Stochastic Processes
- lecture notes -

Matteo Caiaffa

Academic year 2011-2012

Contents

Chapter 1

Introduction to Stochastic processes

1.1 Propaedeutic definitions and theorems

Definition 1.1.1. (of probability space). A *probability space* is a triple $(\Omega, \mathscr{E}, \mathbb{P})$ where

(i) Ω is any (non - empty) set, it is called the *sample space*,

(ii) $\mathscr{E}$ is a σ - algebra of subsets of Ω (whose elements are called *events*). Clearly we have $\{\varnothing, \Omega\} \subset \mathscr{E} \subset \mathscr{P}(\Omega)$ $(= 2^{\Omega})$,

(iii) $\mathbb{P}$ is a map, *called probability measure* on $\mathscr{E}$, associating a number with each element of $\mathscr{E}$. $\mathbb{P}$ has the following properties: $0 \leq \mathbb{P} \leq 1$, $\mathbb{P}(\Omega) = 1$ and it is countably additive, that is, for any sequence A_i of disjoint elements of $\mathscr{E}$, $\mathbb{P}(\bigcup_i^\infty A_i) = \sum_{i=1}^\infty \mathbb{P}(A_i)$.

Definition 1.1.2. (of random variable). A *random variable* is a map $X : \Omega \to \mathbb{R}$ such that $\forall\ I \in \mathbb{R} : X^{-1}(I) \in \mathscr{E}$ (that is, a random variable is an $\mathscr{E} - measurable$ map), where I is a Borel set.

Let us consider the family $\mathscr{F}_X$ of subsets of Ω defined by $\mathscr{F}_X = \{X^{-1}(I), I \in \mathscr{B}\}$, where $\mathscr{B}$ is the σ - algebra of the Borel sets of $\mathbb{R}$. It is easy to see that $\mathscr{F}_X$ is a σ - algebra, so we can say that X is a random variable if and only if $\mathscr{F}_X \in \mathscr{E}$. We can also say that $\mathscr{F}_X$ is the least σ - algebra that makes X measurable.
Definition 1.1.2 gives us the following relation: $\Omega \xrightarrow{X} \mathbb{R}$ and $\mathscr{E} \xleftarrow{X^{-1}} \mathscr{B}$. We now introduce a measure (associated with the random variable X) on the Borel sets of $\mathbb{R}$:

$$\mu_X(I) := \mathbb{P}(X^{-1}(I))$$

and, resuming,

$$
\begin{array}{ccc}
(\Omega, & \mathscr{E}, & \mathbb{P}) \\
\Big\downarrow X & \Big\uparrow X^{-1} & \Big\downarrow \\
(\mathbb{R}, & \mathscr{B}, & \mu_X)
\end{array}
$$

Definition 1.1.3. The measure μ_X is called the *probability distribution*[1] of X and it is the image measure of $\mathbb{P}$ by means of the random variable X.

The measure μ_X is completely and uniquely determined by the values that it assumes on the Borel sets I.[2] In other words, one can say that μ_X turns out to be uniquely determined by its *distribution function*[3] F_X:

$$F_X = \mu_x(]-\infty, t]) = \mathbb{P}(X \leq t). \tag{1.1.1}$$

Definition 1.1.4. (of random vector). A *random vector* is a random variable with values in $\mathbb{R}^d$, where $d > 1$. Formally, a random vector is a map $X : \Omega \to \mathbb{R}^d$ such that $\forall\ I \in \mathscr{B} : X^{-1}(I) \in \mathscr{E}$.

The definition just stated provides the relations $\Omega \xrightarrow{X} \mathbb{R}^d$ and $\mathscr{E}^d \xleftarrow{X^{-1}} \mathscr{B}$, where $\mathscr{B}$ is the smallest σ - algebra containing the open sets of $\mathbb{R}^d$. Moreover, $\mathbb{P}$ corresponds with an image measure μ over the Borel sets of $\mathbb{R}^d$.

Definition 1.1.5. (of product σ - algebra). Consider as $I = I_1 \times I_2 \times I_3 \times \dots$ (rectangles in $\mathbb{R}^d$) where $I_k \in \mathscr{B}$. The smallest σ - algebra containing the family of rectangles $I_1 \times I_2 \times I_3 \times \dots$ is called *product σ-algebra*, and it is denoted by $\mathscr{B}^d = \mathscr{B} \otimes \mathscr{B} \otimes \mathscr{B} \otimes \dots$.

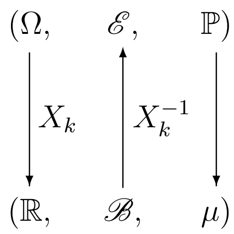
Theorem 1.1.6. *The σ-algebra generated by the open sets of $\mathbb{R}^d$ coincides with the product σ-algebra $\mathscr{B}^d$.*

Theorem 1.1.7. *$X = (X_1, X_2, \dots, X_d)$ is a random vector if and only if each of its component X_k is a random variable.*

Thanks to the previous theorem, given a random vector $X = (X_1, X_2, \dots, X_d)$ and a product σ - algebra $\mathscr{B}^d$, for any component X_k the following relations holds: $\Omega \xrightarrow{X_k} \mathbb{R}$ and $\mathscr{E}^d \xleftarrow{X_k{}^{-1}} \mathscr{B}$. To complete this scheme of relations, we introduce a measure associated

[1] For the sake of simplicity, we will omit the index X where there is no possibility of misunderstanding.
[2] Moreover, μ_X is a σ - additive measure over $\mathbb{R}$.
[3] In italian: Funzione di ripartizione.

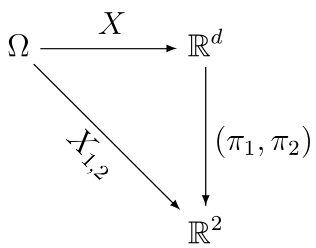
with $\mathbb{P}$. Consider d monodimensional measures μ_k and define $\nu(I1 \times I_2 \times \ldots \times I_d) := \mu_1(I_1)\mu_2(I_2)\ldots\mu_d(I_d)$.



Theorem 1.1.8. *The measure ν can be extended (uniquely) to a measure over $\mathscr{B}^d$. ν is called product measure and it is denoted by $\mu_1 \otimes \mu_2 \otimes \ldots \otimes \mu_d$.*[4]

Definition 1.1.9. If $\mu \equiv \mu_1 \otimes \mu_2 \otimes \ldots \otimes \mu_d$, the k measures μ_k are said to be *independent*.

Example 1.1.10. Consider a 2 - dim random vector $X_{12} := (X_1, X_2)$ and the Borel set $J = I \times I$ with $I \subset \mathbb{R}$. We obviously have $\Omega \xrightarrow{X} \mathbb{R}^d$ and $\mathscr{E}^d \xleftarrow{X_{1,2}{}^{-1}} \mathscr{B}^2$. Moreover, the map $\mathbb{R}^d \xrightarrow{(\pi_1,\pi_2)} \mathbb{R}^2$, where (π_1, π_2) are the projections of $\mathbb{R}^d$ onto $\mathbb{R}^2$, fulfills a commutative diagram between Ω, $\mathbb{R}^d$ and $\mathbb{R}^2$.



By definition of *image measure*, we are allowed to write:

$$\mu_{1,2}(J) = \mathbb{P}(X_{1,2}^{-1}(J)) = \mu(J \times \mathbb{R} \times \ldots \times \mathbb{R})$$

and, in particular

$$\mu_{1,2}(I \times \mathbb{R}) = \mu_1(I), \quad \mu_{2,1}(I \times \mathbb{R}) = \mu_2(I).$$

In addition to what we have recalled, using the well-known definition of *Gaussian measure*, the reader can also verify an important proposition:

Proposition 1.1.11. *μ is a Gaussian measure.*

[4]In general, $\mu \neq \mu_1 \otimes \mu_2 \otimes \ldots \otimes \mu_d$.

1.2 Some notes on probability theory

We recall here some basic definitions and results which should be known by the reader from previous courses of probability theory.
We have already established (definition 1.1.3) how the probability distribution is defined. However we can characterize it by means of the following complex - valued function:

$$\varphi(t) = \int_{\mathbb{R}} e^{itx}\mu(dx) = \int_{\mathbb{R}} e^{itx}dF = \mathbb{E}(e^{itx}). \tag{1.2.1}$$

Where $\mathbb{E}(X) = \int_\Omega X d\mathbb{P}$ is the *expectation value* of X.

Definition 1.2.1. (of characteristic function). $\varphi(t)$ (in equation 1.2.1) is called the *characteristic function* of μ_x (or the characteristic function of the random variable X).

Theorem 1.2.2. (of Bochner). *The characteristic function has the following properties*

(i) $\varphi(0) = 1$,

(ii) φ *is continuous in* 0,

(iii) φ *is positive definite, that is, for any* n *with* n *an integer, for any choice of the* n *real numbers* $t_1, t_2, \dots, t_n$ *and the* n *complex numbers* $z_1, z_2, \dots, z_n$*, results*

$$\sum_{i,j=1}^{n} \varphi(t_i - t_j) z_i \bar{z}_j \geq 0.$$

We are now interested to establish a fundamental relation existing between a measure and its characteristic function. To achieve this goal let us show that, given a random variable X and a Borel set $I \in \mathscr{B}^d$, one can define a measure μ using the following integral

$$\mu(I) = \int_I g(x)dx, \tag{1.2.2}$$

with dx the ordinary Lebesgue measure in $\mathbb{R}^d$, and

$$g_{m,A} = \left(\frac{1}{\sqrt{2\pi}}\right)^d \frac{1}{\sqrt{detA}} e^{-\frac{1}{2}\langle A^{-1}(x-m),(x-m)\rangle}, \tag{1.2.3}$$

where $m \in \mathbb{R}^d$ is a number (*measure value*) and $A = d \times d$ is a covariance matrix, symmetric and positive definite.
Notice that the Fourier transform of the measure is just the characteristic function

$$\varphi(t) = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle}\mu(dx) \tag{1.2.4}$$

and in particular, for a Gaussian measure, the characteristic function is given by the following expression:

$$\varphi_{g_{m,A}}(t) = e^{i\langle m,t\rangle - \frac{1}{2}\langle At,t\rangle}. \tag{1.2.5}$$

Some basic inequalities of probability theory.

(i) **Chebyshev's inequality:** If $\lambda > 0$

$$P(|X| > \lambda) \leq \frac{1}{\lambda^p}\mathbb{E}(|X|^p).$$

(ii) **Schwartz's inequality:**

$$\mathbb{E}(XY) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}.$$

(iii) **Hölder's inequality:**

$$\mathbb{E}(XY) \leq [\mathbb{E}(|X|^p)]^{\frac{1}{p}}[\mathbb{E}(|Y|^q)]^{\frac{1}{q}},$$

where $p, q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$.

(iv) **Jensen's inequality:** If $\varphi : \mathbb{R} \to \mathbb{R}$ is a convex function and the random variable X and $\varphi(X)$ have finite expectation values, then

$$\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi(X)).$$

In particular for $\varphi(x) = |x|^p$, with $p \geq 1$, we obtain

$$|\mathbb{E}(X)|^p \leq \mathbb{E}(|X|^p).$$

Different types of convergence for a sequence of random variables X_n, $n = 1, 2, 3, \dots$.

(i) **Almost sure (a.s.) convergence:** $X_n \xrightarrow{a.s.} X$, if

$$\lim_{n\to\infty} X_n(\omega) = X(\omega), \quad \forall \omega \notin N, \quad \text{where} P(N) = 0.$$

(ii) **Convergence in probability:** $X_n \xrightarrow{p} X$, if

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0, \quad \forall \varepsilon > 0.$$

(iii) **Convergence in mean of order** $p \geq 1$**:** $X_n \xrightarrow{L_p} X$, if

$$\lim_{n\to\infty} \mathbb{E}(|X_n - X|^p) = 0.$$

(iv) **Convergence in law:** $X_n \xrightarrow{L} X$, if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x),$$

for any point x where the distribution function F_X is continuous.

1.3 Conditional probability

Theorem 1.3.1. (of Radon - Nikodym). *Let $\mathscr{A}$ be a σ - algebra over X and $\alpha, \nu : \mathscr{A} \to \mathbb{R}$ two σ - additive real - valued measures. If α is positive and absolutely continuous with respect to ν, then there exists a map $g \in L^1(\alpha)$, called density or derivative of ν with respect to α, such that*

$$\nu(A) = \int_A g \; d\alpha,$$

for any $A \in \mathscr{A}$. Moreover each other density function is equal to g almost everywhere.

Theorem 1.3.2. (of dominated convergence). *Let $\{f_k\}$ be a increasing sequence of measurable non - negative function, $f_k : \mathbb{R}^n \to [0, +\infty]$, then if $f(x) = \lim_{k\to\infty} f_k(x)$ one has*

$$\int_{\mathbb{R}^n} f(x) \; d\mu(x) = \lim_{k\to\infty} f_k(x) \; d\mu(x)$$

Lemma 1.3.3. *Consider the probability space $(\Omega, \mathscr{E}, \mathbb{P})$. Let $\mathscr{F} \subseteq \mathscr{E}$ be a σ - algebra and $A \subseteq \Omega$. Then there exists a unique $\mathscr{F}$ - measurable function $Y : \Omega \to \mathbb{R}$ such that for every $F \subseteq \mathscr{F}$*

$$\mathbb{P}(A \cap F) = \int_F Y(\omega)\mathbb{P}(d\omega).$$

Proof $\quad F \mapsto \mathbb{P}(A \cap F)$ is a measure absolutely continuous with respect to $\mathbb{P}$ thus, by Randon - Nikodym theorem, there exists Y such that

$$\mathbb{P}(A \cap F) = \int_F Y d\mathbb{P}.$$

Finally

$$\int_F Y d\mathbb{P} = \int_F Y' d\mathbb{P} \quad \Rightarrow \quad \int_F (Y - Y') d\mathbb{P} = 0 \quad \text{for any } F \subseteq \Omega,$$

that implies $Y = Y'$ almost everywhere with respect to $\mathbb{P}$. □

Definition 1.3.4. (of conditional probability). We call conditional probability of A with respect to $\mathscr{F}$ and it is denoted by $\mathbb{P}(A|\mathscr{F})$, the only Y such that

$$\mathbb{P}(A \cap F) = \int_F Y(\omega)\mathbb{P}(d\omega).$$

Chapter 2

Gaussian Processes

2.1 Generality

Definition 2.1.1. (of stochastic process). A stochastic process is a family of random variables $\{X_t\}_{t\in T}$ defined on the same probability space $(\Omega, \mathscr{E}, \mathbb{P})$. The set T is called the *parameter set* of the process. A stochastic process is said to be a *discrete parameter process* or a *continuous parameter process* according to the discrete or continuous nature of its parameter set. If t represents time, one can think of X_t as the *state* of the process at a given time t.

Definition 2.1.2. (of trajectory). For each element $\omega \in \Omega$, the mapping

$$t \longrightarrow X_t(\omega)$$

is called the *trajectory* of the process.

Consider a continuous parameter process $\{X_t\}_{t\in[0,T]}$. We can take out of this family a number $d > 0$ of random variables $(X_{t_1}, X_{t_2}, \dots, X_{t_d})$. For example, for $d = 1$ we can take a random variable anywhere in the set $[0, T]$ (this set being continuous) and, from what we know about a random variable and a probability space, we get the maps $\Omega \xrightarrow{X_t} \mathbb{R}$ and $\mathscr{E} \xleftarrow{X_t{}^{-1}} \mathscr{B}$. Moreover, we can associate a measure μ_t to $\mathbb{P}$. Proceeding in this way for all the (infinite) X_t, we obtain a family of 1 - dim measures $\{\mu_t\}_{t\in[0,T]}$ (*1 - dim marginals*).

Example 2.1.3. Let $\{X_t\}_{t\in[0,T]}$ be a stochastic process defined on the probability space $(\Omega, \mathscr{E}, \mathbb{P})$, then consider the random vector $X_{1,2} = (X_{t_1}, X_{t_2})$. We can associate with $\mathbb{P}$ a 2 - dim measure μ_{t_1,t_2} and, for every choice of the components of the random vector, we construct a (infinite) family of 2 - dim measures $\{\mu_{t_1,t_2}\}_{t_1,t_2\in[0,T]}$.

In general, to each stochastic process corresponds a family $\mathscr{M}$ of marginals of finite dimensional distributions:

$$\mathscr{M} = \{\{\mu_t\}, \{\mu_{t_1,t_2}\}, \{\mu_{t_1,t_2,t_3}\}, \ldots, \{\mu_{t_1,t_2,\ldots,t_n}\}, \ldots\}. \tag{2.1.1}$$

A stochastic process is a model for *something*, it is a description of some random phenomena evolving in time. Such a kind of model does not exists by itself, but has to be *built.* In general, one starts from $\mathscr{M}$ to construct it (we shall see later, with some examples, how to deal with this construction). An important characteristic to notice about the family $\mathscr{M}$ is the so-called *compatibility property* (or *consistency*).

Definition 2.1.4. (of compatibility property). If $\{t_{k_1} < \ldots < t_{k_m}\} \subset \{t_1 < \ldots < t_n\}$, then $\mu_{t_{k_1},\ldots,t_{k_m}}$ is the marginal of $\mu_{t_1,\ldots,t_n}$, corresponding to the indexes $k_1, k_2, \ldots, k_m$.

Definition 2.1.5. (of Gaussian process). A process characterized by the property that each finite - dimensional measure $\mu \in \mathscr{M}$ has a normal distribution, is said to be a *Gaussian stochastic process.* It is customary to denote the mean value of such a process by $m(t) := \mathbb{E}(X_t)$ and the covariance function with $\varphi(s,t) := \mathbb{E}[(X_t - m(t))(X_s - m(s))]$, where $s, t \in [0, T]$.

According to the previous definition, we can write the expression of the generic n - dimensional measure as:

$$\mu_{t_1,\ldots,t_n}(I) := \int_I \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sqrt{\det(A)}}\, e^{-\frac{1}{2}\langle A^{-1}(x-m_{t_1,\ldots,t_n}), x-m_{t_1,\ldots,t_n}\rangle}\, dx, \tag{2.1.2}$$

where $x \in \mathbb{R}^n$, I is a Borel set, $m_{t_1,\ldots,t_n} = (m(t_1), \ldots, m(t_n))$ and $A = a_{ij} = \varphi(t_i, t_j)$, $\forall\ i, j = 1 \ldots n$. In equation 2.1.2 the covariance matrix A must be invertible. To avoid this constraint, the characteristic function (which does not depend on A^{-1}) can turn to be helpful:

$$\varphi_{t_1,\ldots,t_n}(z) = \int_{\mathbb{R}^n} e^{\boldsymbol{i}\langle z,x\rangle}\, \mu_{t_1,\ldots,t_n}(dx) = e^{\boldsymbol{i}\langle m_{t_1,\ldots,t_n}, z\rangle - \frac{1}{2}\langle Az, z\rangle}.$$

2.2 Markov process

Definition 2.2.1. (of Markov process). Given a process $\{X_t\}$ and a family of transition probability $p(s, x; t, I)$, $\{X_t\}$ is said to be a Markov process if the following conditions are satisfied.

(i) $(s, x, t) \to p(s, x; t, I)$ is a measurable function,

(ii) $I \to p(s, x; t, I)$ is a probability measure,

(iii) $p(s, x; t, I) = \int_{\mathbb{R}} p(s, x; r, dz) p(r, z; t, I)$,

(iv) $p(s, x; s, I) = \delta_x(I)$,

(v) $\mathbb{E}(\mathbf{1}_I(X_t)|\mathscr{F}_s) = p(s, X_s; t, I) \quad = \mathbb{E}(\mathbf{1}_I(X_t)|(X_s))$, where $\mathscr{F}_t = \sigma\{X_u : u \leq t\}$.

Property (iii) is equivalent to state: $\mathbb{P}(X_t \in I|X_s = x, X_{s_1} = x_1, \ldots, X_{s_n} = x_n) = \mathbb{P}(X_t \in I|X_s = x)$. One can observe that the family of finite dimensional measures $\mathscr{M}$ does not appear in the definition.[1] Anyway $\mathscr{M}$ can be derived from the transition probability. Setting $I = (I_1 \times I_2 \times \ldots \times I_n)$ and $G = (I_1 \times I_2 \times \ldots \times I_{n-1})$:

$$\mu_{t_1,t_2,\ldots,t_n}(I)$$

$$\begin{aligned}
&= \mathbb{P}\left((X_{t_1}, \ldots, X_{t_n}) \in I\right) \\
&= \mathbb{P}\big((X_{t_1}, \ldots, X_{t_{n-1}}) \in G, X_{t_n} \in I_n\big) \\
&= \int_G \mathbb{P}(X_{t_n} \in I_n|X_{t_{n-1}} = x_{n-1}, \ldots, X_{t_1} = x_1)\mathbb{P}(X_{t_1} \in dx_1, \ldots, X_{t_{n-1}} \in dx_{n-1}) \\
&= \int_G p(t_{n-1}, x_{n-1}; t_n, I_n)\mathbb{P}(X_{t_1} \in dx_1, \ldots, X_{t_{n-1}} \in dx_{n-1}) \\
&= \int_G p(t_{n-1}, x_{n-1}; t_n, I_n)p(t_{n-2}, x_{n-2}; t_{n-1}, dx_{n-1})\mathbb{P}(X_{t_1} \in dx_1, \ldots, X_{t_{n-2}} \in dx_{n-2}).
\end{aligned}$$

By reiterating the same procedure:

$$\mu_{t_1,t_2,\ldots,t_n}(I)$$

$$= \int_{\mathbb{R}} \int_G \mathbb{P}(X_0 \in dx_0)p(0, x_0; t_1, dx_1)p(t_1, x_1; t_2, dx_2) \ldots p(t_{n-1}, x_{n-1}; t_n, dx_n),$$

where $\mathbb{P}(X_0 \in dx_0)$ is the initial distribution.

2.2.1 Chapman - Kolmogorov relation

Let X_t be a Markov process and, in particular, with $s_1 < s_2 < \cdots < s_n \leq s < t$

$$\mathbb{P}(X_t \in I|X_s = x, X_{s_n} = x_n, \ldots, X_{s_1} = x_1) = \mathbb{P}(X_t \in I|X_s = x).$$

Denote the probability $\mathbb{P}(X_t \in I|X_s = x)$ with $p(s, x; t, I)$. Then the following equation

$$p(s, x; t, I) = \int_{\mathbb{R}} p(s, x; r, dz)p(r, z; t, I), \quad s < t, \tag{2.2.1}$$

holds and it is called the *Chapman - Kolmogorov* relation. Under some assumptions, one can write this relation in a different way, giving birth to various families of transition probability.

[1] We will see later (first theorem of Kolmogorov) how $\mathscr{M}$ can characterise a stochastic process.

2.2.2 Reductions of transition probability

In general there are two ways to study a Markov process:

(i) a stochastic method: studying the trajectories (X_t)(e.g. by stochastic differential equation),

(ii) an analytic method: studying directly the family $\{p(s,x;t,I)\}$ (transition probability).

Take account of the second way.

1^{st} simplification: Homogeneity in time
Consider the particular case in which we can express the dependence on s,t by means of a single function $\varphi(\tau,x,I)$ which dependes on the difference $t-s=:\tau$. So we have[2]

$$\{p(s,x;t,I)\} \to \{p(t,x,I)\}. \tag{2.2.2}$$

It is easy to understand how propositions (i),(ii) and (iv) in the definition of Markov process change: fixing x we have that the map $(t,x)\to p(t,x,I)$ is measurable, $I\to p(t,x,I)$ is a probability measure and $p(0,x,t)=\delta_x(I)$. As far as the Chapman - Kolmogorov relation is concerned, the previous change of variables produces

$$p(t_1+t_2,x,I)=\int_{\mathbb{R}} p(t_1,x,dz)p(t_2,z,I)$$

or, changing time intervals,

$$p(t_1+t_2,x,I)=\int_{\mathbb{R}} p(t_2,x,dz)p(t_1,z,I).$$

2^{nd} simplification
This is the particular case of a transition probability that admits density, such that $p(s,x;t,I)=\int_I \varphi(y)dy$, where $\varphi\geq 0$ and $\int_{\mathbb{R}}\varphi(y)dy=1$. Again, as is customary, we replace φ with p:

$$p(s,x;t,I)=\int_I p(s,x;t,y)dy,$$

with $p(s,x;t,y)\geq 0$ and $p(s,x;t,I)=\int_{\mathbb{R}} p(s,x;t,y)dy=1$. Under these new hypoteses,

[2]From now on, according to the notation adopted by other authors, the reducted forms of transition probability will be written with the same letters identifying variables of the original expression. However here (in the second member of 2.2.2) t is referred to the difference $t-s$ and p is φ.

Chapman - Kolmogorov relation becomes:

$$p(s, x; t, I) = \int_{\mathbb{R}} p(s, x, r, z) p(r, z, t, y).$$

3^{rd} **simplification**

Starting from the assumption of homogeneity in time, there is another reduction:

$$p(t, x, I) = \int_I p(t, x, y) dy.$$

A straightforward computation yields the new form of the Chapman - Kolmogorov relation

$$p(t_1 + t_2, x, y) = \int_{\mathbb{R}} p(t_1, x, z) p(t_2, z, y)$$

or the equivalent expression replacing t_1 with t_2.

4^{th} **simplification: Homogeneity in space**

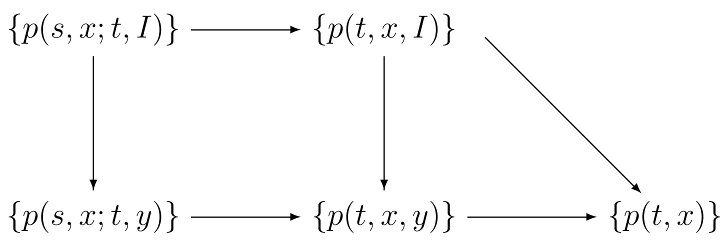
This simplification allows us to write the transition probability as a function defined on $\mathbb{R}^+ \times \mathbb{R}$:

$$p(t, x, y) = p(t, x)$$

where the new x actually represents the difference $y - x$. Proposition (iii) becomes

$$p(t_1 + t_2, x) = \int_{\mathbb{R}} p(t_1, y) p(t_2, x) dy = \int_{\mathbb{R}} p(t_2, y) p(t_1, x) dy$$

and $p(t_1 + t_2) = p(t_1) \star p(t_2)$ is just the convolution. All these simplifications are resumed in the following diagram.



2.3 Wiener process

In 1827 Robert Brown (botanist, 1773 - 1858) observed the erratic and continuous motion of plant spores suspended in water. Later in the 20's Norbert Wiener (mathematician, 1894 - 1964) proposed a mathematical model describing this motion, the *Wiener process* (also called the *Brownian motion*).

Definition 2.3.1. (of Wiener process). A Gaussian, continuous parameter process characterized by mean value $m(t) = 0$ and covariance function $\varphi := min\{s,t\} = s \wedge t$, for any $s, t \in [0, T]$, is called a *Wiener process* (denoted by $\{W_t\}$).

We will often talk about a Wiener process following the characterisation just stated, but it is appreciable to notice that next definition can provide a more intuitive description of the fundamental properties of a process of this kind.

Definition 2.3.2. A stochastic process $\{W_t\}_{t\geq 0}$ is called a Wiener process if it satisfies the properties listed below:

(i) $W_0 = 0$,

(ii) $\{W_t\}$ has, *almost surely*, continuous trajectory,[3]

(iii) $\forall\ s, t \in [0, T]$, with $0 \leq s < t$, the increments $W_t - W_s$ are *stationary*,[4] *independent* random variables and $W_t - W_s \sim \mathscr{N}(0, t-s)$.

Theorem 2.3.3. *Definition 2.3.1 and 2.3.2 are equivalent.*

Proof
(2.3.1) $\Rightarrow$ (2.3.2)
The independence of increments is a plain consequence of the Gaussian structure of W_t. As a matter of fact we can express W_t and $W_t - W_s$ as a linear combination of Gaussian random variables,

$$(W_s, W_t - W_s) = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} W_s \\ W_t \end{pmatrix}$$

moreover, they are not correlated [5]

$$\begin{aligned} a_{ij} = a_{ji} &= E[(W_{t_i} - W_{t_{i-1}})(W_{t_j} - W_{t_{j-1}})] \\ &= \ldots = t_i - t_{i-1} + t_{i-1} - t_i = 0, \end{aligned}$$

thus independent. Now we can write a Wiener process as:

$$W_t = (W_{\frac{t}{N}} - W_0) + (W_{\frac{2t}{N}} - W_{\frac{t}{N}}) + \ldots + (W_{\frac{tN}{N}} - W_{\frac{t(N-1)}{N}}),$$

then

$$W_{\frac{kt}{N}} - W_{\frac{(k-1)t}{N}} \sim N(0, \frac{t}{N}).$$

[3] This statement is equivalent to say that the mapping $t \longrightarrow W_t$ is continuous in t with probability 1.

[4] The term *stationary increments* means that the distribution of the *nth* increment is the same as for the increment $W_t - W_0$, $\forall\ 0 < s, t < \infty$.

[5] Observe that under the assumption $s < t$, the relations $\mathbb{E}(W_t) = m(t) = 0$, $\mathbf{Var}(W_t) = \min\{t,t\} = t$, $\mathbb{E}(W_t - W_s) = 0$ and $\mathbb{E}[(W_t - W_s)^2] = t - 2s + s = t - s$ are valid.

(2.3.2) $\Rightarrow$ (2.3.1)

Conversely, the 'shape' of the increments distribution implies that a Wiener process is a Gaussian process. Sure enough, for any $0 < t_1 < t_2 \ldots < t_n$, the probability distribution of a random vector $(W_{t_1}, W_{t_2}, \ldots, W_{t_n})$ is normal because is a linear combination of the vector $(W_{t_1}, W_{t_2} - W_{t_1}, \ldots, W_{t_n} - W_{t_{n-1}})$, whose components are Gaussian by definition. The mean value and the autocovariance function are:

$$\begin{aligned} \mathbb{E}(W_t) &= 0 \\ \mathbb{E}(W_t W_s) &= \mathbb{E}[W_S(W_t - W_s + W_s)] \\ &= \mathbb{E}[W_s(W_t - W_s)] + \mathbb{E}({W_s}^2) \\ &= s = s \wedge t \quad when \quad s \leq t. \end{aligned}$$

□

Remark 2.3.4. Referring to definition 2.3.2, notice that the *a priori* existence of a Wiener process is not guaranteed, since the mutual consistency of properties (i)-(iii) has first to be shown. In particular, it must be proved that (iii) does not contradict (ii). However, there are several ways to prove the existence of a Wiener process.[6] For example one can invoke the Kolmogorov theorem that will be stated later or, more simply, one can show the existence by construction.

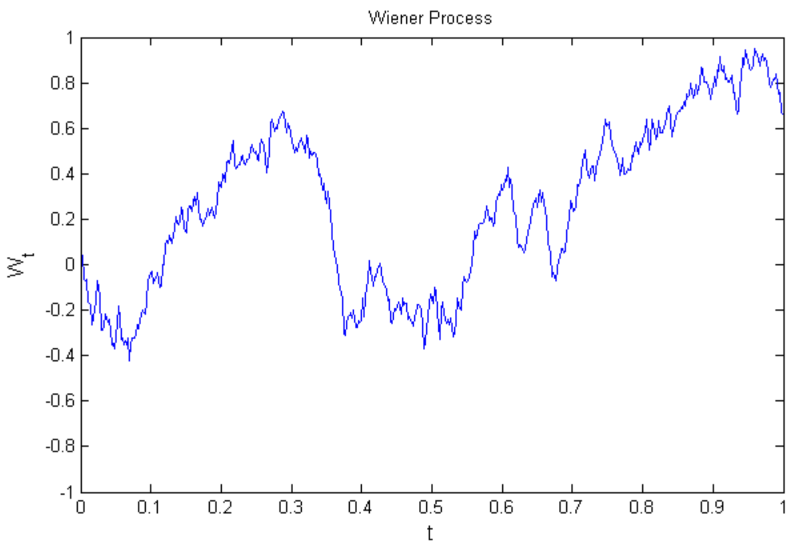


Figure 2.3.1: Simulation of the Wiener process.

[6]Wiener himself proved the existence of the process in 1923.

2.3.1 Distribution of the Wiener Process

Theorem 2.3.5. *The finite dimensional distributions of the Wiener process are given by*

$$\mu_{t_1,\dots,t_n}(B) = \int_B g_{t_1}(x_1) g_{t_2-t_1}(x_2 - x_1) \dots g_{t_n - t_{n-1}}(x_n - x_{n-1}) dx_1 \dots dx_n.$$

Proof Given a Wiener process W_t, so $m(t) = 0$ and $\phi(s,t) = \min(s,t)$, let us compute $\mu_{t_1,\dots,t_n}(B)$. We already know that

$$\mu_{t_1,\dots,t_n}(B) := \int_B \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sqrt{\det A}} e^{-\frac{1}{2}\langle A^{-1}x, x\rangle} dx,$$

where the covariance matrix is

$$A = \begin{pmatrix} t_1 & t_1 & \cdots & \cdots & t_1 \\ t_1 & t_2 & \dots & \cdots & t_2 \\ \vdots & t_2 & t_3 & \cdots & t_3 \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ t_1 & t_2 & \cdots & \cdots & t_n \end{pmatrix}$$

Performing some operations one gets:

$$\det(A) = t_1(t_2 - t_1) \cdots (t_n - t_{n-1}).$$

We can now compute the scalar product $\langle A^{-1}x, x\rangle$. Take $y := A^{-1}x$, then $x = Ay$ and $\langle A^{-1}x, x\rangle = \langle y, x\rangle$, that is

$$\begin{cases} x_1 = t_1(y_1 + \dots + y_n) \\ x_2 = t_1 y_1 + t_2(y_2 + \dots + y_n) \\ \vdots \\ x_n = t_1 y_1 + t_2 y_2 + \dots + t_n y_n \end{cases}$$

Manipulating the system one gains the following expression for the scalar product

$$\langle A^{-1}x, x\rangle = \frac{x_1^2}{t_1} + \frac{(x_2 - x_1)^2}{t_2 - t_1} + \dots + \frac{(x_n - x_{n-1})^2}{t_n - t_{n-1}},$$

then, introducing the function $g_t(\alpha) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{\alpha^2}{t}}$, we finally obtain

$$\mu_{t_1,\dots,t_n}(B) = \int_B g_{t_1}(x_1) g_{t_2-t_1}(x_2 - x_1) \cdots g_{t_n - t_{n-1}}(x_n - x_{n-1}) dx_1 \cdots dx_n.$$

□

2.3.2 Wiener processes are Markov processes

Consider a Wiener process ($m(t) = 0$ and $\phi(s,t) = \min(s,t)$).

Theorem 2.3.6. *A Wiener process satisfies the Markov condition*

$$\mathbb{P}(W_t \in I | W_s = x, W_{s_1} = x_1, \dots, W_{s_n} = x_{s_n}) = \mathbb{P}(W_t \in I | W_s = x). \tag{2.3.1}$$

Proof. Compute first the right-hand side of 2.3.1. By definition

$$\mathbb{P}(W_t \in I, W_s \in B) = \int_B \mathbb{P}(W_t \in I | W_s = x)\mathbb{P}(W_s \in dx), \tag{2.3.2}$$

otherwise

$$\begin{aligned} \mathbb{P}(W_t \in I, W_s \in B) &= \int_{B\times I} g_s(x_1) g_{t-s}(x_2 - x_1) dx_1 dx_2 \\ &= \int_B \left\{ \int_I g_{t-s}(x_2 - x_1) dx_2 \right\} g_s(x_1) dx_1. \end{aligned} \tag{2.3.3}$$

So, comparing 2.3.2 and 2.3.3 and noticing that

$$\mathbb{P}(W_s \in dx) = g_s(x)dx,$$

then

$$\mathbb{P}(W_t \in I | W_s = x) = \int_I g_{t-s}(y - x) dy \quad a.e. \ . \tag{2.3.4}$$

Compunting the left-hand side of 2.3.1, we actually obtain the same result of 2.3.4. Consider an $n+1$ - dim random variable Y and perform the computation for a Borel set $B = J_1 \times \dots J_n \times J$. As before

$$\begin{aligned} \int_B \mathbb{P}(W_t \in I | Y = y)\mathbb{P}(Y \in dy) &= \mathbb{P}(W_t \in I, Y \in B) \\ &= \mathbb{P}(W_t \in I, W_{s_1} \in J_1, \dots, W_{s_n} \in J_n, W_s \in J). \end{aligned}$$

Where $\mathbb{P}(W_t \in I, W_{s_1} \in J_1, \dots, W_{s_n} \in J_n, W_s \in J)$ can be expressed as

$$\int_{B\times I} g_{s_1}(y_1) g_{s_2 - s_1}(y_2 - y_1) \cdots g_{s - s_n}(y_{n+1} - y_s) g_{t-s}(y - y_{n+1}) dy_1 \dots dy_{n+1} dy.$$

Since $\mathbb{P}(Y \in dy_1 \cdots dy_{n+1}) = g_{s_1}(y_1) \cdots g_{s-s_n}(y_{n+1} - y_n) dy_1 \cdots dy_{n+1}$

$$\int_B \mathbb{P}(W_t \in I | Y = y) \mathbb{P}(Y \in dy) = \int_B \left\{ \int_I g_{t-s}(y - y_{n+1}) dy \right\} \mathbb{P}(Y \in dy),$$

that implies

$$\mathbb{P}(W_t \in I | Y = y) = \int_I g_{t-s}(y - y_{n+1}) dy \qquad a.e.$$

and, since in our case $y_{n+1} = x$, the theorem is proved. □

2.3.3 Brownian bridge

As it has already been said, one can think of a stochastic process as a model built from a collection of finite dimensional measures. For example, suppose to have a family $\mathscr{M}$ of Gaussian measures. We know that the 1 - dim measure $\mu_t(I)$ is given by the following integral:

$$\mu_t(I) = \int_I \frac{1}{\sqrt{2\pi\sigma^2(t)}} e^{-\frac{[x-m(t)]^2}{2\sigma^2(t)}} dx. \tag{2.3.5}$$

Now, for every choice of $m(t)$ and $\sigma^2(t)$ we obtain an arbitrary 1 - dim measure.[7] We can also construct the 2 - dim measure, but in that case, instead of $\sigma^2(t)$, we will have a 2×2 covariance matrix that must be symmetric and positive definite for the integral to be computed. In addition, the 2 - dim measure obtained in this way must satisfy the compatibility property (definition 2.1.4).

Example 2.3.7. Consider $t \in [0, +\infty]$ and a Borel set I. Try to construct a process under the hypotheses $m \equiv 0$ and $a(s,t) = s \wedge t - st$ ($a_{11} = t_1 - {t_1}^2, a_{12} = t_1 - t_1 t_2, \ldots$). It can be proved that the covariance matrix $A = a_{ij}$ is symmetric and definite positive, thus equation (2.3.5) makes sense.

In conclusion we have now the idea of how to build a process given a family of finite dimensional measures $\mathscr{M}$. On the other hand, it is interesting and prolific to make over the previous procedure asking ourselves if a process corresponding with a given $\mathscr{M}$ does exists . We will introduce chapter 3 keeping this question in mind.

Example 2.3.7 leads us to the following definition.

Definition 2.3.8. (of Brownian bridge). A Gaussian, continuous parameter process $\{X_t\}_{t \in [0,1]}$ with continuous trajectories, mean value $m(t) = 0$, covariant function $a(s,t) = s \wedge t - st$ and conditioned to have $W(1) = 0$, $\forall\ 0 \le s \le t \le T$, is called a *Brownian bridge* process.

[7]For example, for suitable $m(t)$ and $a(s,t)$ we obtain just the Wiener process.

There is an easy way to construct a Brownian bridge from a Wiener process:

$$X(t) = W(t) - tW(1) \quad \forall\, 0 \le t \le 1.$$

Observe also that given a Wiener process, we do not have to worry about the existence of the Brownian bridge. It is interesting to check that the following process represents a Brownian bridge too:

$$Y(t) = (1-t)\, W\Big(\frac{t}{1-t}\Big) \quad \forall\, 0 \le t \le 1, \quad Y(1) = 0.$$

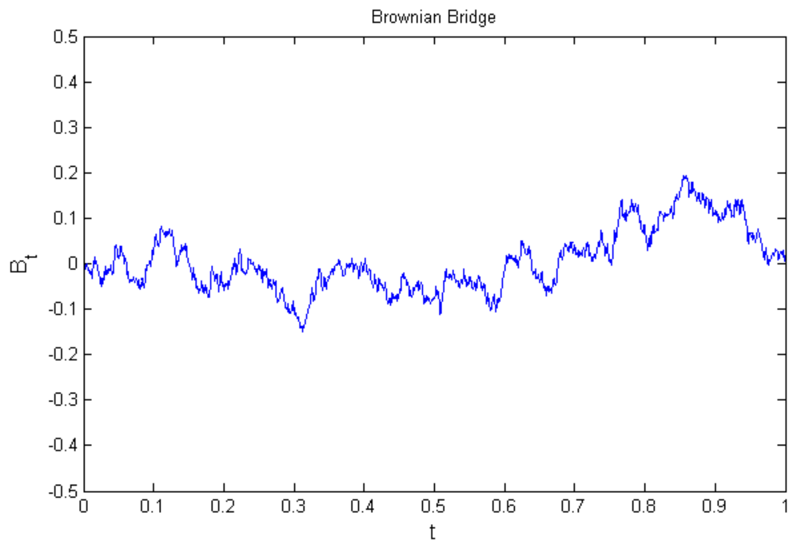


Figure 2.3.2: Simulation of the Brownian bridge.

2.3.4 Exercises

Example 2.3.9. Consider a Wiener process $\{W_t\}$, the vector space $L^2_{loc}(\mathbb{R}^+)$ and a function $f\colon \mathbb{R}^+ \to \mathbb{R}$ locally of bounded variation (BV from now on) and locally in L^2. For any trajectory define

$$X = \int_0^T f(t)dW_t \quad , \quad X(\omega) = \int_0^T f(t)dW_t(\omega). \tag{2.3.6}$$

Observe that we are integrating over time, hence we can formally remove the ω that appears in the integrand as a parameter. Integration by parts yields

$$X(\omega) = \int_0^T f(t)dW_t = f(t)W_t - f(0)W_0 - \int_0^T W_t df(t).$$

You are now wondering how X is distributed. To this purpose fix a T, choose any partition $\pi = \{0, t_1, t_2, \ldots, t_n\}$ of $[0, T]$ and search for the distribution of the sum

$$\sum_{i=1}^{\infty} f(t_i)(W_{t_{i+1}} - W_{t_i}) =: X_\pi, \tag{2.3.7}$$

whose limit is just the first equation in 2.3.6. Look now at the properties of X_π. First of all notice that it is a linear combination of Gaussian processes. Then, because of the properties of a Wiener process,[8] the covariance function of X_π (sum of the covariance functions) is $\sum_{i=0}^{n-1} f(t_i)^2(t_{i+1} - t_i)$, thus

$$X_\pi \sim \left(0, \sum_{i=0}^{n-1} f(t_i)^2(t_{i+1} - t_i)\right) \tag{2.3.8}$$

and, for $n \to \infty$,

$$\sum_{i=0}^{n-1} f(t_i)^2(t_{i+1} - t_i) \longrightarrow \int_0^T f(t)^2 dt = \|f\|^2_{L^2_{(0,T)}}.$$

By construction we have, for $\pi \to 0$, that $X_\pi(\omega \to X(\omega)$. The same result can be achieved independently by calling into account the characteristic function of X_π:

$$\varphi(\lambda)_{X_\pi} = \mathbb{E}(e^{i\lambda X_\pi}) = e^{-\frac{1}{2}\lambda^2 \sum_{i=0}^{n-1} f(t_i)^2(t_{i+1}-t_i)} \xrightarrow{\pi \to 0} e^{-\frac{1}{2}\int_0^T f(t)^2 d_t}.$$

This is a sequence of function $e^{i\lambda X_\pi(\omega)}$ that converges pointwise to $e^{i\lambda X(\omega)}$. Observing that $|e^{i\lambda X_\pi}| = 1$ and calling to action the *Lebesgue dominated convergence theorem*, as $\pi \to 0$ one has

$$\varphi(\lambda)_{X_\pi} \longrightarrow \varphi(\lambda)_X = \mathbb{E}(e^{i\lambda X}) = e^{-\frac{1}{2}\lambda^2 \int_0^T f(t)^2 dt}$$

The integral $X = \int_0^T f(t)dW_t$ has been computed; for a more general expression notice that if its upper bound is variable, one gets the process

$$X_t := \int_0^t f(r)dW_r. \tag{2.3.9}$$

Example 2.3.10. Given a Wiener process, a Borel set I and the rectangle $G = (x - \delta, x + \delta) \times I$, compute the marginal $\mu_{s,t}(G) \in \mathscr{M}$.

[8]Independent increments distributed as $W_t - W_s \sim \mathscr{N}(0, t - s)$.

Due to the features of the Wiener process one has

$$\mu_{s,t}(G) = \mathbb{P}(W_s, W_t \in G) = \int_G g(x)dx = \int_G \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{detA}} e^{-\frac{1}{2}\langle A^{-1}(x-0),(x-0)\rangle} dx \quad (2.3.10)$$

where

$$A = \begin{pmatrix} s & s \\ s & t \end{pmatrix} \quad \text{and} \quad A^{-1} = \frac{1}{detA}\begin{pmatrix} t & -s \\ -s & s \end{pmatrix}.$$

Set $z = (x - 0)$, then

$$(z_1, z_2)\begin{pmatrix} t & -s \\ -s & s \end{pmatrix}\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \ldots = tz_1^2 - 2sz_1z_2 + sz_2^2,$$

so

$$\int_G \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{detA}} e^{-\frac{1}{2}\langle A^{-1}z,z\rangle} dz = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{st-s^2}} \int_G e^{-\frac{tz_1^2-2sz_1z_2+sz_2^2}{2(st-s^2)}} dz_1dz_2$$
$$= \ldots = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{st-s^2}} \int_G e^{-\frac{z_1^2}{2s}} e^{-\frac{(z_2-z_1)^2}{t-s}} dz_1dz_2.$$

At this point, bringing to mind the *transition function* $g_t(\xi) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{\xi^2}{2t}}$ we have in this case:

$$\mathbb{P}\big((W_s, W_t) \in G\big) = \int_{G=I\times(x-\delta)(x+\delta)} g_s(z_1) \cdot g_{t-s}(z_2 - z_1)dz_1dz_2. \quad (2.3.11)$$

Example 2.3.11. Let $\{W_t\}$ be a Wiener process, I a Borel set and $s, t \in [0, T]$ with $s < t$. Compute the probability $\mathbb{P}(W_t \in I|W_s = x)$.

By definition of conditional probability we know that

$$\mathbb{P}(W_t \in I|W_s = x) = \frac{\mathbb{P}(W_t \in I, W_s = x)}{\mathbb{P}(W_s = x)}. \quad (2.3.12)$$

When the denominator of the right hand-side is zero (as in this case), the numerator is zero too[9] and formula 2.3.12 is not useful (0/0). Try then to compute the probability

$$\mathbb{P}(W_t \in I|W_s \in (x-\delta, x+\delta)) = \frac{\mathbb{P}(W_t \in I, W_s \in (x-\delta, x+\delta))}{\mathbb{P}(W_s \in (x-\delta, x+\delta)}, \quad (2.3.13)$$

where $\mathbb{P}(W_s \in (x-\delta, x+\delta) > 0$.

[9]Consider two events A, B and the expression $\frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$. If $\mathbb{P}(B) = 0$, since $\mathbb{P}(A \cap B) \subset P(B)$ also the numerator would be zero.

Using the result of exercise 2.3.10 we have

$$\frac{\mathbb{P}(W_t \in I, W_s \in (x-\delta, x+\delta))}{\mathbb{P}(W_s \in (x-\delta, x+\delta)} = \frac{\displaystyle\int_I \left(\int_{(x-\delta)(x+\delta)} g_s(z_1) g_{t-s}(z_2 - z_1) dz_1 \right) dz_2}{\displaystyle\int_{\mathbb{R}} \left(\int_{(x-\delta)(x+\delta)} g_s(z_1) g_{t-s}(z_2 - z_1) dz_1 \right) dz_2}$$
$$= \frac{\displaystyle\int_{x-\delta}^{x+\delta} g_s(z_1) \left(\int_I g_{t-s}(z_2 - z_1) dz_2 \right) dz_1}{\displaystyle\int_{x-\delta}^{x+\delta} g_s(z_1) \left(\int_{\mathbb{R}} g_{t-s}(z_2 - z_1) dz_2 \right) dz_1}$$
$$= \frac{\frac{1}{2\delta} \displaystyle\int_{x-\delta}^{x+\delta} g_s(z_1) \left(\int_I g_{t-s}(z_2 - z_1) dz_2 \right) dz_1}{\frac{1}{2\delta} \displaystyle\int_{x-\delta}^{x+\delta} g_s(z_1) dz_1}.$$

Latest expression has been achieved by observing that $\int_{\mathbb{R}} g_{t-s} dz_2 = 1$ because W_s is Gaussian (with known variance) and then dividing numerator and denominator by $\frac{1}{2\delta}$. If we computed the limit, we would obtain 0/0 again, but after applying De L'Hospital's rule we have

$$\cdots \longrightarrow \frac{g_s(x) \int_I g_{t-s}(z_2 - x) dz_2}{g_s(x)} = \int_I g_{t-s}(z_2 - x) dz_2$$

and, finally, we get the transition probability

$$\mathbb{P}(W_t \in I | W_s = x) = \mathbb{P}(W_t - W_s \in I - x). \tag{2.3.14}$$

2.4 Poisson process

A random variable $T_\Omega \to (0, \infty)$ has *exponential distribution of parameter* $\lambda > 0$ if, for any $t \geq 0$,

$$\mathbb{P}(T > t) = e^{-\lambda t}.$$

In that case T has a density function

$$f_T(t) = \lambda e^{-\lambda t} \mathbf{1}_{(0,\infty)}(t)$$

and the mean value and the variance of T are given by

$$\mathbb{E}(T) = \frac{1}{\lambda}, \quad \mathbf{Var}(T) = \frac{1}{\lambda^2}.$$

Proposition 2.4.1. *A random variable $T : \Omega \to (0, \infty)$ has an exponential distribution* *if and only if it has the following memoryless property*

$$\mathbb{P}(T > s + t | T > s) = \mathbb{P}(T > t)$$

for any $s, t \geq 0$.

Definition 2.4.2. (of Poisson process.) A stochastic process $\{N_t, t \geq 0\}$ defined on a probability space $(\Omega, \mathscr{E}, \mathbb{P})$ is said to be a Poisson process of rate λ if it verifies the following properties

(i) $N_t = 0$,

(ii) for any $n \geq 1$ and for any $0 \leq t_1 \leq \ldots \leq t_n$ the increments $N_{t_n} - N_{t_{n-1}} \ldots N_{t_2} - N_{t_1}$, are independent random variables,

(iii) for any $0 \leq s < t$, the increment $N_t - N_s$ has a Poisson distribution with parameter $\lambda(t - s)$, that is,

$$\mathbb{P}(N_t - N_s = k) = e^{-\lambda(t-s)} \frac{[\lambda(t-s)]^k}{k!},$$

where $k = 0, 1, 2, \ldots$ and $\lambda > 0$ is a fixed constant.

Definition 2.4.3. (of compound Poisson process.) Let $N = \{N_t, t \geq 0\}$ be a Poisson process with rate λ. Consider a sequence $\{Y_n, n \geq 1\}$ of independent and identically distributed random variables, which are also independent of the process N. Set $S_n = Y_1 + \ldots Y_n$. Then the process

$$Z_t = Y_1 + \ldots + Y_{N_t} = S_{N_t},$$

with $Z_t = 0$ if $N_t = 0$, is called a compound Poisson process.

Proposition 2.4.4. *The compound Poisson process has independent increments and the law of an increment* $Z_t - Z_s$ *has characteristic funcion*

$$e^{\left(\varphi_{Y_1}(x) - 1\right)\lambda(t-s)},$$

where $\varphi_{Y_1}(x)$ *denotes the characteristic function of* Y_1.

Chapter 3

Kolmogorov's theorems

3.1 Cylinder set

Consider a probability space $(\Omega, \mathscr{E}, \mathbb{P})$ and a family of random variables $\{X_t\}_{t\in[0,T]}$. Every $\omega \in \Omega$ corresponds with a trajectory $t \longrightarrow X_t(\omega),\ \forall\ t \in [0,T]$. We want to characterize the mapping X defined from Ω in some space of functions, where X is a function of t: $X(\omega)(t) := X_t(\omega),\ \forall\ \omega \in \Omega$. If we have no information about trajectories, we can assume some hypotheses in order to define the codomain of the mapping X. Suppose, for example, to deal only with constant or continuous trajectories. This means to immerse the set of trajectories into the larger sets of constant or continuous functions, respectively. In general, if also the regularity of the trajectories is unknown, we can simply consider the very wide space $\mathbb{R}^{[0,T]}$ of real function, then $X : \Omega \longrightarrow \mathbb{R}^{[0,T]}$. We want now to find the domain of the mapping X^{-1}, that is, a σ - algebra that makes X measurable with respect to the σ - algebra itself. To this end, let us introduce the *cylinder set*

$$C_{t_1,t_2,\ldots,t_n,B} = \{f \in \mathbb{R}^{[0,T]} : (f(t_1), f(t_2), \ldots, f(t_n)) \in B\},$$

where $B \in \mathscr{B}^d$ (B is called the *base* of the cylinder) and $C \subseteq \mathbb{R}^{[0,T]}$. Then consider the set $\mathscr{C}$ of all the cylinders C:

$$\mathscr{C} = \{C_{t_1,t_2,\ldots t_n,B},\ \forall\ n \geq 1,\ \forall\ (t, t_1, t_2, \ldots, t_n),\ \forall\ B \in \mathscr{B}^n\}.$$

Actually, $\mathscr{C}$ is an algebra but not σ - algebra. Anyway we know that there exists the smallest σ - algebra containing $\mathscr{C}$. We are going to denote it by $\sigma(\mathscr{C})$. Now we can show that $X^{-1} : \sigma(\mathscr{C}) \longrightarrow \mathscr{E}$.

Theorem 3.1.1. *The preimage of $\sigma(\mathscr{C})$ lies in $\mathscr{E}$.*

PROOF Consider the cylinder set $C_{t_1,t_2,\ldots,t_n,B} \in \mathscr{C} \subseteq \sigma(\mathscr{C})$.

$$\begin{aligned} X^{-1}(C_{t_1,t_2,\ldots,t_n,B}) &= \{\omega \in \Omega : X(\omega) \in C_{t_1,t_2,\ldots,t_n}\} \\ &= \{\omega \in \Omega : (X(\omega)(t_1), X(\omega)(t_2), \ldots, X(\omega)(t_n)) \in B\} \\ &= \{X_{t_1}(\omega), X_{t_2}(\omega), \ldots, X_{t_n}(\omega) \in B\} \in \mathscr{E}. \end{aligned} \tag{3.1.1}$$

Now we consider the sigma algebra $\{X^{-1}(E) \in \mathscr{E}\} = \mathscr{F}$ for any $E \subset \mathbb{R}^{[0,T]}$ and, since $\mathscr{C} \subset \mathscr{F}$, the theorem is proved. □

To fulfill the relations between the abstract probability space and the objects which we actually deal with, notice that the map X is measurable as a function $(\Omega, \mathscr{E}) \to (\mathbb{R}^{[0,T], \sigma(\mathscr{C})})$, so we can consider the image measure of $\mathscr{P}$ defined as usual:

$$\mu(E) = \mathbb{P}(X^{-1}(E)), \quad \forall\ E \in \sigma(\mathscr{C}).$$

If we now consider two preocesses X, X' (also defined on different probability spaces) with the same family $\mathscr{M}$, that is

$$\mathbb{P}((X_{t_1}, X_{t_2}, \ldots, X_{t_n})) \in E) = \mathbb{P}((X'_{t_1}, X'_{t_2}, \ldots, X'_{t_n})) \in E),$$

then also $\mu_X = \mu_{X'}$. In other words, μ_X is determined by $\mathscr{M}$ only.

3.2 Kolmogorov's theorem I

Theorem 3.2.1. (of Kolmogorov, I). *Given a family of probability measures* $\mathscr{M} = \{\{\mu_t\}, \{\mu_{t_1,t_2}\}, \ldots\}$ *with the compatibility property 2.1.4, there exists a unique measure* μ *on* $(\mathbb{R}^{[0,T]}, \sigma(\mathscr{C}))$ *such that* $\mu(C_{t_1,t_2,\ldots,t_n,B}) = \mu_{t_1,t_2,\ldots,t_n}(B)$ *and we can define, on the space* $(\mathbb{R}^{[0,T]}, \sigma(\mathscr{C}), \mu)$*, the stochastic process* $X_t(\omega) = \omega(t)$ *(with* $\omega \in \mathbb{R}^{[0,T]}$*) which has the family* $\mathscr{M}$ *as finite dimensional distributions.*

Although this theorem assures the existence of a process, it is not a constructive theorem, so it does not provide enough informations in order to find a process given a family of probability measures. For our purposes it is more useful to take into account another theorem, that we will call *second theorem of Kolmogorov.*[1] From now on, if not otherwise specified, we will always recall the second one.

[1]Note that some authors call it *Kolmogorov's criterion* or *Kolmogorov continuity criterion.*

3.3 Kolmogorov's theorem II

Before stating the theorem, let us recall some notions.

Definition 3.3.1. (of equivalent processes). A stochastic process $\{X_t\}_{t\in[0,T]}$ is equivalent to another stochastic process $\{Y_t\}_{t\in[0,T]}$ if

$$\mathbb{P}(X_t \neq Y_t) = 0 \quad \forall\, t \in [0,T].$$

It is also said that $\{X_t\}_{t\in[0,T]}$ is a *version* of $\{Y_t\}_{t\in[0,T]}$. However, equivalent processes may have different trajectories.

Definition 3.3.2. (of continuity in probability). A real - valued stochastic process $\{X_t\}_{t\in[0,T]}$, where $[0,T] \subset \mathbb{R}$, is said to be continuous in probability if, for any $\epsilon > 0$ and every $t \in [0,T]$

$$\lim_{s\to t} \mathbb{P}(|X_t - X_s| > \epsilon) = 0. \tag{3.3.1}$$

Definition 3.3.3. Given a sequence of events $A_1, A_2, \ldots$, one defines $lim\ supA_k$ as the event $lim\ supA_k = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$.

Lemma 3.3.4. (of Borel Cantelli). *Let $A_1, A_2, \ldots$ be a sequence of events such that*

$$\sum_{k=1}^{\infty} \mathbb{P}(A_k) < \infty,$$

then the event $A = lim\ supA_K$ has null probability $\mathbb{P}(A) = 0$. Of course $\mathbb{P}(A^c) = 1$, where $A^c = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c$.

Proof We have $A \subset \bigcup k = n^{\infty} A_k$. Then

$$\mathbb{P}(A) \leq \sum_{k=n}^{\infty} \mathbb{P}(A_k)$$

which, by hypothesis, tends to zero when n tends to infinity. □

Definition 3.3.5. (of Hölder continuity).[2] A function $f : \mathbb{R}^n \longrightarrow \mathbb{K}$, $(\mathbb{K} = \mathbb{R} \vee \mathbb{C})$ satisfies the Hölder condition, or it is said to be Hölder continuous, if there exist non - negative real constants C and α such that

$$|f(x) - f(y)| \leq C|x - y|^{\alpha}$$

[2]More generally, this condition can be formulated for functions between any two metric space.

for all x, y in the domain of f.[3]

Definition 3.3.6. (of dyadic rational). A *dyadic rational* (or dyadic number) is a rational number $\frac{a}{2^b}$ whose denominator is a power of 2, with a an integer and b a natural number.

Ahead we will deal with the *dyadic set* $D = \bigcup_{m=0}^{\infty} D_m$, where $D_m := \{\frac{k}{2^m} : k = 0, 1, \dots, 2^m\}$.

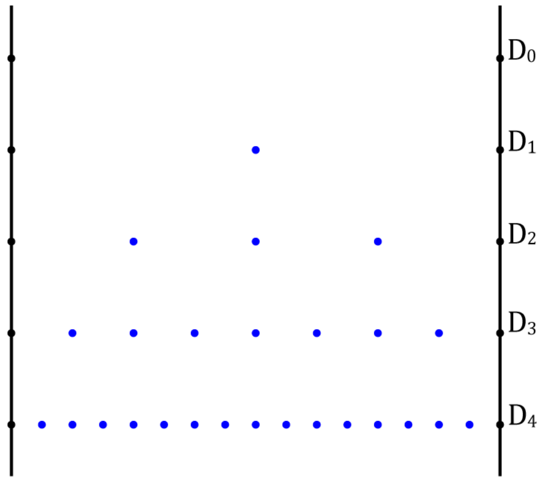


Figure 3.3.1: Construction of dyadic rationals.

Theorem 3.3.7. (of Kolmogorov, II). *Let* $(\Omega, \mathscr{E}, \mathbb{P})$ *be a probability space and consider the stochastic process* $\{X_t\}_{t\in[0,T]}$ *with the property*

$$\mathbb{E}|X_t - X_s|^\alpha \leq c|t-s|^{1+\beta} \quad \forall\ s, t \in [0, T], \tag{3.3.2}$$

where c, α, β *are non - negative real constants. Then there exists another stochastic process* $\{X'_t\}_{t\in[0,T]}$ *over the same space such that*

(i) $\{X'_t\}_{t\in[0,T]}$ *is locally Hölder continuous with exponent* γ, $\forall\ \gamma \in (0, \frac{\beta}{\alpha})$,[4]

(ii) $\mathbb{P}(X_t \neq X'_t) = 0$ *forall* $t \in [0, T]$.

PROOF Chebyshev inequality implies

$$\mathbb{P}(|X_t - X_s| > \varepsilon) \leq \frac{1}{\varepsilon^\alpha}\mathbb{E}(|X_t - X_s|^\alpha) \leq \frac{c}{\varepsilon^\alpha}|t-s|^{1+\beta}. \tag{3.3.3}$$

[3]If $\alpha = 1$ one gets just the *Lipschitz continuity*, if $\alpha = 0$ then the function in simply bounded.
[4]It's not necessary that $\{X_t\}_{t\in[0,T]}$ has Hölder continuous trajectories too.

Using the notion of dyadic rationals (definition 3.3.6), search for a useful valuation of 3.3.3 studying the following inequality:

$$\mathbb{P}(|X_{t_{i+1}} - X_{t_i}| > \varepsilon_m) \leq \frac{c}{\varepsilon_m^\alpha} \frac{1}{2^{m(1+\beta)}}, \qquad t_i \in D_m.$$

Let us consider the event $A = (\max_i |X_{t_{i+1}} - X_{t_i}| > \varepsilon_m)$. It's easy to prove that

$$A = \bigcup_{i=0}^{2^m-1} A_i \;\; where \;\; A_i = (|X_{t_{i+1}} - X_{t_i}| > \varepsilon_m).$$

So, by the properties of probability measure and by the obtained valuations we get:

$$\mathbb{P}(\max_i |X_{t_{i+1}} - X_{t_i}| > \varepsilon_m) \leq 2^m \frac{c}{\varepsilon_m^\alpha} \frac{1}{2^{m(1+\beta)}} = \frac{c}{\varepsilon_m^\alpha 2^{m\beta}}.$$

Set now $A_m = (\max_i |X_{t_{i+1}} - X_{t_i}| > \varepsilon_m)$ and $\varepsilon_m = \frac{1}{2^{\gamma m}}$. These choices yield

$$\sum_{m=0}^{\infty} \mathbb{P}(A_m) \leq c \sum_{m=0}^{\infty} \frac{2^{m\gamma\alpha}}{2^{m\beta}} = c \sum_{m=0}^{\infty} \frac{1}{2^{m(\beta-\gamma\alpha)}} \tag{3.3.4}$$

and, choosing $\gamma < \frac{\beta}{\alpha}$, the sum 3.3.4 converges.

Consider now the set $A = \limsup A_m$. By the Borel-Cantelli lemma we have $\mathbb{P}(A) = 0$. Take an element ω such that $\omega \notin A$, then $\omega \in \bigcup_{m=1}^{\infty} \bigcap_{k=m}^{\infty} A_k^c$. It follows that $\exists\, m^\star$ such that $\omega \in \bigcap_{k=m^\star}^{\infty} A_k^c$. Hence $\omega \in A_k^c \quad \forall k \geq m^\star$. It results

$$\max_i |X_{t_{i+1}}(\omega) - X_{t_i}(\omega)| \leq \frac{1}{2^{m\gamma}} \qquad \forall m \geq m^\star.$$

Resuming, we have constructed a null measure set A such that

$$\omega \notin A \Rightarrow \exists m^\star \;\; \text{with} \;\; \max_i |X_{t_{i+1}}(\omega) - X_{t_i}(\omega)| \leq \frac{1}{2^{m\gamma}} \quad \forall m \geq m^\star$$

and we want to prove that, for $s, t \in D$, the following valuation holds:

$$|X_t - X_s| \leq c|t-s|^\gamma.$$

Thus, suppose that $|t - s| < \frac{1}{2^n} = \delta$. This implies $t, s \in D_m, m > n$. Now try to prove the following inequality

$$|X_t - X_s| \leq 2(\frac{1}{2^{(n+1)\gamma}} + \frac{1}{2^{(n+2)\gamma}} + \cdots + \frac{1}{2^{m\gamma}}) \leq (\frac{1}{2^\gamma})^{n+1} \frac{2}{1-\frac{1}{2^\gamma}}. \quad (3.3.5)$$

We can prove it by induction. To this end let us suppose that inequality 3.3.5 holds true for numbers in D_{m-1}. Take $t, s \in D_m$ and $s^\star, t^\star \in D_{m-1}$ such that $s \leq s^\star, t^\star \leq t$. Hence

$$\begin{aligned} |X_t - X_s| &\leq |X_t - X_{t^\star}| + |X_{t^\star} - X_{s^\star}| + |X_{s^\star} - X_s| \\ &\leq \frac{1}{2^{m\gamma}} + \frac{1}{2^{m\gamma}} + 2(\frac{1}{2^{(n+1)\gamma}} + \cdots + \frac{1}{2^{(m-1)\gamma}}) = 2(\frac{1}{2^{(n+1)\gamma}} + \cdots + \frac{1}{2^{m\gamma}}). \end{aligned}$$

Observe now that inequality 3.3.5 holds for the first step $(n+1)$ because $|X_t - X_s| \leq \frac{1}{2^{(n+1)\gamma}} \leq \frac{2}{2^{(n+1)\gamma}}$. Thus 3.3.5 is proved.

Set now $\delta = \frac{1}{2^{m^\star}}$, $c = 2(\frac{1}{1-\frac{1}{2^\gamma}})$ and consider two instants $t, s \in D$. Then $\exists n$ such that $\frac{1}{2^{n+1}} < |t-s| < \frac{1}{2^n}$. Finally, consider $|t-s| < \delta$. Then

$$|X_t(\omega) - X_s(\omega)| \leq c \left(\frac{1}{2^{n+1}}\right)^\gamma \leq c|t-s|^\gamma.$$

We have obtained the valuation we were looking for, but we have obtained it only for dyadic numbers and we have no information about other instants.
Define a new process $X_t^{'}(\omega)$ as follow:

(i) $X_t'(\omega) = X_t(\omega)$, se $t \in D$,

(ii) $X_t'(\omega) = \lim_n X_{t_n}(\omega)$, if $t \notin D$, where $t_n \to t, t_n \in D$,

(iii) $X_t'(\omega) = z$, where $z \in \mathbb{R}$, when $\omega \in A$.

This is surely a process with continuous trajectories. All we have to prove now is that $X_t'(\omega)$ is a version of $X_t(\omega)$, that is $\mathbb{P}(X_t = X_t') = 1$.
This is true by definition if $t \in D$. If $t \notin D$, consider a sequence $t_n \to t$ such that $X_{t_n} \to X_t$ almost everywhere. This implies that $X_{t_n} \to X_t$ in probability.
Finally, we have to show that the two processes have the same family of finite dimensional measures. Indeed, considering the set $\Omega_t = \{\omega \ : \ X_t(\omega) = X_t'(\omega)\}$, we get, obviously, $\mathbb{P}(\Omega_t) = 1$ and $\mathbb{P}(\bigcup_{t_i} \Omega_{t_i}^c) \leq \sum_{t_i} \mathbb{P}(\Omega_{t_i}^c) = 0$.
Hence

$$\begin{aligned} \mathbb{P}((X_{t_1}, \ldots, X_{t_n}) \in B) \mathbb{P}((X_{t_1}, \ldots, X_{t_n}) \in B, \bigcap_{t_i} \Omega_{t_i}) \\ = \mathbb{P}((X_{t_1}', \ldots, X_{t_n}') \in B, \bigcap_{t_i} \Omega_{t_i}) \\ = \mathbb{P}((X_{t_1}', \ldots, X_{t_n}') \in B). \end{aligned}$$

□

Chapter 4

Martingales

We introduce first the notion of *conditional expectation* of a random variable with respect to a σ - algebra.

4.1 Conditional expectation

Lemma 4.1.1. *Given a random variable $X \in L^1(\Omega)$ there exists a unique $\mathscr{F}$ - measurable map Z such that for every $F \subseteq \mathscr{F}$,*

$$\int_F X(\omega)\mathbb{P}(d\omega) = \int_F Y(\omega)\mathbb{P}(d\omega).$$

Definition 4.1.2. (of conditional expectation). Let X be a random variable in $L^1(\Omega)$ and Z an $\mathscr{F}$ - measurable random variable such that, for every $F \in \mathscr{F}$,

$$\int_F X(\omega)\mathbb{P}(d\omega) = \int_F Z(\omega)\mathbb{P}(d\omega), \quad \text{for any } F \in \mathscr{F}$$

or, similarly,

$$\mathbb{E}(Z\mathbf{1}_A) = \mathbb{E}(X\mathbf{1}_A). \tag{4.1.1}$$

We will denote this function with $\mathbb{E}(X \,|\, \mathscr{F})$. Its existence is guaranteed by Random-Nikodym theorem (1.3.1) and it is unique up to null - measure sets.

Moreover equation 4.1.1 implies that for any bounded $\mathscr{F}$ - measurable random variable Y we have

$$\mathbb{E}(\mathbb{E}(X|\mathscr{F})Y) = \mathbb{E}(XY).$$

Properties of conditional expectation

(i) If X is a integrable and $\mathscr{F}$ - measurable random variable, then

$$\mathbb{E}(X|\mathscr{F}) = X \quad a.e.\ . \tag{4.1.2}$$

(ii) If X_1, X_2 are two integrable and $\mathscr{F}$ - measurable random variables, then

$$\mathbb{E}(\alpha X_1 + \beta X_2|\mathscr{F}) = \alpha\mathbb{E}(X_1|\mathscr{F}) + \beta\mathbb{E}(X_2|\mathscr{F}) \quad a.e.\ . \tag{4.1.3}$$

(iii) A random variable and its conditional expectation have the same expectation

$$\mathbb{E}(\mathbb{E}(X|\mathscr{F})) = \mathbb{E}(X) \quad a.e.\ . \tag{4.1.4}$$

(iv) Given $X \in L^1(\Omega)$, consider the σ-algebra generated by X: $\mathscr{F}_X = \{X^{-1}(I), I \in \mathscr{B}\} \subset \mathscr{E}$. If $\mathscr{F}$ and $\mathscr{F}_X$ are independent,[1] then

$$\mathbb{E}(X|\mathscr{F}) = \mathbb{E}(X) \quad a.e.\ . \tag{4.1.5}$$

In fact, the constant $\mathbb{E}(X)$ is clearly $\mathscr{F}$ - measurable, and for all $A \in \mathscr{F}$ we have

$$\mathbb{E}(X\mathbf{1}_A) = \mathbb{E}(X)\mathbb{E}(\mathbf{1}_A) = \mathbb{E}(\mathbb{E}(X)\mathbf{1}_A). \tag{4.1.6}$$

(v) Given X, Y with X a general random variable and Y is a bounded and $\mathscr{F}$ - measurable random variable, then

$$\mathbb{E}(XY|\mathscr{F}) = Y\mathbb{E}(X|\mathscr{F}) \quad a.e.\ . \tag{4.1.7}$$

In fact, the random variable $Y\mathbb{E}(X|\mathscr{F})$ is integrable and $\mathscr{F}$ - measurable, and for alla $A \in \mathscr{F}$

$$\mathbb{E}(\mathbb{E}(X|\mathscr{F})Y\mathbf{1}_A) = \mathbb{E}(\mathbb{E}(XY\mathbf{1}_A|\mathscr{B})) = \mathbb{E}(XY\mathbf{1}_A). \tag{4.1.8}$$

(vi) Given the random variable $\mathbb{E}(X|\mathscr{G})$ with $X \in L^1(\Omega)$ and $\mathscr{F} \subset \mathscr{G} \subset \mathscr{E}$, then

$$\mathbb{E}(\mathbb{E}(X|\mathscr{G})|\mathscr{F}) = \mathbb{E}(\mathbb{E}(X|\mathscr{F})|\mathscr{G})\mathbb{E}(X|\mathscr{F}) \quad a.e. \tag{4.1.9}$$

[1] For any $F \in \mathscr{F}$ and $G \in \mathscr{G}$, $\mathscr{F}$ and $\mathscr{G}$ are said to be independent if F and G are independent: $\mathbb{P}(F \cap G) = \mathbb{P}(F) \cdot \mathbb{P}(G)$.

4.2 Filtrations and Martingales

Definition 4.2.1. (of filtration.) Given a probability space $(\Omega, \mathscr{E}, \mathscr{P})^2$, a filtration is defined as a family $\{\mathscr{F}_t\} \subset \mathscr{E}$ of σ - algebras increasing with time, that is, if $t < t'$, then $\mathscr{F}_t \subset \mathscr{F}_{t'}$.

Definition 4.2.2. (of martingale.) Given a stochastic process $\{X_t, t \geq 0\}$ and a filtration $\{\mathscr{F}_t\}_{t\geq 0}$ we say that $\{X_t, t \geq 0\}$ is a martingale (resp. a supermartingale, a submartingale) with respect to the given filtration if:

(i) $X_t \in L^1(\Omega)$ for any $t \leq T$

(ii) X_t is $\mathscr{F}_t$ - measurable (we said that the process is adapted to the filtration), that is $\mathscr{F}_t^X \subset \mathscr{F}_t$, where $\mathscr{F}_t^X = \sigma\{X_u \text{ s.t. } u \leq t\}$,

(iii) $\mathbb{E}(X_t|\mathscr{F}_s) = X_s$ (resp. $\leq$, $\geq$) a.e. .

4.3 Examples

Example 4.3.1. Consider the smallest σ - algebra of $\mathscr{E}$: $\mathscr{F}_0 = \{\varnothing, \Omega\} \subset \mathscr{E}$. To compute $Z := \mathbb{E}(X|\mathscr{F})$, notice that by definition we have

$$\int_F X d\mathbb{P} = \int_F (X|\mathscr{F}) d\mathbb{P} \quad \forall\ F \in \mathscr{F}. \tag{4.3.1}$$

So, since $\mathbb{E}(X|\mathscr{F})$ is constant,

$$\int_F X d\mathbb{P} = \int_F (X|\mathscr{F}_0) d\mathbb{P} = const\ \mathbb{P}(F). \tag{4.3.2}$$

Moreover, due to the inclusion $\mathscr{F}_0 \subset \mathscr{F}_s$, we also obtain the following relations.

$$\mathbb{E}(\mathbb{E}(X_t|\mathscr{F}_s)) = \mathbb{E}(\mathbb{E}(X_t|\mathscr{F}_s)|\mathscr{F}_0) = \mathbb{E}(X_t|\mathscr{F}_0) = \mathbb{E}(X_t) = \mathbb{E}(X_s) = \mathbb{E}(X_0)$$

Consider now the first order σ - algebra $\mathscr{F}_1 = \{\varnothing, A, A^c, \Omega\}$ for some subset $A \subset \Omega$. What is the expectation value $\mathbb{E}(X|\mathscr{F}_1)$?
We have $0 < \mathbb{P}(A), \mathbb{P}(A^c) < 1$ and

$$\mathbb{E}(X|\mathscr{F}_1) = \begin{cases} c_1 & \omega \in A \\ c_2 & \omega \in A^c \end{cases}$$

[2] Note that $\mathscr{P}$ is not necessary in this definition.

Applying formula 4.3.1

$$\mathbb{E}(X|\mathscr{F}_1) = \begin{cases} \int_A c_1 d\mathbb{P} = c_1\mathbb{P}(A) & \Rightarrow\ c_1 = \frac{1}{\mathbb{P}(A)}\int_A X d\mathbb{P} \qquad \omega \in A \\ \int_{A^c} c_2 d\mathbb{P} = c_2\mathbb{P}(A^c) & \Rightarrow\ c_2 = \frac{1}{\mathbb{P}(A^c)}\int_{A^c} X d\mathbb{P} \qquad \omega \in A^c \end{cases}$$

We can extend this result to any partition of Ω and, for every subset A_i, c_i will be given by:

$$c_i = \frac{1}{\mathbb{P}(A_i)}\int_{A_i} X d\mathbb{P}.$$

Example 4.3.2. Consider a Wiener process adapted to the filtration $\mathscr{F}_t$, thus $\mathscr{F}_t^W \subset \mathscr{F}_t$ and $\mathscr{F}_t$ and $W_{t+h} - W_h$ are independent. These hypoteses yields to the equalities

$$\begin{aligned}\mathbb{E}(W_t|\mathscr{F}_s) = \mathbb{E}(W_t - W_s + W_s|\mathscr{F}_s) &= \mathbb{E}(W_t - W_s|\mathscr{F}_s) + \mathbb{E}(W_s|\mathscr{F}_s) \\ &= \mathbb{E}(W_t - W_s) + W_s = 0 + W_s = W_s.\end{aligned}$$

These reasoning can be applied to a process W_t^n in the following way:

$$\begin{aligned}\mathbb{E}(W_t^n|\mathscr{F}_s) = \mathbb{E}\left((W_t - W_s + W_s)^n|\mathscr{F}_s\right) &= \mathbb{E}\left(\sum_{k=0}^{n}\binom{n}{k}(W_t - W_s)^k W_s^{n-k}|\mathscr{F}_s\right) \\ &= \sum_{k=0}^{n}\binom{n}{k}\mathbb{E}\left((W_t - W_s)^k W_s^{n-k}|\mathscr{F}_s\right) \\ &= \sum_{k=0}^{n}\binom{n}{k}W_s^{n-k}\mathbb{E}\left((W_t - W_s)^k|\mathscr{F}_s\right) \\ &= \sum_{k=0}^{n}\binom{n}{k}W_s^{n-k}\mathbb{E}((W_t - W_s)^k).\end{aligned}$$

And for odd k the general term vanishes. So, setting $b = [\frac{n}{2}]$,

$$\sum_{k=0}^{b}\binom{n}{2k}W_s^{n-2k}\mathbb{E}((W_t - W_s)^{2k}) = \sum_{k=0}^{b}\binom{n}{2k}\frac{(2k)!}{2^k k!}W_s^{n-2k}(t-s)^k.$$

For example, for $n = 2$, one has

$$\mathbb{E}(W_t^2|\mathscr{F}_s) = W_s^2 - s + t.$$

Thus W_t^2 is not a martingale, but since t is a costant, we can easily recognize that the process $W_t^2 - t$ is actually a martingale:

$$\mathbb{E}(W_t^2 - t|\mathscr{F}_s^W) = W_s - s.$$

We can try to perform the same computation for an exponential function and look for a solution of $\mathbb{E}(e^{\vartheta W_t}|\mathscr{F}_s)$, with $\vartheta \in \mathbb{R}$. In order to compute this conditional expectation value it is useful to write the exponential function in terms of its Taylor expansion ($e^{\vartheta W_t} = \sum_{n=0}^{\infty} \frac{\vartheta^n W_t^n}{n!}$):

$$\begin{aligned}
\mathbb{E}(e^{\vartheta W_t}|\mathscr{F}_s) &= \sum_{n=0}^{\infty} \frac{\vartheta^n}{n!} \sum_{k=0}^{b} \binom{n}{2k} \frac{(2k)!}{2^k k!} W_s^{n-2k} (t-s)^k \\
&= \sum_{n=0}^{\infty} \vartheta^n \sum_{k=0}^{b} \frac{1}{(n-2k)! 2^k k!} W_s^{n-2k} (t-s)^k \\
&= \sum_{n=0}^{\infty} \sum_{k=0}^{b} \frac{1}{(n-2k)! k!} (\vartheta W_s)^{n-2k} [\frac{\vartheta^2}{2}(t-s)]^k \\
&= \sum_{k=0}^{\infty} \frac{[\frac{\vartheta^2}{2}(t-s)]^k}{k!} \sum_{n=2k}^{\infty} \frac{(\vartheta W_s)^{n-2k}}{(n-2k)!} \\
&= \sum_{k=0}^{\infty} \frac{[\frac{\vartheta^2}{2}(t-s)]^k}{k!} e^{\vartheta W_s} = e^{\frac{\vartheta^2}{2}(t-s)+\vartheta W_s}.
\end{aligned}$$

Then the considered function is a martingale because $\mathbb{E}(e^{\vartheta W_t}|\mathscr{F}_s) = e^{\frac{\vartheta^2}{2}(t-s)+\vartheta W_s}$, that is $\mathbb{E}(e^{\vartheta W_t - \frac{\vartheta^2 t}{2}}|\mathscr{F}_s) = e^{\vartheta W_s - \frac{\vartheta^2 s}{2}}$.

Chapter 5

Monotone classes

Talking about Markov processes we introduced the relation:

$$\mathbb{P}(X_t \in I | X_s = x, X_{s_1} = x_1, \ldots, X_{s_n} = x_n) = \mathbb{P}(X_t \in I | X_s = x) \tag{5.0.1}$$

and we know how it can be computed. However, equation 5.0.1 can be proved in the light of the tools we are about to define.
In general, if one wants to prove a certain property of a σ - algebra, it may be helpful to prove it first for a particular subset of the σ - algebra itself (that is a π system) and then extend the result to the whole field.

Definition 5.0.1. (of π system). A family $\mathscr{P}$, with $\mathscr{P} \in \mathscr{E}$, where $\mathscr{E}$ is the σ - algebra of elements of the sample space, is called a π - system if:

$$A, B \in \mathscr{P} \Rightarrow A \cap B \in \mathscr{P}. \tag{5.0.2}$$

Suppose, for example, to have two measures μ_1, μ_2. If we want to prove that $\mu_1 = \mu_2$ in some σ - algebra $\mathscr{F}$, that is $\mu_1(E) = \mu_2(E) \ \ \forall \ E \in \mathscr{F}$, we can start proving it for a particular subset $E' \in \mathscr{P} \subset \mathscr{F}$. This example shows not just the usefulness of π - systems, but also the necessity of a new tool able to put all the subsets with the same property togheter. These are the λ - systems.

Definition 5.0.2. (of λ system). A family $\mathscr{L}$, with $\mathscr{L} \in \mathscr{E}$, where $\mathscr{E}$ is the σ - algebra of elements of the sample space, is called a λ - system if:

(i) if $A_1 \subset A_2 \subset A_3 \ldots$ and $A_i \in \mathscr{L}$, then $\bigcup A_i \in \mathscr{L}$,

(ii) if $A, B \in \mathscr{L}$ and $A \subset B$, then $B \setminus A \in \mathscr{L}$.

Proposition 5.0.3. *If $\mathscr{L}$ is a λ - system such that $\Omega \in \mathscr{L}$ and $\mathscr{L}$ is also a π - system, then $\mathscr{L}$ is a σ - algebra.*

Proof   By hypotesis $\Omega \in \mathscr{L}$; by definition of λ - system $A \in \mathscr{L} \Rightarrow A^c \in \mathscr{L}$; finally, due to the hypotesis that $\mathscr{L}$ is also a π - system, we get $\{A_i\} \Rightarrow \bigcup A_i \in \mathscr{L}$ for any sequence $\{A_i\}$. □

A foundamental relation between π - systems and λ - systems is given by the Dynkin lemma. Before stating and proving it, let us introduce some observations.

Lemma 5.0.4. *Given a π - system $\mathscr{P}$ and a λ - system $\mathscr{L}$ such that $\Omega \in \mathscr{L}$ and $\mathscr{P} \subset \mathscr{L}$, define:*

$$\mathscr{L}_1 = \{E \in \mathscr{L} : \forall C \in \mathscr{P}, \quad E \cap C \in \mathscr{L}\},$$
$$\mathscr{L}_2 = \{E \in \mathscr{L}_1 : \forall A \in \mathscr{L}_1, \quad E \cap A \in \mathscr{L}\}.$$

Then $\mathscr{L}_1$ and $\mathscr{L}_2$ are π - systems and $\mathscr{P} \subset \mathscr{L}_2 \subset \mathscr{L}_1 \subset \mathscr{L}$.

Proof   Consider $\mathscr{L}_1$ first. We want to prove that $\mathscr{L}_1$ is a λ - system and that $\mathscr{P} \subset \mathscr{L}_1 \subset \mathscr{L}$, then we will extend the same kind of reasoning to $\mathscr{L}_2$. Observe first that $\mathscr{L}_1$ is non-empty, for example $\Omega \in \mathscr{L}_1$, $\mathscr{P} \in \mathscr{L}_1$.
We know that if $A_1 \subset A_2 \subset A_3 \dots$ with $A_k \in \mathscr{L}$, then $\bigcup_k A_k \in \mathscr{L}$ surely, because $\mathscr{L}$ is a λ - system; but we want to prove it for elements of $\mathscr{L}_1$, then take $(C \cap \bigcup_k A_k)$. Does it belong to $\mathscr{L}$? Notice that we can write $(C \cap \bigcup_k A_k)$ as $\bigcup_k (C \cap A_k)$. Now define $\tilde{A}_k := (C \cap A_k)$ which belongs to $\mathscr{L}$ by hypotesis. So, taking $\tilde{A}_1 \subset \tilde{A}_2 \subset \dots$ we obtain $\bigcup_k (\tilde{A}_k) \in \mathscr{L}$, that is the first property of a λ - system.
To verify the other property that makes of $\mathscr{L}_1$ a λ - system we need to show that $B \setminus A$ belongs to $\mathscr{L}_1$ when A is a subset of B and $A, B \in \mathscr{L}_1$. This is equivalent to verify if $C \cap (B \setminus A)$ belongs to $\mathscr{L}$ for any $C \in \mathscr{P}$. But $C \cap (B \setminus A) = (C \cap B) \setminus (C \cap A)$, and $C \cap A$, as $C \cap B$, belongs to $\mathscr{L}$. So we conclude that $\mathscr{L}_1$ (and for the same facts $\mathscr{L}_2$) is a λ - system, so $\mathscr{P} \subset \mathscr{L}_2 \subset \mathscr{L}_1 \subset \mathscr{L}$. □

5.1 The Dynkin lemma

We can now state and prove the following lemma.

Lemma 5.1.1. (of Dynkin).[1] *Given a π - system $\mathscr{P}$ and a λ - system $\mathscr{L}$ such that*

[1]This is a probabilistic version of *monotone class method* in measure theory (product spaces). See Rudin Chp. 8.

$\Omega \in \mathscr{L}$ and $\mathscr{P} \subset \mathscr{L}$, then

$$\mathscr{P} \subset \sigma(\mathscr{P}) \subset \mathscr{L}, \tag{5.1.1}$$

where $\sigma(\mathscr{P})$ is the σ - algebra generated by $\mathscr{P}$.

PROOF Consider the family of λ - systems $\mathscr{L}^{(\alpha)}$ such that $\mathscr{P} \subset \mathscr{L}^{(\alpha)}$ and $\Omega \in \mathscr{L}^{(\alpha)}$. Set $\Lambda := \bigcap_\alpha \mathscr{L}^{(\alpha)}$. Λ is a λ - system by lemma 5.0.4.
Since $\Lambda \subset \mathscr{L}$, Λ has the same properties of $\mathscr{L}$, so $\mathscr{P} \subset \Lambda$ and $\Omega \in \Lambda$. We can construct the following hierarchy of inclusions:

$$\mathscr{P} \subset \ldots \subset \Lambda_2,\ \Lambda_1,\ \Lambda \subset \mathscr{L},$$

but $\Lambda_1, \Lambda_2, \ldots$ are all the same because $\Lambda = \bigcap_\alpha \mathscr{L}^{(\alpha)}$ is the smallest λ - system containing Ω. In other words $\mathscr{P} \subset \Lambda \equiv \Lambda_1 \equiv \Lambda_2 \equiv \ldots$.
Moreover, it is easy to prove that Λ is a π - system and, by proposition 5, we get that Λ is a σ - algebra. This fact concludes the proof because[2]

$$\mathscr{P} \subset \sigma(\mathscr{P}) \subset \Lambda \subset \mathscr{L}$$

□

5.2 Some applications of the Dynkin lemma

Example 5.2.1. Consider a π - system $\mathscr{P}$ and suppose $\mathscr{E} = \sigma(P)$, $\Omega \in \mathscr{P}$ and $\mathscr{P}_1(E) = \mathscr{P}_2(E)$ ($\mathscr{P}_1$ and $\mathscr{P}_2$ are probability measures). Using the Dynkin lemma, show that $\mathscr{P}_1 \equiv \mathscr{P}_2$ on $\mathscr{E}$.

PROOF $\mathscr{P}$ is a π - system by hypotesis. Take $\mathscr{L} = \{F \in \mathscr{E} : \mathscr{P}_1(F) = \mathscr{P}_2(F)\}$. Since $\mathscr{P} \subset \mathscr{L}$, $\mathscr{L}$ is non - empty. Considering $F_1 \subset F_2 \subset \ldots$ with $F_k \in \mathscr{L}$ and remembering the additivity property of probability measures, it is easy to show that $\mathscr{L}$ is also a λ - system then, for the Dynkin lemma,

$$\mathscr{P} \subset \sigma(\mathscr{P}) \subset \mathscr{L}$$

and, consequently, the thesis. □

Example 5.2.2. Let us consider two definitions of Markov property.

[2]Notice that $\sigma(\mathscr{P})$ obviously belongs to Λ.

(i) $\mathbb{P}(X_t \in I | X_s = x, X_{s_1} = x_1, \dots, X_{s_n} = x_n) = \mathbb{P}(X_t \in I | X_s = x)$ *a.e.*,

(ii) $\mathbb{P}(X_t \in I | \mathscr{F}_s) = \mathbb{P}(X_t \in I | X_s = x)$ *a.e.*.

We want to show that these equations are equivalent.[3] Take the left-hand side of (i). This is a number depending on t, I, s, x, so we can write the first equation as

$$p(t, I; s, x, s_1, x_1, \dots, s_n, x_n) = p(t, I; s, x) \quad a.e. \,. \tag{5.2.1}$$

Now take the σ - algebras $\mathscr{F}_{X_s}, \mathscr{F}_{X_{s_1}}, \dots, \mathscr{F}_{X_{s_n}}$ generated by $X_s, X_{s_1}, \dots$ and set

$$\mathscr{F}_{X_s, X_{s_1}, \dots, X_{s_n}} := \sigma(\mathscr{F}_{X_s}, \mathscr{F}_{X_{s_1}}, \dots, \mathscr{F}_{X_{s_n}}).$$

We are going to prove that $(i) \Rightarrow (ii)$ by means of 5.2.1. Now observe that the left-hand side of (i) is just

$$\mathbb{P}(X_t \in I | \mathscr{F}_{X_s, X_{s_1}, \dots, X_{s_n}}) \tag{5.2.2}$$

and, as far as (ii) is concerned, we can write:

$$\mathbb{P}(X_t \in I | \mathscr{F}_s) = \mathbb{E}(\mathbf{1}_I(X_t) | \mathscr{F}_s) \tag{5.2.3}$$

where

$$\int_F \mathbf{1}_I(X_t) d\mathbb{P} = \int_F \mathbb{E}(\mathbf{1}_I(X_t) | \mathscr{F}_s) d\mathbb{P} \quad \forall\ F \in \mathscr{F}_s \tag{5.2.4}$$

and

$$\mathbb{P}(X_t \in I | X_s) = \mathbb{E}(\mathbf{1}_I(X_t) | \mathscr{F}_{X_s}), \tag{5.2.5}$$

where

$$\int_{F'} \mathbf{1}_I(X_t) d\mathbb{P} = \int_{F'} \mathbb{E}(\mathbf{1}_I(X_t) | \mathscr{F}_{X_s}) d\mathbb{P} \quad \forall\ F' \in \mathscr{F}_s. \tag{5.2.6}$$

PROOF $(i) \Rightarrow (ii)$

Notice, first of all, that $s_1, s_2, \dots, s_n$ represent instants before s, then $\mathscr{F}_{X_s, X_{s_1}, \dots, X_{s_n}} \subset \mathscr{F}_s$. Moreover, remember that $\mathbb{E}(\mathbb{E}(X|\mathscr{G})|\mathscr{F}) = \mathbb{E}(X|\mathscr{F})$. Now, recalling formulas 5.2.2-5.2.6, one gets

$$\begin{aligned}
\mathbb{E}(\mathbf{1}_I(X_t) | \mathscr{F}_{X_s, X_{s_1}, \dots, X_{s_n}}) &= \mathbb{E}(\mathbb{E}(\mathbf{1}_I(X_t) | \mathscr{F}_s) | \mathscr{F}_{X_s, X_{s_1}, \dots, X_{s_n}}) \\
&= \mathbb{E}(\mathbb{E}(\mathbf{1}_I(X_t) | \mathscr{F}_{X_s}) | \mathscr{F}_{X_s, X_{s_1}, \dots, X_{s_n}}) \\
&= \mathbb{E}(\mathbf{1}_I(X_t) | \mathscr{F}_{X_s}),
\end{aligned}$$

where the latter equality is justified by the inclusion $\mathscr{F}_{X_s} \subset \mathscr{F}_{X_s, X_{s_1}, \dots, X_{s_n}} \subset \mathscr{F}_s$. □

[3] To prove that $(ii) \Rightarrow (i)$ we will need the Dynkin lemma.

Proof $(ii) \Rightarrow (i)$[4]
Consider $F = (X_{s_1}, X_{s_2}, \dots, X_{s_n}, X_s)^{-1}(B)$ and denote with $\mathscr{G}$ the following field

$$\mathscr{G} := \{F_B\} = \mathscr{F}_{X_{s_1}, X_{s_2}, \dots, X_{s_n}, X_s}. \tag{5.2.7}$$

By definition we have $\mathbb{P}(X_t \in I|\mathscr{G}) = \mathbb{E}(\mathbf{1}_I(X_t)|\mathscr{G})$ and, by definition of expectation value and from hypotesis (ii) the following equalities hold

$$\int_{F_B} \mathbf{1}_I(X_t) d\mathbb{P} = \int_{F_B} \mathbb{E}(\mathbf{1}_I(X_t)|\mathscr{G}) d\mathbb{P} = \int_{F_B} \mathbb{E}(\mathbf{1}_I(X_t)|X_s) d\mathbb{P}. \tag{5.2.8}$$

Equation 5.2.8 can be written as

$$\mu_1(F_B) = \mu_2(F_B), \tag{5.2.9}$$

where

$$\mu_1(F_B) = \int_{F_B} \mathbf{1}_I(X_t) d\mathbb{P},$$
$$\mu_2(F_B) = \int_{F_B} \mathbb{E}(\mathbf{1}_I(X_t)|X_s) d\mathbb{P}.$$

For the first memeber in (i), we can write

$$\mu_1(F) = \int_F \mathbf{1}_I(X_t) d\mathbb{P} = \int_F \mathbb{P}(X_t \in I|\mathscr{F}_s) d\mathbb{P} \tag{5.2.10}$$

for any $F \in \mathscr{F}_s$, where $\mathscr{F}_s = \sigma(\mathscr{F}_{X_r}, 0 \le r \le s)$. So our thesis will be proved if we prove the following equality:

$$\mu_1(F) = \int_F \mathbb{P}(X_t \in I|X_s) d\mathbb{P} = \mu_2(F) \quad \forall\ F \in \mathscr{F}_s. \tag{5.2.11}$$

To this end, observe that if 5.2.11 is true for a π - system, its validity can be extended to the whole σ - algebra. Consider $\mathscr{P} = \bigcup \mathscr{G}$ and check if it is a π - system, that is, prove that

$$A, B \in \mathscr{P} \Rightarrow A \cap B \in \mathscr{P}. \tag{5.2.12}$$

[4] Application of the Dynkin lemma.

Remember that $\mathscr{G}$ is defined as $\mathscr{G} = \mathscr{F}_s = \{X^{-1}_{s_1,s_2,\dots,s}(I), I \in B\}$. Take

$$A = (X_{s_1}, X_{s_2}, \dots, X_{s_n}, X_s)^{-1}(I),$$
$$B = (X_{r_1}, X_{r_2}, \dots, X_{r_m}, X_s)^{-1}(J),$$

with $I \in \mathbb{B}^{n+1}$, $J \in \mathbb{B}^{m+1}$, $s_1 < s_2 < \dots < s_n < s$ and $r_1 < r_2 < \dots < r_m < r$. To find a common set to A and B consider the union $\{s_1 < s_2 < \dots < s_n < s\} \bigcup \{r_1 < r_2 < \dots < r_m < r\}$ and call t_i the indexes s_i, r_i. The new set is $\{t_1, t_2, \dots, t_N, t_s\}$.[5] Then A and B are given by:

$$A = (X_{t_1}, X_{t_2}, \dots, X_{t_N}, X_s)^{-1}(I \times \mathbb{R}^{N-n}),$$
$$b = (X_{t_1}, X_{t_2}, \dots, X_{t_N}, X_s)^{-1}(\mathbb{R}^{N-m} \times J).$$

Where with $\mathbb{R}^{N-n}$ and $\mathbb{R}^{N-m}$ we mean that the elements of A and B which do not belong to I or J, will belong to some Borel set in $\mathbb{R}^{N-n}$ or $\mathbb{R}^{N-m}$, respectively. Thus $A \cap B \in \mathscr{G}$, then $\mathscr{P} = \bigcup \mathscr{G}$ is a π - system, so we can apply the Dinkin lemma and finally $(ii) \Rightarrow (i)$.

□

[5]The i - th index can represent s_i, r_i or both. Anyway the N new indexes will be sorted by the order relation $<$.

Chapter 6

Some stochastic differential equation

6.1 Brownian bridge and stochastic differential equation

Consider the following stochastic differential equation

$$dX_t = \frac{b - X_t}{T - t}\, dt + dW_t, \qquad X_0 = a$$

where W_t is a stantard Wiener process. Observe that

$$d\Big(\frac{X_t - b}{T - t}\Big) = \frac{dX_t}{T - t} + \frac{X_t - b}{(T - t)^2} dt = \frac{dW_t}{T - t},$$

that yields

$$\frac{X_t - b}{T - t} - \frac{X_s - b}{T - s} = \int_s^t \frac{dW_r}{T - r}$$

and

$$X_t = \frac{T - t}{T - s}\, X_s + \frac{t - s}{T - s}\, b + \int_s^t \frac{T - t}{T - r}\, dW_r.$$

From latest equation is easy to verify that $X_T = b$.
Transition probabilities are given by

$$p(s, x; t, I) := \mathbb{P}(X_t \in I \,|\, X_s = x),$$

they are Gaussian with mean value and variance equal to

$$\frac{(T - t)x + (t - s)b}{T - s}, \qquad \int_s^t \Big(\frac{T - t}{T - r}\Big)^2 dr = \frac{T - t}{T - s}(t - s),$$

that is

$$p(s,x;t,I) = \mathscr{N}\Big(\frac{(T-t)x+(t-s)b}{T-s}, \frac{T-t}{T-s}(t-s)\Big)(I).$$

We can also say, using the notation

$$g_t(z) = \tfrac{1}{\sqrt{2\pi t}} e^{-\frac{z^2}{t}},$$

that $p(s,x;t,I)$ has the following probability density

$$g_{\frac{T-t}{T-s}(t-s)}\big(y - \tfrac{(T-t)x+(t-s)b}{T-s}\big).$$

In addition, notice that some calculation yields the incoming Lemma.

Lemma 6.1.1. *The following equality holds*

$$g_{\frac{T-t}{T-s}(t-s)}\big(y - \tfrac{(T-t)x+(t-s)b}{T-s}\big) = g_{t-s}(y-x)\frac{g_{T-t}(b-y)}{g_{T-s}(b-x)}. \tag{6.1.1}$$

Theorem 6.1.2. *Brownian Bridge can be defined as a stochastic differential equation:*

$$\dot{X}_t = -\frac{X_t}{1-t} + \dot{W}_t.$$

Proof. Multipling both members by $\frac{1}{1-t}$ we obtain

$$\Big(\frac{X_t}{1-t}\Big)' = \frac{\dot{W}_t}{1-t}$$

and, integrating,

$$X_t = (1-t)X_0 + \int_0^t \frac{1-t}{1-s} dW_s$$

Setting $X_0 = 0$ we get

$$X_t = \int_0^t \frac{1-t}{1-s} dW_s \sim \mathscr{N}(0, t(1-t))$$

that is just the Brownian Bridge.

Now, integrating between s and t, we obtain the following expression

$$X_t = \frac{1-t}{1-s} X_s + \int_s^t \frac{1-t}{1-r} dW_r$$

and, since the integrand is independent from X_s (due to the independence of the increments in the Wiener process), we get the already known relation:

$$\mathbb{P}(X_t \in I | X_s = x) \sim \mathscr{N}\left(\frac{1-t}{1-s}x, \frac{1-t}{1-s}(t-s)\right)$$

□

6.2 Ornstein - Uhlenbeck equation

Theorem 6.2.1. *Ornstein - Uhlenbeck process is governed by equation*

$$\dot{X}_t = -\lambda X_t + \sigma \dot{W}_t.$$

PROOF Multiply both members by $e^{\lambda t}$:

$$(e^{\lambda t} X_t)' = \sigma e^{\lambda t} \dot{W}_t.$$

Integrating we obtain

$$X_t = e^{-\lambda t} X_0 + \sigma \int_0^t e^{-\lambda(t-s)} dW_s.$$

From Stieltjes integration theory we know that

$$\sigma \int_0^t e^{-\lambda(t-s)} dW_s \sim \mathscr{N}\left(0, \sigma^2 \int_0^t e^{-2\lambda(t-s)} ds = \frac{1-e^{-2\lambda t}}{2\lambda}\right),$$

then, assuming that X_0 is independent from the integrand and that $X_0 \sim (m, v^2)$, we get

$$X_t \sim \mathscr{N}\left(e^{-\lambda t} m, e^{-2\lambda t}\left(v^2 - \frac{\sigma^2}{2\lambda}\right)\right)$$

Setting $X_0 \sim \left(0, \frac{\sigma^2}{2\lambda}\right)$, we obtain the same distribution for X_t, a uniform distribution. We could get the same solution solving

$$X_t = X_0 - \lambda \int_0^t X_s ds + \sigma W_t.$$

□

Chapter 7

Integration

Consider the integral

$$X_t = \int_0^t f(r) dW_r, \tag{7.0.1}$$

where $f \in L^2_{loc}$ and f is a funcion of bounded variation. Integral 7.0.1 is the limit of the sum

$$\sum_{i=0}^{n-1} f(r_i)(W_{r_{i+1}} - W_{r_i}),$$

where $0 = r_0 < r_1 < \ldots < r_n = t$. For any ω, the sum converges to something that depends on ω, that is just X_t by definition:

$$\sum_{i=0}^{n-1} f(r_i)(W_{r_{i+1}}(\omega) - W_{r_i}(\omega)) \xrightarrow{\delta \to 0} X_t$$

with $\delta = max_i |r_{i+1} - r_i|$. Indeed, integrating by parts, one has

$$\sum_{i=0}^{n-1} f(r_i)(W_{r_{i+1}}(\omega) - W_{r_i}(\omega)) = f(t)W_t(\omega) - \sum_{i=0}^{n-1} W_{r+1}(\omega)(f_{r_{i+1}} - f_{r_i}),$$

where $f(t)W_t(\omega)$ is fixed and the last sum converges because of the hypoteses on f. In the next subsection we will study an example even when f is not of BV.

Suppose now we want to compute

$$\mathbb{E}(X_t | \mathscr{F}_s^W), \tag{7.0.2}$$

where X_t is defined as in 7.0.1. We know that if we considered W_t (that is a martingale) we would obtain

$$\mathbb{E}(W_t | \mathscr{F}_s^W) = \mathbb{E}(W_t - W_s + W_s | \mathscr{F}_s^W) = \mathbb{E}(W_t - W_s | \mathscr{F}_s^W) + \mathbb{E}(W_s | \mathscr{F}_s^W) = W_s,$$

where $\mathbb{E}(W_t - W_s|\mathscr{F}_s^W) = 0$ because the increments are independent. We are asking if this result holds for 7.0.2 too. Actually

$$\mathbb{E}(X_t|\mathscr{F}_s^W) = \mathbb{E}(X_t - X_s|\mathscr{F}_s^W) + X_s,$$

but

$$X_t - X_s = \int_s^t f(r)dW_r = \lim_{\delta\to 0}\sum_{i=0}^{n-1} f(r_i)(W_{r_{i+1}} - W_{r_i}),$$

where $W_{r_{i+1}}$ and W_{r_i} are independent on $\mathscr{F}_s^W$. Then it results again

$$\mathbb{E}(X_t|\mathscr{F}_s^W) = X_s.$$

7.1 The Itô integral, a particular case

Take into account the integral

$$\int_0^t W_s dW_s, \tag{7.1.1}$$

where neither W_s nor dW_s are of BV. Consider the sums

$$\begin{aligned} &(i) \qquad S_\pi = \sum_{i=0}^{n-1} W_{t_{i+1}}(W_{t_{i+1}} - W_{t_i}), \\ &(ii) \qquad s_\pi = \sum_{i=0}^{n-1} W_{t_i}(W_{t_{i+1}} - W_{t_i}), \end{aligned} \tag{7.1.2}$$

where $0 = t_0 < t_1 < \ldots < t_n = t$ and π identifies a partition. The integral makes sense if these sums converge to the same number but, as we will see, their limits are different. Take the sum of (i) and (ii) in 7.1.2

$$S_\pi + s_\pi = \sum_{i=0}^{n-1}(W_{t_{i+1}} + W_{t_i})(W_{t_{i+1}} - W_{t_i}) = \sum_{i=0}^{n-1}(W_{t_{i+1}}^2 - W_{t_i}^2) = W_t^2,$$

so this sum is constant (notice that $\sum_{i=0}^{n-1}(W_{t_{i+1}}^2 - W_{t_i}^2)$ is a telescopic series). What about the difference? Define

$$S_\pi - s_\pi = \sum_{i=0}^{n-1}(W_{t_{i+1}} - W_{t_i})^2 =: Y_\pi,$$

then

$$S_\pi = \sum_{i=0}^{n-1} W_{t_{i+1}}(W_{t_{i+1}} - W_{t_i}) = \frac{W_t^2 + Y_\pi}{2},$$
$$s_\pi = \sum_{i=0}^{n-1} W_{t_i}(W_{t_{i+1}} - W_{t_i}) = \frac{W_t^2 - Y_\pi}{2}.$$

Before continue, let us state the next proposition.

Proposition 7.1.1. *The sum* $Y_\pi = \sum_{i=0}^{n-1}(W_{t_{i+1}} - W_{t_i})^2$ *converges to* t *in* $L^2(\Omega)$.

So, for $\pi \to \infty$ ($\delta = max_i|t_{i+1} - t_i| \to 0$), one has

$$S_\pi \xrightarrow{L^2} \frac{W_t^2 + t}{2},$$
$$s_\pi \xrightarrow{L^2} \frac{W_t^2 - t}{2}.$$

This result makes the Riemann - Stieltjes integration 7.1.1 impossible. Now try to take a different point, for example the middle point $\tau_i = \frac{t_i + t_{i+1}}{2}$ between t_i and t_{i+1}. With this choice, the corresponding sequence becomes

$$\sum_{i=0}^{n-1} W_{\tau_i}(W_{\tau_{i+1}} - W_{\tau_i}) \xrightarrow{\quad L^2 \quad} \frac{W_t^2}{2}$$

and

$$\int_0^t W_s dW_s = W_t^2 - \int_0^t W_s dW_s \quad \Rightarrow \quad \int_0^t W_s dW_s = \frac{W_t^2}{2}. \tag{7.1.3}$$

This is called the Stratonovic integral.

A more interesting result is achieved if we choose $\tau = t_i$ which defines the so called Itô integral:

$$\int_0^t W_s dW_s := \lim_{L^2} \sum_{i=0}^{n-1} W(t_i)(W(t_{i+1}) - W(t_i)) = \frac{W(t)^2 - t}{2}. \tag{7.1.4}$$

Theorem 7.1.2. *The Itô integral 7.1.4 is a martingale with respect to the filtration* $\mathscr{F}_t = \sigma\{W_u,\ u \leq t\}$.

PROOF Observe that $\mathbb{E}(W_s|\mathscr{F}_s) = W_s$ because W_s is $\mathscr{F}_s$ - misurable and because

$$\mathbb{E}(W_t|\mathscr{F}_s) - \mathbb{E}(W_s|\mathscr{F}_s) = \mathbb{E}(W_s - W_t|\mathscr{F}_s) = \mathbb{E}(W_t - W_s) = 0.$$

Moreover

$$\begin{aligned}\mathbb{E}(W_t^2 - W_s^2|\mathscr{F}_s) =& \mathbb{E}((W_t - W_s)^2|\mathscr{F}_s) + 2\mathbb{E}(W_s(W_t - W_s)|\mathscr{F}_s)\\ =& \mathbb{E}((W_t - W_s)^2) + 2W_s\mathbb{E}(W_t - W_s|\mathscr{F}_s) = t - s,\end{aligned}$$

then, since $W_s^2 - s$ is $\mathscr{F}_s$ - misurable,

$$\mathbb{E}(W_t^2 - t - W_s^2 + s|\mathscr{F}_s) = \mathbb{E}(W_t^2 - W_s^2) - (t - s) = 0.$$

□

Chapter 8

Semigroup of linear operators

In section 2.2.2 we have seen how transition probability can be expressed for several kind of Markov processes. Consider the function $f : \mathbb{R} \to \mathbb{R}$, $f \in L^\infty(\mathbb{R})$ and the integral

$$\int_{\mathbb{R}} f(y)p(s, x, t, dy).$$

We can integrate $f(y)$ with respect to dy because it is in $L^\infty(\mathbb{R})$, so

$$\int_{\mathbb{R}} |f(y)||p(s, x, t, dy)| \leq \|f\|_\infty \int_{\mathbb{R}} p(s, x, t, dy) < +\infty.$$

Now that we have estimated the integral, let us look at the following function

$$u(s, x, t, f) = \int_{\mathbb{R}} f(y)p(s, x, t, dy)$$

which,for fixed s, t, defines a linear operator from L^∞ to L^∞ denoted by $U_{s,t}f$:

$$(U_{s,t}f)(x) = u(s, x, t, f) = \int_{\mathbb{R}} f(y)p(s, x, t, dy).$$

Identity operator.

Recall now the properties of a Markov process, in particular $p(s, x; s, I) = \delta_x(I)$. In this situation the operator U_{ss} is just the identity operator I, indeed

$$(U_{s,s}f)(x) = \int_{\mathbb{R}} f(y)\delta_x(dy) = f(x).$$

Backward evolution operator.

Compute $(U_{s,t}f)(x)$ for any s, t.

If we consider an istant r such that $s < r < t$, we can write the operator as

$$(U_{s,t}f)(x) = U_{s,r}(U_{r,t}f)(x).$$

by definition we have

$$(U_{r,t}f)(x) = \int_{\mathbb{R}} f(y)p(r,x,t,dy) =: g(x),$$

so we may think to compute $U_{s,t}$ as $U_{s,r}$ applied to $g(x)$:

$$(U_{s,r}g)(x) = \int_{\mathbb{R}} g(y)p(s,x,r,dy) := g(y).$$

Let us change notation for $g(y)$:

$$g(y) = \int_{\mathbb{R}} f(z)p(r,y,t,dz).$$

Then

$$(U_{s,r}U_{r,t}f)(x) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(z)p(r,y,t,dz) \right) p(s,x,r,dy)$$

and, by Fubini,

$$(U_{s,r}U_{r,t}f)(x) = \int_{\mathbb{R}} f(z) \int_{\mathbb{R}} p(r,y,t,dz)p(s,x,r,dy), \tag{8.0.1}$$

where the second integral is just $p(s,x;t,dz)$ because of the Chapman - Kolmogorov relation. Due to the order in which operators $U_{s,r}$ and $U_{r,t}$ are applied to f, $U_{s,t}$ is called *backward evolution operator*. Notice that the result in 8.0.1 is valid only for Markov process non homogeneus in time, that is, for the family $\{p(s,x;t,I)\}$ and not for the simplified versions of transition probability seen in section 2.2.2.

Forward evolution operator.

Let us introduce an operator defined as

$$V_{t,s}\mu(I) = \int_{\mathbb{R}} \mu(dx)p(s,x;t,I) \tag{8.0.2}$$

and acting from the space of measures to itself (indeed the result of the integral in 8.0.2 is just another measure). As before, one can prove the relation

$$V_{t,s} = V_{t,r}V_{r,s},$$

that justifies the name of this operator.

We now turn to consider time homogeneus Markov processes. Let us introduce a function $u(t,x)$ defined as

$$u(t,x) = \mathbb{E}(f(X_t)|X_0 = x) = \int_{\mathbb{R}} f(y)\mathbb{P}(X_t \in dy|X_0 = x) = \int_{\mathbb{R}} f(y)p_t(x,dy)$$

and an operator defined as

$$(U_t f)(x) = u(t,x). \tag{8.0.3}$$

Theorem 8.0.1. (of semigroup relation). *The operator U resulting from 8.0.3 is linear over $\mathscr{L}^\infty$ and the following semigroup relation holds:*

$$U_{t+s} = U_t U_s = U_s U_t.$$

PROOF The operator U_s applied to $f(y)$ is

$$g(y) = (U_s f)(y) = \int_{\mathbb{R}} f(z)p_s(y,dz),$$

then

$$\begin{aligned}(U_t g)(x) = \int_{\mathbb{R}} g(y)p_t(x,dy) &= \int_{\mathbb{R}} p_t(x,dy) \int_{\mathbb{R}} f(z)p_s(y,dz) = \\ &= \int_{\mathbb{R}} f(z) \int_{\mathbb{R}} p_s(y,dz)p_t(x,dy) = \\ &= \int_{\mathbb{R}} f(z)p_{s+t}(x,dz) = U_{s+t}f(x)\end{aligned}$$

□

Theorem 8.0.2. (Kolmogorov equation). *The function $u(t,x)$ satisfies*

$$\frac{\partial u}{\partial t} = Au(t,x).$$

PROOF To obtain the derivative of $u(t,x)$ consider

$$u(t+h,x) - u(t,x) = U_{t+h}f(x) - U_t f(x) = (U_h - I)U_t f(x).$$

Then

$$\frac{u(t+h,x) - u(t,x)}{h} = \frac{U_h - I}{h} U_t f(x) = \frac{U_h g(x) - g(x)}{h},$$

where we have set $g(x) = U_t f(x)$. We now consider the set

$$D_A = \left\{ g \in L^\infty : \frac{U_h g - g}{h} \longrightarrow Ag \right\},$$

where A is a linear operator.
Taking $f(x) \in D_A$ we obtain the following partial differential equation

$$\begin{cases} \frac{\partial u}{\partial t} = Au(t,x) \\ u(0,x) = f(x) \end{cases}$$

This is called the Kolmogorov equation (valid for time homogeneus Markov processes) and it can be proved that it is a parabolic PDE and that the operator A is a second order operator such that

$$Au(x) = \frac{1}{2}a(x)\frac{\partial^2 u}{\partial x^2} + b(x)\frac{\partial u}{\partial x}.$$

□

Theorem 8.0.3. (Wiener equation).

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t.$$

PROOF Consider a Wiener process W_t. It satisfies

$$p_t(x,y) = \frac{1}{\sqrt{2\pi t}}e^{-\frac{(y-x)^2}{2t}}$$

from which we obtain the operator

$$Au = \frac{1}{2}\frac{\partial^2 u}{\partial x^2}.$$

In the Wiener equation $b(x)$ is called drift and $\sigma(x) = \sqrt{a(x)}$ is called diffusion. □

Fokker - Plank equation.
Let $V_{s,t}$ be an operator over the space of measures and

$$v(t,I) = V_{t,s}\mu(I),$$

that yields the equation

$$\frac{\partial v(t,I)}{\partial t} = B(t)v(t,I). \quad (8.0.4)$$

Equation 8.0.4 togeter with an initial condition is the Fokker - Plank equation (or forward Kolmogorov equation). It can be shown that forward and backward equation are intimately related, such that

$$A(s)^* = B(s).$$

For example, if

$$A(s)\phi = \frac{1}{2}a(x)\phi_{xx} + b(x)\phi_x,$$

integrating by parts one gets

$$A(s)^*\psi = \frac{1}{2}(a\psi)_{xx} - (b\psi)_x.$$

Consider again time homogeneus Markov property adn $U_{s,t}f(x) = U_{t-s}f(x)$. Fokker - Plank equation becomes

$$\frac{\partial U_{s-t}}{\partial(s-t)} = AU_{s-t}$$

and

$$B = A^*.$$

In the light of what we have said in this last chapter, let us motion some operators asociated with known stochastic processes.

Wiener process.
Wiener process corresponds with the operator

$$A\phi = \frac{1}{2}\phi_{xx}.$$

In addition one can verify that it is a selfadjoint operator.

Brownian bridge.
Starting form the following stochastic differential equation

$$dX_t = -\frac{X_t}{1-t}dt + dW_t,$$

we obtain the (time dependent) operator

$$A(t)\phi = \frac{1}{2}\phi_{xx} - \frac{x}{1-t}\phi_x.$$

Ornstein - Uhlenbeck process.
Consider the process defined by the stochastic differential equation

$$dX_t = -\lambda X_t dt + \sigma dW_t.$$

The operator assumes here the following form

$$A\phi = \frac{1}{2}\sigma^2\phi_{xx} - \lambda x\phi_x$$

and its adjoint is

$$A^*\phi = \frac{1}{2}\sigma^2\phi_{xx} + \lambda(x\phi)_x.$$