August 2023 update: December 11, 2023

Stochastic processes

Master 2 Physics of Complex Systems

Christophe Texier

LPTMS Laboratoire de Physique Théorique et Modèles Statistiques — université PARIS-SACLAY

Prerequisites :

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

Contents

1 Introduction

This set of lectures is devoted to stochastic processes and out-of-equilibirum 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-equilibirum statistical physics requires a statistical treatment of the dynamics, which can be achieved by various approaches, phenomenoligical 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,...). 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 (§ e) page 58).

2 Probability : some useful concepts

I give a physicist's introduction of several useful concepts. For a mathematical monograph, see the standard book of Feller [15].

2.1 Events, probability and random variable

Random events occur when the underlying dynamics is too complex (for example when one plays heads or tails, [1] 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. $\Omega = \{\text{head, tail}\}$ when tossing a coin). We denote by $\mathscr{P}(\omega) \geqslant 0$ the probability of occurence of the event, which is the relative frequency of occurence of ω when a large number of observations is made. Obviously, $\mathscr{P}(\Omega) = 1$.

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 $\mathscr{P}(\omega)$, we will introduce the distribution of the random variable

$$P(x) = \text{Proba}\{X(\omega) = x\} = \sum_{\omega\in\Omega} \mathscr{P}(\omega)\, \delta_{X(\omega),x} \tag{1}$$

Importantly, the distribution satisfies the normalisation condition

$$\sum_x P(x) = \mathscr{P}(\Omega) = 1\,. \tag{2}$$

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.*

[1] "pile ou face" ; "testa o croce".

Example : 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 : binomial distribution.— Play to heads or tails 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

$$\text{Proba}\{H_N = n\} = C_N^n \, p^n \, (1-p)^{N-n} \tag{3}$$

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 : Poisson distribution.— The distribution of the number of occurences of uncorrelated events occuring with constant rate during a finite time

$$\text{Proba}\{X = n\} = \frac{q^n}{n!} \, \mathrm{e}^{-q} \tag{4}$$

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

$$g_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)} \, . \tag{5}$$

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

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

$$\mathcal{C}_{\mu,a}(x) = \frac{a/\pi}{(x-\mu)^2 + a^2} \, . \tag{6}$$

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 $A_1, A_2, \cdots \in \mathcal{A}$, then $\bigcup_n A_n \in \mathcal{A}$. We say that "$\mathcal{A}$ is a σ-algebra on Ω".

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

$$\mathscr{P}\big(\bigcup_i A_i\big) = \sum_i \mathscr{P}(A_i)$$

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

If you like this style, you can look at Øksendal's book [43].

✍ ***Exercice 1 – Student distribution :*** *The student distribution of index $\mu > 0$ is defined as*

$$S^{(\mu)}(x) = c \, \left(1 + x^2/\mu\right)^{-\frac{1+\mu}{2}} \tag{7}$$

a) Compute the normalisation constant

b) We denote by $C^{(\mu)}(x) = \int_{-\infty}^{x} \mathrm{d}t\, S^{(\mu)}(t)$ the cumulative distribution. Give $C^{(1)}(x)$.

c) check that $C^{(2)}(x) = \frac{1}{2} + \frac{x}{2\sqrt{2+x^2}}$.

d) What is $\lim_{\mu\to\infty} S^{(\mu)}(x)$?

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

$$Y = \varphi(X)\,? \tag{8}$$

(the case of a non monotonous φ is not much more difficult). It is important to remember that only probabilities can be made equal (not densities, which bytheway may have different physical dimensions) : $p(x)\,\mathrm{d}x = \mathrm{Proba}\{X \in [x, x+\mathrm{d}x]\} = \mathrm{Proba}\{Y \in [\varphi(x), \varphi(x)+\mathrm{d}\varphi(x)]\} = q(\varphi(x))\,\varphi'(x)\mathrm{d}x$, i.e. $p(x) = q(\varphi(x))\,\varphi'(x)$. To be short, remember

$$p(x)\,\mathrm{d}x = q(y)\,\mathrm{d}y \qquad \text{with } y = \varphi(x)\,. \tag{9}$$

Note that it can also be convenient to write

$$q(y) = \langle \delta(y - \varphi(X)) \rangle = \int \mathrm{d}x\, p(x)\, \delta(y - \varphi(x)) \tag{10}$$

(this form allows to deal with a non monotonous function).

✍ ***Exercice 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*

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

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

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

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

c) Application n°1 : Exponential distribution $p(x) = (1/a)\mathrm{e}^{-x/a}$ for $x > 0$.

d) Application n°2 : Cauchy distribution $p(x) = \pi^{-1}(x^2+1)^{-2}$ for $x \in \mathbb{R}$.

e) Application n°3 : Student distribution $S^{(2)}(x)$ for $x \in \mathbb{R}$.

✍ ***Exercice 3 – Box-Muller algorithm :*** *The exercise explains how to generate a Gaussian random number with a computer (preliminary : you should study at least the first questions of exercise 2).*

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

a) What is the distribution of the radius $R = \sqrt{X^2+Y^2}$? What is the distribution of the angle Θ ?

b) What is the distribution of $\xi = \frac{1}{2}(X^2+Y^2)$?

c) Deduce a method to generate a Gaussian random number from a box distribution.

✍ ***Exercice 4* – : *We show here how to generate easily a random number with power law distribution.***

a) We first consider the distribution

$$p(x) = \mu\, x^{-1-\mu} \qquad \text{for } x \geqslant 1\,. \tag{11}$$

Deduce how to generate such a random variable from a box distribution.

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

2.2 Stochastic independence – Joint and conditional probabilities

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

$$P(A\,|\,B) = \frac{P(A\cap B)}{P(B)} \tag{12}$$

As a consequence it is clear that $P(A\cap B) = P(A\,|\,B)\,P(B) = P(B\,|\,A)\,P(A)$ (Bayes theorem).

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

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

$$P(m,c) = \text{Proba}\{\#\text{ of mice} = m \;\&\; \#\text{ of cats} = c\} \tag{13}$$

the *joint probability distribution* 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

$$p(m) = \sum_c P(m,c) \tag{14}$$

(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 mice. We then introduce the conditional probability

$$P(m\,|\,c) = \frac{P(m,c)}{q(c)} \tag{15}$$

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

$$\sum_m P(m\,|\,c) = 1\,. \tag{16}$$

- If we consider mice in Paris and cats in Torino, we expect the two random variables to be independent, [2] hence

$$\boxed{P(m,c) = p(m)\,q(c) \qquad : \text{stochastic independence}} \tag{17}$$

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

- If instead we focus on Parisian cats and mice, we expect that the two populations interact and their numbers are not independent, [3] hence the conditional probability $P(m\,|\,c)$ does not coincide with $p(m)$ and
$$P(m,c) \neq p(m)\,q(c) \tag{18}$$

✍ ***Exercice 5 – Conditional probability for 2 Gaussian variables :*** *Consider two real random variables distributed according to the Gaussian distribution $P(x,y) = \mathcal{N}\,\exp\big[-\frac{1}{2}ax^2 + bxy - \frac{1}{2}cy^2\big]$.*

a) Compute the normalisation constant $\mathcal{N}$. What is the condition on a, b and c ? Determine the conditional probability $P(x|y)$.

b) Deduce $\langle X\,|\,Y=y\rangle$ the average of X conditioned by $Y=y$.

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 $\big\langle (x_t-x_0)^2\big\rangle = 2Dt$, where D is the diffusion constant.

✍ ***Exercice 6 – 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 $y=\sin\theta$. 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 ?*

2.3 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. $P(x)\,\mathrm{d}x = \mathrm{Proba}\{X\in[x,x+\mathrm{d}x]\}$. The average, or "mean value" [4] of the variable is $\langle X\rangle = \int \mathrm{d}x\,P(x)\,x$. For any known function $\phi(x)$, the average of $\phi(X)$ is
$$\langle\phi(X)\rangle = \int \mathrm{d}x\,P(x)\,\phi(x)\,. \tag{19}$$

a) Moments

We denote by
$$\mu_n = \langle X^n\rangle \tag{20}$$
the n-th moment. It exists if the integral $\int \mathrm{d}x\,P(x)\,x^n$ 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*
$$G(k) \stackrel{\text{def}}{=} \langle \mathrm{e}^{kX}\rangle = \sum_{n=0}^{\infty}\frac{\mu_n}{n!}k^n\,, \tag{21}$$
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

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

[4] The Mathematicians talk about "expectation value", denoted $\mathbb{E}(X)$.... However, unless the distribution is very narrow, there is no reason to "expect" the mean value!

$G(k)$, one can deduce the moments by simple derivations

$$\mu_n = G^{(n)}(0)\ , \tag{22}$$

the function G "*generates*" the moments (by derivations).

✍ ***Exercice 7 – Gamma distribution :*** *The Gamma distribution is defined by*

$$P^{(\mu)}(x) = c\ \theta_{\rm H}(x)\, x^{\mu-1}\, {\rm e}^{-x} \tag{23}$$

where $\mu > 0$. We introduce the Heaviside function $\theta_{\rm H}(x)$ and c is a normalisation constant.

a) compute c.

b) Compute the moments μ_n.

c) Compute the generating function $G(k)$. Specify when it is defined.

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

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

✍ ***Exercice 8 – :*** *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 $\Gamma(z) = \int_0^\infty {\rm d}t\, t^{z-1}\, {\rm e}^{-t}$.

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 funtion*, i.e. the Fourier transform of the distribution

$$\widehat{P}(k) = \langle {\rm e}^{-{\rm i}kX} \rangle \equiv G(-{\rm i}k) \tag{24}$$

which always exists (the integral $\int {\rm d}x\, P(x)\, {\rm e}^{-{\rm i}kx}$ is absolutely convergent).

For example, consider a positive random variable X with distribution with power law tail $P(x) \sim x^{-1-\mu}$ for $x \to \infty$. Clearly $\langle X^n \rangle < \infty$ for $n < \mu$ and $\langle X^n \rangle = \infty$ for $n \geqslant \mu$.

✍ ***Exercice 9 – :*** *Discuss the moments of the Cauchy distribution (6).*
Compute its characteristic function $\widehat{\mathcal{C}}_{\mu,a}(k)$.

When all moments are finite, we can write $\widehat{P}(k) = \sum_{n=0}^\infty \frac{\mu_n}{n!}(-{\rm i}k)^n$, i.e. $\widehat{P}(k)$ is an analytic function of the variable k. When some moments are infinite, such a series representation does not exist. Because μ_n is related to $\widehat{P}^{(n)}(0)$, the derivatives exist as long as moments are finite. For the tail $P(x) \sim x^{-1-\mu}$, moments up to $n = \lfloor \mu \rfloor$ are finite, thus $\widehat{P}(k)$ 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 $P(x) = \frac{4/\pi}{4+x^4}$. 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 Using residue theorem, on gets $\widehat{P}(k) = \langle {\rm e}^{-{\rm i}kX} \rangle = \big[\cos k + \sin|k|\big]{\rm e}^{-|k|}$. Expansion of the function for $k \to 0$ reads $\widehat{P}(k) = 1 - k^2 + \frac{2}{3}|k|^3 + \mathcal{O}(k^4)$, which shows that $\widehat{P}(k) \in \mathscr{C}_2(\mathbb{R})$ (continuous function differentiable twice everywhere).

✍ ***Exercice 10 – :*** *Use the residue theorem to show that $\widehat{P}(k) = \big[\cos k + \sin|k|\big]{\rm e}^{-|k|}$.*

✍ ***Exercice 11 – :*** *Consider a random variable X distributed according to the law*

$$w(x) = \frac{1}{\Gamma(\mu)\, x^{1+\mu}}\, {\rm e}^{-1/x}\ , \tag{25}$$

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

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

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

✍ ***Exercice 12*** *– : In this exercise we show that the $x \to \infty$ behaviour of the distribution is related to the $k \to 0$ of its Laplace transform. We consider a positive random variable $X > 0$, with distribution with tail $p(x) \simeq c\, x^{-1-\mu}$ with $\mu \in]0, 1[$. Then it is possible to define its Laplace transform (i.e. the generating function).*

a) Show that the characteristic function (Laplace transform) presents the behaviour $\tilde{p}(\beta) \stackrel{\text{def}}{=} G(-\beta) \simeq 1 - A\, \beta^\mu$ for $\beta \to 0$ (give the constant A).

Hint : start from $\tilde{p}(\beta) = \int_0^\infty \mathrm{d}x\, p(x)\, e^{-\beta x} = 1 - \int_0^\infty \mathrm{d}x\, p(x)\, (1 - e^{-\beta x})$.

b) Consider now $\mu \in]1, 2[$. How could you adapt the same trick. Without performing the precise calculation, give the $\beta \to 0$ behaviour of $\tilde{p}(\beta)$.

2.4 Variance and cumulants

a) Variance and standard deviation

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

$$\mathrm{var}(X) \stackrel{\text{def}}{=} \left\langle [X - \langle X \rangle]^2 \right\rangle = \left\langle X^2 \right\rangle - \langle X \rangle^2 \tag{26}$$

We also introduce the *standard deviation*

$$\sigma \stackrel{\text{def}}{=} \sqrt{\mathrm{var}(X)} \tag{27}$$

which gives the order of the fluctuations of the variable.

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 carried by the variance ? The answer is yes and corresponds to introduce *cumulants*.

b) Generating function of the cumulants

Cumulants are defined from the generating function

$$W(k) = \ln G(k) = \sum_{n=1}^{\infty} \frac{\kappa_n}{n!} k^n \tag{28}$$

where κ_n is the cumulant of order n.

✍ ***Exercice 13*** *– : 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 are*

$$\kappa_1 = \mu_1 \tag{29}$$

$$\kappa_2 = \mu_2 - \mu_1^2 = \langle [X - \langle X \rangle]^2 \rangle \tag{30}$$

$$\kappa_3 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3 = \langle [X - \langle X \rangle]^3 \rangle \tag{31}$$

$$\kappa_4 = \mu_4 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4 = \langle [X - \langle X \rangle]^4 \rangle - 3\langle [X - \langle X \rangle]^2 \rangle^2 \tag{32}$$

✍ ***Exercice 14*** *– : 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 $\sum_i |x_i y_i| \leqslant \left(\sum_i |x_i|^{1/\alpha}\right)^\alpha \left(\sum_i |y_i|^{1/\beta}\right)^\beta$, with $\alpha+\beta=1$ and $\alpha,\ \beta\in[0,1]$. The substitution $x_i \longrightarrow \mathrm{e}^{\alpha k X_i}$ and $y_i \longrightarrow \mathrm{e}^{\beta k' X_i}$ leads to the form :

$$\ln\left(\sum_i \mathrm{e}^{\alpha k X_i+\beta k' X_i}\right) \leqslant \alpha \ln\left(\sum_i \mathrm{e}^{k X_i}\right) + \beta \ln\left(\sum_i \mathrm{e}^{k' X_i}\right) \tag{33}$$

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

$$W\left(\alpha k+(1-\alpha)k'\right) \leqslant \alpha\, W(k) + (1-\alpha)\, W(k') \qquad \forall\, \alpha\in[0,1] \tag{34}$$

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

The Marcinkiewicz theorem : A theorem proved by Marcinkiewicz in 1939 [37] (see also [7]). 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.

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

$$\text{skewness} = \kappa_3/\kappa_2^{3/2}\,. \tag{35}$$

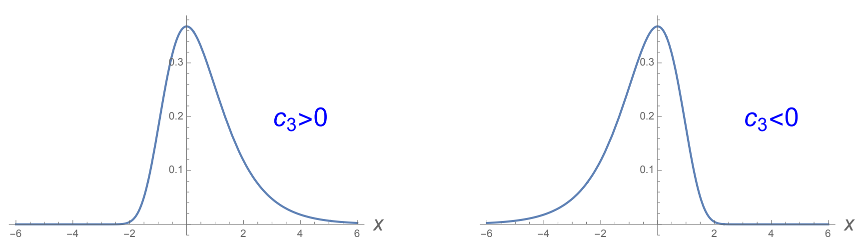


Figure 1: *À gauche : la loi de Gumbel $f(x)=\exp(-x-\mathrm{e}^{-x})$ a pour premiers cumulants : $\kappa_1 = \mathbf{C} = 0.577...$ (constante d'Euler), $\kappa_2=\pi^2/6$, $\kappa_3 = 2\zeta(3)$ (fonction zeta de Riemann), $\kappa_4=\pi^4/15$, etc. Le signe de κ_3 indique si les fluctuations positives ou négatives (relativement à $\langle X\rangle$) sont favorisées. From [51].*

The fourth cumulant κ_4 shows whether fluctuations larger ($\kappa_4>0$) or smaller ($\kappa_4<0$) than Gaussian fluctuations are favoured. One defines the "*kurtosis*"

$$\text{kurtosis} = \kappa_4/\kappa_2^2\,. \tag{36}$$

c) Addition of random variables

✍ ***Exercice 15* – *Addition of independent random variables :*** (IMPORTANT) *Consider X and Y two independent random variables. Show that the cumulants of X and Y are additive.*

Hint: consider the generating function of the sum $S=X+Y$.

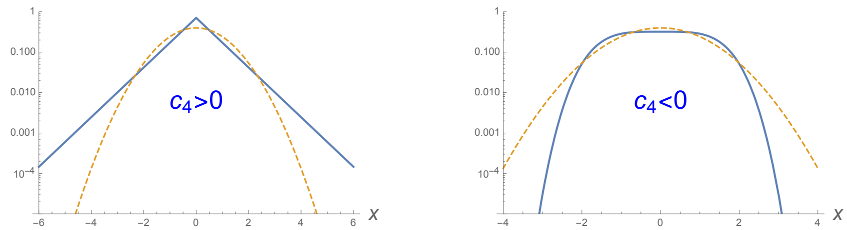


Figure 2: *La distribution* $g_a(x) = \frac{a\sqrt{\Gamma(3/a)}}{2\Gamma(1/a)^{3/2}} \exp\left(-\left|\sqrt{\frac{\Gamma(3/a)}{\Gamma(1/a)}}\,x\right|^a\right)$*, ici tracée pour* $a=1$ *(à gauche) et* $a=4$ *(à droite), a pour premiers cumulants (non nuls) :* $\kappa_2 = 1$, $\kappa_4 = \frac{\Gamma(1/a)\Gamma(5/a)}{\Gamma(3/a)^2} - 3$*, 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 [51].*

2.5 Multivariate Gaussian distribution

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

$$P(X) = \sqrt{\frac{\det A}{(2\pi)^N}}\, \mathrm{e}^{-\frac{1}{2}(X-X_0)^{\mathrm{T}} A (X-X_0)} \tag{37}$$

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

$$G(K) = \left\langle \mathrm{e}^{K^{\mathrm{T}} X}\right\rangle = \mathrm{e}^{K^{\mathrm{T}} X_0 + \frac{1}{2} K^{\mathrm{T}} A^{-1} K} \tag{38}$$

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

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

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

$$\langle x_i x_j\rangle_c = \frac{\partial^2}{\partial k_i \partial k_j} \ln G(K)\big|_{K=0} \tag{39}$$

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

$$\boxed{\langle x_i x_j\rangle_c = \left(A^{-1}\right)_{ij}} \tag{40}$$

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).

[5] with positive e.v.

✍ ***Exercice 16 – Discrete Ornstein-Uhlenbeck process :*** *We consider random Gaussian variables $(\cdots,\phi_n,\cdots)$ with probability weight $P(\phi)\propto\exp[-S]$ where the action is*

$$S=\frac{1}{2}\sum_{t\in\mathbb{Z}}\left[(\phi_{t+1}-\phi_t)^2+\mu^2\phi_t^2\right] \tag{41}$$

Write the action as $S=\frac{1}{2}\phi^{\mathrm{T}}A\phi$ 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 $\mu\to0$.

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

✍ ***Exercice 17 – "discrete Furutsu-Novikov theorem" :*** *We consider $f(X)$, a function of $X=(x_1,\cdots,x_N)^{\mathrm{T}}\in\mathbb{R}^N$. Show that for Gaussian random variables with $\langle x_i\rangle=0$ one has*

$$\langle x_i f(X)\rangle=\sum_j\langle x_ix_j\rangle\left\langle\frac{\partial f}{\partial x_j}\right\rangle\,. \tag{42}$$

✍ ***Exercice 18 – Wick theorem :*** *We consider N Gaussian random variables with distribution $P(X)\propto\mathrm{e}^{-\frac{1}{2}X^{\mathrm{T}}AX}$.*
a) Compute the four point correlation function $\langle x_ix_jx_kx_l\rangle$.
b) Generalize to the $2n$-point correlation function $\langle x_1x_2\cdots x_{2n}\rangle$.

✍ ***Exercice 19 – Gaussian random walk and the Wiener measure :*** *We consider N i.i.d. Gaussian random numbers of zero mean and unit variance, $\xi_1,\cdots,\xi_N$.*

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

We now consider the Gaussian random walk (RW)

$$x_n=x_{n-1}+\xi_n\qquad\textit{with } x_0=0\,. \tag{43}$$

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

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

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

$$P_N(x_N,\cdots,x_1)=\left(\prod_{i=2}^{N}P(x_i|x_{i-1})\right)P_1(x_1) \tag{44}$$

e) Wiener measure : What is the form of the measure $\mathrm{e}^{-\frac{1}{2}X^{\mathrm{T}}A_NX}$ in the continuum limit $x_t\to x(t)$?

2.6 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 $X_1,\cdots,X_N$. We denote their sum

$$S_N=\sum_{n=1}^{N}X_n\,. \tag{45}$$

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

$$\frac{S_N}{N} \underset{N\to\infty}{\longrightarrow} \langle X_i \rangle \tag{46}$$

- 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 $\mathrm{var}(S_N) = N\mathrm{var}(X_i)$.

a) Proof of the theorem

It is useful to recall the proof of the theorem. We denote by $p(x)$ the distribution of one variable X_i. The distribution of the sum is [6]

$$P_N(s) = \left\langle \delta\Big(s - \sum_{i=1}^{N} X_i\Big) \right\rangle \overset{\text{indep.}}{=} (\underbrace{p * \cdots * p}_{N\text{ times}})(s) = \int_{\mathbb{R}} \frac{\mathrm{d}k}{2\pi}\, \mathrm{e}^{\mathrm{i}ks}\, \hat{p}(k)^N \tag{47}$$

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 $N\to\infty$. For this reason we use the expansion in terms of the cumulants

$$\hat{p}(k) = \mathrm{e}^{W(-\mathrm{i}k)} = \mathrm{e}^{-\mathrm{i}\,\kappa_1\,k + \frac{1}{2}\kappa_2(-\mathrm{i}k)^2 + \mathcal{O}(k^3)} \tag{48}$$

This leads to

$$P_N(s) = \int_{\mathbb{R}} \frac{\mathrm{d}k}{2\pi}\, \mathrm{e}^{\mathrm{i}ks - \mathrm{i}\,N\kappa_1\,k - \frac{N}{2}\kappa_2 k^2 + \mathcal{O}(Nk^3)} \tag{49}$$

Then we rescale the argument of the distribution as

$$s = N\kappa_1 + \sqrt{N\kappa_2}\,y \tag{50}$$

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

$$P_N(N\kappa_1 + \sqrt{N\kappa_2}\,y) = \frac{1}{2\pi\sqrt{N\kappa_2}} \int_{\mathbb{R}} \mathrm{d}t\, \mathrm{e}^{\mathrm{i}ty - \frac{1}{2}t^2 + \mathcal{O}(N^{-1/2}t^3)} \tag{51}$$

Finally we use

$$\lim_{N\to\infty} \int_{\mathbb{R}} \mathrm{d}t\, \mathrm{e}^{\mathrm{i}ty - \frac{1}{2}t^2 + \mathcal{O}(N^{-1/2}t^3)} = \sqrt{2\pi}\, \mathrm{e}^{-\frac{1}{2}y^2} \tag{52}$$

leading to

$$\sqrt{N\kappa_2}\, P_N(N\kappa_1 + \sqrt{N\kappa_2}\,y) \underset{N\to\infty}{\simeq} \frac{1}{\sqrt{2\pi}}\, \mathrm{e}^{-\frac{1}{2}y^2} \tag{53}$$

which completes the proof.

b) Beyond the central limit theorem

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.
- 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 7) we will explain a method to study the distribution of $\sum_{t=1}^{N} \psi(x_t)$ when x_t is a stochastic process (like a Brownian motion).

[6] Start with $P_2(s) = \int \mathrm{d}x_1\, p(x_1)\, \mathrm{d}x_2\, p(x_2)\, \delta(s - x_1 - x_2) = \int \mathrm{d}x_1\, p(x_1)\, p(s - x_1) = (p * p)(s)$.

- The last point is the more subtle, as it is somehow hidden in the previous demonstration. The permutation of the limit $\lim_{N\to\infty}$ and the integral $\int_{\mathbb{R}} \mathrm{d}t$ is in fact only licit when y 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 determine the form of the tails of the distribution $P_N(s)$, for $s \to \infty$? To shed light on this subtle point, we will introduce the concept of *large deviations*.

> At this point it is important to check that one is comfortable with the steepest descent method : cf. § in the Appendix page 99.

2.7 Large deviations : typical and atypical fluctuations

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

$$W(k) = \ln\langle \mathrm{e}^{kX}\rangle \tag{54}$$

The distribution of S_N can be written as an inverse Laplace transform : [7]

$$P_N(s) = (\underbrace{p * \cdots * p}_{N\ \text{times}})(s) = \int_{\mathrm{i}\mathbb{R}} \frac{\mathrm{d}k}{2\mathrm{i}\pi}\, \mathrm{e}^{-ks+N\,W(k)}\ . \tag{57}$$

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

$$P_N(s = Ny) \underset{N\to\infty}{\simeq} \frac{1}{\sqrt{2\pi N W''(k_*)}} \mathrm{e}^{-N[k_* y - W(k_*)]} \tag{58}$$

where the saddle point is given by

$$W'(k_*) = y \tag{59}$$

(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

$$\boxed{P_N(s) \underset{N\to\infty}{\sim} \exp\left\{-N\,\Phi\left(\frac{s}{N}\right)\right\} \qquad \text{where} \qquad \Phi(y) = \max_k \left\{k\,y - W(k)\right\}} \tag{60}$$

---

[7] **Laplace transform :** a little reminder (with the standard convention). Consider a function $f(x)$ such that

$$\tilde{f}(p) = \int_{\mathbb{R}} \mathrm{d}x\, f(x)\, \mathrm{e}^{-px} \tag{55}$$

exists. Then the inverse Laplace transform involves an integral over the vertical axis in the complex plane, called the "*Bromwich contour* $\mathscr{B}$" :

$$f(x) = \int_{\mathscr{B}} \frac{\mathrm{d}p}{2\mathrm{i}\pi}\, \tilde{f}(p)\, \mathrm{e}^{+px} \tag{56}$$

$\mathscr{B}$ is the vertical line (parallel to imaginary axis) at the right of all singularities.

Example : consider $\tilde{f}(p) = 1/(p+a)$. Then the Bromwich contour is at the right of $p = -a$ in the complex plane of p (it can be the imaginary axis for $\mathrm{Re}(a) > 0$: $f(x) = \int_{-\mathrm{i}\infty}^{+\mathrm{i}\infty} \frac{\mathrm{d}p}{2\mathrm{i}\pi} \frac{\mathrm{e}^{px}}{p+a}$). Since $|\mathrm{e}^{+px}| = \mathrm{e}^{x\,\mathrm{Re}(p)}$, we can use the residue theorem and close the contour from the left for $x > 0$ and from the right for $x < 0$: this gives $f(x) = \theta_{\mathrm{H}}(x)\, \mathrm{e}^{-ax}$.

This result is known as the *Gärtner-Ellis theorem.*

Let us stress that the central limit theorem corresponds to a limiting case : consider the limit $y \to 0$ (i.e. consider that the argument of the distribution $P_N(s)$ is $s = N\,y \ll N$), we expect that $k_* \to 0$ 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 such $k_* \simeq (y-\kappa_1)/\kappa_2$ and we get

$$\Phi(y) \underset{y\to\kappa_1}{\simeq} \frac{(y-\kappa_1)^2}{2\kappa_2}\,, \tag{61}$$

i.e. the distribution has a Gaussian form

$$P_N(s) \underset{\delta s \ll N}{\sim} \exp\big[-(s-N\kappa_1)^2/(2N\kappa_2)\big] \qquad \text{(central limit theorem).} \tag{62}$$

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. $|\delta S_N| \ll \mathcal{O}(N)$. Above this scale, for $\delta S_N \gtrsim \mathcal{O}(N)$, the distribution deviates from the Gaussian form.

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 characterized by the central limit theorem and have a *universal Gaussian character.* However rare events for $\delta S_N \gtrsim \mathcal{O}(N)$ 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 $|s - N\langle X\rangle| \gtrsim \mathcal{O}(N)$ characterizes *atypical* fluctuations (rare events), which are *non universal.*

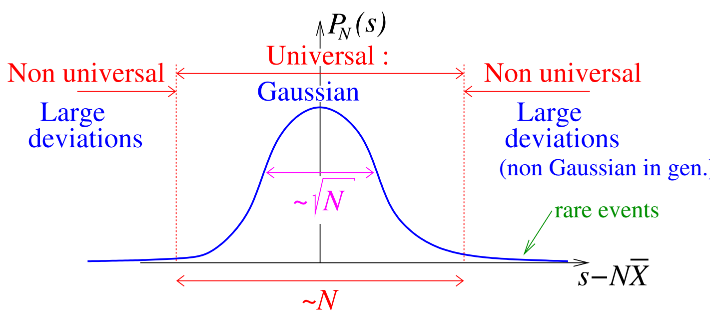


Figure 3: *The distribution of the sum of i.i.d. variables present a Gaussian universal central part of width $\sim\sqrt{N}$ (central limit theorem) and large deviation tails which are in general non Gaussian.*

✍ ***Exercice 20 – Large deviation function of the binomial distribution :*** *Using the Stirling formula, show that the binomial distribution (3) obeys the large deviation principle, i.e. has the form* $\text{Proba}\{H_N = n\} \underset{N\to\infty}{\sim} \exp\left\{-N\,\Phi\left(\frac{n}{N}\right)\right\}$. *Give the large deviation fuction* $\Phi(y)$.

Few remarks on the Legendre transform.— The Gärtner-Ellis 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

$$W(k) = \max_y \{k\,y - \Phi(y)\} \tag{63}$$

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

$$\mathrm{e}^{NW(k)} = \left\langle \mathrm{e}^{kS_N} \right\rangle = \int \mathrm{d}s\, P_N(s)\,\mathrm{e}^{ks} \sim N \int \mathrm{d}y\, \mathrm{e}^{N\left[ky-\Phi(y)\right]} \sim \mathrm{e}^{N\max_y\{k\,y-\Phi(y)\}} \tag{64}$$

Let us now simplify the notation and write the Legendre transforms as

$$\begin{cases} \Phi(y) = ky - W(k) & \text{where } y = W'(k) \\ W(k) = ky - \Phi(y) & \text{where } k = \Phi'(y) \end{cases} \tag{65}$$

It is convenient to write

$$\begin{cases} \Phi(W'(k)) = kW'(k) - W(k) \\ W(\Phi'(y)) = y\Phi'(y) - \Phi(y) \end{cases} \tag{66}$$

Setting $k = 0$ in the first equation gives $\Phi(\kappa_1) = -W(0) = 0$, which is the minimum of the function. 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

$$\begin{cases} \Phi''(W'(k)) = 1/W''(k) \\ W''(\Phi'(y)) = 1/\Phi''(y) \end{cases} \tag{67}$$

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

✍ ***Exercice 21** – **Legendre transform:** Consider the cumulant generating function with limiting behaviour*

$$W(k) \simeq \begin{cases} \kappa_1 k + (1/2)\kappa_2 k^2 & \text{for } k \to 0 \\ c\,|k|^{1+\alpha} & \text{for } k \to \pm\infty \end{cases} \tag{68}$$

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

✍ ***Exercice 22** – **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 : $\Omega(E) = \mathrm{e}^{S(E)}$ (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 $\widetilde{F}(\beta) = -\ln Z(\beta)$ and $S(E)$ are related through a Legendre transform. Argue that $S(E)$ is a large deviation function.

The role of the large deviations in statistical physics has been emphasized in a review article by Touchette [52].

✍ ***Exercice 23** – **:** Consider the symmetric distribution*

$$p(x) = \frac{1}{2}\,\mathrm{e}^{-|x|}\,. \tag{69}$$

Compute the generating function $W(k) = \ln\langle \mathrm{e}^{kX}\rangle$ and discuss where it is defined. Deduce the two first cumulants.
Get the large deviation function from the Legendre transform $\Phi(y) = \max_k \{k\,y - W(k)\}$.
Analyze the $k \to 0$ result to check your result (compare to the central limit theorem).

✍ ***Exercice 24** – **:** For $p(x) = \frac{1}{2}\,\mathrm{e}^{-|x|}$, the distribution of the sum can be expressed in terms of the MacDonald function*

$$P_{n+1}(s) = \frac{1}{2\sqrt{\pi}\,n!}\left|\frac{s}{2}\right|^{n+1/2} K_{n+\frac{1}{2}}(s) = \frac{\mathrm{e}^{-|s|}}{2^{2n+1}n!}\sum_{m=0}^{n}\frac{(n+m)!}{m!(n-m)!}(2|s|)^{n-m}\,. \tag{70}$$

Using the asymptotic of the MacDonald function (exercise 84), check the large deviation form obtained in the previous exercise.

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

The previous exercise reveals some interesting observation : for a distribution with purely exponential tail $p(x) \sim \mathrm{e}^{-|x|}$, the large deviation is also linear $\Phi(y) \sim |y|$ at large y, hence N in 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 $G(k) = \left\langle \mathrm{e}^{kX} \right\rangle$ 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 tail

$$p(x) \underset{x\to\infty}{\sim} \mathrm{e}^{-|x|^\alpha} \qquad \text{for } \alpha \in]0,1] \tag{71}$$

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

$$P_N(s) \underset{N\to\infty}{\sim} \exp\left\{-N^{\frac{\alpha}{2-\alpha}}\,\Phi\Big(N^{-\frac{1}{2-\alpha}}\,s\Big)\right\} \tag{72}$$

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

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

✍ ***Exercice 25* – :** *Consider i.i.d. random variables with distribution*

$$p(x) = \frac{1}{4\sqrt{|x|}}\,\mathrm{e}^{-\sqrt{|x|}} \tag{73}$$

a) Compute the moments μ_n.

b) Compute the convergence radius of the series $\sum_n \frac{\mu_n}{n!} k^n$.

c) Give the large deviation form (72).

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

- For $\alpha \geqslant 1$: the distribution of the sum presents the usual large deviation form $P_N(s) \sim \exp\{-N\,\Phi(s/N)\}$.
- For $0 < \alpha \leqslant 1$: the distribution of the sum presents the form (72), involving non trivial powers of N.

2.8 Generalization of the central limit theorem for power law distribution

The central limit theorem applies when the second moment is finite, $\left\langle X^2 \right\rangle < \infty$, meaning that the characteristic function behaves as $\hat{p}(k) = -\mathrm{i}\mu_1 k - \frac{1}{2}\mu_2 k^2 + o(k^2)$. When $\left\langle X^2 \right\rangle = \infty$ (or $\langle X \rangle = \infty$), we have seen that the characteristic function is non analytic for $k \to 0$ 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 with a large deviation function.

a) Case of symmetric distributions $p(x)$

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 $k \to 0$, all moments are finite, the central limit theorem holds.

(ii) $\hat{p}(k)$ is *not* analytic for $k \to 0$, however the second moment is finite, meaning that $\hat{p}(k) \simeq 1 - \frac{1}{2}\kappa_2 k^2 + \cdots + c\,|k|^\mu$ where $\mu > 2$ is non integer. The central limit theorem still holds.

(iii) $\hat{p}(k)$ is *not* analytic for $k \to 0$, and behaves as

$$\hat{p}(k) \simeq 1 - c\,|k|^\mu \text{ for } \mu \in]0,2[\tag{74}$$

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

We now examine further this last situation. We analyze $P_N(s) = \int_{\mathbb{R}} \frac{\mathrm{d}k}{2\pi}\,\mathrm{e}^{\mathrm{i}ks + N\ln\hat{p}(k)}$and proceed as for the central limit theorem. The behaviour (74) however suggests to introduce a different scaling variable. We now write

$$N^{1/\mu} P_N(s = N^{1/\mu} y) \underset{N\to\infty}{\simeq} N^{1/\mu} \int_{\mathbb{R}} \frac{\mathrm{d}k}{2\pi}\,\mathrm{e}^{N^{1/\mu}\mathrm{i}ky - Nc|k|^\mu} \simeq \int_{\mathbb{R}} \frac{\mathrm{d}q}{2\pi}\,\mathrm{e}^{\mathrm{i}qy - c|q|^\mu} \tag{75}$$

In other terms

$$P_N(s) \underset{N\to\infty}{\simeq} \frac{1}{(cN)^{1/\mu}}\,\mathcal{L}_{\mu,0}\left(\frac{s}{(cN)^{1/\mu}}\right) \quad \text{where} \quad \mathcal{L}_{\mu,0}(x) \stackrel{\text{def}}{=} \int_{\mathbb{R}} \frac{\mathrm{d}k}{2\pi}\,\mathrm{e}^{\mathrm{i}kx - |k|^\mu} \tag{76}$$

is a new universal function called a *Lévy law*. Clearly the distribution also presents the same power law tail as the initial distribution $p(x)$:

$$\mathcal{L}_{\mu,0}(x) \underset{x\to\infty}{\sim} |x|^{-1-\mu}\,. \tag{77}$$

(see exercise 12). Thus $P_N(s) \sim N/|s|^{1+\mu}$ for $s \to \infty$.

Eq. (76) 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

$$S_N \sim N^{1/\mu}\,. \tag{78}$$

(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, $\frac{S_N}{N} \sim N^{1/\mu - 1} \to 0$ for $N \to \infty$, 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 $S_N/N \sim N^{1/\mu-1} \to \infty$ for $N \to \infty$, 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 $N^{1/\mu} \gg N$.

To emphasize this point, I have performed some simulations (Fig. 4) : 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, the random walk seems continuous at large scale. For power law distribution with $\mu > 1$, the random walk exhibits some moderate jumps. For power law distribution with $\mu <$, the random walk is dominated by jumps, and the sum is clearly dominated by very few large values.

Figure 4: *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. 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* $\mu \leqslant 1$*. (the random numbers with power law distribution are generated thanks to the method of exercise 4).*

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 (large tails). For the new universality classes corresponding to Lévy laws, one obtains distributions characterized by power law tails, hence universality is stronger as it extends to large deviations.

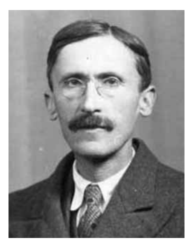


Figure 5: *Paul Lévy (1886-1971).*

b) Stable Lévy laws (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 $\hat{g}_{\mu,\sigma}(k) = \mathrm{e}^{-\mathrm{i}\mu k - \frac{1}{2}\sigma^2 k^2}$, 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]

$$\widehat{\mathcal{L}}_{\mu,\beta}(k) = \begin{cases} \mathrm{e}^{-|k|^\mu \left[1+\mathrm{i}\beta\,\mathrm{sign}(k)\tan(\pi\mu/2)\right]} & \text{for } \mu\in]0,2]\ \&\ \mu\neq 1 \\ \mathrm{e}^{-|k|\left[1-\frac{2\mathrm{i}\beta}{\pi}\,\mathrm{sign}(k)\ln|k|\right]} & \text{for } \mu=1 \end{cases} \tag{79}$$

The index μ controls the power of the tail $\mathcal{L}_{\mu,\beta}(x)\sim|x|^{-1-\mu}$ and the parameter $\beta\in[-1,+1]$ controls the asymmetry. For $\beta=0$, the distribution is symmetric. Few concrete remarks

- For $\beta=0$, there exist expansions for $x\to0$:

$$\mathcal{L}_{\mu,0}(x) = \frac{1}{\pi\mu}\sum_{n=0}^{\infty}\frac{(-1)^n\Gamma\big((2n+1)/\mu\big)}{(2n)!}x^{2n} \tag{80}$$

  and $x\to\infty$

$$\mathcal{L}_{\mu,0}(x) = \frac{1}{\pi}\sum_{n=1}^{\infty}(-1)^{n-1}\frac{\Gamma(1+n\mu)\sin(\pi\mu n/2)}{n!}\,|x|^{-1-\mu n} \tag{81}$$

- $\widehat{\mathcal{L}}_{1,0}(k)=\mathrm{e}^{-|k|}$ corresponds to the Cauchy distribution $\mathcal{L}_{1,0}(x)=\frac{1/\pi}{x^2+1}$.
- For $\beta\neq0$, the asymptotic of the Lévy distribution is (`wikipedia`)

$$\mathcal{L}_{\mu,\beta}(x) \underset{x\to\pm\infty}{\simeq} \frac{\sin(\frac{\pi\mu}{2})\,\Gamma(\mu+1)}{\pi}\,\frac{1+\beta\,\mathrm{sign}(x)}{|x|^{\mu+1}} \tag{82}$$

- For $\beta=+1$, the characteristic function is $\widehat{\mathcal{L}}_{\mu,1}(k)=\mathrm{e}^{-c_\mu(\mathrm{i}k)^\mu}$, where c_μ is a positive constant. For $\mu\in]0,1[$, the support of the distribution is $\mathbb{R}_+$. For $\mu\in]0,1[$ and $\beta=-1$, the support is $\mathbb{R}_-$.
- $\widehat{\mathcal{L}}_{1/2,1}(k)=\mathrm{e}^{-\sqrt{2\mathrm{i}k}}$ corresponds to the Lévy distribution mentioned above [8]

$$\mathcal{L}_{\frac12,1}(x) = \frac{1}{\sqrt{2\pi}x^{3/2}}\mathrm{e}^{-1/2x} \tag{83}$$

- 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 :

$$\mathcal{L}_{\frac32,1}(x) = \frac{\sqrt3}{\sqrt\pi|x|}\,\mathrm{e}^{\frac{x^3}{27}}\begin{cases} W_{\frac12,\frac16}\left(-\frac{2x^3}{27}\right) & \text{for } x\leqslant0 \\ \frac16 W_{-\frac12,\frac16}\left(\frac{2x^3}{27}\right) & \text{for } x\geqslant0\end{cases} \simeq \begin{cases}\frac13\sqrt{\frac{-2x}{\pi}}\,\mathrm{e}^{\frac{2x^3}{27}} & \text{for } x\to-\infty \\ \frac{3}{2\sqrt{2\pi}\,x^{5/2}} & \text{for } x\to+\infty\end{cases} \tag{84}$$

  where $W_{\mu,\nu}(z)$ is the Whittaker function.
- $\mathcal{L}_{2,\beta}(x)=\mathcal{L}_{2,0}(x)$ is the Gaussian distribution.

✍ ***Exercice 26 – :*** *Get the convergence radius of the two series representations (80) and (81).*

Remark : The Lévy law $\mathcal{L}_{2,\beta}(x)=\mathcal{L}_{2,0}(x)$ is the Gaussian distribution. A distribution with power law tail $|x|^{-3}$ 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. is of the type $\mathrm{e}^{-k^2(1-\ln|k|)}$, whose Fourier transform is Gaussian in the central part, with non Gaussian large deviation tails.

[8] To 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 $\int_0^\infty \mathrm{d}x\,\mathcal{L}_{1/2,1}(x)\,\mathrm{e}^{-\beta x}=\mathrm{e}^{-\sqrt{2\beta}}$.

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 [5] contains a nice appendix (I borrowed few remarks from it).

✍ ***Exercice 27*** – ***Marginal case*** $\mu = 1$ ***:*** *We discuss the distribution of the sum corresponding to the case* $\mu = 1$*.*

a) We assume that the random variables have a distribution given by the stable Lévy law $p(x) = \mathcal{L}_{1,\beta}(x)$*, with characteristic function (79). Show that the distribution of* S_N *is*

$$P_N(s) = \frac{1}{N} \mathcal{L}_{1,\beta}\left(\frac{s - \frac{2\beta}{\pi} N \ln N}{N}\right) \tag{85}$$

b) Consider now random variables with a distribution with tail $p(x) \sim x^{-2}$*. Discuss the distribution of the sum* $P_N(s)$*.*

✍ ***Exercice 28*** – ***Random trap model :*** *We consider a line with traps at regular positions. A particle 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 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. We denote by*

$$T = \sum_{\alpha=1}^{N} \tau_\alpha \tag{86}$$

the time after N *jumps (here it is fixed). The times are i.i.d. random variables with distribution* $\psi(\tau)$*.*
a) Assuming the power law tail $\psi(\tau) \sim \tau^{-1-\mu}$ *for* $\tau \to \infty$*, discuss how* T *scales with* N *depending on* μ*.*
b) Deduce the nature of the random walk.

2.9 Extreme value statistics

Study of extreme events has become very important these last decades, in climate (hurricane, typhoon,...), economy (market cracks,...), earthquakes, etc. In statistics, "ordered statistics" or "extreme value statistics" amounts to obtain the distribution of the largest among a collection of random variables (or the distribution of the second largest, etc). The case of independent identically distributed (i.i.d) random variables is well documented since the pioneering work of Fréchet [16] and Gumbel [22]. We will only focus on this case here (see books [23, 24, 10]). For a recent review discussing the case of correlated variables in the context of statistical physics, see Ref. [36].

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

$$M_N = \max(x_1, \cdots, x_N)\,?$$

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

$$\Phi_N(x) = \text{Proba}\{M_N \leqslant x\} = \text{Proba}\{x_1 \leqslant x \ \& \ \cdots \ \& \ x_N \leqslant x\} \tag{87}$$

$$= \int_{-\infty}^{x} \mathrm{d}x_1 \cdots \int_{-\infty}^{x} \mathrm{d}x_N\, \mathcal{P}_N(x_1, \cdots, x_N) \tag{88}$$

which can be extremely difficult to analyze for large N.

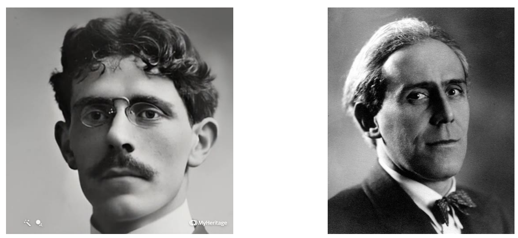


Figure 6: *Maurice Fréchet (1878-1973) and Emil Julius Gumbel (1891-1966).*

If now the random variables are **independent**, introducing the cumulative distribution of one random variable

$$F(x) = \int_{-\infty}^{x} \mathrm{d}y\, f(y) = \text{Proba}\{x_n \leqslant x\} \tag{89}$$

we have the simple form

$$\Phi_N(x) = [F(x)]^N \tag{90}$$

As we will see, the analysis of this distribution for $N \to \infty$ is interesting and reveals some universal character. For this, we introduce a_N, the expected largest value among the N variables :

$$F(a_N) = 1 - \frac{1}{N} \tag{91}$$

Let us give few examples in the following table :

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

For the power law tail, the scale of variation is given by a_N ($b_N \sim a_N$), hence there is only one relevant scale.

✍ ***Exercice 29 – :*** *Argue that* $a_N \simeq \left[\ln(N)\right]^{1/\alpha}$ *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 maximum scales as $M_N \sim \left[\ln(N)\right]^{1/\alpha}$, hence it cannot dominate the sum of N terms.
- 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 slower than N, hence S_N scales as N and the law of large numbers holds.
- 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. 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.

a) Gumbel class

We consider the case where $f(x)$ presents a decay of exponential type at large x. It is useful to introduce the scale b_N which is the typical scale of variation of $F(x)$ around a_N. For large x, $F(x)$ is close to one, hence the decay of $1 - F(x)$ is of exponential type and we can obtain the scale by writing

$$1/b_N = \frac{F'(a_N)}{1 - F(a_N)} = N\, f(a_N)\,. \tag{92}$$

This scale also measures the sensitivity of a_N as a function of N : indeed, differentiation of (91) with respect to N gives $f(a_N)\frac{\mathrm{d}a_N}{\mathrm{d}N} = 1/N^2$ i.e.

$$b_N = \frac{\mathrm{d}a_N}{\mathrm{d}\ln(N)}\,. \tag{93}$$

Examples are given in the table.

From the definition of b_N, we see that the behaviour of the cumulative distribution in the vicinity of a_N is

$$F(x) \underset{x\sim a_N}{\simeq} 1 - \frac{1}{N}\mathrm{e}^{-(x-a_N)/b_N} \tag{94}$$

Note that we don't assume that the distribution is exponential, however it decays *locally* exponentially (consider the example of the Gaussian distribution, cf. exercise 32). Introducing this form in the expression of the cumulative of the maximum, we have

$$\Phi_N(x) \underset{x\sim a_N}{\simeq} \left[1 - \frac{1}{N}\mathrm{e}^{-(x-a_N)/b_N}\right]^N \simeq \exp\left[\mathrm{e}^{-(x-a_N)/b_N}\right] \tag{95}$$

Hence, introducing the variable

$$y = (x - a_N)/b_N \tag{96}$$

we deduce the scaling form

$$\boxed{\Phi_N\left(x = a_N + b_N\, y\right) \underset{N\to\infty}{\simeq} \exp\left[-\mathrm{e}^{-y}\right] \equiv \Psi(y)} \tag{97}$$

in the vicinity of a_N, i.e. for $y \sim \mathcal{O}(1)$. This also shows that b_N measures the fluctuations of the maximum M_N.

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

$$\boxed{\psi(y) = \exp\left[-y - \mathrm{e}^{-y}\right]} \tag{98}$$

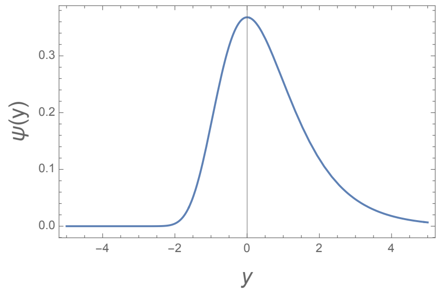


Figure 7: *Gumbel distribution* (98).

It reaches its maximum for $y = 0$ and is strongly asymmetric : for $y \to +\infty$ its decay is reminiscent of the exponential decay of the distribution $f(x)$, while for $y \to -\infty$ it is strongly suppressed (exponential of exponential). Note that this universal form has been obtained by keeping $y \sim \mathcal{O}(1)$, while sending $N \to \infty$, hence it only describes the distribution close to the typical value of the maximum, in the "central part". Going further away, large deviation tails are non universal (cf. exercise 32).

✍ ***Exercice 30 – Maximum for the exponential distribution :*** *Consider the distribution* $f(x) = \mathrm{e}^{-x}$ *for* $x > 0$. *Compute* $F(x)$ *and deduce explicitly* $\Phi_N(x)$. *Deduce* a_N *and* b_N

✍ ***Exercice 31 – Maximum for the Gaussian distribution :*** *Now consider a Gaussian distribution* $f(x)$*? i.e.* $F(x) = 1 - \frac{1}{2}\mathrm{erfc}(x/\sqrt{2})$. *Find* a_N *and* b_N.

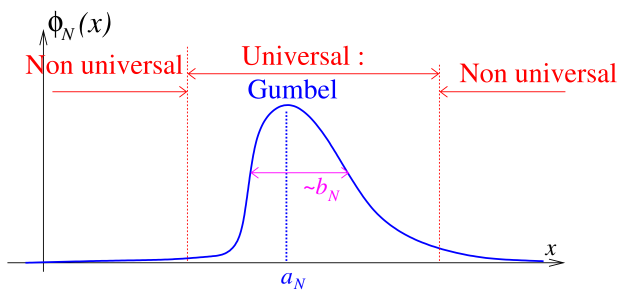


Figure 8: *Sketch of the distribution of the maximum in the Gumbel class. The Gumbel distribution describes the universal (central) part of the distribution of width* $\sim b_N$. *There exists also non universal tails [similar to large deviation tails of* $P_N(s)$*].*

✍ ***Exercice 32 – 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 maximmum under the form $\Phi_N(x) \simeq \exp\big[-\frac{N}{2}\,\mathrm{erfc}(x/\sqrt{2})\big]$.

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* $b_N \simeq 1/a_N$.

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 $k = 1, 2, \cdots$ (the maximum corresponding to $k = 1$). We have now to consider

$$\begin{aligned}\Phi_{N,k}(x) &= \mathrm{Proba}\{x_1 \leqslant x \ \& \ \cdots \ \& \ x_{N-k} \leqslant x \ \& \ x_{N-k+1} \geqslant x \ \& \ \cdots \ \& \ x_N \geqslant x\} \\ &= [F(x)]^{N-k+1}\,[1 - F(x)]^{k-1}\end{aligned} \tag{99}$$

i.e. $\Phi_N(x) \equiv \Phi_{N,1}(x)$ for the cumulative distribution of the largest value. Introducing again (94) we obtain the scaling form

$$\Phi_{N,k}\left(x = a_N + b_N\, y\right) \underset{N\to\infty}{\simeq} (\cdots) \exp\left[-(k-1)\, y - k\, \mathrm{e}^{-y}\right] \equiv \Psi_k(y) \tag{100}$$

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

$$\psi_k(y) = \frac{k^k}{(k-1)!} \exp\left[-k\, y - k\, \mathrm{e}^{-y}\right] \ . \tag{101}$$

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

b) Fréchet class

The second university class corresponds to distributions with power law tail $f(x) \simeq \mu\, A\, x^{-1-\mu}$ for $x \to \infty$. As shown in the table above, the typical position of the maximum among N variable is $a_N = (A\, N)^{1/\mu}$ and therefore we can rewrite $F(x) \simeq 1 - A\, x^{-\mu}$ for $x \to \infty$ as

$$F(x) \underset{x \gtrsim a_N}{\simeq} 1 - \frac{1}{N} \left(\frac{a_N}{x}\right)^{\mu} \tag{102}$$

Inserting this form in $\Phi_N(x) = \left[F(x)\right]^N$ we eventually obtain the scaling form

$$\Phi_N\left(x = a_N\, y\right) \underset{N\to\infty}{\simeq} \mathrm{e}^{-y^{-\mu}} \equiv \Psi(y) \tag{103}$$

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

$$\psi(y) = \frac{\mu}{y^{\mu+1}} \mathrm{e}^{-y^{-\mu}} \ . \tag{104}$$

I.e. we have obtained another universal function describing the distribution of the maximum for variable with power law distribution.

c) Weibull class

The last universality class corresponds to a distribution with a bounded domain of definition. For example, consider that $x_n > 0$ and we now ask the question of the distribution of the *minimum*. Assume that $f(x) \sim x^{\mu-1}$ for $x \to 0$ (with $\mu > 0$). The distribution of the *minimum* is given by the universal scaling function

$$\psi(y) = \mu\, y^{\mu-1}\, \mathrm{e}^{-y^{\mu}} \tag{105}$$

Note that the result can be deduced from the Fréchet one with $\mu \to -\mu$.

☺ Important points

- Basic concepts : random variable, moments, cumulants, generating functions, independence,...
- The case of divergent moments : power law $p(x) \sim |x|^{-1-\mu}$ for $x \to \infty$ is related to $\hat{p}(k) \simeq 1 - c\,|k|^{\mu}$ for $k \to 0$ when $\mu \in]0,1[$, $\hat{p}(k) \simeq 1 - \mathrm{i}k\,\overline{x} - c\,|k|^{\mu}$ when $\mu \in]1,2[$, etc.
- Gaussian random variables : relate correlations to the measure.
- Central limit theorem and its generalisation for power law distributions.
- Large deviations, large deviation function : typical and atypical fluctuations.
- Extreme value statistics.

3 The Langevin equation for a particle in a fluid

The aim of this introductory paragraph is to start the discussion of stochastic processes with a concrete and simple example : consider a particle in a fluid, submitted to a friction force. The usual phenomenological model is friction proportional to the velocity (Stokes regime),

$$F_f = -\gamma\, v\,, \tag{106}$$

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, $\eta \simeq 10^{-3}$ kg.m^{-1}.s^{-1} for water at $T = 20\,^{\circ}$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

$$\gamma = \frac{m}{\tau} \tag{107}$$

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

3.1 Fluctuations and Langevin force

In 1827, the scottish botanist Robert Brown observed with a microscope that pollen grains at the surface of watter move erratically. [9] 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 phenomenom was given much later by Albert Einstein in 1905 [14]. 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 $\tau_{\mathrm{coll}} \sim 10^{-15}$ s, thus we expect that the Brownian particle experiences collisions with the rate $1/\tau_{\mathrm{coll}}$ 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 a force $\xi(t)$ fluctuating in time, called the "**Langevin force**" :

$$m\frac{\mathrm{d}v(t)}{\mathrm{d}t} = -\gamma\, v(t) + \xi(t) \tag{108}$$

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = v(t) \tag{109}$$

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

$$\langle \xi(t) \rangle = 0 \tag{110}$$

where $\langle \cdots \rangle$ denotes *statistical averaging* [10] (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 $\langle \xi(t)\xi(t')\rangle = \frac{C}{\tau_{\mathrm{coll}}}\,\varphi((t-t')/\tau_{\mathrm{coll}})$ where φ is normalised function of width ~ 1 centered on the origin (like $(1/2)\mathrm{e}^{-|x|}$ or $\pi^{-1/2}\mathrm{e}^{-x^2}$) and C the strength of the fluctuations. As we are interested in the dynamics of the Brownian particle over time $\gg \tau_{\mathrm{coll}}$ we can simply consider

$$\big\langle \xi(t)\xi(t') \big\rangle = C\, \delta(t-t') \tag{111}$$

(which corresponds formally to $\tau_{\mathrm{coll}} \to 0$). A random function characterised by such local correlations is called a "*white noise*".

[9] You can find some historical perspectives in the excellent article of Bertrand Duplantier [13].

[10] Statistical averaging corresponds to average over different histories of the particle, with same initial conditions but in a different environments (different realisations of the Langevin force). This is the procedure followed by Jean Perrin in his experiments [44] ; cf. figure below.

Figure 9: *Robert Brown (1773-1858), Albert Einstein (1879-1955), Paul Langevin (1872-1946) and Jean Perrin (1870-1942).*

This model was introduced by Paul Langevin [30]. 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.

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

$$v(t) = v(0)\,\mathrm{e}^{-t/\tau} + \frac{1}{m}\int_0^t \mathrm{d}t'\,\xi(t')\,\mathrm{e}^{-(t-t')/\tau}\,. \tag{112}$$

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

$$\langle v(t)\rangle = v(0)\,\mathrm{e}^{-t/\tau}\,. \tag{113}$$

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 $\langle v(t)v(t')\rangle_c \stackrel{\text{def}}{=} \langle v(t)v(t')\rangle - \langle v(t)\rangle\,\langle v(t')\rangle$:

$$\big\langle v(t)v(t')\big\rangle_c = \frac{C\,\tau}{2m^2}\left(\mathrm{e}^{-|t-t'|/\tau} - \mathrm{e}^{-(t+t')/\tau}\right)\,. \tag{114}$$

The correlations decay in time over the same time scale τ.

✍ ***Exercice 33** – **Comparison between time and statistical averaging :*** *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 $N\to\infty$ and $T\to\infty$) the rate of occurence 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 $\widetilde{C}(t-t') = \overline{\xi(t)\xi(t')}^c = \overline{\xi(t)\xi(t')} - \overline{\xi(t)}\ \overline{\xi(t')}$? Compare to $C(t-t') = \langle\xi(t)\xi(t')\rangle_c = \langle\xi(t)\xi(t')\rangle - \langle\xi(t)\rangle\,\langle\xi(t')\rangle$.

3.2 A fluctuation-dissipation relation

We now introduce *another assumption* : after a sufficient long time, we expect that the particle is at thermal equilibrium, like the fluid, hence $\langle v(t)^2\rangle = k_\mathrm{B}T/m$ (equipartition theorem). This requires a constraint between the strength C of the Langevin noise (the fluctuations in the fluid), the friction coefficient γ and the temperature :

$$\underset{\substack{\downarrow \\ \text{fluctuations}}}{C} = 2 \overset{\substack{\text{dissipation} \\ \uparrow}}{\gamma} k_{\mathrm{B}} T \tag{115}$$

this is a first formulation of the **fluctuation-dissipation theorem** (FDT). We could write the correlator of the noise

$$\left\langle \xi(t)\xi(t') \right\rangle = 2\gamma \, k_{\mathrm{B}} T \, \delta(t-t') \,. \tag{116}$$

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). Below (§ e)), we will introduce a microscopic model of friction and try to clarify the origin of this relation.

✍ ***Exercice 34* – *Langevin equation for random initial velocity :*** *The correlator (114) 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 $\langle v(t)v(t') \rangle^{\mathrm{equil}}$.

b) Can we compare the two correlators ?

✍ ***Exercice 35* – *Measure of the Ornstein-Uhlenceck process :*** *In the stationary regime, compare the correlator with the one obtained in Exercise 16. Deduce what is the measure of the Ornstein-Uhlenbeck process.*

3.3 Diffusion

In the stationary regime the correlator of the speed is a "narrow function" of width τ, with weight

$$\int_{-\infty}^{+\infty} \mathrm{d}(t-t') \left\langle v(t)v(t') \right\rangle_c^{\mathrm{stat}} = \frac{2k_{\mathrm{B}}T}{\gamma} \,. \tag{117}$$

As a result, neglecting the transient regime at short times, we can write [11]

$$\left\langle x(t)^2 \right\rangle \simeq \int_0^t \mathrm{d}t_1 \int_0^t \mathrm{d}t_2 \, \langle v(t_1)v(t_2) \rangle_c^{\mathrm{stat}} \simeq t \int_{-\infty}^{+\infty} \mathrm{d}(t_1 - t_2) \, \langle v(t_1)v(t_2) \rangle_c^{\mathrm{stat}} = 2D\, t \tag{118}$$

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

$$\boxed{D \stackrel{\mathrm{def}}{=} \int_0^\infty \mathrm{d}t \, \langle v(t)v(0) \rangle} \tag{119}$$

For the present model, we get the expression

$$\boxed{D = \frac{k_{\mathrm{B}}T}{\gamma}} \tag{120}$$

which is known as the "*Einstein relation*", obtained in his 1905's article on Brownian motion [14]. 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.

[11] We recall the definition of the diffusion constant $D \stackrel{\mathrm{def}}{=} \lim_{t\to\infty} \frac{1}{2t} \left\langle x(t)^2 \right\rangle_c$.

✍ ***Exercice 36* – :** *Choose initial conditions for a fixed initial velocity $x(0) = 0$ and $v(0) = v_0$. Compute $\langle x(t) \rangle$. Then, 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.*

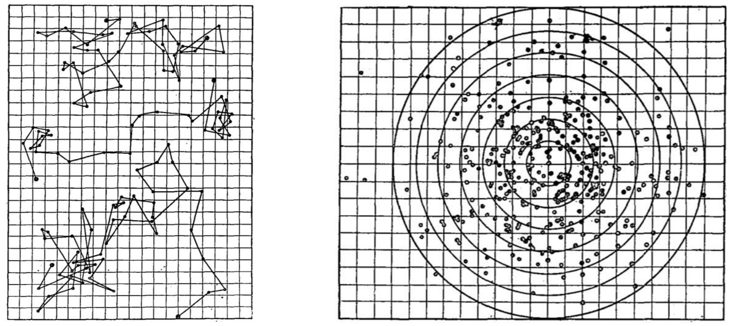


Figure 10: *Measurements of Jean Perrin (1908) ; from [44]. Left : few examples of trajectories. Right : final points after the several histories (for a fixed time).*

✍ ***Exercice 37* – 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 We assume that $x(0) = 0$ and that the particle is initially at equilibrium with the fluid. We apply the method proposed by Langevin in his famous article [30].*

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

b) Give an argument to justify $\langle x(t)\,\xi(t) \rangle = 0$. What is $\langle v(t)^2 \rangle$?

c) Argue that $\frac{\mathrm{d}}{\mathrm{d}t} \langle x(t)^2 \rangle \big|_{t=0} = 0$ and deduce

$$\left\langle x(t)^2 \right\rangle = \frac{2k_\mathrm{B}T}{\gamma} \left[t - \tau \left(1 - \mathrm{e}^{-t/\tau} \right) \right] \tag{121}$$

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

3.4 Large scale properties and the overdamped regime

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

$$\left\langle v(t)v(t') \right\rangle_c^{\mathrm{stat}} = \frac{k_\mathrm{B}T}{m}\,\mathrm{e}^{-|t-t'|/\tau} = \frac{C}{\gamma^2}\,\frac{\mathrm{e}^{-|t-t'|/\tau}}{2\tau} \overset{\text{large scale}}{\approx} \frac{C}{\gamma^2}\,\delta(t-t') = \frac{1}{\gamma^2}\,\left\langle \xi(t)\xi(t') \right\rangle . \tag{122}$$

This corresponds to write

$$v(t) \overset{\text{large scale}}{\approx} \frac{1}{\gamma}\,\xi(t) \tag{123}$$

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

$$0 \approx -\gamma v(t) + \xi(t) \qquad \text{(overdamped regime)} . \tag{124}$$

This approximation is called the "*overdamped regime*", which is achieved either by studying the process over large time scales, $\boxed{t \gg \tau}$, or by formally considering the limit of strong damping, $\gamma \to \infty$. As a result we obtain that the **velocity equals the force**. If an additional (conservative) force $F(x)$ is introduced in the equation of motion, we have

$$\boxed{\frac{\mathrm{d}x(t)}{\mathrm{d}t} \approx \frac{1}{\gamma}\,\left[F(x(t)) + \xi(t)\right]} \qquad \text{(overdamped regime)} . \tag{125}$$

This is similar to the pre-Galileo-Newtonian postulate, proposed by Aristote, which makes sense for the motion of a particle in a viscous fluid. We will come back later to a general analysis of this stochastic differential equation.

3.5 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

$$x(t) = \underbrace{x(0)}_{=0} + \frac{1}{\gamma}\int_0^t \mathrm{d}u\,\xi(u) \tag{126}$$

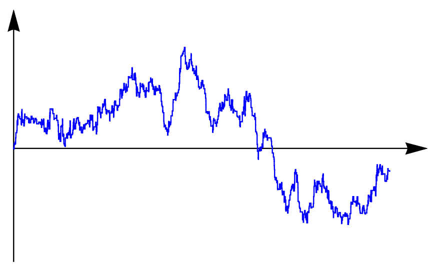


Figure 11: *A Brownian trajectory : $W(t)$ as a function of t.*

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

$$\langle \eta(t)\rangle = 0 \quad \text{and} \quad \big\langle \eta(t)\eta(t')\big\rangle = \delta(t-t')\,. \tag{127}$$

We now consider the normalised free Brownian motion (the "Wiener process")

$$W(t) = \int_0^t \mathrm{d}u\,\eta(u) \tag{128}$$

thus

$$\big\langle W(t)W(t')\big\rangle = \int_0^t \mathrm{d}u \int_0^{t'} \mathrm{d}v\,\delta(u-v) = \int_0^{\min(t,t')} \mathrm{d}u \tag{129}$$

Finally

$$\boxed{\big\langle W(t)W(t')\big\rangle = \min\big(t,t'\big)} \tag{130}$$

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

✍ ***Exercice 38* – :** *Check that the increment depends only on the time difference*

$$\big\langle \big[W(t) - W(t')\big]^2 \big\rangle = |t-t'| \tag{131}$$

In the limit $t \to t'$ we get $\lim_{t\to t'}\big\langle \big[W(t) - W(t')\big]^2 \big\rangle = 0$. One rewrites this property as

$$\underset{t'\to t}{\text{ms-lim}}\big[W(t) - W(t')\big] = 0 \tag{132}$$

where ms-lim is the "mean-square limit". This last equation implies that the curve $W(t)$ is continuous. Mathematicians have proven that this is true for all Brownian curves (atypical discontinuous curves do not exist). Another consequence of (131) is that

$$\left\langle \left(\frac{W(t) - W(t')}{t - t'} \right)^2 \right\rangle = \frac{1}{|t - t'|} \tag{133}$$

which goes to infinity when $t \to t'$, i.e. $W(t)$ is **non differentiable**. The curve is extremely irregular (see Fig. 11). This irregularity is related to *scale invariance* and fractal behaviour (with fractal dimension 1/2). For $\alpha > 0$, it is clear that

$$\langle W(\alpha t) W(\alpha t') \rangle = \alpha \min\left(t, t'\right) = \alpha \langle W(t) W(t') \rangle \tag{134}$$

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*"

$$\boxed{W(\alpha t) \overset{\text{(law)}}{=} \sqrt{\alpha}\, W(t)} \tag{135}$$

(their "laws" are equal). Because $\dot{W}(t) = \eta(t)$, we can also write the equation for the Gaussian white noise as

$$\eta(\alpha t) \overset{\text{(law)}}{=} \frac{1}{\sqrt{\alpha}}\, \eta(t)\,. \tag{136}$$

✍ ***Exercice 39 – From the Wiener process to the Ornstein-Uhlenbeck process :*** *We consider the Wiener process described by the equation* $\frac{\mathrm{d}W(u)}{\mathrm{d}u} = \eta(u)$*, where* $\eta(u)$ *is a normalised Gaussian white noise.*

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

$$\eta(\varphi(u)) \overset{\text{(law)}}{=} \frac{1}{\sqrt{|\varphi'(u)|}}\, \eta(u) \tag{137}$$

b) Deduce the stochastic differential equation for

$$x(t) = \frac{W(u)}{\sqrt{u}} \qquad \text{with} \qquad u = u_0 \mathrm{e}^{2\gamma t} \tag{138}$$

☺ Important points

- Master the analysis of the linear Langevin equation (108) (integrate, average, etc)
- Fluctuation-dissipation relation (different forms) : Langevin force and damping force have the same origin, hence the relation.
- Wiener process : main properties.

4 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 tools (vocabulary) allowing for a general analysis of stochastic processes.

Figure 12: *Marian von Smoluchowski (1872-1917) is considered as the father of the theory of stochastic processes.*

4.1 Generalities : joint probabilities, conditional probabilities

The aim of the section is to introduce some useful tools and concepts needed to describe random processes. In the previous section, we have obtained an integral representation of the trajectory in terms of the Langevin force, which has been used in order to analyze its statistical properties. In general, one considers a random process $X(t)$, i.e. a random function of the time, and one is interested in its statistical properties. Its probability should be given by a *functional* $P[X(t)]$, which is rather complicate to manipulate (for example an explicit calculation of an average might be difficult, e.g. $\langle X(t)\rangle = \int \mathcal{D}X(t)\, P[X(t)]\, X(t)$ where one should define how to perform the integral over the functions). For this reason we will introduce other tools more simple conceptually and practically.

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

$$\underbrace{P_n(x_n,t_n;\cdots;x_2,t_2;x_1,t_1)}_{\overleftarrow{\text{time}}} = \langle \delta(x_n - X(t_n))\cdots\delta(x_1 - X(t_1))\rangle \tag{139}$$

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

$$P_n(x_n,t_n;\cdots;x_2,t_2;x_1,t_1)\,\mathrm{d}x_1\cdots\mathrm{d}x_n = \mathrm{Proba}\{X(t_1)\in[x_1,x_1+\mathrm{d}x_1]\ \&\ \cdots\ \&\ X(t_n)\in[x_n,x_n+\mathrm{d}x_n]\}$$

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

$$\int \mathrm{d}x_k\, P_n(x_n,t_n;\cdots;x_{k+1},t_{k+1};\underline{x_k,t_k};x_{k-1},t_{k-1};\cdots;x_1,t_1) \tag{140}$$

$$= P_{n-1}(x_n,t_n;\cdots;x_{k+1},t_{k+1};x_{k-1},t_{k-1};\cdots;x_1,t_1) \tag{141}$$

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

$$P_{n|m}(x_n, t_n; \cdots; x_1, t_1 \,|\, y_m, \tau_m; \cdots; y_1, \tau_1) = \frac{P_{n+m}(x_n, t_n; \cdots; x_1, t_1; y_m, \tau_m; \cdots; y_1, \tau_1)}{P_m(y_m, t_m; \cdots; y_1, \tau_1)} \tag{142}$$

4.2 Markov processes

A very important class of random processes are *Markov processes.* A Markov process is a random process whose evolution only depends on its initial value, and not in his history before the initial time. Hence we can write

$$P_{n|m}(x_n, t_n; \cdots; x_1, t_1 \,|\, y_m, \tau_m; \cdots; y_1, \tau_1) = P_{n|1}(x_n, t_n; \cdots; x_1, t_1 \,|\, y_m, \tau_m) \tag{143}$$

which expresses that history prior to τ_m does not matter... **only the *last* position y_m at time τ_m determines the future evolution**.

Figure 13: *Andreï Andreïevich Markov (1856-1922).*

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

$$\begin{aligned} P_3(x_3, t_3; x_2, t_2; x_1, t_1) &= P_{1|2}(x_3, t_3|x_2, t_2; x_1, t_1)\, P_2(x_2, t_2; x_1, t_1) && (144)\\ &= P_{1|2}(x_3, t_3|x_2, t_2; x_1, t_1)\, P_{1|1}(x_2, t_2|x_1, t_1)\, P_1(x_1, t_1) && (145)\\ &\overset{\text{Markov}}{=} P_{1|1}(x_3, t_3|x_2, t_2)\, P_{1|1}(x_2, t_2|x_1, t_1)\, P_1(x_1, t_1) && (146)\end{aligned}$$

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

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

Chapman-Kolmogorov equation : Start from the general property

$$\int \mathrm{d}x_2\, P_3(x_3, t_3; x_2, t_2; x_1, t_1) = P_2(x_3, t_3; x_1, t_1)\,. \tag{147}$$

For a Markov process, using (146), one gets the Chapman-Kolmogorov equation

$$\boxed{\int \mathrm{d}x_2\, P(x_3, t_3|x_2, t_2)\, P(x_2, t_2|x_1, t_1) = P(x_3, t_3|x_1, t_1)} \tag{148}$$

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.

Now multiply this equation by $P(x_1, t_1)$ and integrate over x_1. We get (x_3, $t_3 \to x_f$, t_f and x_2, $t_2 \to x_i$, t_i)

$$P(x_f, t_f) = \int \mathrm{d}x_i \, P(x_f, t_f | x_i, t_i) \, P(x_i, t_i) \tag{149}$$

Which shows that the conditional probability relates the distribution at initial time t_i to the distribution at final time t_f. For this reason, $P(x_f, t_f | x_i, t_i)$ is sometimes called the "*propagator*".

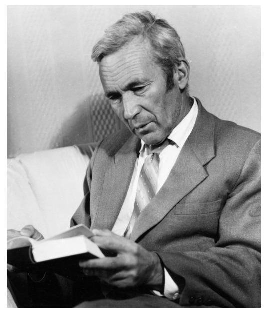


Figure 14: *Andreï Nikolaïevitch Kolmogorov (1903-1987), well-known by physicists for his major contributions to the theory of dynamical systems and probability.*

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

$$P(x_2, t_2 | x_1, t_1) = P(x_2, t_2 - t_1 | x_1, 0) \tag{150}$$

Such random processes are denoted "*homogeneous*". I will sometimes denote the propagator as $P_t(x|x_0)$.

An example of Markov process : we can come back to the Langevin equation (108) for the velocity. This equation is first order and involves a white noise (ξ uncorrelated in time), hence the evolution is fully determined by $v(0) = v_0$ and the process is Markovian. Assuming furthermore that the Langevin noise is Gaussian, [12] we can get easily the two fundamental probabilities characterizing the process. From the above calculations we have, cf. (113,114)

$$\langle v(t) \rangle = v_0 \, \mathrm{e}^{-t/\tau} \tag{151}$$

$$\mathrm{Var}[v(t)] = \frac{k_\mathrm{B}T}{m} \left(1 - \mathrm{e}^{-2t/\tau}\right) \tag{152}$$

The process $v(t)$ is a convolution of the Gaussian Langevin force $\xi(t)$, hence it is also Gaussian. The knowledge of these two moments is sufficient to characterize the full distribution, which is here conditioned by the initial velocity :

$$P_t(v|v_0) = \sqrt{\frac{m}{2\pi k_\mathrm{B}T \, \left(1 - \mathrm{e}^{-2t/\tau}\right)}} \, \exp\left\{ -\frac{m \, (v - v_0 \, \mathrm{e}^{-t/\tau})^2}{2k_\mathrm{B}T \, \left(1 - \mathrm{e}^{-2t/\tau}\right)} \right\} . \tag{153}$$

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

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

$$P_t(v|v_0) \underset{t\to\infty}{\longrightarrow} P(v) = \sqrt{\frac{m}{2\pi k_\mathrm{B}T}}\,\mathrm{e}^{-\frac{m}{2k_\mathrm{B}T}\,v^2} \tag{154}$$

(the 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 (108), or the conditional probability (153), is known as the **Ornstein-Uhlenbeck process**. It is the subject of the Doob theorem (the only homogeneous Gaussian stationary random process is the Ornstein-Uhlenbeck process). [13]

✍ ***Exercice 40*** *– : Recover the correlator $\langle v(t)v(t')\rangle_c$ given by (114) from the conditional probability. Consider both cases of initially fixed velocity and random velocity.*

Let us now discuss two instructive examples of stochastic processes, Markovian and non Markovian. Let us recall that the Markovian nature of the process defined by the Langevin equation (108) originates from the two reasons : (i) the differential equation is first order, hence the solution depends only on some initial value $v(0)$; (ii) the noise is δ-correlated, hence there is no memory.

A non Markovian process : if we consider now the position of the particle described by the Langevin equation (108,109), the differential equation is $m\,\ddot{x} = -\gamma\,\dot{x} + \xi(t)$: the noise is still δ-correlated, however the differential equation is now second order, hence the solution depends on both the initial position $x(0)$ *and* the initial velocity $\dot{x}(0)$, which depends on the history before $t=0$: the process $x(t)$ is not Markovian.

We could also argue that the position obeys a first order stochastic differential equation, $\dot{x} = v(t)$ where $v(t)$ is a "noise", however this latter is characterized by a finite correlation time (memory time), $\langle v(t)v(t')\rangle = \frac{k_\mathrm{B}T}{m}\,\mathrm{e}^{-|t-t'|/\tau}$. Hence $x(t)$ is non Markovian because the noise is not δ-correlated (from this point of view, one says that the SDE for $x(t)$ involves a "colored noise").

These are two different points of view to assert that the process $x(t)$ is *not* Markovian.

A 2D Markov process : $x(t)$ is a non Markovian process, however it can be considered as the first component of a two-dimensional Markovian process $\vec{\psi}(t) = (x(t), v(t))$: the system of differential equations is first order, and can be written under the more general form of a multidimensional Langevin equation

$$\dot{\psi}_i = \Phi_i(\vec{\psi}) + B_{ij}\,\Xi_j(t) \tag{155}$$

where $\vec{\Phi} = (v, F(x) - \gamma\,v)$ is the drift (we have added a conservative force). The noise $\vec{\Xi}(t) = (0, \xi(t))$ is uncorrelated in time and 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" $\partial_t P_t = \big[-\partial_x v - \partial_v(F(x) - \gamma\,v) + \gamma k_\mathrm{B}T\,\partial_v^2\big]P_t$.

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.

[13] The Ornstein-Uhlenbeck process is more frequently introduced as a model for a particle attached to a spring in the overdamped regime : $\dot{x}(t) = -\kappa\,x(t) + \xi(t)$.

4.3 Master equation

Continuous processes.— Let us start with the case of continuous processes, which is more general. As we have seen, Eq. (149), the evolution of the distribution of a Markov process can be represented in terms of the conditional probability which plays the role of a "*propagator*"

$$P(x,t) = \int \mathrm{d}x_0 \, P(x,t|x_0,t_0) \, P(x_0,t_0) \, . \tag{156}$$

However, this equation and (148) 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 : such an evolution equation can be related to the above integral equation by considering the evolution during an infinitesimal time $\delta t \to 0$. In this case we expect

$$P(x,t+\delta t|x_0,t) \simeq \delta(x-x_0) + \delta t \, W_t(x|x_0) + \mathcal{O}(\delta t^2) \tag{157}$$

The linear correction follows from the Markov assumption : at short time, the transition probability is linear with time and involves a transition rate $W_t(x|x_0)\mathrm{d}x$ for performing the transition from x_0 to $[x, x+\mathrm{d}x]$.

In the following, **we will restrict ourselves to homogeneous processes** (time translation invariant) such that

$$W_t(x|x_0) \to W(x|x_0) \qquad \text{(homogeneous process)} \tag{158}$$

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

$$\boxed{\frac{\partial P(x,t)}{\partial t} = \int \mathrm{d}x' \, W(x|x') \, P(x',t)} \tag{159}$$

Note that the conservation of probability requires that

$$\int \mathrm{d}x_f \, W(x_f|x_i) = 0 \qquad \Rightarrow \qquad \frac{\partial}{\partial t} \int \mathrm{d}x \, P(x,t) = 0 \ \ \forall \, t \tag{160}$$

so that probability $\int \mathrm{d}x \, P(x;t) = 1$ is conserved. This condition follows from the normalization condition of the conditional probability, in the expansion (157). Obviously, the conditional probability obeys the same equation

$$\frac{\partial P_t(x|x_0)}{\partial t} = \int \mathrm{d}x' \, W(x|x') \, P_t(x'|x_0) \quad \text{for initial condition } P_0(x|x_0) = \delta(x-x_0) \, . \tag{161}$$

In a specific problem, the transition "rates" $W(x|x_0)$ are given and the aim is to solve the master equation (159), or (161).

In the most general case, a Markov process can combine

- a diffusion : in this case the integral kernel is replaced by a second order differential operator
- jumps : leading to an integral term in the master equation, like in (159).

Below, we will give concrete examples.

Discrete processes.— For simplicity, let us first consider a random process which takes discrete values $X(t) \in \{x_1, \cdots, x_{\mathcal{M}}\}$ and denote $P_n(t) = \text{Proba}\{X(t) = x_n\}$. The Markovian nature of the process implies that $P_n(t+\delta t)$ depends on the state of the process at time t, hence it can be related to the probability $P_n(t)$, i.e. the distribution obeys a **first order** differential equation

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}t} P_n(t) = \sum_m W_{n,m} P_m(t)} \tag{162}$$

where the transition rates $W_{n,m}$ form a $\mathcal{M} \times \mathcal{M}$ matrix (with $W_{n,m} \geqslant 0$ for $n \neq m$), and satisfy

$$\sum_n W_{n,m} = 0 \tag{163}$$

ensuring the conservation of probability $\sum_n P_n(t) = 1 \; \forall\, t$ (hence $W_{n,n} = -\sum_{m\,(\neq n)} W_{m,n}$ is the only negative matrix element). The evolution equation (162) is known as the *master equation.* Note that by using $W_{n,n} = -\sum_{m\,(\neq n)} W_{m,n}$ we can rewrite the master equation as

$$\frac{\mathrm{d}}{\mathrm{d}t} P_n(t) = \sum_{m\,(\neq n)} \left[W_{n,m} P_m(t) - W_{m,n} P_n(t) \right] \tag{164}$$

(we can further replace $\sum_{m\,(\neq n)} \to \sum_n$). This form avoids to add the restriction (163).

Birth and death processes.— A subclass of these 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 eqaution has the form

$$\frac{\mathrm{d}}{\mathrm{d}t} P_n(t) = d_{n+1}\, P_{n+1}(t) + b_{n-1}\, P_{n-1}(t) - (d_n + b_n)\, P_n(t) \tag{165}$$

where $d_n > 0$ and $b_n > 0$ are death and birth rates, respectively. A simple example is the Poisson process studied below.

We discuss below several examples for various Markov processes.

✍ ***Exercice 41** – **Random telegraph process :*** *We 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). The transition rates are λ_1 (from X_1 to X_2) and λ_2 (from X_2 to X_1). We denote by $P_i(t) = \text{Proba}\{X(t) = X_i\}$ with $i \in \{1, 2\}$.*

a) Write the set of differential equations for $P_1(t)$ and $P_2(t)$. Deduce a matricial form $\frac{\mathrm{d}}{\mathrm{d}t} P(t) = W\, P(t)$, where $P = (P_1 \; P_2)^{\mathrm{T}}$ is the column vector ($^{\mathrm{T}}$ denotes tranposition).

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

c) Determine the conditional probability $P_t(i|j)$. 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 $X_2 = 1$. Compute $C(t-t') = \langle X(t) X(t') \rangle - \langle X(t) \rangle \langle X(t') \rangle$.

e) Deduce the power spectrum $S(\omega)$ of the telegraphic noise (use the Wiener-Khintchine theorem and the relation with the correlation function $C(t)$).

a) Example : the Poisson process (statistics of uncorrelated events)

The Poisson process takes integer values $\mathscr{N}(t) \in \mathbb{N}$ with $\mathscr{N}(0) = 0$. With probability rate λ, the process is incremented by one, i.e. during an interval of time of duration $\mathrm{d}t$, the process increases by one with probability $\lambda\, \mathrm{d}t$. We denote $P_n(t) = \text{Proba}\{\mathscr{N}(t) = n\}$ its probability.

The Poisson process (PP) counts the occurences of *independent events*. For instance the number of drops of rain falling on the floor during a time interval t. Or the number of desintegrations in a radioactive material.

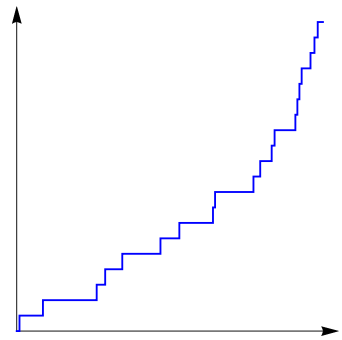


Figure 15: *An instance of Poisson process : $\mathscr{N}(t)$ as a function of t.*

✍ ***Exercice 42 – Master equation for the PP :***
a) Show that the master equation for the Poisson process is

$$\frac{\mathrm{d}}{\mathrm{d}t}P_n(t) = \lambda\, P_{n-1}(t) - \lambda\, P_n(t) \tag{166}$$

for $n > 0$ (and $\frac{\mathrm{d}P_0(t)}{\mathrm{d}t} = -\lambda\, P_0(t)$). In other terms, the rate "matrix" has elements on the diagonal and just below the diagonal $W_{n,m} = \lambda\big(-\delta_{m,n} + \delta_{m,n-1}\big)$.

b) 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.

c) Deduce that

$$P_n(t) = \frac{(\lambda t)^n}{n!}\,\mathrm{e}^{-\lambda t} \tag{167}$$

d) Determine the moments the cumulants $\langle \mathscr{N}(t)^k\rangle_c$ of the Poisson process.

e) Give the distribution $q(\tau)$ of the time separating two successive events [indication : relate $q(\tau)$ and $P_0(t)$].

Note that (167) corresponds to the initial condition $\mathscr{N}(0) = 0$, hence the conditional probability of the Poisson process is $P_t(n|m) = P_{n-m}(t)$ for $n \geqslant m$ and $P_t(n|m) = 0$ for $n < m$. Here we have used translation invairance in "space".

✍ ***Exercice 43 – :*** *Consider the Poisson process with $\mathscr{N}(0) = 0$.*
Check that $\sum_{n_0} P_t(n|n_0)\, P_{n_0}(t_0) = P_n(t + t_0)$.

✍ ***Exercice 44 – Derivative of the Poisson process :*** *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 $\xi(t) = \mathscr{N}'(t)$.*
We introduce the generating function of the noise $G[h] \stackrel{\text{def}}{=} \big\langle \exp\int \mathrm{d}t\, h(t)\, \xi(t)\big\rangle$, where $\langle\bullet\rangle$ is the averaging over the random times t_n's.

a) Show that $G[h] = \exp\big\{\lambda\int \mathrm{d}t(\mathrm{e}^{h(t)} - 1)\big\}$.
Hint: Consider that the N times are not ordered, distributed over $[0,T]^N$ with measure $\mathrm{d}t_1\cdots\mathrm{d}t_N/T^N$.

b) 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, $\xi(t) = \mathscr{N}'(t)$ *is* ***a non Gaussian white noise****.*

b) Another example : the compound Poisson process

A natural generalization of the Poisson process is the *compound Poisson process* (CPP) [14] : we consider now that the process $X(t)$ makes random jumps

$$X(t_n^+) = X(t_n^-) + \eta_n \tag{168}$$

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. After a time t, the number of jumps $\mathscr{N}(t)$ is random (it is a PP). The CPP can be written in terms of this PP as

$$X(t) = \sum_{n=0}^{\mathscr{N}(t)} \eta_n \tag{169}$$

with $\eta_0 = X(0) = 0$.

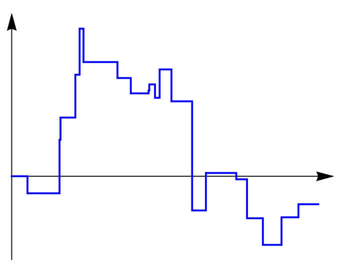


Figure 16: *Compound Poisson process* $X(t)$ *for Gaussian jumps.*

✍ ***Exercice 45 – Master equation for the CPP :***

a) Show that the master equation for the CPP is

$$\frac{\partial P(x,t)}{\partial t} = \lambda \int \mathrm{d}\eta\, w(\eta)\, \left[P(x-\eta,t) - P(x,t)\right] \tag{170}$$

i.e. of the form (159) for $W(x|x_0) = \lambda\left[w(x-x_0) - \delta(x-x_0)\right]$*. Note that here, the transition rate is translation invariant.*

b) Continuum limit.— Study the limit $\lambda \to \infty$ *with* $w \to 0$ *such that* $a = \lambda\langle\eta_n\rangle$ *and* $b = \lambda\langle\eta_n^2\rangle$ *are kept finite (argue that, in this limit,* $\lambda\langle\eta_n^k\rangle \to 0$ *for* $k > 2$*).*

c) Introducing the Fourier transforms $\widehat{P}(k,t) = \int \mathrm{d}x\, \mathrm{e}^{-\mathrm{i}kx}\, P(x;t)$ *and* $\hat{w}(k) = \int \mathrm{d}\eta\, \mathrm{e}^{-\mathrm{i}k\eta}\, w(\eta)$*, show that the solution is*

$$P(x,t) = \int_{-\infty}^{+\infty} \frac{\mathrm{d}k}{2\pi}\, \mathrm{e}^{\lambda t\,[\hat{w}(k)-1]+\mathrm{i}kx}\,. \tag{171}$$

Discuss the continuum limit.

d) When $\langle\eta_n^2\rangle = \infty$*, the process belongs to the class of* ***Lévy flights****. For example, if* $w(\eta) \sim c/\eta^2$ *for* $\eta \to \pm\infty$ *we have* $\hat{w}(k) \simeq 1 - c|k|$ *for* $k \to 0$*. Deduce* $P(x,t)$ *over large scales. Discuss also the more general case where* $\hat{w}(k) \simeq 1 - c|k|^\mu$ *for* $k \to 0$*, with* $\mu \in]0,2[$*.*

[14] in French: "processus de Poisson composé".

✍ ***Exercice 46 – Derivative of the CPP - a non Gaussian white noise :***
a) 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 $\langle \xi(t_1)\cdots\xi(t_n)\rangle_c$.
Hint: follow the same steps as in exercise 44.

b) Show that the noise becomes a Gaussian white noise in a certain limit.

4.4 Markov chains

An important class of random processes are Markov chains, which are homogeneous random 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 random process $X(t) \in \{1, \cdots, \mathcal{M}\}$ and denote $P_n(t) = \text{Proba}\{X(t) = n\}$. The master equation (162) introduced above involves transitions at random times. For Markov chain, the jumps occur at regular discrete times, thus the master equation takes the form

$$\boxed{P_n(t+1) = \sum_m M_{nm}\,P_m(t)} \tag{172}$$

where

$$M_{nm} = \text{Proba}\{m \to n\} \in [0,1] \tag{173}$$

is the $\mathcal{M} \times \mathcal{M}$ matrix of transition probabilities. M is called a "*stochastic matrix*". It satisfies

$$\boxed{\sum_n M_{nm} = 1} \tag{174}$$

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

$$M_{nm} = p\,\delta_{m,n-1} + q\,\delta_{m,n+1} \tag{175}$$

(with $p + q = 1$).

For the following, it is useful to rewrite the master equation (172) in a form closer to the differential equation (164) by using (174)

$$P_i(t+1) - P_i(t) = \sum_{j(\neq i)} \left[M_{ij}\,P_j(t) - M_{ji}\,P_i(t)\right] \tag{176}$$

✍ ***Exercice 47 – Continuum limit of the Markov chain :*** *Consider a Markov chain with jumps occuring every δt. Argue that the master equation (162) is recovered by considering the continuum limit*

$$M_{ij} = \delta_{ij} + \delta t\,W_{ij} \qquad \text{with } \delta t \to 0\,. \tag{177}$$

If M is a stochastic matrix, what is the constraint on the matrix W ?

b) The Perron-Fröbenius theorem and the stationary state

We can interpret the condition (174) as the existence of a left eigenvector $L^{(0)} = (1, \cdots, 1)^{\mathrm{T}}$ for eigenvalue $\lambda_0 = 1$:

$$L^{(0)\mathrm{T}} M = L^{(0)\mathrm{T}} \qquad \text{or} \qquad M^{\mathrm{T}} L^{(0)} = L^{(0)} \,, \tag{178}$$

where $(\cdot)^{\mathrm{T}}$ denotes transposition (the vectors are column vectors). The *Perron-Fröbenius theorem* states that (i) $\lambda_0 = 1$ is non-degenerate, (ii) it is the largest eigenvalue, (iii) the related right eigenvector

$$M\, R^{(0)} = R^{(0)} \,, \tag{179}$$

has positive components. For a finite number of states $\mathcal{M}$, this corresponds to the **stationary solution**, $R^{(0)} = (P_1^*, \cdots, P_{\mathcal{M}}^*)^{\mathrm{T}}$. Normalization condition reads $L^{(0)\mathrm{T}} R^{(0)} = 1$ (the scalar product is the product of the line and the column vectors). We can rewrite equation (179) as

$$\sum_{j(\neq i)} \left[M_{ij}\, P_j^* - M_{ji}\, P_i^* \right] = 0 \,. \tag{180}$$

c) Classification of Markov processes

We now discuss the different scenarii which might occur. We keep considering the case of Markov chains, although a similar discussion could be developed within the master equation.

(i) **Equilibrium.—** Often, the existence of a stationary solution is ensured by a condition *stronger* than (180), called the **detailed balance condition**

$$M_{ij}\, P_j^* - M_{ji}\, P_i^* = 0 \qquad \text{(detailed balance)} \tag{181}$$

If detailed balance is fulfilled, one says that P_i^* is an *equilibrium state*. We can also conveniently relate the ratio of rates to the ratio of probabilities

$$\boxed{\frac{M_{ij}}{M_{ji}} = \frac{P_i^*}{P_j^*}} \qquad \text{(detailed balance} \equiv \text{equilibrium)} \tag{182}$$

The relation should hold $\forall (i,j)$. This is *a probabilistic definition of equilibrium.*

(ii) **NESS (non-equilibrium steady state).—** If the detailed balance condition (181) is *not* fulfilled but the condition

$$\sum_{j(\neq i)} \big[\underbrace{M_{ij}\, P_j^* - M_{ji}\, P_i^*}_{\neq 0} \big] = 0 \qquad \text{(stationarity)} \tag{183}$$

holds, one says that the stationary state is a *non-equilibrium steady state.* Such states are characterised by the existence of non zero *probability* fluxes.

(iii) **Transient process.—** When $\mathcal{M} \to \infty$ it is possible that the eigenvector $(\cdots, P_i^*, \cdots)^{\mathrm{T}}$ is not normalisable, so that there is no stationary state. One says that the process is *transient.*

Depending on the matrix M_{ij} which defines the Markov chain, one encounters one of the three situations.

d) Spectral decomposition - Relaxation

The stochastic matrix M, with positive matrix elements, is not symmetric in general, $M^{\mathrm{T}} \neq M$. We have seen above that its eigenvalue $\lambda_0 = 1$ is associated with a couple of left and right eigenvectors $L^{(0)}$ and $R^{(0)}$. If M is diagonalisable, its eigenvalues $\lambda_n < 1$ are associated with a biorthogonal set of left and right eigenvectors $L^{(n)}$ and $R^{(n)}$. We can choose the orthonormalisation condition as $L^{(n)\mathrm{T}} R^{(m)} = \delta_{n,m}$, which leads to the spectral representation

$$M = \sum_n \lambda_n R^{(n)} L^{(n)\mathrm{T}} \,. \tag{184}$$

This is useful in order to solve the master equation (172). Denoting by $P(0) = (P_1(0), \cdots, P_{\mathcal{M}}(0))^{\mathrm{T}}$ the initial conditions, we can write

$$P(t) = M^t \, P(0) \qquad \text{i.e.} \qquad P_n(t) = \sum_j \lambda_j^t \, R_n^{(j)} \, \underbrace{L^{(j)\mathrm{T}} P(0)}_{c_j \overset{\mathrm{def}}{=}} \tag{185}$$

c_j is the coefficient of the initial vector on the basis of eigenvectors $P(0) = \sum_j c_j R^{(j)}$. Note that $c_0 = L^{(0)\mathrm{T}} P(0) = \sum_j P_j(0) = 1$ carries all the normalisation.

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 (172) is the conditional probability

$$\boxed{P_t(n|m) = \left(M^t\right)_{nm}} \tag{186}$$

Now let us discuss the large time behaviour. Using that $\lambda_0 = 1 > \lambda_1 > \lambda_2 > \cdots$, the large time behaviour takes the form

$$P_n(t) \underset{t\to\infty}{\simeq} \underbrace{P_n^*}_{\equiv R_n^{(0)}} + \underbrace{c_1 \, \lambda_1^t R_n^{(1)}}_{\underset{t\to\infty}{\longrightarrow} 0} \tag{187}$$

where we have used $c_0 = 1$ (normalisation). This shows that $1/\tau_{\mathrm{relax}} = -\ln \lambda_1$ 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 $\lambda_0 = 1$, the eigenvalues are not real in general , however 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

$$\boxed{\frac{1}{\tau_{\mathrm{relax}}} = -\ln |\lambda_1|} \tag{188}$$

Remark 2 : When M is non symmetric, it is not always diagonalisable. However it can always be decomposed in terms of Jordan blocks.

✍ ***Exercice 48* – :** *In the same spirit, solve Eq. (162). If probability is conserved, what is expected for the eigenvalues of W_{nm} ?*

✍ ***Exercice 49* – *Biased random walk on a ring* :** *Consider the random walk on a ring with L sites, such that with $M_{nm} = p\,\delta_{n,m+1} + q\,\delta_{n,m-1}$ for $n, m \in \{1, \cdots, L\}$. Periodic boundary conditions are $M_{1L} = p$ and $M_{L1} = q$.*

a) Argue that the stationary state is an equilibrium state when $p = q = 1/2$ and a NESS for

$p \neq q$.

b) 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. $|\lambda_k| \leqslant 1 \; \forall \, k$.

c) Decompose the conditional probability $P_t(n|m)$ over the eigenvalues and the eigenvectors.

d) Consider the limit $L \to \infty$ and discuss the bottom of the spectrum. Compute $P_t(n|m)$ in the two limiting cases $v = 0$ and $v = \pm 1$.

e) 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 $P_n^* \propto \mathrm{e}^{-\beta\varepsilon_n}$. 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

$$\frac{\Gamma_{n\leftarrow m}}{\Gamma_{m\leftarrow n}} = \mathrm{e}^{-\beta(\varepsilon_n - \varepsilon_m)} \tag{189}$$

i.e. emission is more probable than absorption (the difference is spontaneous emission).

✍ ***Exercice 50** – **Master equation: the three scenarii :*** *Let us consider the master equation describing the one dimensional diffusion on $\mathbb{Z}$ with transitions between nearest neighbour sites*

$$\partial_t P_n(t) = W_{n,n-1} P_{n-1}(t) + W_{n,n+1} P_{n+1}(t) - (W_{n-1,n} + W_{n+1,n})\, P_n(t) \tag{190}$$

i.e. $W_{n,m}$ is a tridiagonal (infinite) matrix with $W_{n,n} = -W_{n-1,n} - W_{n+1,n}$. Hence, this is an example of birth and death process.

a) Current : check that the master equation can be rewritten under the form

$$\partial_t P_n = -J_n + J_{n-1} \tag{191}$$

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

$$W_{n,m} = \mathrm{e}^{[V(m) - V(n)]/2} \tag{192}$$

where $V(x)$ is a known function.

b) Equilibrium ($J = 0$).— Show that

$$P_n^* = C\, \mathrm{e}^{-V(n)} \tag{193}$$

is a stationary solution corresponding to a vanishing probability current. Discuss the normalisability.

c) NESS ($J \neq 0$).— Find the stationary solution corresponding to $J_n = J \; \forall \, n$. Show that it is

$$P_n^* = J\, \mathrm{e}^{-V(n)} \sum_{m=n}^{\infty} \mathrm{e}^{[V(m+1) + V(m)]/2} \tag{194}$$

Discuss the normalisability (consider the continuum limit for simplicity).

d) Provide an example where there is no stationary state.

f) Detailed balance, reversibility and ergodicity

Let us consider a Markov chain with master equation (172), such that detailed balance is fulfilled. We denote by P_i^* the equilibrium solution. Define

$$\mathcal{D}_t \stackrel{\text{def}}{=} \sum_i \frac{(P_i(t) - P_i^*)^2}{P_i^*} = \sum_i \frac{P_i(t)^2}{P_i^*} - 1 \geqslant 0 \tag{195}$$

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 (182) leads to

$$\Delta\mathcal{D}_t = -\frac{1}{2} \sum_{i,j,k} M_{ji} M_{ki} P_i^* \left(\frac{P_j(t)}{P_j^*} - \frac{P_k(t)}{P_k^*} \right)^2 \leqslant 0 \tag{196}$$

Conclusion :

- $\mathcal{D}_t \geqslant 0$
- $\Delta\mathcal{D}_t \leqslant 0$
- We conclude that $\mathcal{D}_t \searrow$ and thus $P_i(t) \to P_i^*$.

This shows that detailed balance ensures that the system reaches equilibrium. This remark is borrowed from [33].

g) 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 $P_\ell \propto \mathrm{e}^{-\beta E_\ell}$. If the number of states N_{state} is too large, it might be difficult to compute numerically the sum

$$\langle \mathcal{O} \rangle_{\text{eq}} = \sum_{\ell=1}^{N_{\text{state}}} P_\ell \, \mathcal{O}_\ell \tag{197}$$

For example, if one considers N Ising spins, the sum runs over $N_{\text{state}} = 2^N$ microstates, which becomes rapidly untracktable if N is large (a square of 10×10 spins 1/2 has $N_{\text{state}} \sim 10^{30}$ microstates).

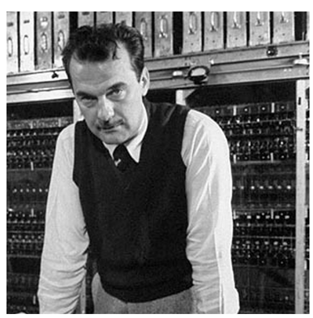


Figure 17: *Nicholas Metropolis (1915-1999).*

The central idea of equilibrium statistical physics is to replace the study of the microscopic (deterministic) dynamics by a probabilistic description. The Monte Carlo method replaces the probabilistic description by a stochastic dynamics defined as follows : if the system is in state $|\,i\,\rangle$ at time t, a move to another state $|\,f\,\rangle$ chosen randomly is made with probability

$$\text{Proba}\{i \to f\} \equiv M_{fi} = \min\left(1,\, \mathrm{e}^{-\beta(E_f - E_i)}\right) . \tag{198}$$

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 time step. Assuming $E_f > E_i$, the stochastic matrix has the form

$$M = \begin{pmatrix} \ddots & & & \\ & 1 - \mathrm{e}^{-\beta(E_f - E_i)} & 1 & \\ & \mathrm{e}^{-\beta(E_f - E_i)} & 0 & \\ & & & \ddots \end{pmatrix} \begin{matrix} \\ \leftarrow \ \ |\,i\,\rangle \\ \leftarrow \ \ |\,f\,\rangle \\ \\ \end{matrix} \tag{199}$$

All other diagonal matrix elements are equal to one and all other non diagonal matrix elements equal to zero. This is the Metropolis algorithm (from the name of the inventor of the method, Nicholas Metropolis). Because $M_{fi}/M_{if} = \mathrm{e}^{-\beta(E_f - E_i)}$, such dynamics converges towards the Gibbs equilibrium. Finally the statistical average is replaced by the time average over the stochastic dynamics involving N_{step}

$$\overline{\mathcal{O}(t)} = \frac{1}{N_{\text{step}}} \sum_{t=1}^{N_{\text{step}}} \mathcal{O}(t)\,. \tag{200}$$

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 [29] for a detailed discussion).

Going from the microscopic dynamics to the Monte Carlo method, the scheme is the following :

Classical mechanics		Equilib. statistical physics		Monte Carlo method
deterministic evolution	$\longrightarrow$	probabilistic description	$\longrightarrow$	stochastic dynamics

4.5 Beyond the master equation : renewal processes

I discuss here another type of stochastic processes, known as *Markov renewal processes*, which generalize the jump processes described by the master equation (159), or its discrete version (162). Consider the master equation (293) given below, which does not assume any constraint on the kernel $W(x|x')$. This last form makes clear that the master equation describes a situation where the process in the state x' performs a jump with rate $\lambda = \int \mathrm{d}y\, W(y|x')$. Hence it stays at its initial position during a random time τ distributed with an exponential law $q(\tau) = \lambda \mathrm{e}^{-\lambda\tau}$ (this was discussed in exercise 42). A specific case where the transition rates are translation invariant $W(x|x') = \lambda\, w(x - x')$ corresponds to the compound Poisson process studied above (exercise 45).

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

$$X(t_n^+) = X(t_n^-) + \eta_n\,. \tag{201}$$

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.).

✍ ***Exercice 51 – Continuous time random walks (CTRW) :*** *For simplicity, we assume in the following that w is a symmetric function.*

a) Justify that the master equation is replaced by the integral equation (in time)

$$P(x,t) = \int_0^t \mathrm{d}\tau\, q(\tau) \int_{\mathbb{R}} \mathrm{d}\eta\, w(\eta)\, P(x-\eta, t-\tau) + \delta(x) \int_t^\infty \mathrm{d}\tau\, q(\tau) \tag{202}$$

Check normalisation

b) If $q(\tau) = \lambda \mathrm{e}^{-\lambda\tau}$, check that one recovers (170) (compound Poisson process).

c) Solve the equation by introducing the Laplace-Fourier transform

$$\widetilde{P}(k,s) \stackrel{\text{def}}{=} \int_0^\infty \mathrm{d}t\, \mathrm{e}^{-st} \int_{\mathbb{R}} \mathrm{d}x\, \mathrm{e}^{-\mathrm{i}kx}\, P(x,t) \tag{203}$$

Deduce $\widetilde{P}(k,s)$ in terms of $\tilde{q}(s) = \int_0^\infty \mathrm{d}\tau\, \mathrm{e}^{-s\tau} q(\tau)$ and $\hat{w}(k) = \int_{\mathbb{R}} \mathrm{d}\eta\, \mathrm{e}^{-\mathrm{i}k\eta}\, w(\eta)$. Express $P(x,t)$ under an integral form.

d) Consider distributions with power law tails $w(\eta) \simeq \frac{c}{|\eta|^{\mu+1}}$ for $\eta \to \pm\infty$ and $q(\tau) \simeq \frac{a}{\tau^{\alpha+1}}$ for $\tau \to +\infty$.
What is the $s \to 0$ 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) Discuss the limiting behaviour of $\widetilde{P}(k,s)$ for $k \to 0$ and $s \to 0$. 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 $\mu = 2\alpha \in]0, 2[$.

4.6 Spectral analysis of stochastic processes – Wiener-Khintchine theorem

Convention for Fourier transform in time : We define the Fourier transform in time as

$$\widetilde{C}(\omega) = \int_{-\infty}^{+\infty} \mathrm{d}t\, C(t)\, \mathrm{e}^{\mathrm{i}\omega t} \quad \text{et} \quad C(t) = \int_{-\infty}^{+\infty} \frac{\mathrm{d}\omega}{2\pi}\, \widetilde{C}(\omega)\, \mathrm{e}^{-\mathrm{i}\omega t}\,. \tag{204}$$

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 [15]

$$\tilde{x}_n = \int_0^T \frac{\mathrm{d}t}{T}\, x(t)\, \mathrm{e}^{+\mathrm{i}\omega_n t} \quad \text{et} \quad x(t) = \sum_n \tilde{x}_n \mathrm{e}^{-\mathrm{i}\omega_n t} \text{ où } \omega_n = \frac{2n\pi}{T} \text{ avec } n \in \mathbb{Z}\,. \tag{205}$$

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 $[\omega, \omega + \Delta\omega]$:

$$\mathcal{S}(\omega) \stackrel{\text{def}}{=} \frac{1}{\Delta\omega} \sum_{\omega_n \in [\omega, \omega+\Delta\omega]} \langle |\tilde{x}_n|^2 \rangle \tag{206}$$

This is precisely the outcome of the device represented in figure 18 : sample → ampli/filter → multiplicator → measurement.

[15] Later, we will define the Fourier transform in space as $f_q = \int_V \mathrm{d}r\, f(r)\, \mathrm{e}^{-\mathrm{i}qr}$, where q is quantized if the volume is finite, and $f(r) = \frac{1}{V} \sum_q f_q\, \mathrm{e}^{+\mathrm{i}qr} \to \int \frac{\mathrm{d}q}{(2\pi)^d} f_q\, \mathrm{e}^{+\mathrm{i}qr}$.

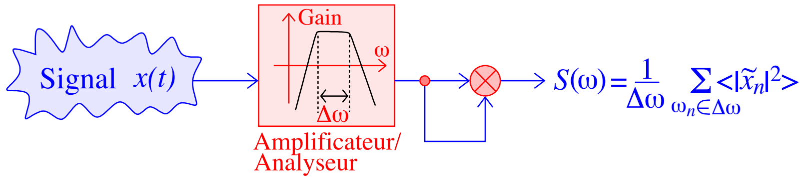


Figure 18: *Measure of the noise : the signal is amplified, duplicated and multiplied by itself. The result is averaged over a long time T.*

Figure 19: *Norbert Wiener (1894-1964) & Aleksandr Yakovlevich Khinchin (1894-1959).*

Wiener-Khintchine theorem.– From the above hypothesis, one can verify that : [16]

$$\langle \tilde{x}_n \tilde{x}_m^* \rangle = \frac{1}{T}\delta_{n,m}\widetilde{C}_{xx}(\omega_n) \tag{207}$$

where $\widetilde{C}_{xx}(\omega) = \int_{-\infty}^{+\infty} \mathrm{d}\tau\, C_{xx}(\tau)\, \mathrm{e}^{\mathrm{i}\omega\tau}$. Only components corresponding to opposite frequencies ω_n and ω_{-n} are correlated [17]. Thus one has : $\sum_{\omega_n\in[\omega,\omega+\Delta\omega]}\langle|\tilde{x}_n|^2\rangle = \frac{\mathcal{N}_{\Delta\omega}}{T}\widetilde{C}_{xx}(\omega)$ where $\mathcal{N}_{\Delta\omega} = \Delta\omega T/2\pi$ is the number of frequencies ω_n in the bandwidth. Finally one gets

$$\boxed{\mathcal{S}(\omega) = \frac{\widetilde{C}_{xx}(\omega)}{2\pi}} \tag{208}$$

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).

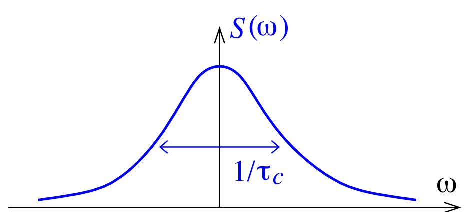


Figure 20: **Wiener-Khintchine theorem :** *width of noise spectrum is inversely proportional to the correlation time of the process.*

[16] Write $\langle \tilde{x}_n \tilde{x}_m^* \rangle = \int_0^T \frac{\mathrm{d}t}{T}\int_0^T \frac{\mathrm{d}t'}{T}\mathrm{e}^{\mathrm{i}\omega_n t-\mathrm{i}\omega_m t'}C_{xx}(t-t') = \frac{1}{T}\int_0^T\frac{\mathrm{d}t'}{T}\mathrm{e}^{\mathrm{i}(\omega_n-\mathrm{i}\omega_m)t'}\int_{-t'}^{T-t'}\mathrm{d}(t-t')\mathrm{e}^{\mathrm{i}\omega_n(t-t')}C_{xx}(t-t')$. Short range correlation allows to write $\int_{-t'}^{T-t'}\mathrm{d}(t-t')\cdots \to \int_{-\infty}^{+\infty}\mathrm{d}(t-t')\cdots$.

[17] One can as well consider a process defined on $\mathbb{R}$ by writing $T\to\infty$. The Fourier transform is then defined as $\tilde{x}(\omega) = \int \mathrm{d}t\, x(t)\mathrm{e}^{\mathrm{i}\omega t}$ and one can show that $\langle \tilde{x}(\omega)\tilde{x}(\omega')\rangle = 2\pi\delta(\omega+\omega')\widetilde{C}_{xx}(\omega)$. Correspondence between the two formulations is ensured by the substitutions $\tilde{x}(\omega) \leftrightarrow T\tilde{x}_n$ and $2\pi\delta(\omega-\omega') \leftrightarrow T\delta_{n,n'}$.

✍ ***Exercice 52 – :*** *As a simple application of the Wiener-Khintchine theorem, we analyze the correlation of the velocity for the process defined by the phenomenological Langevin equation*

$$\frac{\mathrm{d}v(t)}{\mathrm{d}t} = -\int \mathrm{d}t'\,\gamma(t-t')\,v(t') + \xi(t) \tag{209}$$

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

$$C_{vv}(\tau) = \int_{-\infty}^{+\infty} \mathrm{d}\omega\, \underbrace{\frac{\widetilde{C}_{\xi\xi}(\omega)}{2\pi}}_{\mathcal{S}_{\text{Force}}(\omega)} \frac{\mathrm{e}^{-\mathrm{i}\omega\tau}}{|\tilde{\gamma}(\omega)-\mathrm{i}\omega|^2} \tag{210}$$

Consider the limit $C_{\xi\xi}(\tau) = 2D\gamma^2\delta(\tau)$ et $\gamma(t) = \gamma\,\delta(t)$ and compute explicitly the correlator.

✍ ***Exercice 53 – :*** *Consider now the case of a Langevin force correlated over the finite time τ_c (a microscopic time) : $C_{FF}(t) = 2D\gamma^2\frac{1}{2\tau_c}\mathrm{e}^{-|t|/\tau_c}$. We expect the function $\gamma(t)$ to be of finite width τ_m ; Assume $\gamma(t) = \gamma\,\theta(t)\frac{1}{\tau_m}\mathrm{e}^{-t/\tau_m}$. The three time scales fulfill : $\tau_c \lesssim \tau_m \ll 1/\gamma$. Analyze the residus of $|\tilde{\gamma}(\omega)-\mathrm{i}\omega|^{-2}$ justify that one can consider $\tau_m \to 0$ while keeping a finite τ_c, what simplifies the evaluation of the integral. Show then that $C_{vv}(\tau) = \frac{D\gamma}{1-(\gamma\tau_c)^2}[\mathrm{e}^{-\gamma|t|} - \gamma\tau_c\mathrm{e}^{-|t|/\tau_c}]$. 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-Khintchine theorem : relation between the correlation function of an homogeneous process and its noise spectrum.

5 Stochastic differential equations

In § 3, we have discussed a specific case of stochastic differential equation (SDE), the Langevin equation $m\frac{\mathrm{d}}{\mathrm{d}t}v(t) = -\gamma\, v(t) + \xi(t)$ involving a δ-correlated Langevin force. We took advantage of the linearity to obtain an integral representation of the solution, which makes easy the analysis of the statistical properties of the solution. The aim of this paragraph is to consider a more general situation and consider SDE of the form $\frac{\mathrm{d}}{\mathrm{d}t}x = F(x) + \sqrt{2D(x)}\,\eta(t)$, where $\eta(t)$ is a normalised Gaussian white noise.

SDE are particularly well suited for numerical simulations (it is easy to generate many realizations of such processes). Here, the aim is to introduce some tools allowing for a statistical analysis of the solution. Finally, let us stress that by considering that $\eta(t)$ is a Gaussian white noise, in this chapter we restrict ourselves to the study of **continuous Markov processes** (with no jump). [18]

5.1 SDE with drift and additive noise

Let us come back to the analysis of the stochastic process described by Eq. (125). We write

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = F(x(t)) + \sqrt{2D}\,\eta(t) \tag{211}$$

where $\eta(t)$ is a normalised Gaussian white noise with

$$\langle \eta(t) \rangle = 0 \qquad \text{and} \qquad \big\langle \eta(t)\eta(t') \big\rangle = \delta(t-t')\,. \tag{212}$$

An analysis similar to the one of Section 3 is not possible (unless $F(x) \propto x$) due to the nonlinear character of the equation. Being interested in statistical properties of the solution, it is natural to consider its distribution, or at least to build an equation for it, the Fokker-Planck equation. Below we show that the corresponding FPE is

$$\frac{\partial P_t(x)}{\partial t} = -\frac{\partial}{\partial x}\left[F(x)\,P_t(x)\right] + D\,\frac{\partial^2}{\partial x^2}P_t(x) \tag{213}$$

where $P_t(x)$ is the distribution of $x(t)$. In the next section, we will furhter discuss how to solve this equation.

Proof : We introduce the Wiener process

$$W(t) = \int_0^t \mathrm{d}t'\,\eta(t')\,. \tag{214}$$

Introduce the increment $\delta W(t) = W(t+\delta t) - W(t)$. The most important observation is the independence of the increments and the property

$$\big\langle \delta W(t)^2 \big\rangle = \delta t \tag{215}$$

see above, Eq. (131). We now may write

$$\delta x(t) \stackrel{\text{def}}{=} x(t+\delta t) - x(t) \simeq F(x(t))\,\delta t + \sqrt{2D}\,\delta W(t) \tag{216}$$

[18] Remember the end of § 3 : We showed that the Wiener process $W(t) = \int_0^t \mathrm{d}u\,\eta(u)$ is continuous but not differentiable. The solution $x(t)$ of the SDE $\mathrm{d}x = F(x)\,\mathrm{d}t + \sqrt{2D(x)}\,\mathrm{d}W(t)$ is also continuous but not differentiable.
Note that we could also write a differential equation for a Markov process with jumps by considering that $\eta(t)$ is the derivative of a Poisson process or a Compound Poisson process, i.e. of the form $\eta(t) = \sum_n \eta_n\,\delta(t-t_n)$. The analysis would be more complicated because the equation for $P_t(x)$ then involves an *integral* operator instead of a *differential* operator like in the Fokker-Planck equation (see for example (170) describing the simple case of the CPP).

Consider a test function $\varphi(x)$. We study the evolution of $\langle\varphi(x(t))\rangle$.

$$\langle\varphi(x(t+\delta t))\rangle - \langle\varphi(x(t))\rangle$$
$$= \left\langle \varphi'(x)\,\left[F(x)\,\delta t + \sqrt{2D}\,\delta W\right] + \frac{1}{2}\varphi''(x)\,\left[F(x)\,\delta t + \sqrt{2D}\,\delta W\right]^2 + \cdots \right\rangle \tag{217}$$
$$= \left\langle \varphi'(x(t))\,F(x(t))\right\rangle \delta t + \sqrt{2D}\,\left\langle \varphi'(x(t))\,\delta W(t)\right\rangle + D\left\langle \varphi''(x(t))\right\rangle \delta t + \cdots \tag{218}$$

where we have kept terms $\mathcal{O}(\delta t)$. Because $x(t)$ is only correlated with the increment $\delta W(t')$ for $t > t'$, we see that $x(t)$ and $\delta W(t)$ are uncorrelated, thus $\langle\varphi'(x(t))\,\delta W(t)\rangle = \langle\varphi'(x(t))\rangle\,\langle\delta W(t)\rangle = 0$. Finally

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\varphi(x(t))\rangle = \left\langle \varphi'(x(t))\,F(x(t))\right\rangle + D\left\langle \varphi''(x(t))\right\rangle \tag{219}$$

the expansion was performed until second order as the second order term in δW gives some first order contribution in δt, due to (215). We can now rewrite the equation in terms of the distribution

$$\frac{\partial}{\partial t}\int \mathrm{d}x\, P_t(x)\,\varphi(x) = \int \mathrm{d}x\, P_t(x)\,\left[\varphi'(x)\,F(x) + D\,\varphi''(x)\right] \tag{220}$$
$$= \int \mathrm{d}x\,\varphi(x)\,\left(-\frac{\partial}{\partial x}\left[F(x)\,P_t(x)\right] + D\,\frac{\partial^2 P_t(x)}{\partial x^2}\right) \tag{221}$$

Because the equation is valid $\forall\,\varphi$, we can remove the integral, hence (213). Qed.

5.2 SDE with multiplicative noise : Itô or Stratonovich ?

The SDE (211) is not the most general form of stochastic differential equation as it corresponds to the case where the diffusion constant is uniform in space. The aim of the paragraph is to discuss the case of SDE of the form

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = a(x(t)) + b(x(t))\,\eta(t) \qquad \textit{(not well defined!)} \tag{222}$$

where $b(x) = \sqrt{2D(x)}$ can be related to a x-dependent diffusion constant. As we explain now, this form is however not well defined.

The noise is here multiplied by a function of the process : one says that the **noise is multiplicative**, whereas it is said additive in SDE (211). For a mutiplicative noise, the differential equation (222) is ambiguous and it is not fully defined. This is not surprising : if $\eta(t)$ is a Gaussian white noise, $x(t)$ has the same regularity as the Brownian motion, i.e. is continuous but not differentiable. The existence of a difficulty comes from the fact that we manipulate a differential equation involving objects which are not differentiable in the sense of functions !

a) Discretization (numerics)

A first approach could be to discretize time (this is natural for numerical implementation of the stochastic differential equation). We could write

$$x_{t+1} = x_t + a(x_t)\,\delta t + b(x_t)\,\delta W_t \tag{223}$$

where $x(t) = x_t$ is measured every time step δt and $\delta W_t = W(t+\delta t) - W(t)$ is a Gaussian random variable for $\left\langle \delta W_t^2\right\rangle = \delta t$ (all δW_t are i.i.d.). This is perfectly fine, however it turns out that in the limit $\delta t \to 0$, we do not recover the usual rules of differential calculus for regular functions, as we will see below. This is not necessarily a problem, however this definitely deserves clarification.

b) Origin of the ambiguity

To clarify this point, we come back to the continuous description and consider a slightly different type of noise, made of δ-peaks at random times

$$\eta(t) = \sum_n \delta(t - t_n) \tag{224}$$

If the times occur with a finite rate λ, the noise $\eta(t)$ is a white noise since $\langle \eta(t)\eta(t')\rangle_c = \lambda\,\delta(t-t')$, however it is of *non Gaussian* nature because its higher cumulants are non zero (cf. exercise 44, page 38).

In the close neighbourhood of time t_n, we can forget the drift and approximate the evolution as

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \cdots + b(x(t))\,\delta(t - t_n) \qquad \text{for } t \sim t_n \tag{225}$$

This means that $x(t)$ is discontinuous at t_n. In principle, dealing with a continuous function $\psi(t)$ we can write $\psi(t)\,\delta(t-t_n) = \psi(t_n)\,\delta(t-t_n)$. What should we do for a function $\psi(t)$ which is discontinuous at t_n ? Here, this makes the time evolution ambiguous : Eq. (225) shows that $x(t)$ makes a jump whose amplitude is $b(x(t))$, i.e. depends on the process at a time where the process is discontinuous and still unkown ! How to choose this time ? We propose two possible interpretations of the evolution (225) :

(i) **Proposal 1 (↔ "Itô")** Interpret the equation as $\frac{\mathrm{d}}{\mathrm{d}t}x(t) = \cdots + b(x(t_n^-))\,\delta(t-t_n)$, then

$$x(t_n^+) = x(t_n^-) + b(x(t_n^-)) \tag{226}$$

This is a natural choice for numerics. This is analogous to (223).

(ii) **Proposal 2 (↔ "Stratonovich")** As a physicist, one would rather consider that the δ-peak is a mathematical model for a regular narrow function of finite width $\delta(t) \to \delta^\epsilon(t)$, for example $\delta^\epsilon(t) = \frac{1}{2\epsilon}\mathrm{e}^{-|t|/\epsilon}$. Then starting from (225) one writes $\mathrm{d}x(t)/b(x(t)) \simeq \delta^\epsilon(t-t_n)\,\mathrm{d}t$ and integrate around the δ^ϵ, eventually taking the limit $\epsilon \to 0^+$. One gets

$$\int_{x(t_n^-)}^{x(t_n^+)} \frac{\mathrm{d}x}{b(x)} = 1 \tag{227}$$

which obviously differs from (226).

✍ ***Exercice 54*** *– : Consider a multiplicative noise with $b(x) = \alpha\,x$ where α is a constant. Compare the two evolutions (226) and (227) in this case.*

The choice of the prescription, i.e. the precise meaning to give to the multiplicative noise term, determines the evolution and contribute to define the stochastic process with the SDE. We stress that given *two different* interpretations of the *same* equation (225) leads to two different evolutions, (226) or (227), i.e. define *two different processes*. A similar problem occurs with the SDE (222) where $\eta(t)$ is a Gaussian white noise. Several interpretations can be given to the multiplicative noise term.

c) Itô convention

The simpler choice which first comes in mind is to consider that the process and the increment at equal time are independent. This is a natural choice if one discretizes the evolution, as explained above, see Eq. (223) where $\delta W_t = W(t+\delta t) - W(t)$ and x_t are *independent*. This is appropriate for numerical simulations. This is known as the *Itô convention*, corresponding to the "proposal 1" discussed above. In order to specify in which sense the SDE is understood, we write the SDE as

$$\mathrm{d}x(t) = a(x(t))\,\mathrm{d}t + b(x(t))\,\mathrm{d}W(t) \qquad \text{(Itô)}. \tag{228}$$

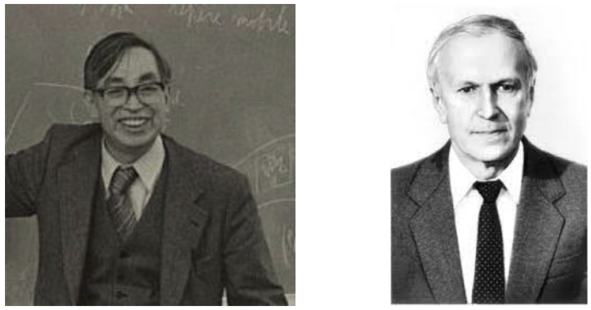


Figure 21: *Kiyoshi Itô (1915-2008) and Ruslan Leont'evich Stratonovich (1930-1997).*

Doblin-Itô calculus and the Itô formula.— The main manipulations can be performed keeping in mind that

$$\boxed{\text{Itô}: \quad x(t) \ \text{ and } \ \mathrm{d}W(t) \ \text{ are statistically independent at coinciding times}}$$

and

$$\boxed{\mathrm{d}W(t)^2 = \mathrm{d}t} \tag{229}$$

(mathematicians omit the averaging, see below the "appendix" on stochastic integrals in order to understand why $\langle\cdots\rangle$ can be omitted). Roughly speeking we have $\mathrm{d}W(t) \sim \mathcal{O}(\sqrt{\mathrm{d}t})$ and for this reason $\mathrm{d}W(t)^{2+n} = 0$ for $n > 0$.

An important formula concerns the change of variable from $x \to \varphi(x)$, where $\varphi(x)$ is a regular function, differentiable a least twice. From (228) we deduce

$$\boxed{\mathrm{d}\varphi(x(t)) = \left[\varphi'(x)\,a(x) + \frac{1}{2}\varphi''(x)\,b(x)^2\right]\mathrm{d}t + \varphi'(x)\,b(x)\,\mathrm{d}W(t) \qquad \text{(Itô)}} \tag{230}$$

This is known as the "*Itô formula*".

Proof : in the expansion of $\mathrm{d}\varphi(x) = \varphi'(x)\,\mathrm{d}x + \frac{1}{2}\varphi''(x)\,\mathrm{d}x^2 + \cdots$, due to (229), a term $\mathcal{O}(\mathrm{d}t)$ is produced by the $\mathcal{O}(\mathrm{d}x^2)$ term : $\mathrm{d}x^2 = \left[a(x)\,\mathrm{d}t + b(x)\,\mathrm{d}W(t)\right]^2 = b(x)^2\,\mathrm{d}W(t)^2 + 2a(x)b(x)\mathrm{d}W(t)\mathrm{d}t + a(x)^2\mathrm{d}t^2 = b(x)^2\,\mathrm{d}t + \mathcal{O}(\mathrm{d}W(t)\mathrm{d}t)$. The correction is $\mathcal{O}(\mathrm{d}W(t)\mathrm{d}t) = \mathcal{O}(\mathrm{d}t^{3/2})$.

✍ ***Exercice 55*** *– : Write the Itô formula for the multiplicative noise $\mathrm{d}x(t) = \kappa\, x\, \mathrm{d}W(t)$. The apply the formula to $\varphi(x) = x^2$*

Itô formula implies that **"Itô calculus" does not correspond with the "usual" differential calculus** when $W(t)$ is a regular function. Indeed, we have

$$\mathrm{d}\varphi(x(t)) \neq \varphi'(x(t))\,\mathrm{d}x(t) \qquad \text{(Itô)}. \tag{231}$$

Remark : With Itô convention, $x(t)$ and $W(t)$ are independent (at equal time). It follows that averaging (230) is straightforward and gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\varphi(x(t))\rangle = \langle\varphi'(x)\,a(x)\rangle + \frac{1}{2}\langle\varphi''(x)\,b(x)^2\rangle\,. \tag{232}$$

Related FPE.– One can immediatly deduce the FPE related to the Itô equation (228). Write (232) as

$$\frac{\partial}{\partial t}\int \mathrm{d}x\, P_t(x)\,\varphi(x) = \int \mathrm{d}x\, P_t(x)\,\left[\varphi'(x)\,a(x) + \frac{1}{2}\varphi''(x)\,b(x)^2\right] \tag{233}$$

In the r.h.s, integrations by part allow to factorize $\varphi(x)$. Because the relation is true $\forall\,\varphi(x)$, we conclude that

$$\boxed{\frac{\partial P_t(x)}{\partial t} = -\frac{\partial}{\partial x}\left[a(x)\,P_t(x)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[b(x)^2\,P_t(x)\right]} \tag{234}$$

Despite its drawbacks (for physicists), the Itô calculus is widely used by probabilists (and justified for certain physical situations). Also in finance, which is not a surprise as the time is discrete in this case, so it corresponds to the discretization scheme mentioned above, Eq. (223).

How to get the FPE from the SDE in a simple manner ? Above, the relation between the Itô SDE and the FPE was demonstrated by introducing a test function. A simpler way is to use $\langle \mathrm{d}W(t)^2\rangle_\text{noise} = \mathrm{d}t$ (physicist's notation) and to remark that the drift and the "diffusion" are given by

$$\langle a(x)\rangle = \frac{\langle \mathrm{d}x\rangle_\text{noise}}{\mathrm{d}t} \qquad \text{and} \qquad \left\langle b(x)^2\right\rangle = \frac{\langle \mathrm{d}x^2\rangle_\text{noise}}{\mathrm{d}t} \tag{235}$$

As an application we consider the multidimensional case

$$\mathrm{d}x_i(t) = a_i(\vec{x})\,\mathrm{d}t + b_{ij}(\vec{x})\,\mathrm{d}W_j(t) \qquad \text{(Itô)}. \tag{236}$$

with $\langle \mathrm{d}W_i(t)\mathrm{d}W_j(t)\rangle_\text{noise} = \delta_{ij}\mathrm{d}t$. Only the diffusion term is more complicated

$$\frac{\langle \mathrm{d}x_i\mathrm{d}x_j\rangle_\text{noise}}{\mathrm{d}t} = \langle b_{ik}b_{jk}\rangle \tag{237}$$

(with Einstein's convention for implicit summation over repeated indices). Then

$$\partial_t P_t(\vec{x}) = -\partial_i\left[a_i(\vec{x})P_t(\vec{x})\right] + \frac{1}{2}\partial_i\partial_j\left[b_{ik}(\vec{x})b_{jk}(\vec{x})P_t(\vec{x})\right]\,. \tag{238}$$

Application : Kramers and Smoluchowski equations.— Consider the equations

$$\begin{cases} \mathrm{d}x = v\,\mathrm{d}t \\ \mathrm{d}v = \left(-\frac{v}{\tau} + \frac{F(x)}{m}\right)\mathrm{d}t + \frac{1}{m}\sqrt{2k_\text{B}T\gamma}\,\mathrm{d}W(t) \end{cases} \tag{239}$$

The drift terms are $a_x = \frac{\langle \mathrm{d}x\rangle_\text{noise}}{\mathrm{d}t} = v$ and $a_v = \frac{\langle \mathrm{d}v\rangle_\text{noise}}{\mathrm{d}t} = -\frac{v}{\tau} + \frac{F(x)}{m}$. The diffusive terms are $b_{xx} = \langle \mathrm{d}x^2\rangle_\text{noise}/\mathrm{d}t = v^2\mathrm{d}t \to 0$, $b_{vv} = \langle \mathrm{d}v^2\rangle_\text{noise}/\mathrm{d}t = 2k_\text{B}T\gamma/m^2 = 2k_\text{B}T/(m\tau)$ and $b_{xv} = \langle \mathrm{d}x\mathrm{d}v\rangle_\text{noise}/\mathrm{d}t \sim \langle v\,\mathrm{d}W(t)\rangle_\text{noise} \to 0$. Finally, the FPE equation is

$$\left(\partial_t + v\,\partial_x + \frac{F(x)}{m}\partial_v\right)P_t(x,v) = \frac{1}{\tau}\partial_v\left(v + \frac{k_\text{B}T}{m}\partial_v\right)P_t(x,v) \tag{240}$$

This equation is called the *Kramers equation.*

✍ ***Exercice 56 – Smoluchowski equation :*** *Using the overdamped limit introduced in § 3 get an equation for* $P_t(x) = \int \mathrm{d}v\,P_t(x,v)$ *in the limit of strong friction.*

d) Stratonovich convention

For physicists, it would be more natural to consider the Gaussian white noise in the SDE (222) as the limit of a regular noise with a finite but small correlation time, for example $\langle\eta^\epsilon(t)\eta^\epsilon(t')\rangle = \frac{1}{2\epsilon}\mathrm{e}^{-|t-t'|/\epsilon}$ with ϵ "small". In this case we expect that the standard rules of differential calculus for regular functions hold. If we follow the same strategy as for the treatment of the multiplicative δ-peak, the construction is more complicated. We just describe how one could proceed. The proof of the results will come afterwards by using different arguments. Consider $W^\epsilon(t) = \int_0^t \mathrm{d}u\, \eta^\epsilon(u)$, which is regular (continuous and differentiable since $\langle(\partial_t W^\epsilon(t))^2\rangle = 1/(2\epsilon)$ is finite. Now the differential equation $\mathrm{d}x(t) = \alpha(x)\,\mathrm{d}t + \beta(x)\,\mathrm{d}W^\epsilon(t)$ is well defined mathematically. The key point is that the limit $\epsilon \to 0^+$ DOES NOT lead to the Itô SDE $\mathrm{d}x(t) = \alpha(x)\,\mathrm{d}t + \beta(x)\,\mathrm{d}W(t)$. Instead, it leads to the Itô SDE for a modified drift

$$\mathrm{d}x(t) = \left(\alpha(x) + \frac{1}{2}\beta(x)\,\beta'(x)\right)\mathrm{d}t + \beta(x)\,\mathrm{d}W(t) \qquad \text{(Itô)} \tag{241}$$

or, rather to the "*Stratonovich SDE*"

$$\mathrm{d}x(t) = \alpha(x)\,\mathrm{d}t + \beta(x)\,\mathrm{d}W(t) \qquad \text{(Stratonovich)} \tag{242}$$

However one must keep in mind that

Stratonovich : $x(t)$ and $\mathrm{d}W(t)$ are in general correlated at coinciding times

This is a bit subtle : $x(t)$ and $W(t')$ are uncorrelated for $t' > t$, as the process depends only on the noise in the *past*. The Stratonovich convention tells something about the correlations at *equal time* (these correlations are studied below in exercise 60).

Itô/Stratonovich connection : In other terms, the two SDE (228) and (242) describe the *same process* if

$$\alpha(x) = a(x) - \frac{1}{2}b(x)\,b'(x) \quad \text{ and } \quad \beta(x) = b(x)\,. \tag{243}$$

Note that Mathematicians follow a different strategy to define the Stratonovich SDE : in 1961, Stratonovich introduced a "symmetrized" form of stochastic integrals and differential forms (cf. appendix on stochastic integrals to have an idea of this strategy).

I emphasize :

- when $b(x)$ is not constant (case of multiplicative noise), the two SDE $\mathrm{d}x = a(x)\mathrm{d}t + b(x)\mathrm{d}W(t)$ (Itô) and $\mathrm{d}x = a(x)\mathrm{d}t + b(x)\mathrm{d}W(t)$ (Stratonovich) describe two *different* processes (related to different FPEs).
- Conversely $\mathrm{d}x = a(x)\mathrm{d}t + b(x)\mathrm{d}W(t)$ (Itô) and $\mathrm{d}x = \alpha(x)\mathrm{d}t + \beta(x)\mathrm{d}W(t)$ (Stratonovich) describe the *same* process provided (243) hold (then, they are related to the same FPE).

The FPE corresponding to the Stratonovich SDE (242) is

$$\frac{\partial P_t(x)}{\partial t} = -\frac{\partial}{\partial x}\left[\alpha(x)\,P_t(x)\right] + \frac{1}{2}\frac{\partial}{\partial x}\left[\beta(x)\,\frac{\partial}{\partial x}\left[\beta(x)\,P_t(x)\right]\right] \tag{244}$$

(this connection is proven in exercise 59).

✍ ***Exercice 57*** *– : Using the connection Itô-SDE/FPE (234) and the relation Itô-SDE/Strato-SDE (243), recover the Strato-SDE/FPE connection (244).*

We stress that, as (243) was not proven, we have not demonstrated either the relation between the Stratonovich SDE (242) and the Fokker-Planck equation (244), which is more tricky if we follow the construction evoked above : consider a regular noise $\eta^\epsilon(t)$ and take the (singular) limit of the Gaussian white noise at the end, which requires "projection method" (see [18, 53] for discussions). A more simpler approach is suggested in the appendix on stochastic integrals (and can be adapted at the level of the FPE). In exercise 59 below, we propose a simple derivation based on the assumption that usual differential calculus holds.

✍ ***Exercice 58 – Stratonovich corresponds to standard differential calculus :*** *Using the relation (243), transform the Itô formula (230) in the Stratonovich convention and check that*

$$\mathrm{d}\varphi(x(t)) = \varphi'(x(t))\,\mathrm{d}x(t) \qquad \textit{(Stratonovich).} \tag{245}$$

I.e. within the Stratonovich's prescription, standard rules of differential calculus for regular functions do apply.

✍ ***Exercice 59 – From the Stratonovich SDE to the FPE :*** *We consider the SDE* $\frac{\mathrm{d}x}{\mathrm{d}t} = \alpha(x) + \beta(x)\,\eta(t)$. *For additive noise* ($\beta(x) = \mathrm{cste}$) *the mapping onto the FPE is simple and has been discussed above. Difficulties have arisen for multiplicative noise. To circumvent this, we perform a transformation of the SDE which leads to additive noise. Using ordinary rules of differential calculus means that we interpret the SDE with the Stratonovich interpretation.*

a) Consider $z(t) = \int^{x(t)} \mathrm{d}\tilde{x}/\beta(\tilde{x})$. *Write the SDE for* $z(t)$.

b) Give the FPE for $Q_t(z)$, *the distribution of* $z(t)$.

c) Deduce the FPE for $P_t(x)$.

✍ ***Exercice 60 – Correlation between the process and the noise (Stratonovich) :*** *Consider the Stratonovich equation (242). Denoting* $\eta(t) = \mathrm{d}W(t)/\mathrm{d}t$, *show that* $\langle\beta(x(t))\,\eta(t)\rangle$ *can be expressed as the average of a function of* $x(t)$.

Hint : use the relation between Itô and Stratonovich SDE.

e) Take home message

- If a SDE appears in a physical model, it should be most frequently interpreted in the Stratonovich sense (if the white noise is the limit of a regular noise with symmetric correlation function).
- Remember how to relate the Stratonovich SDE (242) to the FPE (244) is the most important.
- If you like Itô calculus, keep in mind the relation between Stratonovich SDE and Itô SDE (243) and the relation with FPE, Eq. (234).

Bibliography : More can be found in the book of Gardiner [18]. For a presentation for mathematicians, see the book [43].

Historical note on Doblin-Itô calculus

Until 2000, Itô was considered as the foundator of what is usually denoted today the "Itô calculus". However in 2000, a sealed envelope ("pli cacheté" number 11-668), received in 1940 from a young mathematician named Vincent Doblin (born Wolfgang Doblin), was opened at Académie des Sciences de Paris, which showed that Döblin's contribution anticipated the work of Itô on stochastic calculus. Hence, we should rather name it "Doblin-Itô calculus".

Wolfgang Döblin was the son of a well-known german writer, Alfred Döblin. Because he was jewish and opponent to the nazism, Alfred Döblin escaped Germany to Zürich at the begining of 1933 with part of his family, followed by his son Wolgang. They arrived in Paris in the fall of 1933. Wolfgang obtained the french nationality in 1936, becoming "Vincent Doblin". In 1938 he passed his PhD, under the supervision of the famous mathematician Maurice Fréchet, however, at the end of 1938, he was incorporated in the French army. Refusing to serve as an officer, he was affected to the communications. During this period in the army, at the begining of the war,

Figure 22: *Vincent Doblin (1915-1940). A page of the pli cacheté (from [6]).*

he was sent to the Ardennes and was able to produce important scientific results, which he chose to send to the Académie des Sciences under the form of a "pli cacheté", entitled "sur *l'équation de Kolmogoroff*, par Vincent Doblin". Just after the collapse of the French army, as his company was surrounded by germans in the Vosges region, Vincent Doblin tried unsuccessfully to cross the german lines and eventually preferred to commit suicide rather being captured. It was only possible to open the "pli cacheté" 60 years after his death. Although Vincent Doblin was already known in the mathematics community despite his youth, the importance of his contribution was not anticipated before 2000.

To learn more : look at the article [6] (available on the internet) written by the two probabilists Bernard Bru and Marc Yor, who analyzed the pli cacheté and recognized its scientific importance. Or the book by Marc Petit [45].

✍ ***Exercice 61* – *Electromagnetic noise :*** *We consider a model of electromagnetic noise : the two components of the electric field* $E_x + \mathrm{i}\, E_y$ *obey the SDE*

$$\begin{cases} \mathrm{d}E_x(t) = -\gamma\, E_x(t)\,\mathrm{d}t + \sqrt{D}\,\mathrm{d}W_x(t) \\ \mathrm{d}E_y(t) = -\gamma\, E_y(t)\,\mathrm{d}t + \sqrt{D}\,\mathrm{d}W_y(t) \end{cases} \tag{246}$$

where W_x *and* W_y *are two independent Wiener processes (hence we can write* $\mathrm{d}W_x^2 = \mathrm{d}W_y^2 = \mathrm{d}t$ *and* $\mathrm{d}W_x \mathrm{d}W_y = 0$*, remember that averages can be omitted for elementary differential increments).*

1/ We introduce the intensity and the phase : $E_x = \sqrt{I}\,\cos\theta$ *and* $E_y = \sqrt{I}\,\sin\theta$*. Write a SDE for the intensity* I *within the Stratonovich convention.*

2/ We write $E_x + \mathrm{i}\, E_y = \mathrm{e}^{\lambda+\mathrm{i}\theta}$*, where* $A = \mathrm{e}^{\lambda}$ *is the amplitude of the field and* θ *its phase. Within Itô calculus, express* $\mathrm{d}\lambda + \mathrm{i}\,\mathrm{d}\theta$ *as a function of* λ*,* θ *and the noises* $\mathrm{d}W_x(t)$ *and* $\mathrm{d}W_y(t)$*. Show that*

$$\mathrm{d}W_A(t) = \cos\theta(t)\,\mathrm{d}W_x(t) + \sin\theta(t)\,\mathrm{d}W_y(t) \text{ and } \mathrm{d}W_\theta(t) = -\sin\theta(t)\,\mathrm{d}W_x(t) + \cos\theta(t)\,\mathrm{d}W_y(t)$$

are two independent noises. Deduce two Itô SDE for $\lambda(t)$ *and* $\theta(t)$*.*

3/ Using the Itô formula, deduce the Itô SDE for the amplitude $A = |E_x + \mathrm{i}\, E_y|$ *and then for the intensity* $I = A^2$*. Relate the Itô SDE for* I *to a Stratonovich SDE and compare to the equation obtained in the first question.*

4/ Write the SDE for the amplitude under the form

$$\mathrm{d}A(t) = -V'(A(t))\,\mathrm{d}t + \sqrt{D}\,\mathrm{d}W_A(t) \tag{247}$$

and give the "potential" $V(A)$*. Find its minimum.*
Using a harmonic approximation, deduce the equilibrium distribution for the amplitude and the

correlator $\langle A(t)A(t')\rangle_c$. Discuss the harmonic approximation.

5/ Write the FPE related to the SDE for $A(t)$. Deduce the exact equilibrium distribution and compare $\langle A\rangle$ and $\langle \delta A^2\rangle$ with the one given by the harmonic approximation. Discuss also the distribution of the intensity.

APPENDIX : Stochastic integrals

If you feel unsatisfactory with the above presentation of Itô/Stratonovich convention, you can read this paragraph (borrowed from chapter 4 of [18]). Instead of considering the SDE, one considers integrals of the form $\int_0^t \mathrm{d}W(t')\,G(t')$ which requires the same discussion as for SDE.

Itô integral.— One defines the Itô integral as

$$\text{Itô}\int_0^t \mathrm{d}W(t')\,G(t') \stackrel{\text{def}}{=} \underset{N\to\infty}{\text{ms-lim}} \sum_{i=1}^N \delta W_i\, G(t_{i-1}) \tag{248}$$

where $\delta W_i = W(t_i) - W(t_{i-1})$. Here "ms-lim" stands for "mean-square limit" of a random variable, meaning that :

$$\underset{N\to\infty}{\text{ms-lim}}\, X_N = X_\infty \quad\text{if}\quad \lim_{N\to\infty} \langle [X_N - X_\infty]^2\rangle = 0\,. \tag{249}$$

Let us study an example. Consider the integral Itô $\int_0^t \mathrm{d}W(t')\,W(t')$. One has to analyze the sum

$$\sum_{i=1}^N \delta W_i\, W_{i-1} = \frac{1}{2}\sum_{i=1}^N \left[(\delta W_i + W_{i-1})^2 - W_{i-1}^2 - \delta W_i^2\right] = \frac12\sum_{i=1}^N W_i^2 - \frac12 \sum_{i=0}^{N-1} W_i^2 - \frac12\sum_{i=1}^N \delta W_i^2$$
$$= \frac12\left[W_N^2 - W_0^2\right] - \frac12\sum_{i=1}^N \delta W_i^2 \tag{250}$$

It is easy to show that $\underset{N\to\infty}{\text{ms-lim}} \sum_{i=1}^N \delta W_i^2 = t$ (this is the reason why one writes $\mathrm{d}W(t)^2 = \mathrm{d}t$ without the average). Thus

$$\text{Itô}\int_0^t \mathrm{d}W(t')\,W(t') = \frac12\left[W(t)^2 - W(0)^2 - t\right] \tag{251}$$

which differs (by $-t/2$) from the usual Riemann integral of a regular function.

Stratonovich integral.— Now introduce the definition of the Stratonovich integral

$$\int_0^t \mathrm{d}W(t')\,G(t') \stackrel{\text{def}}{=} \underset{N\to\infty}{\text{ms-lim}} \sum_{i=1}^N \delta W_i\, \frac{G(t_i)+G(t_{i-1})}{2} \tag{252}$$

(I use the standard notation for integration, anticipating that it will coincide with usual Riemann integrals).

Consider now the same integral as before $\int_0^t \mathrm{d}W(t')\,W(t')$ with the new convention. This time, one deals with

$$\sum_{i=1}^N \delta W_i\, \frac{W_i + W_{i-1}}{2} = \frac12\sum_{i=1}^N W_i^2 - \frac12\sum_{i=1}^N W_{i-1}^2 = \frac12 W_N^2 - \frac12 W_0^2 \tag{253}$$

so that we have recovered

$$\int_0^t \mathrm{d}W(t')\,W(t') = \frac12\left[W(t)^2 - W(0)^2\right] \tag{254}$$

as for the integration of regular functions.

From stochastic integral to SDE : in the books [43, 18], stochastic integrals are first discussed along these lines, then SDE are introduced as derivatives of stochastic integrals.

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 deterministic dynamics from which will emerge the effective dynamics described by the Langevin equation. This will allow to identify the microscopic origin of the dissipation.

Model.— From Section 3, one would 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

$$H = \frac{p^2}{2m} + V(x) + \sum_n \left[\frac{p_n^2}{2} + \frac{1}{2}\omega_n^2 \left(q_n - \frac{c_n\,x}{\omega_n^2} \right)^2 \right] \tag{255}$$

i.e.

$$H_{\mathrm{sys}}(x,p) = \frac{p^2}{2m} + V(x) \tag{256}$$

$$H_{\mathrm{env}}(\{q_n,p_n\}) = \sum_n \left[\frac{p_n^2}{2} + \frac{1}{2}\omega_n^2 q_n^2 \right] \tag{257}$$

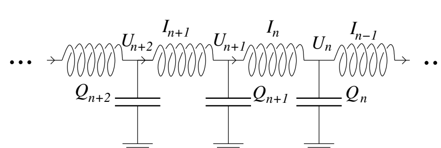
and the coupling is linear (this is very important for the following)

$$H_{\mathrm{int}} = -x \sum_n c_n\, q_n + \frac{1}{2}x^2 \sum_n \frac{c_n^2}{\omega_n^2} \tag{258}$$

Here c_n are coupling constants.

A physical realization is : an electron in an atom, coupled to the electromagnetic modes. Or : an electric device coupled to a L-C line.

✍ ***Exercice 62 – 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 $I_n(t) = \tilde{I}_n \mathrm{e}^{-\mathrm{i}\omega t}$. We recall that the impendance of the capacitance is $Z_C = 1/(-\mathrm{i}\omega C)$ and that of the inductance $Z_L = -\mathrm{i}\omega L$, 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 $I_n(t) = \mathrm{e}^{\mathrm{i}qn-\mathrm{i}\omega(q)t}$ only exist in a finite bandwidth $\omega \in [0,\omega_0]$ where $\omega_0 \stackrel{\mathrm{def}}{=} 2/\sqrt{LC}$. Give the dispersion relation.*

3/ Evanescent modes.– *Study solutions of the form* $I_n(t) = (-1)^n \mathrm{e}^{qn - \mathrm{i}\omega(q)t}$. *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-indinite line* $Z_\infty \equiv Z(\omega)$. *Plot* $\mathrm{Re}\, Z(\omega)$ *and* $\mathrm{Im}\, Z(\omega)$. *Discuss the fact that* $\mathrm{Re}\, Z(\omega) \neq 0$ *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

$$\begin{cases} m\,\ddot{x} = F(x) - x\sum_n \frac{c_n^2}{\omega_n^2} + \sum_n c_n q_n \\ \ddot{q}_n = -\omega_n^2 q_n + c_n\, x \end{cases} \tag{259}$$

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 $\ddot{G}^{\mathrm{R}}(t) + \omega_n^2 G^{\mathrm{R}}(t) = \delta(t)$ is $G^{\mathrm{R}}(t) = \theta_{\mathrm{H}}(t)\,\frac{\sin(\omega_n t)}{\omega_n}$. Thus we can solve the equation of motion for the oscillators

$$q_n(t) = q_n(0)\,\cos(\omega_n t) + \dot{q}_n(0)\,\frac{\sin(\omega_n t)}{\omega_n} + c_n\int_0^t \mathrm{d}t'\,\frac{\sin(\omega_n(t-t'))}{\omega_n}\,x(t') \tag{260}$$

We split the source term in Eq. (259) in two parts

$$\sum_n c_n\, q_n(t) = \overbrace{\sum_n c_n\left(q_n(0)\,\cos(\omega_n t) + \dot{q}_n(0)\,\frac{\sin(\omega_n t)}{\omega_n}\right)}^{\xi(t)} + \int_0^t \mathrm{d}t'\,\Gamma(t-t')\,x(t') \tag{261}$$

where

$$\Gamma(t) \stackrel{\mathrm{def}}{=} \sum_n c_n^2\frac{\sin(\omega_n t)}{\omega_n} \tag{262}$$

is a function depending of the details of the model. We denote $\xi(t)$ "the noise" as it is controlled by a macroscopic number of degrees of freedom of the bath, which is expected to exhibit a complex dynamics. Using $\int_0^\infty \mathrm{d}t\,\mathrm{e}^{\mathrm{i}\omega t} = \frac{1}{0^+ - \mathrm{i}\omega}$ we remark that

$$\int_0^\infty \mathrm{d}t\,\Gamma(t) = \sum_n \frac{c_n^2}{\omega_n^2} \tag{263}$$

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

$$m\,\ddot{x}(t) = F(x(t)) - x(t)\int_0^\infty \mathrm{d}\tau\,\Gamma(\tau) + \int_0^t \mathrm{d}\tau\,\Gamma(\tau)\,x(t-\tau) + \xi(t) \tag{264}$$

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

$$P(\{q_n, p_n\}) \propto \mathrm{e}^{-\beta H_{\mathrm{env}}} \tag{265}$$

so that

$$\langle q_n(0) q_m(0) \rangle = \delta_{n,m} \frac{k_{\rm B}T}{\omega_n^2} \tag{266}$$

$$\langle \dot q_n(0) \dot q_m(0) \rangle = \delta_{n,m}\, k_{\rm B}T \tag{267}$$

Then the noise correlator is

$$C(t-t') = \big\langle \xi(t)\xi(t') \big\rangle = k_{\rm B}T \sum_n \frac{c_n^2\,\cos(\omega_n(t-t'))}{\omega_n^2} \tag{268}$$

At this stage it is useful to define the spectral function

$$J(\omega) \stackrel{\rm def}{=} \pi \sum_n \frac{c_n^2}{2\omega_n}\,\delta(\omega-\omega_n) \tag{269}$$

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

$$\Gamma(t) = \frac{2}{\pi}\int_0^\infty {\rm d}\omega\, J(\omega)\,\sin(\omega t) \tag{270}$$

and the correlator

$$C(t) = \frac{2k_{\rm B}T}{\pi}\int_0^\infty {\rm d}\omega\,\frac{J(\omega)}{\omega}\,\cos(\omega t) \tag{271}$$

in terms of the spectral function. Two remarks :

- In practice, we expect a dense spectrum of oscillators for frequencies $\omega \geqslant 0$ (it is natural to assume that the spectrum of eigenmodes start at $\omega=0$ since there exist low frequency excitations ususally). On the other hand, the eigen-frequencies should be cut off at a characteristic scale ω_D (like the Debye frequency in a solid).
  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 $\omega \to 0$, 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 $g(0)$ =cste.
- The spectrum of frequencies is usually cut off at a frequency ω_D related to the microscopic scale (the lattice spacing for the phonons in a crystal).

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

$$J(\omega) = \gamma_0\,\omega\,\frac{\omega_D^2}{\omega^2+\omega_D^2} \tag{272}$$

gives

$$C(t) = k_{\rm B}T\,\gamma_0\,\omega_D\,{\rm e}^{-\omega_D|t|}\,. \tag{273}$$

Its integral is

$$\int_{-\infty}^{+\infty} {\rm d}t\, C(t) = 2\gamma_0\, k_{\rm B}T \tag{274}$$

which recall us something...

Effective equation of motion.— Let us come back to the analysis of the effective equation of motion (264). If the spectral function 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

$$\gamma(t) = \int_t^\infty \mathrm{d}t'\,\Gamma(t') \tag{275}$$

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

$$\gamma(t) = \theta_{\mathrm{H}}(t) \sum_n \frac{c_n^2 \cos(\omega_n t)}{\omega_n^2} = \frac{2\,\theta_{\mathrm{H}}(t)}{\pi} \int_0^\infty \mathrm{d}\omega\, \frac{J(\omega)}{\omega} \cos(\omega t) \tag{276}$$

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

An integration by parts gives

$$\int_0^t \mathrm{d}\tau\,\Gamma(\tau)\,x(t-\tau) = \gamma(0)\,x(t) - \gamma(t)\,x(0) - \int_0^t \mathrm{d}\tau\,\gamma(\tau)\,\dot{x}(t-\tau) \tag{277}$$

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

$$\boxed{m\,\ddot{x}(t) = F(x(t)) - \int_0^t \mathrm{d}\tau\,\gamma(\tau)\,\dot{x}(t-\tau) + \xi(t)} \tag{278}$$

This makes clear the physical interpretation of the integral term as a friction term, non local in time, as damping needs some time to establish.

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

$$\boxed{C(\tau) = k_{\mathrm{B}}T\,\gamma(\tau) \quad \text{for } \tau > 0 \qquad \text{(FDT)}} \tag{279}$$

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 § 3 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 $P_{\mathrm{sys}}(x,p) \propto \exp[-\frac{\beta}{2}mv^2]$. 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 $C = 2\gamma_0 k_{\mathrm{B}}T$. To some extent, this relation was *assumed* for consistency.

- Here, we have only assumed that, being macroscopic, the *bath* is at thermal equilibrium $P_{\mathrm{bath}}(x,p) \propto \mathrm{e}^{-\beta H_{\mathrm{bath}}}$. As a result of the integration of the conservative dynamics, we have deduced the relation $C = 2\gamma_0 k_{\mathrm{B}}T$. 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. [19] or [50]).

Energetic considerations : We now study the energy of the system

$$\frac{\mathrm{d}}{\mathrm{d}t}H_{\mathrm{sys}} = \dot{x}\,[m\ddot{x} - F(x)] = -v(t)\int_0^t \mathrm{d}\tau\,\gamma(\tau)\,v(t-\tau) + v(t)\,\xi(t) \tag{280}$$

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

$$\frac{đW}{\mathrm{d}t} = v(t)\,\xi(t) \tag{281}$$

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

$$\frac{đQ}{\mathrm{d}t} = -v(t)\int_0^t \mathrm{d}\tau\,\gamma(\tau)\,v(t-\tau)\,. \tag{282}$$

One can consider the model with $\gamma(\tau) = \gamma_0\,\omega_D\,\mathrm{e}^{-\omega_D\tau}$. 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 $v(t) \simeq \frac{1}{m}\int_0^t \mathrm{d}t'\,\xi(t')\,\mathrm{e}^{-(t-t')/\tau}$. We can estimate the average work of the Langevin force

$$\frac{\langle đW\rangle}{\mathrm{d}t} = \frac{1}{m}\int_0^t \mathrm{d}t'\,\left\langle \xi(t)\xi(t')\right\rangle\,\mathrm{e}^{-(t-t')/\tau} = \frac{C}{m}\,\theta_{\mathrm{H}}(0) = \frac{C}{2m} = \frac{k_{\mathrm{B}}T}{\tau} \tag{283}$$

where we have used that $\theta_{\mathrm{H}}(0) = 1/2$ (this is consistent with a symmetric regularised δ function). The averaged heat is

$$\frac{\langle đQ\rangle}{\mathrm{d}t} = -\int_0^t \mathrm{d}t'\,\gamma(t-t')\,\left\langle v(t)v(t')\right\rangle \tag{284}$$

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 $\gamma(t-t')\,\langle v(t)v(t')\rangle \simeq \gamma(t-t')\,\left\langle v(t)^2\right\rangle$, hence

$$\frac{\langle đQ\rangle}{\mathrm{d}t} \simeq -\left\langle v(t)^2\right\rangle\int_0^\infty \mathrm{d}t''\,\gamma(t'') = -\frac{k_{\mathrm{B}}T}{m}\,\gamma_0 = -\frac{k_{\mathrm{B}}T}{\tau} \tag{285}$$

As it should the total energy is conserved on averaged

$$\langle đW\rangle + \langle đQ\rangle = 0 \tag{286}$$

The Langevin force furnishes some work to the particle and the bath receives the heat which is dissipated. The bath receives the entropy $\mathrm{d}S_{\mathrm{bath}} = -đQ/T$, thus the dissipation corresponds to the production of entropy with rate

$$\frac{\mathrm{d}\left\langle S_{\mathrm{bath}}\right\rangle}{\mathrm{d}t} = +\frac{k_{\mathrm{B}}}{\tau}\,. \tag{287}$$

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 [31, 11, 38, 48, 47, 34] or to the lectures of Bernard Derrida at collège de France (2015-2016), `https://www.college-de-france.fr/site/bernard-derrida/`

☺ Important points

- Understand the difference between Itô and Stratonovich conventions (for multiplicative noise).
- For Itô calculus : remember $\mathrm{d}W(t)^2 = dt$ and be careful with differential calculus !
- Be familiar with the relations between SDE (Itô or Stratonovich) and the FPE.

6 The Fokker-Planck equation

When considering a stochastic process, the main goal is usually to determine its statistical properties, i.e. its distribution. Fokker-Planck equation is an important equation for the distribution of Markov processes with no jump. In this chapter, we discuss several applications of the Fokker Planck equation. This will demonstrate the power of the approach, compared to the stochastic differential equation approach, and we will see that we can address more subtle properties of random processes, like exit problem or first passage time.

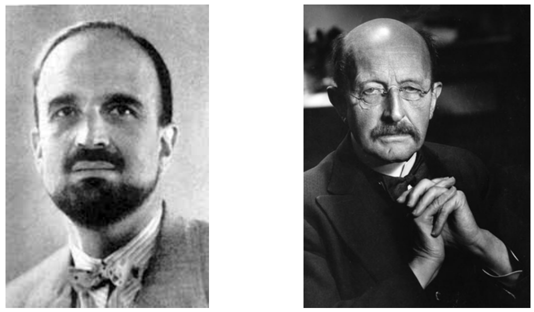


Figure 23: *Adriaan Fokker (1887-1972) and Max Planck (1858-1947)*

6.1 The Fokker-Planck equation

The Fokker-Planck equation is a special form of the master equation (159) for which the kernel can be reduced to a differential operator :

$$\frac{\partial P_t(x)}{\partial t} = -\frac{\partial}{\partial x}\left[F(x)\, P_t(x)\right] + \frac{\partial^2}{\partial x^2}\left[D(x)\, P_t(x)\right] \tag{288}$$

Below, we explain precisely in what limit and under what conditions we can go from (159) to (288). The equation is also known as the "*Kolmogorov equation*"or, for $D(x) \to D$, the "*Smoluchowski equation*". Let us first give the interpretation of the two terms in the Fokker-Planck equation (applications will be discussed below).

a) Drift

Imagine that only the first term is present and that F is uniform

$$\frac{\partial P_t(x)}{\partial t} = -F\,\frac{\partial P_t(x)}{\partial x}\,. \tag{289}$$

The solution is $P_t(x) = \varphi(x - F\,t)$ thus F is the velocity (for uniform F, there is no deformation of the distribution). The first term in (288) is the **drift term**, where "the drift" $F(x)$ is interpreted as the force acting on the particle (remember that velocity=force).

b) Diffusion

Consider now the effect of the second term of (288) for a uniform D.

$$\frac{\partial P_t(x)}{\partial t} = D\frac{\partial^2 P_t(x)}{\partial x^2}\,. \tag{290}$$

As t grows, the distribution increases where the function is convex and diminishes where the function is concave (Fig.24). This leads to a spreading of the distribution. The second term in (288) is the **diffusion term**, $D(x)$ playing the role of a x-dependent diffusion constant.

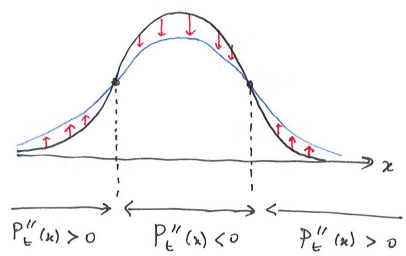


Figure 24: *the spreading of the distribution due to the diffusion term.*

c) Current density

The Fokker-Planck equation can be rewritten under the form of a **conservation equation**

$$\frac{\partial P_t(x)}{\partial t} = -\frac{\partial J_t(x)}{\partial x} \tag{291}$$

where

$$\boxed{J_t(x) = \underbrace{F(x)\,P_t(x)}_{\text{drift current}} - \overbrace{\frac{\partial}{\partial x}\left[D(x)\,P_t(x)\right]}^{\text{diffusion current}}} \tag{292}$$

is the current density. The drift current is the usual velocity×density. The diffusion term accounts for the fact that when the density is non uniform, particles moves from high density regions to low density regions (entropic effect).

6.2 From the master equation to the Fokker-Planck equation

In the previous chapter, we have deduced the FPE from stochastic differential equations (SDE). This connection has relied on the fact that the noise in the SDE is a Gaussian white noise.

In the present paragraph, we would like to understand the emergence of the FPE from a broader perspective. For this purpose we go back to the master equation and show in which limit and under what conditions we can establish the connection to the FPE.

a) Kramers-Moyal expansion

We consider the case where the state of the system is described by a coordinate which varies continuously in $\mathbb{R}$. The master equation (159) can be appropriately written as [19]

$$\frac{\partial P_t(x)}{\partial t} = \int \mathrm{d}x' \left[W(x|x')\,P_t(x') - W(x'|x)\,P_t(x)\right] \tag{293}$$

where the kernel $W(x|x')$ is $\sim$ probability rate to jump from x' to x. Here, I have prefered this form, slightly different from (159), in order to avoid any external constraint on the kernel $W(x|x')$ and to make more explicit the conservation of probability $\partial_t \int \mathrm{d}x\, P_t(x) = 0$. Furthermore we rewrite the kernel as a function of the initial poisition and the jump amplitude $\eta = x - x'$:

$$W(\underset{\text{final}}{\underset{\downarrow}{x}} \mid \underset{\text{initial}}{\underset{\downarrow}{x'}}) \equiv \tilde{w}(\underset{\text{initial}}{\underset{\downarrow}{x'}}\;;\;\underset{\text{jump}}{\underset{\downarrow}{\eta}} = x - x') \tag{294}$$

[19] one can recover the form (159) by introducing $\widetilde{W}(x|x') = W(x|x') - \delta(x-x')\int \mathrm{d}y\, W(y|x)$, which satisfies $\int \mathrm{d}x\, \widetilde{W}(x|x') = 0$. We get $\partial_t P_t(x) = \int \mathrm{d}x'\, \widetilde{W}(x|x')\,P_t(x')$.

We get

$$\frac{\partial P_t(x)}{\partial t} = \int \mathrm{d}\eta\, \tilde{w}(x-\eta;\eta)\, P_t(x-\eta) - P_t(x) \int \mathrm{d}\eta\, \tilde{w}(x;-\eta) \tag{295}$$

where we have used that $W(x'|x) = \tilde{w}(x; x'-x=-\eta)$.

Comparison with the specific case of the CPP : Note that the master equation for the CPP, Eq. (170), is an example of such an equation, corresponding to a translation invariant situation, then

$$W(x'+\eta|x') \equiv \tilde{w}(x';\eta) \stackrel{\mathrm{CPP}}{=} \lambda\, w(\eta) \tag{296}$$

is independent of x' (the CPP is invariant under translation in space). Here w is the normalised distribution of jumps and λ the rate of jumps.

Come back to the general case : This remark makes clear that in general $\tilde{w}(x';\eta)$ can be interpreted as the distribution of the jump amplitude (up to a factor related to the rate of jumps λ), which depends in general on the starting point of the jump x'.

In general, the master equation being an integral equation, it is not very simple to manipulate (unless in simple cases where translation invariance holds, as it was shown in Exercise 45). We now want to show how (under what conditions) it can be replaced by a partial differential equation much more easy to handle. The main assumptions are now

- $\tilde{w}(x';\eta)$ is a sharp function of η (small jumps dominates)
- $\tilde{w}(x';\eta)$ and $P(x';t)$ are smooth functions of x'.

These assumptions should allow an expansion of the function of $x-\eta$ in powers of η

$$\int \mathrm{d}\eta\, \tilde{w}(x-\eta;\eta)\, P_t(x-\eta) = \int \mathrm{d}\eta \sum_{n=0}^{\infty} \frac{(-\eta)^n}{n!} \frac{\partial^n}{\partial x^n} \left[\tilde{w}(x;\eta)\, P_t(x)\right] \tag{297}$$

After introduction of this series in (295), the $n=0$ term is cancelled by the last term of Eq. (295). We can introduce

$$a_n(x) \stackrel{\mathrm{def}}{=} \int \mathrm{d}\eta\, \eta^n\, \tilde{w}(x;\eta) \tag{298}$$

($\sim n$-th moment of the jumps from x'). The condition that $\tilde{w}(\cdot;\eta)$ is a "narrow" function should be rather reformulated as $a_n(x) < \infty\ \forall n$. Permuting integration over η and derivations with respect to x, we end with

$$\frac{\partial P_t(x)}{\partial t} = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x^n} \left[a_n(x)\, P_t(x)\right] \tag{299}$$

which is known as the **Kramers-Moyal expansion**. Of course such an expansion only exists if the distribution of jumps is such that *all moments are finite*. Under this form, the equation if not much more easy to manipulate than the integral form from which we started. However, with the above assumption that $\tilde{w}(x;\eta)$ is a narrow function of η, corresponding to small jumps, we expect the moments $a_n(x)$ to decay fast with n, which allows a truncation of the expansion. The truncated equation

$$\frac{\partial P_t(x)}{\partial t} = -\frac{\partial}{\partial x}\left[a_1(x)\, P_t(x)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[a_2(x)\, P_t(x)\right] \tag{300}$$

corresponds to the Fokker-Planck equation (288). [20] Following exercise 45, we expect that this truncation describes correctly the large scale properties, as long as the distribution of the jumps

[20] A proper justification of the truncation requires a neat rescaling of the jumps and the rate, like it was done in the above exercise 45. The argument follows the spirit of the central limit theorem.

is sufficiently narrow. The FPE describes a *continuous* random process (i.e. the jumps disappear in the continuum limit, which is only possible if the original distribution of jumps is sufficiently narrow). [21]

Remark : Having in mind the analysis of the CPP, we could write $\tilde{w}(x;\eta) = \lambda\, w_x(\eta)$ where $\lambda = \int \mathrm{d}\eta\, \tilde{w}(x;\eta)$ so that $w_x(\eta)$ is normalized. This makes clear that, in the Master equation, the time can be rescaled as $\tilde{t} = \lambda t$, so that large time corresponds to high rate. Hence, the truncation leading to the Fokker-Planck equation describes the large time limit of the Master equation (provided the Kramers Moyal expansion exists).

Bibliography : I have borrowed this discussion from the book of van Kampen (chapter VIII) [53].

Pawula theorem : Can we truncate the Kramers-Moyal expansion (299) at any n ? The Pawula theorem states that it can only be stopped at $n = 1$ or $n = 2$. The positivity of the solution implies that if not stopped at $n = 1$ or $n = 2$, one should keep the infinite series (cf. § 4.3, [46]). This reminds us the Marcinkiewicz theorem about the generating function of cumulants.

b) Conclusion : jump process *versus* diffusion

In general, the master equation can be written under the form

$$\partial_t P_t(x) = \mathscr{L} P_t(x) \tag{301}$$

where $\mathscr{L}$ is a linear operator.

- For a *jump process*, the linear operator is an integral operator, of the form (293) of (295). For example, a simple **jump process** is the CPP studied above, for which $\big[\mathscr{L}\varphi\big](x) = \lambda \int \mathrm{d}\eta\, w(\eta) \big(\varphi(x-\eta) - \varphi(x)\big)$. In the general case, the distribution of the jump amplitude depends on the initial position, Eq. (295).
- If the linear operator is a differential operator, $\mathscr{L} = -\partial_x a(x) + \frac{1}{2}\partial_x^2 b(x)$, one says that "the process is **a diffusion**". From the Pawula theorem, the differential operator can be at most second order. Physically, a diffusion is obtained as the limit of small jumps occuring with high rate. This is the type of stochastic processes discussed in the previous chapter on SDE and the present one on FPE.
- In general, a Markov process can combine a diffusion and jumps.

Go further : Let us discuss further the simple and important case of *homogeneous* and *translation invariant* processes. Consider such a process $X(t)$, known as a **Lévy process** and characterized by its *Lévy exponent* $\Lambda(k)$:

$$\left\langle \mathrm{e}^{-\mathrm{i}kX(t)} \right\rangle = \mathrm{e}^{-t\,\Lambda(k)} \tag{302}$$

This behaviour is related to the Markov property and the property of i.i.d. increments (for mathematicians it is related to the property of "*infinite divisibility*"). This shows that cumulants of $X(t)$ all grow linearly with t (like for the sum of N i.i.d. variables whose cumulants are $\propto N$). The Lévy exponent $\Lambda(k)$ is the generating function of the cumulants of $X(1)$.

[21] The condition "distribution of jumps sufficiently narrow" should have been made clear in exercise 45. It is furhter discussed in exercise 51.

- The case of a diffusion corresponds to the drifted Brownian motion, hence $\Lambda_{\rm BM}(k)=\mathrm{i}k\mu+\frac{1}{2}\sigma^2k^2$, where μ is the drit and σ^2 the variance. The process can be written as $X(t)=\mu t+\sigma\,W(t)$, where $W(t)$ is the Wiener process.
- The compound Poisson process (CPP) studied above corresponds to $\Lambda_{\rm CPP}(k)=\lambda\,(1-\hat{w}(k))$ where λ is the rate of jumps and $\hat{w}(k)=\int\mathrm{d}\eta\,w(\eta)\,\mathrm{e}^{-\mathrm{i}k\eta}$ the characteristic function of the jumps (see exercises 45 or 46).
- Combinations of drifted Brownian motion and CPP belong to the class of "*interlacing processes*" for which the Lévy exponent has the form $\Lambda(k)=\Lambda_{\rm BM}(k)+\Lambda_{\rm CPP}(k)=\mathrm{i}k\mu+\frac{1}{2}\sigma^2k^2+\lambda\,(1-\hat{w}(k))$.
- Remarkably, the class of interlacing processes does not exhaust all possible Lévy processes. The most general Lévy processes are characterized by the Lévy exponent given by the *Lévy-Khintchine formula* [22]

$$\Lambda(k)=\mathrm{i}k\tilde{\mu}+\frac{1}{2}\sigma^2k^2+\int m(\mathrm{d}\eta)\,\left(1-\mathrm{e}^{-\mathrm{i}k\eta}-\frac{\mathrm{i}k\eta}{1+\eta^2}\right) \tag{304}$$

where the measure $m(\mathrm{d}\eta)$ is not always associated with a normalisable density. When it is not, the Lévy process is "*singular*". Roughly speaking, a singular Lévy process corresponds to a situation where the density of small jumps goes to infinity. Let us clarify this point : if $m(\mathrm{d}\eta)=\lambda\,w(\eta)\,\mathrm{d}\eta$ where $w(\eta)$ is normalised, then we can split the integral and the process is clearly an interlacing process for drift $\mu=\tilde{\mu}-\lambda\int\mathrm{d}\eta\,w(\eta)\,\eta/(1+\eta^2)$. However, there are cases where it is not allowed to split the integral, while $\Lambda(k)$ is finite. For example, the measure $m(\mathrm{d}\eta)=(\mathrm{d}\eta/\eta)\,\mathrm{e}^{-\eta}$ is not related to a normalisable distribution of jumps, however it leads to a finite Lévy exponent $\Lambda(k)=\ln(1+\mathrm{i}k)$ (this process is known as the "Gamma subordinator"). Another example of singular Lévy process is the "alpha stable subordinator" with $m(\mathrm{d}\eta)=\mathrm{d}\eta\,\alpha/\big(\Gamma(1-\alpha)\eta^{1+\alpha}\big)$, related to the Lévy exponent $\Lambda(k)=(\mathrm{i}k)^\alpha$.

More can be found in the book [4] or the summary [3] (or the brief introduction in the article [20]).

6.3 Spectral analysis of the Fokker Planck equation

In this section we discuss the the FPE from the spectral point of view. For simplicity, we study processes with additive noise

$$\mathrm{d}x(t)=F(x(t))\,\mathrm{d}t+\sqrt{2D}\,\mathrm{d}W(t)\;, \tag{305}$$

i.e. Eq. (288) for $D(x)\to D$. Because we only consider additive noise, Itô and Stratonovich interpretations of the SDE correspond to the same process. *We restrict ourselves to the one-dimensional case where the drift allows for an equilibrium state*, i.e. when the potential

$$V(x)=-\int^x\mathrm{d}x'\,F(x') \tag{306}$$

is *confining* so that there exists an equilibrium state.

[22] The idea of the Lévy-Khintchine formula is to regularize the $\eta\to0$ behaviour. An alternative version of the Lévy-Khintchine formula is [4, 3]

$$\Lambda(k)=\mathrm{i}k\tilde{\mu}+\frac{1}{2}\sigma^2k^2+\int m(\mathrm{d}\eta)\,\left(1-\mathrm{e}^{-\mathrm{i}k\eta}-\mathrm{i}k\eta\,\theta_{\rm H}(1-|y|)\right)\;. \tag{303}$$

a) Generator of the diffusion

The related FPE was obtained above, Eq. (213). This form is not unfamiliar and recalls the Schrödinger equation in imaginary time $-\partial_t P = H_{\mathrm{FP}} P$, which naturally leads to perform a spectral analysis as we know the importance of spectral analysis in quantum mechanics. Here we have introduced $H_{\mathrm{FP}} = -D\,\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \frac{\mathrm{d}}{\mathrm{d}x}F(x)$, where the notation means that the action of the operator $\frac{\mathrm{d}}{\mathrm{d}x}F(x) = F'(x) + F(x)\frac{\mathrm{d}}{\mathrm{d}x}$ must be understood as acting on a function $\phi(x)$ as $\frac{\mathrm{d}}{\mathrm{d}x}\big[F(x)\phi(x)\big] = F'(x)\phi(x) + F(x)\phi'(x)$. The operator $H_{\mathrm{FP}} = -D\,\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \frac{\mathrm{d}}{\mathrm{d}x}F(x)$ is however not self-adjoint in the presence of the drift,[23] $H_{\mathrm{FP}} \neq H_{\mathrm{FP}}^\dagger$. Instead of the notation H_{FP}, I will prefer a notation used by the mathematicians

$$\partial_t P_t(x) = \mathscr{G}^\dagger P_t(x) \quad \text{where} \quad \mathscr{G}^\dagger = D\,\frac{\mathrm{d}^2}{\mathrm{d}x^2} - \frac{\mathrm{d}}{\mathrm{d}x}F(x) \tag{307}$$

is the "forward generator". By convention, probabilists call the adjoint of this operator

$$\mathscr{G} = D\,\frac{\mathrm{d}^2}{\mathrm{d}x^2} + F(x)\frac{\mathrm{d}}{\mathrm{d}x} \tag{308}$$

the "generator of the diffusion". I will also call it the "backward generator" as we will see that it governs the evolution backward in time. We have used that $\big(\frac{\mathrm{d}}{\mathrm{d}x}\big)^\dagger = -\frac{\mathrm{d}}{\mathrm{d}x}$ (like in quantum mechanics).

We will make several interesting remarks on the operator $\mathscr{G}^\dagger$ by noticing that it can be written as

$$\boxed{\mathscr{G}^\dagger = D\,\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^{-V(x)/D}\,\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^{V(x)/D}} \tag{309}$$

where $F(x) = -V'(x)$.

Remark 1 : equilibrium.— It makes clear that a stationary solution of the FPE, i.e. a solution of $\mathscr{G}^\dagger P = 0$, is

$$P_{\mathrm{eq}}(x) = C_0\,\mathrm{e}^{-V(x)/D} \qquad \Rightarrow \quad \mathscr{G}^\dagger P_{\mathrm{eq}}(x) = 0 \tag{310}$$

where C_0 is a normalisation constant (we assumed above that $V(x)$ is such that this solution is normalizable).

✍ ***Exercice 63*** – **:** *Argue that the solution (310) is an equilibrium solution (hint : analyze the related current).*

✍ ***Exercice 64*** – ***FPE on $\mathbb{R}$ for a non confining potential :*** *Consider the FPE $\partial_t P_t(x) = D\partial_x^2 P_t(x) + \partial_x\big[V'(x)P_t(x)\big]$ on $\mathbb{R}$ such that the drift $F(x) = -V'(x)$ drives the particle from $-\infty$ to $+\infty$. This requires that $V(x\to\pm\infty)\to\mp\infty$.*

1/ Give an example of $V(x)$ and discuss the typical trajectories.

2/ Argue that $\mathscr{G}^\dagger P = 0$ has two independent solutions.

3/ Show that the equilibrium solution is not normalisable and find the expression of the second solution (under the form of an integral).

4/ ***Condition for the NESS***
a) If the stationary solution exists, using the expression found above, show that it presents the asymptotic behaviour $P_{\mathrm{st}}(x) \simeq J/F(x)$ for $x\to+\infty$.

[23] Self-adjointness is not only broken by the drift term ; it can also be broken by the boundary conditions, if they induce a drift at the boundaries.

b) Deduce the condition for existence of the stationary state for the non confining potential.
c) Give a example of non confining drift with a stationary state, and an example without stationary state.

5/ Compare with exercise *50, page 43.*

✍ ***Exercice 65 – The pendulum in the overdamped regime :*** *We consider a pendulum in a fluid, in the overdamped regime, described by the Fokker-Planck equation*

$$\partial_t P_t(\theta) = D\,\partial_\theta^2 P_t(\theta) - \partial_\theta\big[(v - k\,\sin\theta)\,P_t(\theta)\big] \qquad \text{for } \theta \in [-\pi, +\pi] \tag{311}$$

a) Relate the drift to a potential $U(\theta)$ and show that, for $v = 0$, the FPE admits an equilibrium state $P_{\mathrm{eq}}(\theta)$.

b) Argue that this solution is not satisfactory and that there is no equilibrium when $v \neq 0$.

c) To simplify the calculation, we set $D = 1$. Show that the stationary state has the form

$$P_{\mathrm{st}}(\theta) = J\,\psi(\theta)\left[c + \int_\theta^\pi \frac{\mathrm{d}\alpha}{\psi(\alpha)}\right] \qquad \text{where } \psi(\theta) = \mathrm{e}^{-U(\theta)}\,. \tag{312}$$

Find c. What is the physical meaning of J ? Is it a free parameter ? Express J.

d) Analyze the limiting behaviour of J when $v \to 0$; use the modified Bessel function $I_0(z) = \int_0^{2\pi}\frac{\mathrm{d}t}{2\pi}\,\mathrm{e}^{z\cos t}$ (see Appendix for its asymptotic behaviour).

Remark 2 : supersymmetry.— Eigenvalues of $\mathscr{G}^\dagger$ are all *negative*. We can prove this as follows. We first perform the non unitary transformation

$$H_+ = -\mathrm{e}^{V(x)/2D}\mathscr{G}^\dagger\mathrm{e}^{-V(x)/2D} = -D\,\mathrm{e}^{V(x)/2D}\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^{-V(x)/D}\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^{V(x)/2D} \tag{313}$$

which thus relates $-\mathscr{G}^\dagger$ to the self-adjoint operator H_+ (the two operators have the *same spectrum* of eigenvalues). This Hamiltonian has a specific structure

$$\boxed{H_+ = \mathscr{Q}^\dagger\mathscr{Q}} \qquad \text{with } \mathscr{Q} \stackrel{\text{def}}{=} -\sqrt{D}\mathrm{e}^{-V(x)/2D}\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^{V(x)/2D} = \sqrt{D}\left(-\frac{\mathrm{d}}{\mathrm{d}x} + \frac{F(x)}{2D}\right) \tag{314}$$

known as "supersymmetric" (it is possible to introduce the supersymmetric partner $H_- = \mathscr{Q}\mathscr{Q}^\dagger$, the two operators having the same spectrum but the zero mode [24]). The structure $H_+ = \mathscr{Q}^\dagger\mathscr{Q}$ implies that the spectrum of the operator is strictly positive [25]

$$\mathrm{Spec}(H_+) = \mathrm{Spec}(-\mathscr{G}^\dagger) \subset \mathbb{R}_+\,. \tag{315}$$

We have also

$$H_+ = -D\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \frac{F(x)^2}{4D} + \frac{F'(x)}{2} \tag{316}$$

The drift is such that there exists an equilibrium state, hence the effective potential $\mathcal{U}(x) = \frac{F(x)^2}{4D} + \frac{F'(x)}{2}$ is a confining potential and the Hamiltonian has a discrete spectrum :

$$H_+\psi_n(x) = \lambda_n\,\psi_n(x) \qquad \text{with } \lambda_n \geqslant 0\,. \tag{317}$$

In particular, the equilibrium solution $P_{\mathrm{eq}}(x)$ is related to the zero mode of H_+ :

$$H_+\psi_0(x) = 0 \quad \text{with } \psi_0(x) = c_0\,\mathrm{e}^{-V(x)/2D} = \sqrt{P_{\mathrm{eq}}(x)}. \tag{318}$$

where c_0 is a normalisation.

[24] Only one of the two hamiltonians H_+ and H_- may have a zero mode. If none of them possesses a normalizable zero mode, the supersymmetry is said to be broken, cf. book [26].

[25] since $H_+|\,\psi\,\rangle = \lambda|\,\psi\,\rangle$ implies $\lambda = \langle\,\psi\,|\mathscr{Q}^\dagger\mathscr{Q}|\,\psi\,\rangle = ||\mathscr{Q}|\,\psi\,\rangle||^2 \geqslant 0$.

✍ ***Exercice 66 – Ornstein-Uhlenbeck process and the quantum harmonic oscillator :*** *Consider the case $F(x) = -\kappa\, x$. Check that H_+ is the quantum Hamiltonian for the harmonic oscillator. What are the two operators $\mathscr{Q}$ and $\mathscr{Q}^\dagger$?*
Deduce the spectrum of eigenvalues $\{\lambda_n\}_{n\in\mathbb{N}}$ of H_+ and $-\mathscr{G}^\dagger$.

Remark : case of NESS.— We discuss here the case of diffusions with an equilibrium state. If instead the diffusion is characterised by a NESS, the spectral analysis is much more tricky and the spectrum of eigenvalues is not necessary real (hence the mapping on the supersymmetric Hamiltonian is not so helpful as the boundary conditions are not natural from the point of view of the quantum mechanical problem). We can go back to Exercise 49 (or Exercise 70 below). Nonetheless, it is possible to find the expression of the stationary distribution.

✍ ***Exercice 67 – Generalized SUSY :*** *Consider the case of the diffusion for x-dependent diffusion constant*

$$\mathrm{d}x(t) = F(x)\,\mathrm{d}t + \sqrt{2D(x)}\,\mathrm{d}W(t) \qquad \textit{(Stratonovich)} \tag{319}$$

a) Give the generator $\mathscr{G}$ of this diffusion and show that it can be written under a form analogous to (309).

b) How the operators H_+, $\mathscr{Q}$ and $\mathscr{Q}^\dagger$ are generalized ?

c) Deduce the expression of the equilibrium distribution $P_{\mathrm{eq}}(x)$.

d) Assuming the existence of a steady current, give the related stationary distribution $P_{\mathrm{st}}(x)$.

b) Conditional probability (propagator)

An important object characterizing the diffusion is the propagator of the diffusion (the conditional probability), solution of

$$\partial_t P_t(x|x_0) = \mathscr{G}^\dagger P_t(x|x_0) \quad \text{for initial condition } P_0(x|x_0) = \delta(x - x_0)\,. \tag{320}$$

We now consider the case of the drift $F(x)$ such that there exists an equilibrium. Then the spectrum of $\mathscr{G}^\dagger$ (and of H_+) is discrete. Given the spectral information $\{\lambda_n, \psi_n(x)\}$ we can obtain a representation of the propagator.

Method n°1 : we can use the relation to supersymmetric quantum mechanics. The non unitary transformation $P_t(x) = \psi_0(x)\psi(x;t)$, where $\psi_0(x) \propto \mathrm{e}^{-V(x)/2D}$, maps the PDE for the conditional probability onto

$$-\partial_t \Psi_t(x|x_0) = H_+ \Psi_t(x|x_0) \quad \text{for initial condition } \Psi_0(x|x_0) = \delta(x - x_0)\,. \tag{321}$$

The solution can be decomposed over the eigenstates of H_+ (this is the main motivation for spectral analysis!). Starting from the initial condition $\Psi_0(x|x_0) = \sum_{n=0}^\infty \psi_n(x)\psi_n(x_0)$ we get at time t,

$$\Psi_t(x|x_0) = \sum_{n=0}^\infty \psi_n(x)\psi_n(x_0)\,\mathrm{e}^{-\lambda_n t}\,. \tag{322}$$

We go back to the conditional probability. In order to satisfy the initial condition, we must write $P_t(x|x_0) = \psi_0(x)\Psi_t(x|x_0)/\psi_0(x_0)$, thus

$$\boxed{P_t(x|x_0) = \frac{\psi_0(x)}{\psi_0(x_0)} \sum_{n=0}^\infty \psi_n(x)\psi_n(x_0)\,\mathrm{e}^{-\lambda_n t}} \tag{323}$$

Method n°2 : It is more straightforward to manipulate operators : [26]

$$P_t(x|x_0) = \langle\, x\,|\mathrm{e}^{t\mathscr{G}^\dagger}|\, x_0\,\rangle = \langle\, x\,|\mathrm{e}^{-t\psi_0(\hat{x})H_+\psi_0(\hat{x})^{-1}}|\, x_0\,\rangle = \langle\, x\,|\psi_0(\hat{x})\mathrm{e}^{-tH_+}\psi_0(\hat{x})^{-1}|\, x_0\,\rangle$$
$$= \psi_0(x)\langle\, x\,|\mathrm{e}^{-tH_+}|\, x_0\,\rangle\frac{1}{\psi_0(x_0)} \tag{324}$$

where $\hat{x}$ is the "position operator" with $\hat{x}|\,x\,\rangle = x|\,x\,\rangle$. The propagator of H_+ can be decomposed over its eigenstates and we recover (323).

✍ ***Exercice 68*** *– : Check the normalisation* $\int \mathrm{d}x\, P_t(x|x_0) = 1$.
Argue that $\lim_{t\to\infty} P_t(x|x_0) = P_{\mathrm{eq}}(x)$.

This structure makes clear the relation $P_t(x|x_0)\psi_0(x_0)^2 = P_t(x_0|x)\psi_0(x)^2$ i.e.

$$\boxed{P_t(x|x_0)P_{\mathrm{eq}}(x_0) = P_t(x_0|x)P_{\mathrm{eq}}(x)} \tag{325}$$

which is similar to the *detailed balance* condition, which was expected as we are dealing with a situation where an equilibrium exists ($V(x)$ is confining).

Remark : This equation is the consequence of the identity with operators

$$P_{\mathrm{eq}}(\hat{x})^{-1}\,\mathscr{G}^\dagger\,P_{\mathrm{eq}}(\hat{x}) = \mathscr{G}\,. \tag{326}$$

✍ ***Exercice 69 – The Ornstein-Uhlenbeck process and the quantum oscillator :*** *Using the expression of the propagator for the quantum mechanical harmonic oscillator*

$$\langle\, x\,|\mathrm{e}^{-tH_\omega}|\, x_0\,\rangle = \sqrt{\frac{m}{2\pi\omega\sinh\omega t}}\;\exp -\frac{m}{2\omega\sinh\omega t}\left[\cosh\omega t\,(x^2+x_0^2) - 2xx_0\right] \tag{327}$$

for $H_\omega = -\frac{1}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \frac{1}{2}m\omega^2$, *recover the propagator* (153) *of the Ornstein-Uhlenbeck process described by the SDE* $\mathrm{d}x = -\kappa\, x\,\mathrm{d}t + \sqrt{2D}\,\mathrm{d}W(t)$.
Check the condition (325).

c) Solving the FPE without supersymmetry

Above, we have related the FPE $\partial_t P = \mathscr{G}^\dagger P$ to the imaginary time Schrödinger equation $-\partial_t\psi = H_+\psi$ through the non unitary transformation $P_t(x) = \psi_0(x)\psi(x;t)$. We have introduced the self-adjoint operator H_+ for convenience : [27] the spectral analysis could have been performed directly on the non self-adjoint operator $\mathscr{G}^\dagger$, whose spectrum of eigenvalues is obviously the same as H_+.

Let us apply the spectral method to solve directly the FPE $\partial_t P_t(x) = \mathscr{G}^\dagger P_t(x)$. We look for a solution of the "separable" form

$$P_t(x) = \Phi(x)\,\mathrm{e}^{-\lambda t} \quad \text{hence} \quad \mathscr{G}^\dagger\Phi(x) = -\lambda\,\Phi(x)\,. \tag{328}$$

Because the generator is not self adjoint, this last equation is the one for the "right eigenvector" of $\mathscr{G}^\dagger$, whose spectrum involves a bi-orthognoal set of right and left eigenvectors

$$\mathscr{G}^\dagger\Phi_n^{\mathrm{R}}(x) = -\lambda_n\,\Phi_n^{\mathrm{R}}(x) \quad \text{and} \quad \mathscr{G}\Phi_n^{\mathrm{L}}(x) = -\lambda_n\,\Phi_n^{\mathrm{L}}(x) \tag{329}$$

[26] The two operators are related by a non-uniatry transformation of the form $\mathscr{G}^\dagger = -\mathcal{U}H_+\mathcal{U}^{-1}$. Exponentiating the equality we find $\exp\left[t\mathscr{G}^\dagger\right] = \exp\left[-t\mathcal{U}H_+\mathcal{U}^{-1}\right] = \mathcal{U}\exp\left[-t\,H_+\right]\mathcal{U}^{-1}$.

[27] Although this does not occur for confining potential $V(x)$, self adjointness of H_+ may be broken by certain boundary conditions, which makes the spectral analysis of H_+ tricky.

with

$$\int \mathrm{d}x\, \Phi_n^{\mathrm{L}}(x)\Phi_m^{\mathrm{R}}(x) = \delta_{nm} \tag{330}$$

(see the discrete version in Subsection d) page 42). When there exists a stationary state, the spectrum is discrete and the lowest eigenvalue is $\lambda_0 = 0$. The next eigenvalue λ_1 corresponds to the *relaxation rate* toward stationary state (or $\mathrm{Re}(\lambda_1)$ if λ_1 is complex). When there exists an equilibrium, for a confining potential $V(x)$, the spectrum is real.

Let us use this spectral information to solve the FPE for a given initial condition $P_0(x)$. We first decompose this latter on the right eigenvectors

$$P_0(x) = \sum_n c_n\, \Phi_n^{\mathrm{R}}(x) \quad \text{where} \quad c_n = \int \mathrm{d}x\, \Phi_n^{\mathrm{L}}(x) P_0(x) \tag{331}$$

Then, the solution at time t reads

$$P_t(x) = \sum_n c_n\, \Phi_n^{\mathrm{R}}(x)\, \mathrm{e}^{-\lambda_n t}\,. \tag{332}$$

When the initial condition is $P_0(x) = \delta(x - x_0)$, hence $c_n = \Phi_n^{\mathrm{L}}(x_0)$, the solution coincides with the conditional probability. Therefore the spectral decomposition of the conditional probability is

$$P_t(x|x_0) = \sum_{n=0}^{\infty} \Phi_n^{\mathrm{R}}(x)\Phi_n^{\mathrm{L}}(x_0)\, \mathrm{e}^{-\lambda_n t} \tag{333}$$

Relation to supersymmetry.— It is now instructive to make the connection with supersymmetry. Eq. (333) coincides with (323). This shows that the right and left eigenvectors can be simply related to the eigenfunctions of H_+ as follows

$$\Phi_n^{\mathrm{R}}(x) = \psi_0(x)\psi_n(x) \quad \text{and} \quad \Phi_n^{\mathrm{L}}(x) = \frac{\psi_n(x)}{\psi_0(x)} \tag{334}$$

In particular, for $\lambda_0 = 0$,

$$\Phi_0^{\mathrm{R}}(x) = \psi_0(x)^2 = P_{\mathrm{eq}}(x) \quad \text{and} \quad \Phi_0^{\mathrm{L}}(x) = 1\,. \tag{335}$$

Example : We have discussed the relation between the Ornstein-Uhlenbeck process and the QM oscillator, Exercise 69. As a result, in this case

$$\Phi_n^{\mathrm{R}}(x) = c_n\, H_n\left(\sqrt{\frac{k}{2D}}\, x\right) \mathrm{e}^{-\frac{k}{2D}x^2} \quad \text{and} \quad \Phi_n^{\mathrm{L}}(x) = H_n\left(\sqrt{\frac{k}{2D}}\, x\right) \tag{336}$$

where $H_n(x)$ is a Hermite polynomial and c_n a normalisation. Note that the left eigenvectors grow at infinity.

✍ ***Exercice 70 – Diffusion for a uniform drift on a ring :*** *The aim is to obtain the propagator $P_t(x|x_0)$ of the diffusion (213) for a uniform drift $F(x) = F_0$ on a ring, i.e. on the finite interval $[0, L]$ with periodic boundary conditions.*

a) Discuss the spectrum of the forward generator $\mathscr{G}^\dagger = D\partial_x^2 - F_0\partial_x$: eigenvalues, right and left eigenvectors.

b) Write $P_t(x|x_0)$ by using the spectral information. Analyze the $t \to \infty$ limit (identify a characteristic time τ_D).

c) In order to analyze the limit $t \ll \tau_D$, get another representation of the conditional probability from the Poisson formula (521). Discuss the $L \to \infty$ limit.

d) Forward and backward FPE

We have solved above the forward FPE [28]

$$\partial_t P_t(x|x_0) = \mathscr{G}_x^\dagger P_t(x|x_0) \tag{337}$$

where the forward generator is a differential operator acting on the *final* coordinate x. The above discussion makes clear that the generator of the diffusion is involved in the *backward FPE* [29]

$$\partial_t P_t(x|x_0) = \mathscr{G}_{x_0} P_t(x|x_0) \tag{338}$$

where the operator acts on the *initial* coordinate x_0. We will see some applications of this equation below.

✍ ***Exercice 71 – BFPE from FFPE :*** *Deduce (338) from (337) by using (325).*

6.4 Boundary conditions for the FPE

So far we have not discussed the situation where the FPE is solved on a bounded domain. Let us discuss here the question of boundary conditions. For simplicity we consider the FPE $\partial_t P_t(x) = \big[D\partial_x^2 - \partial_x F(x)\big] P_t(x) = -\partial_x J_t(x)$ on $\mathbb{R}^+$ so that we just have one boundary at $x = 0$.

a) Reflecting boundary condition

The first natural boundary condition is the reflecting boundary condition, where the particle coming to $x > 0$ is simply reflected at $x = 0$. This is expressed by the condition of a vanishing current at the origin

$$J_t(0) = F(0)P_t(0) - DP_t'(0) = 0 \tag{339}$$

where $'$ means here derivation with respect to x. In the usual terminology, this corresponds to a "mixed boundary condition". [30]

Remark : reflecting boundary conditions for the conditional probability.— For the following we will have to impose the boundary conditions for the conditional probability $P_t(x|x_0)$. The reflecting boundary condition is

$$\big[D\partial_x - F(x)\big] P_t(x|x_0)\big|_{x=0} = 0 \tag{340}$$

We could aslo ask about the condition with respect to the initial coordinate : Using the relation (325), we have

$$\left(\partial_x - \frac{1}{D}F(x)\right) P_t(x|x_0) = \left(\partial_x + \frac{1}{D}V'(x)\right)\left[\frac{P_\mathrm{eq}(x)}{P_\mathrm{eq}(x_0)} P_t(x_0|x)\right] = \frac{P_\mathrm{eq}(x)}{P_\mathrm{eq}(x_0)}\partial_x P_t(x_0|x) \tag{341}$$

where I used that $P_\mathrm{eq}(x) \propto \exp[-V(x)/D]$. As a consequence, the presence of the reflecting boundary at $x = 0$ implies

$$\partial_{x_0} P_t(x|x_0)\big|_{x_0=0} = 0\,. \tag{342}$$

The reflecting boundary condition is not symmetric for the two coordinates.

[28] This equation, formally $\partial_t P_t = \mathscr{G}^\dagger P_t = P_t \mathscr{G}$, is called the "*Kolmogorov equation*" by mathematicians.

[29] I have used $\langle\,\psi\,|\,A\chi\,\rangle = \langle\,A^\dagger\psi\,|\,\chi\,\rangle$; explicitly $\partial_t P_t(x|x_0) = \langle\,x\,|\mathscr{G}^\dagger \mathrm{e}^{t\mathscr{G}^\dagger}|\,x_0\,\rangle = \mathscr{G}_x^\dagger\langle\,x\,|\mathrm{e}^{t\mathscr{G}^\dagger}|\,x_0\,\rangle = \mathscr{G}_{x_0}\langle\,x\,|\mathrm{e}^{t\mathscr{G}^\dagger}|\,x_0\,\rangle$.

[30] For wave equation for the wave $\psi(x)$, the standard terminology is : (i) Dirichlet boundary condition : $\psi(0) = 0$; (ii) Neumann boundary condition : $\psi'(0) = 0$; (iii) mixed boundary condition : $\psi(0)\cos\theta + \psi'(0)\sin\theta = 0$ (one recovers Dirichlet and Neumann b.c. for $\theta = 0$ and $\theta = \pi/2$, respectively).

b) General boundary condition

Let us now consider a general (mixed) boundary condition

$$\tilde{\lambda}P_t(0) = P_t'(0) \tag{343}$$

For $\tilde{\lambda} = F(0)/D$, this corresponds to the reflecting boundary condition. What is the meaning of this condition for arbitrary real $\tilde{\lambda} \neq F(0)/D$?

Consider

$$\partial_t \int_0^\infty \mathrm{d}x\, P_t(x) = -\int_0^\infty \mathrm{d}x\, \partial_x J_t(x) = J_t(0) \tag{344}$$

The current of probability through $x = 0$ makes the total probability decreases (for $J_t(0) < 0$). Making use of the boundary condition, we get

$$J_t(0) = \left[F(0) - D\tilde{\lambda}\right] P_t(0) \equiv -\lambda\, P_t(0) \tag{345}$$

where we have found convenient to introduce $\lambda = D\tilde{\lambda} - F(0)$. For $\lambda > 0$, the total probability decreases

$$\partial_t \int_0^\infty \mathrm{d}x\, P_t(x) = -\lambda\, P_t(0)\,. \tag{346}$$

Hence for λ has roughly the meaning of the rate of escape, when the particle reaches the boundary at $x = 0$.

c) Absorbing boundary condition

Writing $P_t(0) = \tilde{\lambda}^{-1}\, P_t'(0)$ shows that the limit $\tilde{\lambda} \to \infty$, or $\lambda \to \infty$, corresponds to a Dirichlet boundary condition

$$P_t(0) = 0 \tag{347}$$

corresponding physically to the situation where the particle reaching the boundary is absorbed with probability one.

6.5 Random walk on the lattice : persistence and recurrence

In this section we discuss several important properties of the Brownian motion in d-dimensions. It will be convenient to discretize the problem and consider a random walk on a lattice (for simplicity we consider the square lattice), which will regularize some problems.

a) Propagator for the Markov chain

Denote by $\{\vec{e}_i\}_{i=1,\cdots,d}$ the basis of orthonormal vectors of the lattice : for the square lattice which we consider below, $\vec{e}_i \cdot \vec{e}_j = \delta_{ij}$. At time $t = 0$, the walker is at site $\vec{0}$ and at each time interval ($\delta t = 1$), the walker jumps on one of the $2d$ nearest neighbour sites with equal probability $1/(2d)$. The jumps are independent, hence the walk is an example of *Markov chain*. Denote by $P_t(\vec{x})$ the probabilty for the walker to be on site $\vec{x} \in \mathbb{Z}^d$ at time $t \in \mathbb{N}$. The probability to be at $\vec{x}$ at time $t+1$ is related to the probability to be on one of the neighbouring sites at time t :

$$P_{t+1}(\vec{x}) = \sum_{i=1}^{d} \frac{1}{2d}\Big[P_t(\vec{x}+\vec{e}_i) + P_t(\vec{x}-\vec{e}_i)\Big]\,, \quad \text{with initial condition } P_0(\vec{x}) = \delta_{\vec{x},\vec{0}}\,. \tag{348}$$

Thus this solution is in fact the condional probability $P_t(\vec{x}|\vec{0})$, or the "propagator" of the random walk, however we keep the lighter notation $P_t(\vec{x})$.

The most simple manner to solve the equation is to consider the Fourier transform. Because the position is on the lattice, $\vec{x} \in \mathbb{Z}^d$, we have to deal with the discrete Fourier transform

$$\widetilde{P}_t(\vec{k}) = \sum_{\vec{x}} P_t(\vec{x})\,\mathrm{e}^{-\mathrm{i}\vec{k}\cdot\vec{x}} \tag{349}$$

with the wave vector in the Brillouin zone $\vec{k} \in [-\pi, +\pi]^d \equiv \mathrm{ZdB}$. Taking the Fourier transform of the master equation we obviously get $\widetilde{P}_{t+1}(\vec{k}) = \left(\frac{1}{d}\sum_i \cos k_i\right)\widetilde{P}_t(\vec{k})$ leading to the solution

$$P_t(\vec{x}) = \int_{\mathrm{ZdB}} \frac{\mathrm{d}^d\vec{k}}{(2\pi)^d} \left(\frac{1}{d}\sum_{i=1}^{d} \cos k_i\right)^t \mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{x}}\,. \tag{350}$$

It will be useful to note that, formally, we can rewrite the Master equation (348)

$$P_{t+1}(\vec{x}) = \sum_{\vec{x}\,'} \widetilde{\Delta}_{\vec{x},\vec{x}\,'} P_t(\vec{x}\,') \tag{351}$$

where $\widetilde{\Delta}$ ressembles the discrete "Laplacian" [31]. This matrix plays the role of the stochastic matrix of the Markov chain. We can then rewrite the solution as

$$P_t(\vec{x}) = \left(\widetilde{\Delta}^t\right)_{\vec{x},\vec{0}} \tag{352}$$

which makes clear that Eq. (350) is just the spectral representation of this latter equation.

b) Green's function

Above we have introduced the Fourier transform (over space variable) of the distribution, now we introduce its Laplace transform (over time), which we denote the Green's function :

$$G(\vec{x}; z) \stackrel{\text{def}}{=} \sum_{t=0}^{\infty} z^t\, P_t(\vec{x}) \tag{353}$$

which will be convenient for the following. The reason of this terminology is due to the fact that it obeys

$$G(\vec{x}; z) - z\sum_{\vec{x}\,'} \widetilde{\Delta}_{\vec{x},\vec{x}\,'} G(\vec{x}\,'; z) = \delta_{\vec{x},\vec{0}} \tag{354}$$

which we can deduce from (351). The solution is formally

$$G(\vec{x}; z) = \left(\frac{1}{1 - z\widetilde{\Delta}}\right)_{\vec{x},\vec{0}} \tag{355}$$

which we can also directly obtain from (352), writing : $G(\vec{x}; z) = \sum_{t=0}^{\infty} z^t \left(\widetilde{\Delta}^t\right)_{\vec{x},\vec{0}}$. We now want to analyze the Green's function with the help of its spectral representation

$$G(\vec{x}; z) = \int_{\mathrm{ZdB}} \frac{\mathrm{d}^d\vec{k}}{(2\pi)^d} \frac{\mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{x}}}{1 - \frac{z}{d}\sum_{i=1}^{d} \cos k_i}\,. \tag{356}$$

Let us discuss few cases, when the integral can be computed explicitly.

[31] The discrete Laplace operator is defined as $(\Delta f)(\vec{x}) = \sum_{i=1}^{d}\left[f(\vec{x}+\vec{e}_i) - 2\,f(\vec{x}) + f(\vec{x}-\vec{e}_i)\right]$. Thus $\Delta = \widetilde{\Delta} - 2d\,\mathbf{1}$.

Case $d = 1$ **:** We should compute

$$G(x;z) = \int_{-\pi}^{+\pi} \frac{\mathrm{d}k}{2\pi} \, \frac{\mathrm{e}^{\mathrm{i}kx}}{1 - z\,\cos k} \tag{357}$$

for $x \in \mathbb{Z}$. Using the integral $\int_{-\pi}^{+\pi} \frac{\mathrm{d}k}{2\pi} \frac{\cos(nk)}{\cosh a + \cos k} = \frac{\mathrm{e}^{-a|n|}}{\sinh a}$, with $\cosh a = 1/z$, we find

$$G(x;z) = \frac{1}{\sqrt{1-z^2}} \left(\frac{1}{z} - \sqrt{\frac{1}{z^2} - 1} \right)^{|x|} \tag{358}$$

We can use the formula $(1-x)^{-1/2} = \sum_{n=0}^{\infty} \frac{(2n-1)!!}{2^n n!} \, x^n$ to expand $G(0;z)$ in powers of z and deduce

$$P_t(0) = \frac{(t-1)!!}{2^{t/2}(t/2)!} \qquad \text{for } t \text{ even} \tag{359}$$

and $P_t(0) = 0$ for t odd. Note that $P_t(x)$ is related to the binomial distribution and can be found explicitly by simpler means. [32]

Case $d = 2$ **:** We can at least get the Green's function at the origin

$$G(\vec{0};z) = \frac{2}{z} \int_{-\pi}^{+\pi} \frac{\mathrm{d}k_x \mathrm{d}k_y}{(2\pi)^2} \, \frac{1}{2/z - \cos k_x - \cos k_y} = \frac{2}{\pi} \, \mathrm{K}(z) \tag{360}$$

where $\mathrm{K}(z)$ is the elliptic integral (see the appendix). In particular, it will be useful for the following to notice that

$$G(\vec{0};z) \underset{z\to 1^-}{\simeq} \frac{1}{\pi} \, \ln[8/(1-z)] \tag{361}$$

General case : We can get a simpler integral representation for the Green's function at $\vec{x} = \vec{0}$: the difficulty with the above integral representation is to deal with the multiple integral. They can be decoupled with the following trick

$$G(\vec{0};z) = \int_{\mathrm{ZdB}} \frac{\mathrm{d}^d\vec{k}}{(2\pi)^d} \, \frac{1}{1 - \frac{z}{d}\sum_{i=1}^d \cos k_i} = \int_{\mathrm{ZdB}} \frac{\mathrm{d}^d\vec{k}}{(2\pi)^d} \int_0^\infty \mathrm{d}t \, \mathrm{e}^{-t\left(1 - \frac{z}{d}\sum_{i=1}^d \cos k_i\right)} \tag{362}$$

$$= \int_0^\infty \mathrm{d}t \, \mathrm{e}^{-t} \left[\int_{-\pi}^{+\pi} \frac{\mathrm{d}k}{2\pi} \, \mathrm{e}^{\frac{zt}{d}\cos k} \right]^d \tag{363}$$

We recognize the modified Bessel function (see the appendix)

$$G(\vec{0};z) = \int_0^\infty \mathrm{d}t \, \mathrm{e}^{-t} \left[I_0\left(zt/d\right) \right]^d \ . \tag{364}$$

This integral form will be useful in the following.

[32] Consider a symmetric random walk on $\mathbb{Z}$ with only nearest neighbour jumps. The solution is given by the binomial distribution $P_t(x) = \frac{1}{2^t}\frac{t!}{n_+! n_-!} = \frac{1}{2^t}\frac{t!}{\left(\frac{t+x}{2}\right)!\left(\frac{t-x}{2}\right)!}$.

c) Continuum limit

Propagator.— At this step it is interesting to discuss the continuum limit. For this purpose, we introduce the lattice spacing $a \to 0$ and the time interval $\delta t \to 0$ in the probability and consider the probability density in $\mathbb{R}^d$

$$\mathcal{P}_t(\vec{x}) = \frac{1}{a^d}\, P_{t/\delta t}(\vec{x}/a) = \int_{-\pi/a}^{\pi/a} \frac{\mathrm{d}^d\vec{k}}{(2\pi)^d}\, \exp\left\{ \frac{t}{\delta t} \ln\left(\frac{1}{d}\sum_{i=1}^{d} \cos(k_i a) \right) + \mathrm{i}\vec{k}\cdot\vec{x} \right\} \tag{365}$$

$$\underset{a\to 0}{\simeq} \int \frac{\mathrm{d}^d\vec{k}}{(2\pi)^d}\, \exp\left\{ -\frac{ta^2}{2d\delta t}\vec{k}^{\,2} + \mathcal{O}(a^4) + \mathrm{i}\vec{k}\cdot\vec{x} \right\} \simeq \frac{1}{(4\pi D t)^{d/2}}\, \mathrm{e}^{-\frac{\vec{x}^{\,2}}{4Dt}} \tag{366}$$

where

$$D = \frac{a^2}{2d\delta t} \tag{367}$$

is the diffusion constant. The continuum limit of the random walk on the lattice is the Brownian motion in space, which is obtained by sending both $a \to 0$ and $\delta t \to 0$, keeping the ratio $a^2/\delta t$ finite.

Properties of the paths : on the lattice and in the continuum.— We can more conveniently study the properties of the paths within the continuum description. For example we have

$$\langle \vec{x}^{\,2} \rangle = 2d\, Dt\,. \tag{368}$$

Hence the typical distance covered by the path after time t is $\ell_t \sim \sqrt{d\,Dt}$. This should not be confused with the length of the path $\mathcal{L} = a\,(t/\delta t) \sim d\,Dt/a \sim \ell_t^2/a$ which becomes infinite in the continuum limit $a \to 0$. This reminds us that the continuum limit of the random walk is a *non differentiable* continuous path. Indeed, in 1D, the property $\langle [x(t+\delta t) - x(t)]^2 \rangle = 2D\delta t$ implies

$$\langle \big[(x(t+\delta t) - x(t))/\delta t\big]^2 \rangle \propto 1/\delta t \underset{\delta t\to 0}{\longrightarrow} \infty\,. \tag{369}$$

Finally, let us emphasize that studying the random walk over long time $t \to \infty$ and large scale $|\vec{x}| \to \infty$ is formally equivalent to the continuum limit defined above. In this case, the continuum limit is obtained by a coarse graining procedure.

Green's function.– We can also analyze the Green's function in the continuum limit. In this case I prefer to define

$$\widehat{\mathcal{P}}(\vec{x};s) = \int_0^\infty \mathrm{d}t\, \mathrm{e}^{-st}\, \mathcal{P}_t(\vec{x}) \tag{370}$$

which thus obeys to the equation

$$(s - D\,\Delta)\, \widehat{\mathcal{P}}(\vec{x};s) = \delta(\vec{x}) \tag{371}$$

which is reminiscent of Eq. (354). It is useful to rewrite it as

$$\widehat{\mathcal{P}}(\vec{x};s) = \langle \vec{x}\,| \frac{1}{s - D\Delta} |\,\vec{0}\rangle \tag{372}$$

where Δ is the usual Laplace operator. The relation with the Green's function defined above is $G(\vec{x};\mathrm{e}^{-s}) \simeq \widehat{\mathcal{P}}(\vec{x};s)$ for $s \to 0$. While for the discrete RW it was only possible to get an integral representation for the Green's funtion at the origin, for the continuous BM one can obtain a simple analytic form in any dimension and arbitrary position. Introducing the expression of the propagator in (370) we deduce

$$\widehat{\mathcal{P}}(\vec{x};s) = \frac{1}{(2\pi)^{d/2}D}\left(\sqrt{\frac{s}{D}}\frac{1}{|\vec{x}|}\right)^{\frac{d-2}{2}} K_{\frac{d-2}{2}}\left(\sqrt{\frac{s}{D}}\,|\vec{x}|\right) \tag{373}$$

where $K_\nu(z)$ is the MacDonald function (modified Bessel function of third kind, see the appendix). Using the limiting behaviour given below we deduce

$$\widehat{\mathcal{P}}(\vec{x};0) = \frac{\Gamma\left(\frac{d-2}{2}\right)}{2(2\pi)^{d/2}D}\,|\vec{x}|^{-d+2} \tag{374}$$

for $d \neq 2$. In two dimensions we have $\widehat{\mathcal{P}}(\vec{x};0) = \frac{1}{2\pi D}\ln|\vec{x}|$, which can be obtained by letting $d \to 2$ continuously in the previous expression (this is called "dimensional regularisation").

The Green's function at the initial point will be useful below. We see that, in the continuum limit, it is divergent at $\vec{x}=0$, whereas it is finite on the lattice.

d) Probability of first return

We now turn to a more subtle property of the random trajectories. We have analysed above in great detail the probability $P_t(\vec{0})$ to return at the starting point after time t. We now study the probability of *first* return after time t, which we denote Q_t. The probility $P_t(\vec{0})$ encodes trajectories returning for the first time, second time, etc. Hence it can be written as

$$P_t(\vec{0}) = \delta_{t,0} + Q_t + \sum_{t_1=0}^{t} Q_{t_1}Q_{t-t_1} + \sum_{\substack{t_1,t_2,t_3\\ \text{with } t_1+t_2+t_3=t}} Q_{t_1}Q_{t_2}Q_{t_3} + \cdots = \delta_{t,0} + \sum_{t_1=0}^{t} Q_{t_1}P_{t-t_1}(\vec{0}) \tag{375}$$

where the first term account for the initial condition $P_0(\vec{0}) = 1$. Here we have chosen by convention $Q_0 = 0$ because the particle can only return to its starting point for time $t \geqslant 2$.

We can easily solve the equation by introducing the generating function

$$\widetilde{Q}(z) \stackrel{\text{def}}{=} \sum_{t=0}^{\infty} z^t\,Q_t \tag{376}$$

which is a kind of discrete Laplace transformation (setting $z = \mathrm{e}^{-s}$, s is the Laplce parameter). This convert the convolution in (375) into multiplication. The "Laplace" transform of (375) is

$$G(\vec{0};z) = 1 + \widetilde{Q}(z)\,G(\vec{0};z) \tag{377}$$

thus

$$\widetilde{Q}(z) = 1 - \frac{1}{G(\vec{0};z)}\,. \tag{378}$$

Case $d=1$: From the expression of the Green's function found above, we get

$$\widetilde{Q}(z) = 1 - \sqrt{1-z^2}\,. \tag{379}$$

Note that $\widetilde{Q}(1) = 1$ corresponding to the normalisation condition $\sum_t Q_t = 1$. Using $(1-x)^{1/2} = 1 - \frac{1}{2}x - \sum_{n=2}^{\infty}\frac{(2n-3)!!}{2^n n!}\,x^n$ we deduce the form

$$\widetilde{Q}(z) = \frac{1}{2}z^2 + \sum_{n=2}^{\infty}\frac{(2n-3)!!}{2^n n!}\,z^{2n} \tag{380}$$

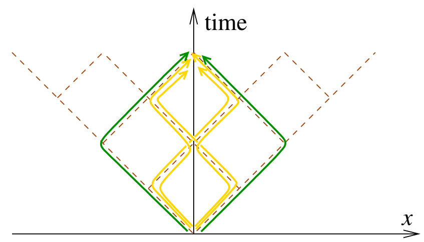


Figure 25: *After time $t=4$, among the $2^4=16$ trajectories, 6 come back to the origin, $P_4(0)=3/8$. Among those six, the four yellow trajectories visit twice the origin and the two green trajectories return for the first time to the starting point, thus $Q_4=1/8$.*

thus $Q_2 = P_2(0) = 1/2$ and

$$Q_t = \frac{(t-3)!!}{2^{t/2}(t/2)!} \qquad \text{for } t \text{ even} \tag{381}$$

and $Q_t = 0$ for t odd. For example, one can check that $Q_4 = 1/8$ as it should (whereas $P_4(0) = 1/4$) : cf. Fig. 25. Using the Stirling formula, we get the asymptotic

$$P_t(0) \simeq \sqrt{\frac{2}{\pi t}} \tag{382}$$

and, since $Q_t = P_t(0)/(t-1)$ we find

$$\boxed{Q_t \underset{t\to\infty}{\simeq} \sqrt{\frac{2}{\pi}}\,\frac{1}{t^{3/2}}} \tag{383}$$

Remark : It is also interesting to get this expression by another method and make the connection with the $z \to 1$ behaviour of $\widetilde{Q}(z)$. For this purpose it is convenient to consider

$$\widetilde{Q}(\mathrm{e}^{-s}) \underset{s\to 0}{\simeq} 1 - \sqrt{2s}\,. \tag{384}$$

One can show that this behaviour is indeed related to the $t^{-3/2}$ tail at large time (we have emphasize this point in § c) page 8). Let us recall the argument : the generating function takes the form of a Laplace transform for $s \to 0$

$$\widetilde{Q}(\mathrm{e}^{-s}) = \sum_t \mathrm{e}^{-st}\, Q_t = 1 - \sum_t Q_t\,(1-\mathrm{e}^{-st}) \simeq 1 - \frac{1}{2}\int_0^\infty \mathrm{d}t\, Q_t\,(1-\mathrm{e}^{-st}) \tag{385}$$

where the $1/2$ stands from the fact that $Q_t = 0$ for t odd.

Assume the power law tail $Q_t \simeq A/t^{\alpha+1}$ for $t \to \infty$. In the $s \to 0$ limit, the integral selects the tail and one has

$$\int_0^\infty \mathrm{d}t\, Q_t\,(1-\mathrm{e}^{-st}) \simeq \frac{A\,s}{\alpha}\int_0^\infty \frac{\mathrm{d}t}{t^\alpha}\,\mathrm{e}^{-st}$$

thus $\widetilde{Q}(\mathrm{e}^{-s}) \simeq 1 - \frac{A}{2\alpha}\Gamma(1-\alpha)\, s^\alpha$ for $s \to 0$. Comparing with Eq. (384) shows that $\alpha = 1/2$ and $A\,\Gamma(1-\alpha)/(2\alpha) = \sqrt{2}$, leading to $A = \sqrt{2/\pi}$. Hence we recover precisely the behaviour $Q_t \simeq \sqrt{2/\pi}\, t^{-3/2}$ obtained above, Eq. (383).

Case $d = 2$: In two dimensions, we have obtained above $G(\vec{0}; \mathrm{e}^{-s}) \simeq \frac{1}{\pi}\ln(8/s)$ for $s \to 0$, thus

$$\widetilde{Q}(\mathrm{e}^{-s}) \simeq 1 - \frac{\pi}{\ln(8/s)} \tag{386}$$

Using similar argument at for the 1D case, i.e. starting from (385), one finds that the first return probability presents the tail [33]

$$\boxed{Q_t \underset{t\to\infty}{\simeq} \frac{2\pi}{t\,[\ln(8t)]^2}} \tag{387}$$

e) Recurrence

In the last §, I have discussed the *persitence* of the random walk : the probability to return at the origin for the first time at time t is Q_t. The probability to return at the starting point *at any time* is $\mathscr{P}_r = Q_2 + Q_4 + Q_6 + \cdots$, thus

$$\mathscr{P}_r \stackrel{\text{def}}{=} \sum_{t=0}^{\infty} Q_t = \widetilde{Q}(1) = 1 - \frac{1}{G(\vec{0};1)} \,. \tag{388}$$

In dimension $d = 1$ and $d = 2$, we have obtained above that $G(\vec{0}; 1) = \infty$, thus

$$\mathscr{P}_r = 1 \qquad \text{in } d = 1,\ 2 \tag{389}$$

The random walk is said to be *recurrent*, as the random walk eventually comes back to the initial point with probability one after a certain time.

In higher dimension, we have $G(\vec{0}; 1) < \infty$ and thus

$$\mathscr{P}_r < 1 \qquad \text{in } d > 2\,, \tag{390}$$

i.e. the random walk has a finite probability to never comes back to its starting point after infinite time. The RW is said to be *transient*.

Using the integral form (364) given above, we obtain

d	$\mathscr{P}_r$
1	1
2	1
3	0.3405
4	0.1932
5	0.1352
⋮	⋮

As the dimension increases, the probability to ever return to the starting point diminishes.

Bibliography : this discussion was borrowed from the book of Itzykson and Drouffe [25] (volume 1).

6.6 First passage and exit problem (in 1D)

We now study the recurrence for general diffusions in one-dimension (the restriction to dimension one makes the calculations simple ; the method extends easily to dimensions $d > 1$).

[33] We write the Laplace transform as $\widetilde{Q}(\mathrm{e}^{-s}) \simeq 1 - \frac{1}{2}\int_0^\infty \mathrm{d}t\, Q_t\,(1 - \mathrm{e}^{-st}) \approx 1 - \frac{1}{2}\int_{1/s}^\infty \mathrm{d}t\, Q_t$.

a) Persistence of the free Brownian motion

Here, I come back to the problem studied in § 6.5 : One studies (in the exercise) the question of the first return of the free Brownian motion. This was studied above for the discrete random walk, this is studied here for a continuous Brownian motion.

✍ ***Exercice 72 – Persistence, first passage time and maximum of the BM :***
1/ Propagator on the half line.– *We consider the free diffusion on $\mathbb{R}_+$ with a Dirichlet boundary condition at the origin. Construct the solution of the diffusion equation*

$$\partial_t P_t(x) = D\partial_x^2 P_t(x) \qquad \text{for } x>0 \text{ with } P_t(0)=0 \tag{391}$$

(use the image method). Apply the method to get the propagator of the diffusion on $\mathbb{R}_+$, denoted $P_t^+(x|x_0)$.

2/ Survival probability.– *Dirichlet boundary condition describes absorption at $x=0$. Compute the survival probability for a particle starting from x_0 :*

$$S_{x_0}(t) = \int_0^\infty \mathrm{d}x\, P_t^+(x|x_0) \tag{392}$$

What would have been the result if $P_t^+(x|x_0)$ would have satisfied a Neunmann boundary condition ?

3/ First passage time.– *We denote by T_{x_0} the first time at which the process starting from $x_0>0$ reaches $x=0$ (it is a random quantity depending on the process), and $\mathscr{P}_{x_0}(T)$ is distribution. Argue that*

$$S_{x_0}(t) = \int_t^\infty \mathrm{d}T\, \mathscr{P}_{x_0}(T) \tag{393}$$

Deduce $\mathscr{P}_{x_0}(T)$ and plot it.

4/ Maximum of a BM.– *We now consider another property of the Brownian motion $x(\tau)$ with $\tau\in[0,t]$ starting from $x_0=0$: we denote by $M=\underset{\tau\in[0,t]}{\mathrm{Max}}(x(\tau))\geqslant 0$ its maximum and $W_t(m)$ the corresponding distribution. Justify the following identity*

$$\int_0^m \mathrm{d}m'\, W_t(m') = S_m(t) \tag{394}$$

Deduce the expression of $W_t(m)$. What does $W_t(0)$ represent ? The exponent of the power law $t^{-\theta}$ is called the persistence exponent. Give θ for the Brownian motion.

Hint : use appendix with properties of error function.

Remark : we have recovered the results obtained within the discrete model of random walk (§ 6.5) : the probability to return to the starting point is $P_t(0|0)\sim t^{-1/2}$ and the probability for the *first* return is $\mathscr{P}_{x_0}(t)\sim t^{-3/2}$ at large time.

Conclusion : Two important points :

- in order to study the first passage time at the origin, one should impose an absorbing boundary condition at $x=0$.
- The definition of the survival probability suggests that one should first find the conditional probability, then integrate it to get the survival probability. In fact, the backward FPE provides a shortcut and allows to find an equation directly for the survival probability, as we will see.

b) First passage time for arbitrary drift

We consider the case of a diffusion with drift and uniform diffusion constant for simplicity

$$\mathrm{d}x(t) = F(x)\,\mathrm{d}t + \sqrt{2D}\,\mathrm{d}W(t) \tag{395}$$

The drift derives from a potential $F(x) = -V'(x)$ and $W(t)$ is the Wiener process. Consider that the diffusion starts from the initial condition $x(0) = x_0$.

The determination of the propagator $P_t(b|x_0)$ allows to answer the question : what is the probability that the process reaches the point $x = b$ in a (fixed) time t (i.e. the final position is the random variable). We now ask a dual question : **what is the time T_{x_0} needed to reach the point $x = b$ for the first time ?** $x(0) = x_0$ and $x(T_{x_0}) = b$ with $x(t) < b$ for $0 < t < T_{x_0}$. Hence we now fix the final position ($x = b$) and study the statistical properties of the random time T_{x_0}. We denote by $\mathscr{P}_{x_0}(T)$ its distribution.

The main idea is to introduce an **absorbing boundary** at $x = b$:

$$P_t(b|x_0) = P_t(x|b) = 0 \tag{396}$$

implying that the particle is absorbed when it reaches $x = b$. For simplicity for future calculations and analysis, we impose a *reflecting boundary condition* at another point $x = a$, i.e. we impose that the current vanishes

$$\left(F(x) - D\partial_x\right) P_t(x|x_0)\big|_{x=a} = \partial_{x_0} P_t(x|x_0)\big|_{x_0=a} = 0 \tag{397}$$

The boundary condition takes a different form with respect to the two arguments (this is expecetd as $P_t(x|x_0)$ is not a symmetric function of its two arguments in general, cf. § 6.4 page 73).

We introduce the survival probability

$$S_{x_0}(t) = \int_a^b \mathrm{d}x\, P_t(x|x_0)\,, \tag{398}$$

the probabilty that the particle has survived up to time t, i.e. has not reached the absorbing boundary at $x = b$. This is also the probability for the particle to be absorbed after time t

$$S_{x_0}(t) = \mathrm{Proba}\{T_{x_0} \geqslant t\} = \int_t^\infty \mathrm{d}T\, \mathscr{P}_{x_0}(T)\,. \tag{399}$$

Then $\mathscr{P}_{x_0}(T) = -\partial_T S_{x_0}(T)$. Because we integrate over the final position x involved in the propagator, we see that it is interesting to make use of the *backward* FPE (338) :

$$\partial_t S_{x_0}(t) = \int_a^b \mathrm{d}x\, \mathscr{G}_{x_0} P_t(x|x_0) = \mathscr{G}_{x_0} S_{x_0}(t) \tag{400}$$

for the initial condition

$$S_{x_0}(0) = \begin{cases} 1 & \text{for } x_0 \in [a, b[\\ 0 & \text{for } x_0 \geqslant b \end{cases}\,. \tag{401}$$

Similarly the first passage time distribution obeys

$$\partial_t \mathscr{P}_{x_0}(t) = \mathscr{G}_{x_0} \mathscr{P}_{x_0}(t) \tag{402}$$

At this point it is useful to introduce the n-th moment of the time :

$$T_n(x_0) \stackrel{\text{def}}{=} \left\langle \left(T_{x_0}\right)^n \right\rangle = \int_0^\infty \mathrm{d}T\, T^n\, \mathscr{P}_{x_0}(T) \tag{403}$$

✍ ***Exercice 73 – Moments of the first passage time :***
a) Show that the moments obey the recurrence

$$\mathscr{G}_{x_0}\, T_n(x_0) = -n\, T_{n-1}(x_0) \quad \text{and} \quad \mathscr{G}_{x_0}\, T_1(x_0) = -1\,. \tag{404}$$

b) Justify that the boundary conditions are $\partial_{x_0} T_n(x_0)\big|_{x_0=a} = 0$ *and* $T_n(b) = 0$.
c) Deduce (calculation requires to solve a first order linear differential equation: easy!)

$$\boxed{T_n(x_0) = \frac{n}{D}\int_{x_0}^{b} \mathrm{d}x\, \mathrm{e}^{V(x)/D} \int_a^x \mathrm{d}x'\, \mathrm{e}^{-V(x')/D}\, T_{n-1}(x')} \tag{405}$$

✍ ***Exercice 74 – :*** *Generalize (405) for a x-dependent diffusion constant $D \to D(x)$.*

c) Arrhenius law

An important application of the above formalism is the analysis of the escape time for a particle trapped in a potential well. This problem is relevant in chemistry where chemical reactions are activated by overcoming some potential (activation) barriers in the configuration space of the molecules. For simplicity we consider a one-dimensional problem of a particle initially in a potential well (Fig. 26) : x_0 is close to the local minimum at x_1. We study the time needed to escape the well, i.e. jump in the region $x > x_2$.

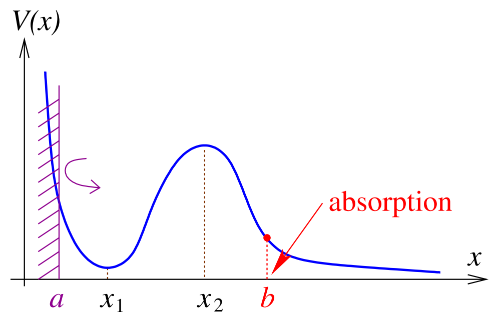


Figure 26: *A particle escapes from a metastable state.*

In a rather arbitrary manner, we introduce a reflecting boundary at $x = a$ at the left of the local minimum, and the absorbing boundary at $x = b$ at the right of the potential barrier (not too close from the top). As we have seen above the average time is given by

$$T_1(x_0) = \frac{1}{D}\int_{x_0}^{b} \mathrm{d}x\, \mathrm{e}^{V(x)/D} \int_a^x \mathrm{d}x'\, \mathrm{e}^{-V(x')/D}\,. \tag{406}$$

The integral can be analysed by using the steepest descent method. For $D \to 0$, the integral over x is dominated by the neighbourhood of $x = x_2$, hence we can replace the upper bound of the second integral $\int_a^x \to \int_a^{x_2}$, expand the potential in the exponential $\mathrm{e}^{V(x)/D} \simeq \exp\left\{\frac{1}{D}[V(x_2) - \frac{1}{\delta_2^2}(x-x_2)^2]\right\}$, where $\delta_2 = 1/\sqrt{-V''(x_2)}$, and perform the remaining Gaussian integral. Similarly, the integral over x' is dominated by the neighbourhood of $x' = x_1$; expanding similarly the integrand as $\mathrm{e}^{-V(x')/D} \simeq \exp\left\{\frac{1}{D}[-V(x_1) - \frac{1}{\delta_1^2}(x'-x_1)^2]\right\}$, where $\delta_1 = 1/\sqrt{V''(x_1)}$, we end with

$$\boxed{\langle T_{x_0}\rangle \equiv T_1(x_0) \simeq 2\pi\,\delta_1\,\delta_2\,\exp\frac{V(x_2)-V(x_1)}{D}} \tag{407}$$

The main result is that the average time is exponentially large in the height of the potential barrier $\Delta V = V(x_2) - V(x_1)$. It is pretty independent of x_0 (provided that it remains in the well) : for x_0 in the well, the particle is rapidly driven at the bottom of the well, where it is submitted to the fluctuations (the time scale is controlled by the curvature at x_1, like for the Ornstein-Uhlenbeck process) ; then it takes a long time to escape the well, thanks to large (and thus rare) thermal fluctuation.

Figure 27: *The swedish chemist Svante August Arrhenius (1859-1927), Nobel prize in chemistry in 1903.*

✍ ***Exercice 75 – :*** *How far from the top of the barrier (at x_2) must be the absorbing boundary b so that the previous analysis is justified ? And how far the reflecting boundary at a should be from the bottom of the well (at x_1) ?*

We can also analyze higher moments : applying the same arguments to (405) we get

$$T_n(x_0) \simeq n\, T_{n-1}(x_1)\, T_1(x_0) \tag{408}$$

for $D \to 0$. Using the independence in the initial position, we conclude that the moments are

$$T_n(x_0) \simeq n!\; [T_1(x_0)]^n \tag{409}$$

i.e. those of a Poisson distribution.

$$\boxed{\mathscr{P}_{x_0}(T) \simeq \frac{1}{\langle T_{x_0}\rangle} \exp - \frac{T}{\langle T_{x_0}\rangle}} \tag{410}$$

The exponential distribution was expected as in the $D \to 0$ limit, the particle is trapped a long time in the well, hence has time to decorrelate : the picture is that, starting from the initial position x_0, the particle falls after a short time in the vicinity of the minimum of the potential well $x(t) \sim x_1$. There, fluctuations are $\delta x \sim \sqrt{D}/\delta_1$. As long as the process remains in the well, it is approximatively described by the Langevin equation $\frac{\mathrm{d}}{\mathrm{d}t}x(t) \approx -\frac{1}{\delta_1^2}(x-x_1) + \sqrt{2D}\,\eta(t)$, where we have linearized $F(x) = -V'(x)$ near x_1. The correlation function is $\langle x(t)x(t')\rangle_c \simeq (D/\delta_1^2)\, \exp\big[-|t-t'|/\delta_1^2\big]$ (cf. chapter 3 on Langevin equation). Indeed, the decorrelation time is $\sim \delta_1^2$, which is exponentially smaller than the typical time to escape the well. This shows that the escape process can be approximatively considered Markovian, hence the exponential distribution (410).

Remark: this discussion is inspired by the book of Gardiner [18] and by the appendix of my paper [49], where an application for the statistics of energy levels in a quantum (Anderson) localisation [34] problem is discussed.

[34] Anderson localisation is the problem of localisation of a wave in a (static) random medium.

✍ ***Exercice 76 – Time needed to fall at the bottom of a harmonic well:*** *We discuss the situation where the initial point x_0 is far from the minimum of the well at $x = x_1$ and clarify a point of the previous discussion. We consider the Ornstein-Uhlenbeck process $\frac{\mathrm{d}}{\mathrm{d}t}x(t) = -\lambda\,(x - x_1) + \sqrt{2D}\,\eta(t)$. What is the typical time needed by a particle initially far from the minimum, $x(0) - x_1 = \Delta$ "large", to fall in the potential well ? Compare to the Arrhenius time.*

✍ ***Exercice 77 – Lifetimes of metastable states:*** *We have obtained above the following formula for the average lifetime of a metastabel state corresponding to the well of Fig. 26 : $\langle T_{x_0}\rangle \simeq \frac{2\pi}{\sqrt{-V''(x_1)V''(x_2)}}\,\exp\left\{\frac{V(x_2)-V(x_1)}{D}\right\}$, valid in the $D \to 0$ limit.*
Derive some analogous formulae for the two potentials of Fig. 28.

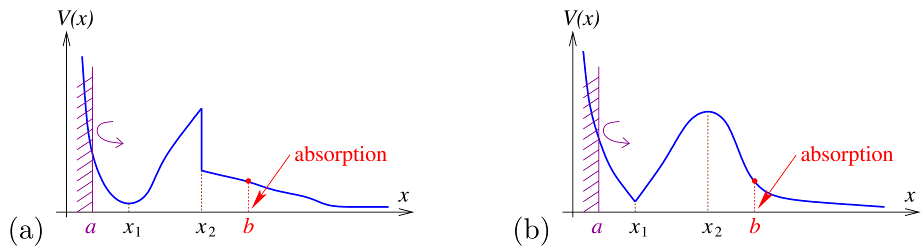


Figure 28: *Two other types of trapping potentials.*

✍ ***Exercice 78 – Escape from the two boundaries:*** *We now consider the problem where a particle starts at $x(0) = x_0 \in]a, b[$ and can escape the interval at one of the two boundaries. In this case one must solve the differential equation (404), i.e.*

$$\mathscr{G}_{x_0} T_n(x_0) = -n\,T_{n-1}(x_0) \qquad \textit{i.e.} \quad \left(D\frac{\mathrm{d}}{\mathrm{d}x_0} - V'(x_0)\right)\frac{\mathrm{d}T_n(x_0)}{\mathrm{d}x_0} = -n\,T_{n-1}(x_0) \tag{411}$$

for two Dirichlet boundary conditions $T_n(a) = T_n(b) = 0$. For simplicity, we consider only the first moment.

1/ *Denoting by $\psi(x) = \exp[-V(x)/D]$ (this is the equilibrium distribution, if normalisable), study the action of the generator $\mathscr{G}_x$ on*

$$\Phi(x) = \int_a^x \frac{\mathrm{d}y}{\psi(y)} \int_x^b \frac{\mathrm{d}x'}{\psi(x')} \int_a^{x'} \mathrm{d}z\,\psi(z) - \int_x^b \frac{\mathrm{d}y}{\psi(y)} \int_a^x \frac{\mathrm{d}x'}{\psi(x')} \int_a^{x'} \mathrm{d}z\,\psi(z) \tag{412}$$

2/ *Deduce $T_1(x_0)$.*

3/ *Study the limit $D \to 0$ for the potential of Fig. 29, when the initial condition is in the well. Introduce $1/\delta_0^2 = V''(x_0)$ and $1/\delta_{1,2}^2 = -V''(x_{1,2})$. Distinguish the general case $V(x_1) \neq V(x_2)$ and the case $V(x_1) = V(x_2)$.*

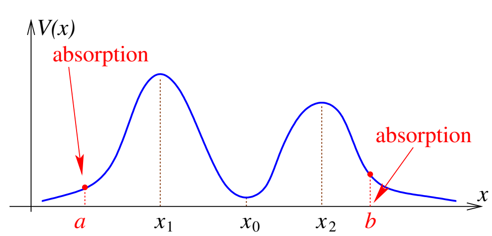


Figure 29: *Two absorbing boundaries.*

✍ ***Exercice 79* – *First passage time in dimension* d *:*** *We consider the problem of first passage time in dimension $d > 1$: a diffusive particle submitted to a centro-symmetric drift $\vec{F}(\vec{r}) = -V(r)\,\vec{u}_r$ where $\vec{u}_r$ is the radial unit vector. The forward generator of the diffusion in $\mathbb{R}^d$ is $\mathscr{G}^\dagger = D\Delta - \vec{\nabla}\cdot\vec{F}$. The particle starts from $\vec{r}_0$ and we ask the question : when does it reachs a sphere of radius $b < r_0 = ||\vec{r}_0||$ for the first time ?*

a) Show that the moments of the first passage time obey the differential equation

$$\left[D\left(\frac{\mathrm{d}^2}{\mathrm{d}r^2} + \frac{d-1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\right) - V'(r)\frac{\mathrm{d}}{\mathrm{d}r}\right] T_n(r) = -n\,T_{n-1}(r) \qquad (413)$$

Find an integral representation for $T_1(r_0)$.

b) When the dimension is increased, does the first passage time increases or decreases ?

☺ Important points

• Have in mind the meaning of the two terms in the FPE (drift and diffusion).
• Conservation equation $\partial_t P = -\partial_x J$ and expression of the current.
• Existence of the BFPE
• Be familiar with the different types of boundary conditions (relecting, absorbing)
• The problem of first passage time (relation between survival probability and the distribution of the first passage time) ; the use of the BFPE.
• Study of the first passage through point $x = b$ requires to consider an absorbing boundary at this place.
• The Arrhenius law.

7 Functionals of stochastic processes

7.1 Introduction/motivations

In the previous chapter, we have explained how to study various properties of stochastic processes : the distribution $P_t(x)$ or the conditional probability $P_t(x|x_0)$, and the first passage time T_{x_0}. This last chapter is devoted to the study of the statistical properties of "functionals" of the form

$$\boxed{S[x(\tau)] = \int_0^t \mathrm{d}\tau\, U(x(\tau))} \tag{414}$$

where $U(x)$ is a regular function and $x(\tau)$, for $\tau \in [0,t]$, is a Brownian motion (the case where it is a more general stochastic process will be also discussed).

A discrete version would be to consider a random sequence of numbers (a Markov chain) $(x_0, x_1, \cdots, x_t)$ and study the distribution of the sum

$$S(x_0, x_1, \cdots, x_t) = \sum_{\tau=0}^{t} U(x_\tau) \tag{415}$$

where $x_{\tau+1} = x_\tau + \eta_\tau$ is a random walk defined controlled by the "noise" η_τ, i.e. η_τ's are i.i.d. random numbers, with for instance $\langle \eta_\tau \rangle = 0$ and $\langle \eta_\tau \eta_{\tau'} \rangle = \delta_{\tau,\tau'}$. The sum $S(x_0, x_1, \cdots)$ for such process is the sum of *strongly correlated variables*, for a special type of correlations, $\langle x_\tau x_{\tau'} \rangle = \min(\tau, \tau')$. Hence, in this last chapter, we are extending the central limit theorem in another direction unexplored so far, when the sum of random variables are not i.i.d. but correlated.

Let us now give few examples of possible functionals which will be of interest.

Time spent in an interval.— An interesting example is the time spent on some interval $[a,b]$

$$T_{[a,b]}[x(\tau)] \stackrel{\text{def}}{=} \int_0^t \mathrm{d}\tau\, \mathbf{1}_{[a,b]}(x(\tau)) \tag{416}$$

where $\mathbf{1}_I(x)$ is the indicator function of the interval I,

$$\mathbf{1}_I(x) = \begin{cases} 1 & \text{for } x \in I \\ 0 & \text{for } x \notin I \end{cases} \tag{417}$$

The case of $T_{\mathbb{R}_+}[x(\tau)]$ is a famous example considered by Paul Lévy in 1939 and will be studied in detail below.

Local time.— Another example is the "local time"

$$\tau_a[x(\tau)] \stackrel{\text{def}}{=} \int_0^t \mathrm{d}\tau\, \delta(x(\tau) - a) \tag{418}$$

which measures the time spent at $x = a$. Exercise 81 is devoted to the derivation of its distribution.

Current in a disordered environment.— A last interesting example arises in the context of *disordered systems*. Consider the 1D classical diffusion in a force field $F(x) = -V'(x)$ in an interval $[0, L]$. We consider the stationary state $P_{\text{st}}(x)$ corresponding to a steady current injected from the left boundary, in the presence of some absorbing boundary condition at the right boundary, $P_{\rm st}(L)=0$. As we have already discussed, the stationary state is solution of $J=-\big[V'(x)+D\partial_x\big]P_{\rm st}(x)$, hence

$$P_{\rm st}(x)=\frac{J}{D}\mathrm{e}^{-V(x)/D}\int_x^L\mathrm{d}y\,\mathrm{e}^{+V(y)/D} \tag{419}$$

If the distribution is normalised on $[0,L]$, the normalisation condition provides the expression of the current J and of the distribution at the boundary, $P_{\rm st}(0)$. If instead we consider a non normalised solution and fix the value $P_{\rm st}(0)=P_0$, then the current is given by

$$1/J=\frac{1}{DP_0}\int_0^L\mathrm{d}x\,\mathrm{e}^{V(x)/D} \tag{420}$$

with $V(x)=\int_0^x\mathrm{d}y\,F(y)$. Now choose a quenched random force field, such that $\overline{F(x)}=\overline{F}$ and $\overline{F(x)F(x')}^c=\sigma\,\delta(x-x')$, the potential is a stochastic process in space (if $F(x)$ is Gaussian, $V(x)$ is the usual Brownian motion). The inverse of the current has the form (414) for a function $U(x)=\mathrm{e}^x$. Such functional was studied in detailed in [41]. Exponential functional of the Brownian motion have also been extensively considered by mathematicians and have many applications (reviews in mathematics [54, 39, 40] or in physics [9, 8]).

An application in finance and risk theory.— Let me close the list of examples with one which has arisen in finance concerning the determination of the present value of annuity or perpetuity : this has been discussed by mathematicians, see [12]. An annuity is a financial product which provides some fixed cash flow C, for example every year until time t. The question for insurance companies is to know the present value of the annuity : after one year, the value of the money will diminish (for example due to inflation), hence the value will not be C but $C/(1+r)$, where r is some discount rate. The value of the money received in two years will be $C/(1+r)^2$, etc. The amount of money C is fixed (determined by the contract), however the discount rate depends on the market fluctuations : let us denote r_t the rate at time t. Hence the present value of the annuity is

$$\mathrm{PV}=\frac{C}{1+r_1}+\frac{C}{(1+r_1)(1+r_2)}+\cdots+\frac{C}{(1+r_1)\cdots(1+r_t)} \tag{421}$$

Introducing the notation $\eta_\tau=-\ln(1+r_\tau)$ we have

$$\mathrm{PV}=\sum_{\tau=1}^t\prod_{i=1}^\tau\mathrm{e}^{\eta_i}=\sum_{\tau=1}^t\mathrm{e}^{x_\tau} \tag{422}$$

where $x_\tau=\sum_{i=1}^\tau\eta_i$ is a random walk. Hence we are considering the discrete version of (420).

7.2 Path integrals

The use of path integrals makes rather straighforward the derivation of the Feynman-Kac formula, a central formula for the study of functionals.

a) The Wiener measure

The starting point is the exercise 19. Let us consider the free Brownian motion on $\mathbb{R}$, described by the Gaussian conditional probability

$$P_t(x|x_0)=\frac{1}{\sqrt{2\pi t}}\mathrm{e}^{-\frac{1}{2t}(x-x_0)^2} \tag{423}$$

(in this section I set the diffusion constant $D = 1/2$ for simplicity). Using the stability of the Gaussian distribution, we can split the Brownian motion in many steps for small time increments $\delta t = t/N$, leading to

$$P_t(x|x_0) = \int \mathrm{d}x_{N-1} \cdots \mathrm{d}x_1 \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\,\delta t}} \mathrm{e}^{-\frac{1}{2\delta t}(x_i - x_{i-1})^2} \quad \text{where } x_N \equiv x \tag{424}$$

This corresponds to the Gaussian random walk studied in Exercise 81. This allows to identify the probabilistic weight of a trajectory passing through all intermediate positions

$$\mathrm{weight}(x_0 \to x_1 \to x_2 \to \cdots \to x_N) = \frac{1}{(2\pi\,\delta t)^{N/2}} \exp\left\{ -\sum_{i=1}^{N} \frac{(x_i - x_{i-1})^2}{2\delta t} \right\} \tag{425}$$

Taking the limit $N \to \infty$, we identify the weight of the continuous Brownian trajectory as the sum is replaced by the integral $\int_0^t \mathrm{d}\tau\, \frac{1}{2} \left(\frac{\mathrm{d}x(\tau)}{\mathrm{d}\tau}\right)^2$. Denoting the product of the differentials with the notation

$$\mathcal{D}x(\tau) = \lim_{N\to\infty} \frac{1}{\sqrt{2\pi\,t/N}} \prod_{i=1}^{N-1} \frac{\mathrm{d}x_i}{\sqrt{2\pi\,t/N}} \tag{426}$$

we conclude that the conditional probability can be formally written as a sum over the "paths" starting at x_0 and ending at x as

$$P_t(x|x_0) = \int_{x(0)=x_0}^{x(t)=x} \mathcal{D}x(\tau)\, \exp\left\{ -\int_0^t \mathrm{d}\tau\, \frac{1}{2} \left(\frac{\mathrm{d}x(\tau)}{\mathrm{d}\tau} \right)^2 \right\} \tag{427}$$

$\mathcal{D}x(\tau)\, \exp\left\{ -\int_0^t \mathrm{d}\tau\, \frac{1}{2} \left(\frac{\mathrm{d}x(\tau)}{\mathrm{d}\tau}\right)^2 \right\}$ is the "*Wiener measure*". This is the weight to associate to a Brownian trajectory for summation. This calls for a comment : summing over the paths with $\int \mathcal{D}x(\tau)$ a priori leads to sum over functions which are not necessarily continuous. This is the Gaussian part which imposes that trajectories are both continuous and non differentiable (meaning that discontinuous paths have weight zero).

b) A useful identity for path integrals

Path integration has very little in common with the usual calculus of integrals. Path integrals rather provide very useful and transparent representations of certain Green's function. The conditional probability is the solution of the equation

$$\partial_t P_t(x|x_0) = \frac{1}{2} \partial_x^2 P_t(x|x_0) \tag{428}$$

hence we have sometimes used the representation (similar to the QM one)

$$P_t(x|x_0) = \langle\, x \,|\mathrm{e}^{-tH_0}|\, x_0 \,\rangle \quad \text{for } H_0 = -\frac{1}{2} \frac{\mathrm{d}^2}{\mathrm{d}x^2} \tag{429}$$

(the minus sign is introduced in order to deal with an operator with positive spectrum).

For the following, we will also encounter the kernel

$$K_t(x|x_0) = \langle\, x \,|\mathrm{e}^{-tH}|\, x_0 \,\rangle \quad \text{for } H = -\frac{1}{2} \frac{\mathrm{d}^2}{\mathrm{d}x^2} + V(x)\,. \tag{430}$$

As we did above for the conditional probability, we can split the time t into N small time intervals

$$K_t(x|x_0) = \int \mathrm{d}x_{N-1} \cdots \mathrm{d}x_1 \prod_{i=1}^{N} \langle\, x_i \,|\mathrm{e}^{-\frac{t}{N}H}|\, x_{i-1} \,\rangle \tag{431}$$

(which corresponds to insert $N-1$ closure relations). In the $(t/N)\to 0$ limit, we can simplify the kernel as $\langle x_i|\mathrm{e}^{-\frac{t}{N}H}|x_{i-1}\rangle \simeq \langle x_i|\mathrm{e}^{-\frac{t}{N}H_0}|x_{i-1}\rangle \mathrm{e}^{-\frac{t}{N}V(x_i)}$ (use the Zassenhaus formula [35]). As a result the kernel can be represented under the form of the following path integral

$$K_t(x|x_0) = \int_{x(0)=x_0}^{x(t)=x} \mathcal{D}x(\tau)\, \exp\left\{ -\int_0^t \mathrm{d}\tau \left[\frac{1}{2}\left(\frac{\mathrm{d}x(\tau)}{\mathrm{d}\tau}\right)^2 + V(x(\tau)) \right] \right\} \tag{433}$$

This corresponds to the kernel $K_t(x|x_0) = \langle x|\mathrm{e}^{-tH}|x_0\rangle$ and obeys the differential equation

$$\partial_t K_t(x|x_0) = -HK_t(x|x_0) \quad \text{for } H = -\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V(x) \tag{434}$$

with initial condition $K_0(x|x_0) = \delta(x-x_0)$. This is all what one has to remember about path integrals : the rule of calculation of (433) is (434)

c) The Green's function

The kernel $K_t(x|y) = \langle x|\mathrm{e}^{-tH}|y\rangle$ is solution of a PDE which might be difficult to solve in general. It is often more convenient to consider its Laplace transform

$$G(x,y;\alpha) = \int_0^\infty \mathrm{d}t\, \mathrm{e}^{-\alpha t}\, K_t(x|y) = \langle x|\,(\alpha+H)^{-1}\,|y\rangle\,. \tag{435}$$

Let us forget the spectral parameter α, which can be incorporated into the potential and simply consider the Green's function $G(x,y) = \langle x|H^{-1}|y\rangle$ (assuming that $\lambda = 0$ does not belong to the spectrum of H), which solves the equation

$$\left(-\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V(x)\right) G(x,y) = \delta(x-y) \tag{436}$$

In 1D, the method to get G is simple. Introduce $\psi_-(x)$ and $\psi_+(x)$, the two solutions of the homogeneous equation,

$$-\frac{1}{2}\psi''_\pm(x) + V(x)\,\psi_\pm(x) = 0 \tag{437}$$

which vanish at $-\infty$ and $+\infty$, respectively. Clearly we have $G(x,y) = A_y\,\psi_-(x)$ for $x<y$ and $G(x,y) = B_y\,\psi_+(x)$ for $x>y$. Ensuring continuity we have

$$G(x,y) = C \begin{cases} \psi_-(x)\psi_+(y) & \text{for } x<y \\ \psi_-(y)\psi_+(x) & \text{for } x>y \end{cases} \tag{438}$$

We deduce a matching condition from the equation for the Green's function (do $\int_{y-\epsilon}^{y+\epsilon}\mathrm{d}x$ and take the limit $\epsilon\to 0^+$) : $\frac{\partial G}{\partial x}\Big|_{x=y-}^{x=y+} = -2$, thus we deduce that the coefficient C is related to the Wronskian

$$W[\psi_-,\psi_+] \overset{\text{def}}{=} \psi_-(x)\psi'_+(x) - \psi'_-(x)\psi_+(x) \tag{439}$$

which is constant for such a differential equation : one easliy checks that $\frac{\mathrm{d}}{\mathrm{d}x}W[\psi_-,\psi_+] = 0$. Finally

$$\boxed{G(x,y) = -\frac{2\,\psi_-(x_<)\psi_+(x_>)}{W[\psi_-,\psi_+]}} \tag{440}$$

where $x_< = \min(x,y)$ and $x_> = \max(x,y)$.

---

[35] Consider two operators A and B :

$$\mathrm{e}^{A+B} = \mathrm{e}^{A}\,\mathrm{e}^{B}\,\mathrm{e}^{-\frac{1}{2}[A,B]}\,\mathrm{e}^{\frac{1}{6}\left(2[B,[B,A]]+[A,[A,B]]\right)}\cdots \tag{432}$$

7.3 Functionals of the Brownian motion and the Feynman-Kac formula

Let us first consider a concrete and simple example, then we will develope the theory in the general case.

a) A simple case : the first Lévy's arcsine law (1939)

We consider the Wiener process, a 1D free Brownian motion starting from the origin $x(0) = 0$ and ask the question : *what is the time spent on the positive real axis ?* The answer to this question was given by Paul Lévy in 1939 [32].

We introduce the functional

$$T[x(\tau)] = \int_0^t \mathrm{d}\tau\ \theta_{\mathrm{H}}(x(\tau))\,. \tag{441}$$

The aim is to derive its distribution $\mathcal{P}_t(T)$ (here t is a parameter).

Characteristic function.— We introduce the characteristic function

$$\widetilde{\mathcal{P}}_t(p) = \int_0^\infty \mathrm{d}T\,\mathcal{P}_t(T)\,\mathrm{e}^{-pT} = \left\langle \mathrm{e}^{-p\,T[x(\tau)]}\,\middle|\,x(0)=0\right\rangle \tag{442}$$

which can we conveniently written with a path integral as

$$\widetilde{\mathcal{P}}_t(p) = \overbrace{\int \mathrm{d}x \int_{x(0)=0}^{x(t)=x}}^{\text{sum over paths}} \underbrace{\mathcal{D}x(\tau)\,\mathrm{e}^{-\frac{1}{2}\int_0^t\mathrm{d}\tau\,\dot{x}(\tau)^2}}_{\text{weight of a path}}\,\mathrm{e}^{-p\,T[x(\tau)]}\,. \tag{443}$$

The path integral is the representation of the propagator

$$K_t(x|x_0) = \int_{x(0)=x_0}^{x(t)=x} \mathcal{D}x(\tau)\,\mathrm{e}^{-\int_0^t\mathrm{d}\tau\,\left[\frac{1}{2}\dot{x}(\tau)^2+p\ \theta_{\mathrm{H}}(x(\tau))\right]} = \langle\, x\,|\mathrm{e}^{-tH_p}|\,x_0\,\rangle \tag{444}$$

where

$$H_p = -\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + p\ \theta_{\mathrm{H}}(x)\,. \tag{445}$$

Thus the kernel solves

$$(\partial_t - H_p)K_t(x|x_0) = 0 \ \text{ for initial condition } K_0(x|x_0) = \delta(x-x_0)\,. \tag{446}$$

As a result, the characteristic function of the time is

$$\widetilde{\mathcal{P}}_t(p) = \int \mathrm{d}x\,\langle\, x\,|\mathrm{e}^{-tH_p}|\,0\,\rangle \tag{447}$$

For technical reasons it is more easy to take another Laplace transform, with respect to the time parameter

$$Q(\alpha,p) \stackrel{\text{def}}{=} \int_0^\infty \mathrm{d}t\,\mathrm{e}^{-\alpha t}\,\widetilde{\mathcal{P}}_t(p) = \int \mathrm{d}x\,\langle\, x\,|\,(\alpha+H_p)^{-1}\,|\,0\,\rangle \tag{448}$$

which involves the Green's function

$$G(x,x_0;\alpha,p) \stackrel{\text{def}}{=} \langle\, x\,|\frac{1}{\alpha+H_p}|\,x_0\,\rangle \tag{449}$$

obeying

$$\left(\alpha - \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + p\,\theta_\mathrm{H}(x)\right) G(x,x_0;\alpha,p) = \delta(x-x_0)\,. \tag{450}$$

The solution is easy to find (see paragraph above on the calculation of the Green function).

We apply the formula (440). For this, we just need to find $\psi_-(x)$ and $\psi_+(x)$, i.e. solve an elementary linear second order differential equation, $-\psi''_\pm(x) + 2\big(\alpha + p\,\theta_\mathrm{H}(x)\big)\psi_\pm(x)$. For $x<0$, the solution vanishing at $-\infty$ is $\psi_-(x) = \mathrm{e}^{\sqrt{2\alpha}\,x}$ and for $x>0$ we have $\psi_-(x) = A\mathrm{e}^{+\sqrt{2(\alpha+p)}\,x} + B\mathrm{e}^{-\sqrt{2(\alpha+p)}\,x}$, where the two coefficients are found by imposing the matching conditions (continuity of ψ_- and ψ'_- at $x=0$). We easily find $A = \frac{1}{2}\big[1+\sqrt{\alpha/(\alpha+p)}\big]$ and $B = \frac{1}{2}\big[1-\sqrt{\alpha/(\alpha+p)}\big]$, hence

$$\psi_-(x) = \begin{cases} \mathrm{e}^{\sqrt{2\alpha}\,x} & \text{for } x<0 \\ \cosh\left[\sqrt{2(\alpha+p)}\,x\right] + \sqrt{\frac{\alpha}{\alpha+p}}\,\sinh\left[\sqrt{2(\alpha+p)}\,x\right] & \text{for } x>0 \end{cases} \tag{451}$$

$\psi_+(x)$ is simply obtained by performing the substitutions $x\to -x$ and $\alpha \leftrightarrow \alpha+p$ in $\psi_-(x)$. Therefore $\psi'_-(0) = \sqrt{2\alpha}$ and $\psi'_+(0) = -\sqrt{2(\alpha+p)}$ and the Wronskian is

$$W[\psi_-,\psi_+] = -\sqrt{2\alpha} - \sqrt{2(\alpha+p)}\,. \tag{452}$$

As a result

$$G(x,0;\alpha,p) = \frac{2}{\sqrt{2\alpha}+\sqrt{2(\alpha+p)}} \begin{cases} \mathrm{e}^{\sqrt{2\alpha}\,x} & \text{for } x<0 \\ \mathrm{e}^{-\sqrt{2(\alpha+p)}\,x} & \text{for } x>0 \end{cases} \tag{453}$$

whose integration gives

$$Q(\alpha,p) = \int \mathrm{d}x\, G(x,0;\alpha,p) = \frac{2}{\sqrt{2\alpha}+\sqrt{2(\alpha+p)}}\left(\frac{1}{\sqrt{2\alpha}} + \frac{1}{\sqrt{2(\alpha+p)}}\right) = \frac{1}{\sqrt{\alpha(\alpha+p)}} \tag{454}$$

Finally the double Laplace transform of the distribution has a simple expression

$$\int_0^\infty \mathrm{d}t\,\mathrm{e}^{-\alpha t}\int_0^\infty \mathrm{d}T\,\mathrm{e}^{-pT}\,\mathcal{P}_t(T) = \frac{1}{\sqrt{\alpha(p+\alpha)}}\,. \tag{455}$$

Arcsine law .— Performing first the inverse Laplace transform with respect to the variable α, we get the integral form

$$\int_0^\infty \mathrm{d}T\,\mathrm{e}^{-pT}\,\mathcal{P}_t(T) = \int_{\mathscr{B}} \frac{\mathrm{d}\alpha}{2\mathrm{i}\pi}\,\frac{\mathrm{e}^{\alpha t}}{\sqrt{\alpha(\alpha+p)}} = \int_0^p \frac{\mathrm{d}x}{\pi}\,\frac{\mathrm{e}^{-xt}}{\sqrt{x(p-x)}} \tag{456}$$

Hence, simply relabelling the integration variable, we get

$$\boxed{\mathcal{P}_t(T) = \frac{1}{\pi\sqrt{T(t-T)}}} \qquad \text{for } T\in[0,t]\,, \tag{457}$$

which is known as the Lévy's "*first arcsine law*" (probabilists always prefer to give the cumulative rather than the density)

$$\int_0^T \mathrm{d}T'\,\mathcal{P}_t(T') = \frac{2}{\pi}\,\arcsin\left(\sqrt{T/t}\right)\,. \tag{458}$$

This result is interesting from the Brownian motion perspective. The distribution of the time has two peaks close to $T\sim 0$ and $T\sim t$, meaning that most probably, the Wiener process spends all time either in $\mathbb{R}_-$ or in $\mathbb{R}_+$. This seems contradictory with our intuition from the recurrent properties, as we have seen that the 1D free Brownian motion returns to its starting point with probability one after long enough time. Hence one could incorrectly think that a typical trajectory frequently comes back to the origin and spends equal time on $\mathbb{R}_-$ and $\mathbb{R}_+$; this latter picture is incorrect. Altough the typical Brownian trajectory can visit many times its starting point, it spends most time either in $\mathbb{R}_-$ or in $\mathbb{R}_+$.

The second and third arcsine laws.— Lévy also proved two other nice properties of the Wiener process :

- The second arcsine law is about the distribution of the last time t_+ when the process changes in sign, which is also distributed according to $\mathcal{P}_t(t_+)$.
- The time t_m where the maximum occurs also obeys the same law $\mathcal{P}_t(t_m)$ (third arcsine law).

A Mathematical tools

A.1 Fourier transform

a) Fourier transform on $\mathbb{R}$

Consider a function f on $\mathbb{R}$. Its Fourier transform is defined as

$$\hat{f}(k) = \int_{\mathbb{R}} \mathrm{d}x\, f(x)\, \mathrm{e}^{-\mathrm{i}kx} \quad \text{and} \quad f(x) = \int_{\mathbb{R}} \frac{\mathrm{d}k}{2\pi} \hat{f}(k)\, \mathrm{e}^{\mathrm{i}kx} \tag{479}$$

the inverse Fourier transform is recovered by using

$$\int_{\mathbb{R}} \frac{\mathrm{d}k}{2\pi}\, \mathrm{e}^{\mathrm{i}kx} = \delta(x) \tag{480}$$

You should now the basic properties of Fourier transform

$$\mathscr{F}_k[f'(x)] = \mathrm{i}k\, \hat{f}(k)\,, \tag{481}$$

$$\mathscr{F}_k[f * g] = \hat{f}(k)\, \hat{g}(k)\,, \tag{482}$$

etc.

b) Discrete Fourier transform (over a finite interval)

Consider a function $f(x)$ defined on a finite interval $x \in [0, L]$ (the function can be a periodic function with period L). For a bounded domain, the wave vectors are quantized, hence the function can be decomposed over a countable basis of plane waves

$$u^{(n)}(x) = \mathrm{e}^{\mathrm{i}k_n x} \quad \text{for } k_n = \frac{2n\pi}{L} \text{ with } n \in \mathbb{Z}\,. \tag{483}$$

We easily check the orthonormalisation condition

$$\int_0^L \frac{\mathrm{d}x}{L}\, u^{(n)}(x)^* u^{(m)}(x) = \int_0^L \frac{\mathrm{d}x}{L}\, \mathrm{e}^{\mathrm{i}(k_m - k_n)x} = \delta_{n,m} \tag{484}$$

Thus Fourier transform takes the form

$$f(x) = \frac{1}{L} \sum_{n\in\mathbb{Z}} \hat{f}_n\, \mathrm{e}^{\mathrm{i}k_n x} \quad \text{where} \quad \hat{f}_n = \int_0^L \mathrm{d}x\, f(x)\, \mathrm{e}^{-\mathrm{i}k_n x} = \langle\, u^{(n)}\, |\, f\, \rangle \tag{485}$$

If we take the limit $L \to \infty$, the spectrum of wave vectors becomes dense and we recover the continuous Fourier transform, using $\frac{1}{L}\sum_{k_n} \to \int_{\mathbb{R}} \frac{\mathrm{d}k}{2\pi}$.

c) Fourier transform of a discrete function

If we consider a function f_x defined over $\mathbb{Z}$ (on a lattice), the basis of harmonic functions are plane waves $u_x^{(k)} = \mathrm{e}^{\mathrm{i}kx}$ where the wave vector varies continuously over a bounded domain, the "Brillouin zone" $k \in]-\pi, +\pi]$ (because $u_x^{(k+2\pi)} = u_x^{(k)}$ for $x \in \mathbb{Z}$). As a result

$$f_x = \int_{-\pi}^{+\pi} \frac{\mathrm{d}k}{2\pi} \hat{f}(k)\, \mathrm{e}^{\mathrm{i}kx} \quad \text{where} \quad \hat{f}(k) = \sum_{x\in\mathbb{Z}} f_x\, \mathrm{e}^{-\mathrm{i}kx} \tag{486}$$

The passage between the two is now ensured by the Poisson formula

$$\sum_{n\in\mathbb{Z}} \delta(k - 2\pi n) = \frac{1}{2\pi} \sum_{x\in\mathbb{Z}} \mathrm{e}^{\mathrm{i}kx} \tag{487}$$

(because k is in the Brillouin zone, we only need the $n=0$ term).

It is also possible to recover the Fourier transform for continuous functions by introducing a lattice spacing a (writing now that $x/a \in \mathbb{Z}$). The Brillouin zone is then $k \in]-\pi/a, \pi/a]$. In the limit $a \to 0$ we recover the Fourier transform.

d) Fourier transform of a discrete function defined over a finite set

If we now define a discrete function f_x over a finite set $x \in \{1,\, 2,\, \cdots, L\}$, the wave vector is both quantized (finite volume) and bounded (lattice problem) :

$$k_n = 2\pi n/L \quad \text{with} \quad n = -L/2+1, \cdots, -1,\, 0,\, 1,\, \cdots, L/2 \in]-\pi, +\pi]$$

(for L even).

Preliminary.— consider the N roots of the identity, solutions of $z^N = 1$,

$$z_k = \mathrm{e}^{2\mathrm{i}\pi k/N} \qquad \text{for } k = 0,\, 1 \cdots, N-1 \tag{488}$$

Using that the characteristic polynomial is

$$P(z) = z^N - 1 = \prod_{k=0}^{N-1} (z - z_k) \tag{489}$$

$\prod_k (z - z_k)$ is a polynomial of degree N with only two terms. The z^1 coefficient being zero, we have $\sum_k z_k = 0$. Similarly we can prove $\sum_k z_k^n = N\,\delta_{n,0}$, i.e. more explicitely

$$\frac{1}{N} \sum_{k=0}^{N-1} \mathrm{e}^{2\mathrm{i}\pi nk/N} = \delta_{n,0}\,. \tag{490}$$

This relation can be used for the Fourier transform.

Fourier transform.— We have now

$$f_x = \frac{1}{L} \sum_{n=1}^{L} \hat{f}_n\, \mathrm{e}^{2\mathrm{i}\pi nx/L} \qquad \text{where} \qquad \hat{f}_n = \sum_{x=1}^{L} f_x\, \mathrm{e}^{-2\mathrm{i}\pi nx/L} \tag{491}$$

A.2 Linear differential equations with source terms

First order.– First consider the simple differential equation

$$y'(x) - \lambda\, y(x) = s(x) \tag{492}$$

where λ and $s(x)$ are given. The solution of the homogeneous equation is $y(x) = A\,\mathrm{e}^{\lambda x}$. Use the "variation of the constant's method" : search a particular solution under the form $y(x) = A(x)\,\mathrm{e}^{\lambda x}$, which leads to $A'(x) = s(x)\,\mathrm{e}^{-\lambda x}$, i.e. $A(x) = \int^x \mathrm{d}t\, s(t)\,\mathrm{e}^{-\lambda t}$. Hence the general solution is

$$y(x) = y(0)\,\mathrm{e}^{\lambda x} + \int_0^x \mathrm{d}t\, s(t)\, \mathrm{e}^{\lambda(x-t)} \tag{493}$$

You can check that the method can be extended to the case where $\lambda \to \lambda(x)$ in the differential equation (492). Then

$$\boxed{y(x) = y(0)\,\mathrm{e}^{\int_0^x \mathrm{d}u\,\lambda(u)} + \int_0^x \mathrm{d}t\, s(t)\, \mathrm{e}^{\int_t^x \mathrm{d}u\,\lambda(u)}} \tag{494}$$

Second order.– The same idea can be applied to the second order linear differential equation

$$y''(x) + a(x)\, y'(x) + b(x)\, y(x) = s(x) \tag{495}$$

The homogeneous equation has two independent solutions denoted $y_1(x)$ and $y_2(x)$. The Wronskian is $W \equiv \mathcal{W}[y_1, y_2] \stackrel{\text{def}}{=} y_1 y_2' - y_1' y_2$. We easily obtain the differential equation $W' = -aW$ i.e. $W(x) = W(0) \exp\big[-\int_0^x \mathrm{d}u\, a(u)\big]$. For $a = 0$ the Wronskian is constant. Applying the "variation of the constant's method", we get the general solution

$$y(x) = A\, y_1(x) + B\, y_2(x) - y_1(x) \int_0^x \mathrm{d}u\, \frac{y_2(u) s(u)}{W(u)} + y_2(x) \int_0^x \mathrm{d}u\, \frac{y_1(u) s(u)}{W(u)}\,. \tag{496}$$

For example, for $a(x) = 0$ and $b(x) = k^2$ we can choose $y_1(x) = \cos kx$ and $y_2(x) = \sin kx$, with Wronskian $W(x) = k$.

A.3 Asymptotics of integrals

One frequently deals with integrals involving large parameters (furthermore, many useful special functions are defined through nice integral representations). Here I recall two methods allowing to extract some asymptotic behaviour.

a) The Laplace method

Consider a monotonously increasing function $\varphi(x)$ and the integral $\int_0^\infty \mathrm{d}x\, \mathrm{e}^{-\lambda\varphi(x)}$. When $\lambda \to +\infty$, one expects the exponential to decay fast, on scale $\sim 1/(\lambda\varphi'(0)) \to 0$, and the integral is dominated by the boundary. Note that in the expansion $\mathrm{e}^{-\lambda\varphi(x)} = \mathrm{e}^{-\lambda\varphi(0)} \exp\big\{ - \lambda\varphi'(0)t - \lambda\varphi''(0)t^2/2 + \cdots \big\}$, the exponential $\mathrm{e}^{-\lambda\varphi''(0)t^2/2}$ decays over a *larger* scale, $\sim 1/(\lambda\varphi''(0))^{1/2}$. In conclusion, adding a smooth function $g(x)$ in the integral, we can write

$$I(\lambda) = \int_0^\infty \mathrm{d}x\, g(x)\, \mathrm{e}^{-\lambda\varphi(x)} \underset{\lambda\to+\infty}{\simeq} \frac{g(0)}{\lambda\varphi'(0)} \mathrm{e}^{-\lambda\varphi(0)} \tag{497}$$

Note that if $g(x)$ vanishes, we can extend the method. For example, a similar explansion gives

$$I(\lambda) = \int_0^\infty \mathrm{d}x\, g(x)\, \mathrm{e}^{-\lambda\varphi(x)} \underset{\lambda\to+\infty}{\simeq} \frac{g'(0)}{\big[\lambda\varphi'(0)\big]^2} \mathrm{e}^{-\lambda\varphi(0)} \quad \text{for } g(x) \simeq g'(0)\, x \text{ for } x \to 0. \tag{498}$$

Remark : a systematic expansion under these lines usually produce an asymptotic series. For example, one can apply the method to get the asymptotic of the exponential integral :

$$E_1(z) \stackrel{\text{def}}{=} \int_z^\infty \frac{\mathrm{d}t}{t}\, \mathrm{e}^{-t} \tag{499}$$

One obtains $E_1(z) \simeq (1/z)\mathrm{e}^{-z}$ for $z \to +\infty$. This is the first term of the asymptotic series (series with zero convergence radius)

$$E_1(z) \underset{z\to\infty}{=} \frac{\mathrm{e}^{-z}}{z} \left(\sum_{k=0}^{n} (-1)^k \frac{k!}{z^k} + \mathcal{O}(|z|^{-n-1}) \right) \tag{500}$$

(valid for $|\arg(z)| \leqslant \pi - \delta$ with $\delta > 0$). [36]

[36] You can get the series easily by introducing $E_k(z) \stackrel{\text{def}}{=} \int_z^{+\infty} \frac{\mathrm{d}t}{t^k}\, \mathrm{e}^{-t}$ and use the recurrence $E_n(z) = \frac{\mathrm{e}^{-z}}{z^n} - \frac{n}{z}\, E_{n+1}(z)$

b) The steepest descent method (saddle point approximation)

Another frequently encountered case is an integral of the form $\int_a^b \mathrm{d}x\,\mathrm{e}^{-\lambda\varphi(x)}$ where $\lambda\in\mathbb{R}$ is a large parameter and $\varphi(x)$ a real function with one absolute minimum at $x_*\in]a,b[$ sufficiently far from the boundaries, with $\varphi''(x_*)>0$. In the $\lambda\to+\infty$ limit, the exponential is sharply peaked around x_* and we can expand the function around it, which leads to a Gaussian integral. As a result

$$I(\lambda)=\int_a^b \mathrm{d}x\,\mathrm{e}^{-\lambda\varphi(x)} \underset{\lambda\to+\infty}{\simeq} \sqrt{\frac{2\pi}{\lambda\,\varphi''(x_*)}}\,\mathrm{e}^{-\lambda\varphi(x_*)}\,. \tag{501}$$

Remarks :

- When $\varphi(x)$ is analytic, contour deformations are allowed, which might permit to pass through a saddle point not on the original contour (the real axis) but somewhere else in the complex plane.
- The method also applies to integral of the type $I(\lambda)=\int_a^b \mathrm{d}x\,\mathrm{e}^{-\mathrm{i}\lambda\varphi(x)}$ where λ and $\varphi(x)$ are real. In this case one should take into account the contributions of all stationary points along the contour, i.e. both maxima and minima of $\varphi(x)$. Then, the method is called the "*stationary phase method*".

✍ ***Exercice 83* – :** *Extend the method when $\varphi''(x_*)=0$ and $\varphi''''(x_*)>0$.*

✍ ***Exercice 84* – *Asymptotic of the MacDonald function* :** *Using the integral representation (517) of the MacDonald function, use the steepest descent method to get the asymptotic $K_\nu(z)\underset{z\to+\infty}{\simeq}\sqrt{\frac{\pi}{2z}}\,\mathrm{e}^{-z}$.*

✍ ***Exercice 85* – :** *Requirement of a contour deformation in the application of the steepest descent method appears when studying the Airy function*

$$\mathrm{Ai}(z)=\int_{-\infty}^{+\infty}\frac{\mathrm{d}t}{2\pi}\,\mathrm{e}^{\mathrm{i}(\frac{t^3}{3}+zt)} \tag{502}$$

solution of the Airy equation $y''(z)=z\,y(z)$. We study the asymptotic for $z\to\pm\infty$. Find the saddle point(s) and deduce the asymptotics

$$\mathrm{Ai}(z)\simeq\begin{cases}\frac{1}{2\sqrt{\pi}\,z^{1/4}}\mathrm{e}^{-\frac{2}{3}z^{3/2}} & \text{for } z\to+\infty\\ \frac{1}{\sqrt{\pi}(-z)^{1/4}}\cos\left(\frac{2}{3}(-z)^{3/2}-\frac{\pi}{4}\right) & \text{for } z\to-\infty\end{cases} \tag{503}$$

argue that only one saddle point controls the asymptotic for $z>0$.
Hint : plot $\varphi(t)$ along the imaginary axis of the complex variable t.

A.4 Special functions

Special functions are very useful. You can find information in standard books [21, 2] or on the internet [1].

Euler Gamma function

any integral of the form $\int_0^\infty \mathrm{d}x\,x^\alpha\,\mathrm{e}^{-c\,x^\beta}$ can be related to the Euler Gamma function

$$\Gamma(z)=\int_0^\infty \mathrm{d}t\,t^{z-1}\,\mathrm{e}^{-t}\quad \text{for } \mathrm{Re}(z)>0\,. \tag{504}$$

It can also be defined for $\mathrm{Re}(z) \leqslant 0$, however it presents simple poles for $-n \in \mathbb{N}$.
Important property : $\boxed{\Gamma(z+1) = z\,\Gamma(z)}$.
Special values : $\Gamma(1) = 1$ and $\Gamma(1/2) = \sqrt{\pi}$.
From these, we immediatly deduce $\Gamma(n+1) = n!$ and $\Gamma(n+\frac{1}{2}) = 2^{-n}\sqrt{\pi}\,(2n-1)!!$.

Asymptotic is given by the Stirling formula for $z \gg 1$

$$\Gamma(z+1) \simeq \sqrt{2\pi z}\, z^z\, \mathrm{e}^z \,. \tag{505}$$

✍ ***Exercice 86** – **:** Demonstrate this result starting from the integral representation and using the steepest descent method (cf. Appendix b) page 99).*

✍ ***Exercice 87** – **:** Express the integral*

$$\int_0^\infty \mathrm{d}x\, x^\mu\, \mathrm{e}^{-\frac{1}{2}a\,x^2} \tag{506}$$

in terms of the Gamma function.

Pochhammer symbol : sometimes it is useful to introduce the Pochhammer symbol

$$(a)_n \stackrel{\text{def}}{=} a(a+1)\cdots(a+n-1) = \frac{\Gamma(a+n)}{\Gamma(a)}\,. \tag{507}$$

It can be used in the following Taylor expansion :

$$(1-x)^{-\alpha} = 1 + \alpha\,x + \frac{\alpha(\alpha+1)}{2!}\,x^2 + \cdots = \sum_{n=0}^{\infty} \frac{(\alpha)_n}{n!}\,x^n \tag{508}$$

$$(1+x)^{\alpha} = 1 + \alpha\,x + \frac{\alpha(\alpha-1)}{2!}\,x^2 + \cdots = \sum_{n=0}^{\infty} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha-n+1)n!}\,x^n \tag{509}$$

Note that this last expansion is stopped when $\alpha \in \mathbb{N}$ due to the divergence of the Gamma function in the denominator when $n > \alpha \in \mathbb{N}$ (the formula then coincides with the binomial formula).

Euler Beta function

The Beta function is defined as

$$B(\mu,\nu) = \frac{\Gamma(\mu)\Gamma(\nu)}{\Gamma(\mu+\nu)}\,. \tag{510}$$

Useful integrals :

$$B(\mu,\nu) = \int_0^1 \mathrm{d}t\, t^{\mu-1}(1-t)^{\nu-1} = 2\int_0^{\pi/2} \mathrm{d}\theta\, \sin^{2\mu-1}\theta\, \cos^{2\nu-1}\theta\,. \tag{511}$$

Error function

The error function is the integral of the Gaussian function :

$$\mathrm{erf}(z) \stackrel{\text{def}}{=} \frac{2}{\sqrt{\pi}} \int_0^z \mathrm{d}t\, \mathrm{e}^{-t^2} \tag{512}$$

We also introduce the complementary error functions $\mathrm{erfc}(z) = 1 - \mathrm{erf}(z)$. Asymptotics :

$$\mathrm{erfc}(z) \underset{z\to\infty}{\simeq} \frac{\mathrm{e}^{-z^2}}{\sqrt{\pi}} \sum_{n=0}^{N} (-1)^n \frac{(1/2)_n}{z^{2n+1}} + R_N(z) \tag{513}$$

where $(a)_n$ is the Pochhammer symbol.

Modified Bessel functions

The modified Bessel equation is $z^2\, y''(z) + z\, y'(z) - (\nu^2 + z^2)\, y(z) = 0$.

• A solution is the modified Bessel function

$$I_\nu(z) = \sum_{n=0}^{\infty} \frac{(z/2)^{2n+\nu}}{n!\Gamma(\nu+n+1)} \tag{514}$$

It grows exponentially at infinity $I_\nu(z) \sim \mathrm{e}^{+z}$ for $z \to +\infty$.

The one for index $\nu = 0$ has been used :

$$I_0(z) = \int_0^\pi \frac{\mathrm{d}t}{\pi}\, \mathrm{e}^{z\cos t} = \int_{-1}^{+1} \frac{\mathrm{d}t}{\pi} \frac{\mathrm{e}^{-zt}}{\sqrt{1-t^2}} \tag{515}$$

Asymptotics : $I_0(z) \simeq 1 + \frac{z^2}{4} + \cdots$ for $z \to 0$ and $I_0(z) \simeq \frac{\mathrm{e}^z}{\sqrt{2\pi z}}$ for $z \to \infty$.

• Another independent solution of the modified Bessel equation is given by the MacDonald function (Bessel function of third kind)

$$K_\nu(z) = \frac{\pi}{2\sin \pi\nu} \left[I_{-\nu}(z) - I_\nu(z)\right] \tag{516}$$

(the expression is valid for $\mu \notin \mathbb{Z}$). Obviously $K_\nu(z) = K_{-\nu}(z)$. For integer index, see [21], of formula 10.31.1 of [1].

The MacDonald function is divergent for $z \to 0$, as $K_\nu(z) \simeq \frac{\Gamma(\nu)}{2}\left(\frac{z}{2}\right)^{-\nu}$, and *decays* exponentially at infinity, $K_\nu(z) \simeq \sqrt{\frac{\pi}{2z}}\,\mathrm{e}^{-z}$ for $z \to +\infty$. A useful integral representation is

$$K_\nu(z) = \frac{1}{2}\left(\frac{z}{2}\right)^\nu \int_0^\infty \frac{\mathrm{d}t}{t^{\nu+1}}\, \mathrm{e}^{-t-z^2/4t} \quad \text{for } \operatorname{Re} z > 0 \tag{517}$$

Elliptic integrals

Elliptic integral of first kind :

$$\mathrm{K}(k) = \int_0^{\pi/2} \frac{\mathrm{d}\alpha}{\sqrt{1 - k^2 \sin^2 \alpha}} \tag{518}$$

Limit $k \to 0$: $\mathrm{K}(k) = \frac{\pi}{2}\left(1 + \frac{k^2}{4} + \cdots\right)$

Limit $k \to 1^-$: $\mathrm{K}(k) = \ln 4/k' + \frac{k'^2}{4}\left(\ln 4/k' - 1\right) + \cdots$ where $k' = \sqrt{1-k^2}$.

A useful integral is

$$\int_{-\pi}^{+\pi} \frac{\mathrm{d}^2\vec{k}}{(2\pi)^2} \frac{1}{2a + \cos k_x + \cos k_y} = \frac{1}{\pi a}\, \mathrm{K}(1/a) \tag{519}$$

A.5 Poisson formulae

Let $\hat{f}(k) = \int_{\mathbb{R}} \mathrm{d}x\, f(x)\, \mathrm{e}^{-\mathrm{i}kx}$:

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{n\in\mathbb{Z}} \hat{f}(2\pi n) \tag{520}$$

In particular

$$\sum_{n\in\mathbb{Z}} \delta(x-n) = \sum_{n\in\mathbb{Z}} \mathrm{e}^{2\mathrm{i}\pi n x} \tag{521}$$

✍ ***Exercice 88 – :*** *Check that*

$$\Theta(y|\alpha,\eta) \stackrel{\text{def}}{=} \sum_{n\in\mathbb{Z}} \mathrm{e}^{2\mathrm{i}\pi n\eta}\,\mathrm{e}^{-\pi^2(n+\alpha)^2 y} = \frac{1}{\sqrt{\pi y}} \sum_{n\in\mathbb{Z}} \mathrm{e}^{2\mathrm{i}\pi(n-\eta)\alpha}\,\mathrm{e}^{-\frac{(n-\eta)^2}{y}}\ , \tag{522}$$

This sum is related to the Jacobi theta function (DLMF, § 20) $\theta_3(z,q) = \sum_n q^{n^2}\mathrm{e}^{2\mathrm{i}nz}$, i.e. $\Theta(y|0,\eta) = \theta_3(\pi\eta, \mathrm{e}^{-\pi^2 y})$.

References

[1] NIST Digital Library of Mathematical Functions (), http://dlmf.nist.gov/, Release 1.1.3 of 2021-09-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain (eds).

[2] M. Abramowitz and I. A. Stegun (editors), *Handbook of Mathematical functions*, Dover, New York (1964).

[3] D. Applebaum, Lévy processes– from probability to finance and quantum groups, Notices Amer. Math. Soc. **51**(11), 1336–1347 (2004).

[4] D. Applebaum, *Lévy processes and stochastic calculus*, Cambridge University Press, Cambridge (2004).

[5] J.-P. Bouchaud and A. Georges, Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications, Phys. Rep. **195**, 127–293 (1990).

[6] B. Bru and M. Yor, La vie de W. Doeblin, Lettre de l'Académie des Sciences **2**, 16–17 (2001).

[7] S. Chaturvedi, Gaussian stochastic processes, in *Stochastic Processes Formalism and Applications*, edited by G. S. Agarwal and S. Dattagupta, pp. 19–29, Berlin, Heidelberg (1983), Springer Berlin Heidelberg, ISBN 978-3-540-40923-6.

[8] A. Comtet, J. Desbois and C. Texier, Functionals of the Brownian motion, localization and metric graphs, J. Phys. A: Math. Gen. **38**, R341–R383 (2005).

[9] A. Comtet, C. Monthus and M. Yor, Exponential functionals of Brownian motion and disordered systems, J. Appl. Probab. **35**, 255 (1998).

[10] H. A. David, *Order Statistics*, John Wiley & Sons, New York, second edition (1981).

[11] A. Dhar, Heat transport in low-dimensional systems, Adv. Phys. **57**, 457–537 (2008).

[12] D. Dufresne, The distribution of a perpetuity, with application to risk theory and pension funding, Scand. Actuarial. J. pp. 39–79 (1990).

[13] B. Duplantier, Le mouvement brownien, "divers et ondoyant", Séminaire Poincaré **1**, 155–212 (2005).

[14] A. 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).

[15] W. Feller, *An Introduction to Probability Theory and its Applications*, Wiley, New York (1971), vol. I & vol. II.

[16] M. Fréchet, Sur la loi de probabilité de l'écart maximum, Ann. Soc. Polonaise Math. Cracovie **6**, 93 (1927).

[17] K. 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).

[18] C. W. Gardiner, *Handbook of stochastic methods for physics, chemistry and the natural sciences*, Springer (1989).

[19] C. W. Gardiner and P. Zoller, *Quantum noise : a handbook of Markovian and non-Markovian quantum stochastic methods with applications to quantum optics*, Springer (2004).

[20] A. Grabsch, C. Texier and Y. Tourigny, One-dimensional disordered quantum mechanics and Sinai diffusion with random absorbers, J. Stat. Phys. **155**(2), 237–276 (2014).

[21] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series and products*, Academic Press, fifth edition (1994).

[22] E. J. Gumbel, Les valeurs extrêmes des distributions statistiques, Annales de l'Institut Henri Poincaré **V**(2), 115–158 (1935).

[23] E. J. Gumbel, Statistical Theory of Extreme Values and Some Practical Applications, National Bureau of Standards Applied Mathematics Series **33** (1954), Issued February 12.

[24] E. J. Gumbel, *Statistics of Extremes*, Columbia University Press, New York (1958).

[25] C. Itzykson and J.-M. Drouffe, *Théorie statistique des champs*, Interéditions–CNRS, Paris (1989), Tomes 1 et 2.

[26] G. Junker, *Supersymmetric methods in quantum and statistical physics*, Springer (1996).

[27] M. Kac, On the distribution of certain Wiener functionals, Trans. Am. Math. Soc. **65**(1), 1–13 (1949).

[28] M. Kac, A class of limit theorems, Trans. Am. Math. Soc. **84**, 459–471 (1957).

[29] W. Krauth, *Statistical Mechanics: Algorithms and Computations*, Oxford Unversity Press (2006).

[30] P. Langevin, Sur la théorie du mouvement brownien, C. R. Acad. Sc. (Paris) **146**, 530–533 (1908).

[31] S. Lepri, R. Livi and A. Politi, Thermal conduction in classical low-dimensional lattices, Phys. Rep. **377**(1), 1–80 (2003).

[32] P. Lévy, Sur certains processus stochastiques homogènes, Compositio Mathematica **7**, 283–339 (1940).

[33] R. Livi and P. Politi, *Nonequilibrium statistical physics: a modern perspective*, Cambridge University Press (2017).

[34] E. Lutz and S. Ciliberto, Information: From Maxwell's demon to Landauer's eraser, Physics Today **68**, 30–35 (2015).

[35] S. N. Majumdar, Brownian Functionals in Physics and Computer Science, Current Science **89**, 2076–2092 (2005), preprint arXiv: cond-mat/05 10 064.

[36] S. N. Majumdar, A. Pal and G. Schehr, Extreme value statistics of correlated random variables: A pedagogical review, Phys. Rep. **840**, 1–32 (2020), Extreme value statistics of correlated random variables: A pedagogical review.

[37] J. Marcinkiewicz, Sur une propriété de la loi de Gauss, Mathematische Zeitschrift **44**(1), 612–618 (1939).

[38] U. M. B. Marconi, A. Puglisi, L. Rondoni and A. Vulpiani, Fluctuation-dissipation: Response theory in statistical physics, Physics Reports **461**(4-6), 111–195 (2008).

[39] H. Matsumoto and M. Yor, Exponential functionals of Brownian motion, I: Probability laws at fixed time, Probab. Surveys **2**, 312–347 (2005).

[40] H. Matsumoto and M. Yor, Exponential functionals of Brownian motion, II: Some related diffusion processes, Probab. Surveys **2**, 348–384 (2005).

[41] C. Monthus and A. Comtet, On the flux distribution in a one-dimensional disordered system, J. Phys. I (France) **4**(6), 635–653 (1994).

[42] 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)].

[43] B. Øksendal, *Stochastic differential equations : an introduction with applications*, Springer-Verlag, Heidelberg, New York, fith edition (2000).

[44] J. Perrin, Mouvement brownien et réalité moléculaire, Annales de Chimie et de Physique **9**, 1–114 (1909).

[45] M. Petit, *L'équation de Kolmogoroff. Vie et mort de Wolfgang Doeblin, un génie dans la tourmente nazie*, Ramsay (2003).

[46] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*, Springer, Berlin (1989).

[47] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, Rep. Progr. Phys. **75**(12), 126001 (2012).

[48] K. Sekimoto, *Stochastic energetics*, volume Series in modern condensed matter physics vol.10, Springer (2010).

[49] C. Texier, Individual energy level distributions for one-dimensional diagonal and off-diagonal disorder, J. Phys. A: Math. Gen. **33**, 6095–6128 (2000).

[50] C. Texier, *Physique statistique (faiblement) hors équilibre : formalisme de la réponse linéaire. Application à l'étude de la dissipation quantique et du transport électronique* (2013), notes de cours du DEA de physique quantique, École Normale Supérieure, disponible à l'adresse http://www.lptms.u-psud.fr/christophe_texier/.

[51] C. Texier and G. Roux, *Physique statistique : des processus élémentaires aux phénomènes collectifs*, Dunod, Paris (2017).

[52] H. Touchette, The large deviation approach to statistical mechanics, Phys. Rep. **478**(1-3), 1–69 (2009).

[53] N. G. van Kampen, *Stochastic processes in physics and chemistry*, North-Holland, Amsterdam (1992).

[54] M. Yor, *Exponential functionals of Brownian motion and related processes*, Springer (2000).

Index