An application of the supervised classification of stochastic process theory

Gabriel Ducrocq

Under the direction of
Christophe Denis

College year 2014/2015

Acknowledgements

I would like to thank the teaching staff of Université Paris-Est Marne la Vallée for the quality of its courses. More specifically, I wish to render thank to Mr Denis to his support during the internship and to Mr Elie for backing me up in my choices of career.

Abstract

In this paper we develop a two-steps method of supervised classification for diffusion paths. In the first, we present the general framework of bayesian classification and provide some important results and definitions of the stochastic processes theory. We also use the results given in [Cad13] to build the bayes rule for the brownian motion with drift, the Ornstein-Uhlenbeck process and the geometric brownian motion. In the second chapter, we discuss some related works such as [Cad13], [Den14] and we give our general procedure of classification: first we estimate the parameters of the processes using a maximum-likelihood estimate. Second, we plug these estimation in the bayes rule of the relevant process. Then, we apply it to the brownian motion with drift and the Ornstein-Uhlenbeck process. Finally, we provide a numerical study of the performances of the two-step method on these two processes.

Contents

Introduction

In this paper we consider a problem of supervised classification. But unlike classic classification problem, our features are functions, and more precisely stochastic processes. Since we aim to classify functionnal data, the traditional methods no longer holds. Hence, we have to develop new classification methods to take into account the trajectorial nature of the features. In our case, we only consider stochastic processes defined as stochastic differential equation. This case was studied in a number of papers, among [Cad13], [Den14]. and Moreover, we restrict our method to the parametric case: that is the drift and the volatility functions are totally parametric. We also assume that we do not observe the entire trajectorie but a discretized version of it. Relying on the results given in [Cad13], we develop a two-step classification method: first, we estimate the parameters of the SDE and then, we plug the estimates of the parameters in the bayes rules for the process.
The paper is organized as follow: in the first chapter we set the general framework of this paper, giving general results of stochastic processes theory and applying results of [Cad13] to well-known examples of stochastic processes. In the second chapter, we present the classification method that we propose and we review some related works. We also give a numerical study of the classification method we developped throughout this paper.

Chapter 1

In this chapter we give the general framework which we will consider throughout this paper. We begin introducing the problem of binary classification and apply it a gaussian mixture. Then we give some important results of stochastic calculus and the solutions of well-known processes. Finally, using the paper of Benoit Cadre [Cad13], we derive the bayes rule of theses processes.

1.1 Supervised Classification

In this section, we present the binary classification problem. The purpose is to build a decision function which attribute only one label to a given observation. More precisely, we study a particular approach of the supervised classification called the bayesian classification. The aim of this approach is to minimize the probability of misclassification.

1.1.1 General Framework

Let (X, Y) be a random vector taking its value in $(\mathbb{R}^p, \{0, 1\})$ and $\mathbb{P}$ its law. In the following, X is called the features and Y the label. We now give the first definition:

Definition 1. *A decision function, also called a classifier, is a measurable function $g : \mathbb{R}^n \to \{0, 1\}$. The set of all classifiers is denoted by $\mathcal{G}$, where $\mathcal{G} = \{g | g : \mathbb{P}^n \to \{0, 1\}\}$.*

Moreover, for each classification function we can define an error cost:

Definition 2. *Let $g \in \mathcal{G}$. The loss function is given by $l(X, Y) = \mathbf{1}_{\{Y \neq g(X)\}}$ The misclassification risk is defined as $R(g) = \mathbf{E}[l(X, Y)] = \mathbb{P}\{Y \neq g(X)\}$.*

Obviously, our goal is to use the classification function with the lowest misclassification risk. We introduce the following function: $\eta(x) = \mathbb{E}[Y|X = x] = \mathbb{P}\{Y = 1|X = x\}$. This function will play a central role in the following of this paper, allowing us to compute the Bayes decision as well as the plug-in decision function. Furthermore, the Bayes decision function is defined by:

Definition 3. *The Bayes decision function is given by* $g^*(x) = \mathbf{1}_{\{\eta(x)>1/2\}}$.

This function is particularly important because of the following theorem:

Theorem 1. *For any decision function* $g : \mathbb{R}^n \in \mathcal{G}$*, we have:*

$$\mathbb{P}\{g^*(X) \neq Y\} \leq \mathbb{P}\{g(X) \neq Y\} \tag{1.1}$$

The interpretation of this theorem is straight: the Bayes decision function is optimal according to the misclassification risk.

Proof. Let $g : \mathbb{R}^n \to \{0, 1\}$ be a classifier and g^* the Bayes classifier. We have:

$$\begin{aligned} R(g) - R(g^*) &= \mathbb{P}(g(X) \neq Y) - \mathbb{P}(g^*(X) \neq Y) \\ &= \textstyle\int_{\mathbb{R}^p} (\mathbb{P}(g(X) \neq Y|X = x) - \mathbb{P}(g^*(X) \neq Y|X = x)) dP_X(x) \end{aligned}$$

Moreover, for any classifier we have:

$$\begin{aligned} \mathbb{P}(g(X) \neq Y|X = x) &= 1 - \mathbb{P}(g(X) = Y|X = x) \\ &= 1 - [\mathbb{P}(g(X) = 1, Y = 1|X = x) + \mathbb{P}(g(X) = 0, Y = 0|X = x)] \\ &= 1 - [\mathbb{E}[\mathbf{1}_{\{Y=1\}}\mathbf{1}_{\{g(X)=1\}}|X = x] + \mathbb{E}[\mathbf{1}_{\{Y=0\}}\mathbf{1}_{\{g(X)=0\}}|X = x]] \\ &= 1 - [\mathbf{1}_{\{g(x)=1\}}\mathbb{E}[\mathbf{1}_{\{Y=1\}}|X = x] + \mathbf{1}_{\{g(x)=0\}}\mathbb{E}[\mathbf{1}_{\{Y=0\}}|X = x]] \\ &= 1 - [\mathbf{1}_{\{g(x)=1\}}\mathbb{P}(Y = 1|X = x) + \mathbf{1}_{\{g(x)=0\}}\mathbb{P}(Y = 0|X = x)] \\ &= 1 - [\mathbf{1}_{\{g(x)=1\}}\eta(x) + \mathbf{1}_{\{g(x)=0\}}(1 - \eta(x))] \end{aligned}$$

Now, we consider the difference:

$$\begin{aligned} &\mathbb{P}(g(X) \neq Y|X = x) - \mathbb{P}(g^*(X) \neq Y|X = x) \\ &= \eta(x)[\mathbf{1}_{\{g^*(x)=1\}} - \mathbf{1}_{\{g(x)=1\}}] + (1 - \eta(x))[\mathbf{1}_{\{g^*(x)=0\}} - \mathbf{1}_{\{g(x)=0\}}] \\ &= \eta(x)[\mathbf{1}_{\{g^*(x)=1\}} - \mathbf{1}_{\{g(x)=1\}}] - (1 - \eta(x))[\mathbf{1}_{\{g^*(x)=1\}} - \mathbf{1}_{\{g(x)=1\}}] \\ &= (2\eta(x) - 1)[\mathbf{1}_{\{g^*(x)=1\}} - \mathbf{1}_{\{g(x)=1\}}] \geq 0 \end{aligned}$$

by definition of g^*.
Then, it is clear that $R(g) - R(g^*) \geq 0$ for any classifier g.

□

1.1.2 Application to Gaussian Mixture

In order to understand the previous developpements, we apply the Bayes classifier to a simple Gaussian mixture model:
In the following, we assume that $X \in \mathbb{R}^n$, $Y \in \{0,1\}$ are random variables such that:

$$\begin{cases} \mathcal{L}(X|Y=0) \sim \mathcal{N}_n(\mu_0, \Sigma) \\ \mathcal{L}(X|Y=1) \sim \mathcal{N}_n(\mu_1, \Sigma) \end{cases}$$

where $\mu_0, \mu_1 \in \mathbb{R}^n$
For simplicity we assume that $\Sigma = Id_p$ and $\mathbb{P}(Y=0) = \mathbb{P}(Y=1) = \frac{1}{2}$

Proposition 1. *The Bayes classifier for the above gaussian mixture is:*

$$g^*(x) = \mathbf{1}_{\{(\mu_1-\mu_0)^T x \geq \frac{1}{2}(\|\mu_1\|^2 - \|\mu_0\|^2)\}} \tag{1.2}$$

Proof. It is clear that $\eta(x) = \dfrac{f_{X|Y=1}(x)}{f_{X|Y=1}(x) + f_{X|Y=0}(x)}$.
Hence: $\eta(x) = \dfrac{1}{1 + e^{-\frac{1}{2}(x-\mu_0)^T(x-\mu_0) + \frac{1}{2}(x-\mu_1)^T(x-\mu_1)}}$
Thus, we have:

$$\begin{aligned} g^*(x) &= \mathbf{1}_{\{1 \geq 1/2 + 1/2 e^{-1/2(x-\mu_0)^T(x-\mu_0) + 1/2(x-\mu_1)^T(x-\mu_1)}\}} \\ &= \mathbf{1}_{\{-(x-\mu_0)^T(x-\mu_0) + (x-\mu_1)^T(x-\mu_1) \leq 0\}} \\ &= \mathbf{1}_{\{(\mu_1-\mu_0)^T x \geq 1/2(\|\mu_1\|^2 - \|\mu_0\|^2)\}} \end{aligned}$$

□

Proposition 2. *The misclassification risk for this gaussian mixture is given by:*

$$R(g^*) = 1 - \Phi\left(\frac{\|\mu_1 - \mu_0\|}{2}\right)$$

where Φ is the distribution function of the standard normal distribution.

Proof. It is clear that we have:

$$\begin{aligned} \mathbb{P}(g^*(X) \neq Y) &= \mathbb{P}((g^*(X) = 1 \cap Y = 0) \cup (g^*(X) = 0 \cap Y = 1)) \\ &= \mathbb{P}((\mu_1 - \mu_0)^T X < 1/2(\|\mu_1\|^2 - \|\mu_0\|^2) | Y = 1)\mathbb{P}(Y = 1) \\ &\quad + \mathbb{P}((\mu_1 - \mu_0)^T X \geq 1/2(\|\mu_1\|^2 - \|\mu_0\|^2) | Y = 0)\mathbb{P}(Y = 0) \end{aligned}$$

Therefore, we can write:

$$\begin{aligned}\mathbb{P}(g^*(X) \neq Y) &= \frac{1}{2}(\mathbb{P}(\frac{(\mu_1-\mu_0)^T X - (\mu_1-\mu_0)^T\mu_1}{\|\mu_1-\mu_0\|} < \frac{1}{2}\|\mu_1-\mu_0\||Y=1) \\ &\quad +\mathbb{P}(\frac{(\mu_1-\mu_0)^T X - (\mu_1-\mu_0)^T\mu_0}{\|\mu_1-\mu_0\|} \geq -\frac{1}{2}\|\mu_1-\mu_0\||Y=0))\end{aligned}$$

Moreover whe have:

$$\begin{aligned}\mathcal{L}((\mu_1-\mu_0)^T X|Y=0) &= \mathcal{N}((\mu_1-\mu_0)^T\mu_0, \|\mu_1-\mu_0\|^2) \\ \mathcal{L}((\mu_1-\mu_0)^T X|Y=1) &= \mathcal{N}((\mu_1-\mu_0)^T\mu_1, \|\mu_1-\mu_0\|^2)\end{aligned}$$

Thus, we have

$$\begin{aligned}\mathcal{L}(Z_0 = \frac{(\mu_1-\mu_0)^T X - (\mu_1-\mu_0)^T\mu_0}{\|\mu_1-\mu_0\|}|Y=0) &\sim \mathcal{N}(0,1) \\ \mathcal{L}(Z_1 = \frac{(\mu_1-\mu_0)^T X - (\mu_1-\mu_0)^T\mu_1}{\|\mu_1-\mu_0\|}|Y=1) &\sim \mathcal{N}(0,1)\end{aligned}$$

Therefore:

$$\begin{aligned}\mathbb{P}(g^*(X) \neq Y) &= \frac{1}{2}(\mathbb{P}(\frac{(\mu_1-\mu_0)^T X - (\mu_1-\mu_0)^T\mu_1}{\|\mu_1-\mu_0\|} < \frac{1}{2}\|\mu_1-\mu_0\||Y=1) \\ &\quad +\mathbb{P}(\frac{(\mu_1-\mu_0)^T X - (\mu_1-\mu_0)^T\mu_0}{\|\mu_1-\mu_0\|} < \frac{1}{2}\|\mu_1-\mu_0\||Y=0))\end{aligned}$$

And finally:

$$R(g^*) = 1 - \Phi(\frac{\|\mu_1-\mu_0\|}{2})$$

□

1.2 Diffusion Processes

In this section we give some definition and usual notations about diffusion processes. We also provide conditions of existence and uniqueness of solution to stochastic differential equation and apply this result to some well-known diffusion processes. Finally, thanks to the markov property of the diffusion processes, we can probvide two ways to approximate this processes: the Euler-Maruyama scheme and the maximum-likelihood estimate.

1.2.1 Notations, definitions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $T > 0$ a constant, $B = (B_t)_{t\in\mathbb{R}^+}$ be a m-dimensional brownian motion, $X_0 \in \mathbb{R}^n$ be a random variable and $\mathcal{F}(t) = \sigma(\{B_s | 0 \leq s \leq t\})$ the σ-algrebra generated by B up to time t. $\{\mathcal{F}(t)\}_{t\in[0,T]}$ is a filtration.

Definition 4. *The progressive σ-algebra is the set of all $A \in \mathcal{F}$ such that:*

$$\forall t \in [0,T], A \cap (\Omega \times [0,T]) \in \mathcal{F}_t \otimes \mathcal{B}(\mathbb{R}_+)$$

A set of this σ-algebra is said progressively measurable.
Moreover, a stochastic process is said progressively measurable if it is measurable with respect to this σ-algebra.

We will denote by $\Lambda^2(\mathbb{R}_+)$ (resp.$\Lambda^2([0,T])$) the subspace of $\mathbb{L}^2(\mathbb{R}_+ \times \Omega, dt \otimes d\mathbb{P})$ (resp. $\mathbb{L}^2([0,T]\times\Omega, dt\otimes d\mathbb{P})$) made of progressively measurable stochastic process.

Definition 5. *We say that a process $X : \Omega \times [0,T] \to \mathbb{R}^n$ is a solution of the following Stochastic Differential Equation (SDE):*

$$\begin{cases} dX_t = & b(X,t)dt + a(X,t)dB_t \\ X(0) = & X_0 \end{cases} \tag{1.3}$$

such that:

$$b : \mathbb{R}^n \times [0,T] \to [0,1]$$

And

$$a : \mathbb{R}^n \times [0,T] \to \mathbb{R}^{m\times n}$$

for $0 \leq t \leq T$ if:

1. *X is progressively measurable with respect to the filtration $\{\mathcal{F}\}_{t\in[0,T]}$*
2. *$b \in \mathbb{L}^1_n(0,T)$*
3. *$a \in \mathbb{L}^2_{n\times m}(0,T)$*
4. *$X_t = X_0 + \int_0^t b(X_s,s)ds + \int_0^t a(X_s,s)dB_s$ a.s for all $0 \leq t \leq T$*

Here, the integral $\int_0^t a(X_s,s)dB_s$ denotes the classic stochastic integral.

The definition of the solution of a Stochastic Differential Equation being introduced, we can now state the following theorem, which gives us the conditions of existence and uniqueness of the solution of an SDE:

Theorem 2. *Suppose that b and B are continuous and satisfy the following conditions:*

1. $|b(x,t) - b(y,t)| \leq L|x-y|$ *for all* $0 \leq t \leq T$, $x, y \in \mathbb{R}^n$
2. $|a(x,t) - a(y,t)| \leq L|x-y|$ *for all* $0 \leq t \leq T$, $x, y \in \mathbb{R}^n$
3. $|b(x,t)| \leq L(1+|x|)$ *for all* $0 \leq t \leq T$, $x \in \mathbb{R}^n$
4. $|a(x,t)| \leq L(1+|x|)$ *for all* $0 \leq t \leq T$, $x \in \mathbb{R}^n$

for some constant L.
Let X_0 be any $\mathbb{R}^n$-valued random variable such that:

- $E(|X_0|^2) < \infty$
- X_0 *is independant of the σ-algebra:* $\sigma(\{B_s - B_0 | s \geq 0\})$

Then, there exist an unique solution $X \in \mathbb{L}_n^2(0,T)$ to the SDE (1.3).

Moreover if we denote by $X_s^{x,t}$ the solution of the following SDE:

$$\begin{aligned} X_s^{x,t} &= x + \int_s^t b(r, X_r^{x,t})dr + \int_s^t a(r, X_r^{x,t})dB_r \\ X_s &= x \; si \; s \leq t \end{aligned}$$

where $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^d$ *and* $s \geq t$. we have:

Proposition 3. *For any bounded measurable function f on $\mathbb{R}^d$, we have:*

$$\forall u \geq 0, \mathbb{E}[f(X_{t+u}^{x,0})|\mathcal{F}_u] = \mathbb{E}[f(X_t^{X_u^{x,0},0})] a.s.$$

We say that theses process is a Markov process.

1.2.2 Examples

We now give the solution of some well-known stochastic differential equations. The Ornstein-Uhlenbeck process and the Black-Scholes equation are widely used in finance. The first one is used to model interests rates as well as currency exchange rate. The second one is used in option pricing.
We begin with the Ornstein-Uhlenbeck process, defined by the following equation:

$$dX_t = \theta(\mu - X_t)dt + \sigma dB_t \tag{1.4}$$

where $\mu, \theta, \sigma > 0$

Proposition 4. *The solution of the equation* (1.4) *is given by:*

$$X_t = \mu + e^{-\theta t}(X_0 - \mu) + \sigma \int_0^t e^{-\theta(t-u)} dB_u \tag{1.5}$$

Proof. We begin making the following change of variable:

$$Y_t = X_t - \mu$$

Then, we can write the Ornstein-Uhlenbeck equation this way:

$$dY_t = -\theta Y_t dt + \sigma dB_t$$

We no apply a second change of variable:

$$Z_t = e^{\theta t} Y_t$$

So, applying the Ito's lemma we have:

$$\begin{aligned} Z_t &= \theta e^{\theta t} Y_t dt + e^{\theta t} dY_t \\ &= \theta e^{\theta t} Y_t dt + e^{\theta t}(-\theta Y_t dt + \sigma dB_t) \\ &= e^{\theta t} \sigma dB_t \end{aligned}$$

Integrating, we have:

$$Z_t = Z_0 + \sigma \int_0^t e^{\theta u} dB_u$$

Hence, reversing the changes of variables:

$$Y_t = e^{-\theta t} Y_0 + \sigma \int_0^t e^{-\theta(t-u)} dB_u$$

Finally:

$$X_t = \mu + e^{-\theta t}(X_0 - \mu) + \sigma \int_0^t e^{-\theta(t-u)} dB_u$$

□

We introduce the Geometric Brownian Motion process:

$$dX_t = \mu X_t dt + \sigma X_t dB_t \tag{1.6}$$

Proposition 5. *The solution of the equation* (1.6) *is given by:*

$$X_t = X_0 e^{(\mu - \frac{\sigma^2}{2})t + \sigma B_t} \tag{1.7}$$

Proof. We first divide the equation (1.6) by X_t, so we have:

$$\frac{dX_t}{X_t} = \mu dt + \sigma dB_t$$

Furthermore, assuming that $B_0 = 0$ we have:

$$\frac{dX_t}{X_t} = \mu t + \sigma B_t$$

Then, applying the Ito's formula to the function $f(t, X_t) = \log(X_t)$ we obtain:

$$d\log(X_t) = \frac{dX_t}{X_t} - \frac{1}{2}\frac{1}{X_t}(\sigma X_t)^2 dt$$

So it's obvious that:

$$d\log(X_t) = (\mu - \frac{1}{2}\sigma^2)dt + \sigma dB_t$$

Integrating, we have:

$$\log(X_t) - \log(X_0) = (\mu - \frac{1}{2}\sigma^2)t + \sigma B_t$$

And eventually:

$$X_t = X_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma B_t}$$

□

We give the graphics of an Ornstein-Uhlenbeck process and a Black-Scholes equation simulated with in Figure (1.2.2) and graphics of a Geometric-Brownian motion in Figure (1.2.2) at the end of this chapter.

As we know the explicit solutions of these processes, we can easily simulate them. Indeed, the processes have the markov property.
In the following, we consider that the interval $[0, 1]$ is subdivided in $t_0 = 0 \leq t_1 \leq \ldots \leq t_N = 1$ such that $\Delta = \frac{1}{N} = t_i - t_{i-1}\ \forall i \in \{1, \ldots, N\}$ We demonstrated above that the Ornstein-Uhlenbeck process can be written as:

$$X_t = \mu + e^{-\theta t}(X_0 - \mu) + \sigma \int_0^t e^{-\theta(t-u)} dB_u$$

So, we can write:

$$\begin{aligned}
X_{t_j} &= \mu + e^{-\theta t_j}(X_0 - \mu) + \sigma \int_0^{t_j} e^{-\theta(t_j - u)} dB_u \\
&= \mu + e^{-\theta(t_j - t_{j-1} + t_{j-1})}(X_0 - \mu) + \sigma \int_0^{t_j} e^{-\theta(t_j - u)} dB_u \\
&= \mu + \mu e^{-\theta\Delta} - \mu e^{-\theta\Delta} + e^{-\theta\Delta}(X_0 - \mu)e^{\theta t_{j-1}} + \sigma \int_0^{t_j} e^{-\theta(t_j - u)} dB_u \\
&= \mu(1 - e^{-\theta\Delta}) + e^{-\theta\Delta}(\mu + (X_0 - \mu)e^{-\theta t_{j-1}}) + \sigma \int_0^{t_j} e^{-\theta(t_j - t_{j-1} + t_{j-1} - u)} dB_u \\
&= \mu(1 - e^{-\theta\Delta}) + e^{-\theta\Delta}(\mu + (X_0 - \mu)e^{-\theta t_{j-1}}) + \sigma e^{-\theta\Delta} \int_0^{t_j} e^{-\theta(t_{j-1} - u)} dB_u \\
&= \mu(1 - e^{-\theta\Delta}) + e^{-\theta\Delta}(\mu + (X_0 - \mu)e^{-\theta t_{j-1}}) + \sigma e^{-\theta\Delta} \int_0^{t_{j-1}} e^{-\theta(t_{j-1} - u)} dB_u \\
&\quad + \sigma e^{-\theta\Delta} \int_{t_{j-1}}^{t_j} e^{-\theta(t_{j-1} - u)} dB_u \\
&= \mu(1 - e^{-\theta\Delta}) + e^{-\theta\Delta} X_{t_{j-1}} + \sigma e^{-\theta\Delta} \int_{t_{j-1}}^{t_j} e^{-\theta(t_{j-1} - u)} dB_u
\end{aligned}$$

Finally, we have:

$$X_{t_j} = \mu(1 - e^{-\theta\Delta}) + e^{-\theta\Delta} X_{t_{j-1}} + \eta_j \tag{1.8}$$

where $(\eta_j)_j \sim \mathcal{N}(0, \sigma^2 \frac{1 - e^{-2\theta\Delta}}{2\theta})$ are i.i.d.

We can follow the same reasoning for the Geometric Brownian-Motion: We know that:

$$Y_t = \log(X_t) = (\mu - \frac{1}{2}\sigma^2)t + \sigma B_t$$

Thus, we can write:

$$\begin{aligned}
Y_{t_j} &= (\mu - \frac{1}{2}\sigma^2)t_j + \sigma B_{t_j} \\
&= (\mu - \frac{1}{2}\sigma^2)\Delta + (\mu - \frac{1}{2}\sigma^2)t_{j-1} + \sigma(B_{t_j} - B_{t_{j-1}}) + \sigma B_{t_{j-1}}
\end{aligned}$$

So, we have:

$$\log(X_{t_j}) = \log(X_{t_{j-1}}) + (\mu - \frac{1}{2}\sigma^2)\Delta + \eta_j \tag{1.9}$$

where $(\eta_j)_j \sim \mathcal{N}(0, \sigma^2\Delta)$ are i.i.d.
The fact that we have explicit solutions to these EDS make possible this easy

way to simulate them. But we can't always compute exact solutions. In this case, we can rely on the following method:

Definition 6. *Euler-Maruyama Scheme:*
Given the following SDE: $dX_t = a(X_t)dt + b(X_t)dB_t$ *with initial condition* $X_0 = x_0$ *on the interval of time* $[0, T]$*, the Euler-Maruyama approximation of the true solution* X *is the markov chain* Y *defined as follow:*

- *partition the interval* $[0, T]$ *into* N *equals subintervals of width* $\Delta t = \dfrac{T}{N}$

$$0 = \tau_0 < \tau_1 < \cdots < \tau_N = T$$

- *set* $Y_0 = x_0$
- *recursively define* Y_n *for* $1 \leq n \leq N$ *by:*

$$Y_{n+1} = Y_n + a(Y_n)\Delta t + b(Y_n)\Delta B_n$$

where

$$\Delta B_n = B_{\tau_{n+1}} - B_{\tau_n} \tag{1.10}$$

Obviously, the variable ΔB_n *are gaussian with expected value* 0 *and variance* Δt*.*

1.3 Girsanov theorem

In order to compute the bayes rule for some stochastic processes, we will need the following two useful results:

Theorem 3. *Cameron-Martin theorem:*
For any function $f \in \mathbb{L}^2(\mathbb{R}_+)$ *we have:*

- *The process* $t \mapsto e^{\int_0^t f(u)dB_u - \frac{1}{2}\int_0^t f^2(u)du}$ *is a martingale.*
- *We define on* $\mathcal{F}_\infty = \sigma(\cup_{t\geq 0}\mathcal{F}_t)$ *a new probability measure:*

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = e^{\int_0^\infty f(u)dB_u - \frac{1}{2}\int_0^\infty f^2(u)du}$$

- *Under* $\mathbb{Q}$*, the process* $t \mapsto \tilde{B}_t = B_t - \int_0^t f(u)du$ *is a* $(\mathcal{F}_t)_t$*- Brownian Motion.*

Proof. 1. It is straight that $\mathbb{E}[e^{\int_0^t f(u)dB_u - 1/2\int_0^t f(u)^2du}] = 1\ \forall t \in [0,T]$ because of the generating function of the normal law ($\int_0^t f(u)dB_u$ follows normal law with mean 0 and variance $\int_0^t f(u)^2du$.
Moreover, for all $0 < s < t < T$ we have:

$$\begin{aligned}\mathbb{E}[e^{\int_0^t f(u)dB_u - 1/2\int_0^t f(u)^2du}|\mathcal{F}_s] &= \mathbb{E}[e^{\int_0^t f(u)dB_u}|\mathcal{F}_s]e^{-1/2\int_0^t f(u)^2du} \\ &= \mathbb{E}[e^{\int_s^t f(u)dB_u}|\mathcal{F}_s]e^{-1/2\int_0^t f(u)^2du + \int_0^s f(u)dB_u}\end{aligned}$$

But we know that:

$$\begin{aligned}\mathbb{E}[\int_s^t f(u)dB_u|\mathcal{F}_s] &= 0 \\ \mathbb{E}[(\int_s^t f(u)dB_u)^2|\mathcal{F}_s] &= \int_s^t f(u)^2du\end{aligned}$$

Hence, using the Laplace Transform of a Gaussian variable, we have:

$$\mathbb{E}[e^{\int_s^t f(u)dB_u}|\mathcal{F}_s] = e^{1/2\int_s^t f(u)^2du}$$

Thus:

$$\mathbb{E}[e^{\int_0^t f(u)dB_u - 1/2\int_0^t f(u)^2du}|\mathcal{F}_s] = e^{\int_0^s f(u)dB_u - 1/2\int_0^s f(u)^2du}$$

2. The fact that $\mathbb{Q}$ is a probability measure is straight
Moreover, let $A \in \mathcal{F}_t$, we have:

$$\begin{aligned}\mathbb{Q}(A) &= \mathbb{E}^{\mathbb{P}}[e^{\int_0^\infty f(u)dB_u - \frac{1}{2}\int_0^\infty f^2(u)du}\mathbf{1}_A] \\ &= \mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}[e^{\int_0^\infty f(u)dB_u - \frac{1}{2}\int_0^\infty f^2(u)du}\mathbf{1}_A|\mathcal{F}_t]] \\ &= \mathbb{E}^{\mathbb{P}}[\mathbf{1}_A\mathbb{E}^{\mathbb{P}}[e^{\int_0^\infty f(u)dB_u - \frac{1}{2}\int_0^\infty f^2(u)du}|\mathcal{F}_t]] \\ &= \mathbb{E}^{\mathbb{P}}[e^{\int_0^t f(u)dB_u - \frac{1}{2}\int_0^t f^2(u)du}\mathbf{1}_A]\end{aligned}$$

Because the martingale is regular. So, $\mathbb{Q}$ admit a Radon-Nykodim derivative with respect to $\mathbb{P}$. Moreover, $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$.

3. • Let $N = \{\omega, t \mapsto B_t(\omega) \text{ is not continuous on } \mathbb{R}_+\}$.
We know that $\mathbb{P}(N) = 0$, so $\mathbb{Q}(N) = 0$.

• Furthermore, let φ be a step function over $\mathbb{R}_+$.

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}[e^{\int_0^\infty \varphi_r d\tilde{B}_t}] &= \mathbb{E}^{\mathbb{P}}[e^{\int_0^\infty \varphi_r dB_r - \int_0^\infty \varphi_r f_r dr} e^{\int_0^\infty f_r dB_r - \frac{1}{2}\int_0^\infty f_r^2 dr}] \\
&= \mathbb{E}^{\mathbb{P}}[e^{\int_0^\infty (\varphi_r + f_r) dB_r}] e^{-\int_0^\infty (\varphi_r f_r + \frac{1}{2} f_r^2) dr} \\
&= e^{\frac{1}{2}\int_0^\infty (\varphi_r + f_r)^2 dr - \int_0^\infty (\varphi_r f_r + \frac{1}{2} f_r^2) dr} \\
&= e^{\frac{1}{2}\int_0^\infty \varphi_r^2 dr}
\end{aligned}$$

Hence, $\tilde{B}$ is a centered gaussian process.

• If $\varphi = \lambda \mathbf{1}_{]a;b]}$ where $a, b \in \mathbb{R}_+$ we have:

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}[e^{\lambda(\tilde{B}_b - \tilde{B}_a)}] &= e^{\frac{\lambda^2}{2}(b-a)} \\
\mathbb{E}^{\mathbb{Q}}[e^{\lambda(\tilde{B}_b - \tilde{B}_a)}] &= e^{\frac{\lambda^2}{2}\mathbb{E}^{\mathbb{Q}}[(\tilde{B}_b - \tilde{B}_a)^2]}
\end{aligned}$$

Because $\tilde{B}_t$ is a gaussian process. Hence:

$$\mathbb{E}^{\mathbb{Q}}[(\tilde{B}_b - \tilde{B}_a)^2] = b - a$$

for $a < b$.
And if $a = 0$, we have $\mathbb{E}^{\mathbb{Q}}[\tilde{B}_b^2] = b$. Therefore, expanding the square, we have:$\mathbb{E}^{\mathbb{Q}}[\tilde{B}_b \tilde{B}_a] = a$.

We just proved that $(\tilde{B}_t)_t$ is a centered process with the covariance function $K(a, b) = min(a, b)$. Hence, it is a brownian motion under $\mathbb{Q}$. □

Theorem 4. *Girsanov theorem:*
We assume that $(b_t)_{t\geq 0}$ is progressively measurable and verifies the Nokikov criterion, i.e $\mathbb{E}[e^{\frac{1}{2}\int_0^\infty |b_r|^2 dr}] < +\infty$. Then:

- *The process $t \mapsto e^{\int_0^t b(u) dB_u - \frac{1}{2}\int_0^t b^2(u) du}$ is a $(\mathcal{F}_t)_t$- martingale regular.*
- *We define a new probability measure $\frac{d\mathbb{Q}}{d\mathbb{P}} = e^{\int_0^\infty b(u) dB_u - \frac{1}{2}\int_0^\infty b^2(u) du}$*

- *Under $\mathbb{Q}$, the process $t \mapsto \tilde{B}_t = B_t - \int_0^t b(u)du$ is a $(\mathcal{F}_t)_t$- Brownian Motion.*

The Girsanov theorem is a fundamental result of stochastic calculus.

1.4 Classification of Stochastic Processes

In this section we derive the bayes rule for several diffusion processes, using the framework of B.Cadre in [Cad13]. In this paper, the author give some results for bayes classifiction of diffusion path with specific hypothesis about the processes. Moreover, B.Cadre prove the consistency of its approach. In the following we recall some of his results and apply them to several examples.

1.4.1 Notations

In this section we assume that $(X, Y) \in C([0,1])\text{x}\{0,1\}$ with $p_0 = \mathbb{P}(Y = 0) = \frac{1}{2}$ and that $(B_t)_{t\in[0,1]}$ is a standard Brownian motion independant of Y. Moreover, we assume that the process $X = (X_t)_{t\in[0,1]}$ and Y are related by the following equation:

$$\begin{cases} dX_t = b_0(t, X_t)dt + \sigma_0(t, X_t)dB_t \text{ when } Y = 0 \\ dX_t = (b_0(t, X_t) + (f_0\sigma_0)(t, X_t))dt + \sigma_0(t, X_t)dB_t \text{ when } Y = 1 \end{cases} \tag{1.11}$$

for all t in $[0,1]$ and where the initial value X_0 is independant of Y and the unknown borel functions $b_0, f_0, \sigma_0 : [0,1]\text{x}\mathbb{R} \to \mathbb{R}$ are such that each equation in (M) has a strong solution given the initial value X_0. We further assume that the Nikokov criterion is fullfiled:

$$\mathbb{E}e^{\frac{1}{2}\int_0^1 f_0(t,X_t)^2 dt} < \infty$$

Finally, in the sequel, for any function $r : [0,1]\text{x}\mathbb{R} \to \mathbb{R}$, we write the integral $\int_0^t r(t, X_t)dB_t$ (respectively $\int_0^t r(t, X_t)dt$) under the implicit condition:

$$\mathbb{E}\int_0^1 r(t, X_t)^2 dt < \infty \, (resp \, \mathbb{E}\int_0^1 |r(t, X_t)|dt < \infty)$$

And for $h = (r_1, r_2) : [0,1]\text{x}\mathbb{R} \to \mathbb{R}^2$, let

$$F(h, X) = \int_0^1 r_1(t, X_t)dX_t - \int_0^1 r_2(t, X_t)dt$$

We will also denote by μ_0 the conditional distribution of X given $Y = 0$ and by μ_1 the conditional distribution of X given $Y = 1$. We also define $\mu = \frac{1}{2}(\mu_0 + \mu_1)$ the distribution of X.

1.4.2 Bayes rule for diffusion processes

We derive the bayes rule for the classification problem describe in the previous subsection.

Theorem 5. *The probability μ_1 is absolutely continuous with respect to μ_0 with density $e^{F(h_0,.)}$, where:*

$$h_0 = (\frac{f_0}{\sigma_0}, \frac{f_0 b_0}{\sigma_0} + \frac{f_0^2}{2})$$

Proof. We now that f_0 fullfil the Nokikov criterion. Then, using the Girsanov's theorem, we know that it exists a new probability measure $\mathbb{Q}$ such that under this measure, $d\tilde{B}_t = dB_t - f_0(t, X_t)dt$ is a brownian motion. Moreover:

$$\frac{d\mathbb{Q}}{d\mu_0}(X) = e^{\int_0^1 f_0(t,X_t)dB_t - \frac{1}{2}\int_0^1 f_0^2(t,X_t)dt}$$

But we have:

$$\begin{aligned} dX_t^0 &\overset{\mathcal{L}}{=} b_0(t, X_t)dt + \sigma_0(t, X_t)dB_t \\ &\overset{\mathcal{L}}{=} \sigma_0(t, X_t)[f_0(t, X_t)dt + \frac{b_0}{\sigma_0}(t, X_t)dt - f_0(t, X_t)dt + dB_t] \\ &\overset{\mathcal{L}}{=} \sigma_0(t, X_t)f_0(t, X_t)dt + b_0(t, X_t)dt + \sigma_0(t, X_t)d\tilde{B}_t \text{ under } \mathbb{Q} \end{aligned}$$

Thus, we have: $\mathcal{L}_{\mathbb{P}}(X^1) = \mathcal{L}_{\mathbb{Q}}(X^0) = e^{\int_0^1 f_0(t,X_t)dB_t - \frac{1}{2}\int_0^1 f_0^2(t,X_t)dt}\mu_0$

But we know that $dB_t \overset{\mathcal{L}_{\mathbb{P}}}{=} \frac{dX_t^0}{\sigma_0(t, X_t^0)} - \frac{b_0}{\sigma_0}(t, X_t^0)dt$. Hence:

$$\mathcal{L}_{\mathbb{P}}(X^1) = e^{\int_0^1 \frac{f_0(t,X_t)}{\sigma_0(t,X_t)}dX_t - \int_0^1 (\frac{f_0 b_0}{\sigma_0}(t,X_t) + \frac{f_0^2(t,X_t)}{2})dt}\mu_0$$

Therefore, we have: $\frac{d\mu_1}{d\mu_0} = e^{\int_0^1 \frac{f_0(t,X_t)}{\sigma_0(t,X_t)}dX_t - \int_0^1 (\frac{f_0 b_0}{\sigma_0}(t,X_t) + \frac{f_0^2(t,X_t)}{2})dt}$ □

We want to compute the Bayes rule in that general case. It is clear that $\mathbb{E}[Y|X] = \mathbb{P}(Y = 1|X) = \frac{1}{2}\frac{d\mu_1}{d\mu}(X)$. And, from the proof of the above theorem, we derive the following equality:

$$\eta(X) = \mathbb{E}[Y|X] = \frac{e^{F(h_0,X)}}{1 + e^{F(h_0,X)}}$$

So, the next proposition is straight:

Proposition 6. *The Bayes rule writes as:*

$$g^*(X) = \mathbf{1}_{\{F(h_0,X)>0\}}$$

where $F(h_0, X) = \int_0^1 \frac{f_0(t, X_t)}{\sigma_0(t, X_t)} dX_t - \int_0^1 (\frac{f_0 b_0}{\sigma_0}(t, X_t) + \frac{f_0^2(t, X_t)}{2})dt$

1.4.3 Examples

It is time to apply the previous developpements to some examples.

Example 1. *Consider the following equation:*

$$\begin{cases} dX_t = \mu_0 dt + \sigma dB_t \text{ when } Y = 0 \\ dX_t = \mu_1 dt + \sigma dB_t \text{ when } Y = 1 \end{cases}$$

Where $\sigma, \mu_0, \mu_1 \in \mathbb{R}$ *and* $\mu_1 > \mu_0$.
We can consider this example as a simpler version of the gaussian mixture model. Indeed, we have:

$$\begin{cases} \mathcal{L}(X_t|Y = 0) \sim \mathcal{N}(\mu_0, \sigma^2 t) \\ \mathcal{L}(X_t|Y = 0) \sim \mathcal{N}(\mu_1, \sigma^2 t) \end{cases}$$

for all $t \in [0, 1]$.
Obviously, the second equation also writes as:

$$dX_t = (\mu_0 + \frac{(\mu_1 - \mu_0)}{\sigma}\sigma)dt + \sigma dB_t$$

So we can write:

$$\begin{aligned} F(h_0, X) &= \int_0^1 \frac{\mu_1 - \mu_0}{\sigma^2} dX_t - \int_0^1 \frac{(\mu_1 - \mu_0)\mu_0}{\sigma^2} + \frac{(\mu_1 - \mu_0)^2}{2\sigma^2} dt \\ &= \frac{(\mu_1 - \mu_0)}{\sigma^2}(X_1 - X_0) - (\frac{(\mu_1 - \mu_0)\mu_0}{\sigma^2} + \frac{(\mu_1 - \mu_0)^2}{2\sigma^2}) \end{aligned}$$

Finally, we compute the Bayes rule for this example:

$$g^*(X) = \mathbf{1}_{\{X_1 - X_0 > 1/2(\mu_1 - \mu_0)\}} \tag{1.12}$$

Example 2. *We can also compute the bayes rule for the Ornstein-Uhlenbeck process:*

$$\left\{ \begin{array}{rcl} dX_t & = & -\theta(X_t - \mu_0)dt + \sigma dB_t \\ dX_t & = & -\theta(X_t - \mu_1)dt + \sigma dB_t \end{array} \right. \tag{1.13}$$

where $\mu_0 < \mu_1$. The second equation also writes:

$$dX_t = -\theta(X_t - \mu_0) + \frac{\theta(\mu_1 - \mu_0)}{\sigma}\sigma dt + \sigma dB_t$$

$$F(h_0, X) = \int_0^1 \frac{\theta(\mu_1 - \mu_0)}{\sigma^2} dX_t + \int_0^1 \frac{2\theta^2(\mu_1 - \mu_0)(X_t - \mu_0)}{2\sigma^2} - \frac{\theta^2(\mu_1 - \mu_0)(\mu_1 - \mu_0)}{2\sigma^2} dt$$

Eventually, whe have the following Bayes rule:

$$g^*(X) = \mathbf{1}_{\{(X_1 - X_0) + \theta \int_0^1 X_t dt > \theta(\mu_1 + \mu_0)\}} \tag{1.14}$$

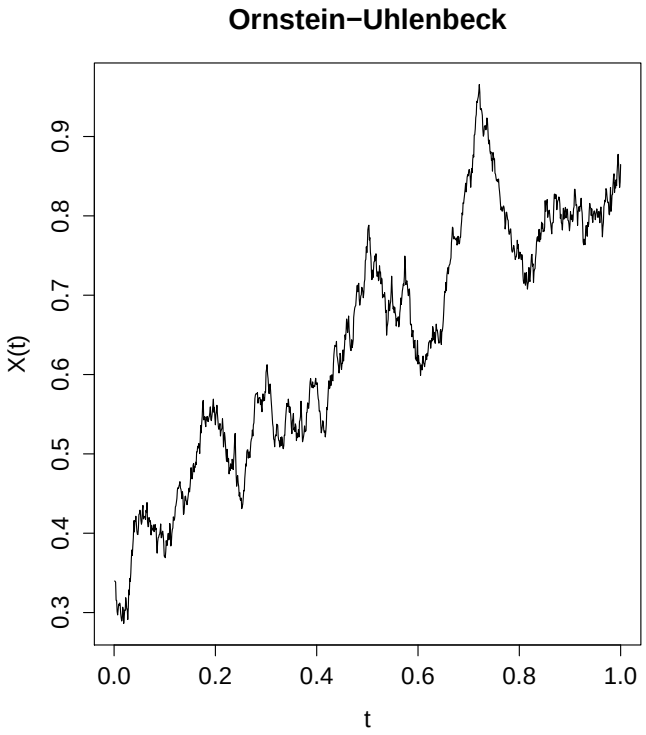

Figure 1.1: Ornstein-Uhlenbeck process with $\theta = 1$, $\mu = 1.2$ and $\sigma = 0.3$ and $N = 1000$.

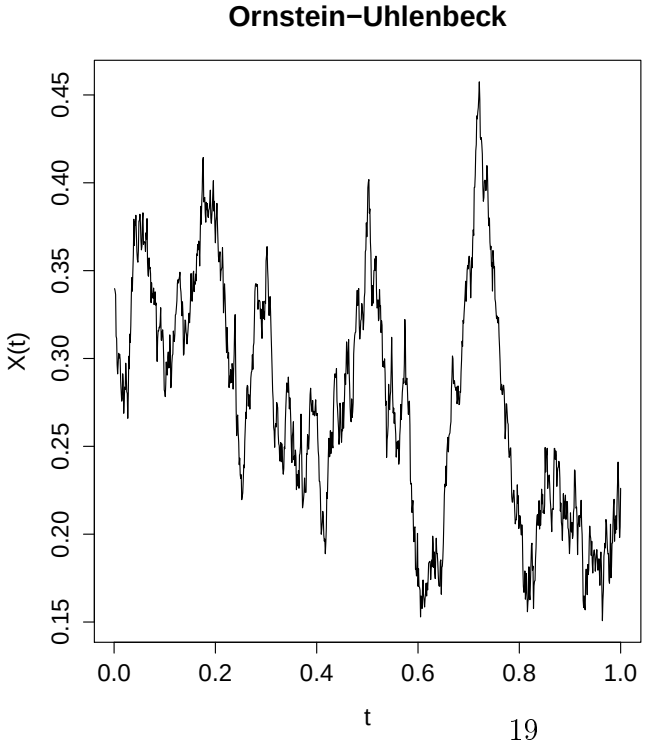

Figure 1.2: Ornstein-Uhlenbeck process with $\theta = 0.1$, $\mu = -0.5$ and $\sigma = 0.3$ and $N = 1000$.

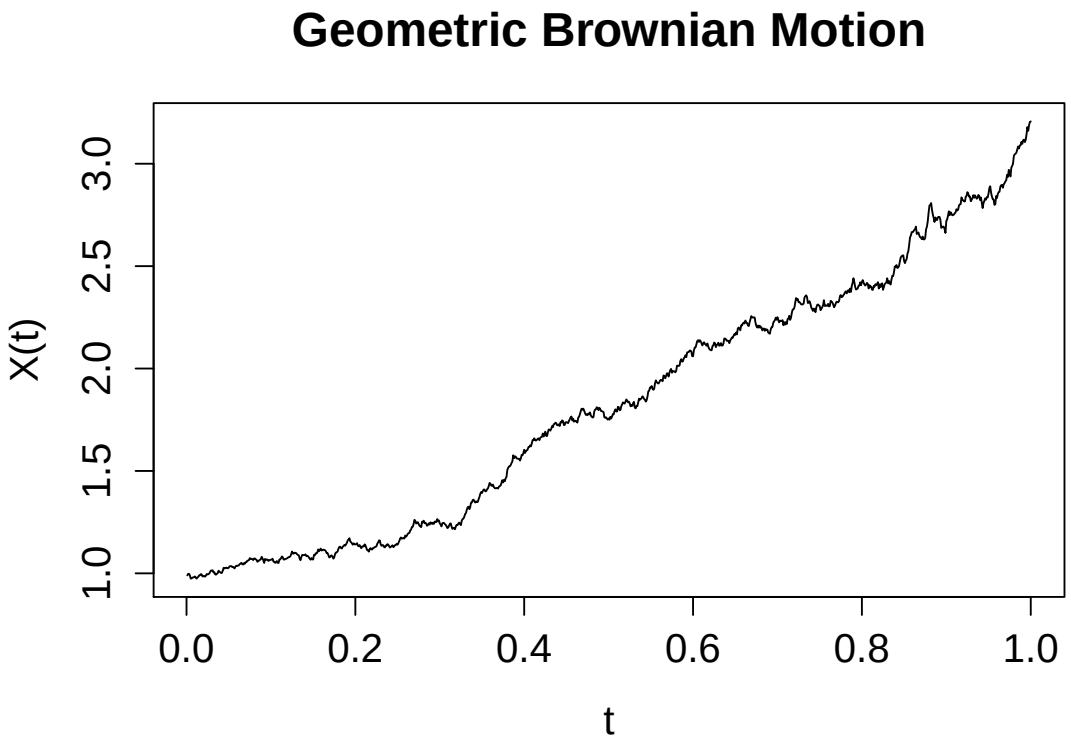

Figure 1.3: Geometric Brownian Motion process with $\mu = 1$,$\sigma = 0.2$ and $N = 1000$.

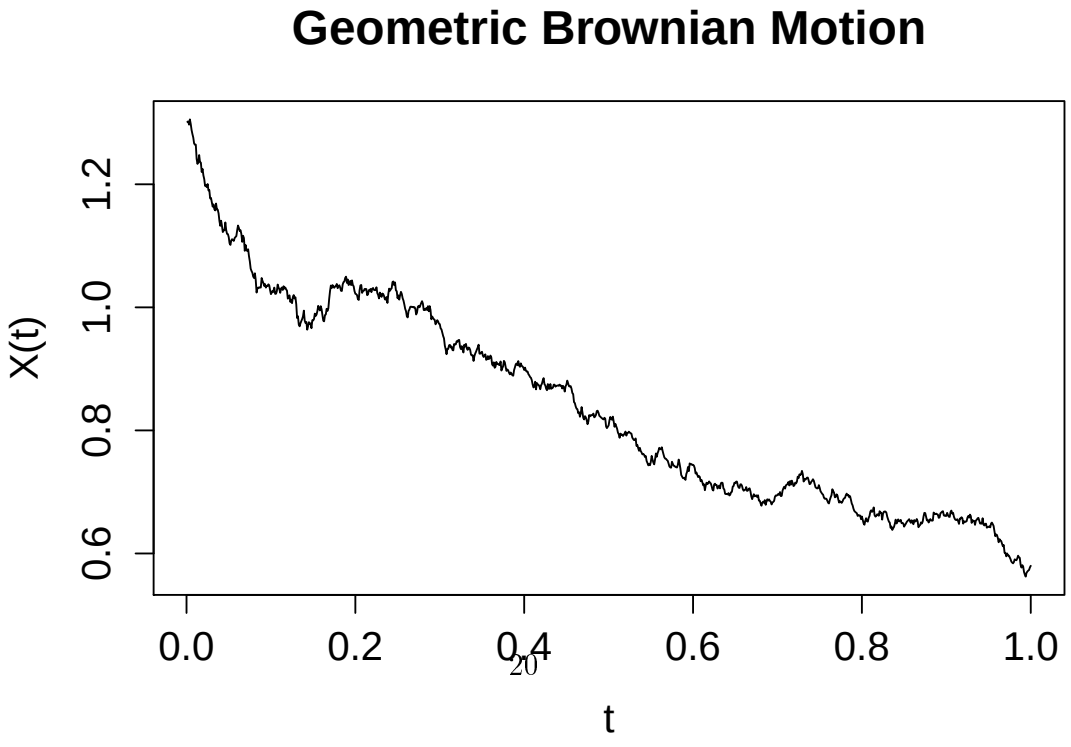

Figure 1.4: Geometric Brownian Motion process with $\mu = -0.4$,$\sigma = 0.2$ and $N = 1000$.

Chapter 2

The purpose of this chapter is to build a classifier over functional data, and more specifically over diffusion processes. More precisely, we assume that we observe a sample $\{(X_i, Y_i), i = 1, \cdots, n\}$ where $X_i = (X^i_{t_j})$ is a sample path observed at discrete time of the ith diffusion process and Y_i is the label of the same process. Moreover, X_i and Y_i are related as explained in the equation (1.11).

2.1 Related works

As we said earlier in this paper, a similar approach was developped in [Den14]. The aim of its paper is to classify the postural style of 70 subjects as hemiplegic, vestibular or normal. In order to do this, each subject completed two protocols involving two different kind of postural perturbations. They recorded the trajectory of each subject during these portocols. Then, they modeled these trajectories using a stochastic process: theCox-Ingersoll-Ross process. But unlike this paper, the authors didn't use the plug-in approach for the construction of a bayes rule in order to classify diffusion paths. Instead, they estimate the parameters of the CIR process and then use a random forest algorithm on these parameters.
On a more theoretical level, Benoît Cadre in his paper [Cad13] developped a theory of bayesian classification of diffusion path. Their approach differs from ours. In this paper, the authors assume that thoe observations are continuous. That is, they observe the entire trajectories and not a discretize version of it. Moreover, the estimation of the Bayes rule is achieved by directly minimizing the empirical risk. Unfortunately, the assumption of continuity of the observations make it very difficult to implement. One the other hand, they prove their estimator concistent.

2.2 Results

2.2.1 General Procedure

We said earlier that the optimal decision function is the Bayes decision:

$$g^*(x) = \mathbf{1}_{\{\eta(x)>1/2\}}$$

The problem is that in general we do not know the function η. But assume that we have access to a nonnegative function $\hat{\eta}(x)$ that approximate $\eta(x)$, it seems natural to use the plug-in decision function:

$$g(x) = \mathbf{1}_{\{\hat{\eta}(x)>1/2\}}$$

to approximate the Bayes decision. In this section we use a maximum-likelihood estimation of their parameters in order to approximate their bayes rule by a plug-in decision function. So, we propose a two step procedure here to determine $g(x)$ by plug-in. First, we have the following equations:

$$\begin{cases} dX_t &= b_{\theta_0}(t, X_t)dt + \sigma(t, X_t)dB_t \; when \, Y = 0 \\ dX_t &= b_{\theta_1}(t, X_t)dt + \sigma(t, X_t)dB_t \; when \, Y = 1 \end{cases}$$

By rewriting the second equation we can stick to the framework developped by Benoit Cadre in [Cad13]:

$$dX_t = b_{\theta_0}(t, X_t)dt + \frac{\sigma(t, X_t)}{\sigma(t, X_t)}(b_{\theta_1}(t, X_t) - b_{\theta_0}(t, X_t))dt + \sigma(t, X_t)dB_t$$

So, we consider the following problem:

$$\begin{cases} dX_t &= b_{\theta_0}(t, X_t)dt + \sigma(t, X_t)dB_t \; when \, Y = 0 \\ dX_t &= b_{\theta_0}(t, X_t)dt + \dfrac{\sigma(t, X_t)}{\sigma(t, X_t)}(b_{\theta_1}(t, X_t) - b_{\theta_0}(t, X_t))dt + \sigma(t, X_t)dB_t \; when \, Y = 1 \end{cases}$$

Thus, for f_θ, where $\theta = (\theta_0, \theta_1)$, has the following general form:

$$f_\theta = \frac{b_{\theta_1}(t, X_t) - b_{\theta_0}(t, X_t)}{\sigma(t, X_t)}$$

So, in order to estimate $F(h, X)$ by plug-in, we go through two stages:

- We use the training set we estimate the parameters θ_0 and θ_1 by $\hat{\theta_0}$ and $\hat{\theta_1}$ respectively.
- Once we have estimated the parameters, we only need to estimate the integrals of the form $\int_0^1 f(t, X_t)dX_t$ by $\Sigma_{i=1}^N f(t_{i-1}, X_{t_{i-1}})(X_{t_i} - X_{t_{i-1}})$

Then, we have $\hat{F}(h, X)$ and we can plug it and have our classifier $g(x)$.

2.2.2 Examples

- The first example we take is the brownian motion with drift, using the example of the section 1.4.3. We suppose that σ is known. Even if we can compute the exact solution of this SDE, we use the Euler-Maruyama Scheme in order to show the method to reader:

  For the first equation, the Euler-Maruyama method writes as:

  $$X_{t_{n+1}} = X_{t_n} + \mu_1 \frac{1}{N} + \sigma\eta_n$$

  for $0 = t_0 \leq t_2, \ldots, t_{N-1} \leq 1 = t_N$ where $\Delta = \frac{1}{N} = t_{n+1} - t_n \ \forall n \in \{1, \ldots, N-1\}$ where the variables η_n are independant normal variable of mean 0 and variance $\frac{1}{N}$.
  Therefore, the negative log-likelihood for one trajectory is:

  $$\begin{aligned} \mathcal{L}(\mu_0, \mu_1, \sigma) &= \log\left(\sqrt{\frac{2\pi}{N}}\right) + \log(\sigma) \\ &\quad + \frac{N}{\sigma^2}\Sigma_{i=0}^{N-1}(X_{t_{i+1}} - X_{t_i} - \frac{1}{N}\mu_0)^2 \end{aligned}$$

  Moreover, minimizing the previous equation with respect to μ_0 is equivalent to minimize the following expression:

  $$\frac{N}{\sigma^2}\Sigma_{i=0}^{N-1}(X_{t_{i+1}} - X_{t_i} - \frac{1}{N}\mu_0)^2$$

  Hence, differentiating this expression with respect to μ_0, we obtain the following estimators:

  $$\begin{aligned} \hat{\mu_0} &= \Sigma_{i=0}^{N-1} X_{t_{i+1}} - X_{t_i} \\ &= X_1 - X_0 \end{aligned}$$

  As we observe n_0 examples of the first process, we can take the mean in order to estimate n_0:

  $$\hat{\mu_0} = \frac{1}{n_0}\Sigma_{l=1}^{n_0} X_1^l - X_0^l \tag{2.1}$$

where $1 \leq l \leq n_0$. So, the estimate of μ_1 is straightfoward:

$$\hat{\mu_1} = \frac{1}{n_1}\Sigma_{k=1}^{n_1} X_1^k - X_0^k \tag{2.2}$$

where n_1 is the number of trajectories observed from the second equation and $1 \leq k \leq n_1$.
Now that we have estimated the parameters, we can plug them in the bayes rule for this example, given by (1.12). So, we have:

$$g(x) = \mathbf{1}_{\{X_1 - X_0 > \frac{(\hat{\mu_1} - \hat{\mu_0})}{2}\}} \tag{2.3}$$

Now that we have the Bayes rule for this simple example, we can compute the Bayes rules for the Ornstein-Uhlenbeck process and the geometric brownian motion more quickly:

- Ornstein-Uhlenbeck:
  The equation (1.8) give us an explicit solution for the Ornstein-Uhlenbeck process. Here, we assume that σ and θ are known. So, we know the following laws:

$$\mathcal{L}(X_{t_j}|X_{t_{j-1}}) = \mathcal{N}(e^{-\theta\Delta}X_{t_{j-1}} + \mu(1 - e^{-\theta\Delta}), \nu(\theta))$$

  where $\nu(\theta) = \sigma^2(\frac{1 - e^{-2\theta\Delta}}{2\theta})$ and $\Delta = \frac{1}{N} = t_{j+1} - t_j \ \forall j \in \{1, \ldots, N-1\}$. So, it is clear that the likelihood of the first equation writes:

$$\mathcal{L}_0 = \prod_{j=1}^{N} \frac{1}{\sqrt{\nu(\theta)}} e^{\frac{-1}{2}\frac{(X_{t_j} - X_{t_{j-1}}e^{-\theta\Delta} - \mu_0(1 - e^{-\theta\Delta})^2}{\nu(\theta)}}$$

  Thus, the log-likelihood is given by:

$$\log \mathcal{L}_0 = \frac{-N}{2}\log(\nu(\theta)) - \frac{1}{2}\Sigma_{j=1}^{N}\frac{(X_{t_j} - X_{t_{j-1}}e^{-\theta\Delta} - \mu_0(1 - e^{-\theta\Delta})^2}{\nu(\theta)}$$

  Maximizing this expression with respect to μ_0, we obtain:

$$\hat{\mu_0} = \frac{\Sigma_{j=1}^{N} X_{t_j} - X_{t_{j-1}}e^{-\theta\Delta}}{N(1 - e^{-\theta\Delta})}$$

  As we observe n_0 trajectories from the first equation, we can take the mean, and we have:

$$\hat{\mu_0} = \frac{\Sigma_{l=1}^{n_0}\Sigma_{j=1}^{N} X_{t_j} - X_{t_{j-1}}e^{-\theta\Delta}}{n_0 N(1 - e^{-\theta\Delta})}$$

Following the same reasoning for the second equation, we obtain:

$$\hat{\mu_1} = \frac{\Sigma_{k=0}^{n_1}\Sigma_{j=1}^{N} X_{t_j}^k - X_{t_{j-1}}^k e^{-\theta\Delta}}{n_1 N(1-e^{-\theta\Delta})}$$

Now that we have estimated μ_0 and μ_1, we can plug them into the bayes rule for the Ornstein-Uhlenbeck process given by (1.13). We obtain:

$$g(x) = \mathbf{1}_{\{(X_1-X_0)+2\int_0^1 \hat{X}_t dt > \hat{\mu}_1+\hat{\mu_0}\}}$$

where $\int_0^1 \hat{X}_t dt = \Sigma_{i=1}^{N}(X_{t_{i-1}})(t_i - t_{i-1})$

2.3 Numerical Study

Now that we have the Bayes rule for the three process of the previous section, we can make a numerical study and observe their behavior according to the chosen parameters.
For each process, we begin by simulating a training set $Tr = \{(x_i^l, y_i^l), i = 1, \ldots, n_1\} \cup \{(x_i^k, y_i^k), i = 1, \ldots, n_2\}$ of n_1 examples of the first equation and n_2 examples of the second one with their labels. We also simulate a test set of n examples with their labels as follow: each element of the test set is generated from the first equation with a probability 1/2 and from the second equation with a probability 1/2. $Te = \{(x_i, y_i), i = 1, \ldots, n\}$. We use the training set to estimate the parameters, as described in the previous section. Once we have those estimates, we perform classification on the examples of the test set. So, for each example $(x, y) \in Te$ of the test set, we can define the misclassification error: $e(x, y) = \mathbf{1}_{\{g(x)\neq y\}}$. Hence, the msiclassification error over the entire test set is given by: $E(Te) = \frac{1}{n}\Sigma_{(x,y)\in Te}\mathbf{1}_{\{g(x)\neq y\}}$.
It is obvious that the performances of the classifier also depend on the parameters. Indeed, we can think that if the parameters make the two equations "too similar", the classifier will give bad results. Now we can verify empirically this hypothesis:
For example, for the simple example of the previous section, we plot the error rate against the value of μ_1, $\mu_0 = 2$ and $\sigma = 1$ being fixed, figure (2.3)
Obviously, for $\mu_1 = 1$, the error rate is 0.67 and decreases until 0 while the value of μ_1 increases. So, we clearly see that while μ_1 is going higher and higher, the second equation become more and more "different" from the first one and thus the sample paths derived from these two equation are easier to distinguish.

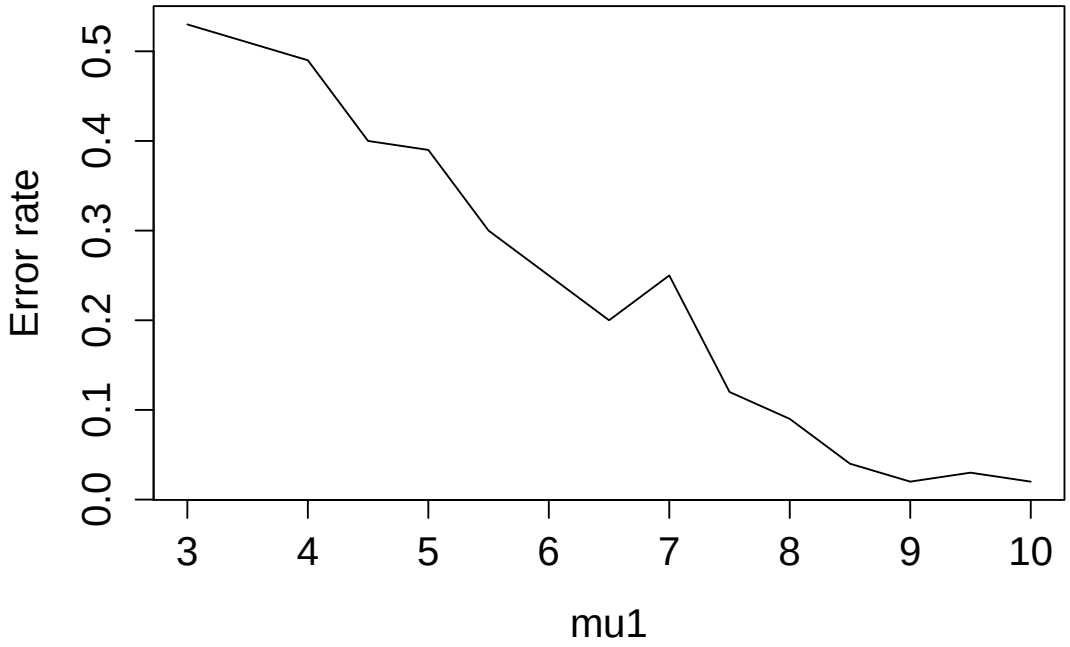

Figure 2.1: Error rate of the simple example with $N = 1000$, $n_0 = 100$, $n_1 = 100$, $\mu_0 = 2$, $\sigma = 1$ and μ_1 varying from 3 to 10

Moreover, from the graphics (2.3), we can see that the variance of the processes can worsen the error rate. In this graphic, we plot the error rate against the value of σ. For $\sigma = 1$, the error rate is 0 and increases until 0.17 while σ increases to 10. One hypothesis to explain this behavior is that even if the respective drifts of the processes are "easily distinguishable", a big variance can add too much noise and make the two processes "less distinguishable". The same behaviors are also true for the Ornstein-Uhlenbeck Process, as shown in figures (2.3) and (2.3).

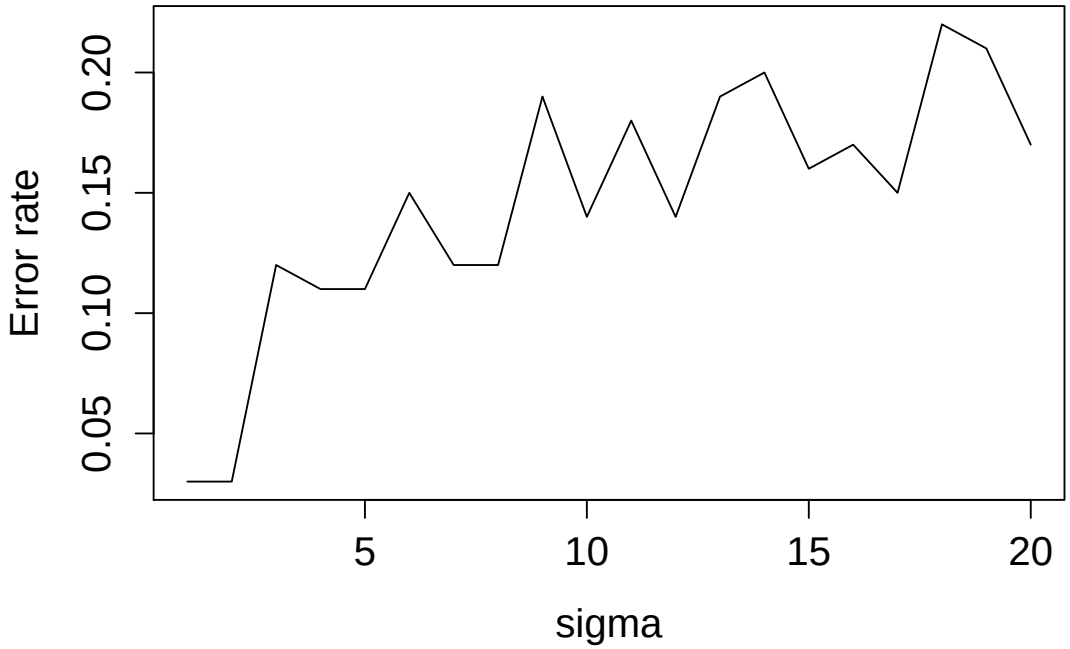

Figure 2.2: Error rate of the simple example with $N = 1000$, $n_0 = 100$, $n_1 = 100$, $\mu_0 = 2$, $\mu_1 = 20$ and σ varying from 1 to 10

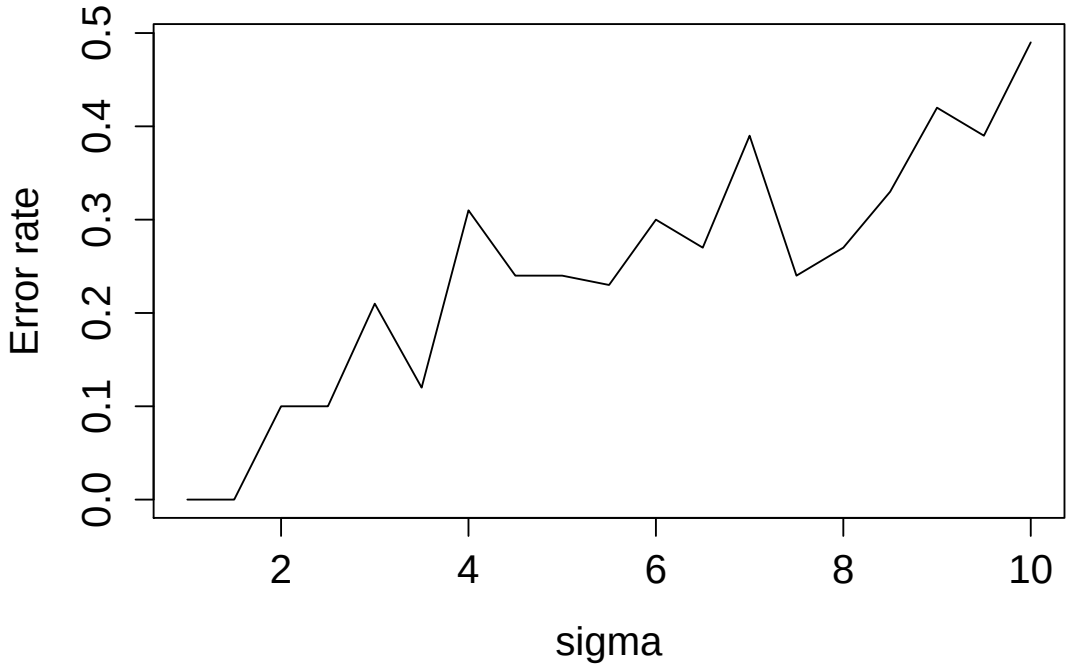

Figure 2.3: Error rate of the Ornstein-Uhlenbeck process with $N = 1000$, $n_0 = 100$, $n_1 = 100$, $\theta = 1$, $\mu_0 = 1$, $\mu_1 = 15$ and σ varying from 1 to 10

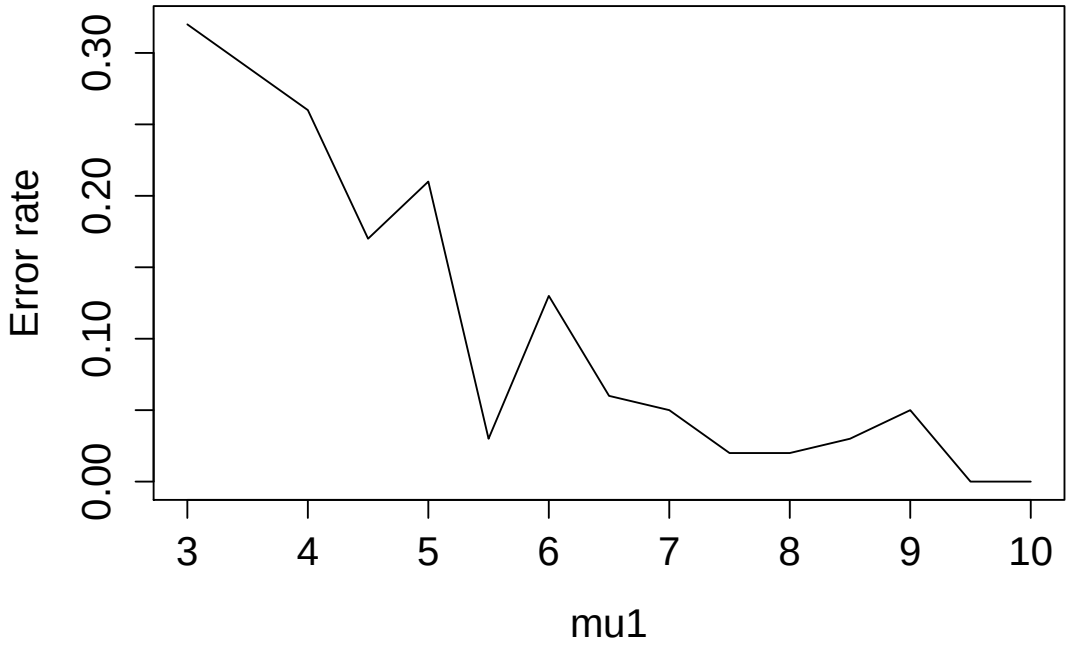

Figure 2.4: Error rate of the Ornstein-Uhlenbeck process with $N = 1000$, $n_0 = 100$, $n_1 = 100$, $theta = 1$, $\mu_0 = 1$, $\sigma = 2$ and μ_1 varying from 3 to 10

Conclusion

In this paper, we developped a way to classify diffusion paths. We started by reminding some of the most important results of the stochastic processes theory and of the bayesian clasification theory. Then, we used the paper [Cad13] to build the bayes rule for diffusion processes. In practice, as we only observe a discretization of the trajectories of the processes, we had to estimate this bayes rule. Our estimation was done by plug-in. So, we had to estimate the parameters of the stochastic differential equations. Several options were at our disposal: when the solution of the equation was explicit, we performed a maximum-likelihood estimation. Otherwise, we us the Euler-Maruyama scheme. To sump-up, we built a two-step process to classify SDE: first, we estimate its parameters by either a maximum-likelihood estimate or using the Euler-Maruyama scheme. The second step is to plug this estimation in the bayes rule of the process. Finally, we made a numerical study of this method of classification.
We also have several ideas for further work: it would be valuable to have a consitency result. Moreover, we only considered the case where the functions in the SDE were paramatric. It could be interesting to develop the non-parametric case.

Bibliography

[Cad13] B. Cadre. Supervised classification of diffusion paths. *Math. Methods Statist.*, pages 213–235, 2013.

[Den14] C. Denis. Classification in postural style based on stochasticmodeling. 2014.