \lambda_2 < \dots$, the large time behaviour takes the form

$$P_n(t) \underset{t \rightarrow \infty}{\simeq} \underbrace{P_n^*}_{\equiv R_n^{(0)}} + \underbrace{c_1 e^{-\lambda_1 t} R_n^{(1)}}_{\xrightarrow{t \rightarrow \infty} 0} \quad (3.62)$$

where we have used $c_0 = 1$ (normalisation). This shows that $1/\tau_{\text{relax}} = \lambda_1$ is the relaxation rate towards the stationary state. Relaxation is usually exponentially fast, unless the gap in the spectrum vanishes and the spectrum is continuous.

Remark 1 : apart $\lambda_0 = 0$, the eigenvalues are not real in general. However, if some eigenvalues are complex, eigenvalues should come in conjugate pairs since M is a real matrix (and the same for eigenvectors). Thus, in the general case, the rate of relaxation towards stationary state is

$$\boxed{\frac{1}{\tau_{\text{relax}}} = \text{Re}(\lambda_1)} \quad (3.63)$$

and the ranking of e.v. is

$$\lambda_0 < \text{Re}(\lambda_1) < \text{Re}(\lambda_2) < \dots \quad (3.64)$$

Remark 2 : a non symmetric is not always diagonalisable. However, W can always be decomposed in terms of Jordan blocks associated to each degenerate eigenvalue. An example will be discussed in Exercise 5.20.

Conditional probability.— If we choose the initial condition $P_n(0) = \delta_{nm}$, the coefficients are $c_j = L_m^{(j)}$. This solution of the master equation is the conditional probability $P_n(t) \equiv P_t(n|m) = \sum_j e^{-\lambda_j t} R_n^{(j)} L_m^{(j)}$, which corresponds to the spectral decomposition of e^{tW} . This shows that the conditional probability can be written as

$$\boxed{P_t(n|m) = (e^{tW})_{nm}} \quad (3.65)$$

✎ **Exercise 3.8 – :** Check that

$$\sum_m P_t(n|m) P_m^* = P_n^* \quad (3.66)$$

3.6 Markov chains

An important class of stochastic processes are *Markov chains*, which are homogeneous stochastic processes, discrete with respect to both the time and the state. This makes such processes rather convenient for numerical analysis.

a) Stochastic matrix

We consider a stochastic process $X(t) \in \{1, \dots, N\}$ and denote $P_n(t) = \text{Proba}\{X(t) = n\}$. The master equation (3.31) introduced above involves transitions at random times. For Markov chain, the jumps occur at regular discrete times, thus the master equation takes the form

$$P_n(t+1) = \sum_m M_{nm} P_m(t) \quad (3.67)$$

where

$$M_{nm} = \text{Proba}\{m \rightarrow n\} \in [0, 1] \quad (3.68)$$

is the $N \times N$ matrix of transition probabilities at each elementary time step. M is called a “*stochastic matrix*”. It has non negative elements and satisfies

$$\boxed{\sum_n M_{nm} = 1} \quad (3.69)$$

Example of Markov chain : the biased RW.— A simple example is the case of the random walk on the line, where, at each time step, the walker jumps to the left with probability q or to the right with probability p . Then

$$M_{nm} = p \delta_{m,n-1} + q \delta_{m,n+1} \quad (3.70)$$

(with $p + q = 1$). The problem is studied in detail in Exercise 3.10 below.

For the following, it is useful to rewrite the master equation (3.67) in a form closer to the differential equation (3.30) by using (3.69)

$$P_i(t+1) - P_i(t) = \sum_{j(\neq i)} [M_{ij} P_j(t) - M_{ji} P_i(t)] \quad (3.71)$$

✎ **Exercise 3.9 – Continuum limit of the Markov chain:** Consider a Markov chain with jumps occurring every δt . Argue that the master equation (3.31) is recovered by considering the continuum limit

$$M_{ij} = \delta_{ij} + \delta t W_{ij} \quad \text{with } \delta t \rightarrow 0. \quad (3.72)$$

If M is a stochastic matrix, what is the constraint on the matrix W ?

b) The Perron-Fröbenius theorem and the stationary state

Discussion is similar as the one of previous paragraph. We can interpret the condition (3.69) as the existence of a left eigenvector $L^{(0)} = (1, \dots, 1)^T$ for eigenvalue $\Lambda_0 = 1$:

$$L^{(0)T} M = L^{(0)T} \quad \text{or} \quad M^T L^{(0)} = L^{(0)}, \quad (3.73)$$

The *Perron-Fröbenius theorem* states that, if M is a stochastic matrix,

- (i) $\Lambda_0 = 1$ is non-degenerate,²³
- (ii) Λ_0 is the largest eigenvalue (and the spectral radius of M is one, i.e. $|\Lambda_n| \leq 1 \forall n$),
- (iii) the related right eigenvector

$$M R^{(0)} = R^{(0)}, \quad (3.74)$$

has positive components. For a finite number of states N , this corresponds to the **stationary solution**, $R^{(0)} = (P_1^*, \dots, P_N^*)^T$. Normalization condition reads $L^{(0)T} R^{(0)} = \sum_i P_i^* = 1$ (the scalar product is the product of the line and the column vectors). We can rewrite the equation (3.74) as

$$\sum_{j(\neq i)} [M_{ij} P_j^* - M_{ji} P_i^*] = 0. \quad (3.75)$$

²³strictly speaking, $\Lambda_0 = 1$ is non-degenerate only if the states are not disconnected in the stationary state, for trivial reason(s). An example where this occurs is $M = \begin{pmatrix} 1 & 0 & p \\ 0 & 1 & q \\ 0 & 0 & 0 \end{pmatrix}$ with $p + q = 1$, which describes the situation where the system ends either on 1 or on 2, where it remains trapped forever. Because the stationary state describes two isolated islands, $\Lambda_0 = 1$ has degeneracy two. The Perron-Fröbenius theorem requires that the matrix M is *irreducible*, so that this situation does not happen.

c) Spectral analysis

We can solve (3.67) by using the spectral information on the matrix M . In the simplest situation, the matrix M is diagonalisable and has a biorthogonal set of right and left eigenvectors $R^{(n)}$, $L^{(n)}$ for eigenvalue Λ_n (with $|\Lambda_n| \leq 1$ according to the Perron-Fröbenius theorem). As a result, the solution of the master equation (3.67) is

$$P(t) = M^t P(0) = \sum_j c_j \Lambda_j^t R^{(j)} \underset{t \rightarrow \infty}{\simeq} R^{(0)} + c_1 \Lambda_1^t R^{(1)} + \dots \quad (3.76)$$

where $c_j = L^{(j)\text{T}} P(0)$ is the component of the initial vector on the eigenvector $R^{(j)}$; in particular $c_0 = 1$. In this case, the relaxation rate is $1/\tau_{\text{relax}} = -\ln |\Lambda_1|$.

Example : Let us illustrate the spectral analysis on a concrete example : I have constructed a random $N \times N$ stochastic matrix as follows : on each column, the number to matrix elements is taken from a binomial distribution $c = B(N, p)$ (there are c non zero matrix elements $M_{.,m}$, i.e. there are c possible transitions from the state m to other states). The c matrix elements are then picked at random so that their sum is one (i.e. the non zero matrix elements are $\sim 1/c \sim 1/(Np)$). Here is an example of matrix obtained in this manner (for $N = 10$ and $p = 1/3$) :

$$M = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.068 & 0.125 & 0 \\ 0 & 0.156 & 0 & 0.322 & 0.400 & 0.060 & 0.641 & 0 & 0 & 0 \\ 0.452 & 0 & 0 & 0 & 0.220 & 0.137 & 0.054 & 0 & 0 & 0 \\ 0.087 & 0 & 0 & 0 & 0 & 0.329 & 0 & 0 & 0 & 0 \\ 0.075 & 0.844 & 0.419 & 0 & 0.157 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.373 & 0.071 & 0 & 0.067 & 0 & 0 & 0 & 0.556 \\ 0 & 0 & 0 & 0.071 & 0 & 0.155 & 0 & 0 & 0.396 & 0 \\ 0.048 & 0 & 0 & 0.536 & 0 & 0.112 & 0 & 0 & 0.268 & 0.366 \\ 0.202 & 0 & 0.208 & 0 & 0 & 0.140 & 0.305 & 0 & 0.116 & 0.078 \\ 0.136 & 0 & 0 & 0 & 0.223 & 0 & 0 & 0.932 & 0.093 & 0 \end{pmatrix} \quad (3.77)$$

One can check that the sum of probabilities on each column is one. Since some transitions are forbidden, this matrix describes an oriented graph. The possible connections are represented in Figure 3.11 : for example, the first column has $c = 6$ non zero elements corresponding to the possible transitions $1 \rightarrow 3$, $1 \rightarrow 4$, $1 \rightarrow 5$ and $1 \rightarrow 8$, $1 \rightarrow 9$, $1 \rightarrow 10$.

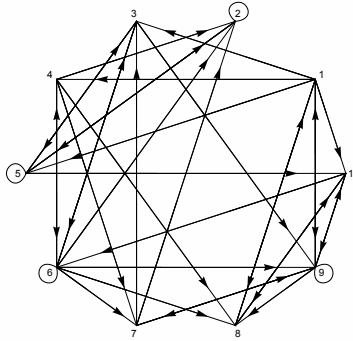


Figure 3.11: $N = 10$ states : connections described by the matrix (3.77).

The spectrum of eigenvalues of such matrix is complex. Examples of spectra are plotted in Fig. 3.12 for $N = 500$ and various values of p .

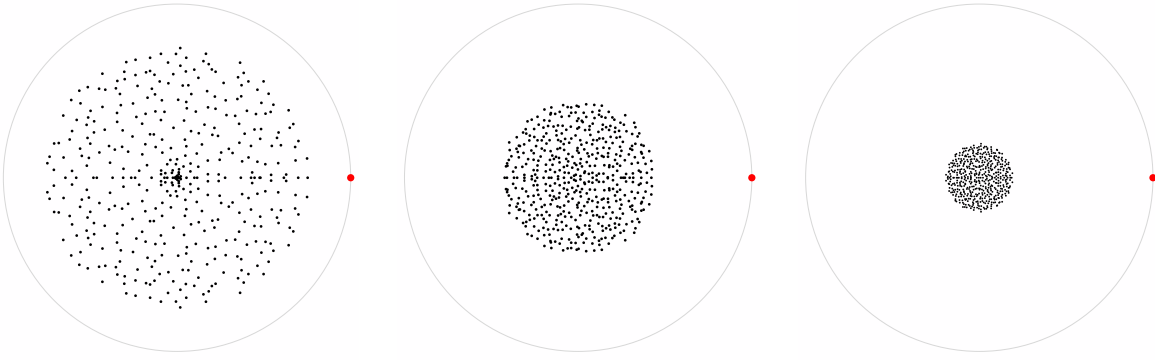


Figure 3.12: *Examples of spectra of some random stochastic $N \times N$ matrices in the complex plane. Here $N = 500$ and $p = 1/250$ (left), $p = 1/50$ (center) and $p = 1/10$ (right). The red dot is $\Lambda_0 = 1$, all others e.v. being inside the circle of unit radius. The symmetry with respect to the real axis is visible (e.v. come in c.c. pairs). The absence of an accumulation of e.v. close to $\Lambda_0 = 1$ ensures the existence of the stationary state. The symmetry of the spectra with respect to the real axis is visible (e.v. come in c.c. pairs).*

Conditional probability.— An example of initial condition is $P_n(0) = \delta_{nm}$ i.e. coefficients $c_j = L_m^{(j)}$. Then, the solution of the master equation (3.67) is the conditional probability

$$P_t(n|m) = (M^t)_{nm} \quad (3.78)$$

which is analogous to (3.65).

Exercise 3.10 – Biased random walk in a ring: Consider the random walk in a ring with L sites, such that with $M_{nm} = p\delta_{n,m+1} + q\delta_{n,m-1}$ for $n, m \in \{1, \dots, L\}$. Periodic boundary conditions are $M_{1L} = p$ and $M_{L1} = q$.

- Argue that the stationary state is an equilibrium state when $p = q = 1/2$ and a NESS for $p \neq q$.
- Give the spectrum of eigenvalues Λ_k and eigenvectors (left/right) of the stochastic matrix M . Write $p = \frac{1+v}{2}$ and $q = \frac{1-v}{2}$ with $v \in [-1, +1]$. Check that the “spectral radius” is unity, i.e. $|\Lambda_k| \leq 1 \forall k$.
- Decompose the conditional probability $P_t(n|m)$ over the eigenvalues and the eigenvectors.
- Consider the limit $L \rightarrow \infty$ and discuss the bottom of the spectrum. Compute $P_t(n|m)$. Discuss the case $v = \pm 1$.

d) Detailed balance, reversibility and ergodicity

Let us consider a Markov chain with master equation (3.67), such that detailed balance is fulfilled. We denote by P_i^* the *equilibrium* solution. Define

$$\mathcal{D}_t \stackrel{\text{def}}{=} \sum_i \frac{(P_i(t) - P_i^*)^2}{P_i^*} = \sum_i \frac{P_i(t)^2}{P_i^*} - 1 \geq 0 \quad (3.79)$$

which measures the distance of the distribution at time t to the equilibrium distribution. One can study the evolution of the quantity by considering $\Delta\mathcal{D}_t = \mathcal{D}_{t+1} - \mathcal{D}_t$. Some algebra making use of detailed balance condition $M_{ij}P_j^* = M_{ji}P_i^*$ leads to

$$\Delta\mathcal{D}_t = -\frac{1}{2} \sum_{i,j,k} M_{ji}M_{ki}P_i^* \left(\frac{P_j(t)}{P_j^*} - \frac{P_k(t)}{P_k^*} \right)^2 \leq 0 \quad (3.80)$$

Conclusion :

- $\mathcal{D}_t \geq 0$
- $\Delta \mathcal{D}_t \leq 0$
- We conclude that $\mathcal{D}_t \searrow$ and thus $P_i(t) \rightarrow P_i^*$.

This shows that detailed balance ensures that the system reaches equilibrium. This remark is borrowed from [45].

e) A practical (and important) application of Markov chains : the Monte Carlo method

Consider a physical observable $\mathcal{O}$. At thermal equilibrium, the probability of a microstate is $P_\ell \propto e^{-\beta E_\ell}$. If the number of states N_{state} is too large, it might be difficult to compute numerically the sum

$$\langle \mathcal{O} \rangle_{\text{eq}} = \sum_{\ell=1}^{N_{\text{state}}} P_\ell \mathcal{O}_\ell \quad (3.81)$$

For example, if one considers N Ising spins, the sum runs over $N_{\text{state}} = 2^N$ microstates, which becomes rapidly untrackable if N is large (a square of 10×10 spins $1/2$ has $N_{\text{state}} \sim 10^{30}$ microstates).

The central idea of equilibrium statistical physics is to replace the study of the microscopic (deterministic) dynamics by a probabilistic description. The main idea of Monte Carlo method is to replace the probabilistic description by a stochastic dynamics which explores reasonably well the states of the system, and perform a time averaging over the stochastic dynamics involving N_{step}

$$\overline{\mathcal{O}(t)} = \frac{1}{N_{\text{step}}} \sum_{t=1}^{N_{\text{step}}} \mathcal{O}(t). \quad (3.82)$$

The number of steps N_{step} can be chosen orders of magnitude smaller than N_{state} , still large enough in order to ensure some ergodicity (see the book [41] for a detailed discussion). There are several possible such dynamics.



Figure 3.13: *Nicholas Metropolis (1915-1999).*

Metropolis-Hastings algorithm.— if the system is in state $|i\rangle$ at time t , a move to another state $|f\rangle$ chosen randomly is made with probability

$$\text{Proba}\{\text{move } i \rightarrow f\} \equiv M_{fi} = \min \left(1, e^{-\beta(E_f - E_i)} \right). \quad (3.83)$$

For example, in a spin system, one chooses a spin randomly and flip it, thus the difference of energy $E_f - E_i$ is due to a local change, and the energy difference is very easy to compute.

This means that the matrix M_{nm} changes randomly at each elementary time step. Assuming $E_f > E_i$, the stochastic matrix has the form

$$M = \begin{pmatrix} \ddots & & & \\ & 1 - e^{-\beta(E_f - E_i)} & 1 & \\ & e^{-\beta(E_f - E_i)} & 0 & \\ & & & \ddots \end{pmatrix} \begin{matrix} \leftarrow |i\rangle \\ \leftarrow |f\rangle \end{matrix} \quad (3.84)$$

All other diagonal matrix elements are equal to one and all other non diagonal matrix elements equal to zero. Because $M_{fi}/M_{if} = e^{-\beta(E_f - E_i)}$, such dynamics converges towards the Gibbs equilibrium $P_n^* \propto e^{-\beta E_n}$.

Glauber dynamics.— Another popular dynamics is the Glauber dynamics, where

$$\text{Proba}\{\text{move } i \rightarrow f\} \equiv M_{fi} = \frac{1}{1 + e^{\beta(E_f - E_i)}} \quad (3.85)$$

while the probability for no move is $M_{ii} = e^{\beta(E_f - E_i)} / (1 + e^{\beta(E_f - E_i)})$. Selecting a pair of states, the stochastic matrix has a 2×2 block with these two probabilities and $M_{if} = M_{ii}$ and $M_{ff} = M_{fi}$. One can also check that Glauber dynamics satisfies detailed balance, hence we expect convergence towards the Gibbs equilibrium as well.

Conclusion.— Going from the microscopic dynamics to the Monte Carlo method, the scheme is the following :

<u>Classical mechanics</u>		<u>Equilib. statistical physics</u>		<u>Monte Carlo method</u>
deterministic	→	probabilistic	→	stochastic
evolution		description		dynamics

3.7 Beyond the master equation : renewal processes

I introduce here another class of stochastic processes, known as *Markov renewal processes*, which generalize the jump processes described by the master equation (3.26). I have pointed out that in the specific case of translation invariant jump processes (CPP, Exercise 3.6), the transition kernel takes the form $W(x|x') = \lambda w(x - x') - \lambda \delta(x - x')$, where λ is the jump rate and $w(\eta)$ the jump distribution (we will see in § 5.2 that this structure is not restricted to translation invariant processes like the CPP, but is very general and encoded in Eq. (3.26)). This observation makes clear that the process stays at its initial position during a random time τ distributed with an exponential law $q(\tau) = \lambda e^{-\lambda\tau}$ (this was discussed in Exercise 3.3).

We consider here a more general type of processes characterized by a *general waiting time distribution* $q(\tau)$. We do not discuss the general renewal theory, instead we concentrate on a situation which is translation invariant in space, i.e. simply generalizes the compound Poisson process. We consider a particle with position $X(t)$ starting from the origin at initial time $X(0) = 0$ and performing random jumps at random times

$$X(t_n^+) = X(t_n^-) + \eta_n. \quad (3.86)$$

The jump amplitudes are distributed according to the distribution $w(\eta)$. For Markov renewal processes, the distribution $q(\tau)$ of time intervals $\tau_n = t_n - t_{n-1} > 0$ is arbitrary (τ_n 's are i.i.d.).

a) Continuous time random walks (CTRW)

CTRW were introduced by Montroll and Weiss [56, 54, 55] in order to generalize normal diffusion and describe anomalous diffusion (sub- or super-diffusion). Let us denote $w(\eta)$ the jump distribution. The master equation for a jump process starting from $X(0) = 0$ is now replaced by the integral equation for CTRW (the “renewal equation”)

$$P(x, t) = \int_0^t d\tau q(\tau) \int_{\mathbb{R}} d\eta w(\eta) P(x - \eta, t - \tau) + \delta(x) \int_t^\infty d\tau q(\tau). \quad (3.87)$$

The last term corresponds to the probability that no jump has occurred up to time t . CTRW are translation invariant renewal processes (the jump distribution $w(\eta)$ does not depend on x , i.e. the point from which the jump is made), which makes the problem easily solvable.

✎ **Exercise 3.11 – CTRW:** We assume that the jump distribution w is a symmetric function.

a) Justify (3.87). Check normalisation.

b) If $q(\tau) = \lambda e^{-\lambda\tau}$, check that one recovers (3.44) (compound Poisson process).

c) Solve the equation by introducing the Laplace-Fourier transform

$$\tilde{P}(k, s) \stackrel{\text{def}}{=} \int_0^\infty dt e^{-st} \int_{\mathbb{R}} dx e^{-ikx} P(x, t) \quad (3.88)$$

Deduce $\tilde{P}(k, s)$ in terms of $\tilde{q}(s) = \int_0^\infty d\tau e^{-s\tau} q(\tau)$ and $\hat{w}(k) = \int_{\mathbb{R}} d\eta e^{-ik\eta} w(\eta)$. Express $P(x, t)$ under an integral form.

d) Consider distributions with power law tails $w(\eta) \simeq \frac{c}{|\eta|^{\mu+1}}$ for $\eta \rightarrow \pm\infty$ and $q(\tau) \simeq \frac{a}{\tau^{\alpha+1}}$ for $\tau \rightarrow +\infty$.

What is the $s \rightarrow 0$ behaviour of $\tilde{q}(s)$ for $\alpha > 1$?

For $\alpha < 1$, show that $\tilde{q}(s) \simeq 1 - A s^\alpha$, where A is a constant.

Same questions for $\hat{w}(k)$ (distinguish $\mu > 2$ and $\mu < 2$).

e) **Phase diagram.**— Discuss the limiting behaviour of $\tilde{P}(k, s)$ for $k \rightarrow 0$ and $s \rightarrow 0$. Deduce the scaling relation between space x and time t [hint : analyze the integral representation for $P(x, t)$].

Draw a “phase diagram” in the plane (μ, α) and identify the regions of normal diffusion, subdiffusion and superdiffusion.

Discuss the case $\mu = 2\alpha \in]0, 2[$.

3.8 Fourier analysis of stochastic processes – Wiener-Khinchin theorem

Let us close the chapter on general properties of stochastic processes with a somehow disconnected problem : the Fourier analysis of stochastic processes.

Consider a homogeneous (time translation invariant) and **stationary** random process $x(t)$ defined on the interval $t \in [0, T]$, where T is the observation time. It is characterised by the correlation function $C_{xx}(\tau) \stackrel{\text{def}}{=} \langle x(t) x(t + \tau) \rangle$, assumed rapidly decreasing (assume $\langle x(t) \rangle = 0$ for simplicity). Because the process is stationary, we prefer to consider its discrete Fourier transform (cf. appendix page 148) ²⁴

$$\tilde{x}_n = \int_0^T \frac{dt}{T} x(t) e^{+i\omega_n t} \quad \text{et} \quad x(t) = \sum_n \tilde{x}_n e^{-i\omega_n t} \quad \text{où} \quad \omega_n = \frac{2n\pi}{T} \quad \text{avec} \quad n \in \mathbb{Z}. \quad (3.89)$$

²⁴Later, we will define the Fourier transform in space as $f_q = \int_V dr f(r) e^{-iqr}$, where q is quantized if the volume is finite, and $f(r) = \frac{1}{V} \sum_q f_q e^{+iqr} \rightarrow \int \frac{dq}{(2\pi)^d} f_q e^{+iqr}$.

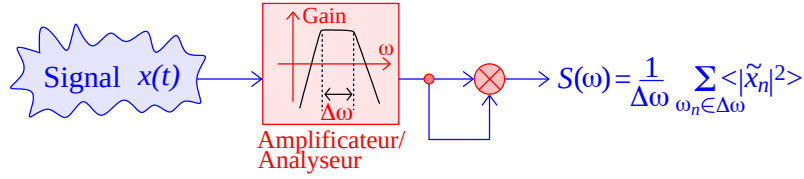


Figure 3.14: *Measure of the noise : the signal is amplified, duplicated and multiplied by itself. The result is averaged over a long time T .*

Noise spectrum.— Let $\Delta\omega$ be the bandwidth of the apparatus (with $\Delta\omega \gg 1/T$). We define the *noise spectrum* as the average of the square modulus of the Fourier components in the bandwidth, i.e. in the interval $[\omega, \omega + \Delta\omega]$:

$$\mathcal{S}(\omega) \stackrel{\text{def}}{=} \frac{1}{\Delta\omega} \sum_{\omega_n \in [\omega, \omega + \Delta\omega]} \langle |\tilde{x}_n|^2 \rangle \quad (3.90)$$

This is precisely the outcome of the device represented in figure 3.14 : sample $\rightarrow$ ampli/filter $\rightarrow$ multiplier $\rightarrow$ measurement. It measures the fluctuations of the Fourier components belonging to the bandwidth.



Figure 3.15: *Norbert Wiener (1894-1964) & Aleksandr Yakovlevich Khinchin (1894-1959).*

Wiener-Khintchin theorem.— From the above hypothesis, one can verify that :

$$\langle \tilde{x}_n \tilde{x}_m^* \rangle = \frac{1}{T} \delta_{n,m} \tilde{C}_{xx}(\omega_n) \quad \text{where } \tilde{C}_{xx}(\omega) \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} d\tau C_{xx}(\tau) e^{i\omega\tau} \quad (3.91)$$

is the Fourier transform of the correlation function. Only components corresponding to opposite frequencies ω_n and ω_{-n} are correlated²⁵. Thus one has : $\sum_{\omega_n \in [\omega, \omega + \Delta\omega]} \langle |\tilde{x}_n|^2 \rangle = \frac{\mathcal{N}_{\Delta\omega}}{T} \tilde{C}_{xx}(\omega)$ where $\mathcal{N}_{\Delta\omega} = \Delta\omega T / 2\pi$ is the number of frequencies ω_n in the bandwidth. Finally one gets

$$\boxed{\mathcal{S}(\omega) = \frac{\tilde{C}_{xx}(\omega)}{2\pi}} \quad (3.92)$$

i.e. a relation between the noise spectrum (fluctuations at frequency ω) and the correlations. A random process characterized by short time correlations thus corresponds to a broad noise spectrum. The limit of correlation with zero range is called a “white noise” (flat spectrum).

²⁵One can as well consider a process defined on $\mathbb{R}$ by writing $T \rightarrow \infty$. The Fourier transform is then defined as $\tilde{x}(\omega) = \int dt x(t) e^{i\omega t}$ and one can show that $\langle \tilde{x}(\omega) \tilde{x}(\omega') \rangle = 2\pi \delta(\omega + \omega') \tilde{C}_{xx}(\omega)$. Correspondence between the two formulations is ensured by the substitutions $\tilde{x}(\omega) \leftrightarrow T \tilde{x}_n$ and $2\pi \delta(\omega - \omega') \leftrightarrow T \delta_{n,n'}$.

Details of derivation of (3.91) : Write

$$\langle \tilde{x}_n \tilde{x}_m^* \rangle = \int_0^T \frac{dt}{T} \int_0^T \frac{dt'}{T} e^{i\omega_n t - i\omega_m t'} C_{xx}(t - t') = \frac{1}{T} \int_0^T \frac{dt}{T} e^{i(\omega_n - i\omega_m)t} \int_{t-T}^t du e^{i\omega_n u} C_{xx}(u)$$

where we have performed the change of variable $(t, t') \rightarrow (t, u = t - t')$ (with Jacobian one). Short range correlation assumption allows to write $\int_{t-T}^t du \cdots \rightarrow \int_{-\infty}^{+\infty} du \cdots$, leading to (3.91). QED.

Note that we have considered the discrete Fourier transform of the process (due to translation invariance in time), however the calculation involves the Fourier transform of the correlation function on $\mathbb{R}$, as it decays fast.

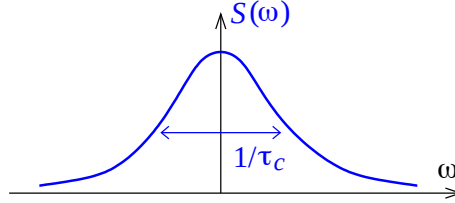


Figure 3.16: **Wiener-Khintchin theorem** : width of noise spectrum is inversely proportional to the correlation time τ_c of the process (τ_c is the width of $C(\tau)$).

✎ **Exercise 3.12** – : As a simple application of the Wiener-Khintchin theorem, we analyze the correlation of the velocity for the process defined by the phenomenological Langevin equation

$$\frac{dv(t)}{dt} = - \int dt' \gamma(t - t') v(t') + \xi(t) \quad (3.93)$$

where $\xi(t)$ is the Langevin force (assumed to be a stationary random process with short time correlations). Here the friction is nonlocal in time, controlled by a causal function $\gamma(t)$, with finite width τ_m . Show that

$$C_{vv}(\tau) = \int_{-\infty}^{+\infty} d\omega \underbrace{\frac{\tilde{C}_{\xi\xi}(\omega)}{2\pi}}_{S_{\text{Force}}(\omega)} \frac{e^{-i\omega\tau}}{|\tilde{\gamma}(\omega) - i\omega|^2} \quad (3.94)$$

Consider the limit $C_{\xi\xi}(\tau) = 2D\gamma^2\delta(\tau)$ et $\gamma(t) = \gamma\delta(t)$ and compute explicitly the correlator.

✎ **Exercise 3.13** – : Consider now the case of a Langevin force correlated over the finite time τ_c (a microscopic time) : $C_{FF}(t) = 2D\gamma^2 \frac{1}{2\tau_c} e^{-|t|/\tau_c}$. We expect the function $\gamma(t)$ to be of finite width τ_m ; Assume $\gamma(t) = \gamma\theta(t) \frac{1}{\tau_m} e^{-t/\tau_m}$. The three time scales fulfill : $\tau_c \lesssim \tau_m \ll 1/\gamma$. Analyze the residus of $|\tilde{\gamma}(\omega) - i\omega|^{-2}$ justify that one can consider $\tau_m \rightarrow 0$ while keeping a finite τ_c , what simplifies the evaluation of the integral. Show then that $C_{vv}(\tau) = \frac{D\gamma}{1-(\gamma\tau_c)^2} [e^{-\gamma|\tau|} - \gamma\tau_c e^{-|\tau|/\tau_c}]$. Analyze the behaviour at short time as well.

☺ Important points

- Markov process (definition).
- Be familiar with the various forms of the master equation (continuous/discrete ; Markov chain).
- A good exercise : recover the properties of the Poisson process (and the CPP).
- Definition of the stochastic matrix. Use of spectral information to solve the master equation.
- Detailed balance and the classification of Markov processes.
- Wiener-Khintchin theorem : relation between the correlation function of an homogeneous process and its noise spectrum.

Additional exercises

✎ **Exercise 3.14 – Solution of the Kramers equation (free case) :** We consider a particle in a fluid described by the two differential equations (2.3,2.4), where we recall that the Langevin force is a Gaussian white noise such that $\langle \xi(t)\xi(t') \rangle = 2\gamma k_B T \delta(t-t')$. The Kramers equation is the related equation for the joint distribution $P_t(x, v)$, which was discussed in this chapter and is also given in the next chapter, Eq. (4.34). Here we consider fixed initial conditions $x(0) = x_0$ and $v(0) = v_0$.

- What is the form of the conditional joint probability $P_t(x, v | x_0, v_0)$?
- Collecting the results obtained in chapter 2 (lecture and exercises), give $\langle x(t) \rangle$, $\langle v(t) \rangle$, $\langle x(t)^2 \rangle_c$ and $\langle v(t)^2 \rangle_c$. Compute also $\langle x(t)v(t) \rangle_c$.
- Deduce $P_t(x, v | x_0, v_0)$.
- Discuss the marginal distribution of the position $W_t(x | x_0, v_0) \stackrel{\text{def}}{=} \int dv P_t(x, v | x_0, v_0)$ for fixed initial velocity and also when the particle is initially at equilibrium, i.e. $Q_t(x | x_0) \stackrel{\text{def}}{=} \int dv dv_0 P_t(x, v | x_0, v_0) P_{\text{eq}}(v_0)$.
- Compute $\int dx_1 Q_t(x | x_1) Q_{t'}(x_1 | x_0)$ and discuss the result.

✎ **Exercise 3.15 – Poisson noise (derivative of the PP) :** We consider the noise $\xi(t) = \sum_n \delta(t - t_n)$, where the times are i.i.d. for a uniform density λ . I.e., when they are ordered, the events occur randomly and independently with rates λ . In other terms, the noise is the derivative of the Poisson process introduced above $\xi(t) = \mathcal{N}'(t)$.

We introduce the generating function of the noise $G[h] \stackrel{\text{def}}{=} \langle \exp \int dt h(t) \xi(t) \rangle$, where $\langle \bullet \rangle$ is the averaging over the random times t_n 's.

- Show that $G[h] = \exp \left\{ \lambda \int dt (e^{h(t)} - 1) \right\}$.
Hint: Consider that the N times are not ordered, distributed over $[0, T]^N$ with measure $dt_1 \cdots dt_N / T^N$.
- Deduce the connex correlation functions (cumulants) : $\langle \xi(t) \rangle = \lambda$ and $\langle \xi(t_1) \cdots \xi(t_n) \rangle_c = \lambda \delta(t_1 - t_2) \cdots \delta(t_1 - t_n)$.
Hint: Consider functional derivatives of $\ln G[h]$.

In conclusion, $\xi(t) = \mathcal{N}'(t)$ is a **non Gaussian white noise**.

✎ **Exercise 3.16 – n -point correlations of the PP :** Using the result of Exercise 3.15, show that the n -point correlation function is

$$\langle \mathcal{N}(t_1) \mathcal{N}(t_2) \cdots \mathcal{N}(t_n) \rangle_c = \lambda \min(t_1, t_2, \dots, t_n). \quad (3.95)$$

✎ **Exercise 3.17 – Derivative of the CPP – a non Gaussian white noise :**

- Using the representation $\xi(t) = X'(t) = \sum_n \eta_n \delta(t - t_n)$, where both the times t_n 's and the coefficients η_n 's are random, derive the connex correlation function of the noise $\langle \xi(t_1) \cdots \xi(t_n) \rangle_c$ (the analogous of a cumulant).

Hint: follow the same steps as in Exercise 3.15.

- Show that the noise becomes a Gaussian white noise in a certain limit.

✎ **Exercise 3.18 – Queueing theory :** We are interested in the state of a queue with $n \geq 0$ individuals, arriving or leaving the queue at random times one by one. When the queue has length n , the rate for arrival of one individual is b_n and the rate of departure is d_n . A natural question is to know what is the state of the queue at time t . Can it exist a stationary state for the queue, or does it keep growing ?

- Write the master equation for the distribution $P_n(t)$ of the length of the queue at time t . Identify the related probability current $J_{n+1/2}(t)$ between states n and $n+1$.
- If there exists a stationary state, is it an equilibrium or a NESS ?

- c) Deduce a condition on the rates for the existence of a stationary state P_n^* .
- d) *M/M/1 queueing system* : For uniform rates, $b_n = b$ and $d_n = d \forall n$, give P_n^* and compute the average length of the queue.
- e) Discuss the length of the queue when $b = d$.

✎ **Exercise 3.19 – Random walk on $\mathbb{Z}$ with a potential : the 3 scenarii of stochastic processes** : Let us consider the master equation describing the one dimensional diffusion on $\mathbb{Z}$ with transitions between nearest neighbour sites

$$\frac{dP_n(t)}{dt} = W_{n,n-1}P_{n-1}(t) + W_{n,n+1}P_{n+1}(t) - (W_{n-1,n} + W_{n+1,n})P_n(t) \quad (3.96)$$

i.e. $W_{n,m}$ is a tridiagonal matrix with $W_{n,n} = -W_{n-1,n} - W_{n+1,n}$ (if $n \in \mathbb{Z}$, the matrix has an infinite size). Hence, this is another example of birth and death process.

- a) *Current* : check that the master equation can be rewritten under the form

$$\frac{dP_n}{dt} = -J_n + J_{n-1} \quad (3.97)$$

and express the probability current $J_n(t)$ related to the distribution $P_n(t)$ (J_n measures the current at time t between sites n and $n+1$).

We now choose the matrix such that

$$W_{n,m} = e^{[V(m)-V(n)]/2} \quad (3.98)$$

where $V(x)$ is a known function.

- b) *Equilibrium* ($J = 0$).— Show that

$$P_n^* = C e^{-V(n)} \quad (3.99)$$

is a stationary solution corresponding to a vanishing probability current. Discuss the normalisability. Give an example of $V(n)$ leading to an equilibrium.

- c) *NESS* ($J \neq 0$).— Find the stationary solution corresponding to $J_n = J \forall n$. Show that it is

$$P_n^* = J e^{-V(n)} \sum_{m=n}^{\infty} e^{[V(m+1)+V(m)]/2} \quad (3.100)$$

Discuss the normalisability (consider the continuum limit for simplicity). Give an example of $V(n)$ leading to a NESS.

- d) Provide an example where there is no stationary state.

✎ **Exercise 3.20 – Broadening of the line shape by phase noise** : A field $\mathcal{E}(t)$ is measured with a detector characterized by the response function $\chi(\omega)$ (the response of the detector at frequency ω) such that

$$\int_{\mathbb{R}} \frac{d\omega}{2\pi} |\chi(\omega)|^2 = 1 \quad (3.101)$$

The outcome of the detector (the signal) is

$$S = \left| \int_{\mathbb{R}} \frac{d\omega}{2\pi} \chi(\omega) \tilde{\mathcal{E}}(\omega) \right|^2 \quad \text{where } \tilde{\mathcal{E}}(\omega) = \int_{\mathbb{R}} dt e^{i\omega t} \mathcal{E}(t) \quad (3.102)$$

We consider a monochromatic field with frequency ω_0 , carrying additionally a random phase $\theta(t)$

$$\mathcal{E}(t) = \mathcal{E}_0 e^{-i\omega_0 t + i\theta(t)} . \quad (3.103)$$

We assume that the phase shows normal diffusion, i.e. $\theta(t)$ is a free Brownian motion starting from $\theta(0) = 0$ (Wiener process) such that $\langle \theta(t)^2 \rangle = 2Dt$, where D is the diffusion constant of the phase (here we can consider that $\theta \in \mathbb{R}$).

a) Compute $\langle [\theta(t) - \theta(t')]^2 \rangle$.

b) Argue that $\langle e^{i[\theta(t) - \theta(t')]} \rangle = e^{-\frac{1}{2}\langle [\theta(t) - \theta(t')]^2 \rangle}$.

c) Compute $\langle \tilde{\mathcal{E}}(\omega) \tilde{\mathcal{E}}(\omega')^* \rangle$. Compare the result to the expected one in the absence of the random phase.

d) Deduce $\langle S \rangle$ as a simple integral over the frequency. Analyse $\langle S \rangle$ in the limit $D \rightarrow 0$ and in the limit $D \rightarrow \infty$.