Elementary notes on probability theory

Michel Bauer and Denis Bernard

November 12, 2008

Abstract

We summarize here some of the basics tools of probability theory with the aim of filling part of the gap between the mathematics and physics languages.

1 Measurable spaces

If Ω is any set, a σ-algebra (or tribe) on Ω is a subset of 2^Ω which contains Ω, is stable by complementation and by denumerable unions.
Thus $\mathcal{F} \subset 2^\Omega$ is a σ-algebra on Ω if
– *i*) $\Omega \in \mathcal{F}$,
– *ii*) if $A \in \mathcal{F}$ its complement $A^c \in \mathcal{F}$,
– *iii*) if $A_n \in \mathcal{F}$ for $n \in \mathbb{N}$ then $\cup_n A_n \in \mathcal{F}$.
Two trivial examples are $\mathcal{F} = \{\Omega, \emptyset\}$, the trivial σ-algebra, and $\mathcal{F} = 2^\Omega$, the total σ-algebra. Because of *ii*), *iii*) can be replaced by
– *iii'*) if $A_n \in \mathcal{F}$ for $n \in \mathbb{N}$ then $\cap_n A_n \in \mathcal{F}$,
i.e. by stability under denumerable intersections.

An arbitrary intersection of σ-algebras is still a σ-algebra. An arbitrary subset $\mathcal{S} \subset 2^\Omega$ is contained in a smallest σ-algebra denoted by $\sigma(\mathcal{S})$, the intersection of all σ-algebras containing $\mathcal{S}$. If Ω is a topological space, the smallest σ-algebra containing the open sets is called the Borel σ-algebra, usually denoted by $\mathcal{B}$.

The pair $(\Omega, \mathcal{F})$ is usually referred as a measurable space. A map f from a measurable space $(\Omega, \mathcal{F})$ to another measurable space $(\Omega', \mathcal{F}')$ is called measurable if $f^{-1}(A') \in \mathcal{F}$ whenever $A' \in \mathcal{F}'$.

Example 1 *: coin tossing.*

Think for example of a simple model of a coin tossing game. Games of length $n \geq 1$ can be modeled to take place in the space $\Omega_n \equiv \{-1, 1\}^{[1,n]}$, where 1 stands for "head" say, and -1 for "tail". Thus Ω_n is made of finite sequences $X_1, \cdots, X_n$ with values in ± 1. Infinite games take place in $\Omega \equiv \{-1, 1\}^{\mathbb{N}^*}$, which is made of infinite sequences $X_1, X_2, \cdots$. Truncation after the n^{th} term gives a map π_n from Ω onto Ω_n for each $n \geq 1$. For each n 2^{Ω_n} is a σ-algebra on Ω_n and $\mathcal{F}_n = \{\pi_n^{-1}(A), A \in 2^{\Omega_n}\}$ is a σ-algebra for Ω, which conveys intuitively the knowledge of what happens in the first n tosses of the coin. Clearly $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$ is an increasing sequence of σ-algebras. The smallest σ-algebra containing all of them, denoted by $\mathcal{F}$, is larger than the union $\cup_n \mathcal{F}_n$ (which is *not* a σ-algebra). The subset of Ω made of sequences in which 1 appears at least once is in $\mathcal{F}$, but in no $\mathcal{F}_n$. The same is true of $\{\omega \in \Omega, S_n(\omega)/n \text{ converges}\}$ where S_n is the sum of the first n steps, $S_n \equiv X_1 + \cdots + X_n$. One way to see it is to write this set as $\cap_{k=1}^{\infty} \cup_{l=1}^{\infty} \cap_{n>m>l} A_{k,m,n}$ where

$$A_{k,m,n} \equiv \left\{ \omega \in \Omega, \left| \frac{S_n(\omega)}{n} - \frac{S_m(\omega)}{m} \right| < \frac{1}{k} \right\} \in \mathcal{F}_n.$$

In probability theory, the members of $\mathcal{F}$ are assigned probabilities in a consistent way (see below), and consistency is one among many of the reasons to consider other σ-algebras than the total σ-algebra (see the coin tossing example to find other reasons).

The statistical mechanics viewpoint.

Let us explain now why σ-algebras remain most of the time behind the scene in statistical mechanics.

In statistical mechanics, the configuration space Ω is often finite (think for example of Ising variables on a finite number of sites) or sometimes denumerable (as in height models) to start with. Then $\mathcal{F} = 2^{\Omega}$ will turn out to be a consistent choice. Taking the thermodynamic limit is in general a nontrivial step from the point of view of probability theory, but the difficulties are most of the time of technical nature, and do not need to be

addressed carefully by physicists to get the correct answer to the questions they are interested in.

If Ω is finite or countable and $\mathcal{F}$ is a σ-algebra, it is not difficult to show that there is a finite or countable index set I and a partition of $\Omega = \cup_{i\in I}\Omega_i$ such that the members of $\mathcal{F}$ are the unions $\cup_{j\in J}\Omega_j$ when J runs over the subsets of I, i.e. $\mathcal{F}$ is the smallest σ-algebra containing all the sets of the partition. The Ω_i's are just the minimal elements of $\mathcal{F}$ for the inclusion. We say that $\Omega = \cup_{i\in I}\Omega_i$ is the partition associated to $\mathcal{F}$.

Hence in the context of finite or countable configuration spaces, there is an equivalence between σ-algebras and partitions. Partitions are the standard approach of statistical mechanics.

2 Probability spaces and random variables

A measure space is a triple $(\Omega, \mathcal{F}, \mu)$ where $\mathcal{F}$ is a σ-algebra on Ω and μ a map from $\mathcal{F}$ to $[0, +\infty]$ such that if A_n, $n \in \mathbb{N}$ is a sequence of disjoint members of $\mathcal{F}$ and $A = \cup_n A_n$ then $\mu(A) = \sum_n \mu(A_n)$ (μ is said to be countably additive).

Among measure spaces, probability spaces are most important in these notes, a notable exception being Poisson random measures to be defined below.

A measure space $(\Omega, \mathcal{F}, p)$ is a probability space if $p(\Omega) = 1$.

If $(\Omega, \mathcal{F}, p)$ is a probability space and $(\Omega', \mathcal{F}')$ a measurable space, a random variable X on $(\Omega, \mathcal{F}, p)$ with values $(\Omega', \mathcal{F}')$ is simply a measurable map from $(\Omega, \mathcal{F})$ to $(\Omega', \mathcal{F}')$. Quite often, random variables take values in $\mathbb{R}$ endowed with the Borel σ-algebra.

An arbitrary collection of random variables on $(\Omega, \mathcal{F}, p)$ (with possibly different target spaces) generates a σ-subalgebra of $\mathcal{F}$, namely the smallest σ-algebra for which all random variables in the collection are measurable.

A random variable X induces a probability p_X on its target space $(\Omega', \mathcal{F}')$ by $p_X(A') \equiv p(X^{-1}(A'))$, for which the notation $p(X \in A')$ is also intuitively appealing. This induced probability is called the probability distribution of X.

Example 2 *: Poisson distribution.*
Fix $\lambda \in [0, +\infty[$, take $\Omega = \{0, 1, \cdots\}$, $\mathcal{F} = 2^{\Omega}$ and, for $A \in \Omega$, $p(A) = e^{-\lambda} \sum_{n \in A} \frac{\lambda^n}{n!}$. It is immediate that $(\Omega, \mathcal{F}, p)$ is a probability space. A slight extension is when X is a random variable on a probability space $(\Omega, \mathcal{F}, p)$ with values in $\{0, 1, \cdots\}$ and probability distribution as above. This distribution is called the Poisson distribution of parameter λ.

Whenever Ω is not countable, probability distributions are usually defined by specifying probabilities of a simple subclass $\mathcal{S}$ of $\mathcal{F}$ such that $\mathcal{F} = \sigma(\mathcal{S})$, and using an extension theorem to get a probability defined on all of $\mathcal{F}$. Doing it by hand for a special case is painful. The extension theorems work when some consistency condition holds for the probabilities specified on $\mathcal{S}$. The reader can look at example 3 and refer to the literature for more details.

Example 3 *: Fair coin tossing.* [1]

Take $\Omega \equiv \{-1, 1\}^{\mathbb{N}^*}$ with σ-algebra $\mathcal{F} = \sigma(\cup_n \mathcal{F}_n)$. Each element ω of Ω is an infinite sequence $X_1, X_2, \cdots$, which we can write in a tautological way $X_1(\omega), X_2(\omega), \cdots$ and the coordinate maps $\omega \mapsto X_n(\omega)$ are measurable for $n = 1, 2, \cdots$. By construction, $\mathcal{F}_n$ is the smallest σ-algebra making $X_1, \cdots, X_n$ measurable.

Define a probability p_n on $\mathcal{F}_n$ by $p_n(A) = |\pi_n(A)|/2^n$ for $A \in \mathcal{F}_n$ (recall that $\pi_n(A)$ is a subset of Ω_n and $|\Omega_n| = 2^n$).

The probabilities p_n are consistent in the following way : if $A \in \mathcal{F}_m$ and $n \geq m$ then $A \in \mathcal{F}_m$ and $p_n(A) = p_m(A)$. So we can assemble the p_n's into a function p on $\mathcal{S} \equiv \cup_n \mathcal{F}_n$.

If A_k is a sequence of disjoint elements of $\mathcal{S}$ such that $\cup_k A_k$ is again in $\mathcal{S}$, then $p(\cup_k A_k) = \sum_k p(A_k)$. The proof is simple if tedious. This condition is enough to guaranty a consistent extension of p to a probability on $\mathcal{F} = \sigma(\mathcal{S})$. This is one of the useful extension theorems in the field.

Example 4 *: The uniform distribution.*

[1] With notations as in example 1.

Take $\Omega = [0,1]$ with the Borel σ-algebra $\mathcal{B}$. Le length $b-a$ of an open interval $]a,b[$, $0 \leq a < b \leq 1$, can be extended in a unique way to a probability measure on $(\Omega, \mathcal{B})$, called the uniform distribution, which is nothing but the well-known Lebesgue measure.

Consider the map f from $\{-1,1\}^{\mathbb{N}^*}$ to $[0,1]$ defined by $f(X_1, X_2, \cdots) = \sum_{n>0} b_n/2^n$, where $b_n \equiv (X_n+1)/2 \in \{0,1\}$ The sequence $(b_1, b_2, \cdots)$ is simply the binary expansion[2] of the real number $f(X_1, X_2, \cdots)$. As a consequence, this map is such that if $A \in \cup_n \mathcal{F}_n$, the image $f(A)$ is a finite union of closed intervals and the Lebesgue measure of $f(A)$ coincides with $p(A)$. This indicates that from a probabilistic viewpoint $([0,1], \mathcal{B}, dx)$ and $(\{-1,1\}^{\mathbb{N}^*}, \mathcal{F}, p)$ are essentially indistinguishable. In fact, one can show that $(\{-1,1\}^{\mathbb{N}^*}, \mathcal{F}, p)$ is in some precise sense equivalent as a probability space to any non-denumerable probability space. Let us give two modest illustrations. If $d \geq 2$ is and integer, on can by split a sequence $X = (X_1, X_2, \cdots)$ in d sequences, $X^{(1)} = (X_1, X_{d+1}, X_{2d+1}, \cdots), \cdots, X^{(d)} = (X_d, X_{2d}, X_{3d}, \cdots)$ to show quickly that $([0,1], \mathcal{B}, dx)$ and $([0,1]^d, \mathcal{B}, d^d x)$ are one and the same probability space. One can also split a sequence $X = (X_1, X_2, \cdots)$ into a denumerable family of sequences $X^{(1)} = (X_1, X_3, X_5, \cdots), X^{(2)} = (X_2, X_6, X_{10}, \cdots), X^{(3)} = (X_4, X_{12}, X_{20}, \cdots), \cdots$, a fact Wiener used for its original definition of Brownian motion (see below).

Example 5 *: The Gaussian distribution.*

Take $\Omega = \mathbb{R}$ with the Borel σ-algebra $\mathcal{B}$, and define $p(A) = \int_A \frac{dx}{\sqrt{2\pi}} e^{-x^2/2} \equiv \int \mathbf{1}_A \frac{dx}{\sqrt{2\pi}} e^{-x^2/2}$ for $A \in \mathcal{B}$. This is the standard Gaussian distribution. If X is a random variable which is either constant or such that $(X-b)/a$ follows the standard Gaussian distribution for some real numbers $a \neq 0$ and b, X is called a Gaussian random variable. When $b = 0$, X is called a centered Gaussian random variable.

If $X_1, \cdots, X_n$ are real random variables, the vector $(X_1, \cdots, X_n)$ is called Gaussian if any linear combination $\sum_i c_i X_i$ is a Gaussian random variable.

[2]The dyadic rationals have in fact two binary expansions, but this in not a problem from the probabilistic viewpoint because they form a set of probability 0.

For instance, if (U, V) is uniformly distributed in the unit square $[0, 1]^2$, one can check that $(\log U \cos 2\pi V, \log U \sin 2\pi V)$ is a Gaussian vector. In fact the two components are independent[3] standard Gaussian random variables. Combining this with our remarks on the "size" of the probability space of fair coin tossing, one sees that it can accommodate a countable family of independent standard Gaussian random variables (which can then be used to define Brownian motion, see below).

The statistical mechanics viewpoint.
In statistical mechanics, an energy function E on the (finite or denumerable) space Ω is given, and there is a simple formula for the relative probability of ω and ω' at temperature $T = 1/\beta$:

$$p(\omega)/p(\omega') = e^{\beta(E(\omega')-E(\omega))}.$$

The partition function $Z = \sum_{\omega\in\Omega} e^{-\beta E(\omega)}$ gives the normalization of the probability. If $Z < +\infty$, this defines a probability on $(\Omega, 2^\Omega)$.

A real random variable is any function from Ω to $\mathbb{R}$, also called an observable. If we look at another σ-algebra $\mathcal{F}$ with associated partition $\Omega = \cup_{i\in I}\Omega_i$, a random variable for $(\Omega, \mathcal{F})$ is a function from Ω to $\mathbb{R}$ constant on each Ω_i.

3 (Conditional) expectations

Let $(\Omega, \mathcal{F}, p)$ be a probability space.

Finite sums and products of random variables with values in $(\mathbb{R}, \mathcal{B})$ are again random variables. A useful quantity associated to a real random variable is its average, usually called expectation in the probabilistic context. It is defined at first only for so called simple random variables, those which can be written as $X = \sum_{i=1}^n x_i \mathbf{1}_{A_i}$ for some integer n, real numbers x_i and measurable sets $A_i \in \mathcal{F}$ for $i = 1, \cdots, n$. This decomposition is in general not unique, but the expectation, defined by $\mathbf{E}[X] \equiv \sum_{i=1}^n x_i p(A_i)$ can be shown to be well defined. If Ω is finite, every random variables is simple.

[3]The general notion of independence is recalled below.

In the other cases, one tries to approximate more general random variables by simple ones, and define the expectation by a limiting procedure. For instance, if Ω is countable, then any σ-algebra $\mathcal{F}$ is the smallest σ-algebra containing all sets of a certain partition $\Omega = \cup_{i\in I}\Omega_i$ into a finite or countable number of pieces. The most general random variable can uniquely be written $X = \sum_{i\in I} x_i \mathbf{1}_{\Omega_i}$. The limiting procedure allows to define the expectation of X under the condition $\sum_{i\in I} |x_i| p(\Omega_i) < \infty$ by the formula $\mathbf{E}[X] \equiv \sum_{i\in I} x_i p(\Omega_i)$, a formula which could also be taken as a definition in this simple case. In the general case, an expectation with values in $[0, +\infty]$ can be defined for any positive random variable, and $\mathbf{E}[X]$ can be defined if $\mathbf{E}[|X|] < +\infty$.

The statistical mechanics viewpoint.
If Ω is countable, we can consider the σ-algebra $\mathcal{F} = 2^\Omega$. A real random variable, or observable, is a function X from Ω to $\mathbb{R}$ and

$$\mathbf{E}[X] \equiv \langle X \rangle \equiv \frac{1}{Z} \sum_{\omega\in\Omega} X(\omega) e^{-\beta E(\omega)},$$

whenever the sum is absolutely convergent.

The reader is probably familiar with the notion of conditional probability : if $(\Omega, \mathcal{F}, p)$ is a probability space, $A, B \in \mathcal{F}$ and $p(B) \neq 0$ the probability of A given that B occurs (or simply the probability of A given B) is defined to be $p(A|B) \equiv p(A\cap B)/p(B)$. The events A and B are called independent if $p(A\cap B) = p(A)p(B)$ and then $p(A|B) = p(A)$. Hence conditional probabilities and independence convey the correct intuitive meaning.

Independence can be formulated at different levels. The events of a family $\{A_\alpha, \alpha \in I\}$ are called independent if $p(\cap_J A_\alpha) = \prod_J p(A_\alpha)$ for any finite subset J of I. The σ-algebras $\{\mathcal{F}_\alpha, \alpha \in I\}$ are called independent if the events $\{A_\alpha, \alpha \in I\}$ are independent whenever $A_\alpha \in \mathcal{F}_\alpha$ for all α's in I. The random variables $\{X_\alpha, \alpha \in I\}$ are called independent if the σ-algebras $\sigma(X_\alpha)$ they generate are independent. If moreover the functions $\{f_\alpha, \alpha \in I\}$ are measurable functions from $\mathbb{R}$ to $\mathbb{R}$ such that $\mathbf{E}[|f_\alpha(X_\alpha)|] < +\infty$ for $\alpha \in I$ and J is a finite subset of I, then $\mathbf{E}\left[(\prod_J f_\alpha(X_\alpha))\right] = \prod_J \mathbf{E}[(f_\alpha(X_\alpha))]$.

Conversely, this multiplicative property for all measurable functions from $\mathbb{R}$ to $\mathbb{R}$ such that $\mathbf{E}[|f_\alpha(X_\alpha)|] < +\infty$ for $\alpha \in I$ ensures that the random variables $\{X_\alpha, \alpha \in I\}$ are independent. This is easy to check for simple random variables[4].

Example 6 *: The Kolmogorov strong law of large numbers.*

Let X_n, $n = 1, 2, \cdots$ is a sequence of real independent identically distributed random variables on $(\Omega, \mathcal{F}, p)$ with partial sums $S_n = X_1 + \cdots, X_n$, $n \geq 1$.

– If $\mathbf{E}[|X_n|] < +\infty$ and $\mathbf{E}[X_n] = \mu$, the sequence S_n/n converges to μ almost surely (i.e. the subset of Ω such that S_n/n does not converges to μ fits into elements of $\mathcal{F}$ of arbitrary small probability).

– If $\mathbf{E}[|X_n|] = +\infty$, the sequence S_n/n diverges almost surely.

Example 7 *: Poisson random measures.*

If $(\Omega, \mathcal{F}, \mu)$ is a measure space and $\mathcal{F}_0 \equiv \{B \in \mathcal{F} \mid \mu(B) < \infty\}$ there exits a collection of integer valued random variables $\{N_B \,|B \in \mathcal{F}_0\}$ such that

– *i*) N_B is a Poisson random variable with mean $\mu(B)$,

– *ii*) if $B_1, \cdots, B_n \in \mathcal{F}_0$ are disjoint, the variables $N_{B_1}, \cdots, N_{B_n}$ are independent,

– *iii*) if $B, B' \in \mathcal{F}_0$, $Cov(N_B, N_{B'}) \equiv \mathbf{E}[N_B N_{B'}] - \mathbf{E}[N_B]\,\mathbf{E}[N_{B'}] = \mu(B \cap B')$.

This collection is called the Poisson random measure on $(\Omega, \mathcal{F}, \mu)$. Intuitively, a sample is a collection of points in Ω, the random variables N_B describe the number of points in region B, which follows a Poisson distribution. Disjoint regions are independent. Conditions *i*), *ii*), *iii*) ensure that the number of points in a disjoint union is (almost surely) the sum of the numbers of points in each separate region.

[4] Variables of the type $X = \sum_{i=1}^{n} x_i \mathbf{1}_{A_i}$ where one can assume without loss of generality that the the x_i's are distinct. Then $\sigma(X)$ is simply the finite σ-algebra generated by the A_i's.

This notions of expectations and conditional probabilities are combined in a very useful way in the concept of conditional expectation.

Let X be an $\mathcal{F}$ random variable with $\mathbf{E}[|X|] < +\infty$ and $\mathcal{F}'$ be a σ-subalgebra of $\mathcal{F}$. A conditional expectation of X given $\mathcal{F}'$ is an $\mathcal{F}'$ measurable random variable Ysuch that $\mathbf{E}[|Y|] < +\infty$ and

$$\mathbf{E}[X\mathbf{1}_A] = \mathbf{E}[Y\mathbf{1}_A]$$

for any $A \in \mathcal{F}'$. The notation $Y = \mathbf{E}[X|\mathcal{F}']$ is standard. Let us stress that $\mathbf{E}[X|\mathcal{F}']$ is by definition $\mathcal{F}'$ measurable. The above definition is not a constructive, but it turns out to be a characterization which is most useful to work with conditional expectations.

Note that if X is $\mathcal{F}'$ measurable, then X itself satisfies the properties of $\mathbf{E}[X|\mathcal{F}']$. One can also see that if $\mathcal{F}''$ is a σ-subalgebra of $\mathcal{F}'$,

$$\mathbf{E}[\mathbf{E}[X|\mathcal{F}']\,|\mathcal{F}''] = \mathbf{E}[\mathbf{E}[X|\mathcal{F}'']\,|\mathcal{F}'] = \mathbf{E}[X|\mathcal{F}'']\,.$$

Thus, when conditional expectations are nested, the smallest σ-algebra wins.

The general abstract argument for the existence of a conditional expectation $\mathbf{E}[X|\mathcal{F}']$ relies on the Radon Nykodim theorem or on projections in Hilbert spaces of square integrable random variables, i.e. on cornerstones of measure theory.
More modestly, we construct conditional expectations in the case when Ω is finite or countable, so that $\mathcal{F}$ is associated to a finite or countable partition $\Omega = \cup_{i\in I}\Omega_i$. Suppose that $\mathcal{F}'$ is a σ-subalgebra of $\mathcal{F}$. Then I can be partitioned into a finite or countable number of pieces $I = \cup_{j\in J} I_j$ in such a way that $\mathcal{F}'$ is the smallest σ-algebra containing all $\Omega'_j \equiv \cup_{i\in I_j}\Omega_i$. If $X = \sum_{i\in I} x_i \mathbf{1}_{\Omega_i}$ is a random variable with expectation (i.e. $\sum_{i\in I}|x_i|p(\Omega_i) < \infty$ as above), and $X' = \sum_{j\in J} x'_j \mathbf{1}_{\Omega'_j}$ is an $\mathcal{F}'$ measurable random variable, $\mathbf{E}\left[X\mathbf{1}_{\Omega'_j}\right] = \mathbf{E}\left[X'\mathbf{1}_{\Omega'_j}\right]$ says that $p(\Omega'_j)x'_j \equiv \sum_{i\in I_j} x_i p(\Omega_i)$. As $p(\Omega'_j) = \sum_{i\in I_j} p(\Omega_i)$, this formula fixes x'_j if $p(\Omega'_j) \neq 0$ but leaves the value of x'_j undetermined if $p(\Omega'_j) = 0$. Then, for any choice of the x'_j satisfying the above conditions and for $A \in \mathcal{F}'$, $\mathbf{E}[X\mathbf{1}_{A'}] = \mathbf{E}[X'\mathbf{1}_{A'}]$, i.e. X' is a conditional expectation of X given $\mathcal{F}'$. So conditional expectations exist,

but in general $\mathbf{E}[X|\mathcal{F}']$ is a class of $\mathcal{F}'$-measurable random variables that coincide except on a set of probability 0.

The statistical mechanics viewpoint.
In statistical mechanics, the starting point is not absolute probabilities, but relative probabilities. This implies the use of partition functions, and trivializes the notion of conditional expectations, in that physicists manipulate them all the time without ever giving them a special name.

If Ω is countable, we can consider the σ-algebra $\mathcal{F} = 2^{\Omega}$ associated with a partition of Ω into singletons. The context may dictate to split Ω in larger pieces. For instance, in the Ising model, we may compare different possible boundary conditions, and partition Ω accordingly. Or fix boundary conditions that imply the presence of an interface and partition the configuration space according to (part of) the position of the interface. As a last illustration in the context of the renormalization group, we may split Ω according to the value of the magnetization of blocks of spin. All these contexts lead to a partition $\Omega = \cup_{j\in J}\Omega'_j$ with associated σ-algebra $\mathcal{F}'$, partial partition functions $Z_j = \sum_{\omega\in\Omega'_j} e^{-\beta E(\omega)}$, and, if X is any observable, partial averages

$$\langle X\rangle_j \equiv \frac{1}{Z_j}\sum_{\omega\in\Omega'_j} X(\omega)e^{-\beta E(\omega)}.$$

The random variable taking the constant value $\langle X\rangle_j$ on Ω'_j is nothing but the conditional expectation of X given $\mathcal{F}'$.

It should be clear from the examples that conditional expectations are a general framework for all situations when one want to concentrate on certain degrees of freedom and average over the others. In some sense, the statistical mechanics framework is the most symmetrical, in that absolute probabilities are only a secondary concept, so that conditioning is transparent. Except for the special role played by the temperature which in principle appears in the Boltzmann weight but not in the energy function, nothing indicates that Ω itself and the associated Boltzmann weights have not been obtained by a previous conditioning.

4 Martingales and stopping times : discrete setting

If $(\Omega, \mathcal{F}, p)$ is a probability space, an increasing sequence $\mathcal{F}_n, n = 0, 1, 2, \cdots$ of σ-subalgebras of $\mathcal{F}$ is called a filtration. A sequence of random variables is also called a (random or stochastic) process. Most often, the target is the same, for all terms in the sequence. If the target is $\mathbb{R}$ one talks of a real process.
Given such a filtration,

– a sequence of random variables $A_n, n = 0, 1, 2, \cdots$ is adapted if A_n is $\mathcal{F}_n$ measurable for each n.

– a real adapted process $M_n, n = 0, 1, 2, \cdots$ is a martingale if $\mathbf{E}[|M_n|] < \infty$ for each n and $\mathbf{E}[M_n|\mathcal{F}_m] = M_m$ for $0 \leq m < n$. Note that this condition by itself implies that the sequence M_n is adapted, but for the problem that conditional expectations have ambiguities (on sets of measure zero).

– a real random variable T with values in $0, 1, 2, \cdots, +\infty$ is said to be a stopping time if the event $T \leq n$ is in $\mathcal{F}_n$ for each n, or equivalently (in this discrete setting) if the event $T = n$ is in $\mathcal{F}_n$ for each n. It is an exercise to show that $\mathcal{F}_T \equiv \{A \in \mathcal{F} : A \cap \{T \leq n\} \in \mathcal{F}_n \text{ for each } n\}$ is a σ-algebra that summarizes the information "collected up to T".

One should view the parameter n as a discrete time, $\mathcal{F}_n$ as the knowledge accumulated up to time n. An adapted sequence is one whose description at time n does not require knowledge of the future. A martingale is such that its expectation in the future given the knowledge accumulated up to now is its present value. A stopping time is a random time for which the present knowledge is enough to decide if it has occurred in the past. Note that if h is an increasing map from $\mathbb{N}$ to $\mathbb{N}$ and X_n is adapted (resp. a martingale) for $\mathcal{F}_n$, then $X_{h(n)}$ is adapted (resp. a martingale) for the filtration $\mathcal{F}_{h(n)}$.

From the general rule of nesting of expectations, $\mathbb{E}M_n = \mathbf{E}[\mathbf{E}[M_n|\mathcal{F}_0]]$, so if the sequence M_n is a martingale, $\mathbf{E}[M_n] = \mathbf{E}[M_0]$: martingales are time independent in average.

If X_n is a sequence of random variables and N is a real random variable

with values in $0, 1, 2, \cdots$, one can construct a new random variable X_N by setting $X_N(\omega) \equiv X_{N(\omega)}(\omega)$ for $\omega \in \Omega$, or equivalently, $X_N = \sum_n X_n \mathbf{1}_{N=n}$.

Part of the usefulness of martingales comes from the following : if $M_n, n = 0, 1, 2, \cdots$ is a martingale, and T is a bounded stopping time (i.e. there is an integer m such that T takes values in $0, 1, \cdots, m$), $\mathbf{E}[M_T] = \mathbf{E}[M_0]$. The proof is simple and instructive. If $T \leq m$

$$\begin{aligned}
\mathbf{E}[M_T] &= \mathbf{E}\left[\sum_{n=0}^{m} M_n \mathbf{1}_{T=n}\right] = \sum_{n=0}^{m} \mathbf{E}[M_n \mathbf{1}_{T=n}] \\
&= \sum_{n=0}^{m} \mathbf{E}[\mathbf{E}[M_m | \mathcal{F}_n] \mathbf{1}_{T=n}] \\
&= \sum_{n=0}^{m} \mathbf{E}[M_m \mathbf{1}_{T=n}] = \mathbf{E}\left[M_m \sum_{n=0}^{m} \mathbf{1}_{T=n}\right] \\
&= \mathbf{E}[M_m] = \mathbf{E}[M_0] .
\end{aligned}$$

The first equality is the definition of M_T, the third is the martingale property of the sequence M_n, the fourth is the defining property of conditional expectations because $\mathbf{1}_{T=n}$ is $\mathcal{F}_n$ measurable. The other equalities are obvious.

This result can be adapted to deal with unbounded stopping times, as we shall see in applications.

Another use of martingales is that they allow to define new probability distributions. Suppose M_n is a martingale such that $M_0 = 1$ and $M_n > 0$ (with probability 1) for $n \geq 0$. If X is an $\mathcal{F}_n$ measurable random variable for some n, define $\tilde{\mathbf{E}}[X] \equiv \mathbf{E}[XM_n]$. This is a consistent definition because the martingale property ensures that $\mathbf{E}[XM_n]$ is the same for all n's such that X is $\mathcal{F}_n$ measurable. In the same spirit, If $A \in \mathcal{F}_n$, define $\tilde{p}(A) \equiv \mathbf{E}[\mathbf{1}_A M_n]$. This defines a consistent family of probability distribution on $(\Omega, \mathcal{F}_n)$. Under technical growth conditions for M_n in n that we shall not try to make precise, $\tilde{p}$ extends to a probability measure on $(\Omega, \sigma(\{\mathcal{F}_n\}))$. Note that this σ-algebra may be strictly smaller than $\mathcal{F}$.

We start by illustrating these concepts for the simple random walk and then turn to statistical mechanics.

Example 8 *: Martingales and fair coin tossing.*[5]

Recall $\Omega \equiv \{-1,1\}^{\mathbb{N}^*}$ is a space of infinite sequences $X_1, X_2, \cdots$. Set $S_0 = 0$ and $S_n = S_{n-1} + X_n = X_1 + \cdots + X_n$ for $n \geq 1$. The σ-algebra $\mathcal{F}_n$ is the smallest σ-algebra making $X_1, \cdots, X_n$ all measurable, and an $\mathcal{F}_n$ measurable random variable is simply an arbitrary function of $X_1, \cdots, X_n$, defined on $\{-1,1\}^n$.

With the probability defined in 3, chosen to model intuitively independent tosses of a fair coin, the X_n's are easily checked mathematically to be independent random variables, and $\mathbf{E}[X_n|\mathcal{F}_m] = 0$ for $m < n$.

An adapted process is simply a sequence $A_n = f_n(X_1, \cdots, X_n)$ where f_n is a function on $\{-1,1\}^n$. As $X_{n+1}^2 = 1$, the most general function of X_{n+1} can be written in a unique way as $aX_{n+1} + b$. Hence, $f_{n+1}(X_1, \cdots, X_n, X_{n+1}) - f_n(X_1, \cdots, X_n)$ can be written in a unique way as

$$f_{n+1}(X_1, \cdots, X_n, X_{n+1}) - f_n(X_1, \cdots, X_n) \;=\; g_n(X_1, \cdots, X_n)X_{n+1} + h_n(X_1, \cdots, X_n)$$

More abstractly, any adapted process A_n can be defined recursively in a unique way by a formula $A_{n+1} - A_n = B_n X_{n+1} + C_n$ where B_n and C_n are adapted processes. This leads us to the topic of stochastic difference equations.

Introducing the notation $\Delta U_n \equiv U_{n+1} - U_n$ for finite differences, we note that $\Delta S_n = X_{n+1}$ and $\Delta n = 1$, so that the above equation can be rewritten $\Delta A_n = B_n \Delta S_n + C_n \Delta n$, which is equivalent to

$$A_n = A_0 + \sum_0^{n-1} B_m \Delta S_m + \sum_0^{n-1} C_m \Delta m.$$

The second sum looks very much like a Riemann-Stieljes sum, but the first one is of another nature because ΔS_m oscillates. In the discrete setting, this is harmless, but a good version for continuous time requires the construction of a new integral, the Itô integral. Integrals are amenable objects mostly due to the change of variable formula and integration by parts. Let

[5] With notations as in examples 1,3.

us look at their discrete counterparts. Suppose k_n is a sequence of functions from $\mathbb{R}$ to $\mathbb{R}$ and look at the process $k_n(A_n)$. The outcome is

$$\Delta k_n(A_n) = B_n^{(k)} \Delta S_n + C_n^{(k)} \Delta n$$

where

$$B_n^{(k)} = \frac{k_{n+1}(A_n + B_n + C_n) - k_{n+1}(A_n - B_n + C_n)}{2},$$

which looks like a first order derivative, and

$$\begin{aligned} C_n^{(k)} &= (k_{n+1}(A_n + C_n) - k_n(A_n)) \\ &+ \frac{k_{n+1}(A_n + B_n + C_n) - 2k_{n+1}(A_n + C_n) + k_{n+1}(A_n - B_n + C_n)}{2} \end{aligned}$$

which looks like the sum of a first derivative due to Δn in the original equation and the explicit n dependence in k_n, and a second derivative due to the oscillating nature of ΔS_n tamed by the fact that $(\Delta S_n)^2 = 1$. We could pedantically call this the discrete Itô formula. The serious Itô formula relies heavily on a continuous time counterpart of $(\Delta S_n)^2 = 1$. We shall come back to this later. For instance, if $\Delta A_n = \alpha A_n \Delta S_n$, one finds $\Delta \log A_n = \frac{1}{2} \log \frac{1+\alpha}{1-\alpha} \Delta S_n + \frac{1}{2} \log(1-\alpha^2) \Delta n$, leading to $\log \frac{A_n}{A_0} = \frac{1}{2} \log \frac{1+\alpha}{1-\alpha} S_n + \frac{1}{2} \log(1-\alpha^2) n$. In the same way, if $A_n^{(i)}$, $i = 1, 2$ are adapted processes, and $A_n = A_n^{(1)} A_n^{(2)}$, one finds that $\Delta A_n = B_n \Delta S_n + C_n \Delta n$ with $B_n = (A_n^{(1)} + C_n^{(1)}) B_n^{(2)} + B_n^{(1)} (A_n^{(2)} + C_n^{(2)})$ and $C_n = A_n^{(1)} C_n^{(2)} + C_n^{(1)} A_n^{(2)} + B_n^{(1)} B_n^{(2)}$. For instance $\Delta S_n^2 = 2 S_n \Delta S_n + \Delta n$.

Stochastic difference equations can be used in several ways. On the one hand, one can take B_n and C_n in full generality as given function of $X_1, \cdots, X_n$. On the other hand, it causes no harm and can be very useful to add a dependence in $A_1, \cdots, A_n$ in B_n and C_n, because then the recursion relation itself ensures that indeed B_n, C_n and A_n are adapted. We shall give illustrations below. An important example is when B_n and C_n are given functions b_n and c_n of A_n, and $A_{n+1} - A_n = b_n(A_n) X_{n+1} + c_n(A_n)$. This defines an adapted Markov process which is called a diffusion.

In particular, $\mathbb{E}(A_{n+1} | \mathcal{F}_n) = A_n + C_n$. So the process A_n cannot be a martingale unless $C_n = 0$, i.e. the process C_n vanishes. The law for nesting conditional expectations (the smallest σ-algebra wins) ensures recursively that this is also sufficient condition.

Hence, we have obtained a characterization of martingales in fair coin tossing : the sequence M_n is a martingale if and only if M_0 is a constant and there is an adapted process B_n such that $M_{n+1} - M_n = B_n X_{n+1}$ for $n \geq 0$. Again B_n can be viewed either as an explicit function of $X_1, \cdots, X_n$ or as defined implicitly via the recursion.

A few examples will illustrate this flexibility better than words.

If we take $B_n = 1$ for $n \geq 0$, and $M_0 = 0$ we find that $M_n = S_n \equiv \sum_{m=1}^{n} X_m$ is a martingale. If we take $B_n = 2S_n$ for $n \geq 0$, and $M_0 = 0$, we find that $M_n = S_n^2 - n$ is a martingale. In these two examples, B_n was given a priori as a function of $X_1, \cdots, X_n$. Now fix a constant α, set $M_0 = 1$ and take $B_n = \alpha M_n$ for $n \geq 0$. Then, it is the difference equation itself that ensures that B_n is adapted and M_n is a martingale. The solution to $M_{n+1} - M_n = \alpha M_n X_{n+1}$ is $M_n = \prod_{m=1}^{n}(1 + \alpha X_n)$. Write $\alpha = \tanh H$ (H may be complex) to get $M_n = e^{HS_n - n \log \cosh H}$. Observe that $M_n = 1 + HS_n + \frac{H^2}{2}(S_n^2 - n) + O(H^3)$ at small H, so that the last example contains the previous two. In these three examples, we have not proved that $\mathbf{E}[|M_n|] < +\infty$ but this is obvious because M_n is bounded on Ω for fixed n.

Example 9 *: The ruin problem, martingales and fair coin tossing.*[6]

One of the standard problems in fair coin tossing is the ruin problem. Fix two integers a, b such that $a < 0 < b$. If $-a$ and b are the initial fortunes of two gamblers and S_n is the gain of the first gambler, then he is ruined if S_n reaches a before it reaches b. Let T be the first time at which S_n reaches a or b. Because $\{T \leq n\} = \cup_{m=1}^{n}\{S_m = a\} \cup \{S_m = b\}$ and $\{S_m = a\} \cup \{S_m = b\} \in \mathcal{F}_m \subset \mathcal{F}_n$ for $m \leq n$, T is a stopping time. Obviously, T is not a bounded stopping time. However, $p(T \geq n)$ can be shown to decreases at least exponentially in n. Indeed, set $c = b - a$ and choose an integer m such that $m \geq c/2$. If I is any interval of length c and one starts the random walk somewhere in I, making m steps in the appropriate direction takes the walk to the boundary or out of I, so if k is an integer and $n \geq km$, $p(T \geq n) \leq (1 - 2^{-m})^k$. In particular, $p(T = +\infty) = 0$.

[6] With notations as in examples 1,3,8.

If we forget about the fact that T is unbounded, we get

$$\mathbf{E}[S_T] = \mathbf{E}\left[S_T^2 - T\right] = 0,$$

i.e.

$$ap(S_T = a) + bp(S_T = b) = a^2 p(S_T = a) + b^2 p(S_T = b) - \mathbf{E}[T] = 0$$

which combined with $p(S_T = a) + p(S_T = b) = 1$ yields

$$p(S_T = a) = \frac{b}{b-a} \quad p(S_T = b) = \frac{-a}{b-a} \quad \mathbf{E}[T] = -ab.$$

These results are indeed correct, but we have not justified them, and indeed, they indicate that some care is needed. Let T' be the first time at which S_n reaches b. Again, T' is a stopping time. Moreover, $p(T' = +\infty) \leq p(S_T = a)$ for any a so $p(T' = +\infty) = 0$: the walk S_n goes through b with probability 1. But $S_{T'} = b$ so obviously $E(S'_T) = b \neq 0 = S_0$. One can analyze the details of this failure by looking carefully at what happens when $a \to -\infty$.

Let us instead illustrate why $\mathbf{E}[M_T] = M_0$ holds for certain martingales despite the fact that T is an unbounded stopping time. The basic trick is to define, for integer m, $T_m = \min(m, T)$. Then for each m, T_m is a bounded stopping time and $\mathbb{E}M_{T_m} = M_0$ for any martingale, showing that it is enough to prove that $\lim_{m\to+\infty} \mathbf{E}[M_T - M_{T_m}] = 0$ for the martingale at hand to conclude. For instance, $a \leq S_n \leq b$ for $0 \leq n \leq T$ and $T_m \leq T$ for all m's. So $|S_T - S_{T_m}|$ is 0 for $m \leq T$ and $\leq b - a$ for $m > T$. Hence $\mathbf{E}[|S_T - S_{T_m}|] \leq (b-a)p(T > m)$ which goes to 0 when $m \to +\infty$. Hence $\mathbf{E}[S_T] = 0$. We get analogously that $\mathbf{E}\left[|S_T^2 - S_{T_m}^2|\right] \leq \max(-a, b)^2 p(T > m)$ which goes to 0 when $m \to +\infty$. On the other hand, $\mathbf{E}[T - T_m] = \sum_{n>m}(n-m)p(T = n)$ which again goes to 0 when $m \to +\infty$. Hence $\lim_{m\to\infty} \mathbf{E}\left[|(S_T^2 - T) - (S_{T_m}^2 - T_m)|\right] = 0$ and $\mathbf{E}\left[S_T^2 - T\right] = 0$ as announced.

As a last example, choose $M_n = e^{HS_n - n\log\cosh H}$ for real H as martingale. For $0 \leq n \leq T$, $0 < M_n \leq e^{|H|\max(b,-a)}$, and $\mathbf{E}[|M_T - M_{T_m}|] \leq p(T > m)e^{|H|\max(b,-a)}$ which goes to 0 when $m \to +\infty$. Hence $\mathbf{E}[M_T] = 1$, which gives enough information to compute the distribution of T. To keep formulæ simple, assume that $a + b = 0$. The martingale property

gives $\cosh(bH)\mathbf{E}\left[(\cosh H)^{-T}\right] = 1$ leading for instance to $\log p(T \geq n) \sim n \log \cos \pi/(2b)$, improving significantly the naive bound obtained above for the exponential decay of $p(T \geq n)$.

Example 10 *: Martingales in statistical mechanics.*

Let $\mathcal{C}$ be the configuration space of a lattice statistical model defined on a domain $\mathbb{D}$. For simplicity we assume $\mathcal{C}$ to be discrete and finite but as large as desired. Let w_c be the Boltzmann weights and $Z_{\mathbb{D}}$ the partition function, $Z_{\mathbb{D}} = \sum_{c \in \mathcal{C}} w_c$.

We imagine having introduced a family $\mathcal{Q}_T$ of partitions of the configuration space whose elements $\mathcal{C}_{\gamma_T}$ are labeled by indices γ_T:

$$\mathcal{C} = \bigcup_{\gamma_T} \mathcal{C}_{\gamma_T}, \qquad \mathcal{C}_{\gamma_T} \text{ disjoints.}$$

The index T, which will be identified with 'time', labels the partitions. By convention $\mathcal{Q}_0$ is the trivial partition with $\mathcal{C}$ as its single piece. We assume these partitions to be finer as T increases, which means that for any $S > T$ and any element $\mathcal{C}_{\gamma_T}$ of the partition at time T there exist elements of $\mathcal{Q}_S$ which form a partition of $\mathcal{C}_{\gamma_T}$.

To define a stochastic process we have to specify the probability space and a filtration on it. By construction the probability space should be the total configuration space $\mathcal{C}$ equipped with the probability measure induced by the Boltzmann weights, ie. $\mathbf{P}[\{c\}] = w_c/Z_{\mathbb{D}}$. To any partition $\mathcal{Q}_T$ is associated a σ-algebra $\mathcal{F}_T$ on $\mathcal{C}$, ie. the one generated by the elements of this partition. Since these partitions are finer as 'time' T increases, it induces a filtration $\mathcal{F}_T$ on $\mathcal{C}$ with $\mathcal{F}_S \subset \mathcal{F}_T$ for $T > S$. Physically $\mathcal{F}_T$ is the set of events, observable at 'time' T, which are unions of the sets $\mathcal{C}_{\gamma_T}$. The fact that we trivially get a filtration simply means that increasing 'time' T increases the knowledge on the system.

We define the conditioned partition function $Z_{\mathbb{D}}[\gamma_T]$ by the restricted sum:

$$Z_{\mathbb{D}}[\gamma_T] \equiv \sum_{c \in \mathcal{C}_{\gamma_T}} w_c \ .$$

Since restricting the summation to a subset amounts to impose some condition on the statistical configurations, $Z_{\mathbb{D}}[\gamma_T]$ is the partition function conditioned by the knowledge specified by $\mathcal{C}_{\gamma_T}$. In particular the probability of the event $\mathcal{C}_{\gamma_T}$ is the ratio of the partition functions

$$\mathbf{P}[\mathcal{C}_{\gamma_T}] = Z_{\mathbb{D}}[\gamma_T]/Z_{\mathbb{D}}. \tag{1}$$

Now, given an observable $\mathcal{O}$ of the statistical model, ie. a function $c \to \mathcal{O}_c$ on the configuration space, we can define its conditional average $\prec \mathcal{O} \succ_T \equiv \mathbf{E}[\mathcal{O}|\mathcal{F}_T]$. By definition of conditioned expectations, $\prec \mathcal{O} \succ_T$ is a function on the configuration space which is constant on any set $\mathcal{C}_{\gamma_T}$ such that $\mathbf{E}\left[\mathbf{1}_{\mathcal{C}_{\gamma_T}} \prec \mathcal{O} \succ_T\right] = \mathbf{E}\left[\mathbf{1}_{\mathcal{C}_{\gamma_T}} \mathcal{O}\right]$ with $\mathbf{1}_{\mathcal{C}_{\gamma_T}}$ the characteristic function of the set $\mathcal{C}_{\gamma_T} \subset \mathcal{C}$. Hence, $\prec \mathcal{O} \succ_T = \sum_{\gamma_T} \prec \mathcal{O} \succ_{|\gamma_T} \mathbf{1}_{\mathcal{C}_{\gamma_T}}$ with

$$\prec \mathcal{O} \succ_{|\gamma_T} \equiv \frac{1}{Z_{\mathbb{D}}[\gamma_T]} \sum_{c \in \mathcal{C}_{\gamma_T}} \mathcal{O}_c \, w_c. \tag{2}$$

This is simply the statistical average conditioned on the knowledge specified by the set $\mathcal{C}_{\gamma_T}$. The unconditioned statistical average is $\prec \mathcal{O} \succ_{\mathbb{D}} = Z_{\mathbb{D}}^{-1} \sum_c \mathcal{O}_c w_c$.

By construction the averages of the conditioned expectation $\prec \mathcal{O} \succ_T$ is time independent and equals to the statistical average:

$$\mathbf{E}[\prec \mathcal{O} \succ_T] = \sum_{\gamma_T} \mathbf{P}[\mathcal{C}_{\gamma_T}] \ \prec \mathcal{O} \succ_{\gamma_T} = \frac{1}{Z_{\mathbb{D}}} \sum_{c \in \mathcal{C}} \mathcal{O}_c \, w_c = \prec \mathcal{O} \succ_{\mathbb{D}} . \tag{3}$$

This is a simple but a key equation. One may be more precise and check that $\prec \mathcal{O} \succ_T$ is a (closed) martingale with respect to $\mathcal{F}_T$. Indeed, for $T > S$,

$$\mathbf{E}[\prec \mathcal{O} \succ_T |\mathcal{F}_S] = \mathbf{E}[\mathbf{E}[\mathcal{O}|\mathcal{F}_T] |\mathcal{F}_S]] = \mathbf{E}[\mathcal{O}|\mathcal{F}_S] = \prec \mathcal{O} \succ_S,$$

where we used standard properties of conditional expectations and the fact that $\mathcal{F}_T \subset \mathcal{F}_S$ for $T > S$.

It is now time to turn to the continuous time setting.

5 Brownian motion

5.1 Random processes in the large

A random process on a probability space $(\Omega, \mathcal{F}, p)$ is a family $\{X_t\}_{t\in I}$ of random variables with values in the same measurable space, where I is a totally ordered set, in concrete examples either $\{0, 1, \cdots, N\}$, $\{0, 1, \cdots\}$ (random process in discrete time), $[0, T]$ or $[0, +\infty[$ (random process in continuous time).

It can be useful to put more structure in the definition of a random process and add a filtration $\mathcal{F}_t$, $t \in I$, i.e. an increasing family of σ-algebras, all included in $\mathcal{F}$, such that $\{X_t\}$ is $\mathcal{F}_t$-measurable. Then, it is $\{X_t, \mathcal{F}_t\}_{t\in I}$ which is referred to as a random process. When no such filtration is assumed, $\mathcal{F}_t$ can be taken to be the smallest σ-algebra making $\{X_{t'}\}$ measurable for all $t' \leq t$.

The notion of adapted process, martingale and stopping time can be easily recopied from the discrete definitions for continuous time processes.

5.2 The definition of Brownian motion

We describe Brownian motion in d dimensions, starting at the origin in $\mathbb{R}^d$.

Our aim is to put a probability measure on $\Omega = \mathcal{C}_0([0, +\infty[, \mathbb{R}^d)$, the space of continuous functions γ from $[0, +\infty[$ to $\mathbb{R}^d$ such that $\gamma(0) = 0$. If $n \geq 1$, $0 < t_1 < \cdots < t_n$ and $A_1, \cdots, A_n$ are Borel subsets of $\mathbb{R}^d$, the subspace $C(\{t_m, A_m\}_{m=1,\cdots,n})$ of Ω consisting of the γ's such that $\gamma(t_i) \in A_i$ for $i = 1, \cdots, n$ is called a cylinder set. We define $\mathcal{F}$ to be the smallest σ-algebra containing cylinder sets and $\mathcal{F}_t$ the smallest σ-algebra containing cylinder sets with $C(\{t_m, A_m\}_{m=1,\cdots,n})$ with $t_n \leq t$.

The basic object to define the probability measure is the heat kernel in d dimensions $K(x, t) \equiv \frac{1}{(2\pi t)^{d/2}} \exp -\frac{\|x\|^2}{2t}$. The measure of the cylinder set $C(\{t_m, A_m\}_{m=1,\cdots,n})$ is defined to be

$$\mu(C(\{t_m, A_m\}_{m=1,\cdots,n})) \equiv$$
$$\int_{A_1} d^d x_1 \cdots \int_{A_n} d^d x_n K(x_1, t_1) K(x_2 - x_1, t_2 - t_1) \cdots K(x_n - x_{n-1}, t_n - t_{n-1}).$$

If $n \geq 1$ and A_l is the whole real line for some l, the integral over $\mathbf{x}_l$ can be performed explicitly, and $\mu(C(\{t_m, A_m\}_{m=1,\cdots,n; m\neq l}))$ is recovered. This is an obvious consistency condition if μ is to extend to a probability measure on $\mathcal{F}$. An extension theorem by Kolmogorov ensures this is also a sufficient condition.

It turns out to be important to extend the $\mathcal{F}_t$'s and $\mathcal{F}$ with sets of measure zero, but we shall remain dumb[7] on that.

Note that for $\lambda \in \mathbb{R}\backslash\{0\}$,

$$\mu(C(\{\lambda^2 t_m, \lambda A_m\}_{m=1,\cdots,n})) = \mu(C(\{t_m, A_m\}_{m=1,\cdots,n})),$$

i.e. μ is scale invariant on cylinder sets. Hence if $\lambda \in \mathbb{R}\backslash\{0\}$, and B_t is a Brownian motion, then $\frac{1}{\lambda} B_{\lambda^2 t}$ is also a Brownian motion.

Let us turn for a while to dimension $d = 1$. A point $\omega \in \Omega$ is a continuous function from $[0, +\infty[$ to $\mathbb{R}$, and the Brownian motion process is denoted by B_t, where $B_t(\omega) \equiv \omega(t)$. From the definition,

BM For $0 < t_1 < \cdots < t_n$, the vector $(B_{t_1}, B_{t_2} - B_{t_1}, \cdots, B_{t_n} - B_{t_{n-1}})$ is centered Gaussian with independent components of variance $(t_1, t_2 - t_1, \cdots, t_n - t_{n-1})$

One can generalize the notion of Brownian motion as any process B_t on a probability space such that B_t is almost surely continuous and starting at 0, and moreover satisfies **BM**.

This extension is useful for instance to construct Brownian samples. Take $0 \leq t' < t''$. First, $B_{t'}$ and $B_{t''}$ being fixed, the distribution of B_t for $t \in [t', t'']$ is independent of what has happened before t' or what will happen after t''. Second, setting

$$t = \frac{t' + t''}{2} \qquad X = \frac{2B_t - B_{t''} - B_{t'}}{(t'' - t')^{1/2}}$$

X is a standard Gaussian random variable because the numerator is the difference of two independent centered Gaussians of variance $(t'' - t')/2$. One can iterate to construct 2^n independent standard Gaussian random

[7] In all acceptations of the word, maybe.

variables from $B_{1/2^n}, \cdots, B_{k/2^n}, \cdots, B_1$. In the passage from n to $n+1$, 2^n new standard Gaussian random variables come into play. Going the other way round, one can construct Brownian samples on $t \in [0,1]$ by iterating as follows. Let h be the "tent" function, $h(t) = 2t$ on $[0, 1/2]$, $h(t) = 1 - 2t$ on $[1/2, 1]$ and $h(t) = 0$ elsewhere. Define, for $n \geq 0$ and $k = 0, \cdots, 2^n - 1$, $h_{n,k}(t) \equiv h(2^n t - k)$ and suppose that Y and $X_{n,k}$ for $n \geq 0$ and $k = 0, \cdots, 2^n - 1$ form a family of independent normalized centered Gaussian random variable on some probability space. Then one can show that $tY + \sum_{n,k} \frac{1}{2^{1+n/2}} X_{n,k} h_{n,k}(t)$ is almost surely convergent to a random continuous function of $t \in [0,1]$, call it W_t, and W_t is a Brownian process in the generalized sense. If the random variables Y and $X_{n,k}$ are sampled from a Brownian sample $B_u, u \in [0,1]$ then W_t and B_t coincide at the dyadic rationals $k/2^n$. By cutting at some n one gets nice simulations of Brownian motion. One can show that if B_t is a Brownian motion, $tB_{1/t}$ is a Brownian motion as well. So gluing in the appropriate way two independent Brownian motions on $[0,1]$ one gets a Brownian motion on $[0, +\infty[$.

The Brownian bridge ($d = 1$) can be defined via a measure on $\mathcal{C}_{0,0}([0,1], \mathbb{R})$, the space of continuous functions γ from $[0,1]$ to $\mathbb{R}$ such that $\gamma(0) = \gamma(1) = 0$. The formula for the measure of a cylinder set $C(\{t_m, A_m\}_{m=1,\cdots,n})$ with $0 < t_1 < \cdots < t_n < 1$ is defined to be

$$\begin{aligned}\mu(C(\{t_m, A_m\}_{m=1,\cdots,n})) \equiv \sqrt{2\pi} \int_{A_1} dx_1 \cdots \int_{A_n} dx_n \\ K(x_1, t_1) K(x_2 - x_1, t_2 - t_1) \cdots K(x_n - x_{n-1}, t_n - t_{n-1}) K(-x_n, 1 - t_n).\end{aligned}$$

The Brownian bridge is simply Brownian motion B_t on $[0,1]$ conditioned to $B_1 = 0$ (via a limiting procedure). If B_t is a Brownian motion, $B_t - tB_1$ is a Brownian bridge. The Brownian bridge can be simulated simply by forgetting about Y and keeping only the $X_{n,k}$'s.

5.3 Some martingales

Our starting point is Brownian motion on $\Omega = \mathcal{C}_0([0, +\infty[, \mathbb{R})$. Remember that $\mathcal{F}$ is the smallest σ-algebra containing cylinder sets and $\mathcal{F}_t$ the smallest σ-algebra containing cylinder sets $C(\{t_m, A_m\}_{m=1,\cdots,n})$ with $t_n \leq t$. As the

increments of Brownian motion are independent, $B_{t+s} - B_t$ is independent of $\mathcal{F}_t$ for $s \geq 0$, i.e. $\mathbf{E}[f(B_{t+s} - B_t)X|\mathcal{F}_t] = X\mathbf{E}[f(B_{t+s} - B_t)]$ if the random variable X is $\mathcal{F}_t$-measurable. For instance

$$\begin{aligned} \mathbf{E}[B_{t+s}|\mathcal{F}_t] &= \mathbf{E}[B_t|\mathcal{F}_t] + \mathbf{E}[B_{t+s} - B_t|\mathcal{F}_t] \\ &= B_t + \mathbf{E}[B_{t+s} - B_t] \\ &= B_t \end{aligned}$$

so B_t is a martingale (note that $\mathbf{E}[|B_t|] = \sqrt{\frac{2t}{\pi}} < +\infty$). In the same spirit, writing $B_{t+s}^2 = B_t^2 + 2B_t(B_{t+s} - B_t) + (B_{t+s} - B_t)^2$ and taking the conditional expectation with respect to $\mathcal{F}_t$ yields $\mathbf{E}\left[B_{t+s}^2|\mathcal{F}_t\right] = B_t^2 + 0 + s$ As $\mathbf{E}\left[B_t^2\right] = t < +\infty$, we conclude that $B_t^2 - t$ is a martingale. Finally, writing $e^{HB_{t+s}} = e^{HB_t}e^{H(B_{t+s}-B_t)}$ and taking the conditional expectation with respect to $\mathcal{F}_t$ yields $\mathbf{E}\left[e^{HB_{t+s}}|\mathcal{F}_t\right] = e^{HB_t}e^{sH^2/2}$. As $\mathbf{E}\left[|e^{HB_t}|\right] = e^{t(\Re e\, H)^2/2} < +\infty$ for complex H we conclude that $e^{HB_t - tH^2/2}$ is a martingale. So we have retrieved continuous time analogs for the simplest random walk martingales defined above. However, the need of a continuous analog of stochastic difference equations and stochastic sums is plain.

6 The quadratic variation of Brownian motion

In the theory of Riemann-Stieljes integrals, one defines $\int_0^t f(u)dg(u)$ by limits of Riemann sums.

If $t_0 = 0 < t_1 < \cdots t_n = t$ is a subdivision,define $V \equiv \sum_{m=0}^{n-1} |g_{t_{m+1}} - g_{t_m}|$, $S_{ret} \equiv \sum_{m=0}^{n-1} f(t_m)(g(t_{m+1}) - g(t_m))$, $S_{adv} \equiv \sum_{m=0}^{n-1} f(t_{m+1})(g(t_{m+1}) - g(t_m))$, and $2S \equiv S_{adv} + S_{ret}$. The function $g(t)$ is said to have bounded variation if V is bounded on the set of all subdivisions. It can be shown that if $g(t)$ has bounded variation and (say) $f(t)$, $g(t)$ are continuous, S_{ret}, S_{adv} and S have a common limit when the mesh $\sup_m(t_{m+1} - t_m)$ of the subdivision goes to 0.

Suppose we want to make sense of $\int_0^t 2B_s dB_s$ in the same way. So $S_{ret} = 2\sum_{m=0}^{n-1} B_{t_m}(B_{t_{m+1}} - B_{t_m})$, $S_{adv} = 2\sum_{m=0}^{n-1} B_{t_{m+1}}(B_{t_{m+1}} - B_{t_m})$ and $S = \sum_{m=0}^{n-1}(B_{t_{m+1}} + B_{t_m})(B_{t_{m+1}} - B_{t_m})$. Now S is a telescopic sum whose value

is always B_t^2. On the other hand $S_{adv} - S_{ret} = 2Q$ where $Q(t_1, \cdots, t_n) \equiv \sum_{m=0}^{n-1}(B_{t_{m+1}} - B_{t_m})^2$ a quadratic sum to be contrasted with the linear sum $V = \sum_{m=0}^{n-1}|B_{t_{m+1}} - B_{t_m}|$.

First suppose that the subdivision is regular, i.e. $t_k = tk/n$. By the scale invariance of Brownian motion, $Q(t/n, 2t/n, \cdots, nt/n)$ is distributed as $Q(t, 2t, \cdots, nt)/n$ and $Q(t, 2t, \cdots, nt)$ is a sum of n independent identically distributed random variables with average t : the strong law of large numbers (see example 6) implies that $Q(t/n, 2t/n, \cdots, nt/n)$ converges almost surely to t.

For a general subdivision, one uses the following combinatorial identity, which is proved by direct computation and is central to Itô's theory.

Assume that for $m = 0, \cdots, n-1$, X_m and Δ_m are random variables on some probability space, with the property that Δ_l and Δ_m are independent for $l \neq m$ and Δ_m is independent of X_l for $l < m$. Define $\mathbb{E}\Delta_m^2 \equiv \delta_m$ and assume that $\mathbb{E}\Delta_m^4 = 3\delta_m^2$. Then

$$\mathbb{E}(\sum_{m=0}^{n-1} X_m\Delta_m^2 - \sum_{m=0}^{n-1} X_m\delta_m)^2 = 2\sum_{m=0}^{n-1} \delta_m^2\mathbb{E}X_m^2.$$

Note that the relation between the second and fourth moment of Δ_m is true for a centered Gaussian.

As a first application, take $X_m = 1$ (a constant random variable) and $\Delta_m = B_{t_{m+1}} - B_{t_m}$. Then $\mathbb{E}Q(t_1, \cdots, t_n) = t$ and the combinatorial identity yields $\mathbb{E}(Q(t_1, \cdots, t_n) - t)^2 = 2\sum_{m=0}^{n-1}(t_{m+1} - t_m)^2 \leq 2t\sup_m(t_{m+1} - t_m)$ which goes to 0 if the mesh of the subdivision goes to 0, so that $Q(t_1, \cdots, t_n)$ converges to t in the $\mathbb{L}^2$ topology.

More generally, if X_t is a random process such that the function $\mathbb{E}X_s^2$ is (Riemann) integrable on $s \in [0, t]$ the quantity $\sum_{m=0}^{n-1}(t_{m+1} - t_m)^2\mathbb{E}X_{t_m}^2$ is $\leq \sup_m(t_{m+1} - t_m)\sum_{m=0}^{n-1}(t_{m+1} - t_m)\mathbb{E}X_{t_m}^2$ which goes to 0 if the mesh of the subdivision goes to 0. If moreover X_s is independent $B_t - B_s$ for $t > s$ and the samples of X_s are (Riemann) integrable on $[0, t]$, we infer from the combinatorial lemma that $\sum_{m=0}^{n-1} X_{t_m}(B_{t_{m+1}} - B_{t_m})^2$ converges to $\int_0^t X_s ds$ in the $\mathbb{L}^2$ topology when the mesh of the subdivision goes to 0. This leads to

the suggestive infinitesimal notation $(dB_t)^2 = dt$ which is the "miraculous" rigid analog to the discrete $X_n^2 = 1$ lying at the origin of Itô's calculus.

One could extend these results in several directions, but the point is that the Brownian motion has, for more than enough definitions of convergence, a well defined quadratic variation which is deterministic and equal to t.

The discretization S_{ret} leads to $\int_0^t 2B_s dB_s = B_t^2 - t$, while the discretization S_{adv} would lead to $\int_0^t 2B_s dB_s = B_t^2 + t$ and the discretization S to $\int_0^t 2B_s dB_s = B_t^2$. This discrepancy shows that B_t has infinite variation with probability one and some consistent convention has to be taken to define stochastic integrals. Not all conventions are equally convenient. The symmetric choice S (Stratanovich's convention) is popular in physics but there are good reasons to prefer the discretization S_{ret} (Itô's convention) because it leads to martingales, as the above special case exemplifies.

6.1 Stochastic integrals and Itô's formula

As usual, integrals are first defined for a special class of integrands, and then extended by a limiting procedure about which we shall say almost nothing. If $(B_t, \mathcal{F}_t)$ is a Brownian motion on a space $(\Omega, \mathcal{F}, p)$, a simple process $U(t)$ is a random function for which there exists an increasing deterministic sequence $0 = t_0 < t_1 < \cdots < t_n$ and a sequence of random variables $U_1, \cdots, U_n$ such that U_i is $\mathcal{F}_{t_i}$ measurable, $U(t) = U_m$ for $t \in [t_m, t_{m+1}[$ and $U(t) = 0$ for $t \geq t_n$. Then $\int U(s)dB_s \equiv \sum_{m=0}^{n-1} U_m(B_{t_{m+1}} - B_{t_m})$. If $T \geq 0$ and $U(t)$ is a simple process, then so is $U(t)\mathbf{1}_{t\in[0,T]}$ Then $\int_0^T U(s)dB_s \equiv \int U(s)\mathbf{1}_{s\in[0,T]}dB_s$. There is a deep relationship with Hilbert space theory here, and it is natural to assume that $\mathbf{E}\left[U_m^2\right] < +\infty$ for each $m = 0, \cdots, n-1$. Then $\mathbf{E}\left[(\int U(s)dB_s)^2\right] = \mathbf{E}\left[\int U(s)^2 ds\right]$, a formula at the heart of the extension of the stochastic integral to more complicated processes. This has to be done with care to avoid a wild (non measurable) behavior of the stochastic integral as a function of ω. It is easy to check that if $X(t)$ is a simple stochastic process, $\int_0^T U(s)dB_s$ is a martingale. If we take for $U(t)$ a piecewise constant interpolation of Brownian motion, we recover the definition of S_{ret}. In general $\int_0^T U(s)dB_s$, even if defined, needs not be a martingale. It is a local martingale, which is almost as useful,

because local martingales can be shown to be martingales when stopped at appropriate sequences of stopping times. The reader is referred to the literature for precise definitions. We shall almost surely make no distinction between local martingales and martingales in these notes.

The example of B_t^2 shows that differentials cannot be computed in the classical way for stochastic integrals. Indeed, we have $B_t^2 = \int_0^t 2B_s dB_s + \int_0^t ds$, where the first integral is an Itô integral and the second one an ordinary (say Riemann) integral. More generally, suppose that some process X_t can be written as $X_t = X_0 + \int_0^t U_s dB_s + \int_0^t V_s ds$ where X_0 is a constant random variable and U_t, V_t are adapted processes (then so is X_t). A short-hand notation (and nothing more) is $dX_t = U_t dB_t + V_t dt$. If $f(t,x)$ is smooth enough (three times continuously differentiable is more than enough), $f(t,X_t)$ can also be represented as an integral $f(t,X_t) = f(0,X_0) + \int_0^t P_s dB_s + \int_0^t Q_s ds$ given by Itô's formula :

$$\begin{aligned} P_t &= U_t \frac{\partial f}{\partial x}(t,X_t) \\ Q_t &= \frac{\partial f}{\partial t}(t,X_t) + V_t \frac{\partial f}{\partial x}(t,X_t) + \frac{U_t^2}{2}\frac{\partial^2 f}{\partial x^2}(t,X_t). \end{aligned}$$

Our hand waving argument goes as follows : first, we can use simple processes as approximations in the integrals defining X_t. The resulting integrals converge to X_t, and as $f(t,x)$ is continuous in x, it is enough to prove Itô's formula when U_t and V_t are simple processes. Take a small $\varepsilon > 0$ and a subdivision $0 = t_0 < t_1 < \cdots < t_n = t$ such that on each interval $[t_m, t_{m+1}[$ the processes U_s and V_s are constant but $\sup_m(t_{m+1} - t_m) < \varepsilon$. Set $\delta_m = t_{m+1} - t_m$, $\Delta_m = B_{t_{m+1}} - B_{t_m}$ and expand

$$f(t_{m+1}, X_{t_{m+1}}) - f(t_m, X_{t_m}) = f(t_m + \delta_m, X_{t_m} + U_{t_m}\Delta_m + V_{t_m}\delta_m)$$

in powers of δ_m and Δ_m. The term involving Δ_m^1 is an Itô sum, the terms involving δ_m^1 are Riemann sums. In the limit $\varepsilon \to 0$, their sum over m have a limit as an Itô or a Riemann integral. The sum over m of term involving Δ_m^2 is converted via the combinatorial identity to the same sum with Δ_m^2 replaced by δ_m plus a correction term which is small in the $\mathbb{L}^2$ topology. These terms account for Itô's formula. The error that arises

from keeping only these contribution is small (even after the sum over m) because it involves sums over m of $O(\delta_m\Delta_m, \Delta_m^3, \delta_m^2)$: though Δ_m and its powers have fluctuations, the sum over m tames these – as for Δ_m^2 via the combinatorial identity – yielding a total error of order at most $\sqrt{\varepsilon}$ (this is a time for some energetic hand waving on our side !).

Once the notion of Itô derivative is around, one can define stochastic differential equations. There are important subtleties between so-called weak and strong solutions of stochastic differential equations, but the basic tool for existence of solutions is Picard iteration as for standard differential equations. A useful particular class is formed by Itô diffusions, i.e. processes which solve an equation of the form $X_t = X_0 + \int_0^t a(s, X_s)dB_s + \int_0^t b(s, X_s)ds$.

To conclude this section, we extend the notion of quadratic variation. If the sum $\sum_{m=0}^{n-1}(X_{t_{m+1}} - X_{t_m})^2$ defined for $0 = t_0 < t_1 < \cdots < t_n = t$ has a (in general random) limit when $\sup_m(t_{m+1} - t_m)$ goes to 0, this limit is called the quadratic variation of X_t, usually denoted by $< X_t >$, a notation chosen almost surely to confuse physicists. Brownian motion has a deterministic quadratic variation t, but this is more the exception than the rule. If it exists, $< X_t >$ is a non-decreasing process. If $X_t = X_0 + \int_0^t U_s dB_s + \int_0^t V_s ds$, the process X_t has a well-defined quadratic variation and $< X_t >= \int_0^t U_s^2 ds$, as suggested by a naive formal computation. Itô's formula can be recast in differential notation as

$$df(t, X_t) = \frac{\partial f}{\partial x}(t, X_t)dX_t + \frac{\partial f}{\partial x}(t, X_t)dt + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, X_t)d < X_t > .$$

If c is a constant and X_t,Y_t have finite quadratic variation, then so does $Z_t = X_t + cY_t$, and

$$\frac{< X_t + cY_t > - < X_t > -c^2 < Y_t >}{2c} \equiv < X_t, Y_t >$$

is independent of c and called the cross-variation of X_t and Y_t.

From Itô's change of variable formula, it is easy to obtain a formula for the Itô derivative of a product : if $X_t = X_0 + \int_0^t U_s dB_s + \int_0^t V_s ds$ and $Y_t = X_0 + \int_0^t R_s dB_s + \int_0^t W_s ds$, and $Z_t = X_t + cY_t$ for some constant c, the Itô derivative of Z_t^2 is quadratic in c and the linear term yields $d(X_tY_t) = X_t dY_t + Y_t dX_t + d < X_t, Y_t >$ and $< X_t, Y_t >= \int_0^t U_s R_s ds$.

On the other hand, if X_t and Y_t are independent Brownian motions, their cross-variation is easily checked to vanish. We leave to the reader the straightforward extension of Itô's formula when $X_t = X_0 + \sum_{i=1}^{d} \int_0^t U_s^{(i)} dB_s^{(i)} + \int_0^t V_s ds$ where $(B_t^{(1)}, \cdots, B_t^{(d)})$ is a d-dimensional Brownian motion, i.e. $B_t^{(1)}, \cdots, B_t^{(d)}$ are d independent Brownian motions.

6.2 A few applications of Itô's formula

Among the things that make Brownian motion such an important and ubiquitous probabilistic object are the following characterizations :

– If X_t is a continuous process with independent stationary increments (not assumed to be Gaussian !), there are constants $\kappa \geq 0$ and c such that $X_t = \sqrt{\kappa} B_t + ct$ for some Brownian B_t.

Note that Brownian motion is a special member in a famous class of processes, those with independent stationary increments. There is a classification of these processes, and it turns out that "continuity implies gaussianity". This result is one of the ingredients in Schramm's proof that conformaly invariant interfaces are related to Brownian motion.

– If X_t is a continuous (local) martingale with quadratic variation t, then X_t is a Brownian motion.

Example 11 *: Bessel processes.*

If $(B_t^{(1)}, \cdots, B_t^{(d)})$ is a d-dimensional Brownian motion, let R_t be the "distance to the origin process",also called "Bessel process in dimension d" : $R_t = \sqrt{(B_t^{(1)})^2 + \cdots + (B_t^{(d)})^2}$. It is known that (almost surely) d dimensional Brownian motion is recurrent (i.e visits all points an infinite number of times) for $d < 2$, dense for $d = 2$ and escapes to infinity for $d > 2$.

Itô's formula yields

$$dR_t = \sum_{i=1}^{d} \frac{\partial R_t}{\partial B_t^{(i)}} dB_t^{(i)} + \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2 R_t}{\partial B_t^{(i)} \partial B_t^{(j)}} d < B_t^{(i)}, B_t^{(j)} > .$$

But $< B_t^{(i)}, B_t^{(j)} >= \delta^{i,j} t$, leading to

$$dR_t = \frac{1}{R_t} \sum_{i=1}^{d} B_t^{(i)} dB_t^{(i)} + \frac{d-1}{2R_t} dt.$$

The stochastic contribution look complicated, but one checks easily that the quadratic variation of the (local) martingale $B_t \equiv \int_0^t \frac{1}{R_s} \sum_{i=1}^{d} B_s^{(i)} dB_s^{(i)}$ is t. Hence B_t is a Brownian motion and we arrive to the conclusion that R_t satisfies the stochastic differential equation

$$dR_t = dB_t + \frac{d-1}{2R_t} dt.$$

Setting $\sqrt{d-1} X_t \equiv 2(R_t - B_t)$ yields $dX_t = \frac{2dt}{X_t - \xi_t}$ where $\sqrt{d-1}\xi_t = -2B_t$. Hence X_t satisfies the restriction of Loewner's chordal equation to the positive real axis, for $\kappa = 4/(d-1)$. This leads immediately to the transition between hulls which are simple curves which do neither hit the real axis nor have self contacts for $\kappa \leq 4$ (i.e. $d \geq 2$, when R_t does not return to the origin), and which are thick hulls for $\kappa > 4$.

Example 12 *: Time change.*

Suppose $\tau_t(\omega)$ is an adapted continuous real non-negative non-decreasing random process with $\tau_0 = 0$ almost surely. Then, for fixed τ, $T_\tau = \inf\{t, \tau_t = \tau\}$ is a stopping time, the first time at which τ_t reaches τ. From the definition of martingales and their good behavior with respect to stopping times we infer the following. If $M_t, \mathcal{F}_t$ is a martingale, then $M_{T_\tau}, \mathcal{F}_{T_\tau}$ is also a martingale (τ is the time parameter, which may not cover the whole positive real axis).

Suppose $M_t = \int_0^t U_s dB_s$ is a (local) martingale and set $\tau_t \equiv \int_0^t U_s^2 ds$, the quadratic variation of M_t. Then $M_{T_\tau}, \mathcal{F}_{T_\tau}$ is a (local) martingale too, with quadratic variation is τ. Hence $M_{T_\tau}, \mathcal{F}_{T_\tau}$ is a Brownian motion, though possibly defined only on a finite interval. This is sometimes loosely rephrased as : Brownian motion is the only continuous local martingale modulo time changes.

For fixed t, the distribution of M_t is not Gaussian at all in general. However, when looked at the stopping time T_τ it is Gaussian. Note in

passing the remarkable Skorohod theorem, which goes in the opposite direction somehow : any distribution function with 0 mean is the distribution function of Brownian motion at an appropriate stopping time.

Example 13 *: Conformal invariance of planar Brownian motion.*

From the definition of Brownian motion, if R is a rotation in d dimensions, and T, λ are positive reals, the map f from $\mathcal{C}_0([0, +\infty[, \mathbb{R}^d)$ to itself defined by

$$(fB)_t = \begin{cases} B_t & 0 \leq t \leq T \\ B_T + R(B_{T+\lambda^2(t-T)} - B_T)/\lambda & T \leq t \end{cases}$$

preserves the Brownian probability measure. This transformation can be iterated for different parameters, and vaguely speaking Brownian motion is invariant under "local" dilations and rotations when time is rescaled appropriately. When $d = 2$, conformal transformations have no shear : they preserve angles and look locally like a dilation-rotation. Hence we expect that 2d Brownian motion is conformally invariant. The simplest way to state this precisely is Itô's formula. Suppose X_t and Y_t are two independent Brownian motions, set $Z_t = X_t + iY_t$ and consider a conformal transformation $f : \mathbb{D} \subset \mathbb{C} \to \mathbb{D}' \subset \mathbb{C}$ fixing the origin. The multidimensional Itô formula combined with the Cauchy equations yields

$$df(Z_t) = f'(Z_t)dZ_t + \frac{1}{2}f''(Z_t)d(< X_t > - < Y_t > +2i < X_t, Y_t >).$$

Observe that $< X_t > - < Y_t > +2i < X_t, Y_t >$ could be seen as the quadratic variation of the complex process Z_t if we would accept to put $c = i$ in our definition of the cross variation, but this would be nothing but a convention. Now $< X_t >=< Y_t >= t$ and $< X_t, Y_t >= 0$ if X_t and Y_t are independent Brownian motions. One infers first that $< X_t > - < Y_t > +2i < X_t, Y_t >= 0$ so that $f(Z_t)$ is a (local) martingale. Second one infers that $d < \mathfrak{Re}\, f(Z_t) >= d < \mathfrak{Im}\, f(Z_t) >= |f'(Z_t)|^2 dt$ and $d < \mathfrak{Re}\, f(Z_t), \mathfrak{Im}\, f(Z_t) >= 0$. Thus, the same time change $\tau = \int_0^t |f'(Z_s)|^2 ds$ turns the real and imaginary parts of $f(Z_t)$ into Brownian motions, which

are Gaussian processes, so that the vanishing of the cross variation ensures independence. Hence $f(Z_t)$ is a two dimensional Brownian motion after a time chance, proving the conformal invariance of the two dimensional Brownian motion.

Example 14 *: Girsanov's theorem.*

We have already seen in the discrete setting that martingales can be used to deform probability laws. Let us illustrate the great flexibility gained in the continuous setting.

Let $M_t, \mathcal{F}_t$ be a nonnegative martingale on $(\Omega, \mathcal{F}, p)$ such that $M_0 = 1$. If X is $\mathcal{F}_s$-measurable and $t \geq s$ then basic rules of conditional expectations yield $\mathbf{E}[XM_t] = \mathbf{E}[XM_s]$ so that we can make a consistent definition $\tilde{\mathbf{E}}[X] \equiv \mathbf{E}[XM_t]$ whenever X is $\mathcal{F}_t$ measurable. Then $\tilde{\mathbf{E}}[\cdots]$ is easily seen to be a positive linear functional with $\tilde{\mathbf{E}}[1] = 1$. Hence the definition $\tilde{p}_t(A) \equiv \tilde{\mathbf{E}}[\mathbf{1}_A]$ for $A \in \mathcal{F}_t$ makes $(\Omega, \mathcal{F}_t, \tilde{p}_t)$ a probability space. Under some technical growth conditions on M_t, $\tilde{p}_t$ extends to a probability law on $\sigma\{\cup_t \mathcal{F}_t\}$

Now suppose that

$$M_t = e^{\int_0^t V_s \, dB_s - \frac{2}{2} V_s^2 \, ds}$$

for some adapted process V_s. Itô's formula shows that M_t satisfies the stochastic integral equation $M_t = 1 + \int_0^t M_s V_s dB_s$ and is a (local) martingale. Let X_t be a process satisfying $X_t = B_t - \int_0^t V_s ds$. Girsanov's theorem states that for each $T > 0$, X_t is a Brownian motion on $[0, T]$ for $(\Omega, \mathcal{F}_T, \tilde{p}_T)$. Here are elements of a heuristic proof.

A simple special case is $M_t \equiv e^{HB_t - tH^2/2}$, which we know is a martingale on the Brownian motion space satisfying the conditions above. Conversely, suppose that W_t is a continuous process such that $e^{HW_t - tH^2/2}$ is a martingale on some probability space with a filtration $\mathcal{F}_t$. If $t_1 < \cdots < t_n < t$, $e^{\sum_{m=1}^n H_m W_{t_m}}$ is $\mathcal{F}_{t_n}$ measurable and

$$\mathbf{E}\left[e^{\sum_{m=1}^n H_m W_{t_m} + HW_t - tH^2/2}\right] = \mathbf{E}\left[e^{\sum_{m=1}^n H_m W_{t_m} + HW_{t_n} - t_n H^2/2}\right].$$

This leads to a recursive formula

$$\mathbf{E}\left[e^{\sum_{m=1}^n H_m W_{t_m} + WB_t}\right] = e^{(t-t_n)H^2/2} \mathbf{E}\left[e^{\sum_{m=1}^n (H_m + H\delta_{m,n}) W_{t_m}}\right]$$

from which it follows that the finite dimensional distributions of the continuous process W_t are those of a Brownian motion, so that W_t is a Brownian motion.

We can now go to the case of a general M_t again. The lesson of the previous computation is that to show that X_t is a Brownian motion for $(\Omega, \mathcal{F}_T, \tilde{p}_T)$, it is enough to show that X_t is continuous in t and that $e^{HX_t - tH^2/2}$ is a martingale for $(\Omega, \mathcal{F}_T, \tilde{p}_T)$ i.e. that $e^{HX_t - tH^2/2} M_t$ is a (local) martingale for the original probability law. We compute its Itô derivative using Itô's change of variable and product formulæ. First $de^{HX_t - tH^2/2} = e^{HX_t - tH^2/2} H dX_t$, then $d(e^{HX_t - tH^2/2} M_t) = e^{HX_t - tH^2/2} M_t H dB_t$ as announced.

Path integrals trivialize (for good or bad) this argument : one writes the Wiener measure as $\mathcal{D}x(s) \exp -\frac{1}{2}\int_0^t \dot{x}(s)^2 ds$ and in the same notation,

$$M_t = \exp[\int_0^t V(s)\dot{x}(s)ds - \frac{1}{2}\int_0^t V(s)^2 ds].$$

This is misleading because in general $V(s)$ may depend on $x(s')$ for all $s' < s$. The full measure becomes $\mathcal{D}x(s) \exp -\frac{1}{2}\int_0^t (\dot{x}(s) - V(s))^2 ds$ and a formal triangular change of variables $y(s) = x(s) - \int_0^t V(s)ds$ combined with translation invariance of the (non existing) Lebesgue measure $\mathcal{D}x(s)$ yields Girsanov's result.

As a simple application, take again $M_t \equiv e^{HB_t - tH^2/2}$. Then $X_t = B_t - Ht$ is a Brownian motion with constant drift, which looks like a Brownian motion again when the original measure is multiplied by $e^{HB_t - tH^2/2}$.

Note that the Bessel process R_t in d dimensions also becomes a Brownian motion when the original measure is multiplied by

$$M_t = \exp[-\frac{d-1}{2}\int_0^t \frac{dB_s}{R_s} - \frac{(d-1)^2}{8}\frac{ds}{R_s^2}].$$