

Teaching Notes #0
Elements of Probability Theory and
Stochastic Processes¹

Pietro Veronesi
Graduate School of Business
University of Chicago
Business 35909 Spring 2005

This Version: March 29, 2005

¹These teaching notes draw heavily on Duffie (1996), Karatzas and Shevre (1991) and Billingsley (1986) . They are intended for students of Business 537 only. Please, do not distribute without my prior consent.

1 Elements of Probability Theory

- $\Omega =$ set of states of nature.
 - Examples: (i) $\Omega = \{\omega_1, \dots, \omega_n\}$; (ii) $\Omega = \mathcal{R}^n$
- $\mathcal{F} = \sigma$ -algebra on Ω . That is, a set of subsets $B_i \subseteq \Omega$ such that:
 1. $\emptyset \in \mathcal{F}$;
 2. $B_i \in \mathcal{F} \implies B_i^C \in \mathcal{F}$;
 3. $B_i \in \mathcal{F}$ for $i = 1, 2, \dots \implies \cup_i B_i \in \mathcal{F}$.
 - Examples: $\mathcal{F} = \{\emptyset, \Omega\}$ is the trivial σ -algebra; For (i), $\mathcal{F} = 2^\Omega =$ set of all subset of Ω ; For (ii) $\mathcal{F} = \mathcal{B}(\mathcal{R}^n) =$ Borel σ -algebra = sigma algebra generated by all the open sets on \mathcal{R}^n .
- $P =$ Probability measure on Ω , that is a function $P : \mathcal{F} \longrightarrow [0, 1]$ such that
 1. $P(\emptyset) = 0$;
 2. $P(\Omega) = 1$;
 3. For every B_1, B_2, \dots , such that $B_i \cap B_j = \emptyset$

$$P(\cup_i B_i) = \sum_i P(B_i)$$

- Examples: For (i) any distribution $p = \{p_1, \dots, p_n\}$ on $\Omega = \{\omega_1, \dots, \omega_n\}$ such that $p_i \geq 0$ and $\sum_{i=1}^n p_i = 1$; For (ii) take any non-decreasing function $F : \mathcal{R} \longrightarrow [0, 1]$ such that $F(-\infty) = 0$, $F(\infty) = 1$ and for every $(a, b) \in \mathcal{B}(\mathcal{R})$ define $P((a, b)) = F(b) - F(a)$.

- (Ω, \mathcal{F}) is called a *measurable space*.
- (Ω, \mathcal{F}, P) is called a *probability space*.
- A probability space (Ω, \mathcal{F}, P) is *complete* if all the subsets of sets with zero probability are part of \mathcal{F} ;
 - The use of complete probability spaces is for technical reasons only. We will discuss this below.
- An event $B_i \in \mathcal{F}$ is *almost sure* if $P(B_i) = 1$;

1.1 Random Variables

- A *random variable* X is a function $X : \Omega \longrightarrow \mathcal{R}$ that is *measurable* with respect to \mathcal{F} , that is with the property that for every $B \in \mathcal{B}(R)$, $\{\omega \in \Omega : X(\omega) \in B\} = X^{-1}(B) \in \mathcal{F}$.
 - We need this property to assign a probability distribution on the realizations of X !
 - For example, if
 - * $\Omega = \{\omega_1, \omega_2, \omega_3\}$;
 - * $\mathcal{F} = \{\emptyset, \Omega, \omega_1, \{\omega_2, \omega_3\}\}$,
 - * and $P(\omega_1) = p_1$ and $P(\{\omega_2, \omega_3\}) = 1 - p_1$
 - if the random variable $X : \Omega \longrightarrow \mathcal{R}$ is such that $X(\omega_2) \neq X(\omega_3)$, we would not be able to assign a probability to either values $X(\omega_2)$ and $X(\omega_3)$.
 - Hence, a random variable defined on the space (Ω, \mathcal{F}, P) must be such that $X(\omega_2) = X(\omega_3)$.

- Let $X : \Omega \longrightarrow \mathcal{R}$ be given. We say that the σ -algebra \mathcal{F} is *generated* by X if it is the smallest σ -algebra that makes X measurable with respect to it. We often denote it $\sigma(X)$.

– Example:

- * Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and $X = \{a, b, b\}$.
 - * Then $\mathcal{F} = \{\emptyset, \Omega, \omega_1, \{\omega_2, \omega_3\}\}$ is the smallest σ -algebra that makes $X : \Omega \longrightarrow \mathcal{R}$ measurable.
 - * Notice that X is also measurable with respect to the σ -algebra $\mathcal{F} = 2^\Omega$.
- A random variable X is *simple* if $X = \sum_{i=1}^n \alpha_i 1_{B_i}$ where 1_{B_i} is the indicator function of the event B_i . Let \mathcal{S} be the set of all simple random variables on Ω .
 - The *expectation of X* is defined by

$$E[X] = \sum_{i=1}^n \alpha_i P(B_i)$$

- Let X be a non-negative random variable: Then, we define:

$$E[X] = \sup_{Y \in \mathcal{S}} E(Y) \text{ subject to } Y \leq X$$

- More generally, any random variable X can be written as $X = X^+ - X^- = \max(X, 0) - \max(-X, 0)$. If $E[X^+]$ and $E[X^-]$ are *finite*, then X is said to be *integrable* and we define

$$E[X] = E[X^+] - E[X^-]$$

1.2 Convergence

- A sequence $\{X_n\}$ of random variables converges to X :
 - (1) *in distribution*, if for every bounded $f : \mathcal{R} \rightarrow \mathcal{R}$, we have $E[f(X_n)] \rightarrow E[f(X)]$;
 - (2) *in probability*, if for every $\varepsilon > 0$, we have $P(|X_n - X| < \varepsilon) \rightarrow 0$;²
 - (3) *almost surely*, if there exists $B \subseteq \Omega$ such that $P(B) = 1$ and $X_n(\omega) \rightarrow X(\omega)$ for every $\omega \in B$.
- We have $(3) \implies (2) \implies (1)$.
- The following results of probability theory find wide applications in finance:
- *Monotone Convergence Theorem*: Let $\{X_n\}$ be a sequence of random variables such that $X_n \geq 0$ and let $X_n \uparrow X$ almost surely. Then

$$E[X_n] \uparrow E[X]$$

²Notice that the notation $P(|X_n - X| < \varepsilon)$ really stands for $P(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \varepsilon\})$, that is, the probability of those states of nature $\omega \in \Omega$ such that the condition holds.

- *Fatou's Lemma*: Let $\{X_n\}$ be $X_n \geq 0$. Then³

$$E [\liminf X_n] \leq \liminf E [X_n]$$

- *Dominated Convergence Theorem*: Let $\{X_n\}$ be a sequence of random variables defined on a probability space with $|X_n| < Y$ for all n , where Y is a random variable with $E [|Y|] < \infty$. Suppose that $X_n \rightarrow X$ almost surely. Then

$$E [X_n] \rightarrow E [X]$$

1.3 Equivalent Probability Measures and the Radon-Nikodym Theorem

- Let (Ω, \mathcal{F}) be a measurable space and let P and Q be two probability measures defined on (Ω, \mathcal{F}) .
- The probability measure Q is *absolutely continuous* with respect to P if for each $A \in \mathcal{F}$, $P(A) = 0$ implies $Q(A) = 0$.
- The measures Q and P are *equivalent* if each one is absolutely continuous with respect to the other.
- An important example of an absolutely continuous probability measure is the following:

³Given a sequence of real numbers $\{x_n\}$, let $E = \{x : \text{there exists a subsequence } \{x_{n_k}\} \text{ with } x_{n_k} \rightarrow x\}$ and let $x^* = \sup E$ and $x_* = \inf E$. x^* and x_* are defined as the upper limit and lower limit of $\{x_n\}$ and denoted as $x^* = \limsup x_n$ and $x_* = \liminf x_n$. Given a sequence of random variables $X_n : \Omega \rightarrow \mathcal{R}$ we define the random variable $X_*(\omega) = \liminf X_n(\omega)$ by $X_* = \liminf X_n$.

- Fix a random variable $\xi : \Omega \longrightarrow \mathcal{R}$ with $\xi \geq 0$ and $E[\xi] = 1$. If we define for all $A \in \mathcal{F}$

$$Q(A) = \int_A \xi(\omega) dP(\omega)$$

Then, $Q(A)$ is a probability measure that is absolutely continuous with respect to P .

- *Radon-Nikodym theorem* goes in the opposite direction:
 - Let Q and P be two measures with Q absolutely continuous with respect to P . Then there exists a *non-negative* random variable ξ such that

$$Q(A) = \int_A \xi(\omega) dP(\omega)$$

for all $A \in \mathcal{F}$.

- In addition, Q and P are equivalent if and only if ξ is strictly positive.
- This random variable ξ is called the *Radon-Nikodym derivative* and it is often denoted by dQ/dP .
- Two properties of the Radon-Nikodym derivative are as follows: Let Z be a random variable such that $E^Q[|Z|] < \infty$. Then

1. $E^Q[Z] = E^P[\xi Z]$

2. If $\mathcal{G} \subset \mathcal{F}$,

$$E^Q[Z|\mathcal{G}] = \frac{E^P[\xi Z|\mathcal{G}]}{E^P[\xi|\mathcal{G}]}$$

1.4 Filtrations and Processes

- Suppose that there are multiple periods and let \mathcal{T} be the set of times (e.g. $\mathcal{T} = \{0, 1, \dots, T\}$ or $\mathcal{T} = [0, T]$ with T finite or infinite).
- Fix the set of states Ω . A *filtration* $\{\mathcal{F}_t : t \in \mathcal{T}\}$ is a family of sub- σ -algebras such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ if $s \leq t$.
- Intuitively, the “filtration” describes the evolution of information over time.

– Example: $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Then a filtration could be

$$\mathcal{F}_0 = \{\emptyset, \Omega\}$$

$$\mathcal{F}_1 = \{\emptyset, \Omega, \omega_1, \{\omega_2, \omega_3\}\}$$

$$\mathcal{F}_2 = 2^\Omega = \left\{ \begin{array}{l} \emptyset, \Omega, \omega_1, \{\omega_2, \omega_3\}, \\ \omega_2, \{\omega_1, \omega_3\}, \omega_3, \{\omega_1, \omega_2\} \end{array} \right\}$$

- In this example, at time $t = 0$ there is no information, at time $t = 1$ we know whether ω_1 realized while at time $t = 2$ we have perfect information.
- The following two notions are often used:
 - A filtration is *left continuous* if $\mathcal{F}_t = \sigma(\cup_{s < t} \mathcal{F}_s)$: That is, if it is the σ -algebra generated by the events occurring strictly before t .
 - A filtration is *right continuous* if $\mathcal{F}_t = \cap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon}$ for every $t \geq 0$: That is, any event occurring exactly at time t is “included” in the information set at time t .
- A *process* X defined on the probability space (Ω, \mathcal{F}, P) is a set of random variables $X_t : \Omega \longrightarrow \mathcal{R}$.

- The process X is called *measurable* if for every $B \in \mathcal{B}(\mathcal{R})$ the set $\{(t, \omega) : X_t(\omega) \in B\}$ belongs to the product σ -algebra $\mathcal{B}(\mathcal{T}) \otimes \mathcal{F}$.
- Given $\omega \in \Omega$, the function $X_t(\omega) : \mathcal{T} \longrightarrow \mathcal{R}$ is called *trajectory* or *realization* of the process, corresponding to ω .
- A process X is *adapted* to the filtration $\{\mathcal{F}_t\}$ if for every t , the random variable X_t is measurable with respect to \mathcal{F}_t .
- That is to say, given our information at time t described by \mathcal{F}_t , we can fully “observe” the value of X_t by observing a realization in \mathcal{F}_t .
- Let X be a given process. Then, a natural choice of filtration is the one *generated* by the process itself

$$\mathcal{F}_t = \sigma(X_s; 0 \leq s \leq t)$$

That is, the smallest σ -algebra with respect to which X_s is measurable for every $s \in [0, t]$.

- Example: If $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and the process X_t is such that

$$X_0(\omega) = \{a, a, a\}$$

$$X_1(\omega) = \{a, b, b\}$$

$$X_2(\omega) = \{a, b, c\}$$

then for every $t = 0, 1, 2$, the smallest σ -algebra that makes X_s measurable is given by

$$\mathcal{F}_0 = \{\emptyset, \Omega\},$$

$$\begin{aligned}\mathcal{F}_1 &= \{\emptyset, \Omega, \omega_1, \{\omega_2, \omega_3\}\} \\ \mathcal{F}_2 &= 2^\Omega\end{aligned}$$

- In fact, X_0 is measurable with respect to \mathcal{F}_0 ; X_1 and X_0 are measurable with respect to \mathcal{F}_1 and X_0, X_1 and X_2 are measurable with respect to \mathcal{F}_2 .
- By construction, the process is adapted to the filtration $\{\mathcal{F}_t\}$.
- Notice that the process X it is *not adapted* to the alternative filtration

$$\begin{aligned}\mathcal{F}_0 &= \{\emptyset, \Omega\} \\ \mathcal{F}_1 &= \{\emptyset, \Omega, \omega_3, \{\omega_1, \omega_2\}\} \\ \mathcal{F}_2 &= 2^\Omega\end{aligned}$$

- At time $t = 1$, we know whether ω_3 realized or not.
 - If ω_3 did not realize, we cannot tell what is the value of X_1 : Is it ‘ a ’ or ‘ b ’? We *cannot measure* X_1 given our information \mathcal{F}_1 .
- *Remark:* Although intuitively the filtration generated by the process X is what we need for our finance applications (it describes the information available at every time t , depending on the path of X from 0 to t only), for technical reason this filtration is not rich enough. We need to complete it, that is, augment the filtration by the subsets of sets with zero probability. This is necessary to make the filtration $\{\mathcal{F}_t\}$ right continuous.

- Sometimes a stronger notion of measurability is necessary (but the intuition is the same).
- A process X is *progressively measurable* with respect to the filtration $\{\mathcal{F}_t\}$ if for every $t \geq 0$ and $A \in \mathcal{B}(\mathcal{R})$, the set $\{(s, \omega) : 0 \leq s \leq t, \omega \in \Omega, X_s(\omega) \in A\}$ belongs to the σ -algebra $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$.
- From results in analysis, we have that if X is progressively measurable, then it is measurable and adapted. The converse is not always true.
- However, we have the following: If X is adapted to $\{\mathcal{F}_t\}$ and every sample path is right continuous (or left continuous), then X is also progressively measurable with respect to \mathcal{F}_t .
- Finally, a very useful result is the following:
- *Fubini Theorem*: Let X_t be a measurable process defined on a probability space (Ω, \mathcal{F}, P) . If

$$E \left(\int_0^T |X_t| dt \right) < \infty$$

Then

$$E \left(\int_0^T X_t dt \right) = \int_0^T E [X_t] dt$$

- That is, we can change the order of integration.

1.5 Stopping Times

- Suppose we are interested in the occurrence of a certain phenomenon: An earthquake for example.

- We are then interested in instant T at which this occurs *for the first time*.
- Clearly, the event $\{\omega : T(\omega) \leq t\}$ should be part of the information accumulated up to t , that is \mathcal{F}_t .
- A *random time* T is a random variable with values in $[0, \infty)$;
- Let (Ω, \mathcal{F}) be a measurable space endowed with the filtration $\{\mathcal{F}_t\}$. A random time T is a *stopping time* of the filtration, if for every $t \geq 0$ the event $\{\omega : T(\omega) \leq t\}$ belongs to the σ -algebra \mathcal{F}_t . It is instead an *optional time* if the event $\{\omega : T(\omega) < t\}$ belongs to the σ -algebra \mathcal{F}_t .
 - If the filtration is right-continuous, then stopping time = optional time.
 - If T and S are stopping times, then so are $T \wedge S$, $T \vee S$ and $T + S$.⁴
- Finally, we have the following: Given a stopping time T and a process $\{X_t, \mathcal{F}_t\}$, define the random variable on the set $\{T < \infty\}$

$$X_T(\omega) \equiv X_{T(\omega)}(\omega)$$

- Then, random variable X_T is \mathcal{F}_T -measurable, and the “*stopped process*” $\{X_{T \wedge t}, \mathcal{F}_t\}$ is progressively measurable.
 - The use of “stopped process” is very common in continuous time finance, because it allows us to define a stochastic integral under mild conditions on the integrand function.

⁴The notation is standard: $T \wedge S = \min(T, S)$ and $T \vee S = \max(T, S)$.

1.6 Conditional Expectations⁵

- For any integrable random variable X on (Ω, \mathcal{F}, P) and given a sub- σ -algebra $\mathcal{G} \subseteq \mathcal{F}$, there exists a random variable Y that satisfies

1. Y is integrable and measurable with respect to \mathcal{G} ;
2. Y satisfies the functional equation

$$\int_G Y dP = \int_G X dP \text{ for all } G \in \mathcal{G}$$

- Y is called *conditional expectation* of X given the sub- σ -algebra $\mathcal{G} \subseteq \mathcal{F}$ and is denoted by $E[X|\mathcal{G}]$.
- This random variable has several properties, which also allow us to derive the conditional probability. In the following, all random variables are assumed integrable.
 1. $E[X|\{\emptyset, \Omega\}] = E[X]$;
 2. $E[X|\mathcal{F}] = X$;
 3. If $X = a$ almost surely, then $E[X|\mathcal{G}] = a$;
 4. If a and b are constants, then $E[aX + bY|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}]$;
 5. If $X \leq Y$ almost surely, then $E[X|\mathcal{G}] \leq E[Y|\mathcal{G}]$;
 6. $|E[X|\mathcal{G}]| \leq E[|X||\mathcal{G}]$;
 7. If $\lim_n X_n = X$ a.s., $|X_n| \leq Y$, and Y is integrable, then $\lim_n E[X_n] = E[X]$ a.s.
 8. If X is measurable w.r.t. \mathcal{G} , and if Y and XY are integrable, then

$$E[XY|\mathcal{G}] = XE[Y|\mathcal{G}]$$

⁵This section is from Billingsley (1986)

9. If X is integrable and the σ -fields \mathcal{G}_1 and \mathcal{G}_2 satisfy $\mathcal{G}_1 \subset \mathcal{G}_2$, then almost surely

$$E[E[X|\mathcal{G}_2]|\mathcal{G}_1] = E[X|\mathcal{G}_1]$$

- These last properties imply the following. Let T and S be two stopping times. Then
 1. $E[X|\mathcal{F}_T] = E[X|\mathcal{F}_{T \wedge S}]$ on $\{T \leq S\}$;
 2. $E[E[X|\mathcal{F}_T]|\mathcal{F}_S] = E[X|\mathcal{F}_{T \wedge S}]$.
- Notice that all these properties imply equivalent properties for the *conditional probability measures*. In fact, given the sub- σ -algebra \mathcal{G} , for all $A \in \mathcal{G}$ we have

$$P(A|\mathcal{G}) = E[1_A|\mathcal{G}]$$

1.7 Martingales

- An adapted, integrable process X_t is
 - a martingale if $E[X_t|\mathcal{F}_s] = X_s$ for $t \geq s$;
 - a sub-martingale if $E[X_t|\mathcal{F}_s] \geq X_s$ for $t \geq s$;
 - a super-martingale if $E[X_t|\mathcal{F}_s] \leq X_s$ for $t \geq s$;
- A martingale can also be defined as an adapted, integrable process such that for any bounded stopping time T , $E[X_T] = E[X_0]$;
- The following results are often used in asset pricing:
 - Let $\{X_t, F_t\}$ be a martingale (submartingale) and let $\phi : R \rightarrow R$ be a convex (convex, non-decreasing) function, such that $E[\phi(X_t)] < \infty$ for $t \geq 0$. Then $\{\phi(X_t), F_t\}$ is a submartingale.

- *Doob-Meyer Decomposition Theorem*: Let the filtration $\{\mathcal{F}_t\}$ be right-continuous and such that \mathcal{F}_0 contains all the P -null sets of \mathcal{F} . If the right continuous submartingale $X = \{X_t, \mathcal{F}_t\}$ satisfies some regularity conditions (loosely, they are uniformly integrable), then there is a unique decomposition

$$X_t = M_t + A_t$$

where $\{M_t, \mathcal{F}_t\}$ is a right-continuous martingale and A_t is an increasing process, that is, such that for each $\omega \in \Omega$, $A_t(\omega)$ is increasing in t .

1.7.1 Local Martingales

- The notion of martingale is also local.
- Let $\{X_t, \mathcal{F}_t\}$ be a process. If there exists a non-decreasing sequence $\{T_n\}_{n=1}^{\infty}$ of stopping times of $\{\mathcal{F}_t\}$ such that the stopped process $\{X_t^n = X_{t \wedge T_n}, \mathcal{F}_t\}$ is a martingale for each $n \geq 1$ and $P[\lim_{n \rightarrow \infty} T_n = \infty] = 1$, then we say that X is a *local martingale*.
 - Clearly, any martingale is also a local martingale. However, the opposite is not true. We will see some examples where local martingales are not martingales.
 - It turns out that these properties are useful to study issues related to no-arbitrage.
- **Result 1**: any non-negative local martingale is a super martingale.

- This is an immediate consequence of Fatou’s Lemma. In fact, since $\{X_t^n\}$ are non-negative, for any t and $s > t$

$$E_t[X_s] = E_t[\liminf X_s^n] \leq \liminf E_t[X_s^n] = X_t$$

- An immediate consequence is also
- **Result 2:** Any local martingale bounded from below is a super-martingale.
- A final important definition before moving to Brownian motions:
- A right-continuous martingale $\{X_t, \mathcal{F}_t\}$ is *square-integrable* if $E[X_t^2] < \infty$. If $X_0 = 0$ a.s., we write $X \in \mathcal{M}_2$.
- Since $\{X_t^2, \mathcal{F}_t\}$ is a non-negative sub-martingale, it can be decomposed as

$$X_t^2 = M_t + A_t$$

where $\{M_t, \mathcal{F}_t\}$ is a right-continuous martingale and $\{A_t, \mathcal{F}_t\}$ is an increasing process.

- For $X \in \mathcal{M}_2$, we define the *quadratic variation* of X to be the process $\langle X \rangle_t \equiv A_t$. That is, $\langle X \rangle$ is the unique, adapted (natural) increasing process for which $\langle X \rangle_0 = 0$ a.s. and $X_t - \langle X \rangle_t$ is a martingale.
- This terminology is due to the following result: Fix t and let $\Pi = \{t_0, t_1, \dots, t_m\}$ with $0 = t_0 \leq t_1 \leq \dots \leq t_m = t$ be a finite partition of $[0, t]$. Define the *p-variation* of the process X over the partition Π as

$$V_t^{(p)}(\Pi) = \sum_{k=1}^m |X_{t_k} - X_{t_{k-1}}|^p$$

- Let $|\Pi| = \max(t_k - t_{k-1})$. If the process X is a continuous martingale, then

$$\langle X \rangle_t = \lim_{|\Pi| \rightarrow 0} V_t^{(2)}(\Pi) \text{ (in probability)}$$

2 Brownian Motions and Stochastic Integrals

- Fix a probability space (Ω, \mathcal{F}, P) . A *standard, one dimensional Brownian motion* (or Wiener process) is a continuous, adapted process $B = \{B_t, \mathcal{F}_t\}$ defined on the probability space (Ω, \mathcal{F}, P) with the properties
 1. $B_0 = 0$;
 2. $B_t - B_s$ is independent of \mathcal{F}_s ;
 3. $B_t - B_s \sim \mathcal{N}(0, t - s)$.
- From its definition the Brownian motion $\{B_t, \mathcal{F}_t\}$ is a square-integrable martingale.
- It also has the following two properties

1. Infinite variation

$$\lim_{|\Pi| \rightarrow 0} V_t^{(1)}(\Pi) = \infty \text{ a.s.}$$

2. Finite quadratic variation

$$\lim_{|\Pi| \rightarrow 0} V_t^{(2)}(\Pi) = t \text{ a.s.}$$

- Property (1) implies that the standard BM is nowhere differentiable.

- A process X is a Brownian motion if

$$X_t = X_0 + \mu t + \sigma B_t$$

where B_t is a standard Brownian motion.

- A d -dimensional standard Brownian motion is a vector $\{\mathbf{B}_t, \mathcal{F}_t\}$ with $\mathbf{B}_t = (B_t^1, \dots, B_t^d)$ of standard Brownian motions such that $\mathbf{B}_t - \mathbf{B}_s$ is independent of \mathcal{F}_s and $\mathbf{B}_t - \mathbf{B}_s \sim \mathcal{N}(0, (t - s) \mathbf{I}_d)$.

2.1 Stochastic Integration

- Let $\{B_t, \mathcal{F}_t\}$ be a standard Brownian motion.
- An adapted process $\theta : [0, T] \times \Omega \rightarrow \mathcal{R}$ is simple if there is a partition $\Pi = \{t_0, t_1, \dots, t_m\}$ with $0 = t_0 \leq t_1 \leq \dots \leq t_m = T$ such that $\theta_t = \theta_{t_n}$ for $t \in [t_n, t_{n+1})$.
- In this case, it is natural to define the stochastic integral $\int \theta dB$ for any $t \in [t_n, t_{n+1})$ as

$$\int_0^t \theta_s dB_s = \sum_{i=0}^{n-1} \theta_{t_i} [B_{t_{i+1}} - B_{t_i}] + \theta_{t_n} [B_t - B_{t_n}]$$

- We will denote the following spaces: Let \mathcal{L} be the set of adapted processes. Then

$$\begin{aligned} \mathcal{L}^1 &= \left\{ \theta \in \mathcal{L} : \int_0^T |\theta_t| dt < \infty \right\} \\ \mathcal{L}^2 &= \left\{ \theta \in \mathcal{L} : \int_0^T |\theta_t|^2 dt < \infty \right\} \\ \mathcal{H}^1 &= \left\{ \theta \in \mathcal{L}^2 : E \left[\left(\int_0^T \theta_t^2 dt \right)^{\frac{1}{2}} \right] < \infty \right\} \\ \mathcal{H}^2 &= \left\{ \theta \in \mathcal{L}^2 : E \left(\int_0^T \theta_t^2 dt \right) < \infty \right\} \end{aligned}$$

- Then, we have that for any $\theta \in \mathcal{H}^2$, there is a sequence of $\{\theta^n\}$ of adapted simple processes such that

$$E \left[\int_0^T [\theta_t^n - \theta_t]^2 dt \right] \longrightarrow 0$$

- Also, there is a unique random variable Y_θ such that for any such sequence $\{\theta^n\}$ we have

$$E \left[\left(Y_\theta - \int_0^T \theta_t^n dB_t \right)^2 \right] \longrightarrow 0$$

- This random variable will be called stochastic integral of the process $\{\theta_t, \mathcal{F}_t\}$ and it will be denoted

$$Y_\theta = \int_0^T \theta_t dB_t$$

- The stochastic integral has some nice properties: Let $\theta, \vartheta \in \mathcal{L}^2$

1. If $\theta \in \mathcal{H}^1$ or \mathcal{H}^2 , then $\int_0^t \theta_s dB_s$ is a martingale;
2. If $\theta \in \mathcal{L}^2$, then $\int_0^t \theta_s dB_s$ is a local martingale;
 - That is, for every n , define the stopping time $\tau(n) = \inf \{t : \int_0^t \theta_s^2 ds = n\}$, let $\theta_t^{(n)} = \theta_t 1_{\{t \leq \tau(n)\}}$.
 - Then, $\theta_t^{(n)} \in \mathcal{H}^2$ which implies $\int_0^t \theta_s^{(n)} dB_s$ is a martingale.

- Since $\tau(n) \rightarrow T$ almost surely, the stochastic integral $\int_0^t \theta_s dB_s$ is defined as the limit of $\int_0^t \theta_s^{(n)} dB_s$ as $n \rightarrow \infty$.

$$3. \int_0^T (\alpha \theta_t + \beta \vartheta_t) dB_t = \alpha \int_0^T \theta_t dB_t + \beta \int_0^T \vartheta_t dB_t$$

- Other properties will be discussed later.
- The following result links martingales to Brownian motions and it is of great use:
- *The Martingale Representation Theorem:* Let $\{M_t, \mathcal{F}_t\}$ be a local martingale, then there exists a d -dimensional process $\theta_t \in (\mathcal{L}^2)^d$ such that

$$M_t = M_0 + \int_0^t \theta_s d\mathbf{B}_s$$

2.2 Ito Processes and Ito's formula

- A Ito process is a stochastic process X_t of the form

$$X_t = X_0 + \int_0^t \theta_s ds + \int_0^t \sigma_s dB_s$$

where σ_s is such that

$$P \left[\int_0^t \sigma_s^2 ds < \infty \text{ for all } t \geq 0 \right] = 1$$

- A short-hand for this notation is the stochastic differential equation (SDE)

$$dX_t = \theta_t dt + \sigma_t dB_t \tag{1}$$

- From the definition of Brownian motion , we have the following “rules”

$$dt \cdot dt = dt \cdot dB_t = 0 \text{ and } dB_t \cdot dB_t = dt$$

- *Ito’s formula:* Let X_t be the Ito process in (1) and let $g : [0, T] \times \mathcal{R} \longrightarrow \mathcal{R}$ be twice continuously differentiable, then

$$Y_t = g(t, X_t)$$

is also a Ito’s process with

$$dY_t = \frac{\partial g}{\partial t}(t, X_t) dt + \frac{\partial g}{\partial X}(t, X_t) dX_t + \frac{1}{2} \frac{\partial^2 g}{\partial X^2}(t, X_t) (dX_t)^2$$

- *Multidimensional Ito’s formula:* Let \mathbf{B}_t be a d -dimensional Brownian motion and let $\boldsymbol{\theta}_t$ and $\boldsymbol{\Sigma}_t$ be a $1 \times n$ vector and $n \times n$ matrix of processes in \mathcal{L}^2 . An n -dimensional Ito’ process is defined as

$$d\mathbf{X}_t = \boldsymbol{\theta}_t dt + \boldsymbol{\Sigma}_t d\mathbf{B}_t \tag{2}$$

- Let $g : [0, T] \times \mathcal{R}^d \longrightarrow \mathcal{R}^p$ be a twice continuously differentiable function. Then, the p -dimensional vector

$$\mathbf{Y}_t = g(t, \mathbf{X}_t)$$

is an Ito’s process, whose component number k is given by

$$\begin{aligned} dY_t^k &= \frac{\partial g_k}{\partial t}(t, \mathbf{X}_t) dt + \sum_{i=1}^n \frac{\partial g_k}{\partial X^i}(t, \mathbf{X}_t) dX_t^i \\ &+ \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 g_k}{\partial X^i \partial X^j}(t, \mathbf{X}_t) (dX_t^i) (dX_t^j) \end{aligned}$$

where

$$dB_t^i \cdot dB_t^j = \rho_{ij} dt$$

- The following property is of great use:
- *The unique decomposition property of Ito's processes:* Consider two processes X and Y such that $X_0 = Y_0$ and with

$$\begin{aligned} dX_t &= \mu_t dt + \sigma_t dB_t \\ dY_t &= a_t dt + b_t dB_t \end{aligned}$$

- We then have that $X_t = Y_t$ almost surely if and only if $\mu_t = a_t$ almost everywhere and $\sigma_t = b_t$ almost everywhere.⁶

2.3 Girsanov's Theorem

- A vector $\boldsymbol{\theta} = (\theta^1, \dots, \theta^d) \in (\mathcal{L}^2)^d$ satisfies the *Novikov's condition* if

$$E \left[\exp \left(\frac{1}{2} \int_0^T \boldsymbol{\theta}_s \cdot \boldsymbol{\theta}'_s ds \right) \right] < \infty$$

- If $\boldsymbol{\theta}$ satisfies the Novikov condition, then a martingale ξ^θ is defined by

$$\xi_t^\theta = \exp \left(- \int_0^t \boldsymbol{\theta}_s d\mathbf{B}_s - \frac{1}{2} \int_0^t \boldsymbol{\theta}_s \boldsymbol{\theta}'_s ds \right)$$

⁶Two processes a_t and μ_t are said to be equal almost everywhere if

$$E \left[\int_0^T |a_t - \mu_t| dt \right] = 0.$$

- Notice that $\xi_0^\theta = 1$. Since this is a martingale, we then have $E[\xi_T^\theta] = 1$. But then, we can use the Radon-Nikodym theorem to define a probability measure that is equivalent to P as follows

$$\frac{dQ^\theta}{dP} = \xi_T^\theta$$

- Notice also that by Ito's lemma we have

$$d\xi_t^\theta = -\xi_t^\theta \boldsymbol{\theta}_s d\mathbf{B}_s$$

- *Girsanov's Theorem*: Let $\boldsymbol{\theta} \in (\mathcal{L}^2)^d$ be given and let ξ_t^θ be a martingale. Then a standard Brownian motion that is a martingale under Q is defined by

$$B_t^\theta = B_t + \int_0^t \boldsymbol{\theta}_s ds$$

- In addition, B^θ has the martingale representation property under Q^θ : For any local Q^θ -martingale M , there is a process $\varphi \in (\mathcal{L}^2)^d$ such that

$$M_t = M_0 + \int_0^t \varphi_s dB_s^\theta$$

- We shall use this property often. An additional property is the following:
- **Corollary 1**: Let X be an Ito process in \mathcal{R}^N :

$$\mathbf{X}_t = \mathbf{x} + \int_0^t \boldsymbol{\mu}_s ds + \int_0^t \boldsymbol{\sigma}_s d\mathbf{B}_s$$

where $\boldsymbol{\sigma}_s$ is $N \times d$.

- Suppose $\boldsymbol{\nu} = (\nu^1, \dots, \nu^N)$ is a vector process in \mathcal{L}^1 such that there exists some $\boldsymbol{\theta} \in (\mathcal{L}^2)^d$ such that

$$\boldsymbol{\sigma}_t \boldsymbol{\theta}_t = \boldsymbol{\mu}_t - \boldsymbol{\nu}_t$$

- Then, if ξ^θ is a martingale, the \mathbf{X} is also an Ito process with respect to the probability space $(\Omega, \mathcal{F}, Q^\theta)$ and

$$\mathbf{X}_t = \mathbf{x} + \int_0^t \boldsymbol{\nu}_s ds + \int_0^t \boldsymbol{\sigma}_s d\mathbf{B}_s^\theta$$

- That is to say, the Girsanov's Theorem gives us a way to adjust probability assessments so that a given process can be rewritten as an Ito process with almost arbitrary drift.
- Notice the following implications.
- **Corollary 2:** Let Q be *any* probability measure equivalent to P and let $\{\mathcal{F}_t\}$ be a filtration. By the law of iterated expectations, the process

$$M_t = E \left[\frac{dQ}{dP} \middle| \mathcal{F}_t \right]$$

is martingale.

- Hence, by the Martingale Representation Theorem, we can write

$$dM_t = \boldsymbol{\vartheta}_t d\mathbf{B}_t$$

for some process $\boldsymbol{\vartheta}_t \in (\mathcal{L}^2)^d$. Hence, M_t is continuous.. It is also strictly positive because Q and P are equivalent. Hence,

we can define

$$\boldsymbol{\theta}_t = -\frac{\boldsymbol{\vartheta}_t}{M_t}$$

- By defining also the process

$$\xi_t^\theta = \exp\left(-\int_0^t \boldsymbol{\theta}_s d\mathbf{B}_s - \frac{1}{2} \int_0^t \boldsymbol{\theta}_s \boldsymbol{\theta}_s' ds\right)$$

From Ito's Lemma we have $d\xi_t^\theta = -\xi_t^\theta \boldsymbol{\theta}_t d\mathbf{B}_t$. If $M_t = \xi_t^\theta$, then $dM_t = -M_t \boldsymbol{\theta}_t d\mathbf{B}_t = \boldsymbol{\vartheta}_t d\mathbf{B}_t$ (and viceversa) Hence, $M_t = \xi_t^\theta$ and $Q = Q^\theta$.

- **Diffusion Invariance Principle:** Let X be an Ito process with $dX_t = \mu_t dt + \sigma_t dB_t$. If X_t is a martingale with respect to an equivalent probability measure Q , then there exists a Brownian motion \widehat{B} in \mathcal{R}^d under Q , such that $dX_t = \sigma_t d\widehat{B}_t$, for $t \in [0, T]$.