Markov Processes and Stochastic Calculus

René Caldentey

In this notes we revise the basic notions of Brownian motions, continuous time Markov processes and stochastic differential equations in the Itô sense.

1 Introduction

A sample space $(\Omega, \mathcal{F})$ consists on:

- Ω: An abstract space of points $\omega \in \Omega$.
- $\mathcal{F}$: a σ-field (or σ-algebra) on Ω. That is, a collection of subsets of Ω satisfying:

  1. $\Omega \in \mathcal{F}$.
  2. Let $A \subseteq \Omega$ such that $A \in \mathcal{F}$ then $A^c = \Omega - A \in \mathcal{F}$.
  3. Let $A_1, A_2, A_3, \dots \in \mathcal{F}$ then $A_1 \cup A_2 \cup A_3 \cup \dots \in \mathcal{F}$.

  The elements of $\mathcal{F}$ are called *events*. Condition (1) above simply states that the space Ω is necessarily an event. Conditions (2) and (3) state that the collection of events is closed under the set operations of complement and countable union.

The space $(\Omega, \mathcal{F})$ satisfying these conditions is called a *measurable space*. For a fixed point $\omega \in \Omega$, we called the *sample path* of the process X associated with ω the function $t \to X_t(\omega) \ : \ t \geq 0$.

Definition 1 *Let $(\Omega, \mathcal{F})$ be a measurable space. A set function μ on $\mathcal{F}$ is called a* measure *if it satisfies the following conditions:*

1. $\mu(\emptyset) = 0$,

2. $A \in \mathcal{F}$ *implies* $0 \leq \mu(A) \leq \infty$, *3.* $A_1, A_2, \dots \in \mathcal{F}$ *and* $\{A_n\}$ *are pairwise disjoint* ($A_i \cap A_j = \emptyset$ *for* $i \neq j$)*, then*

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

For example, the *Lebesgue measure* commonly denoted by λ and defined on the class of Borel sets $\mathcal{R}^1$ of the real line, is given by $\lambda(a, b] = b - a$.

We call a measure μ a *probability measure* if $\mu(\Omega) = 1$, in this case instead of μ, we use the notation P. If P is a probability measure then $(\Omega, \mathcal{F}, P)$ is called a *probability space*.

Definition 2 *Let* $(\Omega, \mathcal{F})$ *and* $(\Omega', \mathcal{F}')$ *two measurable spaces. A mapping* $Y : \Omega \to \Omega'$ *is* ($\mathcal{F}$, $\mathcal{F}'$)*-measurable if for each* $A' \in \mathcal{F}'$

$$Y^{-1}(A') = \{\omega \in \Omega : Y(\omega) \in A'\} \in \mathcal{F}.$$

According to this definition, the mapping Y is measurable if for any well-defined event $A' \in \mathcal{F}'$ the pre-image of A' by Y (denoted by $Y^{-1}(A')$) is also a well-defined event in $\mathcal{F}$. If $(\Omega', \mathcal{F}') = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ then Y is called a real-valued random variable. The probability distribution of a random variable Y, is a mapping $P_Y : \mathcal{B}(\mathbb{R}) \to [0, 1]$ defined as follows

$$P_Y(B) = P(\{\omega \in \Omega \text{ such that } Y(\omega) \in B\}) := P(Y \in B).$$

Definition 3 *A* stochastic process *is a collection of random variables* $X = \{X(t) : t \geq 0\}$ *defined on a probability space* $(\Omega, \mathcal{F}, P)$*.*

For simplicity, we will assume on this notes that the index t above represents time. We can view a stochastic process as mapping that at each time $t \geq 0$ associated the occurrence of a random phenomenon represented by $X(t)$. The function $X(\cdot, \omega)$ for a fixed $\omega \in \Omega$ is called a *sample path* or *sample function*. The stochastic process X is called *continuous* if the P-almost every sample path is continuous.

An important consideration is related to the way we collect and use information over time. In particular, we would like to be able to isolate past and present from future. For example, let us consider two events $A, B \in \mathcal{F}$ such that $X_s(\omega_1) = X_s(\omega_2)$; $\forall \omega_1 \in A$, $\forall \omega_2 \in B$, $\forall s, 0 \leq s \leq t$. Then during the period $[0, t]$ the events A and B cannot be distinguished based on the knowledge of X. For this reason, we complement our sample space $(\Omega, \mathcal{F})$ with a *filtration*, i.e., a nondecreasing family $\{\mathcal{F}_t : t \geq 0\}$ of sub-σ-fields of $\mathcal{F}$ such that $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for $0 \leq s \leq t$. The idea is that each $\mathcal{F}_t$ contains the information available up to time t.

Definition 4 *A* stochastic process X *is* adapted *to the filtration* $\mathcal{F}_t$ *if* $X(t)$ *is* $\mathcal{F}_t$*-measurable for all* $t \geq 0$*.*

Given a stochastic process X we define the filtration $\mathcal{F}_t^X$ generated by X as the smallest filtration such that $X(t) \in \mathcal{F}_t^X$ for all $t \geq 0$.

Definition 5 *A* stopping time *is a measurable function T from $(\Omega, \mathcal{F})$ to $[0, \infty)$ such that $\{\omega \in \Omega : T(\omega) \leq t\} \in \mathcal{F}_t$, for all $t \geq 0$.*

2 Brownian Motion

The name *Brownian motion* (BM) comes from the studies done by the botanist Robert Brown in 1828 on the irregular movement (Brownian movement) of pollen suspended in water. A mathematical definition a Brownian motion is the following.

Definition 6 *A standard, one-dimensional* Brownian motion *(or Wiener process) is a continuous, adapted stochastic process $X = \{X_t, \mathcal{F}_t : 0 \leq t < \infty\}$, defined on some probability space $(\Omega, \mathcal{F}, P)$ with the properties that $X_0 = 0$ almost surely and for $0 \leq s < t$, the increment $X_t - X_s$ is independent of $\mathcal{F}_s$ and is normally distributed with mean zero and variance $t - s$.*

A few points about the previous definition:

(a) We say that $X_0 = 0$ *almost surely* (a.s.) in $(\Omega, \mathcal{F}, P)$ if the set $A = \{\omega \in \Omega : X_0(\omega) \neq 0\}$ has probability 0, i.e., $P(A) = 0$.

(b) The stochastic process X is adapted to the filtration $\{\mathcal{F}_t\}$ if for each $t \geq 0$, X_t is an $\mathcal{F}_t$-measurable random variable.

Two extremely important features that characterizes a Wiener process presented in the following property:

Property 1 *If X is a Wiener process then X has* independent increments*, that is, for any positive integer n and any sequences of times $0 \leq t_0 < t_1 < \cdots < t_n < \infty$ the random variables $Y_i = X(t_i) - X(t_{i-1})$, $i = 1, 2, \ldots, n$ are independents.*
In addition, X has stationary increments*, that is, for any $0 \leq s < t < \infty$ the distribution of $X_t - X_s$ depends only on $t - s$.*

These two properties are so attached to the Wiener process that they can be used as an alternative definition of a standard Brownian motion (together with the requirement of continuous sample paths). We notice here that for discrete time stochastic processes the two properties above characterize the Poisson process.

Once we have defined the Wiener process, we can extend its definition and define the general (μ, σ) Brownian motion process Y as follows:

$$Y(t) = Y(0) + \mu t + \sigma X(t)$$

where X is a Wiener process and $Y(0)$ (the initial value) is independent of X. We call μ the *drift* and σ^2 the *variance* or *diffusion* of Y. It follows directly from the definition of X that $Y(t+s) - Y(t)$ is normally distributed with mean μs and variance $\sigma^2 s$. Finally, we say that Z is a *geometric Brownian motion* if $Z_t = e^{Y_t}$, where Y is a Brownian motion.

> **Exercise 1**: Prove that if Z is a process with stationary independent increment, such that $\sigma^2 := \mathsf{Var}(Z(1))$, then $\mathsf{Var}(Z(t)) = \sigma^2\, t$ for all $t \geq 0$.

3 Properties of Brownian Motions

In this section we present the main properties that make Brownian motions a very attractive modelling tool. However, we start ironically presenting some results showing the extremely erratic behavior of Brownian motion processes.

3.1 Basic Properties

Property 2 *Let X be a Brownian motion in $(\Omega, \mathcal{F}, P)$. Then except for a set of probability 0, the sample path $X_t(\omega)$ is nowhere differentiable.*

Even though the variation of X over time is particular unstable, some measure of its variability can be computed. In fact, let define the random variable (quadratic variation) Q_t as follows:

$$Q_t \equiv \lim_{n\to\infty} \sum_{k=0}^{2^n-1} \left[X\left(\frac{(k+1)t}{2^n}\right) - X\left(\frac{kt}{2^n}\right) \right]^2 . \qquad (1)$$

Then, we have the following result.

Property 3 *For almost every $\omega \in \Omega$ we have $Q_t(\omega) = \sigma^2 t$ for all $t \geq 0$.*

This last result implies that Brownian motion have infinite ordinary variation almost surely. In addition, as we will see later, property 3 contains the essence of the Itô's formula.

Property 4 *If X is a (μ, σ) Brownian motion then:*

- $E(X_t) = X_0 + \mu t$,
- $Var(X_t) = \sigma^2 t$,
- $Cov(X_t, X_s) = \sigma^2 (t \wedge s) = \sigma^2 min\{t, s\}$.

The following theorem is a very important result that reflect the memoryless property that characterizes Brownian motion processes.

Theorem 1 (Strong Markov Property)

Let X be a (μ, σ) Brownian motion and T be a finite stopping time. Then $Y_t = X_{T+t} - X_T$ is a (μ, σ) Brownian motion starting at 0 and it is independent of $\mathcal{F}_T$.

Property 5 (Brownian martingales)

Let X be a (μ, σ) Brownian motion then:

(a) *If $\mu = 0$ then X is a martingale, i.e., $E(X_t - X_s | \mathcal{F}_s) = 0$.*

(b) *If $\mu = 0$ then $X_t^2 - \sigma^2 t$ is martingale.*

(c) *Let $q(\beta) = \mu\beta + \frac{1}{2}\sigma^2\beta^2$ and $V_\beta(t) = e^{\beta X_t - q(\beta)t}$. Then, V_β is a martingale.*

3.2 Wiener Measure and Donsker's Theorem

In this subsection we explore the nature of the Wiener process as a type of central limit theorem for stochastic processes. The notation and results are based on the textbook *Convergence of Probability Measures* by P. Billingsley (1999).

We start by introducing the Wiener measure, W, which is a probability measure on $(C, \mathcal{C})^*$ having two properties. First, each X_t is normally distributed under W with mean 0 and variance t, that is:

$$W\left[X_t \leq \alpha\right] = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\alpha} e^{\frac{-u^2}{2t}} du.$$

For $t = 0$, we have $W[X_0 = 0] = 1$. The second property is that the stochastic process X has independent increments under W.

In order to state the main result of this section (Donsker's theorem), we introduce the sequences $\{X^n : n = 0, 1, \dots\}$ of stochastic processes as follows. Let $\Xi = \{\xi_1, \xi_2, \dots\}$ be a sequence of IID

*Where $C \equiv C[0, \infty)$ is the space of all continuous functions $x : [0, \infty) \to \mathcal{R}$ and $\mathcal{C}$ is the Borel σ-algebra on C.

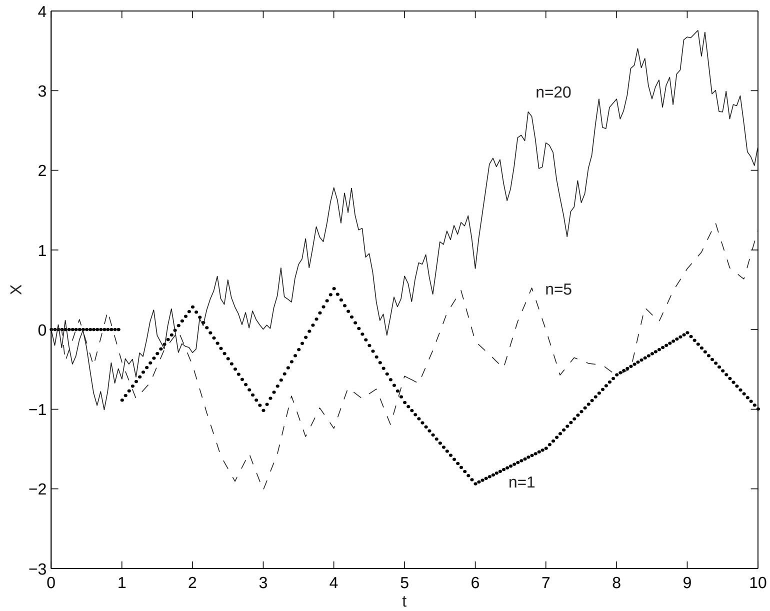

Figure 1: *Behavior of* $\{X_t^n\}$ *for* $n = 1$, 5, *and* 20.

random variables having mean 0 and finite variance σ^2. Let $S_n = \xi_1 + \cdots + \xi_n$ ($S_0 = 0$) be the partial sums of Ξ. We define X^n as follows:

$$X_t^n(\omega) = \frac{1}{\sigma\sqrt{n}} S_{\lfloor nt \rfloor}(\omega) + (nt - \lfloor nt \rfloor) \frac{1}{\sigma\sqrt{n}} \xi_{\lfloor nt \rfloor + 1}(\omega) \tag{2}$$

Figure 1 shows the behavior of the process X_t^n for three values of n †. We can see that as n increases, the behavior of X^n resembles a Wiener process. This result is in fact Donsker's theorem. We can see the non differentiability of $X^n(t)$ as n increases.

Theorem 2 (Donsker's Theorem) *If* $\xi_1, \xi_2, \ldots$ *are independent and identically distributed random variables with mean 0 and variance* σ^2*, and if* X^n *is the random process defined by (2), then* $X^n \Longrightarrow_n W$*, a Wiener process.*

(Where the symbol $\Longrightarrow_n$ stands for *convergence in distribution* as $n \to \infty$.) This result can be understood as a generalization of the standard *central limit theorem* for random variables.

The previous result is intuitive in the sense that S_n -being the sum of IID random variables- converges in distribution to a $N(0, n\sigma^2)$. Another interesting property of the Wiener process, and more generally of any $(0, \sigma)$ Brownian motion is their *scale invariance* that can partially be observed in (2).

† In the construction of the three sample paths the ξ_n are uniformly distributed in $[-1, 1]$.

Property 6 (Scale Invariance) *Let X be a $(0,\sigma)$ Brownian motion, then for any $c > 0$:*

$$\left\{ \frac{X(ct)}{\sqrt{c}} : t \geq 0 \right\} \stackrel{D}{=} \{X(t) : t \geq 0\}. \tag{3}$$

(Where $\stackrel{D}{=}$ stands for equality in distribution.)

This scaling property, that is of course related to the normal distribution, is specially important (together with Donsker's theorem) on the use of heavy traffic approximations for queueing systems.

We can now use Donsker's theorem to find the distribution of $M \equiv \sup W$, however, we need before an additional result.

Theorem 3 (Mapping Theorem) *Let $\{X_n\}$ be a sequence of processes such that $X_n \Rightarrow X$. Let h be a measurable function and let D_h be the set of its discontinuities. If D_h has probability 0, then $h(X_n) \Rightarrow h(X)$.*

Since $h(X) := \sup_t X_t$ is a continuous function on C, then from the mapping theorem and the fact that $X^n \Rightarrow W$, we have that:

$$\sup_t X_t^n \Rightarrow \sup_t W_t.$$

Let $M_n = \max_{0 \leq i \leq n} S_i$, then it is not hard to show that $\sup_t X_t^n = \frac{M_n}{\sigma\sqrt{n}}$. Thus,

$$\frac{M_n}{\sigma\sqrt{n}} \Rightarrow \sup_t W_t. \tag{4}$$

Since we can peak any sequence $\{\xi_n\}$ such that $E(\xi_n) = 0$ and $E(\xi_n^2) < \infty$, let assume that ξ_n takes the values ± 1 with probability $\frac{1}{2}$. Therefore, $S_0, S_1, \ldots$ represents a symmetric random walk starting at 0. We first prove that

$$P(M_n \geq a, S_n < a) = P(M_n \geq a, S_n > a) \; a \geq 0.$$

This should be clear from the fact that the behavior of the random walk is independent of its history and it is symmetric, thus if the random walk reach a at time $\hat{n} < n$ then the value of S_n is symmetric with respect to $S_{\hat{n}} = a$. In other words, for each path of the random walk $(S_0, S_1, \ldots, S_n)$ such that $M_n \geq a, S_n = a - k < a$ there exists another path such that $M_n \geq a, S_n = a + k > a$. This symmetry is an example of the *reflection principle*. Given this result, we have that:

$$\begin{aligned} P(M_n \geq a) &= P(M_n \geq a, S_n < a) + P(M_n \geq a, S_n = a) + P(M_n \geq a, S_n > a) \\ &= 2P(M_n \geq a, S_n > a) + P(M_n \geq a, S_n = a) \\ &= 2P(S_n > a) + P(S_n = a) \end{aligned}$$

By the central limit theorem $P(S_n > a\sqrt{n}) \to P(N > a)$ and $P(S_n = a\sqrt{n}) \to 0$, where N is a standard $(0,1)$ normally distributed random variable . In addition $2P(N > a) = P(|N| > a)$. Thus, combining this results we have that $M = \sup_t W_t$ has the same distribution of $|N|$ and

$$P(M \leq a) = \frac{2}{\sqrt{2\pi}} \int_0^a e^{\frac{-u^2}{2}} du. \tag{5}$$

3.3 Reflection Principle

In this subsection, we look with more detail at the distribution of $M_t = \sup_{0 \leq s \leq t} X_s$, where X is a general (μ, σ) Brownian motion. We first start the analysis for the special case of $\mu = 0, \sigma = 1$. In this case, we can apply a similar argument that the one used in the previous subsection based on the *reflection principle* to show that

$$P(M_t \geq x) = 2P(X_t \geq x) = P(|X_t| \geq x).$$

We can also compute the joint distribution for (X_t, M_t), that is

$$F_t(x, y) = P(X_t \leq x, M_t \leq y).$$

Since $X_0 = 0$ and $M_t \geq X_t$ w.p.1, we can focus our attention to the case $x \leq y$ and $y \geq 0$. First of all, we notice that

$$\begin{aligned} F_t(x, y) &= P(X_t \leq x) - P(X_t \leq x, M_t > y) \\ &= \Phi(xt^{-\frac{1}{2}}) - P(X_t \leq x, M_t > y), \end{aligned}$$

where $\Phi(\cdot)$ is the $N(0,1)$ distribution function. From the reflection principle $P(X_t \leq x, M_t > y) = P(X_t \geq 2y - x) = P(X_t \leq x - 2y)$. Thus, we have the following result.

Property 7 *If $\mu = 0$ and $\sigma = 1$, then*

$$P(X_t \leq x, M_t \leq y) = \Phi(xt^{-\frac{1}{2}}) - \Phi((x - 2y)t^{-\frac{1}{2}}).$$

The previous result depends heavily on the assumption $\mu = 0$ or in other words on the reflection principle. In order to extend the result to general Brownian motion, it is required first to understand how making a change of measure can lead to a change of drift.

Let P and Q be two probability measures on the same space $(\Omega, \mathcal{F})$ with the important property that P is *dominated* by Q. That is, $Q(A) = 0 \implies P(A) = 0$. Then, there exists a non-negative random variable ξ (also denoted by $\frac{dP}{dQ}$) such that

$$P(A) = \int_A \xi \, dQ, \; \forall A \in \mathcal{F}.$$

An important implication of the above relation is that if Y is a random variable and $E_Q(|\xi Y|) < \infty$ then $E_P(Y)$ exists and $E_P(Y) = E_Q(\xi Y)$. The random variable ξ is usually called the *density* or Radon-Nikodym derivative (or likelihood ratio) of P with respect to Q. In order to find the density ξ associated to two Brownian motion measures of different drift, we use an heuristical approach.

Let consider now a (μ, σ) Brownian motion and a sequence of instants $0 = t_0 < t_1 < \cdots < t_n = t$ such that $t_i - t_{i-1} = \delta,\ i = 1, \ldots, n$. The density associated to that particular sequence of instance is given by:

$$\frac{1}{(\sigma\sqrt{2\pi\delta})^n} \prod_{i=1}^{n} e^{-\frac{(X_{t_i} - X_{t_{i-1}} - \mu\delta)^2}{2\sigma^2\delta}}.$$

If the drift were instead $\mu+\theta$ then density is obtained replacing μ by $\mu+\theta$ above. Thus, the density is given by:

$$e^{\frac{1}{2\sigma^2\delta}\sum_{i=1}^{n}(X_{t_i} - X_{t_{i-1}} - \mu\delta)^2 - (X_{t_i} - X_{t_{i-1}} - (\mu+\theta)\delta)^2}.$$

After some algebra, we have that the density is given by:

$$\begin{aligned} \xi(t) &= e^{\frac{\theta}{\sigma^2}(X_t - \mu t - \frac{\theta t}{2})} \\ &= V_{\frac{\theta}{\sigma^2}}(t). \end{aligned} \tag{6}$$

Where $V_\beta(t)$ is Wald martingale defined in property (5). We can compute the distribution of M_t for the case of $\mu \neq 0$ as follows (we start with $\sigma = 1$):

$$\begin{aligned} P_\mu(M_t \geq x) &= E_0(V_\mu(t); M_t \geq x) \\ &= 1 - \Phi\left(\frac{x - \mu t}{\sqrt{t}}\right) + e^{2\mu x}\Phi\left(\frac{-x - \mu t}{\sqrt{t}}\right). \end{aligned}$$

Finally, for the general case (μ, σ), we can rescale the probability measure to obtain:

$$P(M_t \leq x) = \Phi\left(\frac{x - \mu t}{\sigma\sqrt{t}}\right) - e^{\frac{2\mu x}{\sigma^2}}\Phi\left(\frac{-x - \mu t}{\sigma\sqrt{t}}\right), \tag{7}$$

which is called the *inverse Gaussian distribution.*

3.4 Forward and Backward Equations

An important extension to the standard Brownian motion is related to the initial condition X_0. Previously, we have imposed the restriction that $X_0 = 0$ w.p.1. We now turn to the general case $X_0 = x$ w.p.1, where x is any real number. In order to make explicit this new value of the initial state, we introduce the notation P_x to refer to the probability measure that satisfies $P_x(X_0 = x) = 1$ (the same is valid for E_x, the expected value operator under P_x).

A first important result is related to the way we represent Brownian motions (BM). Of course, we have already given a concrete definition of a BM, however, let look at an alternative representation.

We know that $X_{t+s} - X_t$ has a $N(\mu s, \sigma^2 s)$ distribution. Thus, the transition density

$$p(t,x,y)dy \equiv P_x(X_t \in dy) = \frac{1}{\sigma\sqrt{t}}\phi\left(\frac{y-x-\mu t}{\sigma\sqrt{t}}\right)dy$$

satisfies the following differential equation:

$$\frac{\partial}{\partial t}p(t,x,y) = \left(\frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2} + \mu\frac{\partial}{\partial x}\right)p(t,x,y),$$

with initial condition

$$p(0,x,y) = \delta(x-y) = \begin{cases} 1 & \text{if } y = x, \\ 0 & \text{otherwise.} \end{cases}$$

The differential equation above characterizes BM and is called *Kolmogorov's backward equation.* If instead of differentiating with respect to the initial state x, we differentiate with respect y, the final state, we get the *Kolmogorov's forward equation*

$$\frac{\partial}{\partial t}p(t,x,y) = \left(\frac{1}{2}\sigma^2\frac{\partial^2}{\partial y^2} - \mu\frac{\partial}{\partial y}\right)p(t,x,y).$$

In the special case when $\mu = 0$, the previous equation reduces to the traditional *heat equation* (or diffusion equation), for this reason Brownian motion are usually called *diffusion processes.*

3.5 Hitting Time Problem

An important application of BM is the *Hitting Time* problem. That is, the problem of determining the first time when the process reaches a predefined state. Let define $T(y) = inf\{t \geq 0 : X_t = y\}$, i.e., the first time at which X reaches the value y. Suppose that the process start at $x \geq 0$ and let $0 < x \leq b$. Then, we are interested in finding the distribution of $T \equiv T(0) \wedge T(b)$. A first step is the following result:

Property 8

$$E_x(T) < \infty,\ 0 \leq x \leq b.$$

The proof is based on the *martingale stopping* theorem, that is

Theorem 4 Martingale Stopping Theorem

Let T be a stopping time and X a martingale (with right-continuous sample paths) on certain filtered probability space. Then the stopped process $\{X(t \wedge T), t \geq 0\}$ is also a martingale.

Thus, if we apply this result to $M_t = X_t - \mu t$, which is clearly a martingale, we have that:

$$E_x(M(T \wedge t)) = E_x(M(0)) = x.$$

But $E_x(M(T \wedge t)) = E_x(X(T \wedge t)) - \mu E_x(T \wedge t)$. Thus, for $\mu \neq 0$, the result in (8) follows directly. For the case, $\mu = 0$, we have to apply the martingale stopping theorem to the martingale $X_t^2 - \sigma^2 t$.

Let us now recall *Wald martingale* introduced in property (5), that is, $V_\beta(t) = e^{\beta X_t - q(\beta)t}$ where the function $q(\cdot)$ is given by $q(\beta) = \mu\beta + \frac{\sigma^2\beta^2}{2}$. Now, it can be shown that

$$E_x(V_\beta(T)) = E_x(V_\beta(0)) = e^{\beta x},\ 0 \leq x \leq b.$$

Therefore, we have the following decomposition:

$$\begin{aligned} e^{\beta x} &= E_x(V_\beta(T); X_T = 0) + E_x(V_\beta(T); X_T = b), \\ &= \psi_*(x|q(\beta)) + e^{\beta b}\psi^*(x|q(\beta)), \end{aligned} \tag{8}$$

where $\psi_*(x|\lambda) \equiv E_x(e^{-\lambda T}; X_T = 0)$ and $\psi^*(X|\lambda) \equiv E_x(e^{-\lambda T}; X_T = b)$. Solving the equation $q(\beta) = \lambda$, we get:

$$\beta_*(\lambda) = \frac{\mu + \sqrt{\mu^2 + 2\sigma^2\lambda}}{\sigma^2} > 0\ ;\ \beta^*(\lambda) = \frac{\mu - \sqrt{\mu^2 + 2\sigma^2\lambda}}{\sigma^2} < 0.$$

Thus, combining this result and (8), we get the following system of equation:

$$\begin{aligned} e^{-\beta_*(\lambda)x} &= \psi_*(x|\lambda) + e^{-\beta_*(\lambda)b}\psi^*(x|\lambda), \\ e^{-\beta^*(\lambda)x} &= \psi_*(x|\lambda) + e^{-\beta^*(\lambda)b}\psi^*(x|\lambda). \end{aligned}$$

The solution of this system gives the following result.

Property 9 *Let $\lambda > 0$ be fixed. For $0 \leq x \leq b$,*

$$\begin{aligned} \psi^*(x|\lambda) &= \frac{\theta^*(x,\lambda) - \theta_*(x,\lambda)\theta^*(0,\lambda)}{1 - \theta_*(b,\lambda)\theta^*(0,\lambda)}, \\ \psi_*(x|\lambda) &= \frac{\theta_*(x,\lambda) - \theta^*(x,\lambda)\theta_*(b,\lambda)}{1 - \theta_*(b,\lambda)\theta^*(0,\lambda)}, \\ \theta_*(x,\lambda) &= e^{-\beta_*(\lambda)x}, \\ \theta^*(x,\lambda) &= e^{\beta^*(\lambda)(b-x)}. \end{aligned}$$

Finally, from the previous result, we can obtain the distribution (or more precisely the Laplace transform) of T.

Proposition 1 *Let θ^* and θ_* be defined as above. Then,*

$$E_x(e^{-\lambda T(0)}; T(0) < \infty) = \theta_*(x,\lambda)\ \ ;\ \ E_x(e^{-\lambda T(b)}; T(b) < \infty) = \theta^*(x,\lambda),\ \ 0 \leq x \leq b.$$

In addition, if $\mu = 0$ then $P_x(X_T = b) = \frac{x}{b}$. Otherwise,

$$P_x(X_T = b) = \frac{1 - \xi(x)}{1 - \xi(b)}\ \ ;\ \ \xi(z) \equiv e^{\frac{-2\mu z}{\sigma^2}},\ \ 0 \leq x \leq b.$$

4 Stochastic Calculus

The main goals of this section is to present Itô's lemma and the use of stochastic differential equations as important tools for modelling stochastic processes. We start the analysis introducing heuristically *Itô's Stochastic differential equation*.

4.1 Motivation

It is a common practice when modelling physical systems to express the dynamics of the system, i.e., its evolution over time through a difference or differential equation. For example, when describing the position ($y(t)$) of a certain object at time t, we might use the relation

$$\frac{dy(t)}{dt} = v(t),$$

where $v(t)$ is the instantaneous velocity of the object at time t. In general, differential equations have been used extensively in science, and probably one of their biggest advantages is that they are able to capture the essence of the physical system without incorporating the natural and necessary difficulties that are imposed by border conditions.

Let us now look at the "general" (deterministic) differential equation:

$$\frac{dx(t)}{dt} = f(t, x(t)).$$

Solving this equation is an old problem in mathematics, and it is not the purpose of this note to go into the details of how to solve it. We would like, however, to introduce some type of uncertainty into the model. One easy way of doing this is to use the traditional trick used by econometricians, that is, to simply add an stochastic term to the above relation. In order to do that, we proceed as follows. We first approximate the dynamics by:

$$x(t + \Delta t) - x(t) = f(t, x(t))\Delta t + o(\Delta t),$$

where $o(t)$ is function such that $t^{-1}o(t) \to 0$ as $t \to 0$. If we assume now that uncertainty can be model by an stochastic process $v(t)$ that we simply add in to the dynamics of the system, we have:

$$x(t + \Delta t) - x(t) = f(t, x(t))\Delta t + v(t + \Delta t) - v(t) + o(\Delta t).$$

In particular, we might think on this uncertainty as being the sum of independent and small perturbations. Thus, a reasonable model is to suppose that

$$v(t + \Delta t) - v(t) = \sigma(t, x(t))(z(t + \Delta t) - z(t)),$$

where $z(t)$ is Wiener process and σ accounts for the variance of v. We can then rewrite the dynamics of the system as follows:

$$dx = f(t, x)dt + \sigma(t, x)dz,$$

which is called *Itô's stochastic differential equation*. Notice that we can note divide by dt above since z is nowhere differentiable.

4.2 Stochastic Integration

Since the Wiener process is nowhere differentiable, Itô's differential equation does not have a clear meaning *perse*. In this subsection, we will give it one, which is based on the notion of stochastic integral. The idea is the following, we use the notion:

$$dx = f(t,x)dt + \sigma(t,x)dz,$$

as a shorthand for

$$x(t) = x(0) + \int_0^t f(s,x(s))ds + \int_0^t \sigma(s,x(s))dz(s).$$

The first integral in the right-hand side is understood in the usual Riemann sense. The second integral, however, does not have a clear meaning for the reasons that we have already mentioned. Let us then focus in the following stochastic process:

$$I_t(X) = \int_0^t X_s \, dW_s, \quad t \geq 0, \tag{9}$$

that we called the *stochastic integral*. Here X is any stochastic process and W is a wiener process. Stochastic integration was first presented by Itô (1944) and extended later by Doob (1953). Here, we will not go into the formal details behind the theory of stochastic integration. We will rather give a more simpler and intuitive analysis.

From traditional calculus, we know that if a function is relatively well-behaved in the interval $[0,t]$ (i.e., it is integrable), then we can approximate the value of

$$I = \int_0^t f(s)\, ds,$$

as follows. We first introduce a sequence of partitions $\{P_n : n \geq 1\}$ where $P_n = \{t_i : 0 \leq i \leq n\}$ is a partition of the interval $[0,t]$, i.e., $0 = t_0 < t_1 < \cdots < t_n = t$. We denote by $||P_n|| = \max\{t_i - t_{i-1} : 1 \leq i \leq n\}$. Then, if $\lim_{n\to\infty} ||P_n|| = 0$, we have that

$$I = \lim_{n\to\infty} \sum_{t_i \in P_n} f(\xi_i)(t_{i+1} - t_i),$$

where $\xi_i \in [t_i, t_{i+1}]$. In particular, $\xi_i = t_i$ or $\xi_i = t_{i+1}$ does not make much difference for deterministic real-valued function. This analysis is exactly the one that we will apply to compute stochastic integral, however, the analysis requires some extra attention.

Let H^2 the space of $\mathcal{F}_t$-adapted stochastic processes X such that $\mathrm{E}[\int_0^T X^2(s)\,\mathrm{d}s] < \infty$. We introduce a special class of stochastic processes that we call *simple*. A process $X \in H^2$ is *simple* if there exist a sequence of times $\{t_k\}$ such that

$$0 = t_0 < t_1 < \cdots < t_k \to \infty$$

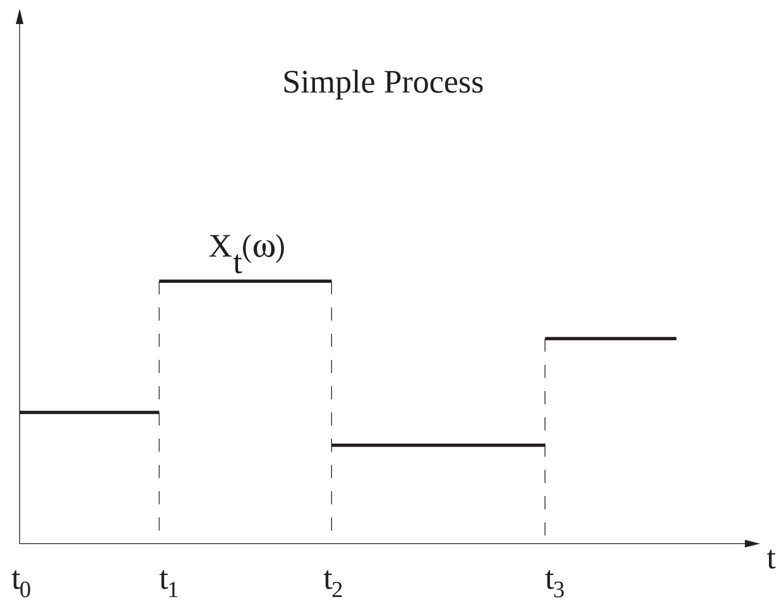

Figure 2: *Simple process.*

and

$$X(t,\omega) = X(t_{k-1},\omega), \ \ \forall t \in [t_{k-1}, t_k) \ \ k = 1, 2, \ldots.$$

We notice that the sequence $\{t_k\}$ is independent of ω. Given the special form of simple processes, it is possible to give a clear definition of the stochastic integral in this case. In fact, if X is simple, then

$$I(X) = \int_0^t X\,dW = \sum_{k=0}^{n-1} X(t_k)[W(t_{k+1}) - W(t_k)],$$

where $t_0 = 0$ and $t_n = t$. The importance of simple processes is not only that we are able to compute easily their integrals but also that they can be use to approximate other more complex processes. That is,

Property 10 *Let consider the stochastic process $X \in H^2$, i.e.,*

$$E\left[\int_0^t X^2(s)\,ds\right] < \infty, \ \ \forall t \geq 0.$$

Then, there exists a sequence of simple processes $\{X_n\} \in H^2$ such that

$$X_n \Rightarrow_n X.$$

Moreover, the value of $I_t(X)$ can be obtained from the fact that

$$I_t(X_n) \Rightarrow_n I_t(X).$$

Let define the norm $||\cdot||$ in H^2 as follows:

$$||X|| = E\left[\int_0^t X^2(s)\,ds\right]^{\frac{1}{2}}.$$

Then we have the following result

Proposition 2 *Let $X \in H^2$, then $E[I_t(X)] = 0$ and $||I_t(X)|| = ||X||$.*

Example 1:

Let consider the case when $X = W$, it is not hard to show that $W \in H^2$, moreover, $||W|| = \frac{t}{\sqrt{2}}$. Now, in order to compute $I(W)$ we introduce the simple processes

$$X_n(s) = W\left(\frac{kt}{2^n}\right), \quad s \in \left[\frac{kt}{2^n}, \frac{(k+1)t}{2^n}\right).$$

Let define $t_k = \frac{kt}{2^n}$. Then, for these simple processes we have:

$$\begin{aligned} I_t(X_n) &= \sum_{k=0}^{2^n-1} W(t_k)[W(t_{k+1}) - W(t_k)] \\ &= \frac{1}{2}\sum_{k=0}^{2^n-1}[W^2(t_{k+1}) - W^2(t_k)] - \frac{1}{2}\sum_{k=0}^{2^n-1}[W(t_{k+1}) - W(t_k)]^2 \\ &= \frac{1}{2}W^2(t) - \frac{1}{2}\sum_{k=0}^{2^n-1}[W(t_{k+1}) - W(t_k)]^2. \end{aligned}$$

But in equation (1), we saw that the summation above converges to t. Therefore, we conclude that:

$$I_t(W) \equiv \int_0^t W\,dW = \frac{1}{2}W^2(t) - \frac{t}{2}. \tag{10}$$

4.3 Itô's Lemma

In this section, we define the notion of *stochastic differential* and state and prove Itô's lemma which is the fundamental rule for computing stochastic differentials.

We start by defining some notation. As usual, we consider a probability space $(\Omega, \mathcal{F}, P)$, a Wiener process $W(t,\omega)$, a process $Y(t,\omega)$ that is jointly measurable in t and ω with respect to $\mathcal{F}_t$, is adapted and satisfies $\int_0^T |Y(t,\omega)|dt < \infty$ w.p.1. We also consider a process X that is non-anticipating on $[0,T]$. We say that Z is an *Itô process* if it has the following functional form:

$$Z(t,\omega) = Z(0,\omega) + \int_0^t X(s,\omega)dW(s,\omega) + \int_0^t Y(s,\omega)ds. \tag{11}$$

The first integral in the right-hand side is call the *Brownian component* of Z and it has to be computed according to the analysis that we did in the previous section. The second integral is called the *drift component* (or *VF component*) of Z and it is evaluated in the usual Reimann sense. Instead of using (11) to represent Z, we say that Z has an *Itô differential* (or *stochastic differential*) dZ given by:

$$dZ = XdW + Ydt.$$

Proposition 3 (Itô's Lemma)

Let $u(t,x)$ be a continuous non-random function with continuous partial derivates u_t, u_x, and u_{xx}. Suppose that Z is a process with stochastic differential $dZ = XdW + Ydt$. Let define the process $V(t) = u(t, Z(t))$, then V has a stochastic differential given by:

$$dV = \left[\frac{\partial}{\partial t}u(t,Z) + \frac{\partial}{\partial Z}u(t,Z)Y + \frac{1}{2}\frac{\partial^2}{\partial Z^2}u(t,Z)X^2\right]dt + \frac{\partial}{\partial Z}u(t,Z)XdW. \quad (12)$$

(The proof of the Lemma uses a second order Taylor expansion of the function $u(t,x)$.)

Let us take a look at Itô's lemma in a particular case. Let suppose that $u(t,Z) = f(Z)$, for some twice continuously differentiable function f. Then (12) implies:

$$df(Z) = \left[f'(Z)Y + \frac{1}{2}f''(Z)X^2\right]dt + f'(Z)XdW.$$

Rearranging terms we get:

$$\begin{aligned} df(Z) &= f'(Z)\left[Ydt + XdW\right] + \frac{1}{2}f''(Z)X^2 dt \\ &= f'(Z)dZ + \frac{1}{2}f''(Z)(dZ)^2. \end{aligned} \quad (13)$$

Relation (13) is a simplify way of expressing Itô's lemma and uses the convention $(dZ)^2 = (Ydt + XdW)^2 = X^2dt$. The idea is that in differential terms only $(dW)^2 \neq 0$, this is consistent with our finding in (1) about the quadratic variation of Wiener processes. Let notice that for ordinary differentials $df(Z) = f'(Z)dZ$, thus the second term in (13) is the main difference for stochastic differential that, as we have already mentioned, reflects the infinite variation of Brownian paths.

Example 2:

Solve the SDE

$$\mathrm{d}V_t = \alpha V_t \,\mathrm{d}t + \beta V_t \,\mathrm{d}W_t.$$

In order to solve the stochastic differential above, we introduce the following change of variable:

$$A = ln(V).$$

Then, using Itô's lemma we get:

$$dA = \underbrace{\frac{1}{V}dV}_{[1]} - \frac{1}{2V^2}\underbrace{(dV)^2}_{[2]}.$$

Thus, replacing [1] by $\alpha\,\mathsf{d}t + \beta \mathsf{d}W_t$ and [2] by $\beta^2 V^2 dt$ we get:

$$dA \;=\; \left(\alpha - \frac{1}{2}\,\beta^2\right)\,\mathsf{d}t + \beta\,\mathsf{d}W_t.$$

This linear differential implies $A_t = (\alpha - \frac{1}{2}\,\beta^2)\,t + \beta W_t$. Finally, combining this result and the transformation $A = ln(V)$ we conclude:

$$V(t) = e^{(\alpha - \frac{1}{2}\,\beta^2)t + \beta W(t)}.$$

That is, V is a *geometric Brownian motion.*

Exercise 2: Compute $\mathsf{E}[V_t]$ and $\mathsf{E}[V_t^2]$.

Theorem 5 (Integration by Parts)
Suppose that $f(s)$ is a deterministic continuous function of bounded variation in $[0, t]$. Then

$$\int_0^t f(s)\,dW_s = f(t)\,W_0 - \int_0^t W_s\,df(s).$$

We finish this section with an important result about the existence of solutions for stochastic differential.

Theorem 6 *Let consider the Itô process Z defined through the following stochastic differential:*

$$dZ(t) = a(t, Z(t))dt + \sigma(t, Z(t))dW(t),$$

with initial condition $Z(0, \omega) = c(\omega) = c$. If

1. *a and σ are both measurable with respect to all their arguments,*
2. *There exists a constant $K > 0$ such that*

$$|a(t, x) - a(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq K|x - y|,$$

$$|a(t, x)|^2 + |\sigma(t, x)|^2 \leq K^2(1 + |x|^2),$$

3. *The initial condition $Z(0, \omega)$ does not depend on $W(t)$ and $E[Z(0, \omega)^2] < \infty$.*

Then, there exists a solution $Z(t)$ satisfying the initial condition which is unique w.p.1, has continuous paths and $sup_t E[Z(t)^2] < \infty$.

5 Continuous Time Markov Processes

In this section, we define and characterize Markov diffusion processes in continuous time and establish their connection to SDE's and PDE's.

5.1 Definitions

Definition 7 (Markov Process)

A stochastic process X is a Markov process if

1. *For any sequence of times $t_1 < t_2 < \cdots < t_m < t$ and $B \in \mathcal{B}(\mathbb{R})$*

$$P(X(t) \in B \,|\, X(t_1), X(t_2), \ldots, X(t_m)) = P(X(t) \in B \,|\, X(t_m)).$$

2. *The function $\hat{P}(s, y, t, B)$ defined as*

$$\hat{P}(s, y, t, B) = P(X(t) \in B \,|\, X(s) = y)$$

   is $\mathcal{B}(\mathbb{R})$-measurable for fixed s, t, B and a probability measure on $\mathcal{B}(\mathbb{R})$ for fixed s, y, t.

3. *The Chapman-Kolmogorov equation*

$$\hat{P}(s, y, t, B) = \int_{\mathbb{R}} \hat{P}(r, x, t, B)\, \hat{P}(s, y, r, dx)$$

   holds for $s < r < t$.

The function $\hat{P}(s, y, t, B)$ is called the transition probability *of X.*

Let X be a one-dimensional Markov process with transition probability $\hat{P}(s, x, t, A)$.

Definition 8 *A Markov process $X(t)$ is called a* diffusion process *if the following conditions hold.*

1. *For every x and $\epsilon > 0$*

$$\int_{|x-y|>\epsilon} \hat{P}(s, x, t, dy) = o(t - s)$$

   uniformly over $s < t$.

2. *There exist functions $a(t, x)$ and $b(t, x)$ such that for every x and $\epsilon > 0$*

$$\int_{|x-y|\leq\epsilon} (y - x)\hat{P}(s, x, s, dy) = a(s, x)(t - s) + o(t - s),$$

$$\int_{|x-y|\leq\epsilon} (y - x)^2\hat{P}(s, x, s, dy) = b(s, x)(t - s) + o(t - s),$$

   uniformly over $s < t$.

The function $a(t, x)$ is called the local drift *and the function $b(t, x) := \sigma^2(t, x)$ is called* local diffusion*.*

The following results establishes the relationship between diffusion processes and stochastic differential equations.

Theorem 7 *Let $a(t,x)$ and $\sigma(t,x)$ denote two functions that satisfy the assumptions of Theorem 6 and let $X(s)$ denote a process defined for $s \in [t,T]$ that is a solution of the stochastic differential equation*

$$X(s) = X(t) + \int_t^s a(\tau, X(\tau))\, d\tau + \int_t^s \sigma(\tau, X(\tau))\, dB(u).$$

Then the process $X(s)$ is a diffusion process whose transition probabilities are given by the relation

$$\hat{P}(t,x,s,A) = P(X(s) \in A \,|\, X(t) = x) = P(X_{t,x}(s) \in A),$$

where the process $X_{t,x}(s)$ satisfies the SDE

$$X_{t,x}(s) = x + \int_t^s a(\tau, X_{t,x}(\tau))\, d\tau + \int_t^s \sigma(\tau, X_{t,x}(\tau))\, dB(u).$$

5.2 Connection between Diffusion Processes and PDE's

We will now establish the connection between diffusion processes and PDE's. As before, we will restrict the exposition to one-dimensional diffusions. The extension to higher dimensions does not involve new ideas.

As before let a and σ^2 be the drift and diffusion coefficient of diffusion process X with transition probability $\hat{P}(s,x,t,\mathrm{d}y)$. Given a $s < t$, we define the linear operator $S_{s,t}$ associated to a Markov process X as follows.

For all bounded, real valued, measurable function f, $S_{s,t}(f(y)) = \displaystyle\int_{\mathbb{R}} f(x)\, \hat{P}(s,y,t,\mathrm{d}x) = \mathrm{E}_{sy}[f(X(t))]$.

We note that the Chapman-Kolmogorov condition implies

$$S_{s,t} = S_{s,r}\, S_{r,t}, \qquad \text{for } s < r < t.$$

Note that $S_{t,t}(f) = f$.

Definition 9 *For a given Markov process X, we define its* generator $\mathcal{A}(t)$ *as follows:*

$$\textit{For all bounded, real valued, measurable function } f, \quad \mathcal{A}(t)\, f = \lim_{h \downarrow 0} \frac{S_{t,t+h}(f) - f}{h}.$$

Consider a function $\psi(t,x)$. Then under suitable restrictions on ψ we can write

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} S_{s,t}\, \psi(t,y) &= \lim_{h \downarrow 0} \frac{\mathrm{E}_{sy}[\psi(t+h, X(t+h)) - \psi(t, X(t))]}{h} \\ &= \lim_{h \downarrow 0} \frac{\mathrm{E}_{sy}[\psi_t(t, X(t+h))\, h] + \mathrm{E}_{sy}[\psi(t, X(t+h)) - \psi(t, X(t))]}{h} \\ &= E_{sy}[\psi(t, X(t)) + \mathcal{A}(t)\psi(t, X(t))] \end{aligned}$$

So we can rewrite this last equality as follows

$$\frac{\mathrm{d}}{\mathrm{d}t} S_{s,t}\,\psi = S_{s,t}\,[\psi + \mathcal{A}(t)\psi].$$

Integrating this last expression over t and interchanging integral and expectation (no formal prove of this interchange is attempted here) we get

$$\mathrm{E}_{sy}[\psi(t, X(t))] - \psi(s, y) = \mathrm{E}_{sy}\left[\int_s^t (\psi_t(\tau, X(\tau)) + \mathcal{A}(\tau)\psi(\tau, X(\tau)))\ \mathrm{d}\tau\right].$$

Proposition 4 *Let $X(t) = X_{s,x}(t)$ be a diffusion process of the form*

$$X_{s,x}(t) = x + \int_s^t u(\tau, X(\tau))\, d\tau + \int_s^t v(\tau, X(\tau))\, dB(\tau).$$

where u and v are bounded functions in L^2. Then for a stopping time $T > s$ and a twice-continuously differentiable function f we have

$$E_{sx}[f(X(T)] = f(x) + E_{sx}\left[\int_s^T \left(u(\tau, X(\tau))\,\frac{\partial f(X(\tau))}{\partial x} + \frac{1}{2} v^2(\tau, X(\tau))\,\frac{\partial^2 f(X(\tau))}{\partial x^2}\right)\ d\tau\right].$$

Corollary 1 *For a diffusion process*

$$dX(t) = a(t, X(t))\, dt + \sigma(t, X(t))\, dB(t)$$

the infinitesimal generator is given by

$$\mathcal{A} = a(t,x)\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2(t,x)\,\frac{\partial^2}{\partial x^2}.$$

The following results, due to Kolmogorov, is similar to those results presented on section (3.4).

Theorem 8 (Kolmogorov Backward Equation) *Let $\varphi(x)$ denote a continuous bounded function such that the function $u(t,x) = \int \varphi(y)\hat{P}(s,x,t,dy) = S_{st}\varphi(X(t))$ has bounded continuous first and second derivatives with respect to x and let the function $a(t,x)$ and $b(t,x)$ be continuous. Then $u(t,x)$ has a derivative $\partial u/\partial t$ which satisfies the equation*

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

and $u(t,x)$ satisfies the boundary condition $\lim_{t\uparrow s} u(t,x) = \varphi(x)$.