Lecture notes on Stochastic Differential Equations and Applications (draft version)

Nuno M. Brites, nbrites@iseg.ulisboa.pt

November 16, 2023

Contents

Preface

This notes must be used as a secondary bibliography to clarify punctual doubts and/ or as a compiler of the fundamental concepts regarding Stochastic Differential Equations. Please do not use this as a reference book.

All errors found are the sole responsibility of the authors and must be communicated. Furthermore, comments and suggestions are welcome.

Chapter 1

Computational Simulation

The present chapter is related with the simulation of random numbers using computational tools, specially important in mathematics when treating the estimation of unknown parameters, for example.

The structure of the chapter is based on a summarisation of fundamental concepts, followed by the presentation of explained examples.

1.1 Monte Carlo simulation

Monte Carlo method is constituted by a wide range of algorithms used in the simulation of parameters for a given random variables. This method is based on the simulation of various independent and identically distributed copies of the random variable we are in search of parameters. The resulting values from the simulated random variables are in turned used to estimate the value for the desired parameter.

The most ubiquitous of these parameters to estimate is the mean, $\mu = \mathbb{E}[x]$, of a given random variable X. Equipped with the Law of large numbers, i.e. $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$, the approximation thus becomes the generation of a large number of values for X and the subsequent computation of their average $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Recall that according to the central limit theorem, given a sequence of independent and identically distributed random variables, $X_1, X_2, \cdots, X_n$, with mean μ, variance σ^2, and finite moments ($\mathbb{E}[x^n] < \infty$ for any $n \in \mathbb{N}$), then

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \xrightarrow{P} N(0,1). \tag{1.1}$$

In other words, when $n \to \infty$, the distribution function of Z_n follows asymptotically a standard normal distribution, $N(0,1)$.

Example: Consider an experience consisting in removing one ball from a basket with balls numbered from 1 to 4. What is the expected number of the removed ball?

Let X be the random variable defining the number of the ball removed. Since we have 4 balls, the probability function has value $\frac{1}{4} = 0.25$.

The real Expected Value is:

```
E_X <- 1*0.25 + 2*0.25 + 3*0.25 + 4*0.25
print(E_X)
```

```
## [1] 2.5
```

Using the Monte Carlo Method, we have the following simulated expected value:

```
sample_1 <- round(runif(5, min=1, max=4)) # Small Sample
# Round fundamental to guarantee the number is integer
# print(sample_1) # Uncomment to visualize sample
MC_E_X <- sum(sample_1) / length(sample_1)
print(MC_E_X) # Poor approximation
```

```
## [1] 2.8
```

```
sample_2 <- round(runif(20000, min=1, max=4)) # Larger sample
#print(sample_2)
MC_E_X_2 <- sum(sample_2) / length(sample_2)
print(MC_E_X_2) # CLoser to the real parameter
```

```
## [1] 2.49415
```

Monte Carlo method may be applied to other parameters, proceeding supported by a large simulated sample.

1.2 Simulation of random variables

1.2.1 Uniform random variables

Independent random numbers uniformly distributed in the interval $[0,1]$, u_n, can be generated by first selecting three positive integers $m, b, x_0 \in \mathbb{Z}^+$, where $m > b, x_0 > 0$, and then applying the algorithm:

$$\begin{aligned} x_n &\equiv b\, x_{n-1} \mod m \\ u_n &= \frac{x_n}{m}, \end{aligned}$$

for any $n \in \{1, 2, ..., m\}$.

Example:

Let $m = 7$, $b = 3$, and $x_0 = 2$, we have,

$$\begin{aligned} x_1 &\equiv 3 \times 2 \equiv 6 \mod 7, & u_1 &= 0.857 \\ x_2 &\equiv 3 \times 6 \equiv 4 \mod 7, & u_2 &= 0.571 \\ x_3 &\equiv 3 \times 4 \equiv 5 \mod 7, & u_3 &= 0.714 \\ x_4 &\equiv 3 \times 5 \equiv 1 \mod 7, & u_4 &= 0.143 \\ x_5 &\equiv 3 \times 1 \equiv 3 \mod 7, & u_5 &= 0.429 \\ x_6 &\equiv 3 \times 3 \equiv 2 \mod 7, & u_6 &= 0.286. \blacksquare \end{aligned}$$

Notice, that parameters b and m require careful selection, in order to ensure that the cycle length is m, because if b, m, and x_0 are not coprime, i.e. $\gcd(b, m), \gcd(m, x_0), \gcd(b, x_0) \neq 1$, then the aforementioned algorithm will not complete a cycle.

Example:

Let us consider $m = 14$, $b = 6$, and $x_0 = 5$, where $\gcd(m, b) = 2$, we have

$$\begin{aligned} x_1 &\equiv 6 \times 5 \equiv 2 \mod 14 \\ x_2 &\equiv 6 \times 2 \equiv 12 \mod 14 \\ x_3 &\equiv 6 \times 12 \equiv 2 \mod 14 \end{aligned}$$

The cycle is only of length 2. ■

Example:

Let us consider $m = 7$, $b = 4$, and $x_0 = 6$, where $\gcd(b, x_0) = 2$, we have

$$\begin{aligned} x_1 &\equiv 4 \times 6 \equiv 3 \mod 7 \\ x_2 &\equiv 4 \times 3 \equiv 5 \mod 7 \\ x_3 &\equiv 4 \times 5 \equiv 6 \mod 7 \end{aligned}$$

The cycle is only of length 3. ■

Hence, by the previous examples, m must preferably be prime to avoid having a greatest common divisor different from 1 with b, and b and x_0 must be coprime. Otherwise, at a certain point the method will start returning 0 (if $m = b^i \times x_0$ for some $i \in \mathbb{Z}^+$), or the cycle will be of a length lower than m, and hence repeat generated values u_n, which may ruin the approximation.

There exists a built-in function in R to generate random uniformly distributed numbers, using a strategy similar to the presented above: runif(n, min=a, max=b) returns a sample of n random numbers in the interval $[a, b]$.

Due to the fact that theses methods generate predictable random numbers when x_0, b and m are known, the obtained numbers are also called pseudorandom numbers.

1.2.2 Bernoulli and Binomial random variables

A Bernoulli trial is an experiment with two possible outcomes, success or failure, with probability p and $1 - p$, respectively, for some $p \in [0, 1]$.

$$P(X = 0) = 1 - p, \quad P(X = 1) = p, \tag{1.2}$$

A Binomial random variable is the sum of m independent Bernoulli random variables.

$$P(X = x) = \binom{m}{x} p^x (1 - p)^{m-x}, \quad \text{for any } x \in \{0, ..., m\}, \tag{1.3}$$

In both cases, the random variables count the number of successes.

One possible way to simulate a Binomial random variable is through the simulation of uniform random variables.

Example: Considering a basket with balls numbered from 1 to 4, let the success be the observation of ball number "2". How many times do we observe ball number "2"?

```
ball <- runif(10, min=0, max=1)
correct_ball <- (ball<=0.25) #Success if correct_ball <= 0.25
table(correct_ball) # TRUE represent the number of "2" observed
```

```
## correct_ball
## FALSE  TRUE
##     7     3
```

Another is using R built-in functions:

Generator: rbinom(n, size, prob); Returns n random variables of a Binomial distribution with parameter m and p.

Probability function: dbinom(x, size, prob); Returns $P(X = x)$ when *size* represents the number of trials and *prob* represents the probability of a success.

Distribution function: pbinom(x, size, prob); Returns $P(X \leq x)$ when *size* represents the number of trials and *prob* represents the probability of a success.

Quantile function: qbinom(x, size, prob); Returns a such that $P(X \leq a) = x$, when *size* represents the number of trials and *prob* represents the probability of a success.

Example:

```
rbinom(5, 10, 0.25) # 5 samples of a binomial distribution
```

```
## [1] 0 2 2 3 2
```

```
dbinom(5, 10, 0.25) # P(X = 5)
```

```
## [1] 0.0583992
```

```
pbinom(5, 10, 0.25) # P(X =< 5)
```

```
## [1] 0.9802723
```

```
qbinom(0.5, 10, 0.25) # Median
```

```
## [1] 2
```

1.2.3 Poisson random variables

A Poisson random variable is the limit of a sequence of binomial distribution with parameters m going to infinity and p_m to zero. Both expected value and variance converge to a constant, λ, referred as *rate*.

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad \text{for any } x \in \mathbb{N}, \tag{1.4}$$

Functions

Generator: rpois(n, lambda); Returns n random variables of a Poisson random variable with rate λ.

Probability function: dpois(x, lambda); Returns $P(X = x)$ where λ represents the rate.

Distribution function: ppois(x, lambda); Returns $P(X \leq x)$ where λ represents the rate.

Quantile function: qpois(x, lambda); Returns a such that $P(X \leq a) = x$, where λ represent the rate.

Example:

```
rpois(10, 1.5)
```

```
##  [1] 1 0 1 3 3 2 4 2 0 2
```

```
dpois(2, 1.5)
```

```
## [1] 0.2510214
```

```
ppois(2, 1.5)
```

```
## [1] 0.8088468
```

```
qpois(0.5, 1.5)
```

```
## [1] 1
```

A homogeneous Poisson random variable can be generated considering λ as the rate of a poisson for one single period and T as the number of periods. The next algorithm generates the number of events in each period:

1. Generate the number of events, N, in the time interval T, using rpois(1, λ*T);
2. Generate the moment when the events occur in the time interval [0,T], using runif(N, max=T).

Example

```
set.seed(123456789) # fix the results
T <- 5 # Number of periods
N_events <- rpois(1, 1.5*T) # lambda=1.5
N_events # number of events in 5 periods
```

```
## [1] 9
```

```
Instant <- sort(runif(N_events, max=5)) # "Sort" to be crescent in time
Instant# return the moments when events occur!
```

```
## [1] 1.175238 1.340850 2.266536 3.269508 3.364405 3.594946 4.495499 4.607723
## [9] 4.654993
```

```
Poisson_vector <- rep(0,T)

for (i in 1:T){
   Poisson_vector[i] <- sum(Instant<i & Instant>i-1)
}
print(Poisson_vector) # Poisson random variable
```

```
## [1] 0 2 1 3 3
```

Poisson random variables disposed on a line can be simulated using the generator of a exponential distribution, representing the time to perform a task or the duration until failure, which will be summed in a cumulative way in order to obtain the instants where the studied event occurs.

Example:

```
t <- rexp(5, rate=2)
cumsum(t)
```

```
## [1] 1.577225 2.366788 2.735835 4.189652 4.239626
```

1.2.4 Exponential random variables

Exponential distribution is commonly associated to the Poisson distribution but its scope is larger. Exponential random variables are used to model waiting or failure times, or the time to complete a task. With probability density function $f(t)$ and distribution function $F(t)$, respectively, given by

$$f(t) = \lambda e^{-\lambda t}, \tag{1.5}$$
$$F(t) = P(T \leq t) = 1 - e^{-\lambda t}. \tag{1.6}$$

The parameter that defines the distribution is again λ.

Functions

Generator: rexp(n, lambda); Returns n exponential random variables with rate λ.

Probability density function: dexp(t, lambda); Returns $P(T = t)$ where λ represents the rate, $f(t) = \lambda e^{-\lambda t}$

Distribution function: pexp(t, lambda); Returns $P(T \leq t)$ where λ represents the rate, $F(t) = 1 - e^{-\lambda t}$

Quantile function: qexp(x, lambda); Returns a such that $P(T \leq a) = t$, where λ represent the rate.

Simulation of exponential random variables can be done by inverting function F,

$$F(T) = U \Leftrightarrow T = F^{-1}(U) \Leftrightarrow T = -\frac{\log(1-U)}{\lambda}, \tag{1.7}$$

then generating values for a uniform distribution $U([0,1])$ and finally imputing those values on the expression for F^{-1}.

1.2.5 Normal random variables

Normal distribution is among the most important distributions and it's widely applied. Mean, or expected value, and variance characterise the normal distribution. When $\mu = 0$ and $\sigma^2 = 1$, it is said that the random variable follow a standard normal distribution. Central limit theorem contributes to the relevance of this distribution.

Functions

Generator: rnorm(n, mean=0, sd=1); Returns n random variables normally distributed with mean 0 and variance 1.

Probability density function: dnorm(n, mean=0, sd=1); Returns $P(X = x)$ where X is a random variable normally distributed with mean 0 and variance 1.

Distribution function: pnorm(n, mean=0, sd=1); Returns $P(X \leq x)$ where X is a random variable normally distributed with mean 0 and variance 1.

Quantile function: qnorm(n, mean=0, sd=1); Returns a such that $P(X \leq a) = x$, where X is a random variable normally distributed with mean 0 and variance 1.

1.2.6 Other built-in random variables

R software has many standard distributions following the convention of *rxxx* to generate random variables, *dxxx* to obtain the probability (density) function, *pxxx* to the cumulative distribution function and *qxxx* to obtain the quantiles. To visualise all these distributions, please consult Braun and Murdoch (2021).

1.3 Multivariate random number generation

When the problem to model is based in more than one random variable, which may happen frequently when working on real problems, it's required to simulate random variables with different distributions.

One possible technique is sequential conditional generation, where random variables are simulated conditioned by the previous. Considering $X = (X_1, X_2, \cdots, X_n)$, and knowing the distributions of each random variable, the first step is to simulate X_1, then simulate X_2 conditioned on X_1, X_3 conditioned on X_1 and X_2 and so on until X_n, which is conditioned by $X_1, \cdots, X_{n-1}$. This method may not be feasible when distributions are not known.

Even though R has some multivariate distributions built in, in general it is necessary to construct our own distribution.

1.4 Markov chain simulation

Markov chains are sequences of random variables where, at time n, the random variable depends only on the previous time instant, i.e. $n-1$. Markov chains are used in modeling systems with short memory, like stock market or in genome passage from parents to offspring.

When there are $1, \cdots, n$ states, the transition probabilities are represented by a $n \times n$ matrix P, where i represent the current state, j a possible future state and $P(X_{n+1} = j \mid X_n = i) = P_{i,j}$ represents the probability of moving from state i to state j.

Two things to note is that the sum of the rows of a transition probability matrix adds to one, i.e. $\sum_{j=1}^{n} P_{i,j} = \sum_{j=1}^{n} P(X_{n+1} = j \mid X_n = i) = 1$, and since time increments are independent, the matrix will be invariant over time, hence the transition matrix remains unchanged.

An important property of a Markov chain, is that they possess an invariant distribution, that is a distribution $\pi_i = [P(X = i)]_i$, such that,

$$\pi = \pi P.$$

Notice that π is simply an eigenvector of P^T, such that, $\sum_{i=1}^{n} \pi_i = 1$. An even more interesting property is that for some transition probability matrices P, X_n will converge to the invariant distribution π.

Example: Let's consider a two state case: weather is sunny (1) or rainy (2). The transition probability matrix associated to this trial is:

```
MC_weather <- matrix( c(0.9, 0.1, 0.5, 0.5), nrow=2, ncol = 2, byrow = TRUE)
MC_weather
```

```
##      [,1] [,2]
## [1,]  0.9  0.1
## [2,]  0.5  0.5
```

We can extract from the matrix that the probability of having a sunny day tomorrow if today was sunny is 90%. Oppositely, the probability of raining after a sunny day is 10%. If today was raining, the probability of having a rainny or sunny day tomorrow is equal to 50%.

1.5 Monte Carlo integration

Integration through Monte Carlo method is useful to approximate the value of an integral resorting to the simulation of various points uniformly distributed along the interval of integration, which are then evaluated on the integrating function to finally take the estimate of the integral as the mean of that last points times the length of that interval.

The previously explained strategy uses the law of large numbers to support the use of the sample mean and is based on the formula of the expected value to obtain its final expression:

$$\mathbb{E}(g(U_i)) = \int_a^b g(U_i)\,\frac{1}{b-a}\,du \Leftrightarrow \int_a^b g(U_i)\,du = (b-a)\,\mathbb{E}(g(U_i))$$

Note that we are using uniform random variables, so the density distribution is equal to $\frac{1}{b-a}$.

Example: Considering the integral $\int_0^3 x^2\,dx$, with value 9. Computing this integral with Monte Carlo we have:

```
set.seed(1234567656) # Fixing results
aux <- runif(100, min=0, max=3)
```

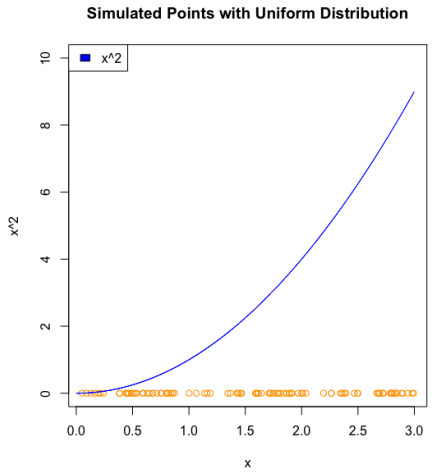


Figure 1.1: Simulation of points with an uniform distribution.

```
points <- aux^2
```

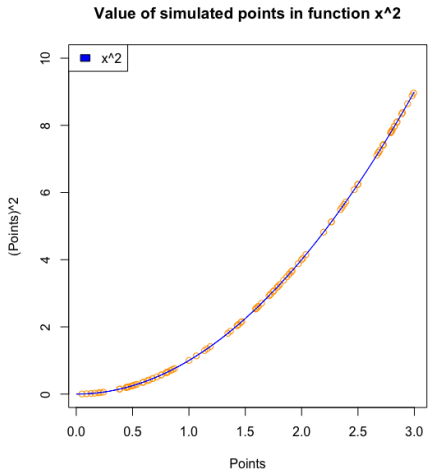


Figure 1.2: Valued of simulated points in the target function.

```
integral_approx <- mean(points)*(3-0)
integral_approx
```

```
## [1] 9.918777
```

```
# or: (augmented sample)
set.seed(1234567656) # Fixing results
aux <- runif(10000, min=0, max=3)
points <- aux^2
integral_approx <- mean(points)*(3-0)
integral_approx
```

```
## [1] 9.013217
```

As one can see, the approximation of the integral through Monte Carlo provides a better result as the number of simulated points increase.

When dealing with a multiple integration on m variables, $\int_{a_1}^{b_1} \cdots \int_{a_m}^{b_m} g(u_1, \cdots, u_m)\, du_1 \cdots du_m$, the strategy is similar but with the adjustment of simulating as many uniform random variables, $U_i^j \sim U[a_j, b_j]$ for any $i \in \{1, \cdots, n\}$ and $j \in \{1, \cdots, m\}$, as the number of variables in the integral and the length of the integral is substituted by the product of the length of the integrals.

$$\mathbb{E}(g(U_i^1, \cdots, U_i^m)) = \int_{a_1}^{b_1} \cdots \int_{a_m}^{b_m} g(u_1, \cdots, u_m)\, \frac{du_1}{b_1 - a_1} \cdots \frac{du_m}{b_m - a_m} \tag{1.8}$$

$$\Leftrightarrow \int_{a_1}^{b_1} \cdots \int_{a_m}^{b_m} g(u_1, \cdots, u_m)\, du_1 \cdots du_m = \prod_{j=1}^{m} (b_j - a_j)\, \mathbb{E}(g(U_i^1, \cdots, U_i^m)) \tag{1.9}$$

Example: Considering the integral $\int_3^{10} \int_1^7 sin(x - y)\, dx\, dy$, with approximated value 0.1185. Computing this integral with Monte Carlo we have:

```
x_aux <- runif(10000000, min=1, max=7)
y_aux <- runif(10000000, min=3, max=10)
points_2 <- sin(x_aux-y_aux)
integral_approx_2 <- mean(points_2)*(10-3)*(7-1)
integral_approx_2
```

```
## [1] 0.1129121
```

The Uniform distribution is not the unique distribution that can be generated to compute an integral. If, for instance, we consider a random variable X that has a distribution function f then $\mathbb{E}\left[\frac{g(X)}{f(X)}\right] = \int \left[\frac{g(X)}{f(X)}\right] f(x)\, dx = \int g(x)\, dx$. Hence we can compute $\int g(x)\, dx$ by generating a very large number of values for X, inputting them in $\frac{g(X)}{f(X)}$ and calculating the average of the resulting values.

1.6 Advanced simulation methods

Up to now we have discussed particular, and well-known, distributions. However, there exists two simulation methods commonly used: rejection sampling and importance sampling.

Rejection sampling from a set S consists on the simulation of values for an appropriate random variable X, such that $S \subset \{X(\omega) : \omega \in \Omega\}$, and selecting the simulated values contained in S, discarding the remaining ones.

Importance sampling is a technique to generate randomly both the sample and the weights,aiming to approximate the expected values to the wanted density function. In this technique we start by choosing a convenient density distribution, $f(x)$, to obtain a sample, then compute the weights dividing the target distribution, $g(x)$, by the selected distribution in order to, finally, approximate the expectation of a function $h(X)$ where X has the probability distribution function $g(x)$ using averages of $h(x_i)$ weighted by $w_i = \frac{g(x_i)}{f(x_i)}$.

Chapter 2

Mathematical Foundations

This chapter covers the fundamental contents required to the topic of Stochastic Differential Equations, which is a subject that requires a broad knowledge and many times is necessary to review or address any deficiency in background. In this sense, our intention is to reproduce a succinct yet complete summary comprehending the most important topics.

2.1 Stochastic Processes

A stochastic process is a collection of random variables $\{X_t, t \in T\}$ defined on a given probability space $(\Omega, \mathcal{A}, \mathcal{P})$, where the set of values assumed by X_t, called states, form the state space S, for each t in a parameter space T.

Therefore, a stochastic process is a real-valued random function $X_t : T \to S$. For a fixed $t \in T$, the map $w \in \Omega \to X_t(w)$ is a random variable. For each $w \in \Omega$ fixed, the map $t \in T \to X_t(w)$ is a real-valued function called the trajectory, realization or sample path of the process.

When the parameter space is designated as the time space, T can be either discrete, $T = \mathbb{N}$, or continuous, $T = [0, +\infty)$. State space S can be a countable set, $S = \mathbb{N}$ for example, or an interval in $\mathbb{R}$, $S = [a, b), a, b \in \mathbb{R}$.

Two stochastic processes, $\{X_t\}_{t\in T}$ and $\{Y_t\}_{t\in T}$ defined on the same probability space $(\Omega, \mathcal{A}, \mathcal{P})$, are said to be stochastically equivalent or indistinguishable if

$$X_t = Y_t \ a.s., \forall\, t \in T, \text{or equivalently, } P\Big(X_t(w) = Y_t(w) \mid w\Big) = 1.$$

The process $\{X_t\}$ is a version or modification of $\{Y_t\}$, and vice versa, and their finite-dimensional distributions coincide.

The finite-dimensional distribution of a stochastic process $\{X_t\}_{t\in T}$ defined on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$ is the joint distribution of the random vector $(X(t_1), \cdots, X(t_n))$, given by

$$F_{t_1,\cdots,t_n}(x_1, \cdots, x_n) = P(X_{t_1} \leq x_1, \cdots, X_{t_n} \leq x_n), \tag{2.1}$$

with $t \in T$, $x_i \in \mathbb{R}$, $n \geq 1$.

The family of distributions that define a stochastic process are important to comprehend its behavior and to determine the $P((X(t_1), \cdots, X(t_n)) \in A$, with $n \in \mathbb{N}_+$, $(t_1, \cdots, t_n) \in T^n$ and $A \subseteq S^n$ a measurable set.

Sometimes it is convenient to construct a stochastic process with a specific finite-dimensional distribution. When this construction is not readily obtained from a sequence of random variables $\{X_i, i = 1, \cdots, n\}$, one might verify the Kolmogorov Existence Theorem.

Theorem 2.1 (Kolmogorov Existence Theorem)**.** *Let P to be a family of finite-dimensional distributions satisfying the following consistency conditions:*

1. *If $\{\pi(1), \cdots, \pi(n)\}$ is a permutation of the numbers $1, \cdots, n$, then for arbitrary time points $t_1, \cdots, t_n \in T$ and $n \geq 1$, $F_{t_{\pi(1)}, \cdots, t_{\pi(n)}}(x_{\pi(1)}, \cdots, x_{\pi(n)}) = F_{t_1, \cdots, t_n}(x_1, \cdots, x_n)$;*
2. *For $m < n$ and arbitrary $t_{m+1}, \cdots, t_n \in T$, $F_{t_1, \cdots, t_m, t_{m+1}, \cdots, t_n}(x_1, \cdots, x_m, \infty, \cdots, \infty) = F_{t_1, \cdots, t_m}(x_1, \cdots, x_m)$.*

Then, there exists a probability space $(\Omega, \mathcal{A}, \mathcal{P})$ and a stochastic process $\{X_t\}_{t \in T}$ defined on a that space such that P is the collection of finite-dimensional distributions of $\{X_t\}_{t \in T}$.

A stochastic process $\{X_t\}_{t \in T}$ defined on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$ has the following characteristics:

1. **Continuous**, if all its paths are continuous, i.e., for almost all $w \in \Omega$ the mapping $t \in T \to X_t(w)$ is continuous;
2. **Integrable**, if X_t is an integrable random variable $\forall t \geq 0$;
3. **Measurable**, if the mapping $(t, w) \to X_t(w)$ is measurable with respect to the σ-algebra $B(T) \times A$, with $B(T)$ the family of all Borel subsets of T;
4. **Square-integrable**, if $E(\mid X_t \mid^2) < \infty, \forall t \geq 0$;
5. **Increasing**, if for any $s < t$, $X_s \leq X_t$;
6. **(Strictly) Stationary**, if the finite-dimensional distributions are invariant under time displacements, i.e., for $t_i, t_{i+h} \in [t_o, n]$ and $h \in \mathbb{R}$, $F_{t_{1+h}, \cdots, t_{n+h}}(x_1, \cdots, x_n) = F_{t_1, \cdots, t_n}(x_1, \cdots, x_n)$;
7. **Gaussian**, if finite-dimensional distributions are normally distributed;
8. **Continuous in probability**, if $P(\mid X_t - X_s \mid > \varepsilon) \to 0$, with any $s \in T$, $s \to t$ and $\varepsilon > 0$;
9. **Quadratic variation process**, if $X_t(w) : \Omega \to \mathbb{R}$ is continuous and the second is defined as

$$(X, X)_t^2(w) = lim_{\Delta t_k \to 0} \sum_{t_k \leq t} \mid X_{t_{k+1}}(w) - X_{t_k}(w) \mid^2,$$

   with $0 < t_1 < t_2 < \cdots < t_n = t$ and $\Delta t_k = t_{k+1} - t_k$;
10. **Markov**, if the process has no memory in the sense that the change at time t is only determined by the value of the process in that specific moment, and not by its past values (Markov Property).

A σ-algebra on the probability space $(\Omega, \mathcal{A}, \mathcal{P})$ is defined as $\mathcal{A}_t = \sigma(X_s, s \in T, s \leq t)$ and comprises all the information regarding the process available at time $t \in T$. The σ-algebra $\mathcal{A}_\infty = \sigma(\cup_{t \in T} \mathcal{A}_t)$ gathers all information of all time points t and, for any $t \in T$, $\mathcal{A}_t \subseteq \mathcal{A}_\infty$.

Given a probability space $(\Omega, \mathcal{A}, \mathcal{P})$, a filtration $\mathcal{J} = \{\mathcal{A}_t\}_{t \in T}$ is an increasing family of sub-σ-algebras of $\mathcal{A}$, that is, for $s, t \in T$ with $0 \leq s \leq t \leq \infty$, $\mathcal{A}_s \subseteq \mathcal{A}_t \subseteq \mathcal{A}$, which implies that $\mathcal{A}_t$ contains the same or more information about the process than $\mathcal{A}_s$, both bounded above by $\mathcal{A}$, and the information of $\mathcal{A}_s$ is a subset of the $\mathcal{A}_t$ information.

A probability space endowed with a filtration is called a filtered probability space and is denoted as $(\Omega, \mathcal{A}, \mathcal{J}, \mathcal{P})$. A stochastic process $\{X_t\}_{t \in T}$ is adapted to the filtration $\mathcal{J}$ if, for each $t \in T$, $\sigma(X_s, s \in T, s \leq T) \subseteq \mathcal{A}_t$, or, equivalently, it is $\mathcal{A}_t$-measurable, that is, the information held at time t by $\mathcal{A}_t$ is enough to determine the value of X_t.

A nonanticipating process is a process that does not present the capability to anticipate the future, that is, the value of X_t depends on the previous behavior of the system, without losing information, and it is only known if $\mathcal{A}_t$ is also known. An important property, denominated tower property, states that $E(E(X_t \mid \mathcal{A}_s) \mid \mathcal{A}_r) = E(X_t \mid \mathcal{A}_r)$ with $r < s < t$, since $\mathcal{A}_r$ comprises less information than $\mathcal{A}_s$ ($\mathcal{A}_r \subset \mathcal{A}_s$).

Given a stochastic process $\{X_t\}_{t\in T}$, one can construct its natural filtration denoted as $\mathcal{F}_t^X = \{\mathcal{A}_t^X\}_{t\in T}$, which consists in a family of σ-algebras generated by the values of process up to time t, generating the smallest filtration to which $\{X_t\}_{t\in T}$ is adapted.

Filtrations are of special importance because they model the information produced by a stochastic process over time. Thus, one can check if a given event has occurred up to the time under analysis, that is, if $A \in \sigma(X_s, s \leq t)$ with $\sigma(X_s, s \leq t)$ the natural filtration of $\{X_t\}_{t\in T}$.

A probability space $(\Omega, \mathcal{A}, \mathcal{P})$ is complete with respect to a measure μ if all sets of measure zero belong to the σ-algebra $\mathcal{A}$.

A filtration $\mathcal{J} = \{\mathcal{A}_t\}_{t\in T}$ is complete if the probability space $(\Omega, \mathcal{A}, \mathcal{P})$ is complete and if the collection of all subsets with zero probability, $\mathcal{A}_0$, contains all the sets with null probability- $A \in \mathcal{A}$, $P(A) = 0$ then $A \in \mathcal{A}_0$.

A filtration $\mathcal{J} = \{\mathcal{A}_t\}_{t\in T}$ is right-continuous if $\mathcal{A}_t = \mathcal{A}_{t+}$, with $\mathcal{A}_{t+} = \cap\{A_s, s \in T, s > t\}\ \forall\ t \in T$.

A continuous stochastic process adapted to a filtration is also progressively measurable with respect to that filtration.

2.2 Martingales

Let $(\Omega, \mathcal{A}, \mathcal{P})$ be a probability space and $\mathcal{J} = \{\mathcal{A}_t\}_{t\in T}$ a filtration on that probability space. A martingale defined on a filtered probability space $(\Omega, \mathcal{A}, \mathcal{J}, \mathcal{P})$ is a sequence of random variables $\{X_t\}_{t\in T}$ on the probability space that are adapted to the filtration. Hence, if $\{X_t\}_{t\in T}$ satisfies

1. X_t is $\mathcal{A}_t$-measurable (X_t's are adapted to the filtration)
2. $E(\mid X_t \mid) < \infty$ (X_t's are integrable)
3. $E(X_t \mid \mathcal{A}_s) = X_s,\ \forall\ 1 \leq s < t < \infty$ (martingale property)

is discrete-time martingale.

The martingale property states that the expected value of X_{t+1} given the information available up to time t is X_t. The best next forecast for the fortune of a gambler on a fair game after t^{th} trials is the current fortune at time t, for example. If the "=" sign in the martingale property is substituted by a "$\leq$" ("$\geq$"), the process $\{X_t\}_{t\in T}$ is called supermartingale (submartingale).

One interest thing that worth mentioning regarding martingales is that one can construct new martingales based on an old one. The general approach is to consider a new process $\{\tilde{X}_t, 0 \leq t < \infty\}$ defined as $\tilde{X}_0 = X_0$ and $\tilde{X}_t = X_0 + A_1(X_1 - X_0) + A_2(X_2 - X_1) + \cdots + A_n(X_t - X_{t-1}),\ t \geq 1$. The process $\tilde{X}_t$ is the martingale transform of $\{X_t\}_{t\in T}$ by $\{A_t\}_{t\in T}$. The following theorem states the conditions appropriately.

Theorem 2.2 (Martingale Transform Theorem)**.** *Let $\{X_t\}_{t\in T}$ be a martingale adapted to the filtration $\mathcal{J} = \{\mathcal{A}_t\}_{t\in T}$ with $\{A_t, 1 \leq t < \infty\}$ a sequence of bounded random variables that are nonantecipating with respect to $\mathcal{J}$. Then, the sequence of martingale transforms $\{\tilde{X}_t\}_{t\in T}$ is itself a martingale adapted to $\mathcal{J}$.*

A stopping time for the filtration $\mathcal{J} = \{\mathcal{A}_t\}_{t\in T}$ is a random variable $\tau : \Omega \to T = \mathcal{N} \cup \{+\infty\}$ where $\{w \mid \tau(w) \leq t\} \in \mathcal{A}_t, \forall\ 0 \leq t < \infty$, that is, a stopping time constitutes a rule to stop the process if the criterion τ was verified based on the information about the filtration $\mathcal{J}$. Stopping times are related with martingales by the noticeable fact that a stopped process is a martingale adapted to $\mathcal{J}$ if the original process was also a martingale adapted to $\mathcal{J}$.

The study of the behavior of a process is interesting to check if this process converges to a certain threshold as time evolves. The following theorems cover these theoretical aspects.

Theorem 2.3. *If a process $\{X_t\}_{t=1}^{\infty}$ is a L^1-bounded martingale on a probability space $(\Omega, \mathcal{A}, P)$, i.e., if $sup_{k\in K} E(\mid X_k \mid) < +\infty$, then there exists an integrable random variable Z such that $lim_{t\to+\infty} X_t = Z$ a.s.*

Convergence in L^1 can be proved using uniform integrability, verifying if for a sequence of random variables $\{X_t\}_{t=1}^{\infty}$, either 1 or 2 verifies.

1. $lim_{m\to\infty}(sup_t E(|X_t|; |X_t| > m) = lim_{m\to\infty}(sup_t \int_{\{|X_t|>m\}} |X_t| dP) = 0$.
2. $\{X_t\}_{t=1}^{\infty}$ is bounded on $L^1(\Omega, \mathcal{A}, P)$ and for all $\varepsilon > 0$ there exists a $\delta > 0$ such that, for an event $A \in \mathcal{A}$, $P(A) < \delta$ implies $\int_A |X_t|\, dP < \delta \; \forall\, t$.

Theorem 2.4. *Let $\{X_t\}_{t=1}^{\infty}$ be a martingale on the probability space $(\Omega, \mathcal{A}, P)$ adapted to the filtration $\{\mathcal{A}\}_{t=1}^{\infty}$. If $\{X_t\}_{t=1}^{\infty}$ is uniformly integrable, then (1) exists an integrable r.v. Z on $(\Omega, \mathcal{A}, P)$ such that $X_t \to Z$ a.s. in $L^1(\Omega, \mathcal{A}, P)$ as $t \to \infty$ and (2) $X_m = E(Z \mid \mathcal{A}_m)$a.s. in $(\Omega, \mathcal{A}, P)$.*

Theorem 2.5. *If $(\Omega, \mathcal{A}_0, P)$, $X \in L^1$ and $\mathcal{A}_t$ are arbitrary σ-algebras with $\mathcal{A}_0$, then the r.v. $E(X \mid \mathcal{A}_\infty)$ are uniformly integrable.*

Theorem 2.6 (Dominated Convergence Theorem for Conditional Expectations)**.** *Suppose $\{X_t\}_{t=1}^{\infty}$ is a martingale on a probability space $(\Omega, \mathcal{A}, P)$. Let $X_t \to X$ a.s. and $|X_t| \leq Z \; \forall t$, with $E(Z) < +\infty$. If $\mathcal{A}_t \uparrow \mathcal{A}_\infty$, then $E(X_t \mid \mathcal{A}_t) \to E(X \mid \mathcal{A}_\infty)$ a.s.*

A martingale $\{X_t\}_{t\geq 0}$ is L^2-bounded if $E(|X_t|^2) < B$, $B \in \mathcal{R}$, for all $t \geq 1$.

Theorem 2.7 (L^2-bounded Martingale Convergence Theorem)**.** *Let $\{X_t\}_{t\geq 0}$ be a L^2-bounded martingale. Then there exists a r.v. X with $E(|X_t|^2) < B$, $B \in \mathcal{R}$, such that $P(lim_{t\to\infty} X_t = X) = 1$ and $lim_{t\to\infty} \|(X_t - X\|_2 = 0$.*

Let´s now extend the previous concepts to the case of continuous-time martingales. Through the use of the correct conversion, it's possible to generate that results from discrete-time martingales.

Given a stochastic process $\{X_t\}_{t\geq 0}$ adapted to the filtration $\{\mathcal{A}_t\}_{t\geq 0}$, defined on the probability space $(\Omega, \mathcal{A}, P)$, it is a continuous-time martingale if $E(X_t \mid \mathcal{A}_s) = X_s \; \forall\, 0 \leq s \leq t$.

Considering $\{X_t\}_{t\geq 0}$ a continuous-time martingale adapted to the filtration $\{\mathcal{A}_t\}_{t\geq 0}$ and $\{X_{t_n}\}_{n=1}^{\infty}$ uniformly integrable, with $\{t_n\}_{n=1}^{\infty}$ a strictly increasing sequence of real numbers verifying $t_n \to \infty$ as $n \to \infty$. Hence, by theorem 2.4, exists a r.v. Z such that $X_{t_n} \to Z$ a.s. as $n \to \infty$ and $E(Z \mid \mathcal{A}_{t_n}) = X_{t_n}$. Furthermore, $E(Z \mid \mathcal{A}_t) = E(E(Z \mid \mathcal{A}_{t_n}) \mid \mathcal{A}_t) = E(X_{t_n} \mid \mathcal{A}_t) = X_t, \; \forall\, t_n > t > 0$.

The random variable $\tau : \Omega \to \mathcal{R} \cup \{+\infty\}$ is a stopping time adpated to the filtration $\{\mathcal{A}_t\}_{t\geq 0}$ since $\{w \in \Omega \mid \tau(w) \leq t\} \in \mathcal{A}_t, \; \forall t \geq 0$.

Theorem 2.8 (Doob's Continuous-Time Stopping Theorem)**.** *Let $\{X_t\}_{t\geq 0}$ be a continuous martingale adapted to the filtration $\{\mathcal{A}_t\}_{t\geq 0}$ that is right-continuous and $\mathcal{A}_0$ contains all P-null sets. If τ is a stopping time for $\mathcal{J}$, then the process $\{Y_t\}_{t\geq 0} = \{X_{min(\tau,t)}\}_{t\geq 0}$ is also a continuous martingale adapted to $\mathcal{J}$.*

The convergence aspect of a continuous time process requires the definition of a *Càdlàg* process (from the french continue à droite, limite à gauche), i.e., a process that is right continuous, if for $w \in \Omega$ the function $X_t(w)$ is right continuous, with left limits, if for $w \in \Omega$ the $lim_{s\uparrow t} X_s(w)$ exists and is finite for $t \geq 0$.

Theorem 2.9 (Doob's Convergence Theorem)**.** *Let $\mathcal{J} = \{\mathcal{A}_t\}_{t\geq 0}$ be a filtration of the probability space $(\Omega, \mathcal{A}, P)$ and let $\{X_t\}_{t\geq 0}$ be a martingale with respect to the filtration $\mathcal{J}$ whose paths are right-continuous and left limit. Then, the following are equivalent:*

1. *$\{X_t\}_{t\geq 0}$ converges in L^1 when $t \to \infty$;*
2. *As $t \to \infty$, $\{X_t\}_{t\geq 0}$ converges a.s. to an integrable and $\mathcal{A}_t$-measurable r.v. Z that satifies $X_t = E(Z \mid \mathcal{A}_t)$, $t \geq 0$;*
3. *$\{X_t\}_{t\geq 0}$ is uniformly integrable.*

2.3 Path Regularity of Stochastic Processes

A function $f : \mathcal{R}_+ \to \mathcal{R}$ is Hölder with exponent γ if there exists constants $c, \gamma > 0$ such that $|f(t) - f(s)| \leq c|t - s|^\gamma$ with $s, t \in \mathcal{R}_+$. Hölder continuity allows the determination of the rate of convergence: $|f(t) - f(s)| \to 0$ at least as fast as $|t - s|^\gamma \to 0$.

The following theorem is related to the regularity of the paths of a stochastic process.

Theorem 2.10 (Kolmogorov Continuity Theorem). *For $\alpha, \varepsilon, c > 0$, if the stochastic process $\{X_t\}_{t\in[0,1]}$ defined on the probability space $(\Omega, \mathcal{A}, P)$ satisfies $E(|X_t - X_s|^\alpha) \leq c|t - s|^{1+\varepsilon}$ for $s, t \in [0, 1]$ then there exists a modification of $\{X_t\}_{t\in[0,1]}$ that is a continuous process and whose paths are γ-Hölder for every $\gamma \in [0, \frac{\gamma}{\alpha}]$.*

Regarding martingales, their regular versions depend on the regularity properties of the associated filtration, thus some conditions on the filtration are mandatory. Thereon, the following theorem guarantees the existence of modified martingales with regular paths.

Theorem 2.11 (Doob's Regularization Theorem). *Let $(\Omega, \mathcal{A}, \{\mathcal{A}_t\}_{t\geq 0}, P)$ be a filtered probability space that satisfies:*

1. *For $A \in \mathcal{A}$ with $P(A) = 0$, then every subset of A is in $\mathcal{A}_0$;*
2. *the filtration $\{\mathcal{A}_t\}_{t\geq 0}$ is right-continuous ($\mathcal{A}_t \cap \mathcal{A}_{t+\varepsilon}\ \forall\ t \geq 0, \varepsilon > 0$).*

Let $\{X_t\}_{t\geq 0}$ be a supermartingale with respect to the filtration $\mathcal{J} = \{\mathcal{A}_t\}_{t\geq 0}$. Assuming the function $t \mapsto E(X_t)$ is right-continuous, then there exists a modified process $\{\tilde{X}_t\}_{t\geq 0}$ of $\{X_t\}_{t\geq 0}$ with the following properties:

1. *$\{\tilde{X}_t\}_{t\geq 0}$ is adapted to the filtration $\mathcal{J}$;*
2. *The paths of $\{\tilde{X}_t\}_{t\geq 0}$ are locally bounded, right-continuous and left limited;*
3. *$\{\tilde{X}_t\}_{t\geq 0}$ is a supermartingale with respect to the filtration $\mathcal{J}$.*

2.4 Markov Processes

A stochastic process $\{X_t\}_{t\geq 0}$ defined on $(\Omega, \mathcal{A}, P)$ provided with the filtration $\mathcal{A}_t = \sigma(X_s, 0 \leq s < t)$ that is real-valued and adapted is a Markov process if

$$P(X_t \in B \mid \mathcal{A}_s) = P(X_t \in B \mid X_s),\ \forall\ 0 \leq s < t < \infty,\ B \in \mathcal{R},\ \text{(Markov property)}. \tag{2.2}$$

From the Markov property is possible to conclude that, given the present value, the future value of a Markov process is independent of the past values, which is the reason why Markov processes are classified as "memoryless". Moreover, from the definition is possible to conclude that any process with independent increments, either discrete or continuous, is a Markov process.

A remark to the fact that 2.2 is defined for a continuous state space, being the process denominated as a Markov Process. However, it is possible to adjust the definition for the case of a discrete state space, obtaining a process called Markov Chain.

The transition probability of a process is denoted by $P(t, B|s, x) = P(X_t \in B|X_s = x)$ and represents the probability of a process that is at position x at time s to move to a state in B at time t. Transition probabilities of a Markov process has the following properties:

1. $P(t, \cdot|s, x)$ is a probability measure on the family of Borel sets $\mathcal{B}$ for every $0 \leq s < t < \infty$ and $x \in \mathcal{R}$;
2. $P(t, B|s, \cdot)$ is Borel measurable for every $0 \leq s < t < \infty$ and $B \in \mathcal{B}$;
3. $P(t, B|s, x) = \int_R P(t, B|r, y)P(r, dy|s, x)$ for $s < r < t$, $x \in \mathcal{R}$ and $B \in \mathcal{B}$ (Chapman-Kolmogorov equation).

To understand property 3, let´s divide the transition from x to a state in B in two steps and consider an intermediate point at moment r. More steps are allowed if it is intended to, but when aiming to some intuition two steps are enough.

The process moves firstly from x to y (at moment r), and then from y to a state in B. The Markov property says that past information should be disregarded and considered only the present value of the process, so the useful information at time r is that the process is in state y.

The absence of memory that characterize this process impose that transitions of moments $s \to r$ and $r \to t$ are independent and the probability of moving from x to a state in B is the product of the probabilities of the intermediate movements.

Since y can assume any value, through the total probability theorem, this transition probability from x to a state in B verified between moments s and t should consider all possible trajectories for each different y, which is the reason for the presence of the integral.

The transition probabilities are stationary if

$$P(X_{t+\tau} \in B | X_{s+\tau} = x) = P(X_t \in B | X_s = x),\ \tau\ constant,\ s < t \tag{2.3}$$

is satisfied. That is, the initial and final moments are no longer relevant but, instead, the critical point is the duration of the interval, along with states x and B. A Markov process is termed homogeneous (in time) when their transition probabilities are stationary.

A homogeneous Markov process satisfies the strong Markov property if, for any $\tau \geq 0$ and B a Borel set,

$$P(X_{t+\tau} \in B \mid \mathcal{A}_t) = P(X_\tau \in B \mid X_0) \tag{2.4}$$

holds for every Markov instants of time. An equivalent way of defining the strong Markov property presented in 2.4 is through the verification of

$$E[h(X_{t+\tau}) \mid \mathcal{A}_t] = E[h(X_\tau) \mid X_0], \tag{2.5}$$

for all Borel-measurable bounded functions h. If the strong Markov property is verified, a homogeneous Markov process is said to be a strong Markov process. Please note that a strong Markov process is also Markov process, but a Markov process is not always a strong Markov process.

2.5 Brownian motion

A stochastic process $\{W_t\}_{t \geq 0}$ defined on a probability space $(\Omega, \mathcal{A}, P)$ is a standard Brownian Motion if the following properties are satisfied:

1. the process begins at 0, i.e., $t_0 = 0$ or $P(W_0 = 0) = 1$;
2. for $0 = t_0 \leq t_1 \leq \cdots \leq t_n$, the increments $W(t_1), W(t_2) - W(t_1), \cdots, W(t_n) - W(t_{n-1})$ are independent random variables;
3. for $0 < s < t$, the increments $W_t - W_s \sim N(0, t - s)$.

Standard Brownian Motions are also called Wiener Processes, due to the contribution of Norbert Wiener on the mathematical definition.

Property (i) is a convention. Property (ii) translates the absence of memory of a standard Brownian motion, since the increment verified on the time interval $[t_k, t_{k+1}]$ is independent of the increment on $[t_{k+n}, t_{k+n+1}]$, with $n \in \mathcal{N}$. Finally, property (iii) states the distribution followed by the increments, which has zero mean and variance proportional to the duration of the interval- the longer the interval, the larger the variance, so a process like Brownian motion does not evidence propensity to return to a previous position.

Standard Brownian motion may be used to represent the cumulative effect of noise. The difference between the value of a standard Brownian motion at instants t and s, $W_t - W_s \; \forall \; 0 \leq s < t$, reflects the noise on the interval $[s, t]$.

The following characteristics of standard Brownian motion worth mentioning:

1. The increments are **stationary**, i.e., the distribution of $W_t - W_s$, $0 \leq s < t$, only depends on $t - s$;
2. It is a **Quadratic Variation process** with quadratic variation equal to t: $E(W_t) = 0$ and $Var(W_t) = t$, $0 \leq s < t$;
3. $E(W_s W_t) = E(W_s W_t + W_s^2 - W_s^2) = E(W_s(W_t - W_s) + W_s^2) = E(W_s)E(W_t - W_s) + E(W_s^2) = E(W_s^2) = s$, or in general, $E(W_s W_t) = min\{s, t\}$;
4. The mapping $t \to W(t, w)$ is a.s. continuous in t and $W(0, w) = 0$, for each w;
5. If a standard Brownian motion process $\{W_t\}_{t \geq 0}$ is measurable, then each sample path function $W(\cdot, w)$ is $\mathcal{A}$-measurable;
6. Its sample paths are highly irregular and a.s. of unbounded variation;
7. Its sample paths are a.s. non differentiable;
8. The symmetric process $\{-W_t\}_{t \geq 0}$ is also a Brwonian motion (**symmetric property**);
9. $\{W_{ct}\}_{t \geq 0}$ follows the same probability law as $\{\sqrt{c}\, W_t\}_{t \geq 0}$ (**scalling property**);
10. $lim_{t \to 0}(W_t / t) = 0$ a.s., i.e., $\{tW_{1/t}\}_{t \geq 0}$ follows the same probability law as $\{W_t\}_{t \geq 0}$ (**Time-inversion property**);
11. $\{W_{t+h} - W_t\}_{t \geq 0}$ with $h > 0$ is also a standard Brownian motion;
12. A standard Brownian motion process is a continuous-time martingale (given $\{W_t\}_{t \geq 0}$ and $\mathcal{A}_t = \sigma(W_s,\ 0 \leq s < t)$, $E(W_t \mid \mathcal{A}_s) = E(W_t - W_s \mid \mathcal{A}_s) + E(W_s \mid \mathcal{A}_s) = 0 + W_s = W_s$ because $W_t - W_s$ is independent of $\mathcal{A}_s$ and W_s is $\mathcal{A}_s$-measurable, and $E((W_t - W_s)^2 \mid \mathcal{A}_s) = t - s$ since the increment is independent of the past and normally distributed with mean zero- a standard Brownian motion with independent increments is a square integrable martingale);

A random process $\{W_t = (W_t^1, \cdots, W_t^d)\}_{t \geq 0}$ is a d-dimensional standard Brownian motion if every $\{W_t^i\}_{t \geq 0}$ is a one-dimensional standard Brownian motion and independent, for $i = 1, \cdots, d$. For $\{W_t = (W_t^1, \cdots, W_t^d)\}_{t \geq 0}$ a d-dimensional standard Brownian motion and $i, j = 1, \cdots, d$,

1. $E(W_t - W_s \mid \mathcal{A}_s) = 0,\ 0 \leq s < t < \infty$
2. $E(W_t^i W_s^j) = min\{t, s\}\delta_{ij},\ 1 \leq i,\ j \leq d$;
3. $E((W_t^i - W_s^i)(W_t^j - W_s^j)) = (t - s)\delta_{ij},\ 0 \leq s < t$;

with

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases},$$

the Dirac delta function.

When $W_t - W_s$ is normally distributed with mean $\mu(t - s)$ and variance $\sigma^2(t - s)$, we have a general Brownian motion. The coefficient μ is called drift coefficient and σ^2 is the variance coefficient, with $\sigma^2 \neq 0$. When $\mu = 0$ and $\sigma^2 = 1$ we have process termed normalized Brownian motion.

Given a Brownian motion $\{W_t\}_{t \geq 0}$ process with drift μ and variance σ^2,

1. $\{-W_t\}_{t \geq 0}$ is a Brownian motion process with drift $-\mu$ and σ^2;
2. $\{aW_{bt}\}_{t \geq 0}$ is a Brownian motion process with drift $ab\mu$ and variance $a^2 b\sigma^2$, with constants $a, b > 0$;
3. $\frac{(W_t - W_s) - \mu(t-s)}{\sigma}$ is a normalized Brownian motion process.

A standard Brownian motion process $\{W_t\}_{t\geq 0}$ is a Markov process with stationary transition probability $P(W_t \in B \mid W_s) = \frac{1}{\sqrt{2\pi(t-s)}} \int_B e^{\frac{-(x-W_s)^2}{2(t-s)}} dx$ *a.s.*, homogeneous with respect to state space and time and it is possible to present continuous sample paths with probability 1- theorem 2.10 guarantees it. Furthermore, $\{W_t\}_{t\geq 0}$ is also a strong Markov process.

The derivative with respect to time of a standard Brownian motion is a constructed notion, since Brownian motion are not differentiable due to its unbounded variation, called a white noise and denoted as $\dot{W}_t = dW_t/dt$. This process is Gaussian and wide-sense stationary (X_t is wide-sense stationary if $E(X_t X_s) = h(t-s)$ for $h : R \to R$ and $E(X_t) = E(X_s)$) with

$$h(\cdot) = \delta_0(A) = \begin{cases} 1 & \text{if } 0 \in A \\ 0 & \text{if } 0 \notin A \end{cases}.$$

Using R software, one can simulate a Brownian motion path through an Euler scheme, using normally distributed random variables, as in the following example. There exist other strategies that allow the construction of this paths.

Example:

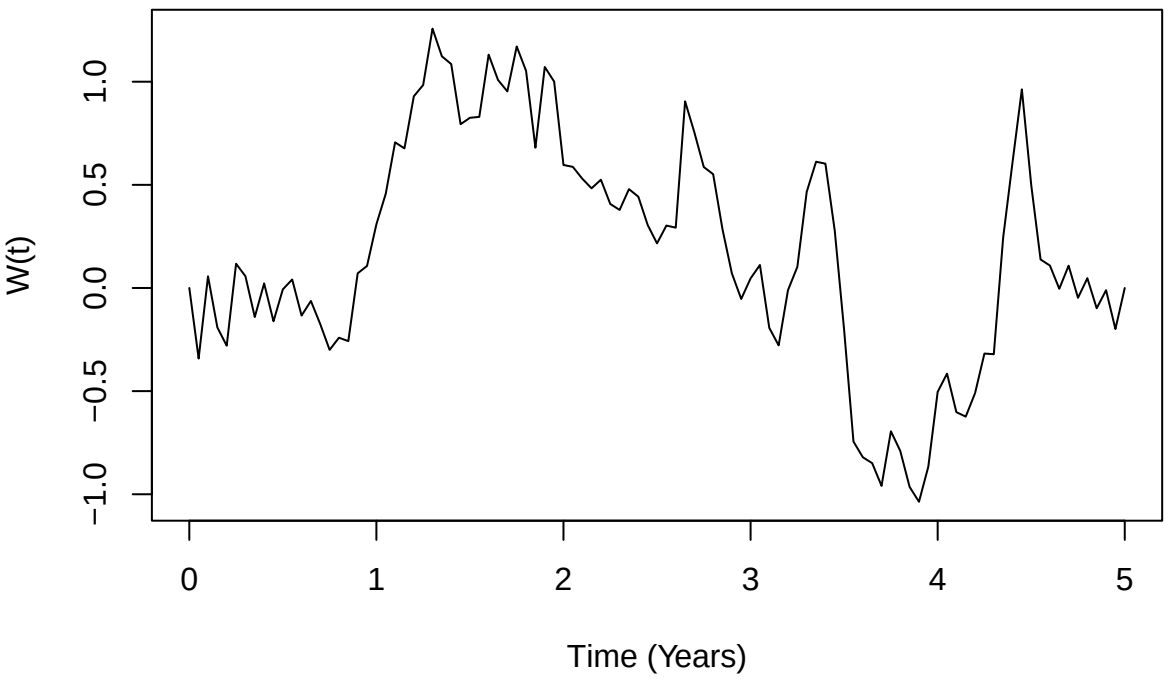
```
n <- 100 # number of steps
T <- 5 # Time

delta_t <- T/n # time step
t <- seq(0, T, by=delta_t) # time vector

W <- rep(0, length(t)) # Brownian motion
set.seed(535563663) # Fix results
N <- rnorm (length(t)-1) # N(0,1) r.v.

for (i in 2:n){
  W[i] <- W[i-1] + N[i]*sqrt(delta_t)
}

plot (t, W, type="l", ylim=range(W), xlab="Time (Years)",
      ylab="W(t)", cex.lab=1, lwd=1, lty=1, col="black")
```



2.6 Itô's integral

Consider the filtered probability space $(\Omega, \mathcal{A}, \{\mathcal{A}_t\}_{t\geq 0}, \mathbb{P})$, and a one dimensional Brownian motion W_t adapted to the filtration $\mathcal{J} = \{\mathcal{A}_t\}_{t\geq 0}$. Our objective in this section is to define a stochastic integral of the form

$$\int_a^b f(s,\omega)dW(s,\omega) \text{ where } \omega \in \Omega, \text{ or more succinctly } \int_a^b f(s)dW_s,$$

for any $a, b \in \mathbb{R}$.

Definition 2.1 (Natural domain of the Itô integral)**.** For any $a, b \in \mathbb{R}_0^+$, we define the **natural domain of the Itô integral** as the space $\mathfrak{M}^2([a,b])$ of all real-valued processes $f = \{f(t)\}_{t\in[a,b]}$ on the product space $\Omega \times [a,b]$ such that

1. the process f is measurable with regards to the σ- algebra on $\Omega \times [a,b]$;
2. the process f is adapted with respect to the filtration $\{\mathcal{A}_t\}_{[a,b]}$ i.e. for any $t \in [a,b]$, $f(t)$ is $\mathcal{A}_t$-measurable;
3. $\|f\|_2^2([a,b]) = \mathbb{E}\left(\int_a^b |f(t)|^2\, dt\right) < \infty$.

Note, that under the norm $\|\cdot\|_2^2([a,b])$ the space $\mathfrak{M}^2([a,b])$ is complete, and any two processes $f, g \in \mathfrak{M}^2([a,b])$ are **equivalent** if $\|f-g\|_2^2([a,b])$.

Definition 2.2 (Itô Integral for Elementary Functions)**.** A stochastic process $\phi = \{\phi(t)\}_{t\in[a,b]} \in \mathfrak{M}_0^2([a,b]) \subset \mathfrak{M}^2([a,b])$ is said to be an elementary process (where $\mathfrak{M}_0^2([a,b])$ is the subspace of elementary functions of $\mathfrak{M}^2([a,b])$) if

$$\phi(t,\omega) = \sum_{i=1}^{k-1} e_i(\omega), \mathbb{1}_{(t_i,t_{i+1}]}(t), \tag{2.6}$$

where $a = t_0 < t_1 < \cdots < t_n = b$, and e_i are square integrable random variables and $\mathcal{A}_t$-measurable. Its stochastic integral in respect to W_t is thus defined as

$$I(\phi) = \int_a^b \phi(t)\, dW(t) = \sum_{i=1}^{k-1} e_i[W_{t_{i+1}} - W_{t_i}]. \tag{2.7}$$

Theorem 2.12 (Itô Isometry)**.** *For any elementary function $\phi \in \mathfrak{M}_0^2([a,b])$, $\|f\|^2_{\mathfrak{M}_0^2([a,b])} = \|f\|^2_{L^2(\mathbb{P})} = \|f\|_2^2([a,b]) = \mathbb{E}(I(\phi))$, i.e. the mapping $I : \mathfrak{M}_0^2([a,b]) \to L^2(\mathbb{P})$ is a continuous isometry.*

Lemma 2.1. *Let $f \in \mathfrak{M}^2([a,b])$, then there exists a sequence of elementary functions $\phi_n \in \mathfrak{M}_0^2([a,b])$, such that $\lim_{n\to\infty} \|f-\phi_n\|^2_{\mathfrak{M}_0^2([a,b])} = \lim_{n\to\infty} \mathbb{E}\left[\int_a^b |f-\phi_n|^2\, dt\right] = 0$.*

Definition 2.3 (Itô Integral)**.** For any stochastic process $f \in \mathfrak{M}^2([a,b])$, and a sequence of elementary functions $\phi_n \in \mathfrak{M}_0^2([a,b])$ such that $\phi_n \in \mathfrak{M}_0^2([a,b])$, such that $\lim_{n\to\infty} \|f-\phi_n\|^2_{\mathfrak{M}_0^2([a,b])} = 0$, the Itô integral of f is defined as

$$I(f) = \lim_{n\to\infty} \int_a^b \phi_n\, dW(t) \text{ in } L^2(\mathbb{P}). \tag{2.8}$$

Consider now, $f, g \in \mathfrak{M}^2([a,b])$ and $\alpha, \beta \in \mathbb{R}$, then the following 9 properties can be easily proven, by first proving the properties for an elementary function $\phi \in \mathfrak{M}_0^2([a,b])$ and then using the fact that there exists a sequence of elementary functions $\phi_n \in \mathfrak{M}_0^2([a,b])$, such that $\lim_{n\to\infty} \mathbb{E}\left[\int_a^b |f-\phi_n|^2\, dt\right] = 0$:

1. **(Measurability)** $\int_a^b f(t)\,dW(t)$ is $\mathcal{A}_a$-measurable;
2. $\mathbb{E}\left(\int_a^b f(t)\,dW(t)\right) = 0$ and $\mathbb{E}\left(\int_a^b f(t)\,dW(t)|\mathcal{A}_a\right) = 0$;
3. $\mathbb{E}\left(\left|\int_a^b f(t)\,dW(t)\right|^2\right) = \mathbb{E}\left(\int_a^b |f(t)|^2\,dt\right)$ and
   $\mathbb{E}\left(\left|\int_a^b f(t)\,dW(t)\right|^2 |\mathcal{A}_a\right) = \mathbb{E}\left(\int_a^b |f(t)|^2\,dt\,|\mathcal{A}_a\right) = \int_a^b \mathbb{E}\left(|f(t)|^2\,|\mathcal{A}_a\right) dt$;
4. **(Linearity)** $\int_a^b (\alpha f(t) + \beta g(t))\,dW(t) = \alpha \int_a^b f(t)\,dW(t) + \beta \int_a^b g(t)\,dW(t)$;
5. $\mathbb{E}\left[\int_a^b f(t)\,dW(t) \int_a^b g(t)\,dW(t)\right] = \mathbb{E}\left[\int_a^b f(t)g(t)\,dW(t)\right]$;
6. If Z is a real value, bounded $\mathcal{A}_b$-measurable random variable then $Zf \in \mathfrak{M}^2([a,b])$ and $\int_a^b Zf(t)\,dW(t) = Z\int_a^b f(t)\,dW(t)$;
7. **Gaussian** If f is deterministic (i.e. f is independent of the value $\omega \in \Omega$, then $\int_a^b f(t)\,dW(t) \sim N\left(0, \int_a^b |f(t)|^2\,dt\right)$;
8. **(Additivity)** $\int_a^b f(t)\,dW(t) = \int_c^b f(t)\,dW(t) + \int_a^c f(t)\,dW(t)$, for any $c \in (a,b)$;
9. For any set $A \subset [a,b]$, we have $\int_A f(t)\,dW(t) = \int_a^b f(t)\mathbb{1}_A\,dW(t)$.

Definition 2.4 (Itô Integral)**.** For any stochastic process $f \in \mathfrak{M}^2([a,b])$, the undefinided Itô stochastic integral of f is defined as

$$I(t) = \int_0^t f(t)\,dW(t), \text{ for any } t \in [0,T], \tag{2.9}$$

where $f(0) = 0$, $\{I(t)\}_{t\in[0,T]}$ is $\mathcal{A}_t$-adapted, square integrable and martingale with respect to the filtration $\{\mathcal{A}_t\}_{t\geq 0}$, i.e.

$$\mathbb{E}[I(t)|\mathcal{A}_s] = \mathbb{E}[I(s)|\mathcal{A}_s] + \mathbb{E}\left[\int_s^t f(r)\,dW(r)|\mathcal{A}_s\right] = I(s) + \mathbb{E}\left[\int_s^t f(r)\,dW(r)\right] = I(s), \tag{2.10}$$

for any $0 \leq s < t \leq T$. By the Doob's martingale property we have:

$$\mathbb{E}\left[\max_{0\leq t\leq T}\left|\int_0^t f(s)\,dW(s)\right|^2\right] \leq 4\mathbb{E}\left[\left|\int_0^t f(s)\,dW(s)\right|^2\right] = 4\mathbb{E}\left(\int_0^T |f(s)|^2\,ds\right). \tag{2.11}$$

Moreover, $\{I(t)\}_{t\in[0,T]}$ has a continuous version a.s., that is, there exists a t-continuous stochastic process J_t on $(\Omega, \mathcal{A}, \{\mathcal{A}\}_t, \mathbb{P})$ with continuous paths such that $\mathbb{P}\left(J_t = \int_0^t g(s)\,dW(s)\right) = 1$, for any $t \in [0,T]$.

Let us write the Itô's integral of a stochastic process f as $X_t = \int_0^t f(s)\,dW(s)$, for every $t \in [0,T]$ the following four properties are easily shown:

1. $\mathbb{E}(X_t) = 0$ and $\mathbb{V}(X_t) = \mathbb{E}(X_t^2) = \int_0^t \mathbb{E}(f(s)^2)\,ds$, for any $t \in [0,T]$;
2. $\mathbb{C}ov(X_t, X_s) = \mathbb{E}(X_t X_s) = \int_0^{\min\{t,s\}} \mathbb{E}(f(u)^2)\,du$, for any $t, s \in [0,T]$;
3. $\mathbb{E}\left(|X_t|^2\right) = \int_0^t \mathbb{E}\left(|f(u)|^2\right)\,du$, for any $t \in [0,T]$;
4. X_t has the orthogonal property, that is $\mathbb{E}\left[(X_u - X_t)(X_s - X_r)\right] = 0$ for $0 \leq r \leq s \leq t \leq u \leq T$.

2.7 Itô formula

Definition 2.5 (Itô's process)**.** Let $\{W_t\}_{t\geq 0}$ be a one-dimensional Brownian motion defined on the complete probability space $(\Omega, \mathcal{A}, \{\mathcal{A}_t\}_{t\geq 0}, \mathbb{P})$. A continuous and adapted process X_t is called an Itô Process if it is of the form

$$X_t = X_0 + \int_0^t f(s)\,ds + \int_0^t g(s)\,dW(s), \tag{2.12}$$

where $f \in L^1_{\mathbb{R}_0^+}$ and $g \in L^2_{\mathbb{R}_0^+}$. Its **stochastic differential** is

$$dX_t = f(s)\,ds + g(s)\,dW(s) \tag{2.13}$$

Theorem 2.13 (One-Dimensional Itô Formula)**.** *Let X_t be an Itô process as previously defined. Consider a function $V = V(X,t) : \mathbb{R}\times\mathbb{R}_0^+ \to \mathbb{R}$ twice continuously differentiable in X and once in t. Then the process $Y(t) = V(X(t),t)$ is an Itô process that can be expressed as*

$$\begin{aligned} Y(t) - Y(0) &= V(X(t),t) - V(X(0),0) \\ &= \int_0^t \left(V_t(X(s),s) + V_X(X(s),s)f(s) + \frac{1}{2}V_{XX}(X(s),s)g(s)^2 \right) ds \\ &\quad + \int_0^t V_X(X(s),s)g(s)\,dW_s \end{aligned} \tag{2.14}$$

We can all write the Itô formula for $Y(t)$ in the differential form:

$$\begin{aligned} dY(t) &= \left(V_t(X(t),t) + V_X(X(t),t)f(t) + \frac{1}{2}V_{XX}(X(t),t)g(t)^2 \right) dt \\ &\quad + V_X(X(t),t)g(t)\,dW_t \end{aligned} \tag{2.15}$$

The proof of the Itô formula while simple is quite laborious. It is a consequence of the application of the two dimensional Taylor series to $V(X,t)$, which results in the equation

$$V(X(t),t) - V(X(0),0) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \frac{1}{n!m!} \int_0^t \int_{X(0)}^{X(t)} V_{X^n,t^m}\,(dX)^n\,(ds)^m, \tag{2.16}$$

and after every component of the sum is computed we arrive at the Itô formula. The computation for each term of the sum can be heuristically expressed by the Itô multiplication table:

$\cdot$	dt	$dW(t)$
dt	0	0
$dW(t)$	0	dt

Theorem 2.14 (Integration by parts Formula)**.** *Let $f(s)$ be independent of Ω, with continuous and bounded variation on $s \in [0,T]$. Then*

$$\int_0^t f(s)\,dW(s) = f(t)W(t) - \int_0^t W(s)\,df(s). \tag{2.17}$$

The previous theorem 2.17 is the result of an application of the Itô formula to the function $V(X,t) = f(t)X$, where X is $W(t)$.

Note that if $\{X_t\}_{t\geq 0}$ is an Itô process $X_t = X_0 + \int_0^t f(s)\,ds + \int_0^t g(s)\,dW(s)$ where both f and g are deterministic functions of t, with X_0 constant, then

$$X(t) \sim N\left[\mathbb{E}(X(t)), \mathbb{V}(X(t))\right] = N\left[X_0 + \int_0^t f(s)\,ds, \int_0^t g(s)^2\,ds\right]$$

with independent increments.

The following tables present the solution to common stochastic integrals and the function $V(X,t)$, where X is $W(t)$, that was applied to the Itô formula, if any, to solve the integral.

	Stochastic Integral
1.	$\int_0^t dW(s) = W(t)$
2.	$\int_0^t W(s)\, dW(s) = \frac{1}{2}W(t)^2 - \frac{1}{2}t$
3.	$\int_a^b c\, dW(s) = c\,(W(b) - W(a))$,
4.	$\int_a^b W(s)\, dW(s) = \frac{1}{2}\left(W(b)^2 - W(a)^2\right) - \frac{1}{2}(b-a)$
5.	$\int_0^t s\, dW(s) = tW(t) - \int_0^t W(s)ds$
6.	$\int_0^t W(s)^2\, dW(s) = \frac{1}{3}W(t)^3 - \int_0^t W(s)ds$
7.	$\int_0^t e^{W(s)}\, dW(s) = e^{W(t)} - 1 - \frac{1}{2}\int_0^t e^{W(s)}ds$
8.	$\int_0^t W(s)e^{W(s)}\, dW(s) = 1 + W(t)e^{W(t)} - e^{W(t)} - \frac{1}{2}\int_0^t e^{W(s)}\,(1 + W(s))\, ds$
9.	$\int_0^t sW(s)\, dW(s) = \frac{1}{2}\left(W(t)^2 - \frac{t}{2}\right) - \frac{1}{2}\int_0^t W(s)^2\, ds$
10.	$\int_0^t \left(W(s)^2 - s\right)\, dW(s) = \frac{1}{3}W(t)^3 - tW(t)$
11.	$\int_0^t e^{\mu s + \sigma W(s)}\, dW(s) = \frac{1}{\sigma}\left(e^{\mu t + \sigma W(t)} - 1\right) - \left(\frac{\mu}{\sigma} + \frac{\sigma}{2}\right)\int_0^t e^{\mu s + \sigma W(s)}ds$, with μ, σ constants
12.	$\int_0^t \sin\left(W(s)\right)\, dW(s) = 1 - \cos(W(t)) - \frac{1}{2}\int_0^t \cos(W(s))ds$
13.	$\int_0^t \cos\left(W(s)\right)\, dW(s) = \sin(W(t)) + \frac{1}{2}\int_0^t \sin(W(s))ds$
14.	$\mathbb{E}\left[\int_0^t dW(s)\right]$
15.	$\mathbb{E}\left[\int_0^t W(s)\, dW(s)\right] = 0$
16.	$\mathbb{V}\left[\int_0^t W(s)\, dW(s)\right] = \frac{t^2}{2}$

	$V(X,t)$
1.	direct
2.	$V(X,t) = X^2$
3.	direct
4.	$V(X,t) = X^2$
5.	$V(X,t) = tX$
6.	$V(X,t) = X^3$
7.	$V(X,t) = e^X$
8.	$V(X,t) = Xe^X$ and the result from 2
9.	$V(X,t) = tX^2$
10.	from results 5 and 6
11.	$V(X,t) = e^{\mu t + \sigma X}$
12.	$V(X,t) = \cos(X)$
13.	$V(X,t) = \sin(X)$
14.	direct
15.	direct from 2
16.	direct from 2

2.8 Martingale Representation Theorem

Theorem 2.15 (Itô Representation Theorem). *Let $\{W_t\}_{t\geq 0}$ be a Brownian motion in the probability space $(\Omega, \mathcal{A}_t, \{\mathcal{A}_t\}_{t\geq 0}, \mathbb{P})$. For every $f \in L^2(\mathbb{P})$ and adapted to the filtration $\{\mathcal{A}_t\}_{t\geq 0}$, there exists a unique stochastic process $\mathcal{U}_t$, adapted to $\{\mathcal{A}_t\}_{t\geq 0}$, where $\mathbb{E}\left(\int_0^t \mathcal{U}(s)^2 ds\right) < \infty$, such that*

$$f(t) = \mathbb{E}[f(0)] + \int_0^t \mathcal{U}_s \, dW(s). \tag{2.18}$$

Theorem 2.16 (Martingale Representation Theorem). *Let $\{W_t\}_{t\geq 0}$ be a Brownian motion in the probability space $(\Omega, \mathcal{A}_t, \{\mathcal{A}_t\}_{t\geq 0}, \mathbb{P})$. For every $\{X_t\}_{t\geq 0}$ a square martingale adapted to the filtration $\{\mathcal{A}_t\}_{t\geq 0}$, there exists a unique stochastic process $\mathcal{U}_t$, adapted to $\{\mathcal{A}_t\}_{t\geq 0}$, where $\mathbb{E}\left(\int_0^t \mathcal{U}(s)^2 ds\right) < \infty$, such that*

$$X(t) = \mathbb{E}[X(0)] + \int_0^t \mathcal{U}_s \, dW(s). \tag{2.19}$$

2.9 Multidimensional Itô Formula

Definition 2.6 (Multidimensional Itô Process). Suppose $\{\mathbf{W}_t\}_{t\geq 0} = (W_1(t), \ldots, W_m(t))$ is an m-dimensional Brownian motion defined in the probability space $(\Omega, \mathcal{A}, \{\mathcal{A}_t\}_{t\geq 0}, \mathbb{P})$, adapted to the filtration $\{\mathcal{A}_t\}_{t\geq 0}$, where the components W_r, for $r = 1, \ldots, m$, are independent one dimensional Brownian motions. A $\mathbb{R}^d$-valued continuous adapted process $\{\mathbf{X}(t)\}_{t\geq 0} = (X_1(t), \ldots, X_d(t))'$ is called a **d-dimensional Itô process** if it is of the form,

$$\mathbf{X}_t = \mathbf{X}_0 + \int_0^t \mathbf{f}(s) \, ds + \int_0^t \mathbf{g}(s) \, d\mathbf{W}(s), \tag{2.20}$$

or in a different notation,

$$X_i(t) = X_i(0) + \int_0^t f_i(s) \, ds + \sum_{j=1}^m \int_0^t g_{i,j}(s) \, dW_j(s), \text{ for any } i = 1, 2, \ldots, d., \tag{2.21}$$

where $\mathbf{f}(t) = (f_1(t), \ldots, f_d(t))' \in L^1_{\mathbb{R}_0^+}\left(\mathbb{R}^d\right)$ and $\mathbf{g}(t) = \left[g_{i,j}\right]_{d\times m} \in L^2_{\mathbb{R}_0^+}\left(\mathbb{R}^{d\times m}\right)$. Its **stochastic differential** is

$$d\mathbf{X}_t = \mathbf{f}(s) \, dt + \mathbf{g}(t) \, d\mathbf{W}(t), \tag{2.22}$$

or in a different notation,

$$dX_i(t) = f_i(t) \, dt + \sum_{j=1}^m g_{i,j}(t) \, dW_j(t), \text{ for any } i = 1, 2, \ldots, d. \tag{2.23}$$

Theorem 2.17 (Multidimensional Itô Formula). *Let $\boldsymbol{X}_t$ be an Itô process as previously defined . Consider a function $V = V(\boldsymbol{X}, t) : \mathbb{R}^d \times \mathbb{R}_0^+ \to \mathbb{R}$ twice continuously differentiable in $\boldsymbol{X}$ and once in t. Then the process $\boldsymbol{Y}(t) = V(\boldsymbol{X}(t), t)$ is an Itô process that can be expressed as*

$$\begin{aligned} dY(t) = &\left[V_t(\boldsymbol{X}(t), t) + V'_{\boldsymbol{X}}(\boldsymbol{X}(t), t)\boldsymbol{f}(t) + \frac{1}{2} tr\left(\boldsymbol{g}'(t) V_{\boldsymbol{X}\boldsymbol{X}}(\boldsymbol{X}(t), t)\boldsymbol{g}(t)\right)\right] dt \\ &+ V'_{\boldsymbol{X}}(\boldsymbol{X}(t), s)\boldsymbol{g}(t) \, d\boldsymbol{W}(t), \end{aligned} \tag{2.24}$$

or

$$\begin{aligned} dY(t) = &\left[V_t + \sum_{i=1}^d V_{X_i} f_i(t) + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^m V_{X_i X_j} g_{ik}(t) g_{jk}(t)\right] dt \\ &+ \sum_{i=1}^d \sum_{j=1}^m V_{X_i} g_{ij}(t) \, dW_j(t). \end{aligned} \tag{2.25}$$

The proof of the multidimensional Itô formula is analogous to the one-dimensional case. It is a consequence of applying the $d+1$ dimensional Taylor series to $V(X_1, X_2, \ldots, X_d, t)$, which results in the equation

$$V(\mathbf{X}(t), t) - V(\mathbf{X}(0), 0) \tag{2.26}$$
$$=$$
$$\sum_{n_1=0}^{\infty} \cdots \sum_{n_d=0}^{\infty} \sum_{m=0}^{\infty} \frac{1}{n_1! \ldots n_d! m!} \int_0^t \int_{X_d^{n_d}(0)}^{X_d^{n_d}(t)} \cdots \int_{X_1^{n_1}(0)}^{X_1^{n_1}(t)} V_{X_1^{n_1}, \ldots, X_d^{n_d}, t^m} (dX_1)^{n_1} \ldots (dX_d)^{n_d} (ds)^m,$$

and after every component of the sum computed we arrive at the multidimensional Itô formula. Just as we did previously, the computation for each term of the sum can be heuristically expressed by the multidimensional Itô multiplication table:

$\cdot$	dt	dW_i	dW_j
dt	0	0	0
dW_i	0	dt	0
dW_j	0	0	dt

Chapter 3

Solution of Stochastic Differential Equations

The present chapter is dedicated to the study of the solutions of Stochastic Differential Equation, namely the conditions to its existence and uniqueness, how to determine the exact solution and its stability.

3.1 Introduction

We are interested in the solution of a Stochastic Differential Equation of the form

$$\begin{cases} dX(t) = f(X(t),t)dt + g(X(t),t)dW(t),\ t \in [t_0, T],\ T > 0 \\ X(t_0) = X_0 \end{cases}, \tag{3.1}$$

where X_0 is independent of the $W(t) - W(t_0)$ and $E(|X_0|^2) < +\infty$. Function f is called the drift coefficient and can be interpreted as the short-term growth. Function g is the diffusion coefficient and can be viewed as the short-term variability. If $g(X(t),t) = 0$, the equation 3.1 is converted to an Ordinary Differential Equation.

Let $(\Omega, \mathcal{A}, P)$ be a complete probability space endowed with the filtration $\mathcal{J} = \{\mathcal{A}_t\}_{t\geq 0}$ that is right-continuous and $\{\mathcal{A}_0\}$ contains all P-null sets, $\{W_t\}_{t\geq 0}$ a one-dimensional Brownian motion defined on the filtered probability space $(\Omega, \mathcal{A}, \{\mathcal{A}_t\}_{t\geq 0}, P)$ and, for $0 \leq t_o < T < +\infty$, X_0 is a $\mathcal{A}_{t_0}$-measurable random variable with $E(|X_0|^2) < +\infty$. Furthermore, consider $f, g : \mathcal{R} \times [t_o, T] \to \mathcal{R}$ Borel-measurable functions. One can represent the solution of 3.1 in the integral form as follows:

$$X(t) = X(0) + \int_{t_0}^{t} f(X(s),s)ds + \int_{t_0}^{t} g(X(s),s)dW(s),\ t \in [t_0, T]. \tag{3.2}$$

A one-dimensional stochastic process $\{X_t\}_{t\in[t_o,T]}$ is the solution to equation 3.1 if:

1. $\{X_t\}$ is continuous and $\mathcal{A}_t$-adapted;
2. $\{f(X(t),t)\} \in L^1([t_0, T])$ and $\{g(X(t),t)\} \in L^2([t_0, T])$;
3. Equation 3.2 holds for every $t \in [t_o, T]$ almost surely (a.s.).

Finally, a solution $\{X_t\}_{t\in[t_o,T]}$ is unique if it is indistinguishable from another solution $\{\tilde{X}_t\}_{t\in[t_o,T]}$, i.e., $P(X_t = \tilde{X}_t,\ t \in [t_0, T]) = 1$.

3.2 Existence and Uniqueness

The following theorem emphasize the conditions to have a solution that is unique to equation 3.1.

Theorem 3.1 (Existence and Uniqueness Theorem). *Let the coefficient functions f, g of the SDE 3.1 satisfy the following conditions:*

1. *Uniform Lipschitz condition: $|f(x,t) - f(y,t)| + |g(x,t) - g(y,t)| \leq k|x-y|$, for some $k \in \mathcal{R}$ constant, $t \in [t_0, T],\ T > 0$;*
2. *Linear growth condition: $|f(x,t)|^2 + |g(x,t)|^2 \leq c^2(1+|x|^2)$, for some constant c and $t \in [t_0, T],\ T > 0$.*

If these conditions hold, then there exists a continuously adapted solution X_t of equation 3.1 such that $X(0) = X_0$ and X_t is uniformly bounded in $\mathcal{M}^2([t_0, T])$ ($\sup_{t_0 \leq t \leq T} E(|X_t^2|) < +\infty$). Furthermore, if both X_t and $\tilde{X}_t$ are continuous $\mathcal{M}^2([t_0, T])$ bounded solutions of 3.1, then $P\left(\sup_{t_0 \leq t \leq T} |X_t - \tilde{X}_t| = 0\right) = 1$, and thus X_t is the unique solution.

One proof of the Existence Uniqueness Theorem involves the use of a norm $\|X\| = \left(\int_0^t e^{\lambda s}\mathbb{E}[|X_s|^2]\, ds\right)^{1/2}$, where $\lambda > k^2(t+1)$, and the operator $\Psi : \mathfrak{M}^2([t_0, T]) \to \mathfrak{M}^2([t_0, T])$, where $\Psi(X)_t = X_0 + \int_0^t f(s, X_s)\, ds + \int_0^t g(s, X_s)\, dW_s$. The linear growth condition ensures that the operator Ψ is well defined (i.e. $\Psi(X)_t \in \mathfrak{M}^2([t_0, T])$ for any $X \in \mathfrak{M}^2([t_0, T])$), and the uniform Lipschitz condition guarantees the operator Ψ is contractive under the aforementioned norm (i.e. $\|X - Y\| > \|\Psi(X) - \Psi(Y)\|$). Hence, by the Fixed Point Theorem, (3.2) has a unique solution.

Another way to look at the theorem is by noticing that the first condition guarantees that functions f and g do not change faster than a linear function of X relative to changes in X, implying the Lipschitz continuity of $f(\cdot, t)$ and $g(\cdot, t)$ for all $t \in [t_0, T]$. The second condition restricts the coefficient functions f and g to linear increases with respect to X, guaranteeing a.s. that the solution X_t does not explode in the interval $[t_0, T]$ independently of X_0.

A solution X_t is termed a strong solution if it has strongly unique sample path and the filtered probability space, the Brownian motion and the coefficient functions are specified in advance. When only f and g are specified in advance and the pair $(\tilde{X}_t, \tilde{W}_t)$ is defined on a suitably probability space the solution $\tilde{X}_t$ is called a weak solution, besides the verification of the conditions of theorem 3.1. If two weak solutions are indistinguishable, then path uniqueness holds. Two weak solutions are called weakly unique if they have the same probability distribution function. A strong solution is also weak, but generally a weak solution is not strong.

Given that the conditions of theorem 3.1 are quiet restrictive, the first condition can be relaxed to a local uniform Lipschitz condition: for every $n \geq 1$, exists a positive constant k_n such that, for all $x, y \in \mathcal{R},\ t \in [t_0, T],\ T > 0$ and $max\{|x|, |y|\} \leq n$,

$$|f(x,t) - f(y,t)| + |g(x,t) - g(y,t)| \leq k_n|x-y|. \tag{3.3}$$

Thus, if this local uniform Lipschitz condition holds simultaneously with the second condition of theorem 3.1, then there exists a unique solution X_t, $X_t \in \mathcal{M}^2([t_0, T])$ to the SDE 3.1. This holds every time when f and g are continuous differentiable in x on $\mathcal{R} \times [t_0, T]$.

Moreover, it is also possible to replace the linear growth condition by a monotone condition: let k be a positive constant such that

$$xf(x,t) + \frac{1}{2}|g(x,t)|^2 \leq k(1+|x|^2),\ \forall\ (X(t), t) \in \mathcal{R} \times [t_0, T],\ T > 0. \tag{3.4}$$

Now, if 3.3 and 3.4 holds, then there exists a solution X_t to 3.1, with $X_t \in \mathcal{M}^2([t_0, T])$. The linear growth condition implies the monotonic condition, but the opposite may not be true.

It worth mention that for f and g defined on $\mathcal{R} \times [t_0, T]$ and when the conditions of theorem 3.1 are verified on every finite subinterval $[t_0, T] \subset [t_o, +\infty)$, the equation 3.1 has a unique solution X_t define on $[t_0, T)$ called global solution.

An autonomous SDE of the form

$$\begin{cases} dX(t) = f(X(t))dt + g(X(t))dW(t),\ t \in [t_0, T],\ T > 0 \\ X(t_0) = X_0 \end{cases}, \tag{3.5}$$

with X_0 a random variable independent of $W(t) - W(t_0)$, that verifies the assumptions of theorem 3.1 has a unique, continuous global solution X_t on $[t_0, +\infty)$ such that $X(t_0) = X_0$ given that

$$|f(x) - f(y)| + |g(x) - g(y)| \leq k|x - y| \tag{3.6}$$

is satisfied and

$$|f(x)|^2 + |g(x)|^2 \leq k^2(1 + |x|^2) \tag{3.7}$$

holds for a fixed y.

3.3 Linear SDEs

Equation 3.1 is a linear SDE if the coefficient functions f and g are both linear with respect to X on $\mathcal{R} \times [t_0, T]$. Thus, if $f(X(t), t) = a(t) + A(t)X(t)$ and $g(X(t), t) = b(t) + B(t)X(t)$ hold with $a(t), b(t), A(t), B(t) : [t_o, T] \to \mathcal{R}$, the linear SDE is as follows:

$$\begin{cases} dX(t) = (a(t) + A(t)X(t))dt + (b(t) + B(t)X(t))dW(t),\ t \in [t_0, T],\ T > 0 \\ X(t_0) = X_0 \end{cases}. \tag{3.8}$$

A linear SDE is homogeneous if $a(t) = b(t) = 0$ and linear in the narrow sense if $B(t) \equiv 0$.

If functions $a(t), b(t), A(t), B(t)$ are measurable and bounded on $[t_0, T]$,$T > 0$, i.e.,

$$\sup_{t_0 \leq t \leq T} (|a(t)| + |b(t)| + |A(t)| + |B(t)|) < +\infty,$$

then functions $f(X(t), t) = a(t) + A(t)X(t)$ and $g(X(t), t) = b(t) + B(t)X(t)$ satisfy the hypothesis of theorem 3.1, guaranteeing that equation 3.8 has a unique continuous solution X_t over $[t_0, T]$, provided $E(|X_0|^2) < \infty$ and X_0 is independent of $W(t) - W(t_0)$. Again, if the previous is verified in all subsets of $[t_0, +\infty)$, the solution X_t is classified as a unique, continuous and global solution.

An autonomous linear SDE requires functions $a(t),\ b(t),\ A(t),\ B(t)$ independent of time, with SDE represented by

$$dX(t) = (a + AX(t))dt + (b + BX(t))\, dW(t), \tag{3.9}$$

where $a,\ b,\ A,\ B$ are real constants and a unique, continuous and global solution X_t is always achievable for this type of SDE.

Itô formula is extremely useful to determine the solution to a SDE of the form $X_t = (W_t, t)$. One can obtain that solution by writing $X_t = (W_t, t)$ in terms of second-order Taylor Expansion and applying the rule of products established on Itô multiplication rule. After the relevant simplifications, it remains to match the coefficients of the first SDE with the coefficients of the SDE obtained by the previous explained strategy. Finally, the function that satisfies both equations is the solution to the SDE. This methodology is called Itô coefficient matching and the class of SDEs that are particularly well suited for solving applying this method are characterized as exact.

Theorem 3.2 (Test for exactness). *A SDE is exact if the coefficient functions $f(x, t)$ and $g(x, t)$ satisfy the following condition:*

$$f_x = g_t + \frac{1}{2} g_{xx}.$$

Note, that the test for exactness, is just an application of the Itô's form to a stochastic function $h(t, W_t)$, where $h(t, W_t) = X_t$.

Example To solve the Geometric BM one can apply directly Itô formula or Itô matching coefficient strategy.

$$\begin{cases} dX_t = \mu X_t dt + \sigma X_t dW_t, \ t \in [0, +\infty) \\ X(0) = X_0 \end{cases} \tag{3.10}$$

Let's start by Itô formula: it's possible to rewrite Geometric BM as

$$\frac{dX_t}{X_t} = \mu dt + \sigma dW_t. \tag{3.11}$$

Left hand side of the equation suggest the existence of $ln(X_t)$ in the solution. Thus, considering $V(X_t, t) = ln(X_t)$, $f = \mu X_t$ and $g = \sigma X_t$, Itô formula returns

$$dV(X_t, t) = \Big(0 + \frac{1}{X_t}\mu X_t + \frac{1}{2}\frac{-1}{X_t^2}\sigma^2 X_t^2\Big)dt + \Big(\frac{1}{X_t}\sigma X_t\Big)dW_t, \tag{3.12}$$

which can be developed as

$$\int_0^t dV(X_s, s) = \int_0^t \Big(\mu - \frac{1}{2}\sigma^2\Big)ds + \int_0^t \Big(\sigma\Big)dW_s. \tag{3.13}$$

Finally, substituting $V(X_t, t)$ and integrating, one obtains

$$ln(X_t) - ln(X_0) = (\mu - \frac{1}{2}\sigma^2)t + \sigma^2 Wt, \tag{3.14}$$

with $W_0 = 0$. The solution of Geometric BM SDE is

$$ln\Big(\frac{X_t}{X_0}\Big) = (\mu - \frac{1}{2}\sigma^2)t + \sigma^2 Wt \iff X_t = X_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma^2 Wt}. \tag{3.15}$$

To apply Itô matching coefficient strategy let's impose the conditions to find function $h(x, t)$:

1. $\mu h(x, t) = h_t(x, t) + \frac{1}{2}h_{xx}(x, t)$
2. $\sigma h(x, t) = h_x(x, t)$

From 2., one can obtain

$$\frac{h_x(x, t)}{h(x, t)} = \sigma \iff h(x, t) = e^{\sigma x + c(t)}, \tag{3.16}$$

with c(t) arbitrary. Replacing that information in 1., one obtains

$$\mu e^{\sigma x + c(t)} = c'(t)e^{\sigma x + c(t)} + \frac{1}{2}\sigma^2 e^{\sigma x + c(t)} \iff \mu e^{\sigma x + c(t)} = \Big(c'(t) + \frac{1}{2}\sigma^2\Big)e^{\sigma x + c(t)} \tag{3.17}$$

The previous equation imposes that

$$c'(t) = \mu - \frac{1}{2}\sigma^2 \iff c(t) = \Big(\mu - \frac{1}{2}\sigma^2\Big)t, \tag{3.18}$$

obtaining $h(x, t) = e^{\sigma x + (\mu - \frac{1}{2}\sigma^2)t}$. Finally, the solution to the Geometric BM SDE is $X_t = X_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma^2 Wt}$, since it is necessary to leave X_t on the left hand side of the SDE, as in the Itô formula. ■

Below we present the strong solutions of specific types of SDE, given the **General form** $dX(t) = f(X(t), t)dt + g(X(t), t)dW(t)$.

1. **Linear** $\rightarrow dX(t) = (a_t + b_t X_t)dt + (c_t + e_t X_t)dW(t)$
   Solution: $X(t) = \Phi(t)\{X_0 + \int_0^t (a_s - e_s c_s)\Phi_s^{-1} ds + \int_0^t c_s \Phi_s^{-1} dW_s\}$, with $\Phi(t) = \exp\{\int_0^t e_s dW_s + \int_0^t (b_s - \frac{1}{2}e_s^2)ds\}$ and $\Phi_s^{-1} = exp\{-\int_0^s e_u dW_u - \int_0^s (b_u - \frac{1}{2}e_u^2)du\}$

   (a) **Homogeneous** $\rightarrow dX(t) = b_t X_t dt + e_t X_t dW(t)$ $(a_t = c_t \equiv 0)$
   Solution: $X_t = X_0 \exp\{\int_0^t (b_s - \frac{1}{2}e_s^2)ds + \int_0^t e_s dW_s\}$

   (b) **Narrow sense** $\rightarrow dX(t) = (a_t + b_t X_t)dt + c_t dW(t)$ $(e_t \equiv 0)$
   Solution: $X_t = e^{\int_0^t b_s ds}\{X_0 + \int_0^t e^{-\int_0^u b_u du} a_s ds + \int_0^t e^{-\int_0^u b_u du} c_s dWs\}$

2. **Autonomous** $\rightarrow dX(t) = f(X_t)dt + g(X_t)dW(t)$ $(f(X_t, t) \equiv f(X_t); g(X_t, t) \equiv g(X_t))$

   (a) **Linear** $\rightarrow dX(t) = (a + bX_t)dt + (c + eX_t)dW(t)$
   Solution: $X_t = exp\{e\int_0^t dW_s + (b - \frac{1}{2}e^2)\int_0^t ds\} \times \{X_0 + (a - ec)\int_0^t \Phi_s^{-1} ds + c\int_0^t \Phi_s^{-1} ds\}$, with $\Phi_s^{-1} = exp\{-e\int_0^s dW_u - (b - \frac{1}{2}e^2)\int_0^s du\}$

   (b) **Homogeneous** $\rightarrow dX(t) = bX_t dt + eX_t dW(t)$ $(a = c \equiv 0)$
   Solution: $X_t = X_0 exp\{\int_0^t b_s ds + \int_0^t e_s dW_s - \frac{1}{2}\int_0^t e_s^2 ds\}$

   (c) **Narrow sense** $\rightarrow dX(t) = (a + bX_t)dt + cdW(t)$ $(e \equiv 0)$
   Solution: $X_t = X_0 e^{bt} + a\int_0^t e^{b(t-s)} ds + c\int_0^t e^{b(t-s)} dW_s$

The solution to the most general case, i.e. $dX(t) = (a_t + b_t X_t)dt + (c_t + e_t X_t)dW(t)$, can be computed by presupposing X_t is the product of two well known stochastic processes $U_t = \alpha_t\, dt + \beta_t\, dW_t$, and $V_t = b_t V_t\, dt + e_t V_t\, dW_t$, applying the Itô formula to the product $U_t V_t$, developing the resulting equation, and find the expressions for α_t and β_t that satisfy the assumed equality, $dX_t = d(U_t\, V_t)$.

A solution X_t, with $X(t_0) = X_0$, is stochastically bounded if for each $\epsilon > 0$ there exists a $\gamma_\epsilon = \gamma_\epsilon(t_0, X_0)$ such that

$$\inf_{t \in [t_0, T]} P(|X_t| \leq \gamma_\epsilon) > 1 - \epsilon. \tag{3.19}$$

If γ_ϵ only depends on X_0, then X_t is termed uniformly stochastically bounded. A sufficient condition for stochastic boundedness of a process $\{X_t\}_{t \in [t_0, T]}$ is a $p > 0$ such that the pth-order moment $E|X_t|^p$ is bounded on the interval $[t_0, T]$.

Theorem 3.3 (Existence of Moments)**.** *Let the assumptions of theorem 3.1 hold, $E|X_0|^{2p} < +\infty$ for p integer and X_t a solution to SDE 3.1 on interval $[t_0, T]$, $T < +\infty$. Thus,*

1. $E|X_t|^{2p} \leq (1 + E(X_0)^{2p})e^{c(t - t_0)}$
2. $E|X_t - X_0|^{2p} \leq D(1 + E(X_0)^{2p})(t - t_0)^p e^{c(t-t_0)}$,

where $c = 2p(2p + 1)K^2$ and D constants dependent on p, k, $T - t_0$.

The two previous inequalities can be proved by the use of other well known inequalities, being the two most worthy of note among them, the Gronwall's inequality, and the assumed linear growth condition.

Theorem 3.4. *Let $p \geq 2$, $g(t) \in \mathcal{M}^2([0, T])$ and $E\int_0^T |g(s)|^p ds$. Then*

$$E|\int_0^t g(s)dW(s)|^p \leq (\frac{p(p-1)}{2})^{\frac{p}{2}} T^{\frac{p-2}{2}} E\int_0^T |g(s)|^p ds. \tag{3.20}$$

Let f, g satisfy the conditions stated on the theorem 3.1 and X_t be the solution of SDE 3.1. Then, X_t is continuous on $[t_0, T]$ if there is a constant $c > 0$ such that $|X_t - X_r|^2 \leq c|t - r|$, $t_0 \leq r, t \leq T$, where $X_t - X_r = \int_r^t f(X_s, s)ds + \int_r^t g(X_s, s)dW_S$ holds.

In some cases, the solution of a SDE is a Markov process, as follows.

Theorem 3.5. *Let X_t be a solution to SDE 3.1 with coefficient functions f, g satisfying the conditions of theorem 3.1. Then X_t is a Markov process with initial probability given by $P_0(B) = P(X_0 \in B)$, $B \in \mathcal{B}$, transition probability $P(x, s; B, t) = P(X_t(x, s) \in B)$ and $X_t = X_t(x, s)$ is a solution to $X_t = x + \int_s^t f(X_r, r)dr + \int_s^t g(X_r, r)dW_r$, $s \leq t$.*

If instead of the conditions of theorem 3.5 f, g are uniformly Lipschitz continuous with linear growth condition satisfied, then the solution X_t is called a strong Markov process.

When transition probabilities are stationary, $P(X_{t+\tau} \in B \mid \mathcal{A}_t) = P(X_\tau \in B \mid X_0)$, $0 \leq u \leq T - t$, a Markov process is said to be homogeneous.

When in the case of an autonomous SDE, if f, g satisfy the conditions of theorem 3.1 with stationary transition probabilities the solution X_t is a homogeneous Markov process. Again if f, g are uniformly Lipschitz continuous and satisfy the linear growth condition, stronger conditions, X_t is called homogeneous strong Markov process.

A solution of a SDE can be expressed as special stochastic process called Itô diffusion. A one-dimensional Markov process with continuous sample paths with probability 1 is a diffusion process if transition probabilities satisfy:

1. $lim_{t \downarrow s} \frac{1}{t-s} \int_{|y-x|<\epsilon} p(X_t = y \mid X_s = x)dy = 0$;
2. $lim_{t \downarrow s} \frac{1}{t-s} \int_{|y-x|<\epsilon} (y - x)p(X_t = y \mid X_s = x)dy = a(x, s)$;
3. $lim_{t \downarrow s} \frac{1}{t-s} \int_{|y-x|<\epsilon} (y - x)^2 p(X_t = y \mid X_s = x)dy = b(x, s)^2$,

with $a(x, s)$ representing the conditional infinitesimal mean and $b(x, s)^2$ the variance of the process.

Theorem 3.6. *Given a SDE of the form of 3.1, if f and g satisfy the conditions of theorem 3.1, then any solution X_t is a diffusion process on interval $[0, T]$ with drift $f(X_t, t)$ and diffusion $g(X_t, t)^2$.*

3.4 Stability

Due to some constraints in the representation of solutions of SDEs, its importance is more related to the qualitative information one can extract, rather than the presentation of the explicit solution.

A relevant aspect of solutions is the concept of stability, which evaluates the impact of a small change in the initial conditions on the solution of the SDE. When small changes in the initial conditions lead to small changes in the solution we are in the presence of a stable solution. When small changes lead to large changes in the solution, we have an unstable solution.

Considering equation 3.1, let (a) the assumptions of theorem 3.1 hold, (b) the initial value $X(t_0) = X_0 \in \mathcal{R}$ is constant with probability one, (c) equation 3.1 has a unique global solution $X_t(t_0, X_0)$ with continuous sample paths and finite moments for any X_0 independent of W_t, $t \geq t_0$ and (d) $f(0, t) = g(0, t) = 0$ such that the unique solution $X_t \equiv 0$ corresponds to the initial value $X(t_0) = X_0 = 0$. The trivial or equilibrium solution to SDE 3.1 is $X_t \equiv 0$.

It is important to state some concepts related to stability. A continuous scalar-valued function $v(x)$ defined on a suitably restricted neighbourhood of zero $N_h = \{x||x| < h, h > 0\}$ is positive definite if $v(0) = 0$ and $v(x) > 0$, $x \in N_h \; \forall \; x \neq 0$. A function v is negative definite if $-v$ is positive definite. Function v is descending if exists a positive definite function $u(x)$ such that $v(x, t) \leq u(x)$,

that is, function $u(x)$ is an upper bounded for $v(x,t)$. Function $v(x,t)$ is radially unbounded if $\lim_{|x|\to+\infty} \inf_{\{t\geq t_0\}} v(x,t) \to +\infty$.

Let $V_t = v(X_t,t)$ be a continuous positive definite function twice differentiable in X_t and once in t defined on $N_h \times \mathcal{R}$, $N_h = \{X_t | |X_t| < h,\ h > 0\}$. The operator L associated with SDE 3.1 is

$$L = \frac{\delta}{\delta t} + f(X_t,t)\frac{\delta}{\delta X_t} + \frac{1}{2}g(X_t,t)^2\frac{\delta^2}{\delta X_t^2}. \tag{3.21}$$

Applying L on V_t we get

$$Lv(X_t,t) = \frac{\delta v}{\delta t} + f(X_t,t)\frac{\delta v}{\delta X_t} + \frac{1}{2}g(X_t,t)^2\frac{\delta v^2}{\delta X_t^2}. \tag{3.22}$$

Using Itô formula, if $X_t \in N_h$ then

$$dv(X_t,t) = Lv(X_t,t)dt + g(X_t,t)\frac{\delta v}{\delta X_t}dW_t. \tag{3.23}$$

A stable solution to 3.1 requires that, on average, $E(dV_t) = E(Lv(X_t,t)dt) < 0$, which is satisfied if $Lv(X_t,t) \leq 0,\ \forall t \geq 0$. Thus, $V_t = v(X_t,t)$ represents the stochastic Lyapunov function for SDE 3.1, which allows the application of extended Lyapunov method to determine if the trivial solution is stable. The main issue with Lyapunov method is Lyapunov function, since it is often hard to find.

Assuming that conditions (a)-(d) are satisfied, the trivial or equilibrium solution is stochastically stable or stable in probability if

$$\lim_{X_0\to 0} P\Big(\sup_{t\in[t_0,+\infty)} |X_t(t_0,X_0)| \geq \varepsilon\Big) = 0, \tag{3.24}$$

otherwise the solution is called stochastically unstable.

Trivial solution is stochastically asymptotically stable if it is stochastically stable and

$$\lim_{X_0\to 0} P\Big(\lim_{t\to+\infty} X_t(t_0,X_0) = 0\Big) = 1. \tag{3.25}$$

Moreover, the trivial solution is globally stochastically asymptotically stable, or stochastically asymptotic stable in large if it is stochastically stable and $\forall X_0 \in \mathcal{R}$

$$P\Big(\lim_{t\to+\infty} X_t(t_0,X_0) = 0\Big) = 1. \tag{3.26}$$

Theorem 3.7. *1. Suppose assumptions (a)-(d) hold and there exists a positive definite stochastic Lyapunov function $v(X_t,t)$ defined on $N_h \times [t_0,+\infty)$ that is everywhere continuous, twice differentiable in X_t and once in t. Let $Lv(X_t,t) \leq 0,\ t \geq t_0,\ 0 \leq X_t \in N_h$, where $Lv(X_t,t)$ is given by equation 3.22. Then, the trivial solution is stochastically stable.*

2. *If, in addition, $v(X_t,t)$ is decreasing function and $Lv(X_t,t)$ is negative definite, then the trivial solution is stochastically asymptotically stable.*

3. *If point 2 hold for a radially bounded function $v(X_t,t)$ defined everywhere on $[t_0,+\infty) \times \mathcal{R}$, then the trivial solution is globally stochastically asymptotically stable.*

Example: Consider the following SDE: $dX_t = \mu X_t dt + \sigma X_t dW_t$. Let the selected stochastic Lyapunov function be $V_t = v(X_t,t) = X_t^2$. Then, applying equation 3.22,

$$\begin{aligned} Lv &= \Big(0 + \mu X_t(2X_t) + \frac{1}{2}(\sigma X_t)^2(2)\Big) \\ &= \Big(2\mu X_t^2 + \frac{1}{2}\sigma^2 X_t^2 2\Big) \\ &= 2\Big(\mu + \frac{1}{2}\sigma^2\Big)X_t^2 \\ &= 2\Big(\mu + \frac{1}{2}\sigma^2\Big)v_t \leq 0 \end{aligned}$$

if $\mu + \frac{1}{2}\sigma^2 < 0$. Hence, if the inequality holds, the trivial solution $X_t = 0$ is globally stochastically asymptotically stable, or stochastically asymptotic stable in large. ■

A. M. Lyapunov developed the stability concept for ODEs: please consult this author and others with relevant contribution in this topic to better understand the general idea for the case of ODEs.

Chapter 4

Stochastic Differential Equation models

This chapter intends to present some models based on Stochastic Differential Equations with applications on biological and financial areas. This models are derived from the simpler Ordinary Differential Equation models, with the introduction of a stochastic component.

4.1 Introduction

A deterministic model ignores the noise presented on the behaviour of a modeled variable. Thus, the stochastic component aims to address this issue, namely through the introduction of a Brownian motion process.

Models have a deterministic component that captures the growth rate of the dependent variable, measurable given the initial conditions, and a stochastic component that represents the variability provoked by random phenomena.

Regarding biology, some sources of noise are the demographic stochasticity, comprising all factors related to reproducible and mortality of each individual, environmental stochasticity for issues like weather, natural disasters and others, mensuration stochasticity originated from the misleading counting and, finally, informational stochasticity for all possible errors or omissions made during the process of modelling.

In the financial case, noise can be generated when a big institution/ investor program their trading, creating a suddenly increase in the volume of trade, or when a group of traders bet on the same industry or company originating a bubble.

The conversion of a non-linear SDE into a linear, by a changing of variable, constitutes a great advantage since the solution to a linear SDE is already known (please look to 3.3).

Given a non-linear SDE of the form

$$dX_t = f(X_t, t)dt + g(X_t, t)dW_t, \tag{4.1}$$

consider $Y_t = \mathcal{U}(X_t, t)$ with $\delta\mathcal{U}/\delta X_t \neq 0$. Then, the solution of 4.1 has the form $X_t = V(Y_t, t)$, provided the coefficients $a_1(t)$, $a_2(t)$, $b_1(t)$, $b_2(t)$ of

$$dY_t = (a_1(t) + a_2(t)Y_t)dt + (b_1(t) + b_2(t)Y_t)dW_t. \tag{4.2}$$

From the application of Itô formula to $\mathcal{U}(X_t, t)$, one can verify that the following expressions make 4.1 reducible.

1. $\mathcal{U}_t + f\mathcal{U}_X + \frac{1}{2}g^2\mathcal{U}_{XX} = a_1(t) + a_2(t)\mathcal{U}(X_t, t)$;
2. $g\mathcal{U}_X = b_1(t) + b_2(t)\mathcal{U}(X_t, t)$.

The next step uses the equalities above to determine the value of $a_1(t)$, $a_2(t)$, $b_1(t)$, $b_2(t)$, $\mathcal{U}(X_t,t)$ to further plug in the linear equation 4.2 to find the value of Y_t. Then, when Y_t is known, it just follows to compute X_t by the inverse operations.

For the special case of autonomous SDE, we have

1. $f\mathcal{U}_X + \frac{1}{2}g^2\mathcal{U}_{XX} = a_1 + a_2\mathcal{U}(X_t)$;
2. $g\mathcal{U}_X = b_1 + b_2\mathcal{U}(X_t)$.

If $g(X_t) \neq 0$ and $b_2 \neq 0$, then by 2 we have $\mathcal{U}(X_t) = Ce^{b_2\int_{x_0}^{X_s}\frac{ds}{g(s)}} - \frac{b_1}{b_2}$, with C and X_0 constants. Applying $\mathcal{U}(X_t)$ in 1, one obtains $C\Big(b_2A(X_t) + \frac{1}{2}b_2^2 - a_2\Big)e^{b_2\int_{x_0}^{X_s}\frac{ds}{g(s)}} = a_1 - a_2\Big(\frac{b_1}{b_2}\Big)$, with $A(X_t) = \frac{f(X_t)}{g(X_t)} - \frac{1}{2}\frac{dg(X_t)}{dX_t}$. Skipping computations, $b_2\frac{dA}{dX_t} + \frac{d}{dX_t}\Big(g(X_t)\frac{dA}{dX_t}\Big) = 0$ and it represents the necessary and sufficient condition to convert an autonomous non-linear SDE into an autonomous linear SDE. Moreover,

- if $b_2 \neq 0$, $\mathcal{U}(X_t) = e^{b_2\int_{x_0}^{X_s}\frac{ds}{g(s)}}$
- if $b_2 = 0$, $\mathcal{U}(X_t) = b_1\int_{x_0}^{X_t}\frac{ds}{g(s)} + C$.

It's mandatory to chose b_1 in a way that it satisfies 2. Given $\mathcal{U}(X_t)$, b_1, b_2 properly defined, 1 allows us to determine a_1, a_2. Substituting this coefficients in 4.2, the solution Y_t is direct. Finally, the solution X_t to the non-linear problem 4.1 is obtained from the application of the inverse operation on Y_t.

Example: To find the solution of the non-linear SDE

$$dX_t = -\frac{1}{2}e^{-2X_t}dt + e^{-X_t}dW_t, \tag{4.3}$$

with $f(X_t) = -\frac{1}{2}e^{-2X_t}$ and $g(X_t) = e^{-X_t}$, let's first evaluate $A(X_t)$,

$$A(X_t) = \frac{f(X_t)}{g(X_t)} - \frac{1}{2}\frac{dg(X_t)}{dX_t} = -\frac{1}{2}e^{-X_t} + \frac{1}{2}e^{-X_t} = 0. \tag{4.4}$$

The result 4.4 implies that

$$b_2\frac{dA}{dX_t} + \frac{d}{dX_t}\Big(g(X_t)\frac{dA}{dX_t}\Big) = 0 \tag{4.5}$$

holds for any b_2. For $b_2 = 0$ and $b_1 = 1$,

$$g\mathcal{U}_X = b_1 \Leftrightarrow e^{-X_t}\mathcal{U}_X = 1 \Leftrightarrow \mathcal{U}_X = e^{X_t} \Rightarrow \mathcal{U} = e^{X_t}. \tag{4.6}$$

Then, $f\mathcal{U}_X + \frac{1}{2}g^2\mathcal{U}_{XX} = a_1 + a_2\mathcal{U}(X_t)$ implies

$$-\frac{1}{2}e^{-2X_t}\cdot e^{X_t} + \frac{1}{2}\Big(e^{-X_t}\Big)^2e^{X_t} = a_1 + a_2e^{X_t} \Leftrightarrow 0 = a_1 + a_2e^{X_t}, \tag{4.7}$$

thus a_1, a_2 can assume any value. To simplify, let $a_1 = a_2 = 0$, check 4.2 and note that the linear SDE is

$$dY_t = dW_t \Rightarrow \int_0^t dY_s = \int_0^t dW_s \Leftrightarrow Y_t = Y_0 + W_t. \tag{4.8}$$

Since $Y_t = \mathcal{U}(X_t) = e^{X_t}$, then the solution to 4.3 is

$$Y_t = Y_0 + W_t \Leftrightarrow e^{X_t} = e^{X_0} + W_t \Leftrightarrow X_t = ln(e^{X_0} + W_t). \blacksquare \tag{4.9}$$

When $Y_t = \mathcal{U}(X_t)$ is known in advance, we resort to the second-order Taylor expansion of $\mathcal{U}$ to write $dY_t = d\mathcal{U}(X_t) = \mathcal{U}_X dX_t + \frac{1}{2}\mathcal{U}_{XX}(dX_t)^2$ and substitute dX_t by the corresponding SDE, which will lead to an autonomous linear SDE.

Example: If one is working on SDE 4.3 but $Y_t = \mathcal{U} = e^{X_t}$ is given, then

$$\begin{aligned} dY_t &= e^{X_t}dX_t + \frac{1}{2}e^{X_t}(dX_t)^2 \\ &= e^{X_t}\Big(-\frac{1}{2}e^{-2X_t}dt + e^{-X_t}dW_t\Big) \\ &+ \frac{1}{2}e^{X_t}\Big(-\frac{1}{2}e^{-2X_t}dt + e^{-X_t}dW_t\Big)^2 \\ &= -\frac{1}{2}e^{-X_t}dt + dW_t + \frac{1}{2}e^{X_t}(e^{-X_t})^2(dW_t)^2 \\ &= -\frac{1}{2}e^{-X_t}dt + dW_t + \frac{1}{2}e^{-X_t}dt \\ &= dW_t \end{aligned}$$

To finalize the answer, one must proceed as in the end of the previous example. ■

4.2 Deterministic and Stochastic Populational Growth Models

The deterministic population growth model is based on the ODE

$$dN_t = rN_t dt, \tag{4.10}$$

where r is a constant and represents the instantaneous relative growth rate that can be rewritten as the difference of birds (b) and deaths (d) for a single species with no migration, i.e. $r = b - d$, termed as the net bird rate or the constant per capita growth rate. The solution to 4.10 is

$$N(t) = N_0 e^{rt},\ N(0) = N_0, \tag{4.11}$$

so when $r > 0$ the population grows exponentially and when $r < 0$ the decreases exponentially.

If instead of considering relative population growth rate just as the difference between birds and deaths, one considers a new relative population growth rate defined as the net bird rate plus a noise component, in the form of $r(t) - \alpha(t) \cdot noise$ with $r,\ \alpha$ constants and $noise = \frac{dW_t}{dt}$, then the ODE is now as follows:

$$\frac{dN_t}{dt} = (r(t) - \alpha(t) \cdot \frac{dW_t}{dt})N_t,\ N(0) = N_0,\ t \geq 0 \tag{4.12}$$

or, alternatively,

$$dN_t = r(t)N_t dt - \alpha(t)N_t dW_t,\ N(0) = N_0,\ t \geq 0. \tag{4.13}$$

Noting that SDE 4.13 is autonomous, linear and homogeneous, the solution is a geometric Brownian motion process of the form $N(t) = N_0 e^{(r - \frac{1}{2}\alpha^2)t + \alpha W_t}$ with

1. $E(N(t)) = N_0 e^{rt}$;
2. $Var(N(t)) = N_0 e^{2rt}(e^{\alpha^2 t} - 1)$, if N_0 is independent of W_t.

In the long-term:

1. if $r > \frac{1}{2}\alpha^2$, then $N(t) \longrightarrow +\infty$ as $t \longrightarrow +\infty$;
2. if $r < \frac{1}{2}\alpha^2$, then $N(t) \longrightarrow -\infty$ as $t \longrightarrow +\infty$;
3. if $r = \frac{1}{2}\alpha^2$, then $N(t)$ fluctuates between arbitrarily large and small small values as $t \longrightarrow +\infty$.

4.3 Deterministic and Stochastic Logistic Growth Models

One possible critic addressed to the Stochastic Population Growth model is the possibility of populations to expand without limit.

To fix this issue, Verhulst proposed a self-regulating mechanism that ponders the population relative growth rate according to the current size of the population. The new r is given by $r\Big(1-\frac{N(t)}{K}\Big)$, where K represents the limiting size of the population or carrying capacity. Thus, the deterministic Logistic Growth Model is

$$dN(t) = r\Big(1-\frac{N(t)}{K}\Big)N(t)dt, \tag{4.14}$$

with solution

$$N(t) = \frac{KN_0}{N_0+(K-N_0)e^{-rt}},\ N(0) = N_0,^t \geq 0. \tag{4.15}$$

If births are defined as $b = b_1 - b_2N(t)$, deaths as $d = d_1 + d_2N(t)$ and r as the difference between births and deaths, then

$$\begin{aligned}\frac{dN_t}{dt} &= \Big(b_1 - b_2N(t) - d_1 - d_2N(t)\Big)N(t) \\ &= \Big((b_1-d_1) - (b_2+d_2)N(t)\Big)N(t) \\ &= (b_1-d_1)\Big(1-\frac{N(t)}{(b_1-d_1)/(b_2+d_2)}\Big)N(t).\end{aligned} \tag{4.16}$$

Hence, $K = (b_1-d_1)/(b_2+d_2)$.

The stochastic Logistic Growth Model appears with the introduction of the multiplicative noise term, originating

$$dN(t) = r\Big(1-\frac{N(t)}{K}\Big)N(t)dt + \alpha N(t)dW_t, \tag{4.17}$$

A strategy similar to the used to convert deterministic population growth model into a stochastic one may also be applied here.

The solution to the stochastic logistic growth model is obtained resorting to the Itô equation, considering $V(N_t,t) = N_t^{-1}$ and $V_N = -N_t^{-2}$, $V_{NN} = 2N_t^{-3}$, $f = r(1-\frac{N_t}{K})N_t$, $g = \alpha N_t$. Following the procedure of Itô formula, one may arrive to the solution

$$d(N_t^{-1}) = \Big\{r\Big(-N_t^{-1}+\frac{1}{k}\Big)+\alpha^2N_t^{-1}\Big\}dt - \alpha N_t^{-1}dW_t. \tag{4.18}$$

The previous solution can be expressed as:

$$N_t = \frac{exp\{(r-(1/2)\alpha^2)t+\alpha W_t\}}{N_0^{-1}+(r/K)\int_0^t exp\{(r-(1/2)\alpha^2)s+\alpha W_s\}ds}. \tag{4.19}$$

The mean and variance of the process N_t are, respectively,:

1. $E(N(t)) = \frac{K}{1+[(K/N_0)-1]e^{-rt}}$;
2. $Var(N(t)) = K^2e^{2rt}\Big\{\frac{K}{N_0}+(e^{rt}-1)\Big\}^{-2}(e^{\alpha^2t}-1)$.

4.4 Deterministic and Stochastic Generalized Logistic Growth Models

The Generalized Logistic Growth model is obtained considering the a new growth equation to the logistic growth model. Hence, the deterministic Generalized Logistic Growth Model is

$$\frac{dN_t}{dt} = \beta N_t\Big[1 - \Big(\frac{N_t}{N_\infty}\Big)^r\Big]. \tag{4.20}$$

N_∞ denotes the horizontal asymptote when $t \longrightarrow \infty$, termed **saturation level**. β is a **feedback term** that ponders growth rate.

The solution to 4.20 is

$$N_t = N_\infty\Big[1 + \Big(\Big(\frac{N_0}{N_\infty}\Big)^{-r} - 1\Big)e^{-\beta r t}\Big]^{-\frac{1}{r}}. \tag{4.21}$$

The stochastic Generalized Logistic Growth model is obtained introducing to 4.20 the multiplicative noise term. The SDE is

$$dN_t = \beta N_t\Big[1 - \Big(\frac{N_t}{N_\infty}\Big)^r\Big]dt + \alpha N_t dW_t. \tag{4.22}$$

To solve SDE 4.22 one must start by diving all terms by N_∞, defining $Y_t = \frac{N_t}{N_\infty}$ and observe that 4.20 the multiplicative noise term. The SDE is

$$dY_t = \beta Y_t(1 - Y_t^r)dt + \alpha Y_t dW_t \tag{4.23}$$

is an autonomous non-linear SDE. To proceed, one should apply the reduction method presented in subsection 4.1, starting by determining a suitable transformation function $Z_t = \mathcal{U}(Y_t)$ and then using the results presented in 3.3.

We have $f = \beta Y_t(1 - Y_t^r)$ and $g = \alpha Y_t$, so let's proceed to find $\mathcal{U}(Y_t)$.

$$A(Y_t) = \frac{\beta Y_t(1 - Y_t^r)}{\alpha Y_t} - \frac{1}{2}\alpha = \frac{\beta(1 - Y_t^r)}{\alpha} - \frac{\alpha}{2} \neq 0 \tag{4.24}$$

$$b_2\Big(- (\beta/\alpha) r Y_t^{r-1}\Big) - r^2\beta Y_t^{r-1} = 0 \Leftrightarrow b_2 = -r\alpha. \tag{4.25}$$

Hence,

$$\mathcal{U}(Y_t) = Ce^{-r\alpha\int_{Y_0}^{Y_t}\frac{1}{\alpha Y_t}ds} = \mathcal{U}(Y_t) = Ce^{-r\alpha\frac{1}{\alpha}ln(Y_t - Y_0)} = Ce^{-rln(Y_t/Y_0)} = Y_t^{-r}, \tag{4.26}$$

for C, Y_0 arbitrary.

Now let´s find $a_1,\ a_2,\ b_1,\ b_2$.

$$\alpha Y_t\Big(- rY_t^{-r-1}\Big) = b_1 - r\alpha Y_t^{-r} \Leftrightarrow b_1 = 0, \tag{4.27}$$

$$\beta Y_t(1 - Y_t^r)\cdot\Big(- rY_t^{-r-1}\Big) + \frac{1}{2}\Big(\alpha Y_t\Big)^2\cdot(-r)(-r-1)Y_t^{-r-2} = a_1(t) + a_2(t)Y_t^{-r} \tag{4.28}$$

$$\Longleftrightarrow$$

$$r\beta + \Big(\frac{1}{2}r(r+1)\alpha^2 - r\beta\Big)Y_t^{-r} = a_1(t) + a_2(t)Y_t^{-r}, \tag{4.29}$$

so $a_1 = r\beta$ and $a_2 = \frac{1}{2}r(r+1)\alpha^2 - r\beta$.

Let $Z_t = Y_t^{-r}$. The linear SDE is

$$dZ_t = \Big(r\beta + \Big(\frac{1}{2}r(r+1)\alpha^2 - r\beta\Big)Z_t\Big)dt - r\alpha Z_t dW_t. \tag{4.30}$$

Resorting to 3.3 and inverting the change of variables, the final solution to 4.22 is achieved. This solution is

$$N_t = N_\infty\Phi_t^{-\frac{1}{r}}\Big\{\Big(\frac{N_0}{N_\infty}\Big)^{-r} + \beta r\int_0^t \Phi_s^{-1}ds\Big\}^{-\frac{1}{r}}, \tag{4.31}$$

with $\Phi_t = exp\Big\{ - \alpha r W_t + \Big(\frac{1}{2}\alpha^2 r - \beta r\Big)t\Big\}$.

4.5 Deterministic and Stochastic Gompertz Growth Models

If the instantaneous rate of growth is proportional to $lnN_\infty - lnN_t$ we have

$$dN_t = \beta(lnN_\infty - lnN_t)N_t dt. \tag{4.32}$$

The logarithm presents a vertical asymptote at $x = 0$ and the growth rate has the shape of an inverse logarithm function.

The solution to ODE 4.32 is the deterministic Gompertz population growth model

$$N_t = N_\infty e^{-ln(N_\infty/N_t)e^{-\beta t}},\ N(0) = N_0,\ t \geq 0. \tag{4.33}$$

Considering $X_t = \frac{N_t}{N_\infty}$, one can write ODE 4.32 as

$$\frac{dN_t}{N_\infty} = \beta(ln\frac{N_\infty}{N_t})\frac{N_t}{N_\infty}dt \Leftrightarrow dX_t = \beta ln(X_t^{-1})X_t dt \Leftrightarrow dX_t = -\beta X_t lnX_t dt. \tag{4.34}$$

By introducing a noise term, the stochastic model arise, as presented below,

$$dX_t = -\beta X_t lnX_t dt + kX_t dW_t, \tag{4.35}$$

which is an autonomous non-linear SDE.

Once again, the method applied to find the solution of 4.35 is the reduction model presented in 4.1. First, let's find the transformation function, given that $f = -\beta X_t lnX_t$ and $g = kX_t$:

$$A(X_t) = \frac{-\beta X_t}{kX_t}lnX_t - \frac{1}{2}k = \frac{-\beta}{k}lnX_t - \frac{k}{2}, \tag{4.36}$$

$$b_2\Big(-\frac{\beta}{kX_t}\Big) + \frac{d}{dX_t}\Big(kX_t\frac{-\beta}{kX_t}\Big) = 0 \Leftrightarrow b_2\Big(-\frac{\beta}{kX_t}\Big) + 0 = 0 \Leftrightarrow b_2 = 0, \tag{4.37}$$

$$B(X_t) = \int_{X_0}^{X_t}\frac{1}{g(s)}ds = \int_{X_0}^{X_t}\frac{1}{kX_s}ds = \frac{1}{k}lnX_t, \tag{4.38}$$

$$\mathcal{U}(X_t) = b_1 B(X_t) + C = \frac{b_1}{k}lnX_t + C, \tag{4.39}$$

$$g\mathcal{U}_X = b_1 + b_2\mathcal{U}(X_t) \Longleftrightarrow kX_t\frac{b_1}{kX_t} = b_1 + 0 \Longleftrightarrow b_1 = b_1, \tag{4.40}$$

so that b_1 can assume any value. Thus, with $b_1 = k$ and $C = 0$, we have

$$\mathcal{U}(X_t) = \frac{k}{k}lnX_t + 0 = lnX_t. \tag{4.41}$$

To find a_1, a_2 we use

$$f\mathcal{U}_X + \frac{1}{2}g^2\mathcal{U}_{XX} = -\frac{\beta X_t lnX_t}{X_t} - \frac{k^2X_t^2}{2X_t^2} = -\beta lnX_t - \frac{k^2}{2} = -\beta Z_t - \frac{k^2}{2}, \tag{4.42}$$

with $Z_t = \mathcal{U}(X_t) = lnX_t$.

The reduced form of 4.35 is

$$dZ_t = \Big(-\frac{1}{2}k^2 - \beta Z_t\Big)dt + kdW_t, \tag{4.43}$$

whose solution is given by 3.3, since now we are working with an autonomous linear SDE. After inverting the change of variables, the final solution to 4.35 is the stochastic Gompertz growth function

$$X_t = exp\Big\{e^{-\beta t}\Big[lnX_0 - \frac{1}{2\beta}k^2(e^{\beta t} - 1) + k\int_0^t e^{\beta s}dW_s\Big]\Big\}. \tag{4.44}$$

4.6 Deterministic and Stochastic Negative Exponential Growth Models

The deterministic negative exponential model is

$$dN_t = \beta\Big(\frac{N_\infty}{N_t} - 1\Big)N_t dt, \tag{4.45}$$

where $\frac{N_\infty}{N_0} - 1$ is called **feedback term** because it compares the limit level of the variable N with its current state, returning a decrease when the variable is above the limit ($N_t > N_\infty$) or supporting an increase when this value is below the limit ($N_t < N_\infty$). β represents the growth rate.

The representation of the evolution of a variable under the deterministic negative exponential model is given by

$$N_t = N_\infty\Big(1 - e^{-\beta t}\Big),\ t \geq 0,\ N_\infty, \beta > 0. \tag{4.46}$$

Considering noise as αN_t, the stochastic negative exponential growth model is

$$dN_t = \beta(N_\infty - N_t)dt + \alpha N_t dW_t, \tag{4.47}$$

which, fortunately, is an autonomous linear SDE, so its solution is direct.

As in 3.3, the solution to SDE 4.47 is

$$N_t = exp\Big\{\alpha W_t - \Big(\beta + \frac{1}{2}\alpha^2\Big)t\Big\} \times \Big\{N_0 + \beta N_\infty \int_0^t exp\Big\{\Big(\beta + \frac{1}{2}\alpha^2\Big)s - \alpha W_s\Big\}ds\Big\}. \tag{4.48}$$

4.7 Deterministic and Stochastic Linear Growth Models

A deterministic linear growth model is characterized by the ODE

$$dN_t = \beta dt, \tag{4.49}$$

with solution

$$N_t = N_0 + \beta t,\ t \geq 0. \tag{4.50}$$

The addition of a multiplicative noise term generates the SDE

$$dN_t = \beta dt + \alpha N_t dW_t, \tag{4.51}$$

the general form of a stochastic linear growth model with multiplicative noise. This SDE is autonomous and linear, with parameter $a = \beta,\ b = c = 0$ and $e = \alpha$, and its solution is

$$N_t = exp\Big\{\alpha W_t - \frac{1}{2}\alpha^2 t\Big\} \times \Big\{N_0 + \beta \int_0^t exp\Big(\frac{1}{2}\alpha^2 s - \alpha W_s\Big)ds\Big\}, \tag{4.52}$$

as defined on 3.3.

If instead of the multiplicative noise one introduces the additive noise, the SDE becomes

$$dN_t = \beta dt + \alpha dW_t, \tag{4.53}$$

an autonomous and linear SDE in the narrow sense ($a = \beta,\ b = e = 0,\ c = \alpha$) with solution

$$N_t = X_0 + \beta \int_0^t ds + \alpha \int_0^t dW_s = X_0 + \beta t + \alpha W_t. \tag{4.54}$$

4.8 Stochastic Square-Root Growth Model with Mean Reversion

The mean-reverting square-root SDE is defined as follows:

$$dX_t = (\alpha - \beta X_t)dt + \sigma\sqrt{X_t}dW_t,\ X(0) = X_0 > 0. \tag{4.55}$$

This model is famous as the Cox-Ingersoll-Ross (CIR) model to the dynamics of short-run interest rates. In that sense, α is the speed of adjustment and $\beta, \sigma > 0$. The drift $\alpha - \beta X_t$ ensures mean revertion and the diffusion $\sigma\sqrt{X_t}$ describes volatility.

Given an integrating factor $e^{\beta t}$, if one rewrites 4.55 as $dX_t + \beta X_t dt = \alpha dt + \sigma\sqrt{X_t}dW_t$, then

$$e^{\beta t}dX_t + \beta e^{\beta t}X_t dt = \alpha e^{\beta t}dt + \sigma e^{\beta t}\sqrt{X_t}dW_t \Leftrightarrow d(e^{\beta t}X_t) = \alpha e^{\beta t}dt + \sigma e^{\beta t}\sqrt{X_t}dW_t. \tag{4.56}$$

Integrating 4.56, one obtains

$$\int_0^t d(e^{\beta s}X_s) = \int_0^t \alpha e^{\beta s}ds + \int_0^t \sigma e^{\beta s}\sqrt{X_s}dW_s \tag{4.57}$$

$$\Leftrightarrow$$

$$e^{\beta t}X_t - e^{\beta\cdot 0}X_0 = \alpha\left(\frac{e^{\beta t}}{\beta} - \frac{e^{\beta\cdot 0}}{\beta}\right) + \sigma\int_0^t e^{\beta u}\sqrt{X_u}dW_u \tag{4.58}$$

$$\Leftrightarrow$$

$$e^{\beta t}X_t = X_0 + \frac{\alpha}{\beta}\left(e^{\beta t} - 1\right) + \sigma\int_0^t e^{\beta u}\sqrt{X_u}dW_u \tag{4.59}$$

$$\Leftrightarrow$$

$$X_t = X_0 e^{-\beta t} + \frac{\alpha}{\beta}e^{-\beta t}\left(e^{\beta t} - 1\right) + \sigma e^{-\beta t}\int_0^t e^{\beta u}\sqrt{X_u}dW_u \tag{4.60}$$

$$\Leftrightarrow$$

$$X_t = \frac{\alpha}{\beta} + e^{-\beta t}\left(X_0 - \frac{\alpha}{\beta}\right) + \sigma e^{-\beta t}\int_0^t e^{\beta u}\sqrt{X_u}dW_u, \tag{4.61}$$

the stochastic mean-reverting square-root growth equation or CIR growth equation.

Due to its application to interest rates, mean and variance must be highlighted.

1. $E[X_t] = E\left[\frac{\alpha}{\beta} + e^{-\beta t}\left(X_0 - \frac{\alpha}{\beta}\right) + \sigma e^{-\beta t}\int_0^t e^{\beta u}\sqrt{X_u}dW_u\right] = \frac{\alpha}{\beta} + e^{-\beta t}\left(X_0 - \frac{\alpha}{\beta}\right)$;
2. $Var[X_t] = E[(X_t - E(X_t))^2] = E\left[\left(\sigma e^{-\beta t}\int_0^t e^{\beta u}\sqrt{X_u}dW_u\right)^2\right] = \sigma^2 e^{-2\beta t}E\left[\left(\int_0^t e^{\beta u}\sqrt{X_u}dW_u\right)^2\right] = \sigma^2 e^{-2\beta t}E\left[\int_0^t e^{2\beta u}X_u du\right] = \sigma^2 e^{-2\beta t}\int_0^t e^{2\beta u}E(X_u)du = \sigma^2 e^{-2\beta t}\int_0^t e^{2\beta u}\left(\frac{\alpha}{\beta} + e^{-\beta t}\left(X_0 - \frac{\alpha}{\beta}\right)\right)du = X_0\left[\frac{\sigma^2(e^{-\beta t} - e^{2\beta t})}{\beta}\right] + \frac{\alpha\sigma^2(1 - e^{-2\beta t})}{2\beta^2}$.

CIR model in 4.55 was subject to an extension, defined as CKLS mean-reverting gamma SDE in honour to the authors, and is defined as

$$dX_t = (\alpha - \beta X_t)dt + \sigma X_t^{\gamma}dW_t,\ X(0) = X_0 > 0, \tag{4.62}$$

where γ measures the degree of non-linearity of X_t and its volatility.

Chapter 5

Approximation and Estimation of Solutions to Stochastic Differential Equations

5.1 Introduction

In a lot of cases, it is not possible to determine a close form solution to the Itô initial value problem, given by the stochastic differential equation:

$$dX_t = f(t, X_t)\,dt + g(t, X_t)\,dW_t,\ X(t_0) = X_0,\ t \in [t_0, +\infty). \tag{5.1}$$

Hence, similar to the deterministic differential equations case, numerical methods are employed. We start by discretizing the interval $[t_0, T]$ into N smaller subintervals $t_0 < t_1 < \cdots < t_N = T$, then using a normal pseudorandom number generator, compute increments for $\Delta W_i = W(t_{i+1}) - W(t_i) \sim N(0, \Delta t_i) = \sqrt{\Delta t_i} N(0,1)$, where $\Delta t_i = t_{i+1} - t_i$, and applying a recursive algorithm to calculate the approximations Y_i of $X(t_i; X_0, t_0)$, where $Y_0 = X_0$, for each $i = 0, 1, \cdots, N$.

Definition 5.1 (Weak and Strong Convergence)**.** Let us consider a discretization t_i of $[t_0, T]$ into N subintervals of equal length, i.e. $t_i = \frac{i}{N}(T - t_0) + t_0$, hence $\Delta t_i = \frac{T - t_0}{N}$, for $i = 0, 1, \ldots, N$.

1. A time discretization approximation Y converges with strong order γ to a continuous-time stochastic process X at time T, if there exists a constant C, independent of Δt, where for N large enough such that $\Delta t < 1$, we have:

$$\mathbb{E}[|X(T) - Y(T)|] \leq C(\Delta t)^{\gamma} \tag{5.2}$$

2. A time discretization approximation Y converges with weak order β to a continuous-time stochastic process X at time T, if there exists a continuously differentiable polynomial function h and a constant C_h, independent of Δt, where for N large enough such that $\Delta t < 1$, we have:

$$|\mathbb{E}[h(X(T))] - \mathbb{E}[h(Y(T))]| \leq C_k(\Delta t)^{\beta}. \tag{5.3}$$

5.2 Taylor Expansions

For the deterministic case, the approximation of a certain continuous process of the form $\frac{dX_t}{dt} = f(X_t)$ can be calculated by the application of the deterministic Taylor formula until a certain r-term, resulting in the approximation $h(X_t) \approx h(X_{t_0}) + \sum_{k=1}^{r} \frac{(t - t_0)^k}{k!} Lh(X_0)$, where L is the operator $L = f\frac{\partial}{\partial X}$. Analogously, for the stochastic case where $f(t, X_t) = f(X_t)$ and $g(t, X_t) = g(X_t)$ (that is $dX_t = f(X_t)\,dt + g(X_t)\,dW_t$), and f and g are sufficiently smooth real-value functions, satisfying the linear growth condition, then

any $h(X_t) : \mathbb{R} \to \mathbb{R}$ twice continuously differentiable, can be expressed as

$$h(X_t) = h(X_{t_0}) + \int_{t_0}^{t} \left[f(X_s)\frac{\partial}{\partial X}h(X_s) + \frac{1}{2}g(X_s)^2\frac{\partial^2}{\partial X^2}h(X_s) \right] ds + \int_{t_0}^{t} g(X_s)\frac{\partial}{\partial X}h(X_s)\, dW_s,\ t \in [t_0, T] \tag{5.4}$$

or as

$$h(X_t) = h(X_{t_0}) + \int_{t_0}^{t} L^0 h(X_s)\, ds + \int_{t_0}^{t} L^1 h(X_s)\, dW_s \tag{5.5}$$

where

$$L^0 = f\frac{\partial}{\partial X} + \frac{1}{2}g^2\frac{\partial^2}{\partial X^2}, \quad L^1 g\frac{\partial}{\partial X} \tag{5.6}$$

are the operators.

If we apply this formula to the functions f and g within the formula we obtain the Itô-Taylor expansion

$$X_t = X_{t_0} + f(X_t)(t - t_0) + g(X_t)(W_t - W_{t_0}) + R_t, \tag{5.7}$$

where the remainder term R_t is

$$R_t = \int_{t_0}^{t}\int_{t_0}^{s} L^0 f(X_z)\, dz\, ds + \int_{t_0}^{t}\int_{t_0}^{s} L^1 f(X_z)\, dW_z\, ds + \int_{t_0}^{t}\int_{t_0}^{s} L^0 g(X_z)\, dz\, dW_s + \int_{t_0}^{t}\int_{t_0}^{s} L^1 g(X_z)\, dW_z\, dW_s. \tag{5.8}$$

Note that, in the context of numerical methods, we can continuously apply the Itô-Taylor expansion and getting better algorithms for approximation of a stochastic X_t, with each subsequent expansion.

5.3 Iterative Schemes for Approximating SDE's

5.3.1 The Euler-Maruyama Approximation

The Euler-Maruyama approximation for SDEs can either be viewed as the generalization of the Euler method for ODEs to the SDE case, or as the first term of the Itô-Taylor expansion. Given a time discretization of $[t_0, T]$, $t_0 < t_1 < \cdots < t_N = T$, the EM approximation for a stochastic process X_t, is derived from

$$\begin{aligned} X_{t_{i+1}} &= X_{t_i} + \int_{t_i}^{t_{i+1}} f(X_s, s)\, ds + \int_{t_i}^{t_{i+1}} g(X_s, s)\, dW_s \\ &\approx X_{t_i} + f(X_{t_i}, t_i)\int_{t_i}^{t_{i+1}} ds + g(X_{t_i}, t_i)\int_{t_i}^{t_{i+1}} dW_s \\ &= X_{t_i} + f(X_{t_i}, t_i)\Delta t_i + g(X_{t_i}, t_i)\Delta W_{t_i}. \end{aligned} \tag{5.9}$$

Thus, the Euler-Maruyama approximation is defined as a continuous-time stochastic process $\{Y_t\}_{t\in[t_0,T]}$ that satisfies the iterative scheme:

$$Y_{i+1} = Y_i + f(Y_i, t_i)\Delta t + g(Y_i, t_i)\sqrt{\Delta t}Z, \tag{5.10}$$

where $Z \sim N(0, 1)$. The EM routine converges with strong order $\gamma = 1/2$ and weak order $\beta = 1$.

5.3.2 The Milstein (Second-Order) Approximation

As the name implies, the Milstein Approximation is derived from the second order Itô-Taylor expansion. Given a time discretization of $[t_0, T]$, $t_0 < t_1 < \cdots < t_n = T$, the Milstein approximation for a stochastic process X_t, is derived from:

$$\begin{aligned} X_{t_{i+1}} &= X_{t_i} + \int_{t_i}^{t_{i+1}} f(X_s, s)\, ds + \int_{t_i}^{t_{i+1}} g(X_s, s)\, dW_s + \int_{t_i}^{t_{i+1}} \int_{t_i}^{s} g(X_z, z) \frac{\partial g}{\partial X_z}(X_z, z)\, dW_z\, dW_s + R_t \quad (5.11) \\ &\approx X_{t_i} + f(X_{t_i}, t_i) \int_{t_i}^{t_{i+1}} ds + g(X_{t_i}, t_i) \int_{t_i}^{t_{i+1}} dW_s + g(X_{t_i}, t_i) \frac{\partial g}{\partial X_{t_i}}(X_{t_i}, t_i) \int_{t_i}^{t_{i+1}} \int_{t_i}^{s} dW_z\, dW_s \\ &= X_{t_i} + f(X_{t_i}, t_i)\Delta t_i + g(X_{t_i}, t_i)\Delta W_{t_i} + g(X_{t_i}, t_i) \frac{\partial g}{\partial X_{t_i}}(X_{t_i}, t_i) \frac{1}{2}((\Delta W_{t_i})^2 - \Delta t_i) \end{aligned}$$

Hence, the Milstein approximation is defined as a continuous-time stochastic process $\{Y_t\}_{t\in[t_0,T]}$ satisfying the iterative scheme:

$$Y_{i+1} = Y_i + f(Y_i, t_i)\Delta t + g(Y_i, t_i)\sqrt{\Delta t} Z + \frac{1}{2} g(Y_i, t_i) \frac{\partial g}{\partial Y_i}(X_{t_i}, t_i) \left(\Delta t Z^2 - \Delta t\right), \quad (5.12)$$

where $Z \sim N(0, 1)$. The Milstein routine converges with strong order $\gamma = 1$.

Example

Given the SDE for the Geometric BM model

$$dX_t = \mu X_t\, dt + \sigma X_t\, dW_t,\ X(0) = X_0$$

consider a n-partition $0 = t_0 \leq t_1 \leq \cdots \leq t_n = T$ of $[0, T]$. The EM and Milstein iteration schemes for approximating the geometric BM are given by

- *EM:* $Y_{t_{i+1}} = Y_{t_i} + \mu Y_{t_i}\Delta t_i + \sigma Y_{t_i}\sqrt{\Delta t_i} Z,\ i = 0, 1, ..., n-1,$
- *Milstein:* $Y_{t_{i+1}} = Y_{t_i} + \mu Y_{t_i}\Delta t_i + \sigma Y_{t_i}\sqrt{\Delta t_i} Z + \frac{1}{2}\sigma^2\Delta(Z^2 - 1),\ i = 0, 1, ..., n-1,$

where $Z \sim N(0, 1)$.

If we use the Geometric BM model SDE in its logarithmic form,

$$d\log(X_t) = \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma\, dW_t,$$

then EM and Milstein iteration schemes become identical with expression

$$Y_{t_{i+1}} = Y_{t_i} \exp\left[\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t_i + \sigma\sqrt{\Delta t_i} Z\right], \quad i = 0, 1, ..., n-1,$$

again with $Z \sim N(0, 1)$. ■

5.3.3 Variations on the EM and Milstein Schemes

As we stated previously, the Itô-Taylor expansion can be expanded to derive other iterative methods aside from the EM and Milstein schemes. We now present 5 iteration techniques for an approximation Y_t for SDEs of the form $dX_t = f(X_t)\, dt + g(X_t)\, dW_t$.

1. **Strong Order 1.5 Taylor Scheme** is given by

$$\begin{aligned} Y_{t_{i+1}} &= Y_{t_i} + f(Y_{t_i})\Delta t_i + g(Y_{t_i})\Delta W_{t_i} + \frac{1}{2}g(Y_t)g_X(Y_t)\left((\Delta W_{t_i})^2 - \Delta t_i\right) \\ &+ \frac{1}{2}f_X(Y_{t_i})g(Y_{t_i})\Delta t_i\left(\Delta W_{t_i} + \frac{\Delta V_{t_i}}{\sqrt{3}}\right) \\ &+ \frac{1}{2}\left(f(Y_{t_i})f_X(Y_{t_i}) + \frac{1}{2}(g(Y_{t_i}))^2 f_{XX}(Y_{t_i})\right)(\Delta t_i)^2 \\ &+ \frac{1}{2}\left(f(Y_{t_i})g_X(Y_{t_i}) + \frac{1}{2}(g(Y_{t_i}))^2 g_{XX}(Y_{t_i})\right)\Delta t_i\left(\Delta W_{t_i} - \frac{\Delta V_{t_i}}{\sqrt{3}}\right) \\ &+ \frac{1}{2}g(Y_{t_i})\left(g(Y_{t_i})g_{XX}(Y_{t_i}) + (g(Y_{t_i}))^2\right)\left(\frac{1}{3}(\Delta W_{t_i})^2 - \Delta t_i\right)\Delta W_{t_i}, \end{aligned} \tag{5.13}$$

   where $\Delta W_{t_i}, \Delta V_{t_i} \sim N(0, \Delta t_i)$ are independent random variables, for all $i = 1, \ldots, N$.

2. **Weak Order 2.0 Taylor Scheme** is defined as

$$\begin{aligned} Y_{t_{i+1}} &= Y_{t_i} + f(Y_{t_i})\Delta t_i + g(Y_{t_i})\Delta W_{t_i} + \frac{1}{2}g(Y_t)g_X(Y_t)\left((\Delta W_{t_i})^2 - \Delta t_i\right) \\ &+ \frac{1}{2}f_X(Y_{t_i})g(Y_{t_i})\Delta t_i\left(\Delta W_{t_i} + \frac{\Delta V_{t_i}}{\sqrt{3}}\right) \\ &+ \frac{1}{2}\left(f(Y_{t_i})f_X(Y_{t_i}) + \frac{1}{2}(g(Y_{t_i}))^2 f_{XX}(Y_{t_i})\right)(\Delta t_i)^2 \\ &+ \frac{1}{2}\left(f(Y_{t_i})g_X(Y_{t_i}) + \frac{1}{2}(g(Y_{t_i}))^2 g_{XX}(Y_{t_i})\right)\Delta t_i\left(\Delta W_{t_i} - \frac{\Delta V_{t_i}}{\sqrt{3}}\right), \end{aligned} \tag{5.14}$$

   where $\Delta W_{t_i}, \Delta V_{t_i} \sim N(0, \Delta t_i)$ are independent random variables for all $i = 1, \ldots, N$.

3. **Weak Order 2.0 Predictor-Corrector Scheme** is divided into two steps

   - *Step 1:* Consider a predictor $\overline{Y}_{t_i}$ given by the scheme

$$\begin{aligned} \overline{Y}_{t_{i+1}} &= Y_{t_i} + f(Y_{t_i})\Delta t_i + \lambda_i + \frac{1}{2}f_X(Y_t)g(Y_t)\left(\Delta\hat{W}_{t_i}\right)\Delta t_i \\ &+ \frac{1}{2}\left(f(Y_{t_i})f_X(Y_{t_i}) + \frac{1}{2}(g(Y_{t_i}))^2 f_{XX}(Y_{t_i})\right)(\Delta t_i)^2, \end{aligned} \tag{5.15}$$

     where

$$\begin{aligned} \lambda_i &= g(Y_{t_i})\Delta\hat{W}_{t_i} + \frac{1}{2}g(Y_t)g_X(Y_t)\left((\Delta\hat{W}_{t_i})^2 - \Delta t_i\right) \\ &+ \frac{1}{2}\left(f(Y_{t_i})g_X(Y_{t_i}) + \frac{1}{2}(g(Y_{t_i}))^2 g_{XX}(Y_{t_i})\right)\Delta\hat{W}_{t_i}\Delta t_i, \end{aligned} \tag{5.16}$$

     and $\Delta W_{t_i} \sim N(0, \Delta t_i)$.

   - *Step 2:* We define the corrector Y_{t_i} as

$$Y_{t_{i+1}} = Y_{t_i} + \frac{1}{2}\left[\overline{Y}_{t_{i+1}} + f(Y_{t_i})\right]\Delta t_i + \lambda_i \tag{5.17}$$

4. **Derivative-Free Weak Order 2.0 Predictor-Corrector Scheme** is defined in two steps

   - *Step 1:* We define the predictor $\overline{Y}_{t_i}$

$$\overline{Y}_{t_{i+1}} = Y_{t_i} + \frac{1}{2}\left[f(u_i) + f(Y_{t_i})\right]\Delta t_i + O_i \tag{5.18}$$

where the supporting values u_i and O_i are defined by

$$\begin{aligned} u_i &= Y_{t_i} + f(Y_{t_i})\Delta t_i + g(Y_{t_i})\Delta \hat{W}_{t_i} \\ O_i &= \frac{1}{4}[g(u^+) + g(u^-) + 2g(Y_{t_i})]\Delta \hat{W}_{t_i} \\ &\quad + \frac{1}{4}[g(u^+) + g(u^-)]\left[(\Delta \hat{W}_{t_i})^2 - \Delta t_i\right](\Delta t_i)^{-\frac{1}{2}}, \end{aligned} \tag{5.19}$$

where

$$\begin{aligned} u_i^+ &= Y_{t_i} + f(Y_{t_i})\Delta t_i + g(Y_{t_i})\sqrt{\Delta t_i} \\ u_i^- &= Y_{t_i} + f(Y_{t_i})\Delta t_i - g(Y_{t_i})\sqrt{\Delta t_i} \end{aligned} \tag{5.20}$$

and $\Delta \hat{W}_{t_i} \sim N(0, \Delta t_i)$.

- *Step 2:* We compute the corrector Y_{t_i} as

$$Y_{t_{i+1}} = Y_{t_i} + \frac{1}{2}\left[f(\overline{Y}_{t_{i+1}}) + f(Y_{t_i})\right]\Delta t_i + O_i \tag{5.21}$$

5. **Weak Order 2.0 Milstein Scheme** is given by

$$\begin{aligned} Y_{t_{i+1}} &= Y_{t_i} + \left(f(Y_{t_i}) - \frac{1}{2}g(Y_t)g_X(Y_t)\right)\Delta t_i + g(Y_{t_i})\sqrt{\Delta t_i}Z + \frac{1}{2}g(Y_t)g_X(Y_t)\Delta t_i Z^2 \\ &\quad + \frac{1}{2}\left(f(Y_{t_i})g_X(Y_{t_i}) + f_X(Y_{t_i})g(Y_{t_i}) + (g(Y_{t_i}))^2 g_{XX}(Y_{t_i})\right)(\Delta t_i)^{3/2}Z \\ &\quad + \frac{1}{2}\left(f(Y_{t_i})f_X(Y_{t_i}) + \frac{1}{2}(g(Y_{t_i}))^2 f_{XX}(Y_{t_i})\right)(\Delta t_i)^2, \end{aligned} \tag{5.22}$$

where $Z \sim N(1, 0)$.

5.3.4 The Lamperti Transformation

One desired aspect of any approximation Y_t of any SDE is their stability. A way to ensure that feature is to remove the state Y_t as factor in the diffusion coefficient $g(t, Y_t)$. For this end, and if possible, we may want to deploy the Lamperti transformation.

Theorem 5.1 (Lamperti transformation)**.** *Consider an SDE of the form*

$$dX_t = f(X_t, t)\,dt + g(X_t)\,dW_t, \quad X(t_0) = X_0 \tag{5.23}$$

with diffusion coefficient $g(X_t)$ dependent only on X_t. The Lamperti transformation, given by

$$Y_t = F(X_t) = \int \frac{1}{g(u)}\,du\Big|_{u=X_t}, \tag{5.24}$$

transforms process X_t into the process Y_t, which solves the SDE

$$\begin{aligned} dY_t &= \left[\frac{f(t, X_t)}{g(X_t)} + \frac{1}{2}g_X(X_t)\right]dt + dW_t \\ &= \left[\frac{f(t, F^{-1}(Y_t))}{g(F^{-1}(Y_t))} + \frac{1}{2}g_X(F^{-1}(Y_t))\right]dt + dW_t. \end{aligned} \tag{5.25}$$

Notice that the previous theorem is a direct result of applying the Itô formula to $F(X_t)$.

5.4 Local Linearization techniques

5.4.1 The Ozaki Method

Consider the continuous time process X_t defined by the SDE of the form, $dX_t = f(X_t)dt + g(X_t)dW_t$ where f is twice continuously differentiable, and g is continuously differentiable. As we seen previously, if we apply the Lamperti transformation, $Y_t = F(X_t) = \int \frac{\sigma}{g(u)}\, du\Big|_{u=X_t}$, we get an SDE with constant diffusion coefficient σ

$$dY_t = b(Y_t)\, dt + \sigma\, dW_t. \tag{5.26}$$

The Ozaki method is developed by taking $b_Y(Y_s)$ to be constant for an interval $s \in [t, t+\Delta t]$. In this case, for $Z_s = \mathbb{E}[Y_s|Y_t]$ the previous SDE is transformed into the ODE $\frac{dZ_s}{ds} = [b_Y(Z_t)s + b(Y_t)]$, which has the solution

$$Z_{t+\Delta t} = Y_t + \frac{b(Y_t)}{b_Y(Y_t)}\left(e^{b_Y(Y_t)\Delta t} - 1\right). \tag{5.27}$$

Recall that we supposed $b_Y(Y_s)$ to be a constant on the interval $s \in [t, t+\Delta t]$, given this, the SDE for Y_t becomes the autonomous linear narrow sense (ALNS) variety $dY_t = K_t Y_t\, dt + \sigma\, dW_t$, which we know the solution to be

$$Y_{t+\Delta t} = Y_t e^{K_t \Delta t} + \sigma \int_t^{t+\Delta t} e^{K_t(t+\Delta t - s)}\, dW_s. \tag{5.28}$$

On the other hand $Z_{t+\Delta t} = \mathbb{E}[Y_{t+\Delta t}|Y_t] = Y_t e^{K_t \Delta t}$, hence we compute K_t to be

$$K_t = \frac{1}{\Delta t}\log\left[1 + \frac{b(Y_t)}{Y_t b_Y(Y_t)}\left(e^{b_Y(Y_t)\Delta t} - 1\right)\right] \tag{5.29}$$

Notice that the stochastic integral of Y_t is a normal distribution with mean 0 and standard deviation $\sigma\sqrt{\frac{e^{2K_t\Delta t}-1}{2K_t}}$, thus the discrete time model given by the Ozaki method approximates Y_t as follow:

$$\begin{aligned} Y_{t+\Delta t} &= A_t Y_t + B_t W_{t+\Delta t} \\ A_t &= e^{K_t \Delta t} \\ B_t &= \sigma\sqrt{\frac{e^{2K_t\Delta t} - 1}{2K_t}} \\ K_t &= \frac{1}{\Delta t}\log\left[1 + \frac{b(Y_t)}{Y_t b_Y(Y_t)}\left(e^{b_Y(Y_t)\Delta t} - 1\right)\right]. \end{aligned} \tag{5.30}$$

5.4.2 The Shoji-Ozaki Method

The Shoji-Ozaki local linearization method is a generalization of the Ozaki method to a continuous-time process X_t defined by the SDE of the form, $dX_t = f(X_t, t)dt + g(X_t)dW_t$, where f is twice continuously differentiable on X_t, continuous differentiable in t and g is continuously differentiable. As we seen previously, if we apply the Lamperti transformation, $Y_t = F(X_t) = \int \frac{\sigma}{g(u)}\, du\Big|_{u=X_t}$, we get an SDE with constant diffusion coefficient σ

$$dY_t = b(Y_t, t)\, dt + \sigma\, dW_t. \tag{5.31}$$

Analogously to the previous method, the Shoji-Ozaki method is developed by supposing $b_t(Y_t, t)$, $b_Y(Y_t, t)$ and $b_{YY}(Y_t, t)$ all to be constant in an interval $t \in [s, s+\Delta s]$. Therefore if we apply the Itô formula to b, we get

$$db(Y_t, t) = L_s Y_t + M_s t + N_s, \tag{5.32}$$

where

$$\begin{aligned} L_s &= b_Y(Y_s, s), \\ M_s &= b_t(Y_s, s) + \frac{\sigma^2}{2} b_{YY}(Y_s, s), \\ N_s &= b(Y_s, s) - b_Y(Y_s, s) Y_s - \left(b_t(Y_s, s) + \frac{\sigma^2}{2} b_{YY}(Y_s, s) \right) s \\ &= b(Y_s, s) - L_s Y_s - M_s s, \end{aligned} \tag{5.33}$$

hence we can rewrite the SDE for the process Y_t as

$$dY_t = (L_s Y_t + M_s t + N_s)\, dt + \sigma\, dW_t, \quad t \geq s. \tag{5.34}$$

If we consider the transformed process $Z_t = e^{-L_s t} Y_t$, apply the Itô formula to it, compute the integral formula of Z_t and reincorporate the result back into Y_t, we arrive at the discretized process of Y_t given by

$$Y_{s+\Delta s} = Y_s + \frac{b(Y_s, s)}{L_s} \left(e^{L_s \Delta s} - 1 \right) + \frac{M_s}{L_s^2} \left[e^{L_s \Delta s} - 1 + L_s \Delta s \right] + \sigma \int_s^{s+\Delta s} e^{L_s (s+\Delta s - u)}\, dW_u. \tag{5.35}$$

The stochastic integral of Y_t is a normal distribution with mean 0 and standard deviation $\sigma \sqrt{\frac{e^{2L_s \Delta s} - 1}{2L_s}}$, thus the discrete time model, given by the Shoji-Ozaki method, which approximates Y_s, at $Y_s \neq 0$ and $L_s \neq 0$, is given by:

$$\begin{aligned} &Y_{s+\Delta s} = A(Y_s) Y_s + B(X_s) W_{s+\Delta s}, \text{ with} \\ &A(Y_s) = 1 + \frac{b(Y_s, s)}{Y_s L_s} \left(e^{L_s \Delta s} - 1 \right) + \frac{M_s}{Y_s L_s^2} \left[e^{L_s \Delta s} - 1 + L_s \Delta s \right] \text{ and} \\ &B(Y_s) = \sigma \sqrt{\frac{e^{2L_s \Delta s} - 1}{2L_s}}. \end{aligned} \tag{5.36}$$

5.4.3 The Rate of Convergence of the Local Linearization Method

Theorem 5.2 (Shoji's Theorem 1)**.** *Consider a continuous-time stochastic process X_t and its discretized approximation $\tilde{X}_t$ given by the Shoji-Ozaki local linearization method. If we define a p-th order error one-step-ahead prediction as $\mathbb{E}_s |X_t - \tilde{X}_t|^p$, for $s \leq t \leq T$. Then the rate of convergence of $\left(\mathbb{E}_s |X_t - \tilde{X}_t|^p \right)^{1/p}$ is 2, i.e.*

$$\mathbb{E}_s |X_t - \tilde{X}_t|^p = \mathcal{O}\left((t-s)^{2p} \right) \tag{5.37}$$

where $\mathcal{O}$ represents the big O notation, and $\mathbb{E}_s$ the conditional expectation at s.

Theorem 5.3 (Shoji's Theorem 2)**.** *Consider a fixed $t \in [s, T]$, a n-partition $s = t_0 \leq t_1 \leq \cdots \leq t_n = T$ of $[s, T]$, where $\Delta t_i = t_{i+1} - t_i$, and a step-by-step approximation of the integration*

$$X_t - X_s = \int_s^t f(X_s, s)\, ds + \sigma (W_t - W_s), \tag{5.38}$$

denoted by $\tilde{X}_t - \tilde{X}_s$, using the Shoji-Ozaki local linearization method. Then the convergence of the step-by-step approximation is $\mathcal{O}(\Delta t)$ in L_p, that is

$$\mathbb{E} \left| \left(X_{t+\Delta t} - X_t \right) - \left(\tilde{X}_{t+\Delta t} - \tilde{X}_t \right) \right|^p = \mathcal{O}\left((\Delta t)^p \right). \tag{5.39}$$

Chapter 6

Estimation of Parameters of Stochastic Differential Equations

6.1 Introduction

Consider the time-homogeneous stochastic differential equation

$$dX_t = f(X_t;\theta)\,dt + g(X_t;\theta)\,dW_t,\ t \geq 0 \tag{6.1}$$

our objective is to estimate $\theta \in \Theta$ where Θ is an open subset of $\mathbb{R}^n$. Since, continuous-time diffusions can only be observed at discrete time points, the transition probability density function is not explicitly computable. Thus, in order to overcome this quandary, we will make use of the **pseudo-maximum likelihood (PML) method**, which is undergirded by the well-known **maximum likelihood estimation** technique.

6.2 The Maximum Likelihood Technique

If we want to estimate the value of an unknown parameter $\theta \in \mathbb{R}^n$ of a given population, where the identical independent random variables $X_1, \ldots, X_n$ have been sampled from the population with a probability density function $p(x;\theta)$, then the joint probability function is given by

$$L\,(x_1, \ldots, x_n; \theta, n) = \prod_{i=1}^{n} p(x_i;\theta). \tag{6.2}$$

In this equation θ is constant and x_i, for $i = 1, \ldots, n$ are variables. The inverse is true for the likelihood function of a sample

$$\mathcal{L}\,(\theta; x_1, \ldots, x_n, n) = \prod_{i=1}^{n} p(x_i;\theta) \tag{6.3}$$

where x_i, for $i = 1, \ldots, n$ are realized samples of the random variables and θ is now the variable. For ease of computation the logarithmic-likelihood function is sometimes used:

$$\log \mathcal{L}\,(\theta; x_1, \ldots, x_n, n) = \sum_{i=1}^{n} \log p(x_i;\theta). \tag{6.4}$$

The maximum likelihood estimate $\hat{\theta} \in \mathbb{R}^n$ of θ, for the given x_i realizations for the independent random variables, results from a well know result in analysis, that the local maximum $\hat{\theta}$ of $\mathcal{L}$ is given when

$$\left.\frac{\partial \log \mathcal{L}}{\partial \theta}\right|_{\theta=\hat{\theta}} = 0, \tag{6.5}$$

and $\left.\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2}\right|_{\theta=\hat{\theta}}$ is negative definite.

6.3 The Markov Property, Transitional Densities, and the Likelihood Function of the Sample

Let $\{X_t\}_{t\in[t_0,T]}$ be a continuous Markov process on the probability space $(\Omega, \mathcal{A}, \{\mathcal{A}\}_{t\in[t_0,T]}, \mathbb{P})$, and consider a time discretization $t_0 \leq t_1 \leq \cdots \leq t_n = T$, and let us define $X_i = X(t_i)$, for $i = 0, 1, \ldots, n$. Let $\mathbb{P}(X_0, \ldots, X_n|\theta)$ be the joint probability of observing the path, for the a given value for parameter θ. Recall that if the process X_t is Markov, it has short memory, that is $\mathbb{P}(X_n|X_0, \ldots, X_{n-1}; \theta) = \mathbb{P}(X_n|X_{n-1}; \theta)$, therefore by the continuous application of the conditional probability definition and the Markov property $\mathbb{P}(X_0, \ldots, X_n|\theta)$ is simplified into,

$$\mathbb{P}(X_0, \ldots, X_n; \theta) = \mathbb{P}(X_n|X_{n-1}; \theta) \cdot \mathbb{P}(X_{n-1}|X_{n-2}; \theta) \cdots \mathbb{P}(X_1|X_0; \theta) \cdot \mathbb{P}(X_0; \theta). \tag{6.6}$$

The joint probability $\mathbb{P}(X_0, \ldots, X_n; \theta)$ of the stochastic process $\{X_t\}_{t\in[t_0,T]}$ has the probability density function, therefore the Markov property can be expressed as $p(X_n|X_1, \ldots, X_{n-1}; \theta) = p(X_n|X_{n-1}; \theta)$. Again, the continuous application of the conditional probability definition and the Markov property, results in

$$p(X_0, \ldots, X_n; \theta) = p(X_n|X_{n-1}; \theta) \cdot p(X_{n-1}|X_{n-2}; \theta) \cdots p(X_1|X_0; \theta) \cdot p(X_0; \theta). \tag{6.7}$$

Thus the likelihood function, for the θ variable, of a realized sample x_i, for $i = 1, \ldots, n$, is given by the formula

$$\mathcal{L}(\theta; x_1, \ldots, x_n, n) = p(x_0; \theta) \prod_{i=1}^{n} p(x_{i+1}|x_i; \theta) \tag{6.8}$$

6.4 Change of Variables

We know from analysis that given two open sets $U, V \subset \mathbb{R}^n$, a differentiable function $a : U \to \mathbb{R}$, and a diffeomorphism $f : V \to U$ between open sets U, V, i.e. a change of variables $f(y_1, \ldots, y_n) = (x_1(y_1, \ldots, y_n), \ldots, x_n(y_1, \ldots, y_n))$, we have

$$\begin{aligned} \int a(x_1, \ldots, x_n)\, dx_1 \ldots dx_n &= \int a \circ f(y_1, \ldots, y_n) |\det df|\, dy_1 \ldots dy_n \\ &= \int a \circ f(y_1, \ldots, y_n) \left| \det \left[\frac{\partial x}{\partial y} \right] \right| dy_1 \ldots dy_n. \end{aligned} \tag{6.9}$$

Thus, the change of variables for a probability density function $p_X(x_1, \ldots, x_n)$ of X for a probability density function $p_Y(y_1, \ldots, y_n)$ of Y, where Y is defined as $X = f(Y)$, where $f(y_1, \ldots, y_n) = (x_1(y_1, \ldots, y_n), \ldots, x_n(y_1, \ldots, y_n))$ is an homeomorphism, is simply a generalization of the previous property. In fact, for every measurable A set we have

$$\begin{aligned} \mathbb{E}[\mathbb{1}_A(X)] &= \int_{X(A)} p_X(x_1, \ldots, x_n)\, dx_1 \ldots dx_n \\ &= \int_{Y(A)} p_X \circ f(y_1, \ldots, y_n) \left| \det \left[\frac{\partial x}{\partial y} \right] \right| dy_1 \ldots dy_n. \end{aligned} \tag{6.10}$$

Therefore, the probability density function p_Y is given by

$$p_Y(y_1, \ldots, y_n) = p_X \circ f(y_1, \ldots, y_n) \left| \det \left[\frac{\partial x}{\partial y} \right] \right|. \tag{6.11}$$

6.5 The Transition Probability Density Function is Known

To have a better perspective of the PML method let us start by tackling a case where equation 6.1 is solvable, hence the PML is not needed

Example:

Consider the SDE

$$dX_t = \mu X_t\, dt + \sigma X_t\, dW_t,\ t \geq 0$$

and n realized samples for X_t, at increasing time points t_i, for $i = 1, \ldots, n$. By the Itô formula we have:

$$X(t_i) = X(t_{i-1}) e^{\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t_i + \sigma\sqrt{\Delta t_i} Z_i}$$

where $Z_i \sim N(0,1)$ are independent and $\Delta t_i = t_i - t_{i-1}$. If we denote $x_i = \log(X(t_i) - X(t_{i-1}))$, then $x_i \sim N\left(\left(\mu - \frac{1}{2}\sigma\right)\Delta t_i, \sigma^2 \Delta t_i\right)$. By the change of variables equation (6.11), we conclude that the probability density function of $X(t_i) - X_(t_{i-1})$ is given by

$$p(x_i, \Delta t_i; \mu, \sigma) = \frac{1}{\sigma(X(t_i) - X_(t_{i-1}))\sqrt{2\pi\Delta t_i}} \exp\left\{ - \frac{[x_i - (\mu - (1/2)\sigma^2)\Delta t_i]^2}{2\sigma^2 \Delta t_i} \right\}$$

Thus, the log-likelihood function for this probability density function is

$$\begin{aligned} \log \mathcal{L} = &- \sum_{i=1}^{n} \log(X(t_i) - X(t_{i-1})) - \frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2 \Delta t_i) \\ &- \frac{1}{2\sigma^2 \Delta t_i} \sum_{i=1}^{n} [x_i - (\mu - (1/2)\sigma^2)\Delta t_i]^2. \end{aligned}$$

If we compute $\frac{\partial \log \mathcal{L}}{\partial \mu} = 0$, and $\frac{\partial \log \mathcal{L}}{\partial \sigma} = 0$, we arrive at the expressions $\hat{m} = \frac{\sum_{i=1}^{n} x_i}{n} = \left(\hat{\mu} - \frac{1}{2}\hat{\sigma}^2\right)\Delta t_i$ and $\hat{v} = \frac{\sum_{i=1}^{n}(x_i - \hat{m})^2}{n} = \hat{\sigma}^2 t_i$ which in turn yield the estimates for μ and σ:

$$\hat{\mu} = \frac{\hat{m} + (1/2)v}{\Delta t_i}, \text{ and } \hat{\sigma} = \sqrt{\frac{\hat{v}}{\Delta t_i}}. \blacksquare \tag{6.12}$$

6.6 The Transition Probability Density Function is Unknown

When the probability density function p is unknown but some of the conditional moments of the diffusion process are known, then under certain restrictions it might be expedient to estimate θ from a pseudo-transition density function h, which while different from p has consonant moments. Therefore, the method for obtaining a pseudo-likelihood function, for a Markov process $\{X_t\}_{t\geq 0}$ where $n+1$ historical observations of $X(t)$ at non stochastic times $t0 < t1 < \cdots < t_n$, where we denote $X_i = X(t_i)$, is analogous to the ML case, resulting in

$$\mathcal{L}(\theta, X_1, \ldots, X_n, \Delta t_i, n) = h(X_0; \theta) \prod_{i=1}^{n} h(\Delta t_i, X_{i-1}, X_i; \theta) \tag{6.13}$$

or if X_0 is independent from θ

$$\mathcal{L}(\theta, X_1, \ldots, X_n, \Delta t_i, n) = \prod_{i=1}^{n} h(\Delta t_i, X_{i-1}, X_i; \theta). \tag{6.14}$$

Thus, the PML estimate of θ is the value that maximizes the logarithm of the pseudo-likelihood function, i.e.

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} \log h(\Delta t_i, X_{i-1}, X_i; \theta). \tag{6.15}$$

In the following, we shall observe how to estimate the PML for the SDE (6.1) using 3 three different iterations schemes: the Euler-Maruyama scheme, the Ozaki linearization scheme, and the Shoji-Ozaki linearization scheme. Recall, the every SDE of the form (6.1) can be transformed to an SDE with constant diffusion term, using the Lamperti transformation, hence we only need to consider SDE's of the form

$$dX_t = f(X_t; \theta)\, dt + \sigma\, dW_t, t \geq 0, X(0) = X_0. \tag{6.16}$$

6.6.1 Euler-Maruyama scheme

The Euler-Maruyama scheme for SDE (6.16), is given by the expression,

$$X_{t+\Delta t} = X_t + f(X_t;\theta)\Delta t + \sigma \Delta W_t. \tag{6.17}$$

Since, $X_{t+\Delta t_i} - X_t$ are independent Normal random variables their pseudo-transition density function are thus

$$h(\Delta t_i, X_{i-1}, X_i; \theta, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2\Delta t_i}} e^{-\frac{(X_i - X_{i-1} - f(X_{i-1};\theta)\Delta t_i)^2}{2\sigma^2\Delta t_i}},$$

where $X_i = X_{t_i}$. Therefore, its pseudo log-likelihood function is given by

$$\begin{aligned}\log \mathcal{L}(\theta, \sigma; X_1, \ldots, X_n, \Delta t_i, n) = &\log h(X_0;\theta,\sigma) - \frac{n}{2}\log(2\pi\sigma^2\Delta t_i) \\ &- \frac{1}{2}\sum_{i=1}^{n} \frac{(X_i - X_{i-1} - f(X_{i-1};\theta)\Delta t_i)^2}{\sigma^2\Delta t_i}.\end{aligned} \tag{6.18}$$

Example:

Consider the Ornstein-Uhlenbeck process given by the SDE

$$dX_t = \alpha(\mu - X_t)\, dt + \sigma\, dW_t.$$

By the previous results, the pseudo log-likelihood function associated with the EM-scheme is

$$\begin{aligned}\log \mathcal{L}(\theta, \sigma; X_1, \ldots, X_n, \Delta t_i, n) = &\log h(X_0;\theta,\sigma) - \frac{n}{2}\log(2\pi\sigma^2\Delta t_i) \\ &- \frac{1}{2}\sum_{i=1}^{n} \frac{(X_i - X_{i-1} - \alpha(\mu - X_{i-1})\Delta t_i)^2}{\sigma^2\Delta t_i}\end{aligned}$$

The approximation afforded by the EM scheme is considered good if:

1. if Δt is very small;
2. if the polynomial growth condition holds, i.e. there exist $L, m > 0$, such that $|f(X;\theta)| \leq L\,(1 + |X|^m)$, for any $\theta \in \Theta$; and
3. if $n(\Delta t_n)^3 \to 0$ (this condition ensures the PML estimator from EM-scheme is consistent and asymptotically efficient). ■

6.6.2 Ozaki Linearization Method

The Ozaki Linearization scheme for SDE (6.16), is given by the expression

$$X_{t+\Delta t_i} = A_t X_t + B_t W_{t+\Delta t}, \tag{6.19}$$

where

$$A_t = e^{K_t \Delta t}, \quad B_t = \sigma \left(\frac{e^{2K_t\Delta t} - 1}{2K_t}\right)^{\frac{1}{2}}, \tag{6.20}$$

and

$$K_t = \frac{1}{\Delta t}\log\left[1 + \frac{f(X_t)}{X_t f_X(X_t)}\left(e^{f_X(X_t)\Delta t} - 1\right)\right]. \tag{6.21}$$

Similarly to the EM scheme case, $X_{t+\Delta t_i} - X_t$ are independent Normal random variables with expected value, and variance given by

$$E_i = \mathbb{E}(X_{t_i}) = e^{K_{t_i}\Delta t_i} X_{t_i}, \quad V_i = \mathbb{V}(X_{t_i}) = \sigma^2 \left(\frac{e^{2K_{t_i}\Delta t_i} - 1}{2K_t} \right), \tag{6.22}$$

respectively. Hence, their pseudo-transition density function are

$$h(\Delta t_i, X_{i-1}, X_i; \theta, \sigma) = \frac{1}{\sqrt{2\pi V_{i-1}}} e^{-\frac{(X_i - E_{i-1})^2}{2V_{i-1}}} \tag{6.23}$$

where $W_i = W_{t_i}$ and $X_i = X_{t_i}$. Therefore, its pseudo log-likelihood function is

$$\log \mathcal{L}(\theta, \sigma; X_1, \ldots, X_n, \Delta t_i, n) = \log h(X_0; \theta, \sigma) - \frac{n}{2} \log(2\pi V_{i-1}) - \frac{1}{2} \sum_{i=1}^{n} \frac{(X_i - E_{i-1})^2}{\sigma^2 \Delta t_i} \tag{6.24}$$

Example:

Consider the SDE, $dX_t = \theta X_t^3\, dt + \sigma\, dW_t$, by the previous results, the pseudo log-likelihood function resulting from the Ozaki Linearization scheme, is completely defined by

$$K_i = \frac{1}{\Delta t_i} \log \left[1 + \frac{1}{3} \left(e^{3\theta X_i^2 \Delta t_i} - 1 \right) \right]$$

because, in turn, K_i defines the expected value and variance as

$$E_i = e^{K_{t_i}\Delta t_i} X_{t_i} = X_i + \frac{X_i}{3} \left(e^{3\theta X_i^2 \Delta t_i} - 1 \right), \quad V_i = \sigma^2 \left(\frac{\left[1 + \frac{1}{3} \left(e^{3\theta X_i^2 \Delta t_i} - 1 \right) \right]^2 - 1}{\frac{2}{\Delta t_i} \left[1 + \frac{1}{3} \left(e^{3\theta X_i^2 \Delta t_i} - 1 \right) \right]} \right). \blacksquare$$

6.6.3 Shoji-Ozaki Linearization Method

The Shoji-Ozaki Linearization scheme for SDE (6.16), is given by the expression

$$X_{t+\Delta t} = A(X_t) X_t + B(X_t) W_{t+\Delta t}, \tag{6.25}$$

where

$$A_t = 1 + \frac{f(X_t)}{X_t L_t} \left(e^{L_t \Delta t} - 1 \right) + \frac{M_t}{X_t L_t^2} \left[\left(e^{L_t \Delta t} - 1 \right) - L_t \Delta t \right], \tag{6.26}$$

$$B_t = \sigma \left(\frac{e^{2L_t \Delta t} - 1}{2L_t} \right)^{\frac{1}{2}}, \quad L_t = f_X(X_t), \text{ and } \quad M_t = \frac{\sigma^2}{2} f_{XX}(X_t).$$

Similarly to the previous two schemes, $X_{t+\Delta t_i} - X_t$ are independent Normal random variables with expected value, and variance given by

$$E_i = \mathbb{E}(X_{t_i}) = A(X_{t_i}) X_{t_i}, \quad V_i = \mathbb{V}(X_{t_i}) = \sigma^2 \left(\frac{e^{2L_{t_i}\Delta t_i} - 1}{2L_{t_i}} \right) \tag{6.27}$$

respectively. Hence, their pseudo-transition density function are

$$h(\Delta t_i, X_{i-1}, X_i; \theta, \sigma) = \frac{1}{\sqrt{2\pi V_{i-1}}} e^{-\frac{(X_i - E_{i-1})^2}{2V_{i-1}}} \tag{6.28}$$

where $W_i = W_{t_i}$ and $X_i = X_{t_i}$. Therefore, its pseudo log-likelihood function is

$$\log \mathcal{L}(\theta, \sigma; X_1, \ldots, X_n, \Delta t_i, n) = \log h(X_0; \theta, \sigma) - \frac{n}{2} \log(2\pi V_{i-1}) - \frac{1}{2} \sum_{i=1}^{n} \frac{(X_i - E_{i-1})^2}{\sigma^2 \Delta t_i} \tag{6.29}$$

Example:

Consider the SDE, $dX_t = \theta X_t^3\, dt + \sigma\, dW_t$, by the previous results, the pseudo log-likelihood function resulting from the Shoji-Ozaki Linearization scheme, is completely defined by

$$L_i = 3\theta X_i^2, \quad M_i = \sigma^2(3\theta X_i)$$

because, in turn, L_i and M_i define the expected value and variance as

$$E_i = X_i + \frac{X_i}{3}\left(e^{3\theta X_i^2 \Delta t_i} - 1\right) + \frac{\sigma^2}{3\theta X_t^3}\left(e^{3\theta X_i^2 \Delta t_i} - 1 - 3\theta X_i^2 \Delta t_i\right), \quad V_i = \sigma^2\left(\frac{e^{6\theta X_i^2 \Delta t_i} - 1}{6\theta X_i^2}\right). \blacksquare$$

References

Braumann, Carlos A. 2019. *Introduction to Stochastic Differential Equations with Applications to Modelling in Biology and Finance*. John Wiley & Sons.

Braun, W John, and Duncan J Murdoch. 2021. *A First Course in Statistical Programming with r*. Cambridge University Press.

Capiński, Marek, and Peter E Kopp. 2004. *Measure, Integral and Probability*. Vol. 14. Springer.

Øksendal, Bernt, and Bernt Øksendal. 2003. *Stochastic Differential Equations*. Springer.

Panik, Michael J. 2017. "Stochastic Differential Equations: An Introduction with Applications in Population Dynamics Modelling."

Pires, Gabriel E. 2012. *Cálculo Diferencial e Integral Em Rn*. Vol. 45. IST Press.