MATHEMATICS OF FINANCE

CHAPTER 3

THE TOOLS FOR CONTINUOUS-TIME MODELING OF ASSET PRICES

JOSÉ E. FIGUEROA-LÓPEZ

INSTRUCTOR'S CLASS NOTES
SPRING 2009

CONTENTS

Date: April 1, 2009.

1. The Wiener process and the Black-Scholes model

1.1. **Definition.**

In this part we will define the Wiener process, which is the fundamental building block of continuous-time price models. We show a construction procedure for the Wiener process based on random walks appropriately rescaled in time and space. As a byproduct, we will revisit the Binomial implementation described at the end of the previous chapter and show that the sequence of Binomial models and their associated option prices converge when the number of time steps become large. The limit process of the Binomial model is the celebrated Black-Scholes model.

A *stochastic process* is a collection of random variables $X = \{X_t\}_{t\geq 0}$ indexed by time t defined on a common sample space Ω representing all possible states ω of the “market" or “nature". The sample space is endowed with a probability measure P. We can think of a stochastic process $\{X_t\}_{t\geq 0}$ as a random measurement evolving in time (e.g. the price of a stock). Another natural interpretation is to see X is as a random function: for each state of nature $\omega \in \Omega$, we observe the random path $t \to X_t(\omega)$. Let us give the formal definition of a *Wiener process*.

Definition 1.1. A stochastic process $\{W_t\}_t$ defined on (Ω, P) is a *Wiener process* if the following conditions hold:

(1) $W_0 = 0$;
(2) The increment or change of the process during a time period $[s, t]$, namely $W_t - W_s$ with $s < t$, is Normally distributed with mean 0 and variance $t - s$;
(3) For any $0 \leq t_0 < \cdots < t_n$, the corresponding increments $W_{t_1} - W_{t_0}, \ldots, W_{t_n} - W_{t_{n-1}}$ are mutually independent;
(4) The graph of the function $t \to W_t$, called the path of the process, is continuous with probability 1.

Notice that condition (3) means that the increments in disjoint time intervals are independent. A process with this property is said to have *independent increments*. Condition (2) can be weaken substantially. It can be shown that all what is needed is the following:

(2a) The distribution of the increment $W_t - W_s$ depends only on the time span $t - s$;
(2b) $E(W_t) = 0$.

A process satisfying (2a) is said to have *stationary increments*.

Exercise 1. Compute the covariance function $\mathrm{Cov}(W_t, W_s)$ of the Wiener process W.

One can statistically corroborate a Wiener process by analyzing the statistical properties of a large number of evenly spaced sampling observations W_Δ, $W_{2\Delta}$, etc. The independence of increments can be corroborated by plotting the increments $W_{k\Delta} - W_{(k-1)\Delta}$ against the time $k\Delta$. The set of points should be centered around the origin without any special patterns and constant variability along time. The empirical covariance

$$\hat{r}(\Delta) := \frac{1}{n-1}\sum_{k=0}^{n-2}(W_{k+1\Delta} - W_{k\Delta})(W_{(k+2)\Delta} - W_{(k+1)\Delta})$$

should be close to 0 and the histogram of the increments should exhibit the traditional bell-shaped form of a Gaussian distribution.

Notice that the path of the Brownian motion should look very jagged, erratic, and non-smooth. Indeed, if we only plot the Wiener process at times Δ, 2Δ, etc. and Δ is small, the change of the process from one time step to the next can be positive or negative with the same probability. Due to this erratic behavior, a Wiener process is a also called a *Brownian motion* in honor of the botanist Robert Brown who observe for the first time a “Wiener-like" process in nature. He observed that a pollen particle immerse in a liquid exhibit such a erratic behavior. One can modeled the position of the particle in time as (W_t^1, W_t^2, W_t^3) were the W^i's are independent Wiener processes.

1.2. **Construction by random walks.**

In this part we give a procedure to simulate a path of the Brownian motion. Suppose that for each $n \geq 1$, $\{U_i^n\}_{i\geq 0}$ are independent random variable with common distribution such that $E(U_i^n) = 0$ and $\mathrm{Var}(U_i^n) = 1$. Consider the process

$$X_t^n = \frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]} U_k^n$$

where $[x]$ is the integer part of x. The above expression looks fancy, but it is easy to describe its graph. This is a piece-wise constant process that jumps at times $t = 1/n, 2/n, \ldots$ by the respective sizes $U_1^n/\sqrt{n}, U_2^n/\sqrt{n}, \ldots$. In other words, the graph of X^n takes the graph of the *random walk*

$$X_t^1 = \sum_{k=1}^{[t]} U_k^1 = \begin{cases} U_1^1, & 0 \leq t < 1 \\ U_1^1 + U_2^1, & 2 \leq t < 3 \\ \vdots & \vdots \end{cases}$$

and shrink the time by a factor of n and the space by a factor of $1/\sqrt{n}$.

Notice that the increment

$$X_v^n - X_u^n = \frac{1}{\sqrt{n}}\sum_{k=[nu]+1}^{[nv]} U_k^n,$$

is such that

$$EX_u^n = 0, \quad \mathrm{Var}(X_v^n - X_u^n) = \frac{[nv]-[nu]}{n} \to v - u, \quad u < v,$$

and also, $X_v^n - X_u^n$ is independent of $X_t^n - X_r^n$ whenever $r < t < u < v$. Also, for large n, $X_v^n - X_u^n$ has approximately the same distribution as $X_t^n - X_r^n$, whenever $0 < v - u = t - r$. In conclusion, for large n, X^n seems to have almost all conditions of a Wiener process and hence should look like a Wiener process. Indeed, one can prove that

$$\{X_t^n\}_{t\geq 0} \xrightarrow{\mathfrak{D}} \{W_t\}_{t\geq 0}, \quad \text{as} \quad n \to \infty. \tag{1}$$

The above result is called the *Donsker's invariance theorem*. In particular, we have that for a fixed time T,

$$\lim_{n\to\infty} Eg(X_T^n) = Eg(W_T), \tag{2}$$

for any continuous bounded function g. Let us remark that the sequence of random variables $\{U_k^n\}_k$ might not even be defined on the same probability space and hence the expectations on the left-hand side of (2) could actually be different for each n, and also, different from the expectation on the right-hand side.

1.3. An application: Convergence of the Binomial model to the Black-Scholes model.

Consider a discrete-time market model with evenly spaced "trading" times $0 = t_0 < t_1 < \cdots < t_n = 1$, where $\Delta_n := t_i - t_{i-1} = 1/n$ (for simplicity, we take $T = 1$). Suppose that $\{S_t^n\}_{t\geq 0}$ is the stock price process that jumps only at times t_i according to a Binomial model and remains unchanged between trading times $t_0, \dots, t_n$. Thus,

$$\widetilde{S}_0^n = S_0, \quad S_{t_i}^n = \begin{cases} u_n S_{t_{i-1}}^n & \text{with prob. } p_u^n, \\ d_n S_{t_{i-1}}^n & \text{with prob. } p_d^n, \end{cases} \quad \text{and} \quad S_t^n = S_{t_{i-1}}, \quad \text{if} \quad t \in [t_{i-1}, t_i)$$

Let us assume the CRR Binomial implementation where

$$u_n = \frac{1}{d_n} = e^{\sigma\sqrt{\Delta_n}}, \quad p_u^n = \frac{1}{2} + \frac{\mu}{2\sigma}\sqrt{\Delta_n}. \tag{3}$$

Our first task is to show that the Binomial stock prices $\{S_t^n\}_{t\geq 0}$ converge to a process $\{S_t\}_{t\geq 0}$, which we shall identify explicitly. Remember that the parameters (3) are chosen so that the log return $Z_k^n = \log \frac{S_{t_k}^n}{S_{t_{k-1}}^n}$ satisfies

$$\frac{1}{\Delta_n} \cdot E\left(Z_k^n\right) \xrightarrow{n\to\infty} \mu, \quad \frac{1}{\Delta_n} \cdot \mathrm{Var}\left(Z_k^n\right) \xrightarrow{n\to\infty} \sigma^2. \tag{4}$$

The key is to observe that

$$S_t^n = S_0 \exp\left\{\sum_{k=1}^{[nt]} Z_k^n\right\}.$$

In terms of the normalized log returns

$$U_k^n = \frac{Z_k^n - EZ_k^n}{\sqrt{\mathrm{Var}(Z_k^n)}},$$

the exponent can be written as follows:

$$\sum_{k=1}^{[nt]} Z_k^n = \sqrt{\frac{1}{\Delta_n}\mathrm{Var}(Z_k^n)} \cdot \left\{ \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} U_k^n \right\} + \frac{[nt]}{n} \left(\frac{1}{\Delta_n} \cdot E\left(Z_k^n\right) \right),$$

where we used that $\Delta_n = T/n = 1/n$. In light of Donsker's invariance theorem and (4), the above process converges to $\sigma W_t + \mu t,$. We conclude that $\{S_t^n\}_{t\geq 0} \xrightarrow{\mathfrak{D}} \{S_t\}_{t\geq 0}$ where

$$S_t := S_0 e^{\sigma W_t + \mu t}, \quad t \geq 0. \tag{5}$$

Such a process is called the *Geometric Brownian motion* or the *Black-Scholes model*, though was originally proposed by Samuelson (1965).

If we are working with the risk-neutral probability measure Q_n, then we will obtain that

$$\{S_t^n\}_{t\geq 0} \xrightarrow{\mathfrak{D}} \{S_0 e^{\sigma W_t + (r - \frac{\sigma^2}{2})t}\}_{t\geq 0}$$

since we saw that

$$\frac{1}{\Delta_n} \cdot E^{Q_n}\left\{ \log \frac{S_{t_i}^n}{S_{t_{i-1}}^n} \right\} \overset{n\to\infty}{\longrightarrow} r - \frac{1}{2}\sigma^2, \quad \frac{1}{\Delta_n} \cdot \mathrm{Var}^{Q_n}\left\{ \log \frac{S_{t_i}^n}{S_{t_{i-1}}^n} \right\} \overset{n\to\infty}{\longrightarrow} \sigma^2.$$

One important consequence is that the CRR Binomial option prices of a simple contingent claim, say

$$\Pi_n(0) := e^{-rT} E^{Q_n} \left\{ \Phi\left(S_T^n\right) \right\},$$

converges to

$$\Pi(0) = e^{-rT} E^{Q} \left\{ \Phi\left(S_0 e^{\sigma W_T + (r - \frac{\sigma^2}{2})T} \right) \right\},$$

whenever Φ is bounded and continuous. Above, $\{W_t\}_t$ is Wiener process on certain sample space equipped with the probability measure Q.

As a matter of an example, consider a put option with strike K. Taking $\Phi(x) = (K - x)_+$, when $x \geq 0$ and $\Phi(x) = K$ for $x < 0$, we conclude that the CRR Binomial prices of a put option with strike K converges to

$$\Pi^{\mathrm{put}}(0) = e^{-rT} E^{Q} \left\{ \left(K - S_0 e^{\sigma W_T + (r - \frac{\sigma^2}{2})T} \right)_+ \right\}. \tag{6}$$

Exercise 2. Justify that the CRR Binomial prices of a call option with strike K converges to

$$\Pi^{\mathrm{call}}(0) = e^{-rT} E^{Q} \left\{ \left(S_0 e^{\sigma W_T + (r - \frac{\sigma^2}{2})T} - K \right)_+ \right\}. \tag{7}$$

The formulas (6)-(7) lead to the celebrated *Black-Scholes option prices*, which became the gold standard in industry.

2. INFORMATION PROCESS AND CONDITIONAL EXPECTATION

Information process plays a crucial role in defining a time-continuous model for option pricing. For instance, the minimal requirement of a trading strategy is that at any given time, the position on an asset is determined only by the information generated by the asset's price up to that time.

2.1. **Definition.**

We fix a time horizon $[0,T]$. Let $Z := \{Z_t\}_{t\leq T}$ be a process defined on a sample space Ω. Typically, Z is the stock prices process $S := \{S_t\}_{t\leq T}$ itself. We first define the concept of "information generated by Z". Notice that if we knew what the state of nature $\omega \in \Omega$ is, then we will know the whole path of $u \in [0,T] \to Z_u$. However, by time $t \in [0,T]$, all the information about the state of nature ω is contained only in the past evolution of the process Z.

If it is possible to decide whether a given event $A \subset \Omega$ occurred or not based upon the observation of the path of Z up to time t, then we say that A is part of the information generated by Z up to time t. The collection of all sets A generated by Z up to time t is denoted by $\mathcal{F}_t^Z$. If $A \in \mathcal{F}_t^Z$, we say that the event A is $\mathcal{F}_t^Z$-measurable. The most fundamental examples of events $A \in \mathcal{F}_t^Z$ are of the form

$$A := \{Z_{t_1} \in C_1, \dots, Z_{t_n} \in C_n\} := \{\omega \in \Omega : Z_{t_1}(\omega) \in C_1, \dots, Z_{t_n}(\omega) \in C_n\}, \tag{8}$$

for "nice" sets of real numbers $C_1, \dots, C_n$ and times $0 \leq t_0 < \cdots < t_n \leq t$. Notice however that events such as

$$A := \{\omega : \sup_{u\leq t} Z_u(\omega) \geq H\}$$

belong to $\mathcal{F}_t^Z$ but can't be written in the form (8). In the case $\mathcal{F}_0^Z$, also denoted hereafter $\mathcal{F}^{Z_0}$, all sets $A \in \mathcal{F}_0^Z$ are of the form

$$A := \{Z_0 \in C\} := \{\omega \in \Omega : Z_0(\omega) \in C\},$$

for a set of real numbers C. Notice that if Z_0 is constant, then

$$\mathcal{F}^{Z_0} = \{\Omega, \emptyset\}.$$

Exercise 3. Suppose that Z is a continuous process (that is, the paths $t \in [0,T] \to Z_t(\omega)$ are continuous for all states of nature $\omega \in \Omega$). Does the event

$$A := \{\omega : \lim_{u\downarrow t} Z_u(\omega) \geq 1\},$$

belong to $\mathcal{F}_t^Z$?

We say that a random variable X is $\mathcal{F}_t^Z$*-measurable*, and write $X \in \mathcal{F}_t^Z$, if the value of X can be completely determined from the values of Z_u for $u \leq t$. The the most fundamental $\mathcal{F}_t^Z$*-measurable* random variables X are of the form

$$X = f(Z_{t_1}, \dots, Z_{t_n}),$$

for a “nice" function f and times $0 \leq t_0 < \cdots < t_n \leq t$. In the case $\mathcal{F}_0^Z$, a random variable X is measurable if

$$X = f(Z_0).$$

Roughly speaking, we can imagine that a r.v. $X \in \mathcal{F}_t^Z$ is of the form

$$X := \Phi(Z_u, u \leq t),$$

where Φ is a “nice" function that assigns a real number to each function φ defined on $[0, t]$. For instance,

$$X := \sup_{u \leq t} Z_u,$$

is $\mathcal{F}_t^Z$-measurable.

The collection $\{\mathcal{F}_t^Z\}_{t \leq T}$ is called the *filtration* generated by Z. A process $H = \{H_t\}_{t \leq T}$ is said to be *adapted* to Z or to the filtration $\{\mathcal{F}_t^Z\}_{t \leq T}$ if H_t is $\mathcal{F}_t^Z$-measurable, for all $t \in [0, \infty)$.

2.2. **Conditional expectation with respect to a filtration.**

Suppose that $\{X_t\}_{t \geq 0}$ is the underlying information process and Y is a random variable such that $EY^2 < \infty$. The conditional expectation of Y given the filtration $\mathcal{F}_t^X$ is that random variable $E\left(Y | \mathcal{F}_t^X\right)$ such that

(1) $E\left(Y | \mathcal{F}_t^X\right)$ is $\mathcal{F}_t^X$-measurable or in simpler terms, $E\left(Y | \mathcal{F}_t^X\right)$ is a function of the values of X up to time t;
(2) $Y^* = E\left(Y | \mathcal{F}_t^X\right)$ is closest to Y, in a the mean-square sense, than any other $\mathcal{F}_t^X$-measurable variable; that is,

$$E\left(Y^* - Y\right)^2 \leq E\left(Y' - Y\right)^2,$$

for any $Y' \in \mathcal{F}_t^X$.

We can say that $E\left(Y | \mathcal{F}_t^X\right)$ is the best predictor of Y which is a function of the values of X up to time t. Another common terminology is the following:

$$E\left(Y | \mathcal{F}_t^X\right) = E\left(Y \left| X_u, u \leq t\right.\right),$$

which is read as the conditional expectation of Y given the values of the process X up to time t.

Example 2.1. Suppose that $Y \in \mathcal{F}_t^X$. What is $E\left(Y | \mathcal{F}_t^X\right)$?

Obviously, it is hard to solve the minimization problem of (2). It turns out that the following is an equivalent condition:

(2’) $Y^* = E\left(Y | \mathcal{F}_t^X\right)$ is such that

$$E\left(Y^* \mathbf{1}_A\right) = E\left(Y \mathbf{1}_A\right),$$

for any A of the form $A := \{X_{t_0} \in C_0, \ldots, X_{t_n} \in C_n\}$ with $t_0, \ldots, t_n \leq t$ and arbitrary sets $C_0, \ldots, C_n$.

The following properties are not hard to see from (2’):

(i) $E\left(E\left(Y|\mathcal{F}_t^X\right)\right) = EY$;

(ii) Suppose that $X_0 = 0$. The, $E\left(Y|\mathcal{F}_0^X\right) = E\left(Y\right)$;

(iii) In general, if $X_0 = Z$ then $E\left(Y|\mathcal{F}_0^X\right) = E\left(Y|Z\right)$.

(iv) $E(YU\left|\mathcal{F}_t^X\right.) = UE(Y\left|\mathcal{F}_t^X\right.)$, for any $U \in \mathcal{F}_t^X$.

(v) *Linearity:* $E(aY_1 + bY_2\left|\mathcal{F}_t^X\right.) = aE(Y_1\left|\mathcal{F}_t^X\right.) + bE(Y_2\left|\mathcal{F}_t^X\right.)$.

(vi) *Positivity:* $E(Y\left|\mathcal{F}_t^X\right.) \geq 0$, *if* $Y \geq 0$.

(vii) *Independence properties: If Y is independent of $\{X_u\}_{u\leq t}$ (which means that Y is independent of $X_{t_0},\dots,X_{t_n}$ for any $t_0,\dots,t_n \leq t$), then*

$$E(Y\left|\mathcal{F}_t^X\right.) = E(Y).$$

(viii) *<u>Tower Property</u>: For any $0 \leq s \leq t$,*

$$E(E(Y\left|\mathcal{F}_t^X\right.)\left|\mathcal{F}_s^X\right.) = E(Y\left|\mathcal{F}_s^X\right.).$$

Example 2.2. In order to get more familiar with the operation $E(\cdot|\mathcal{F}_t^X)$, let us check (viii) above. Since $E(Y\left|\mathcal{F}_s^X\right.)$ is $\mathcal{F}_s^X$-measurable, all what we need to show is that

$$E\left[E(Y\left|\mathcal{F}_t^X\right.)\mathbf{1}_{\{X_{t_0}\in C_0,\dots,X_{t_n}\in C_n\}}\right] = E\left[E(Y\left|\mathcal{F}_s^X\right.)\mathbf{1}_{\{X_{t_0}\in C_0,\dots,X_{t_n}\in C_n\}}\right],$$

for all $t_0,\dots,t_n \leq s$. Justify the following steps:

$$\begin{aligned}
E\left[E(Y\left|\mathcal{F}_t^X\right.)\mathbf{1}_{\{X_{t_0}\in C_0,\dots,X_{t_n}\in C_n\}}\right] &= E\left[E(Y\mathbf{1}_{\{X_{t_0}\in C_0,\dots,X_{t_n}\in C_n\}}\left|\mathcal{F}_t^X\right.)\right] \\
&= E(Y\mathbf{1}_{\{X_{t_0}\in C_0,\dots,X_{t_n}\in C_n\}}) \\
&= E(E(Y\left|\mathcal{F}_s^X\right.)\mathbf{1}_{\{X_{t_0}\in C_0,\dots,X_{t_n}\in C_n\}}).
\end{aligned}$$

Recall that $\mathcal{F}^Z$ stands for all events A which occurrence can be determined from the value of Z. These events are of the form $A = \{\omega : Z(\omega) \in C\}$ and we can say that $\mathcal{F}^Z$ is the information generated by Z. The class $\mathcal{F}_t^X \cup \mathcal{F}^Z$ denotes all the events A which belong to $\mathcal{F}_t^X$ or $\mathcal{F}^Z$ and symbolizes all the information generated by X up to time t or by Z. One can define

$$E\left(Y \mid \mathcal{F}_t^X \cup \mathcal{F}^Z\right) = E\left(Y \mid \mathcal{F}_t^X, Z\right),$$

as a random variable Y^* whose value depends only on the value of X up to time t or the value of Z and such that

$$(2') \quad E\left(Y^*\mathbf{1}_A\right) = E\left(Y\mathbf{1}_A\right),$$

for any A of the form $A := \{X_{t_0} \in C_0,\dots,X_{t_n} \in C_n, Z \in D\}$ with $t_0,\dots,t_n \leq t$ and arbitrary sets $C_0,\dots,C_n,D$. The following *tower property* can be deduced as above:

$$E(E(Y\left|\mathcal{F}_t^X \cup \mathcal{F}^Z\right.)\left|\mathcal{F}_t^X\right.) = E(Y\left|\mathcal{F}_t^X\right.).$$

Exercise 4.

(1) Suppose that Y and Z are independent of $\{X_t\}_{t\geq 0}$. Show that

$$E(Y\left|\mathcal{F}_t^X \cup \mathcal{F}^Z\right.) = E(Y\left|Z\right.).$$

(2) Given two processes $\{X_t\}_{t\geq 0}$ and $\{Z_t\}_{t\geq 0}$, what is the natural definition of $E(Y \,|\mathcal{F}_t^X \cup \mathcal{F}_u^Z)$?

3. Self-financing trading strategies: motivation towards stochastic integration*

The adoption of a continuous-time stochastic model $\{S_t\}_{t\leq T}$ for the stock, such as the Black-Scholes model (5), necessitates to redefine the fundamental concept of self-financing trading strategy. Let us recall that the fundamental problems of finance, namely arbitrage-freeness, hedging and pricing, are defined in terms of self-financing strategies.

Roughly speaking, we can think of a self-financing trading strategy in continuous-time as the limit of discrete-time self-financing trading strategies when the time step Δt between trades gets smaller. So, in principle, the final wealth of the trading strategy can be replicated as close as we want by taking Δt very small. This concept will be made precise in this part.

We consider a market with two assets: A stock with price process $\{S_t\}_{t\geq 0}$ and a money market account for borrowing and lending money. The *money market account* is determined by its *instantaneous interest rate* process $\{r(t)\}_{t\geq 0}$. The later concept means that the if we borrow (resp. lend) x_t dollars at time t, we shall pay (resp. received) the interest $x_t r(t) dt$ at time $t+dt$. Notice that the instantaneous interest rate at a time $t>0$ could be unknown today at time 0. In light of this risk-free investment, one dollar today is worth

$$B_t = e^{\int_0^t r(u)du}$$

at time t. Indeed, having one dollar today, one can invest it in the money market account and rolling over the interest every time step $\Delta t := t/n$. By time t, the final wealth will be

$$\widetilde{B}_t^n := \prod_{k=0}^{n} \left(1 + r\left(\frac{k}{n}\right)\cdot \frac{t}{n}\right).$$

Taking logarithm and using that $\ln(1+x) \sim x$ as $x \to 0$, we get that

$$\lim_{n\to 0} \widetilde{B}_t^n = B_t.$$

That is, having one dollar today, we can generate a risk-free wealth as close to B_t as we wish by taking a small time step Δ between trades. Hence, $1 dollar today is worth B_t at time t or the other way around, the present value of $1 dollar at time t is $B_t^{-1} := e^{-\int_0^t r(u)du}$.

We can represent a *continuous-time trading strategy or portfolio* (also called *dynamical portfolio*) by a pair of processes $h := \{h_t := (x_t, y_t) : t \leq T\}$, where y_t shall determine the number of shares of the stock held at time t and x_t shall determine the money invested in the money market account at time t. Then, the value of the portfolio at time t is

$$V_t^h = x_t + y_t S_t. \tag{9}$$

The first natural restriction of a trading strategy is that it shouldn't be "anticipative". In other words, the positions x_t and y_t should be determined based only upon the information available at time t. Suppose that the information is generated by a process $\{Z_t\}_{t\geq 0}$. Typically, $\{Z_t\}_{t\geq 0}$ is the stock price

$\{S_t\}_{t\geq 0}$ itself but the minimal requirement is that $\{S_t\}_{t\geq 0}$ is adapted to the filtration generated by Z (why?). Then, we assume that

(A) $\{x_t\}_{t\leq T}$ and $\{y_t\}_{t\leq T}$ are adapted to the filtration generated by $\{Z_t\}_{t\geq 0}$.

What does a self-financing trading strategy mean? Suppose we decide to rebalance the portfolio every $\Delta t := T/n$ during the time period $[0, T]$. Is the trading strategy $\{h_t = (x_t, y_t) : t \leq T\}$ self-financing? Concretely, we want to know if the strategy that rebalances the portfolio at times $t_0 = 0, t_1 = T/n, \ldots, t_n = T$ according to $h_{t_0}, \ldots, h_{t_n}$ is self-financing. In order for this to happen, the value $\widetilde{V}_{t_i}^{n,h} := x_{t_i} + y_{t_i} S_{t_i}$ of such a discrete-time strategy must satisfy that

$$\widetilde{V}_{t_i}^{n,h} = x_{t_{i-1}}(1 + r(t_{i-1})\Delta t) + y_{t_{i-1}} S_{t_i}$$

for each $i = 1, \ldots, n$. Equivalently,

$$\widetilde{V}_{t_i}^{n,h} = \widetilde{V}_0 + \sum_{j=1}^{i} x_{t_{j-1}} r(t_{j-1})\Delta t + \sum_{j=1}^{i} y_{t_{j-1}}(S_{t_j} - S_{t_{j-1}}),$$

for each $i = 1, \ldots, n$, or what is the same,

$$\widetilde{V}_{t}^{n,h} = \widetilde{V}_0 + \sum_{t_j \leq t} x_{t_{j-1}} r(t_{j-1})\Delta t + \sum_{t_j \leq t} y_{t_{j-1}}(S_{t_j} - S_{t_{j-1}}), \tag{10}$$

for all $t \leq T$. Since the concept of trading strategy is understood as the limit of discrete-time strategies, we should understood (10) in the limit as the time step Δt between trades gets smaller and smaller.

The above considerations suggest to define a *self-financing trading strategy or portfolio* as a process $\{h_t = (x_t, y_t) : t \leq T\}$ adapted to the filtration generate by $Z := \{Z_t\}_{t\leq T}$ such that the

(B) The limits

$$(i) \quad \int_0^t x_u r(u) du := \lim \sum_{t_j \leq s} x_{t_{j-1}} r(t_{j-1})(t_j - t_{j-1}),$$

$$(ii) \quad \int_0^t y_u dS_u := \lim \sum_{t_j \leq t} y_{t_{j-1}}(S_{t_j} - S_{t_{j-1}})$$

exist for any $t \in [0, T]$, whenever the mesh $\bar{\pi} := \max_j (t_j - t_{j-1})$ of the discrete-times $\pi : 0 = t_0 < t_1 < \cdots < t_n = T$ converges to 0;

(C) The process $V_t^h := x_t + y_t S_t$ satisfies that

$$V_t^h = V_0^h + \int_0^t x_u r(u) du + \int_0^t y_u dS_u, \tag{11}$$

for all $t \leq T$.

In financial jargon, the limit process $\int_0^t x_u r(u) du$ is called the cumulative net *gain in the money-market account*, while $\int_0^t y_u dS_u$ is called the cumulative net *gain in the stock*. Equation (11) says that the value of the portfolio at any time $t \leq T$ equals the sum of the cumulative net gains in the money market account and the stock. This is precisely consistent with the idea of a self-financing portfolio as one where there is no infusion or withdrawn of money.

Mathematically, the limits in (B) above are integrals. The integral (i) is the well-known *Riemann* integral of basic calculus, while the integral in (ii) is the so-called *Riemann-Stieltjes* integral of y_t with respect to S. It turns out that in the case of processes of the form

$$S_t = S_0 e^{\sigma W_t + \mu t},$$

as in the Black-Scholes model, or simply when $S_t = W_t$, the limit

$$\lim_{\bar{\pi}\to 0} \sum_{t_j \le t} y_{t_{j-1}}(\omega)(S_{t_j}(\omega) - S_{t_{j-1}}(\omega)) \tag{12}$$

does not exist for all $\omega \in \Omega$ (not even, with probability 1) in general. This fact was well-known almost from the beginning that the mathematical foundations of Wiener process were established. It was Itô the first one who gave a precise meaning to the limit (12) and created the stochastic calculus for this type of integrals that now are commonly called *Itô integrals*. We shall develop the basis of this theory in the rest of this Chapter.

4. Stochastic integration

Our goal is to give a precise meaning to the integral of a process $\{f_t\}_{t\ge 0}$ with respect to the Wiener process $\{W_t\}_{t\ge 0}$. This operator is denoted as follows:

$$I_{[a,b]}(f) := \int_a^b f_t \, dW_t. \tag{13}$$

As it was mentioned before, the natural definition of (13) is as the limit of the Riemann-Stieljes sums

$$I^n_{[a,b]}(f) := \sum_{j=0}^{n-1} f_{t_j}(\omega)(W_{t_{j+1}}(\omega) - W_{t_j}(\omega)), \tag{14}$$

when the mesh $\bar{\pi} := \max\{t_j - t_{j-1}\}$ of the partition $\pi : t_0 = a < t_1 < \cdots < t_n = b$ vanishes. However, such a limit does not exist in general due to the extremely erratic path behavior of W. We shall see that if f is “nice", (14) will converge to (13) in the mean-square sense:

$$\lim_{n\to\infty} \mathbb{E}\left(I_{[a,b]}(f) - I^n_{[a,b]}(f)\right)^2 = 0. \tag{15}$$

The previous mode of convergence implies in particular that for any $\varepsilon > 0$, there exists a $\delta > 0$ such that if the mesh $\bar{\pi} < \delta$, then

$$P\left(\left|I_{[a,b]}(f) - I^n_{[a,b]}(f)\right| < \varepsilon\right) \ge 1 - \varepsilon. \tag{16}$$

Coming back to the financial interpretation, we shall interpret $I_{[a,b]}(f)$ as the gain of a trading strategy in continuous-time, and hence, (16) tells us that it is possible to replicate the gain within ε with very high probability if the time step between trades is smaller that certain threshold δ.

4.1. **Construction and basic properties.**

The idea towards (13) is to define the integral for simple processes in a natural manner and then, consider processes that can be approximated by simple processes. Consider a process $\{f_t\}_{t\in[a,b]}$ as follows:

$$f_t(\omega) := \sum_{j=0}^{n-1} \xi_j(\omega)\, \mathbf{1}_{[t_j,t_{j+1})}(t), \tag{17}$$

for a partition $\pi : a = t_0 < \cdots < t_n = b$ and $\xi_j \in \mathcal{F}^W_{t_j}$. Notice that $f_{t_j} = \xi_j$ and thus, the later condition means that f_t is not anticipative (namely, its values at any given time depends only on the information available at that time). For this type of processes, called *simple processes*, the limit of the Riemann-Stieljes sums (14) actually exists and is given by

$$\int_a^b f_t\, dW_t := \sum_{j=0}^{n-1} \xi_j(\omega) \left(W_{t_j} - W_{t_{j-1}}\right). \tag{18}$$

Next, suppose that $\{f_t\}_{t\in[a,b]}$ is a process such that there exists a sequence f^n of simple processes satisfying that

$$\lim_{n\to\infty} E \int_a^b (f_t - f_t^n)^2\, dt = 0. \tag{19}$$

Then, it can be proved that there exists a random variable $\int_a^b f_t\, dW_t$ such that

$$\lim_{n\to\infty} E \left(\int_a^b f_t\, dW_t - \int_a^b f_t^n\, dW_t\right)^2 = 0. \tag{20}$$

The random variable $\int_a^b f_t\, dW_t$ is called the *Itô-integral* of f with respect to W on $[a,b]$. It was Itô the first one to identify that the following class of functions f admits the approximation (19) by simple functions $\{f^n\}_{n\geq 1}$:

Definition 4.1. The process f belongs to the class $\mathcal{L}^2[a,b]$ if f is adapted to the information $\{\mathcal{F}^W_t\}_{t\geq 0}$ generated by W and

$$E \int_a^b f_t^2 dt < \infty.$$

If $f \in \mathcal{L}^2[0,t]$ for all $t \geq 0$, then we just write that $f \in \mathcal{L}^2$ and say that f is *Itô square-integrable*.

We summarize the previous claims in the following theorem:

Theorem 4.1. *If $f \in \mathcal{L}^2[a,b]$, then there exist a random variable $\int_a^b f_t\, dW_t$ and a sequence $\{f^n\}_{n\geq 1}$ of simple processes such that (19) and (20) hold true. Moreover, $\int_a^b f_t dW_t$ is identical for any sequence f_n satisfying (19).*

The main properties of the stochastic integral are given in the following result:

Proposition 4.1. *Let $f, g \in \mathcal{L}^2[a,b]$ and $c > 0$ be a real number. Then,*

(1) ***(Linearity)*** $\int_a^b (f_t + cg_t)dW_t = \int_a^b f_t dW_t + c\int_a^b g_t dW_t$.

(2) ***(Mean and variance)*** $E\left[\int_a^b f_t dW_t\right] = 0$ *and* $E\left[\left(\int_a^b f_t dW_t\right)^2\right] = E\int_a^b f_t^2 dt$.

(3) ***(Covariance)*** $E\left[\int_a^b f_t dW_t \int_a^b g_t dW_t\right] = E\int_a^b f_t g_t dt$.

We finish with an important remark. It turns out that if f is left-continuous and adapted then one can take $f_t^n(\omega) := f_a(\omega)\mathbf{1}_{\{a\}}(t) + \sum_{j=0}^{n-1} f_{t_j}(\omega)\,\mathbf{1}_{(t_j,t_{j+1}]}(t)$, and hence, the Riemman-Stieljes sums $I^n_{[a,b]}(f)$ in (14) converges to $I_{[a,b]}(f) := \int_a^b f_t\, dW_t$ in the mean-square sense (15).

The mean and variance formulas of Proposition 4.1 are actually valid even if we have at hand certain information of the past evolution of W. Concretely, we have that

$$E\left[\int_s^t f_u dW_u \middle| W_{t_0}, \dots, W_{t_n}\right] = 0, \quad \text{and} \quad E\left[\left(\int_s^t f_u dW_u\right)^2 \middle| W_{t_0}, \dots, W_{t_n}\right] = E\int_s^t f_u^2 du$$

for any $t_0, \dots, t_n \leq s$. These facts suggest that the conditioning expectation of the integral given "all" the past information generated by W is still 0. We now introduce these properties in the following section.

4.2. **Martingales properties of the stochastic integral.**

The treatment of the multi-period Binomial model showed that the concept of martingale appears naturally in the problems of mathematical finance. For instance, a risk-neutral measure Q can be defined as a probability measure such that the discounted stock-price process

$$S_t^* = \frac{S_t}{(1+R)^t}$$

is a martingale; namely, such that

$$E^Q\left(S_u^* \,|S_0^*, \dots, S_t^*\right) = S_t^*,$$

for any $0 \leq t \leq u \leq T$. The natural continuous-time version of the above formula is

$$E^Q\left(S_u^* \,|S_v^*, v \leq t\right) = S_t^*,$$

for any $0 \leq t \leq u$, where the above expectation is the conditional expectation of S_u^* given the information in $\mathcal{F}_t^{S^*}$. One can of course take a general informations process and define a martingale as follows.

Definition 4.2. Let $Y = \{Y_t\}_{t\geq 0}$ and $X = \{X_t\}_{t\geq 0}$ be a stochastic processes defined on a common sample space Ω, equipped with a probability measure P. The process Y is said to be a martingale relative to the filtration $\{\mathcal{F}_t^X\}_{t\geq 0}$ (or relative to the information generated by X) if

(1) Y_t is $\mathcal{F}_t^X$-measurable, for any t;

(2) $E|Y_t| < \infty$, for any t;

(3) $E\left(Y_t \,\middle|\mathcal{F}_s^X\right) = Y_s$, for all $0 \leq s \leq t$.

Example 4.1. Let Y_∞ be a random variable such that $E|Y_\infty| < \infty$. The process $Y_t := E\left(Y_\infty | \mathcal{F}_t^X\right)$ is a martingale relative to $\mathcal{F}_t^X$.

The following are fundamental martingales related to stochastic integrals.

Theorem 4.2. *Let $f, g \in \mathcal{L}^2$ be Itô-square integrable processes and $M_t = \int_0^t f_s dW_s$ and $N_t := \int_0^t g_s dW_s$. Then,*

(1) *M_t is a martingale relative to $\mathcal{F}_t^W$;*
(2) *$M_t^2 - \int_0^t f_s^2 ds$ is a martingale relative to $\mathcal{F}_t^W$;*
(3) *$M_t N_t - \int_0^t f_s g_s ds$ is a martingale relative to $\mathcal{F}_t^W$.*

Notice that the statement (1) in Theorem 4.2 is equivalent to any of the following two equivalent statements:

(1a) $E\left[\int_0^u f_s dW_s \,\middle|\, \mathcal{F}_t^W\right] = \int_0^t f_s dW_s$, for any $0 \le t \le u$;
(1b) $E\left[\int_t^u f_s dW_t \,\middle|\, \mathcal{F}_u^W\right] = 0$, for any $0 \le t \le u$.

Similarly, (2) in Theorem 4.2 is equivalent to any of the following two equivalent statements:

(2a) $E\left[\left(\int_0^u f_s dW_s\right)^2 - \int_0^u f_s^2 ds \,\middle|\, \mathcal{F}_t^W\right] = \left(\int_0^t f_s dW_s\right)^2 - \int_0^t f_s^2 ds$, for any $0 \le t \le u$;
(2b) $E\left[\left(\int_t^u f_s dW_s\right)^2 \,\middle|\, \mathcal{F}_t^W\right] = E\left[\int_t^u f_s^2 ds \,\middle|\, \mathcal{F}_t^W\right]$, for any $0 \le t \le u$.

Exercise 5. Verify that (2a) and (2b) are equivalent.

The processes $\{\int_0^t f_s^2 ds\}_{t \ge 0}$ is called the *quadratic variation* of M and is denoted by $\langle M \rangle$. It has the property that the discrete-time quadratic variation process, defined by

$$\langle M \rangle_t^\pi := \sum_{t_k \le t} \left(M_{t_k} - M_{t_{k-1}}\right)^2,$$

converges to $\langle M \rangle_t$, in the mean-square sense, when the mesh of the partition $\pi : t_0 = 0 < t_1 < \cdots < t_n$ becomes smaller:

$$\lim_{\mathrm{mesh}(\pi) \to 0} E\left(\langle M \rangle_t^\pi - \langle M \rangle_t\right)^2 = 0. \tag{21}$$

Similarly, the process $\{\int_0^t f_s g_s ds\}_{t \ge 0}$ is called the quadratic covariation of M and N and is denoted by $\langle M, N \rangle$.

4.3. **Itô processes.**

Our goal in the following section is to show a chain rule for stochastic integrals. A consequence will be that a smooth functions of an Itô integral can be expressed as the superposition of an Itô integrals and a Riemann integral. Such processes are called Itô processes.

Definition 4.3. We say that $\{X_t\}_{t\geq 0}$ is an *Itô process* if for all $t \geq 0$,

$$X_t := X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s, \tag{22}$$

for a constant X_0 and processes μ and σ adapted to $\{\mathcal{F}_t^W\}_{t\geq 0}$. The term $\int_0^t \sigma_s dW_s$ is called the Wiener or Brownian component of the Itô process.

The process (30) is well-defined if

$$E\left\{\int_0^t |\mu_s| ds + \int_0^t \sigma_s^2 ds\right\} < \infty, \quad \text{for all} \quad t \geq 0,$$

in which case one can construct $\{X_t\}_{t\geq 0}$ to have *continuous* paths with probability 1.

We often write the process X in the following “differential form" and say that X has the following dynamics

$$dX_t = \mu_t dt + \sigma_t dW_t, \quad X_0 = a. \tag{23}$$

It is important to be conscious that (23) has no meaning itself except as a “convenient" way to write (30). In particular, it is wrong to interpret (23) as the differential form of the following differential equation:

$$\frac{dX_t}{dt} = \mu_t + \sigma_t \frac{dW_t}{dt},$$

where dX_t/dt and dW_t/dt are derivatives. Such an interpretation is wrong (and certainly, not true) since for instance W is known to be no-where differentiable.

Nevertheless, the terminology (23) is useful to make the following definition plausible and easy to remember. We define the integral of a process f with respect to the Itô processes (30) by

$$\int_0^t f_s dX_s := \int_0^t f_s \mu_s ds + \int_0^t f_s \sigma_s dW_s, \tag{24}$$

provided that the Riemann integral and the Itô integrals on the right-hand side are well-defined (for instance if f is bounded).

The above definition of integral with respect to an Itô process is consistent with the Riemann-Stieljes integral (defining the limits in a proper manner). The following result clarifies this point:

Theorem 4.3. *Suppose that $\{f_t\}_{t\geq 0}$ is adapted to W, left-continuous, and bounded on each finite-time interval. Then, $\{\int_0^t f_s dX_s\}_{t\geq 0}$ in (24) is well-defined and satisfies that, for any $\varepsilon > 0$ and $T > 0$,*

$$P\left(\sup_{t\leq T}\left|\int_0^t f_s dX_s - \sum_{t_i \leq t} f_{t_i}(X_{t_{i+1}} - X_{t_i})\right| < \varepsilon\right) \geq 1 - \varepsilon,$$

whenever the mesh of the partition $\pi : t_0 = 0 < \cdots < t_n = T$ is smaller than certain $\delta > 0$ (which depends on the precision $\varepsilon > 0$ and the time horizon T).

We conclude that the Riemann-Stieljes sum $\sum_{t_i \leq T} f_{t_i}(X_{t_{i+1}} - X_{t_i})$ can be made arbitrarily close to $\int_0^T f_s dX_s$ with arbitrarily high probability if the mesh is high-enough. When $\{X_t\}_{t\leq T}$ represents the price process of an asset and f_t represents the position in that asset at time t, $\int_0^t f_s ds$ is interpreted as the net gain in the stockby time t. For instance, suppose that

$$f_t = y\mathbf{1}_{[u,v]}(t),$$

where $u < v$ are constants and y depends on the information up to time u. Then, this trading strategy buys y units of the asset at time u, hold them until time v, and then sell them. Clearly, we expect that the net gain at time t will be

$$\int_0^t f_t d\, X_t = \begin{cases} 0 & t < u \\ y(X_t - X_u) & u \leq t < v \\ y(X_v - X_u) & t \geq v \end{cases} .$$

Indeed, if for instance $t \geq v$, the Riemann-Stieljes sum takes the form

$$\sum_{t_i \leq t} f_{t_i}(X_{t_{i+1}} - X_{t_i}) = y\left(X_{v(\pi)} - X_{u(\pi)}\right),$$

where $v(\pi)$ is the largest t_i smaller than v and $u(\pi)$ is the smallest t_i larger than u. Since the paths of X are continuous with probability 1, the Riemann-Stieljes sum converge to our guess of gain process. Such a trading strategy is called a *buy-and-hold strategy*. The sum $\sum_{t_i \leq T} f_{t_i}(X_{t_{i+1}} - X_{t_i})$ can be interpreted as the net gain of a combination of buy-and-hold trading strategies. Concretely, this strategy changes the position in the asset at times $t_0 < \cdots < t_n$ according to $f_{t_0}, \ldots, f_{t_n}$.

4.4. **The Itô formula.**

This section is one of the most important topics of this chapter, at least for applications in finance. The following theorem introduces the fundamental Itô formula.

Theorem 4.4. *Let $X = \{X_t\}_{t\geq 0}$ be an Itô process of the generic form (30). Let f be a function that is twice continuously differentiable. Then, $f(X_t)$ satisfies the equation:*

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)dX_s + \frac{1}{2}\int_0^t f''(X_s)\sigma_s^2 ds, \tag{25}$$

for all $t \geq 0$. In particular, $f(X_t)$ is also an Itô process with the representation:

$$f(X_t) = f(X_0) + \int_0^t \left\{ f'(X_s)\mu_s + \frac{1}{2}f''(X_s)\sigma_s^2 \right\} ds + \int_0^t f'(X_s)\sigma_s dW_s.$$

Remark 4.1. The differential form of (25) looks like

$$df(X_t) = f'(X_t)dX_t + \frac{1}{2}f''(X_t)\sigma_t^2 dt,$$

which is why Itô formula is sometimes called a chain rule for Itô processes. It is also common to use the terminology $(dX_t)^2 = \sigma_t^2 dt$. In that case, we can write (25) in the following form which evokes a Taylor's expansion of order 2:

$$df(X_t) = f'(X_t)dX_t + \frac{1}{2}f''(X_t)(dX_t)^2.$$

One can think of $(dX_t)^2 = \sigma_t^2 dt$ as the square of the differential $dX_t = \mu_t dt + \sigma_t dW_t$ with the convention that $(dt)^2 = 0$ and $dtdW_t = 0$.

Example 4.2. Itô's formula shows that the calculus for Itô integrals differs from the standard calculus. For instance, it is not true that $\int_0^t f'(W_s)dW_s = f(W_t) - f(W_0)$. Rather, it is necessary a “correction" term as follows:

$$\int_0^t f'(W_s)dW_s = f(W_t) - f(W_0) - \frac{1}{2}\int_0^t f''(W_s)ds.$$

Hence, the following two cases follows:

$$\int_0^t W_s dW_s = \frac{1}{2}W_t^2 - t, \qquad \int_0^t e^{W_s}dW_s = e^{W_t} - 1 - \frac{1}{2}\int_0^t e^{W_s}ds.$$

Example 4.3. Take $f(x) = x^2$ and $X_t = \int_0^t \sigma_s dW_s$. Then, (25) becomes

$$X_t^2 = 2\int_0^t X_s dX_s + \int_0^t d\left\langle X\right\rangle_s.$$

Replacing X and $\left\langle X\right\rangle_t = \int_0^t \sigma_s^2 ds$,

$$X_t^2 = 2\int_0^t X_s\sigma_s dW_s + \int_0^t \sigma_s^2 ds.$$

Since $\int_0^t X_s\sigma_s dW_s$ is a martingale, we conclude that $\int_0^t \sigma_s dW_s - \int_0^t \sigma_s^2 ds$ is a martingale, which is consistent with Theorem 4.2.

Example 4.4. Suppose that σ is a non-random function. The aim of this example is to show that $X_t = \int_0^t \sigma_s dW_s$ is a Normal random variable. In order to do so, we find the moment generating function

$$m(t,u) = E\left(e^{uX_t}\right).$$

Using Itô's formula with $f(x) = e^{ux}$,

$$e^{uX_t} = 1 + u\int_0^t e^{uX_s}\sigma_s dW_s + \frac{u^2}{2}\int_0^t e^{uX_s}\sigma_s^2 ds.$$

Taking expectations on both sides and using that the expectation of an Itô integral is 0 and that σ is deterministic, we get

$$m(t,u) = 1 + \frac{u^2}{2}\int_0^t \sigma_s^2 m(s,u)ds.$$

Writing this equation in differential form (treating u as a constant) and “separating variables",

$$dm(t,u) = \frac{u^2\sigma_t^2}{2}m(t,u)dt \Longrightarrow \frac{dm(t,u)}{m(t,u)} = \frac{u^2\sigma_t^2}{2}dt \Longrightarrow \ln m(t,u) - \ln m(0,u) = \frac{u^2}{2}\int_0^t \sigma_s^2 ds.$$

We finally obtain that

$$m(t,u) = e^{\frac{u^2}{2}\int_0^t \sigma_s^2 ds}.$$

This is precisely the moment generating functions of a Gaussian random variable with mean 0 and variance $\int_0^t \sigma_s^2 ds$.

Exercise 6. The goal of this problem is to show an Integration by parts formula for Itô processes. Suppose that $X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s$ and $Y_t = Y_0 + \int_0^t \tilde{\mu}_s ds + \int_0^t \tilde{\sigma}_s dW_s$ are two Itô processes.

(a) Find the following differentials:

$$d(X_t + Y_t)^2, \quad dX_t^2, \quad dY_t^2.$$

(b) Using parts (a) and that $xy = \{(x+y)^2 - x^2 - y^2\}/2$, show that

$$X_t Y_t = X_0 Y_0 + \int_0^t X_s dY_s + \int_0^t Y_s dX_s + \int_0^t \sigma_s \tilde{\sigma}_s ds. \tag{26}$$

4.5. **The Black-Scholes model and its dynamics.**

As an application of Itô's formula and due to its great relevance in mathematical finance, let us study the dynamics of the Black-Scholes model

$$S_t = S_0 e^{\sigma W_t + \mu t}, \tag{27}$$

where $\sigma \geq 0$ and μ are constants, and W is a Wiener process. Taking $f(x) = S_0 e^x$ and $X_t = \sigma W_t + \mu t$, we obtain that

$$\begin{aligned} S_t &= S_0 + \int_0^t S_u dX_u + \frac{\sigma^2}{2}\int_0^t S_u du \\ &= S_0 + \int_0^t S_u \left(\mu + \frac{\sigma^2}{2}\right) du + \frac{\sigma^2}{2}\int_0^t S_u dW_u. \end{aligned}$$

Hence, the dynamics of the Black-Scholes model is given by

$$dS_t = S_t \left\{ \left(\mu + \frac{\sigma^2}{2}\right) dt + \sigma dW_t \right\}. \tag{28}$$

An important consequence of the previous equation is that we can now formally define the stochastic integrals with respect to S as follows:

$$G(t) := \int_0^t h_u dS_u := \int_0^t h_u S_u \left(\mu + \frac{\sigma^2}{2}\right) du + \int_0^t h_u \sigma dW_u.$$

Recall that, being S an Itô process, this integral will be the limit of Riemann-Stieltjes sums in the sense of Theorem 4.3. Again, in the case that S and h are interpreted as the price process of a stock and a trading strategy, respectively, $G(t)$ is interpreted as the gain process in the stock by time t. This gain process can be realized, within any desired precision ε with very high probability, by running a discrete-time trading strategy at times $t_0 = 0 < \cdots < t_n = t$ according to the positions $h_{t_0}, \ldots, h_{t_n}$ with high enough trading frequency.

It turns out that the equation (28) characterizes the geometric Brownian motion in the sense that the only process that satisfies (28) is (27) as the following exercise proves:

Exercise 7. Suppose that, besides (27), there exists another process $\tilde{S}$ such that

$$d\tilde{S}_t = \tilde{S}_t \left\{ \left(\mu + \frac{\sigma^2}{2} \right) dt + \sigma dW_t \right\}, \tag{29}$$

and $\tilde{S}_0 = S_0$. Apply the integration by parts formula (26) and Itô's formula to show that

$$d(S_t^{-1} S_t) = 0.$$

Conclude that $S_t = \tilde{S}_t$, for all t.

4.6. **Multidimensional Itô formula.**

It is useful to generalize Itô formula to deal with more than one Itô process and with more sources of randomness. Suppose that $W^1, \ldots, W^d$ are d independent Wiener processes. We then say that $W := (W^1, \ldots, W^d)$ is a d*−dimensional Wiener process*. An Itô process is then defined a process of the form

$$X_t = X_0 + \int_0^t \mu_s ds + \sum_{j=1}^d \int_0^t \sigma_s^j dW_s^j,$$

where X_0 is a constant and μ and $(\sigma^1, \ldots, \sigma^d)$ are processes adapted to the information generated by the processes $(W^1, \ldots, W^d)$. It turns out that a smooth function $f(X^1, \ldots, X^m)$ of Itô processes is an Itô process.

Theorem 4.5. *Let $f(t, x_1, \ldots, x_m)$ be a continuous function such that $\frac{\partial f}{\partial t}$, $\frac{\partial f}{\partial x_i}$, and $\frac{\partial^2 f}{\partial x_i \partial x_k}$ are continuous. Let*

$$X_t^i = X_0^i + \int_0^t \mu_s^i ds + \sum_{j=1}^d \int_0^t \sigma_s^{i,j} dW_s^j, \quad i = 1, \ldots, m,$$

be Itô processes. Then, $f(t, X_t^1, \ldots, X_t^m)$ is an Itô process which differential form is given by

$$\begin{aligned} df(t, X_t^1, \ldots, X_t^m) = {} & \frac{\partial f}{\partial t}(t, X_t^1, \ldots, X_t^m)\, dt + \sum_{i=1}^m \frac{\partial f}{\partial x_i}(t, X_t^1, \ldots, X_t^m)\, dX_t^i \\ & + \frac{1}{2} \sum_{i,k=1}^m \frac{\partial^2 f}{\partial x_i \partial x_k}(t, X_t^1, \ldots, X_t^m)\, dX_t^i \cdot dX_t^k, \end{aligned}$$

where $dX_t^i \cdot dX_t^k$ can be found as a standard multiplication operation with the following formal multiplication rules:

$$(dt)^2 = dt \cdot dt = 0, \quad dt \cdot dW_t^j = 0, \quad dW_t^k \cdot dW_t^j = 0, \quad \text{if } k \neq j, \quad dW_t^j \cdot dW_t^j = (dW_t^j)^2 = dt.$$

Example 4.5. Let us apply Itô's formula to find the differential of the process $Z_t = tW_t$. Taking $f(t, x) = tx$,

$$d(tW_t) = tdW_t + W_t dt.$$

Hence, $Z_t = tW_t = \int_0^t W_s ds + \int_0^t s dW_s$.

Example 4.6. Let W^1 and W^2 be independent Wiener processes. Define $R_t = \sqrt{(W_t^1)^2 + (W_t^2)^2}$ be the distance from the origin to (W_1, W_2). Then, we can find the differential dR_t as follows:

$$dR_t = \frac{1}{R_t}\left(\frac{1}{2}dt + W_t^1 dW_t^1 + W_t^2 dW_t^2\right).$$

4.7. **Integration by parts and the quadratic covariation process.**

Let us consider two Itô processes

$$X_t = X_0 + \int_0^t \mu_s ds + \sum_{j=1}^d \int_0^t \sigma_s^j \, dW_s^j, \quad \text{and} \quad Y_t = Y_0 + \int_0^t \tilde{\mu}_s ds + \sum_{j=1}^d \int_0^t \tilde{\sigma}_s^j \, dW_s^j. \tag{30}$$

Then, by Itô formula,

$$d(X_t Y_t) = X_t dY_t + Y_t dX_t + dX_t \cdot dY_t. \tag{31}$$

Using the multiplication rules for differentials, it is easy to see that

$$dX_t \cdot dY_t = \left(\sum_{j=1}^d \sigma_t^j \tilde{\sigma}_t^j\right) dt.$$

Plugging in (31), we obtain the following *integration by parts formula*:

$$\int_0^t X_s dY_s = X_t Y_t - X_0 Y_0 - \int_0^t Y_s dX_s - \langle X, Y \rangle_t \,, \tag{32}$$

where

$$\langle X, Y \rangle_t = \sum_{j=1}^d \int_0^t \sigma_s^j \tilde{\sigma}_s^j ds.$$

The above process is called the *quadratic covariation* of the processes X and Y.

Example 4.7. Apply the integration by parts formula to justify the following formula:

$$\int_0^t \int_0^s \mu_u dW_u ds = \int_0^t (t-u)\mu_u dW_u.$$

How can you justify the previous formula in terms of a change of integration arguments (namely, a Fubini's formula)?

The quadratic covariation has very important properties. First, notice that

$$\langle X, Y \rangle_t = X_t Y_t - X_0 Y_0 - \int_0^t Y_s dX_s - \int_0^t X_s dY_s. \tag{33}$$

If we approximate the last two integrals on the right-hand side using Riemann-Stieltjes sums based on a partition $\pi : t_0 = 0 < t_1 < \cdots < t_n = t$, then it can be shown that

$$X_t Y_t - X_0 Y_0 - \sum_{i=0}^{n-1} Y_{t_i}(X_{t_{i+1}} - X_{t_i}) - \sum_{i=0}^{n-1} X_{t_i}(Y_{t_{i+1}} - Y_{t_i}) = \sum_{i=0}^{n}(X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i}).$$

Since the Riemann-Stieltjes sums converges to the corresponding integral in the sense of Theorem 4.3, we expect that for any $\varepsilon > 0$,

$$P\left(\left|\langle X,Y\rangle_t - \sum_{i=0}^{n}(X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i})\right| < \varepsilon\right) \geq 1-\varepsilon,$$

whenever the mesh of the partition π is smaller than certain $\delta > 0$ (which depends on $\varepsilon > 0$). Hence, the discrete-time quadratic covariation $\sum_{i=0}^{n}(X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i})$ can be made arbitrarily close to $\langle X,Y\rangle_t$ with high probability if the mesh of the partition is small enough.

Anothet outstanding property arises when X and Y are martingales. More specifically, suppose that $\mu = \tilde{\mu} = 0$ and the σ^j's and $\tilde{\sigma}^j$'s are Itô-square integrable processes. Then,

$$X_tY_t - \langle X,Y\rangle_t = \sum_{j=1}^{d}\int_0^t \left(Y_s\sigma_s^j + X_s\tilde{\sigma}_s^j\right) dW_s^j,$$

which is a martingale. In fact, this property characterizes $\{\langle X,Y\rangle_t\}_{t\geq 0}$ as the following result states:

Proposition 4.2. *Suppose that X and Y are as in (30) with $\mu = \tilde{\mu} = 0$ and Itô-quare integrable processes σ and $\tilde{\sigma}$. Then,*

(1) *$M_t := X_tY_t - \langle X,Y\rangle_t$ is martingale relative to $\{\mathcal{F}^W\}_{t\geq 0}$;*
(2) *If $A_t = \int_0^t b_s ds$ is such $X_tY_t - A_t$ is a martingale, then $A_t = \langle X,Y\rangle_t$, for all $t \geq 0$.*

The following property is quite outstanding. It is called the Lévy's martingale characterization of Wiener process.

Theorem 4.6. *Let $\{X_t\}_{t\geq 0}$ and $\{Y_t\}_{t\geq 0}$ be continuous processes adapted to the information $\{\mathcal{F}_t^W\}_{t\geq 0}$ generated by Wiener processes $W = (W^1,\ldots,W^d)$.*

(1) *If $\{X_t\}_{t\geq 0}$ is a martingale relative to $\{\mathcal{F}_t^W\}_{t\geq 0}$ such that $\langle X,X\rangle_t = t$, for all $t \geq 0$, then X is a Wiener process.*
(2) *If $\{X_t\}_{t\geq 0}$ and $\{Y_t\}_{t\geq 0}$ are martingales relative to $\{\mathcal{F}_t^W\}_{t\geq 0}$ such that $\langle X,X\rangle_t = \langle Y,Y\rangle_t = t$ and $\langle X,Y\rangle_t = 0$, for all $t \geq 0$, then X and Y are independent Wiener processes.*

Example 4.8. Consider the Itô process

$$dX_t = \mu_t dt + \sum_{j=1}^{d}\hat{\sigma}_t^j dW_t^j.$$

Applying the Lévy's Theorem 4.6, $B_t := \sum_{j=1}^{d}\int_0^t \frac{1}{\|\hat{\sigma}_s\|}\hat{\sigma}_s^j W_s^j$ is a Wiener process. Then, X admits an Itô representation with respect to only one Wiener process (rather than d Wiener processes):

$$dX_t = \mu_t dt + \sigma_s dB_t,$$

with $\sigma_t = \|\hat{\sigma}_t\|$. As an application, suppose that b and $\sigma = (\hat{\sigma}^1, \ldots, \hat{\sigma}^d)$ are constant processes. Consider the Itô process S

$$dS_t = S_t b dt + S_t \sum_{j=1}^{d} \hat{\sigma}^j dW_t^j.$$

Then, S follows that Black-Scholes model with volatility $\sigma = \|\hat{\sigma}\|$ and mean rate of return $\mu = b - \frac{1}{2}\sigma^2$ (see relation between (27) and (28)).

4.8. **Correlated Wiener processes.**

Let $W = (W^1, W^2)$ be a bivariate Wiener process and let $\{\rho_t\}_{t\geq 0}$ be a process adapted to the information generated by W taking values in $[-1, 1]$. Define the processes

$$B_t^1 = W_t^1, \quad B_t^2 = \int_0^t \rho_s dW_s^1 + \int_0^t \sqrt{1-\rho_s^2} dW_s^2. \tag{34}$$

In light of Lévy's Theorem 4.6, B^1 and B^2 are Wiener processes such that

$$\left\langle B^1, B^2 \right\rangle_t = \int_0^t \rho_s ds. \tag{35}$$

A pair of Wiener processes B^1 and B^2 satisfying (35) are said to be *correlated Wiener processes* with (instantaneous) correlation process $\{\rho_t\}_{t\geq 0}$. The transformations (34) is a devise to built correlated Wiener processes from independent Wiener processes W^1 and W^2.

Exercise 8. From Proposition 4.2, it follows that

$$E\left(B_{t+\Delta}^1 B_{t+\Delta}^2 - \int_0^{t+\Delta} \rho_s ds \,\middle|\, \mathcal{F}_t^W \right) = B_t^1 B_t^2 - \int_0^t \rho_s ds,$$

whenever $t, \Delta \geq 0$. Use this fact to show that

$$E\left((B_{t+\Delta}^1 - B_t^1)(B_{t+\Delta}^2 - B_t^2) \,\middle|\, \mathcal{F}_t^W \right) = E\left(\int_t^{t+\Delta} \rho_s ds \,\middle|\, \mathcal{F}_t^W \right).$$

The previous exercise show that under certain regularity in ρ,

$$\lim_{\Delta\to 0} \frac{1}{\Delta} E\left((B_{t+\Delta}^1 - B_t^1)(B_{t+\Delta}^2 - B_t^2) \,\middle|\, \mathcal{F}_t^W \right) = E\left(\lim_{\Delta\to 0} \frac{1}{\Delta} \int_t^{t+\Delta} \rho_s ds \,\middle|\, \mathcal{F}_t^W \right) = \rho_t. \tag{36}$$

The following exercise show that we can interpret ρ_t as the instantaneous condition correlation between the increments of B^1 and B^2 at time t given the past information up to time t:

Exercise 9. We define the instantaneous correlation of two random variables given $\mathcal{F}_t^W$ as follows:

$$\rho\left(X, Y \,\middle|\, \mathcal{F}_t^W\right) = \frac{\operatorname{Cov}\left(X, Y \,\middle|\, \mathcal{F}_t^W\right)}{\sqrt{\operatorname{Var}\left(X \,\middle|\, \mathcal{F}_t^W\right) \cdot \operatorname{Var}\left(Y \,\middle|\, \mathcal{F}_t^W\right)}},$$

where, as it should be expected,

$$\operatorname{Cov}\left(X, Y \,\middle|\, \mathcal{F}_t^W\right) = E\left[\left(X - E\left[X \,\middle|\, \mathcal{F}_t^W\right]\right)\left(Y - E\left[Y \,\middle|\, \mathcal{F}_t^W\right]\right) \,\middle|\, \mathcal{F}_t^W \right],$$
$$\operatorname{Var}\left(X \,\middle|\, \mathcal{F}_t^W\right) = E\left[\left(X - E\left[X \,\middle|\, \mathcal{F}_t^W\right]\right)^2 \,\middle|\, \mathcal{F}_t^W \right].$$

Using (36), show that

$$\lim_{\Delta\to 0}\rho\left(B^1_{t+\Delta}-B^1_t, B^2_{t+\Delta}-B^2_t\,\middle|\mathcal{F}^W_t\right)=\rho_t$$

In general, we have the following concept.

Definition 4.4. Let $\rho=[\rho^{i,j}]^d_{i,j=1}$ be a $d\times d$ matrix of adapted processes taking values in $[-1,1]$. We say that a d-dimensional process $B=(B^1,\dots,B^d)^T$ of Itô processes are correlated Wiener processes with correlation matrix ρ if each B^i is a Wiener process, and if for each i,j,

$$\left\langle B^i,B^j\right\rangle_t=\int_0^t\rho^{i,j}_s ds,$$

or equivalently, if

$$d\,B^i_t\cdot d\,B^j_t=\rho^{i,j}_t dt.$$

We remark that ρ has some constraints. Concretely, for each t, ρ_t is a symmetric nonnegative definite matrix such that $\rho^{i,i}=1$, for any $i=1,\dots,d$.

Example 4.9. Let $W=(W^1,\dots,W^d)$ be $d-$dimensional Wiener process and consider a $d\times d$ matrix $\Sigma=[\hat\sigma^{i,j}]^d_{i,j=1}$ of processes adapter to $\mathcal{F}^W$. Denote $\hat\sigma^{i\cdot}$ the i^{th}-row of Σ which is assumed to be nonzero for each i. Define the Itô processes

$$B^i=\sum_{j=1}^d\int_0^t\frac{1}{\|\hat\sigma^{i\cdot}_s\|}\hat\sigma^{i,j}_s dW^j_s.$$

Using the multiplication rules for differentials and Lévy's theorem 4.6, we can easily see that $B=(B^1,\dots,B^d)$ is a correlated Wiener process with correlations

$$\rho^{i,j}_t=\frac{\hat\sigma^{i\cdot}_t\left(\hat\sigma^{j\cdot}_t\right)^T}{\|\hat\sigma^{i\cdot}_t\|\|\hat\sigma^{j\cdot}_t\|}.$$

The following result provides a devise to construct correlated Wiener processes from independent Wiener processes (the analog to the construction in (34) above):

Proposition 4.3. *Let $W=(W^1,\dots,W^d)$ be $d-$dimensional Wiener process and let $\rho=[\rho^{i,j}]^d_{i,j=1}$ be a $d\times d$ symmetric (non-negative definite) matrix of adapted processes. Then, there exists a $d\times d$ lower triangular matrix $\bar\Sigma=[\bar\sigma^{i,j}]^d_{i,j=1}$ of adapted processes such that*

$$\rho=\bar\Sigma\cdot\bar\Sigma^T.$$

Furthermore, the processes

$$B^i=\sum_{j=1}^d\int_0^t\bar\sigma^{i,j}_s dW^j_s,\quad i=1,\dots,d,$$

are correlated Wiener processes with correlation matrix ρ.

4.9. **The generalized multidimensional Black-Scholes model.**

First, consider the following generalized Black-Scholes model for the price process of an asset:

$$S_t = S_0 e^{\int_0^t \mu_s ds + \int_0^t \sigma_s dW_s}. \tag{37}$$

Applying Itôo, it is clear tha S satisfies the dynamics

$$dS_t = S_t \left\{ \left(\mu_t + \frac{\sigma_t^2}{2} \right) dt + \sigma_t dW_t \right\}.$$

Let us briefly digress on the financial interpretation of σ and μ. Notice that the log returns of the asset are

$$\ln \frac{S_{t+\Delta}}{S_t} = \int_t^{t+\Delta} \sigma_s dW_s + \int_t^{t+\Delta} \mu_s ds.$$

Then,

$$\mu_t = \lim_{\Delta \to 0} \frac{1}{\Delta} E \left\{ \ln \frac{S_{t+\Delta}}{S_t} \middle| \mathcal{F}_t^W \right\}, \quad \sigma_t^2 = \lim_{\Delta \to 0} \frac{1}{\Delta} \mathrm{Var} \left\{ \ln \frac{S_{t+\Delta}}{S_t} \middle| \mathcal{F}_t^W \right\}.$$

Hence, we can say that μ_t is the instantaneous mean rate of return at time t conditional on the past information, while σ_t is the instantaneous volatility at time t conditional on the past information.

We can generalize (37) to cope with mutiple assets. Suppose that there are N assets in the market with price processes given by the dynamics:

$$S_t^i = S_0^i \exp \left\{ \int_0^t \mu_s^i ds + \sum_{j=1}^d \int_0^t \hat{\sigma}_s^{i,j} dW_s^j \right\}, \tag{38}$$

driven by a d-dimensional Wiener process $W = (W^1, \ldots, W^d)$ (uncorrelated). We first notice that each asset stock price process S^i is of the form (37) with instantaneous conditional mean rate of return μ^j and instantaneous conditional volatility

$$\sigma_t := \| \hat{\sigma}_t^{i,\cdot} \| = \sqrt{\sum_{j=1}^d (\hat{\sigma}_t^{i,j})^2}.$$

Also, from Examples 4.9, it follows that the log returns of asset i and j at time t has instantaneous correlation

$$\rho_t^{i,j} = \frac{\hat{\sigma}_t^{i\cdot} \left(\hat{\sigma}_t^{j\cdot} \right)^T}{\| \hat{\sigma}_t^{i\cdot} \| \| \hat{\sigma}_t^{j\cdot} \|}.$$

In other words,

$$\rho_t^{i,j} = \lim_{\Delta \to 0} \frac{\mathrm{Cov} \left\{ \ln \frac{S_{t+\Delta}^i}{S_t^i}, \ln \frac{S_{t+\Delta}^j}{S_t^j} \middle| \mathcal{F}_t^W \right\}}{\sqrt{\mathrm{Var} \left\{ \ln \frac{S_{t+\Delta}^i}{S_t^i} \middle| \mathcal{F}_t^W \right\} \cdot \mathrm{Var} \left\{ \ln \frac{S_{t+\Delta}^j}{S_t^j} \middle| \mathcal{F}_t^W \right\}}}.$$

5.1. **Existence and properties.**

In this part we study a d-dimensional vector $X_t := (X_t^1, \dots, X^n)^T$ of Itô processes that satisfies the following system of equations:

$$X_t^i = x_0^i + \int_0^t \mu^i(s, X_s)ds + \sum_{j=1}^{d} \int_0^t \sigma^{i,j}(s, X_s)dW_s^j, \quad \text{for all } i = 1, \dots, d, \tag{39}$$

where

$$W_t = \begin{bmatrix} W_t^1 \\ \vdots \\ W_t^d \end{bmatrix}, \quad \mu(t,x) = \begin{bmatrix} \mu^1 \\ \vdots \\ \mu^d \end{bmatrix}, \quad \sigma(t,x) = \begin{bmatrix} \sigma^{1,1}, \dots, \sigma^{1,d} \\ \vdots \\ \sigma^{n,1}, \dots, \sigma^{n,d} \end{bmatrix},$$

are respectively a d-dimensional Wiener process, a $d \times 1$ vector of functions in $[0,\infty) \times \mathbb{R}^d$, and a $n \times d$ matrix of functions in $[0,\infty) \times \mathbb{R}^d$. It is convenient to write (39) in the following form

$$X_t = x_0 + \int_0^t \mu(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s. \tag{40}$$

If there exists a process X satisfying (40), we say that this process is a (strong) solution to the following *stochastic differential equations* (SDE):

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)\, dW_t, \tag{41}$$

with initial condition $X_0 = x_0$. The functions μ and σ are respectively called the *drift* and the *diffusion* terms of the SDE.

The next result gives sufficient conditions for the existence and uniqueness of the solution.

Proposition 5.1. *Suppose that there exists a constant K such that the following conditions holds for all x, y, t:*

(i) Lipschitz conditions:

$$\|\mu(t,x) - \mu(t,y)\| \le K\|x - y\|, \quad \text{and} \quad \|\sigma(t,x) - \sigma(t,y)\| \le K\|x - y\|;$$

(ii) Linear growth condition:

$$\|\mu(t,x)\| + \|\sigma(t,x)\| \le K\left(1 + \|x\|\right).$$

Then, for any x_0, there exists a unique solution to (41) with initial condition $X_0 = x_0$. Furthermore, this solution satisfies the following properties:

(a) $E\|X_t\|^2 \le Ce^{Ct}\left(1 + \|x_0\|^2\right);$
(b) $\{X_t\}_{t\ge0}$ *is adapted to* $\{\mathcal{F}_t^W\}_{t\ge0}$ *with continuous paths;*
(c) *For any $t \le T$ and set C,*

$$P\left(X_T \in C \,\middle|\, \mathcal{F}_t^W\right) = P\left(X_T \in C \,|X_t\right). \tag{42}$$

Example 5.1. Consider the following SDE:

$$dS_t = S_t b dt + S_t \sigma dW_t, \tag{43}$$

where b and σ are constants. This equation is essentially the same equation as in (28). As in Section 4.5, $S_t = S_0 e^{\sigma W_t + (b - \sigma^2/2)t}$ is a solution by Itô's formula. In Exercise (7), it is shown that the solution is unique. Alternatively, uniqueness is a byproduct of the previous theorem since the coefficients:

$$\mu(t,x) = bx, \quad \sigma(t,x) = tx,$$

satisfy the Lipschitz and linear growth conditions.

The condition (42) is called the *Markov property*. Intuitively, this property means that the future statistical behavior of the process depends on the past $\mathcal{F}_t^W$ only through the present value of the process. The following are equivalent formulations of the Markov property (42):

(1) For "any" function f and $0 \le t \le T$:

$$E\left(f(X_T)\,\middle|\mathcal{F}_t^W\right) = E\left(f(X_T)\,|X_t\right); \tag{44}$$

(2) For $t \le t_1 < \cdots < t_n$ and sets $C_1, \dots, C_n$,

$$P\left(X_{t_1} \in C_1, \dots, X_{t_n} \in C_n \,\middle|\mathcal{F}_t^W\right) = P\left(X_{t_1} \in C_1, \dots, X_{t_n} \in C_n \,|X_t\right);$$

(3) If Y is a random variable whose value depends only on $\{X_u : u \ge t\}$ (the "future" path of X) then

$$E\left(Y \,\middle|\mathcal{F}_t^W\right) = E\left(Y\,|X_t\right);$$

(4) For each $(t,x) \in [0,\infty) \times \mathbb{R}$, let $(\Omega_{t,x}, P_{t,x})$ be a probability space equipped with a Wiener process $\{W_s^{t,x}\}_{s\ge 0}$. Suppose that $\{X_s^{t,x}\}_{s \ge t}$ is a solution of the following SDE:

$$dX_s^{t,x} = \mu(s, X_s^{t,x})ds + \sigma(s, X_s^{t,x})\,dW_s^{t,x}, \quad s \ge t \tag{45}$$

$$X_t^{t,x} = x. \tag{46}$$

We say that $\{X_s^{t,x}\}_{s \ge t}$ starts at time t at the value x according to the drift μ and diffusion σ. Then, X has the Markov property (42) if and only if

$$E\left(f(X_T)\,\middle|\mathcal{F}_t^W\right) = F_T(t, X_t), \tag{47}$$

where $F_T(t,x) = E_{t,x}\left(f(X_T^{t,x})\right)$ and $E_{t,x}$ denotes the expectation with respect to $P_{t,x}$.

Notice that in light of (44) and (47),

$$F_T(t,x) = E(f(X_T)\,|X_t = x).$$

We can interpret (47) as follows. Given the past information $\mathcal{F}_t^W$, the future evolution of the process is the same as if the process starts at time t at the value X_t (independently from whatever happen with the process before time t). It is useful to see the system (45-46) as a black box which input data are the initial position x and time t, and the final product is a realization of the stochastic process $\{X_s^{t,x}\}_{s \ge t}$.

In the case that the coefficients of the SDE are time independent,

$$dX_t = \mu(X_t)dt + \sigma(X_t)\, dW_t, \tag{48}$$

the solution will be time-homogeneous and the Markov property will reduces even further:

$$E\left(f(X_T)\,\middle|\mathcal{F}_t^W\right) = E\left(f(X_T)\,|X_t\right) = F_{T-t}(0, X_t), \tag{49}$$

where $F_T(t,x)$ is as in (47).

Example 5.2. Let $\{S_t\}_{t\geq 0}$ be the solution of (43) under a probability P. Then,

$$E\left(f(S_T)\,\middle|\mathcal{F}_t^W\right) = F_{T-t}(0, S_t),$$

where $F_{T-t}(0,x) = E_{0,x}\left(f(X_{T-t}^{0,x})\right)$ and $\{X_s^{0,x}\}_{s\geq 0}$ is the solution of the SDE

$$\begin{aligned} dX_s^{0,x} &= X_s^{0,x} b du +, X_s^{0,x}\sigma\, dW_s^{0,x}, \quad s \geq 0 \\ X_0^{0,x} &= x. \end{aligned}$$

As before, $X_s^{0,x} = x\exp\{\sigma W_s^{0,x} + (\mu - \sigma^2/2)s\}$, where $W^{0,x}$ is a Wiener process. Then,

$$F_{T-t}(0,x) = E_{0,x}\left\{f\left(xe^{\sigma W_{T-t}^{0,x} + (\mu-\sigma^2/2)(T-t)}\right)\right\}.$$

Of course, we could use the same Wiener process W, that drives S, and the expectation under P:

$$F_{T-t}(0,x) = E\left(f\left(xe^{\sigma W_{T-t} + (\mu-\sigma^2/2)(T-t)}\right)\right).$$

5.2. **Feynman-Kac stochastic representation of PDEs.**

The Feyman-Kac theorem provides a connection between two types of problems: a *Partial Differential Equation (PDE)* and a *Stochastic Differential Equation (SDE)*. The below version gives a *"stochastic representation"* for the solution of a PDE:

Theorem 5.1. *Consider the following setting:*

(1) *Let $F : [0,T] \times \mathbb{R} \to \mathbb{R}$ be a continuous solution to the following PDE with boundary conditions:*

$$\frac{\partial F}{\partial t}(t,x) + \mu(t,x)\frac{\partial F}{\partial x}(t,x) + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2 F}{\partial x^2}(t,x) = 0, \quad (t,x) \in (0,T) \times \mathbb{R} \tag{50}$$

$$F(T,x) = \Phi(x), \tag{51}$$

for given functions σ, μ, and Φ.

(2) *For each $(t,x) \in [0,T] \times \mathbb{R}$, let $(\Omega_{t,x}, P_{t,x})$ be a probability space equipped with a Wiener process $\{W_s^{t,x}\}_{s\geq 0}$ and let $\{X_s^{t,x}\}_{s\geq t}$ be a solution of the SDE:*

$$dX_s^{t,x} = \mu(s, X_s^{t,x})du + \sigma(s, X_s^{t,x})\, dW_s^{t,x}, \quad s \geq t \tag{52}$$

$$X_t^{t,x} = x. \tag{53}$$

Suppose also that $\sigma(s, X_s^{t,x}) \frac{\partial F}{\partial t}(s, X_s^{t,x})$ *is Itô square-integrable for each* t, x*. Then,* $F(t,x)$ *satisfies the* stochastic representation*:*

$$F(t,x) = E_{t,x}\left[\Phi\left(X_T^{t,x}\right)\right], \tag{54}$$

where $E_{t,x}$ *is the expectation under* $P_{t,x}$*.*

Example 5.3. Though its apparently complex form, the previous theorem has important applications in finance. For instance, suppose that one wishes to determine the value $F(t,x)$ where F is the solution of (50-51). One method is to use the stochastic representation of the previous theorem. It turns out that it is not hard to write a computer program that generates realizations (simulations) of a process $\{X_s^{t,x}\}_{s\in[t,T]}$ satisfying (at least approximately) (52)-(53). Suppose that we have at hand such a program and that we generate N realizations of the process. Let us denote $v_1, \ldots, v_N$ the final value $X_T^{t,x}$ for each of the N realizations. Then, in light of (54) and the Law of Large Numbers, we can approximate $F(t,x)$ by

$$F(t,x) \approx \frac{1}{N}\sum_{k=1}^{N} v_k.$$

The verification of Theorem 5.1 is actually quite illustrative. These are the main steps:

(1) Applying Itô's formula, the process $Z_s = F(s, X_s^{t,x})$ has the following differential form, where we have dropped the superscript (t,x):

$$\begin{aligned} dF(s,X_s) &= \frac{\partial F(s,X_s)}{\partial t}ds + \frac{\partial F(s,X_s)}{\partial x}dX_s + \frac{1}{2}\frac{\partial F^2(s,X_s)}{\partial x^2}(dX_s)^2 \\ &= \left\{\frac{\partial F(s,X_s)}{\partial t} + \mu(s,X_s)\frac{\partial F(s,X_s)}{\partial x} + \frac{1}{2}\sigma^2(s,X_s)\frac{\partial^2 F(s,X_s)}{\partial x^2}\right\} ds \\ &\quad + \frac{\partial F(s,X_s)}{\partial x}\sigma(s,X_s)dW_s. \end{aligned}$$

The terms in the second line vanish since F solves (50).

(2) Integrating from $s = t$ to $s = T$ and using (51) and (53), we have

$$\int_t^T \frac{\partial F(s,X_s)}{\partial x}\sigma(s,X_s)dW_s = \int_t^T dZ_s = F(T,X_T) - F(t,X_t) = \Phi(X_T) - F(t,x).$$

(3) Taking expectation both sides, this simplifies to

$$\mathbb{E}_{t,x}\left[\Phi(X_T)\right] - F(t,x) = 0,$$

which is (54).

The representation of Theorem 5.1 can be extended to cover other PDE. The following is particularly useful in finance:

$$\frac{\partial F}{\partial t}(t,x) + \mu(t,x)\frac{\partial F}{\partial x}(t,x) + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2 F}{\partial x^2}(t,x) - rF(t,x) = 0, \tag{55}$$

$$F(T,x) = \Phi(x), \tag{56}$$

for given functions σ, μ, Φ, and a constant r. Suppose that $\{X_s^{t,x}\}_{s\geq t}$ is a solution to equations (52-53). In this case, the trick is to consider the process $Z_s = e^{-r(s-t)}F(s, X_s^{t,x})$ on the interval $s \in [t, T]$.

(1) Applying the product formula and Itô's formula,

$$
\begin{aligned}
dZ_s &= e^{-r(s-t)}dF(s, X_s) + F(s, X_s)d(e^{-r(s-t)}) \\
&= e^{-r(s-t)}\left\{\frac{\partial F(s, X_s)}{\partial t} + \mu(s, X_s)\frac{\partial F(s, X_s)}{\partial x} + \frac{1}{2}\sigma^2(s, X_s)\frac{\partial^2 F(s, X_s)}{\partial x^2} - rF(s, X_s)\right\} ds \\
&\quad + e^{-r(s-t)}\frac{\partial F(s, X_s)}{\partial x}\sigma(s, X_s)dW_s.
\end{aligned}
$$

Clearly, the term in the second line vanishes.

(2) Integrating from $s = t$ to $s = T$ and using (51) and (53), we have

$$\int_t^T e^{-r(s-t)}\frac{\partial F(s, X_s)}{\partial x}\sigma(s, X_s)dW_s = e^{-r(T-t)}\Phi(X_T) - F(t, x).$$

(3) Taking expectation both sides,

$$F(t, x) = e^{-r(T-t)}E_{t,x}\left[\Phi(X_T)\right].$$

Exercise 10. Show that if r in (55) is time dependent, then the Feynman-Kac representation takes the form

$$F(t, x) = e^{-\int_t^T r(u)du}E_{t,x}\left[\Phi(X_T)\right].$$

The following Feynman-Kac formula begins with a SDE X starting at $t = 0$ and shows that the function $F(t, x) = E\left(f(X_T)\,|X_t = x\right)$ satisfies a PDE.

Theorem 5.2. *Let W be a Wiener process under a probability measure P. Suppose that X satisfies the following equation*

$$X_t = X_0 + \int_0^t \mu(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \quad t \geq 0.$$

Define the function

$$F(t, x) = E\left(\Phi\left(X_T\right)|X_t = x\right), \quad (t, x) \in [0, T] \times \mathbb{R},$$

which we assume is continuous in $[0, T] \times \mathbb{R}$ and such that $\frac{\partial F}{\partial t}$, $\frac{\partial F}{\partial x}$, and $\frac{\partial^2 F}{\partial x^2}$ exist and are continuous in $(0, T) \times \mathbb{R}$. Then, F satisfies the PDE (50) with the boundary condition (51).

The verification of the above version is also relevant. By definition, the process $Z_t = F(t, X_t)$ is the conditional expectation

$$Z_t = E\left(\Phi\left(X_T\right)|X_t\right).$$

By the Markov's property of X, this is the same as

$$Z_t = E\left(\Phi\left(X_T\right)\middle|\mathcal{F}_t^W\right),$$

and hence, Z is a martigale relative to $\{\mathcal{F}_t\}_{t\geq 0}$. On the other hand, applying Itô's formula,

$$dZ_t = dF(t, X_t) = \left\{ \frac{\partial F(t, X_t)}{\partial t} + \mu(t, X_t)\frac{\partial F(t, X_t)}{\partial x} + \frac{1}{2}\sigma^2(t, X_t)\frac{\partial^2 F(t, X_t)}{\partial x^2} \right\} dt + \frac{\partial F(t, X_t)}{\partial x}\sigma(t, X_t)dW_t.$$

Since Z_t is a martingale, the only possibility is that the first term on the right-hand side of the above equation vanishes. Since this holds true for any (t, X_t), we conclude that $F(t, x)$ satisfied (55). Also, clearly $F(T, x) = E\left(\Phi\left(X_T\right)|X_T = x\right) = \Phi(x)$.