1 Precis for first Chapter

Options- financial contracts/instruments used for hedging risk in a way similar to insurance policies. It is also used for speculation.

Definition 1.1 (European Call Option) A European Call option gives the holder the right, but not the obligation to buy a stock

1. on a certain date $T$, called exercise date or date of maturity; ($T - t$ is thus the time to maturity)

2. for a given price $K$, called a strike price.

The payoff of the EU call option at maturity is

$$C_T = C(S_T, T) = (S_T - K)^+ = \max\{0, S_T - K\},$$

where $S = (S_t, 0 \leq t \leq T)$ is the price of the underlying stock.

A put option is the same animal, but with buy replace by sell. Major problem of Financial Calculus is to determine the price of the option at times prior to maturity:

$$C(S_t, t) = ?, \quad t < T.$$ 

This will allow us to set a price for the Option at time $t = 0$. Thus a speculator will make a profit of

$$S_T - K - C_0$$

if the option is in the money. Plot payoff and profit diagram.

Definition 1.2 (Forward Contract) Exactly the same as a call option, but both parties are obliged to follow the contract.
The forward contract is easier to price. What should the price be?
Seller of forward contract agrees to sell 1 unit of stock at price \( K \) at time \( T \).
His strategy is:

1. Borrow \( S_0 \) $ to buy stock now \( t = 0 \).
2. At time \( t = T \), deliver the stock as per contract. Get \( K \) $ and pay \( S_0 e^{rT} \) for the loan.

A rational seller will want \( K - S_0 e^{rT} \geq 0 \). Thus we know that necessarily

\[
K \geq S_0 e^{rT}.
\]

The buyer of the contract agrees to buy 1 unit of stock at price \( K \) at time \( T \).
Consider his strategy:

1. Short sell the stock now and get \( S_0 \) $.
2. Invest money at risk free rate now.
3. At maturity get \( S_0 e^{rT} \), buy the stock for \( K \) and deliver it to cover the obligation that comes with short selling.

A rational buyer will demand that \( K \) be such that

\[
S_0 e^{rT} - K \geq 0.
\]

Hence the only price at which both parties can agree is

\[
K = S_0 e^{rT}.
\]

If it were otherwise there would be an opportunity by one of the parties to make a riskless profit. This is called an arbitrage opportunity. Profit may not happen with certainty, but there should be no loss. The assumption that there does not exist such an opportunity at the market is called the arbitrage free assumption.

1.1 Combining Option to achieve a given portfolio

bear/bull spread example + Example 3.

**Put-Call Parity** For a call and put option with the same expiry and strike price, we have:

\[
P_t = C_t - S_t + K, \quad t \in [0, T].
\]
American Options the only difference is that they can be exercised at any time up until maturity. Since it gives the holder more options, it must be worth at least as much as the EU version. The holder now has to choose the optimal time to exercise the option. If too early, then you might miss making a bigger profit, if too late the optimum payout may be missed. This leads to stochastic optimum stopping time problems.

Asian Option Payoff depends on the average stock price over the period, not just the stock price at maturity. They have less volatility. Two possibilities for payoff:

\[(A - S_T)^+\]

or

\[(A - K)^+,\]

where the average stock price is defined as either

\[A = \frac{1}{T} \int_0^T S_t dt\]

or

\[A = \frac{1}{T} \int_0^T \ln S_t dt.\]

The latter one is used only because one gets analytical results.

Multi-reset options The strike price may be reset (equal to the current stock value) during the contracts duration:

1. EU style: the dates of change are fixed in advance. So every time one is allowed to reset \(S_T > K_{\text{current}}\), then one resets the strike price to be equal to the stock.

2. AM style: the dates of change are up to the holder to decide.

2 Rational pricing model

Time value of money We assume the availability of a risk-free interest rate \(r\) (what you get when you put a deposit at the bank), which we assume to be a constant. Thus

\[M(1 + r/n)^n\]

is the value of money \(M\) after 1 year compounded \(n\) times. For continuous compounding we have

\[Me^r\]

after 1 year and

\[Me^{rT}\]
after \( T \) years. Alternatively the value of \( M \$ \) in \( T \) years is now given by 
\[
M e^{-rT}.
\]
This is the discounted value of the money.

**Short Selling** Short selling - you sell a stock that you do not actually yet have. It is like a promise. I pay you now for stock to be delivered in a months time. It does not really matter if you have the stock now, all you need to have the stock only at the time of delivery. We assume there are no restriction on the amount of stock that can be sold short and the time of delivery.

**Arbitrage** Go back to the Forward contract example.

**Complete Market** A complete market is one where the pay-off of any financial contract can be replicated by setting up a portfolio containing only stocks and bonds. The value of this contract is the same as the value of setting up the portfolio at time 0.

**Dominant Portfolios** If \( \% \) return on portfolio \( A \) is larger than that of \( B \) with probability 1, then we say \( A \) dominates \( B \). Investors would prefer \( A \). We assume there \( \not\exists \) dominating portfolios. This is the same as no arbitrage opportunity. Let \( V_t \) be the value of a portfolio at time \( t \geq 0 \), then

1. \( V_t = 0 \Rightarrow V_\delta = 0 \) for all \( \delta > t \).
2. \( V_t > 0 \Rightarrow V_s = 0 \) for all \( s < t \).

### 2.1 Option inequalities

Let \( C(S_t, T - t, K) \) be the price behavior of an American Call and \( c(S_t, T - t, K) \) be the price behavior of an EU Call.

1. \( C(S_t, T - t, K) = c(S_t, T - t, K) = (S_T - K)^+ \).
2. \( C(S_t, T - t, K) \geq c(S_t, T - t, K) \geq 0 \)
3. If \( S_t = 0 \) then \( S_s = 0 \) for all \( s > t \) and we have that \( C(0, T - t, K) = 0 = c(0, T - t, K) \), hence never exercise call.
4. if \( T_1 > T_2 \), then \( C(S_t, T_1 - t, K) > C(S_t, T_2 - t, K) \) for all \( t \) and hence
\[
\frac{\partial}{\partial t} C(S_t, T - t, K) \leq 0.
\]
5. If \( K_1 > K_2 \), then \( C(S_t, T - t, K_1) \leq C(S_t, T - t, K_2) \).
6. Claim

\[ C(S_t, T - t, K) \geq c(S_t, T - t, K) \geq (S_t - Ke^{-r(T-t)})^+ \geq (S_t - K)^+ . \]

Proof: Set up the portfolio

\[ V_t = c(S_t, T - t, K) - S_t + Ke^{-r(T-t)}, \]

that is, we have an EU call option, we are short of one stock and have bonds. At time \( t = T \), it has value \( V_T = c(S_T, 0, K) - S_T + K \geq 0 \), then we know that \( V_s \geq 0 \) for \( \forall s < T \), that is,

\[ c(S_t, T - t, K) - S_t + Ke^{-r(T-t)} \geq 0, \quad 0 \leq t \leq T. \]

7. From the fact that \( C(S_t, T - t, K) \geq (S_t - K)^+ \), we conclude that it is never optimal to exercise an American Call option early. Thus it must be worth as much as an EU option. Therefore,

\[ C_t = c_t. \]

2.2 Put Options

1.

\[ p(S_t, T - t, K) = c(S_t, T - t, K) - S_t + Ke^{-r(T-t)}. \]

Set up the portfolio: \( V_t = p(S_t, T - t, K) - c(S_t, T - t, K) + S_t - Ke^{-r(T-t)} \), i.e., we have a EU put, we are short of a EU call, have a stock, and we are short of bonds. Then

\[ V_T = (K - S_T)^+ - (S_T - K)^+ + S_T - K = 0, \]

hence \( V_t = 0 \) for \( t < T \) and hence the result.

2.

\[ P(S_t, T - t, K) \geq (K - S_t)^+. \]

3.

\[ P(S_t, T - t, K) \geq p(S_t, T - t, K). \]

3 Binomial Asset Pricing Model

Discrete models as prelude to continuous models. Advantages of discrete models over continuous financial models include.

1. Simplicity and ease of computer implementation

2. Agrees with the discrete nature of the observable prices.
3.1 A single step model

The elements of the model are:

1. $\psi$ units of bonds (riskless borrowing and lending)
2. $\phi$ units of stock (one type of stock; risky)
3. no transaction costs
4. can trade unlimited amounts of stocks and bonds
5. time tick $\delta t$ is constant
6. 1$ now is worth $e^{rT}$ time units later.
7. assume we can hold fractions of stocks and bonds
8. $S_0 = s_0$ is known and fixed
9. $S_{\delta t} = S_T$ is a random variable, such that

$$S_T = \begin{cases} s_u, & \text{with probability } p \\ s_d, & \text{with probability } (1 - p). \end{cases}$$

We want to price an EU call option $c(S_t, T, K)$.

To price the option we set up a replicating portfolio of bonds and stocks at time $t = 0$ such that the value of the portfolio matches the payoff of the Call at $t = T$. The value of the portfolio at $t = 0$ will then give the price for the option. Any other price for the Call will give rise to an arbitrage opportunity. Suppose the value of the portfolio at $t = 0$ is

$$V_0 = \phi s_0 + \psi,$$

where $\psi$ and $\phi$ are yet unknown. We know however that at maturity $V_T = (S_T - K)^+ = C_T$. Hence,

$$\phi s_u + \psi e^{r\delta t} = S_{\delta t} - K = s_u - K$$
$$\phi s_d + \psi e^{r\delta t} = 0$$

Therefore,

$$\phi = \frac{S_{\delta t} - K}{s_u - s_d},$$
$$\psi = e^{-r\delta t} \frac{s_d (K - S_{\delta t})}{s_u - s_d}.$$

If we bought that much stock and bonds initially, then at maturity the portfolio of stocks and bonds will replicate the value of the call option. Any other price for
the call option will give rise to an arbitrage opportunity to make money. Thus the risk-neutral price for the option is the value of the replicating portfolio at time \( t = 0 \):

\[
V_0 = \frac{S_u - K}{s_u - s_d} s_0 + e^{-r\delta t} s_d (K - s_u) .
\]

**Example 3.1 (Roulette example)** Explain how the price of the call does not depend on the transition probabilities. Explain how to make a risk-free profit.

### 3.2 Multi-step model

**Assumptions:**

1. Two types of securities are traded: bonds (riskless) and stocks (risky).
2. the times of trade are discrete:

\[
0, \delta t, 2\delta t, \ldots, N\delta t = T
\]

3. \( r_t \) is the risk-free interest rate for the period \( t \in [(i - 1)\delta t, i\delta t) \). Thus for the whole period \$1 is worth

\[
B_T = e^{\Sigma_{i=1}^{N} r_i \delta t},
\]

because there are \( N \) periods of accrual.

4. Initially the stock price is \( S_0 \) and each time tick it can either move up to \( s_u \) or down \( s_d \) along a branch of a binary graph. Each branch can be traversed with a given probability specified by a set of transition probabilities \( P \). Thus in \( N \) steps we could be in \( 2^N \) possible states. We will assume that \( s_u = us_{\text{now}} \) and \( s_d = ds_{\text{now}} \), where

\[
0 < d < 1 + r < u.
\]

**Example 3.2 (6, page 31)** Compute expected value of stock price. Use both formulas to compute the value of the option. The simpler formula for the stock price is as follows. Let \( V_{\text{now}} - V_0 \) denote the current price of the contract, which we wish to price, and let \( V_{\text{up}} \) (respectively \( V_{\text{down}} \)) be the price of the contract when the stock price goes up (respectively down), then over one time tick:

\[
V_{\text{now}} = V_0 = e^{-r\delta t} (q V_u + (1 - q)V_d)
\]

\[
= e^{-rT} \mathbb{E}_Q[V_T] = B_T^{-1} \mathbb{E}_Q[V_T],
\]

\[
q = \frac{S_0 e^{rT} - S_d}{s_u - s_d}.
\]

**Note that** \( q \in [0, 1] \), **because** \( d < 1 + r < u \).
Example 3.3 (Example 7, page 35) The bank has sold a call option \( C(S_0 = 100, K = 100, T = 3\delta t) = 15\$ \) and wishes to hedge against the risk of losing from the option. Assume the stock prices move as follows: \( S_{\delta t} = 80 \), \( S_{2\delta t} = 100 \), \( S_{3\delta t} = 120 \), then the actions of the bank are:

1. At \( t = 0 \), the bank gets 15\$ and uses it to buy

\[
\phi = \frac{25 - 5}{120 - 80} = \frac{1}{2}
\]

shares of the underlying stock, and sells bonds (to borrow money) to get 35\$, so that

\[
\psi = -35\$.
\]

The portfolio initially is worth:

\[
V_0 = \frac{1}{2}100 - 35 = 15\$.
\]

So the bank uses the 15\$ from the call option to set up a portfolio worth 15\$.

2. At step \( t = \delta t \), the bank has to hold

\[
\phi = \frac{1}{4}
\]

worth of shares and

\[
\psi = -15\$
\]

worth of bonds. This is accomplished as follows. The bank sells 1/4 of the stock (gets 20\$) and repays part of the 35\$ debt to end up with a debt of 15\$ only. Again the portfolio is worth

\[
V_{\delta t} = \frac{1}{4} \times 80 - 15 = 20 - 15 = 5\$ = C_{\delta t}.
\]

3. At time \( t = 2\delta t \), the portfolio needs readjustment to

\[
\phi = \frac{1}{2}
\]

shares of stock and

\[
\psi = -40\$
\]

worth of bonds. That is, the bank sells 25\$ worth of bonds and uses to money to buy 1/4 of a stock worth 100\$ per share. Thus the bank holds the portfolio:

\[
V_{2\delta t} = \frac{1}{2} \times 100 - 40 = 10\$.
\]
4. At maturity, the person who holds the call option cashes in on it. He/She buys the stock from the bank for $K = 100$ and sells it for 120$, making a profit of 20$. To meet this obligation the bank does the following. It gets the $K = 100$ from that person and uses the money to repay the debt in bonds (so it now has 60$) and with the remaining 60$, it buys 1/2 a share of the underlying stock. Now the bank owns 1 unit of stock and just passes it on to cover the obligation from the Call option.

Of course, the bank does not make its huge profits from speculation, it makes it from transaction fees! The risk is borne by the the person who has bought the call option. That person could be hedging his own business risk or just speculating.

4 Derivation of Black-Sholes model

We divide $[0, T]$ into $N$ intervals $[(i-1)\delta t; i\delta t]$, $\delta t = T/N$ and assume
\[
\frac{d}{dt} \ln S_t = h_i, \quad S_t = S_{(i-1)\delta t}, t = (i-1)\delta t
\]
within each time interval. Then we have:
\[
S_{i\delta t} = S_{(i-1)\delta t}e^{h_i\delta t},
\]
or more generally:
\[
S_T = S_0e^{\delta t\sum_{k=1}^{N}h_k}.
\]
We also assume that $h_i$ is a Bernoulli r.v. such that
\[
P(h_i = \mu + \frac{\sigma}{\sqrt{t}}) = p
\]
and
\[
P(h_i = \mu - \frac{\sigma}{\sqrt{t}}) = 1 - p.
\]
Note that $\ln(S_T/S_0) = \delta t \sum_{k=1}^{N}h_k$. Hence $\frac{\ln(S_T/S_0) - N\delta t \mu}{\sigma\sqrt{N\delta t}} \sim N(0,1)$ as $N \uparrow \infty$ and $N\delta t = T$ under $\mathbb{P}$. Therefore,
\[
\ln(S_T/S_0) = T\mu + \sigma\sqrt{T}Z
\]
or in general
\[
S_t = S_0e^{t\mu + \sigma\sqrt{T}Z}
\]
under $\mathbb{P}$. This is the continuous time behavior of the stock price obtain as the time tick in the binomial tree model is shrunk to zero. What about the martingale measure $\mathbb{Q}$?
\[
q = \frac{S_t e^{\delta t} - S_t e^{\mu \delta t - \sigma \sqrt{\delta t}}}{S_t e^{\mu \delta t + \sigma \sqrt{\delta t}} - S_t e^{\mu \delta t - \sigma \sqrt{\delta t}}} \sim \frac{1}{2} - \frac{\mu - r + \sigma^2/2}{2\sigma} \sqrt{\delta t}.
\]
as \( \delta t \downarrow 0 \). Then under \( \mathbb{Q} \) we have
\[
\mathbb{E}_{\mathbb{Q}}[\ln(S_T/S_0)] = \delta t N(r - \sigma^2/2) = T(r - \sigma^2/2),
\]
\[
\text{Var}_{\mathbb{Q}}[\ln(S_T/S_0)] = T\sigma^2 + O(\delta t).
\]
Therefore, by the Central Limit Theorem, under \( \mathbb{Q} \), we have
\[
\ln(S_t/S_0) = t(r - \sigma^2/2) + \sqrt{t}\sigma Z, \quad Z \sim \mathcal{N}(0,1),
\]
or alternatively
\[
S_t = S_0 e^{t(r - \sigma^2/2) + \sqrt{t}\sigma Z}.
\]

**Remark 4.1 (Where is \( \mu \))** *The really surprising thing about the stock process under \( \mathbb{Q} \) is that it does not depend on \( \mu \).*

Now we are ready to apply the risk-neutral pricing formula and obtain the value of the contracts at any time \( t \). Suppose we wish to price an EU Call option \( C(S_t, T, K) \), then
\[
V_0 = B_0 \mathbb{E}_{\mathbb{Q}}[B_T^{-1}V_T | \mathcal{F}_0] = e^{-rT} \int_{\mathbb{R}} (S_0 e^{(r - \sigma^2/2) + \sqrt{T}\sigma z} - K)^+ \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = S_0 \Phi(d_1) - e^{-rT} K \Phi(d_2),
\]
where
\[
d_1 = \frac{\ln(S_0/K) + T(r + \sigma^2/2)}{\sigma \sqrt{T}},
\]
\[
d_2 = \frac{\ln(S_0/K) + T(r - \sigma^2/2)}{\sigma \sqrt{T}}.
\]

### 4.1 Hedging in continuous time

In the discrete-time models we could trade stocks and bonds at each time tick, thus setting up a replicating portfolio which follows the changes in the stock price (and the underlying option) exactly. In a continuous-time model, the stock/option price changes continuously over time, but we are not able to trade stocks and bonds continuously in order to adjust the replicating portfolio, thus we need local approximations based on Taylor’s theorem. (Continuous trading also ignores transaction costs!) In practice, to avoid large discrepancies between the portfolio and the value of the options, we can readjust the portfolio at small time steps \( \delta t \). There will always be a tiny difference but it will not be significant.

Suppose we have sold a Call option \( C_t = C(S_t, t) \) and wish to hedge against the risk/liability by purchasing a replicating portfolio \( V_t \). The total worth of our portfolio is thus given by:
\[
\Pi_t(S_t, t) = V_t - C_t.
\]
In order to hedge against the liability on the option in the time interval \([i\delta t; (i+1)\delta t]\) = \([t, t+\delta t]\), we purchase the following replicating portfolio (with the revenue from the option) at time \(t = i\delta t\):

\[
V_t = \xi_i F_t + \phi_i S_t + \psi_i,
\]

where

1. \(F_t\) is again an option on the same underlying stock, same strike price, but maturity date \(T\)
2. \(\xi_i\) is the number of options (units) that we have to hold.

So we have to choose the parameter set \((\xi_i, \phi_i, \psi_i)\) so that:

1. \(V_t = C_t\), i.e., the value of the portfolio matches the value of the option initially.
2. \(\frac{\partial}{\partial S_t} V_t = \frac{\partial}{\partial S_t} C_t\) - Delta hedge
3. \(\frac{\partial^2}{\partial S_t^2} V_t = \frac{\partial^2}{\partial S_t^2} C_t\), i.e., \(\Gamma\)–hedge. (matching the convexity of the option)

This gives three equations in three unknowns, so that it is reasonable to expect a unique solution for \((\xi_i, \phi_i, \psi_i)\). Thus, we need to hold

\[
\begin{align*}
\xi_i &= \frac{\partial^2 C_t}{\partial S_t^2} \frac{\partial^2 F_t}{\partial S_t^2} \\
\phi_i &= -\frac{\partial C_t}{\partial t} \frac{\partial F_t}{\partial t} \xi_i \\
\psi_i &= C_t - S_t \phi_i - F_t \xi_i
\end{align*}
\]

at time \(t\) so that we can (approximately) match the value of the option at the instant \(t + \delta t\) and continue to adjust the portfolio through self-financing.

**Example 4.1 (Black-Sholes)** Let \(C_t = C(S_t, T, K)\) be a EU Call, for which the Black-Sholes pricing formula holds at time \(t = i\delta t\), that is:

\[
C_t = S_t \Phi(d_1) - K e^{-r(T-t)} \Phi(d_2).
\]

Then we can verify that to \(\Delta\)–hedge in the time interval \([i\delta t; (i+1)\delta t]\), we need to hold \(\phi_i = \frac{\partial C_t}{\partial S_t} = \Phi(d_1)\) parts of the underlying stock at time \(t\). (This \(\phi_i\) is part of our replicating portfolio \(V_t = \phi_i S_t + \psi_i\) of stocks and bonds only.)
5 Probability Revision

Definition 5.1 (Probability Space \((\Omega, F, \mathbb{P})\)) Let \(\Omega\) be a set, then \(F\) is called a \(\sigma\)-algebra of subsets of \(\Omega\) if:

1. \(\Omega \in F\)
2. \(A \in F \Rightarrow \Omega \setminus A \in F\)
3. \(\{A_n\} \subset F \Rightarrow \bigcup_n A_n \in F\).

\(\mathbb{P}\) is called a probability measure on \(F\) if

1. \(\mathbb{P} : F \to [0, 1]\),
2. \(\mathbb{P}(\Omega) = 1\),
3. \(\{A_n\} \subset F\) and disjoint implies \(\mathbb{P}(\bigcup_n A_n) = \sum_n \mathbb{P}(A_n)\).

We often say that \(F\) is closed under countable unions and \(\mathbb{P}\) is countably additive. \(\Omega\) is often called a sample space (space of outcomes), the elements of \(F\) are called events, and \(\mathbb{P}(A)\) for \(A \in F\) is the probability or relative likelihood of event \(A\). Usually, \(\mathbb{F}\) is not the collection of all possible subsets of \(\Omega\).

Example 5.1 Toss 3 coins (heads/tails) \(\Omega = \{HHH, HHT, HTH, THH, \ldots, TTT\}\). \(A = \{HHT, HTH, THH\}\) is the event that one tail has occurred. \(\mathbb{P}(A) = 3/8\) (\(\Omega\) has 8 elements and \(A\) has 3). \(F\) here is the collection of all subsets of \(\Omega\) and \(A \in F\).

We often assume that the probability space \((\Omega, F, \mathbb{P})\) is complete, that is, if \(A \subseteq B\) (subsets of \(\Omega\)) with \(B \in F\) and \(\mathbb{P}(B) = 0\), then \(A \in F\) and \(\mathbb{P}(A) = 0\).

Example 5.2 \(\Omega = \mathbb{R}\) (real line), \(F = \mathcal{B}(\mathbb{R})\) the Borel subsets \((-\infty, x], x \in \mathbb{R}\) of \(\mathbb{R}\). Then for any probability measure \(\mathbb{P}\) on \(F\) we can define

\[ F(x) = \mathbb{P}((-\infty, x]), \quad x \in \mathbb{R}. \]

Then:

1. \(F\) is non-decreasing,
2. \(F(-\infty) = \lim_{x \downarrow -\infty} F(x) = 0, \quad F(\infty) = \lim_{x \uparrow \infty} F(x) = 1\)
3. \(F\) is continuous from the right and has limits on the left at each \(x \in \mathbb{R}\). So

\[ \lim_{y \uparrow x} F(y) = F(x), \]

and

\[ \lim_{y \downarrow x} F(y) \]

exists.
Such an \( F \) is called a distribution function.

**Definition 5.2 (Random Variables on \((\Omega, \mathcal{F}, P)\))** \( X \) is a random variable on \( \Omega \) if \( X : \Omega \to \mathbb{R} \) and the set \( \{ \omega \in \Omega | X(\omega) \leq x \} \in \mathcal{F} \) for all \( x \in \mathbb{R} \). We say that \( X \) is \( \mathcal{F} \) measurable function on \( \Omega \). The distribution function of \( X \) is defined by

\[
F_X(x) = \mathbb{P}(\{ \omega \in \Omega | X(\omega) \leq x \}).
\]

If a distribution function \( F_X \) is differentiable and \( F'_X(x) = f_X(x) \), then \( f \) is called the probability density function (pdf) of \( X \).

An important theorem relating to the distribution of a r.v.

**Theorem 5.1** A r.v. \( X \) has pdf \( f \) iff

\[
\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(z) f(z) dz
\]

for any bounded continuous function \( g \).

**Definition 5.3 (Independence)** If \( A, B \in \mathcal{F} \), \( A \) and \( B \) are independent events if \( \mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B) \). If \( X, Y \) are r.v. on \((\Omega, \mathcal{F}, \mathbb{P})\), then \( X, Y \) are independent if

\[
F_{XY}(x,y) = \mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x) \mathbb{P}(Y \leq y).
\]

Here \( F_{XY} \) denotes the joint distribution of \( X \) and \( Y \).

**Lemma 5.1 (Independence)** \( X, Y \) are independent iff one of the following holds:

1. \( f_{XY}(x,y) = f_X(x)f_Y(y) \),

where \( f_X \) and \( f_Y \) are the pdfs of \( X \) and \( Y \), and \( f_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y) \) is the joint pdf of \( X, Y \).

2. For all bounded and continuous functions \( g, h : \mathbb{R} \to \mathbb{R} \)

\[
\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)] \mathbb{E}[h(Y)].
\]

**Definition 5.4 (Conditional Expectation)** Let \((\Omega, \mathbb{P}, \mathcal{F})\) be a probability space and \( \mathcal{G} \subseteq \mathcal{F} \) an event space, \( Y \) a random variable on \((\Omega, \mathbb{P}, \mathcal{F})\) (so \( Y \) is \( \mathcal{F} \) measurable)

Then \( \mathbb{E}[Y | \mathcal{G}] \) is defined through the properties:

1. \( \mathbb{E}[Y | \mathcal{G}] \) is \( \mathcal{G} \)-measurable,
2. \( \int_A \mathbb{E}[Y \mid G]d\mathbb{P} = \int_A Y d\mathbb{P} \) for all \( A \in \mathcal{G} \). In other words, 
\[ \mathbb{E}_P[YI(A)] = \mathbb{E}_P[\mathbb{E}_P[Y \mid G]I(A)] \quad \forall A \in \mathcal{G}, \]
and \( I \) is the indicator function.

A special case of this general definition includes \( \mathbb{E}[Y \mid X] \) defined as 
\[ \mathbb{E}[Y \mid X] = \mathbb{E}[Y \mid \mathcal{G}], \quad \mathcal{G} = \sigma(X), \]
i.e., \( \mathcal{G} \) is all the information generated by (relevant to) \( X \). In the case where the joint density of \( X, Y \) exists, we have
\[ \mathbb{E}[X \mid Y] = \int_{\mathbb{R}} x f(x \mid y) dx = \int_{\mathbb{R}} x \frac{f(x, y)}{f(y)} dx. \]

Properties of the conditional expectation include:

1. If \( \mathcal{G} \equiv \{\emptyset, \Omega\} \), then 
   \[ \mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X], \]
   because \( \mathcal{G} \) does not contain any extra information about \( X \), which we already do not know. We already know that \( X \) exists on the sample space \( \Omega \).

2. If \( X \) is \( \mathcal{G} \) measurable, then 
   \[ \mathbb{E}[XY \mid \mathcal{G}] = X \mathbb{E}[Y \mid \mathcal{G}]. \]
   Thus if \( \mathcal{G} \) contains all the information about \( X \), then given \( \mathcal{G} \), \( X \) is known, can be treated as a constant and can be moved outside the expectation sign.

3. If \( \mathcal{G}_1 \subset \mathcal{G}_2 \) then 
   \[ \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1] = \mathbb{E}[X \mid \mathcal{G}_1]. \]
   If \( \mathcal{G}_1 \equiv \{\emptyset, \Omega\} \), then we get the tower property:
   \[ \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_2]] = \mathbb{E}[X]. \]

4. If \( \sigma(X) \) and \( \mathcal{G} \) are independent, that is, \( \mathcal{G} \) does not contain any information related to \( X \), then 
   \[ \mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]. \]

5. If \( g(\cdot) \) is a convex function on \( I \) and \( X \) has range \( I \), then 
   \[ g(\mathbb{E}[X \mid \mathcal{G}]) \leq \mathbb{E}[g(X) \mid \mathcal{G}]. \]
6. (Monotone Convergence Property) If $X_n \geq 0$ and $X_n \uparrow X$ (almost sure convergence, that is for every $\varepsilon > 0$, $\lim_{n \uparrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$) with $\mathbb{E}[|X|] < \infty$, then
$$
\lim_{n \uparrow \infty} \mathbb{E}[X_n \mid \mathcal{G}] = \mathbb{E}[X \mid \mathcal{G}].
$$
In words, if a sequence $\{X_n\}$ of integrable non-negative random variables converges in probability to $X$, then the sequence of conditional expectations $\{\mathbb{E}[X_n \mid \mathcal{G}]\}$ converges to the conditional expectation $\mathbb{E}[X \mid \mathcal{G}]$.

7. (Fatou’s lemma) If $X_n \geq 0$, then
$$
\mathbb{E}[\liminf_n X_n \mid \mathcal{G}] \leq \liminf_n \mathbb{E}[X_n \mid \mathcal{G}].
$$

8. (Dominated Convergence Property) If $X_n \uparrow X$ almost surely and $|X_n| < Y$ with $\mathbb{E}[Y] < \infty$, then
$$
\lim_{n \uparrow \infty} \mathbb{E}[X_n \mid \mathcal{G}] = \mathbb{E}[X \mid \mathcal{G}].
$$
6 Risk Neutral Pricing and Martingales

Suppose we are given the continuous time stochastic process (that is, a collection of random variables) \( (S_t, 0 \leq t \leq T) \) (in discrete time we have \( \{S_0, S_1, \ldots, S_T\} \)). Denote \( \Omega \) to be the set of possible trajectories/outcomes for this process (the set of all possible stock prices in the interval \([0, T]\)).

Let \( F_t = \sigma(S_u, u \in \{0, 1, \ldots, t\}) \) be the \( \sigma \)-algebra generated by the process up to time \( t \). \( F_t \) represents all the information, available to an observer, generated by the process up to time \( t \). In our context, \( F_t \) is the information available to investors at time \( t \) - the stock prices before and at time \( t \). A \( \sigma \)-algebra defines the set of events that can be measured.

**Example 6.1 (Field/algebra of events)** Consider a model of trading a stock for times \( t = 1, 2 \), where at each time tick the stock can go up by a factor \( u \) and down by a factor of \( d \). The set of all possible outcomes is described via the sample space \( \Omega = \{\omega_1 = (u, u), \omega_2 = (u, d), \omega_3 = (d, u), \omega_4 = (d, d)\} \). Initially all we can measure is \( F_0 = \{\emptyset, \Omega\} \). Now suppose that at \( t = 1 \) the stock goes up. Let \( A = \{\omega_1, \omega_2\} \), then the information generated by the process is \( F_1 = \{\emptyset, \Omega, A, A^c\} \).

**Definition 6.1 (Definition of filtration.)** The collection of fields/\( \sigma \)-algebras \( F = (F_t, t \geq 0) \), where \( F_t \subseteq F_s \Leftrightarrow t \leq s \), is called a filtration. It specifies how information about the underlying process is revealed in time. It keeps track of the information about the process as time progresses.

The property \( F_t \subseteq F_s \Leftrightarrow t \leq s \) corresponds to the fact that, once the information is revealed, it is not forgotten.

We have the following important definition.

**Definition 6.2 (Measurability)** If based on the information contained in \( F_t \), is it possible to determine whether a given event \( A \) has occurred, then we say that \( A \) is \( F_t \) measurable and write \( A \in F_t \).

**Example 6.2** Let \( A = \{S_t < 5, \forall t \leq 18\} \) be the event that the stock price is below 5$ up to time \( t = 18 \). Then \( A \in F_{18} \), but \( A \notin F_{16} \). Similarly the event \( A = \{S_{10} > 5\} \) is \( F_t \) measurable only when \( t \geq 10 \).

**Definition 6.3 (Adaption)** The stochastic process \( \{S_t, 0 \leq t \leq T\} \) is said to be adapted to the filtration \( \mathbb{F} \) if \( S_t \in F_t, \forall t \geq 0 \) (\( S_t \) is \( F_t \) measurable for all time), that is, \( F_t \) contains all the information generated by \( S_t \) up to time \( t \) (and perhaps some extra information as well).

**Example 6.3** Suppose \( (F_t, t \geq 0) \) is the natural filtration associated with the stock process \( (S_t, t \geq 0) \), then the process \( M_t = \max_{s \leq t} S_s \in F_t \) for all \( t \), and hence is adapted to \( (F_t, t \geq 0) \). In contrast, \( M_t = \max_{s \leq t+1} S_s \in F_t \) is not \( F_t \) measurable and hence not adapted to \( (F_t, t \geq 0) \).
Definition 6.4 (Previsibility) The continuous time stochastic process is said to be previsible/predictable with respect to the filtration $\mathbb{F}$ if it is adapted and continuous (that is, give the current state of the process it is possible to predict its behavior in the next instant in an infinitesimal neighborhood.) A discrete time process is called previsible with respect to $\mathcal{F}_t$ if it is $\mathcal{F}_{t-1}$ measurable.

Definition 6.5 (Discrete-time martingale) A discrete-time stochastic process $\{S_t\}_{t=0}^T$ is called a martingale with respect to the filtration $(\mathcal{F}_t, t \geq 0)$ and the probability measure $\mathbb{P}$ if

1. $S_t \in \mathcal{F}_t$ for all $t$.
2. $\mathbb{E}_\mathbb{P}[|S_t|] < \infty$ for all $t$.
3. $\mathbb{E}_\mathbb{P}[S_j \mid S_0, S_1, \ldots, S_i] = S_i$, for all $j \geq i$.

Definition 6.6 (Continuous-time Martingale) A stochastic process $(S_t, t \geq 0)$ is called a supermartingale (submartingale) with respect to the filtration $\mathbb{F}$ and the probability measure $\mathbb{P}$ if

1. $S_t \in \mathcal{F}_t$, $\forall t \geq 0$, that is, $S_t$ is $\mathcal{F}_t$-measurable for all $t$.
2. $\mathbb{E}_\mathbb{P}[|S_t|] < \infty$ for all $t$.
3. $\mathbb{E}_\mathbb{P}[S_t \mid \mathcal{F}_s] \leq S_s$, for all $t \geq s$ ($\mathbb{E}_\mathbb{P}[S_t \mid \mathcal{F}_s] \geq S_s$, for all $t \geq s$).

A process which is both a supermartingale and submartingale is called a martingale.

The process is usually referred to as an $(\mathcal{F}_t, \mathbb{P})$ supermartingale (submartingale) for short.

Example 6.4 Let $X_1, X_2, \ldots$ be a sequence of iid integrable random variables with zero mean (under $\mathbb{P}$). Prove that $S_n = \sum_{i=1}^n X_i$ is a martingale with respect to $\mathbb{P}$ and the natural filtration defined by the process.

Example 10, page 46 in notes.

Example 6.5 (Doob’s martingale) Let $X, Y_1, Y_2, \ldots$ be arbitrary random variables such that $\mathbb{E}[|X|] < \infty$, and let

$$Z_n = \mathbb{E}[X \mid Y_1, \ldots, Y_n].$$

Then $\{Z_n, n = 1, 2, 3, \ldots\}$ is a martingale.
Theorem 6.1 (Martingale property of an Arbitrage-free market) A market is arbitrage-free iff there exists a unique probability measure \( \mathbb{Q} \) under which the discounted stock price \( \tilde{S}_t = B_t^{-1} S_t \) is a martingale. In such cases, the risk-neutral price at any time \( 0 \leq t < T \) of an European style contract with payoff \( V_T \) is given by

\[
V_t = B_t \mathbb{E}_{\mathbb{Q}}[B_s^{-1}V_s | \mathcal{F}_t],
\]

where \( (\mathcal{F}_t, t \geq 0) \) is the filtration generated by the stock process.

Example 8/ page 37

Example 6.6 (Example 9/page 38)

7 Brownian motion

Definition 7.1 (Standard Brownian Motion) The process \( (B_t, t \geq 0) \) is called standard Brownian motion (or Wiener process, in which case we use \( W_t \), instead of \( B_t \)) if

1. \( B_0 = 0 \);
2. \( B \) has independent increments, that is, for any \( t_1 < t_2 < t_3 < t_4 \), then

\[
B_{t_4} - B_{t_3} \quad \text{and} \quad B_{t_2} - B_{t_1}
\]

are independent;
3. if \( t \geq s \geq 0 \), then

\[
B_t - B_s \sim \mathcal{N}(0, t - s).
\]
4. \( B_t \) is a continuous function if \( t \).

Some properties include:

1. \( \mathbb{E}[B_t B_s] = \min(t, s) \) and \( (B_t, B_s) \) has a bivariate normal distribution; This crucial property is equivalent to defining property 2. above and can be used as an alternative definition. That is, given that \( \mathbb{E}[B_t B_s] = \min(t, s) \) and \( (B_t, B_s) \) has a bivariate normal distribution, property 2 follows.
2. The sample paths of BM are continuous, yet nowhere differentiable. This means than no matter how closely we look into a smaller and smaller neighborhood of a sample path, it will never appear smooth (the jaggedness will persist on whatever scale we examine the path).
3. The sample paths of BM on $[0,T]$ do not have bounded variation, that is:

$$\lim_{\delta_n \to 0} \sum_{i=1}^{n} |B_{t_i} - B_{t_{i-1}}| = \infty,$$

where $0 = t_0 < t_1 < \cdots < t_n = T$ is a partition of $[0,T]$ and $\delta_n = \max_i \{t_i - t_{i-1}\}$. Thus the 'length' of a sample path of BM is infinite.

4. $\mathbb{E}[(B_t - B_s)^4] = 3(t-s)^2$.

5. $B_{ts}$, $s > 0$ has the same distribution as $\sqrt{s}B_t$, namely $\mathcal{N}(0,t)$.  

6. If $t > s$, then $B_{t-s}$ has the same distribution as $B_t - B_s$, namely $\mathcal{N}(0,t-s)$.

Construction of BM. Let $\{Z_n\}$ be iid $\mathcal{N}(0,1)$ sequence and let $\{h_n(x)\}$ be a sequence that forms a complete orthonormal basis for $L^2(\mathbb{R})$. In other words, any $f \in L^2(\mathbb{R})$ can be expanded in the form

$$f(x) = \sum_{k=1}^{\infty} c_n h_n(x),$$

where $c_m = \int_{\mathbb{R}} f(x)h_m(x)dx$ for $m = 1, 2, \ldots$. I like to use the Hermite functions for $h_n$ (defined in the Appendix). Then BM can be constructed in the following way:

$$B_t = \sum_{n=1}^{\infty} Z_n \int_{0}^{t} h_n(x)dx, \quad Z_1, Z_2, \ldots \sim \text{iid } \mathcal{N}(0,1). \quad (1)$$

To verify that the above sum defines BM, let

$$B_m(t) = \sum_{n=1}^{m} Z_n \int_{0}^{t} h_n(x)dx,$$

then:

1. For a fixed time $t$, $B_m(t)$ is Gaussian;

2. $B_m(t) - B_k(t) = \sum_{n=k+1}^{m} Z_n \int_{0}^{t} h_n(x)dx$, $m > k$

is also Gaussian;

so that

$$\mathbb{E}[(B_m - B_k)^2] = \mathbb{E} \sum_{i=k+1}^{m} \sum_{j=k+1}^{m} Z_i Z_j \int_{0}^{t} h_i(x)dx \int_{0}^{t} h_j(x)dx$$

$$= \sum_{i=k+1}^{m} \left( \int_{0}^{t} h_i(x)dx \right)^2 \to 0, \quad k, m \to \infty,$$
since $E[Z_n Z_m] = \delta_{nm}$ and (verify!) \( \sum_{k=1}^{\infty} \left( \int_0^t h_i(x) \, dx \right)^2 = t < \infty \). Therefore,

$$B_t = \lim_{m \to \infty} B_m(t)$$

exists for each $t$ and is Gaussian for each $t$. We can now verify that:

1. $B(0) = B_0 = 0$.

2. Let $t \geq s$, then (verify!)

$$E[B_t B_s] = \sum_{i=1}^{\infty} \int_0^t h_i(x) \, dx \int_0^s h_i(x) \, dx = s.$$ 

3. $E[B(t)] = 0$ and

$$E[(B_{s+\delta t} - B_s)^2] = \sum_{k=1}^{\infty} \left( \int_{\mathbb{R}} h_k(x) \, dx \right)^2 = \delta t.$$ 

Thus, according to the definition of BM, (1) defines a BM. In fact, a we can construct an infinite number of independent BMs on $(\Omega, \mathcal{F}, \mathbb{P})$ as follows.

- $B_1^t = \sum_{k=1}^{\infty} Z^1_k \int_0^t h_i(x) \, dx, \quad Z^1_1 = Z_1, Z^1_2 = Z_3, Z^1_3 = Z_6, Z^1_4 = Z_{10}, \ldots$
- $B_2^t = \sum_{k=1}^{\infty} Z^2_k \int_0^t h_i(x) \, dx, \quad Z^2_1 = Z_2, Z^2_2 = Z_5, Z^2_3 = Z_9, Z^2_4 = Z_{11}, \ldots$
- $B_3^t = \cdots$ and so on.

### 7.1 Properties of Brownian Motion

Let $\mathcal{F}_t = \sigma\{B_s, 0 \leq s \leq t\}$ be the $\sigma-$algebra generated by the Brownian motion process. Recall that $\mathcal{F}_t$ represents all the information available to an observer of the stochastic process. An event $A \in \mathcal{F}_t$ (is measurable with respect to $\mathcal{F}_t$) if based on the information contained in $\mathcal{F}_t$, it is possible to determine if event $A$ has occurred or not. We then have the following Lemma.

**Lemma 7.1**

1. $B_t$ is an $(\mathcal{F}_t, \mathbb{P})$ martingale.

2. $B^2_t - t$ is an $(\mathcal{F}_t, \mathbb{P})$ martingale.

3. $e^{u B_t - u^2 t/2}$ is an $(\mathcal{F}_t, \mathbb{P})$ martingale.

We can demonstrate 1. as follows.
1. Let $t > s$, then $\mathbb{E}[B_t \mid F_s] = \mathbb{E}[(B_t - B_s) + B_s \mid F_s] = \mathbb{E}[(B_t - B_s) \mid F_s] + \mathbb{E}[B_s \mid F_s] = 0 + B_s = B_s$, as $B_t - B_s$ is independent of $F_s$. Note that all expectations must be taken with respect to $\mathbb{P}$.

2. $B_t$ is $F_t$ measurable for each $t \geq 0$, i.e., the process $(B_t, t \geq 0)$ is adapted to $F = (F_t, t \geq 0)$.

3. $\mathbb{E}_P[|B_t|] \leq \sqrt{\mathbb{E}_P[B_t^2]} = \sqrt{t} < \infty$

We can demonstrate 2. as follows.

1. Let $t > s$, then $\mathbb{E}[B_t^2 - t \mid F_s] = \mathbb{E}[(B_t - B_s + B_s)^2 \mid F_s] - t = \mathbb{E}[(B_t - B_s)^2 \mid F_s] + 2\mathbb{E}[B_s(B_t - B_s) \mid F_s] + \mathbb{E}[B_s^2 \mid F_s] - t = (t - s) + 2B_s\mathbb{E}[B_t - B_s] + B_s^2 - t = B_s^2 - s$. Note that all expectations must be taken with respect to $\mathbb{P}$.

2. $B_t^2 - t$ is $F_t$ measurable for each $t \geq 0$, i.e., the process is adapted to the filtration $F = (F_t, t \geq 0)$.

3. $\mathbb{E}_P[|B_t^2 - t|] \leq \mathbb{E}_P[B_t^2] + t = 2t < \infty$.

Try proving 3. by yourself.

Concerning 2., we have the following converse result.

**Theorem 7.1 (Levy’s Characterization theorem)** A continuous process $\{X_t, t \geq 0\}$ with $F_t = \sigma(X_u, 0 \leq u \leq t)$ is a $(F_t, \mathbb{P})$ Brownian motion iff:

1. $\{X_t, t \geq 0\}$ is an $(F_t, \mathbb{P})$ martingale.

2. $\{X_t^2 - t, t \geq 0\}$ is an $(F_t, \mathbb{P})$ martingale.

### 8 Ito’s Integrals (1947) in $L_2$

We want to define the integral

$$\int_0^T f(t) \, dB_t,$$

where $f$ could be random. We first define the meaning of the integral for simple functions $f$ and the move on to more general functions.

1. If $f(t) = I\{a \leq t \leq b\} = I_{[a,b]}$, and $0 \leq a \leq b \leq T$, then

$$\int_0^T f(t) \, dB_t = B(b) - B(a).$$
2. If
\[
f(t) = \sum_{k=0}^{m-1} a_k I(t_k, t_{k+1}),
\]
where \(0 = t_0 < \ldots < t_m = T\), then
\[
\int_0^T f(t) \, dB_t = \sum_{k=0}^{m-1} a_k [B(t_k + 1) - B(t_k)]. \tag{2}
\]
Note that \(f(t_k) = a_k, \ k = 0, 1, 2, \ldots, m - 1\). We can allow \(\{a_k\}\) to be random, but each \(a_k\) needs to be \(\mathcal{F}_{t_k}\) measurable, where \(\mathcal{F}_t = \sigma\{B_u : 0 \leq u \leq t\}\). We denote this set of simple functions by \(\mathcal{S}\).

Properties of the integral when \(f \in \mathcal{S}\):

1. \(\mathbb{E} \int_0^T f(t) \, dB_t = 0\).

Proof:
\[
\mathbb{E} \int_0^T f(t) \, dB_t = \sum_{k=0}^{m-1} \mathbb{E}[a_k [B_{t_{k+1}} - B_{t_k}]]
\]
\[
= \sum_{k=0}^{m-1} \mathbb{E}[\mathbb{E}[a_k (B_{t_{k+1}} - B_{t_k}) | \mathcal{F}_{t_k}]]
\]
\[
= \sum_{k=0}^{m-1} \mathbb{E}[a_k \mathbb{E}[(B_{t_{k+1}} - B_{t_k}) | \mathcal{F}_{t_k}]]
\]
\[
= \sum_{k=0}^{m-1} \mathbb{E}[f(t_k) 0] = 0.
\]

2. \(\mathbb{E} \left( \int_0^T f(s) \, dB_s \right)^2 = \mathbb{E} \int_0^T (f(s))^2 \, ds\)

Proof:
\[
\mathbb{E} \left( \int_0^T f(s) \, dB_s \right)^2 = \sum_{k=0}^{m-1} \sum_{l=0}^{m-1} \mathbb{E} \left[ a_k a_l (B_{t_{k+1}} - B_{t_k})(B_{t_{l+1}} - B_{t_l}) \right]
\]
\[
= \sum_{k>l} \mathbb{E} [\cdots] + \sum_{k=l} \mathbb{E} [\cdots] + \sum_{k<l} \mathbb{E} [\cdots].
\]
Now note that for the case \(k > l\), we have:
\[
\sum_{k>l} \mathbb{E} \left[ a_k a_l (B_{t_{k+1}} - B_{t_k})(B_{t_{l+1}} - B_{t_l}) \right] = \sum_{k>l} \mathbb{E} \left[ \mathbb{E}[a_k a_l (B_{t_{k+1}} - B_{t_k})(B_{t_{l+1}} - B_{t_l}) | \mathcal{F}_{t_k}] \right]
\]
\[
= \sum_{k>l} \mathbb{E} \left[ a_k a_l (B_{t_{l+1}} - B_{t_l}) \mathbb{E}[(B_{t_{k+1}} - B_{t_k}) | \mathcal{F}_{t_k}] \right]
\]
\[
= \sum_{k>l} \mathbb{E} \left[ a_k a_l (B_{t_{l+1}} - B_{t_l}) 0 \right] = 0.
\]
Similarly, we can show that \( \sum_{k<l} E \left[ \cdots \right] \) is also zero. We thus obtain:

\[
E \left( \int_0^T f(s) dB_s \right)^2 = \sum_{k=l} m^{-1} E \left[ a_k a_l (B_{t_{k+1}} - B_{t_k})(B_{t_{l+1}} - B_{t_l}) \right]
\]

\[
= \sum_{k=0}^{m-1} E \left[ f^2(t_k)(B_{t_{k+1}} - B_{t_k})^2 \right]
\]

\[
= \sum_{k=0}^{m-1} E \left[ E[f^2(t_k)(B_{t_{k+1}} - B_{t_k})^2 | \mathcal{F}_{t_k}] \right]
\]

\[
= \sum_{k=0}^{m-1} E \left[ f^2(t_k)E[(B_{t_{k+1}} - B_{t_k})^2 | \mathcal{F}_{t_k}] \right]
\]

\[
= \sum_{k=0}^{m-1} f^2(t_k)(t_{k+1} - t_k)
\]

\[
= E \sum_{k=0}^{m-1} f^2(t_k)(t_{k+1} - t_k) = E \int_0^T f^2(t) \, dt.
\]

3. If both \( f, g \in \mathcal{S} \) and \( E \int_0^T f^2(t) \, dt < \infty, E \int_0^T g^2(t) \, dt < \infty \), then

\[
E \int_0^T f(t) \, dB_t \int_0^T g(s) \, dB_s = E \int_0^T f(t)g(t) \, dt.
\]

Proof: From the previous result we have:

\[
E \left( \int_0^T (f(t) + g(t)) \, dB_t \right)^2 = E \int_0^T (f(t) + g(t))^2 \, dt.
\]

The result follows by expanding both sides and canceling terms on both sides of the equation.

4. Let \( E \int_0^T f^2(t) \, dt < \infty \) and define

\[
Y_t = \int_0^t f(u) \, dB_u, \quad 0 \leq t \leq T,
\]

then \( \{Y_t, t \geq 0\} \) is an \((\mathcal{F}_t^B, \mathbb{P})\) martingale. Here, of course, \( B_t \) is SBM under \( \mathbb{P} \) and \( \mathcal{F}_t^B \) is the \( \sigma \)-algebra generated by the BM.

(a) \( E[|Y_t|] \leq \sqrt{E[Y_t^2]} = \sqrt{E \int_0^T f^2(t) \, dt} < \infty \)

(b) \( Y_t \) is clearly \( \mathcal{F}_t^B \) measurable for each \( t \in [0, T] \).
(c) For \( t > s \),
\[
\mathbb{E}[Y_t \mid \mathcal{F}_s^B] = \mathbb{E} \left[ Y_s + \int_s^t f(u) \, dB_u \bigg| \mathcal{F}_s^B \right] \\
= Y_s + \mathbb{E} \left[ \int_s^t f(u) \, dB_u \bigg| \mathcal{F}_s^B \right].
\]

Thus we have to show that \( \mathbb{E} \left[ \int_t^s f(u) \, dB_u \bigg| \mathcal{F}_s^B \right] = 0 \). Let \( 0 = t_0 < t_1 < \cdots < t_k = s < \cdots < t_n = t \), we have:
\[
\mathbb{E} \left[ \int_t^s f(u) \, dB_u \bigg| \mathcal{F}_s^B \right] = \sum_{i=k}^{n-1} \mathbb{E} \left[ a_i (B_{t_{i+1}} - B_{t_i}) \bigg| \mathcal{F}_s^B \right] \\
= \sum_{i=k}^{n-1} \mathbb{E} \left[ \mathbb{E}\left[a_i (B_{t_{i+1}} - B_{t_i}) \bigg| \mathcal{F}_{t_i}^B \right] \bigg| \mathcal{F}_s^B \right], \text{ since } \mathcal{F}_s^B \subseteq \mathcal{F}_{t_i}^B, \forall i \\
= \sum_{i=k}^{n-1} \mathbb{E} \left[ f(t_i) \, \mathbb{E}\left[B_{t_{i+1}} - B_{t_i} \bigg| \mathcal{F}_{t_i}^B \right] \bigg| \mathcal{F}_s^B \right] = 0.
\]

Thus \( Y_t \) is a martingale. In fact, since
\[
\max_{t \in [0,T]} \mathbb{E}[Y_t^2] < \infty,
\]
it is a square integrable martingale.

After considering the integral for simple functions, we are ready to define it for more general class of functions.

**Definition 8.1** Let \( \mathcal{V}[0,T] \) be the class of functions \([0,T] \times \Omega \rightarrow \mathbb{R}\) on \((\Omega, \mathcal{F}, \mathbb{P})\) so that:

1. \((t, \omega) \rightarrow f(t, \omega)\) is \( \mathcal{B} \times \mathcal{F} \)-measurable, where \( \mathcal{B} \) denotes the Borel \( \sigma \)-algebra on \([0,T] \).

2. \( f(t, \omega) \) is \( \mathcal{F}_t^B \) adapted.

3. \( \mathbb{E}_\mathbb{P} \left[ \int_0^T f^2(t, \omega) \, dt \right] < \infty. \)

We then have:

**Lemma 8.1** If \( f \in \mathcal{V}[0,T] \), then there exists a sequence \( \{f_n\} \in \mathcal{I} \cap \mathcal{V}[0,T] \) so that
\[
\lim_{n \to \infty} \mathbb{E} \left[ \int_0^T (f_n(t) - f(t))^2 \, dt \right] = 0
\]
We can now define $\int_0^T f(t) dB_t$ for $f \in \mathcal{V}[0,T]$. Let $f_n \rightarrow f$ as in the Lemma and set

$$I_n = \int_0^T f_n(t) dB_t,$$

then

$$\mathbb{E}[(I_n - I_m)^2] = \mathbb{E} \left[ \left( \int_0^T (f_n(t) - f_m(t)) dB_t \right)^2 \right]$$

$$= \mathbb{E} \left[ \int_0^T (f_n(t) - f_m(t))^2 dt \right]$$

$$= \mathbb{E} \left[ \int_0^T [f_n(t) - f(t) + (f(t) - f_m(t))]^2 dt \right]$$

$$\leq 2 \mathbb{E} \left[ \int_0^T [f_n(t) - f(t)]^2 dt \right] + 2 \mathbb{E} \left[ \int_0^T [f_m(t) - f(t)]^2 dt \right] \rightarrow 0,$$

where both $n$ and $m$ tend to infinity and we have used the inequality $(a + b)^2 \leq 2(a^2 + b^2)$. Therefore, there exists $I \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ so that

$$\lim_{n \rightarrow \infty} \mathbb{E} [(I_n - I)^2] = 0.$$

In other words, the random variable $I_n$ converges in $L_2$ norm to $I$. This is one of the many possible modes of convergence for random variables. Here $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is the collection of random variables $X$ on $(\Omega, \mathcal{F}, \mathbb{P})$ so that $\mathbb{E}[|X|^2] < \infty$. It is a Hilbert space, just like $L^2(\mathbb{R})$. Note also that convergence of $I_n$ in norm implies convergence in probability, that is:

$$\lim_{n \rightarrow \infty} \mathbb{P}(|I_n - I| > \varepsilon) = 0,$$

for some $\varepsilon > 0$.

**Example 8.1 (Integral of $B_t$ with respect to itself)** Define

$$I_n = \sum_{k=0}^{n-1} B_{t_k} (B_{t_{k+1}} - B_{t_k}) = \int_0^T f_n(t) dB_t,$$

where

$$f_n(t) = \sum_{k=0}^{n-1} B_{t_k} I_{[t_k, t_{k+1}]} \in \mathcal{H}$$

for some partition

$$0 = t_0 < t_1 < \ldots < t_n = T$$

of the interval $[0,T]$ with $r = \max_k \{ t_{k+1} - t_k , k = 1, \ldots, n - 1 \}$. We insist the partition is such that $\lim_{n \rightarrow \infty} r = 0$. Observe that $\lim_{n \rightarrow \infty} f_n(t) = f(t) = B_t$, so
that we know there exists a limit of $I_n$, denoted $I = \int_0^T B_t \, dB_t$, where convergence to this limit is in the sense of

$$\mathbb{E} \left[ (I_n - I)^2 \right] \to 0, \quad \text{as } n \to \infty.$$ 

To prove this, consider:

$$2I_n = \sum_{k=0}^{n-1} 2B_{t_k} (B_{t_{k+1}} - B_{t_k}) \quad \text{"completion of square"}$$

$$= \sum_{k=0}^{n-1} B_{t_k}^2 + 2B_{t_k} (B_{t_{k+1}} - B_{t_k}) + (B_{t_{k+1}} - B_{t_k})^2 - B_{t_k}^2 - (B_{t_{k+1}} - B_{t_k})^2$$

$$= \sum_{k=0}^{n-1} B_{t_{k+1}}^2 - B_{t_k}^2 - (B_{t_{k+1}} - B_{t_k})^2$$

$$= B_T^2 - \sum_{k=0}^{n-1} (B_{t_{k+1}} - B_{t_k})^2 = B_T^2 - \sum_{k=0}^{n-1} (t_{k+1} - t_k)Z_k^2,$$

where $Z_1, Z_2, \ldots \sim \text{iid } \mathcal{N}(0, 1)$. Hence,

$$\mathbb{E} \left[ I_n - \frac{1}{2} B_T^2 \right] = -\frac{1}{2} T.$$

Moreover,

$$\text{Var} \left[ I_n - \frac{1}{2} B_T^2 \right] = \sum_{k=0}^{n-1} (t_{k+1} - t_k)^2 \text{Var}(Z_k^2) \leq \text{const.} \times T r \to 0,$$

as $r \to \infty$. Hence, we conclude that

$$\mathbb{E} \left[ \left( I_n - \frac{1}{2} B_T^2 + \frac{T}{2} \right)^2 \right] = \text{Var} \left[ I_n - \frac{1}{2} B_T^2 \right] \to 0$$

as $n \to \infty$. In other words, $I_n$ converges to $\frac{1}{2} B_T^2 - \frac{T}{2}$ in $L_2$ norm or mean square sense.

**Example 8.2 (Exercise)** Replicating the arguments above we can show that

$$\sum_{k=0}^{n-1} f(t_k)(B_{t_{k+1}} - B_{t_k})^2 \to \int_0^T f(t) \, dt,$$

where the convergence is in $L_2$ and hence in probability. All we need to show is that:

$$\mathbb{E} \left[ \sum_{k=0}^{n-1} f(t_k)([B_{t_{k+1}} - B_{t_k}]^2 - t_{k+1} + t_k) \right]^2 \to 0, \quad n \to \infty.$$
We can further show that
\[ \int_0^T f(t) \, dB_t = \sum_{n=1}^{\infty} Z_n \int_0^T f(t) h_n(t) \, dt. \]

## 8.1 Stratonovich Integral

An alternative definition to (2) is the following stochastic integral:

\[ \int_0^T f(s; B_s) \circ dB_s = \lim_{r \to 0} \frac{1}{n-1} \sum_{k=0}^{n-1} \left\{ f(t_k; B_{t_k}) + f(t_k; B_{t_{k+1}}) \right\} \] 

This is known as the Stratonovich stochastic integral.

**Remark 8.1** Note the special form of the integrand. It is not a general adapted process. For example

\[ f(t, B_t) = \frac{1}{t} \int_0^t B_u \, du, \]

while allowed with Ito’s definition, is not allowed here.

The reason for this definition is the following formula, which looks like the standard Calculus integration formula:

\[ \int_0^T f'(s; B_s) \circ dB_s = f(T, B_T) - f(0, B_0). \]

In fact, we can express the Stratonovich integral in terms of the Ito integral:

\[ \int_0^T Y_s \circ dX_s = \lim_{r \to 0} \frac{1}{2} \sum_{i=1}^{n} \left( Y_{t_i} + Y_{t_{i-1}} \right) \left( X_{t_i} - X_{t_{i-1}} \right) \]

\[ = \lim_{r \to 0} \frac{1}{2} \sum_{i=1}^{n} \left( Y_{t_i} - Y_{t_{i-1}} + Y_{t_{i-1}} \right) \left( X_{t_i} - X_{t_{i-1}} \right) \]

\[ = \frac{1}{2} \lim_{r \to 0} \sum_{i=1}^{n} \left( Y_{t_i} - Y_{t_{i-1}} \right) \left( X_{t_i} - X_{t_{i-1}} \right) + \lim_{r \to 0} \sum_{i=1}^{n} Y_{t_{i-1}} \left( X_{t_i} - X_{t_{i-1}} \right) \]

\[ = \frac{1}{2} \langle Y, X \rangle_t + \int_0^T Y_s \, dX_s, \]

where \( \langle Y, X \rangle_t \) is the covariation of the processes \( Y_t \) and \( X_t \) (when \( X_t \equiv Y_t \), then it is called the variation of the process). Note that in differential form the above becomes:

\[ Y_t \circ dX_t = Y_t dX_t + \frac{1}{2} d\langle Y, X \rangle_t. \]
Example 8.3 Let \( f(t, x) = x \), then
\[
\int_0^T B_s \circ dB_s = \int_0^T B_s dB_s + \frac{T}{2},
\]
which agrees with what we have done so far.

9 Ito’s formula

An Ito process is any process \((X_t, t \geq 0)\) that can be written in the form:
\[
X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dB_s,
\]
where \( \mu_s \) and \( \sigma_s \) are adapted to \((\mathcal{F}_s, t \geq 0)\), and \( \int_0^T |\mu_s| ds < \infty, \int_0^T \sigma_s^2 ds < \infty \).
The above integral equation is usually written in the short-hand form:
\[
dX_t = \mu_t dt + \sigma_t dB_t, \quad X(0) = X_0.
\]
An Ito process is an example of a semi-martingale, that is, it can be written as a local martingale (the “noise” part of the process) plus a finite variation process (the “trend” part \( \int_0^t \mu_s ds \) of the process). A local martingale is the same as a martingale except that the integrability condition is relaxed.

Lemma 9.1 (Ito’s lemma) Let \( X_t \) be any Ito process and \( g : [0, \infty) \times \mathbb{R} \to \mathbb{R} \) be such that \( g \in C^{1,2} \) (that is, a twice continuously differentiable function with respect to the second argument and a once continuously differentiable function with respect to the first argument), then \( Y_t = g(t, X_t) \) is also an Ito process with
\[
dY_t = g_t(t, X_t) dt + g_x(t, X_t) dX_t + \frac{1}{2} g_{xx}(t, X_t) (dX_t)^2,
\]
where we use the rules
\[
dt^2 = 0, \quad dt dB_t = 0, \quad (dB_t)^2 = dt.
\]
We can thus calculate the “trend” and “noise” parts of the new process \((Y_t, t \geq 0)\):
\[
dY_t = \left( g_t + g_x \mu_t + \frac{1}{2} g_{xx} \sigma_t^2 \right) dt + \sigma_t g_x dB_t
\]
Figure 1: As seen from the photo, pop singer Avril Lavigne has been interested in stochastic calculus and probability theory as a high-school student. Her interest was inspired by stochastic models for acoustic attenuation. These models make it possible for engineers to make high-fidelity recording of her signing and music. She also says the most common way to create compositions through mathematics is to use stochastic processes. In stochastic models a piece of music is composed as a result of non-deterministic methods. The compositional process is only partially controlled by the composer by weighting the possibilities of random events. She is still struggling with the fundamentals of Ito’s Calculus. To remember Ito’s formula, she says she has written it all over the walls of her Hollywood mansion and on all mirrors (she looks at mirrors quite often!). For other pop stars who think maths is cool, see http://britneyspears.ac/lasers.htm.
Example 9.1 Compute the following stochastic differentials

1. $X_t = B_t$ and $g(t, x) = x^2$. So we obtain $\int_0^t B_s dB_s = \frac{1}{2} B_t^2 - t/2$.

2. Let $X_t = B_t$ and $g(t, x) = f(t)x$. So we obtain

$$\int_0^t f(s) dB_s = f(t)B_t - \int_0^t f'(s)B_s ds,$$

which could be used to define the stochastic integral on the left for deterministic $f(t)$.

3. Let $Y_t = \sigma e^{-\alpha t} \int_0^t e^{\alpha s} dB_s$, where $\sigma$ and $\alpha$ are constants (The stochastic differential is is called Langevin’s equation and it described the Ornstein-Uhlenbeck process, which is an alternative, smoothed version, of the Brownian motion particle model. In fact, $\alpha = 0$ yields the original Wiener process.). Find $dY_t$ and compute $E[Y_t]$ and $E[Y_t^2]$, $E[Y_t Y_s]$, $t > s$. Show that $Y_t \sim N(0, \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t}))$.

4. Example 17+18/page 104;

5. $Y_t = e^{\lambda B_t}$, $0 \leq t \leq T$. Find $dY_t$ and use the result to verify $E[Y_t] = e^{\lambda^2/2}$, $0 \leq t \leq T$.

Outline of proof of Ito’s lemma

In this proof we assume that $g$ has bounded 3-rd derivatives. These conditions can be relaxed by technical arguments. Here $0 = t_0 < t_1 < \cdots < t_n = t$, $\delta t_j = t_{j+1} - t_j = t/n$, $\delta X_j = X_{t_{j+1}} - X_{t_j}$, $\mu_j = \mu(t_j)$, $\sigma_j = \sigma(t_j)$, $g^{(j)}(t; X_{t_j})$,  

30
then:

\[ Y_t - Y_0 = \sum_{j=0}^{n-1} g^{(j+1)} - g^{(j)} \]

\[ = \sum_{j=0}^{n-1} g_x^{(j)} \delta X_j \quad (A) \]

\[ + \sum_{j=0}^{n-1} g_t^{(j)} \delta t_j \quad (B) \]

\[ + \frac{1}{2} \sum_{j=0}^{n-1} g_{tt}^{(j)} (\delta t_j)^2 \quad (C) \]

\[ + \frac{1}{2} \sum_{j=0}^{n-1} g_{xx}^{(j)} (\delta X_j)^2 \quad (D) \]

\[ + \sum_{j=0}^{n-1} g_{xt}^{(j)} \delta t_j \delta X_j \quad (E) \]

\[ + \sum_{j=0}^{n-1} o(\mid \delta t_j \mid^2 + \mid \delta X_j \mid^2) \]

We examine the convergence of each term as \( n \to \infty \) in \( \mathcal{V}[0,T] \) (that is, convergence in MISE). We focus on interesting terms in detail. We have

\[ A \to \int_0^t g_x(s,X_s)(\mu_s ds + \sigma_s dB_s) \]

\[ B \to \int_0^t g_s(s,X_s)ds. \]

For term \( C \) we have:

\[ \mathbb{E} \left[ \sum_{j=0}^{n-1} g_{tt}^{(j)} (\delta t_j)^2 \right]^2 \to 0. \]

Similarly \( E \) goes to 0. Now consider term \( D \):

\[ D = \frac{1}{2} \sum_{j=0}^{n-1} g_{xx}^{(j)} (\mu_j^2 (\delta t_j)^2 + 2 \sigma_j \mu_j B_j \delta t_j + \sigma_j^2 (\delta B_j)^2). \]

The only part that does not converge to zero is

\[ \frac{1}{2} \sum_{j=0}^{n-1} g_{xx}^{(j)} \sigma_j^2 (\delta B_j)^2 \to \frac{1}{2} \int_0^t g_{xx}(s,X_s) \sigma_s^2 ds, \]
which is demonstrated as in Example 9.2. Once we have established the lemma assuming bounded derivatives, we can obtain the general result by using $C_2$ bounded functions which converge uniformly on compact subsets.

The quadratic variation and covariation is can be used to gain further insight into Ito’s Lemma.

**Example 9.2 (Quadratic Variation)** Let $dX_t = \mu_t dt + \sigma_t dB_t$ be an Ito process, then using the definition of quadratic variation and Taylor’s theorem:

$$\langle X \rangle_t = \lim_{r \to 0} \frac{1}{n} \sum_{i=0}^{n-1} (X_{t_{i+1}} - X_{t_i})^2$$

$$= \lim_{r \to 0} \frac{1}{n} \sum_{i=0}^{n-1} (\mu_t \delta t_i + \sigma_t (B_{t_{i+1}} - B_{t_i}))^2$$

$$= 0 + \lim_{r \to 0} \frac{1}{n} \sum_{i=0}^{n-1} \sigma_t^2 (B_{t_{i+1}} - B_{t_i})^2 = \lim_{r \to 0} I_n.$$ 

We now verify that

$$E \left[ I_n - \sum_{i=0}^{n-1} \sigma_t^2 \delta t_i \right]^2 \to 0, \quad r \to 0.$$ 

Hence $I_n$ converges to $\int_0^t \sigma_s^2 ds$ in $L^2$, that is, we have shown that:

$$E \left[ \sum_{i=0}^{n-1} (X_{t_{i+1}} - X_{t_i})^2 - \sum_{i=0}^{n-1} \sigma_t^2 \delta t_i \right]^2 \to 0, \quad r \to 0.$$ 

**Example 9.3 (Covariation of two processes)** Let $dX_t = \mu_t dt + \sigma_t dB_t$ and $dY_t = \gamma_t dt + \varrho_t dB_t$ be two Ito processes, then it is easy to verify that:

$$\frac{\langle X + Y \rangle_t - \langle X - Y \rangle_t}{4} = \frac{1}{4} \lim_{r \to 0} \frac{1}{n} \sum_{i=0}^{n-1} (X_{t_{i+1}} + Y_{t_{i+1}} - X_{t_i} + Y_{t_i})^2 - (X_{t_{i+1}} - Y_{t_{i+1}} - X_{t_i} + Y_{t_i})^2$$

$$= \lim_{r \to 0} \frac{1}{n} \sum_{i=0}^{n-1} (X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i})$$

$$= \langle X, Y \rangle_t.$$ 

It then follows that the covariation of the $X$ and $Y$ is given by:

$$\langle X, Y \rangle_t = \int_0^t (\sigma_s + \varrho_s)^2 ds - \int_0^t (\sigma_s - \varrho_s)^2 ds = \int_0^t \sigma_s \varrho_s ds.$$ 

By convention, the differential form is written as

$$d\langle X, Y \rangle_t = \sigma_t \varrho_t dt.$$ 

32
9.1 Ito formula for function of two variables

Let \( dX_t = \mu_t dt + \sigma_t dB_t \) and \( dY_t = \gamma_t dt + \theta_t dB_t \) be two Ito processes, then if \( f(x, y) \) is smooth: 
\[
f(X_{t+1}, Y_{t+1}) = f(X_t, Y_t) + f_x(X_t, Y_t)(X_{t+1} - X_t) + f_y(X_t, Y_t)(Y_{t+1} - Y_t) + f_{xy}(X_t, Y_t)(Y_{t+1} - Y_t)(X_{t+1} - X_t) + \frac{1}{2}f_{xx}(X_t, Y_t)(X_{t+1} - X_t)^2 + \frac{1}{2}f_{yy}(X_t, Y_t)(Y_{t+1} - Y_t)^2 + o(|dX_t|^2 + |dY_t|^2).
\]
Hence we can write:
\[
f(X_t, Y_t) - f(X_0, Y_0) = \sum_{j=0}^{n-1} f(X_{t_{i+1}}, Y_{t_{i+1}}) - f(X_t, Y_t)
= \sum_{j=0}^{n-1} f_x(X_t, Y_t)(X_{t+1} - X_t)
+ \sum_{j=0}^{n-1} f_y(X_t, Y_t)(Y_{t+1} - Y_t)
+ \sum_{j=0}^{n-1} f_{xy}(X_t, Y_t)(Y_{t+1} - Y_t)(X_{t+1} - X_t)
+ \frac{1}{2} \sum_{j=0}^{n-1} f_{xx}(X_t, Y_t)(X_{t+1} - X_t)^2
+ \frac{1}{2} \sum_{j=0}^{n-1} f_{yy}(X_t, Y_t)(Y_{t+1} - Y_t)^2
+ \sum_{j=0}^{n-1} o(|dX_t|^2 + |dY_t|^2).
\]
The first term converges to a limit in mean square sense, which we denote by
\[
\int_0^t f_x(X_s, Y_s) dX_s := \int_0^t f_x(X_s, Y_s) \mu_s ds + \int_0^t f_x(X_s, Y_s) \sigma_s dB_s.
\]
Similarly for the second term. Just like in Example 9.2, we can show that the third term converges in \( L_2 \) to a quantity which we denote
\[
\int_0^t f_{xy}(X_s, Y_s) dX_s dY_s = \int_0^t f_{xy}(X_s, Y_s) \sigma_s \theta_s ds.
\]
The fourth term converges to
\[
\int_0^t f_{xx}(X_s, Y_s) \sigma_s^2 ds,
\]
and similarly for the fifth. The rest converges to 0 in mean square sense. Again, for convenience, we write the formula in differential form:
\[
df(X_t, Y_t) = f_x dX_t + f_y dY_t + f_{xy} d\{X, Y\}_t + \frac{1}{2} f_{xx} d\{X\}_t + \frac{1}{2} f_{yy} d\{Y\}_t.
\]
where the derivatives of $f$ are evaluated at $(X_t, Y_t)$. The differential form is just a notational convention, which is given meaning only through the integral form.

**Example 9.4 (Product Rule)** If we set $f(x, y) = xy$ above, then we get the product rule:

$$d(X_tY_t) = Y_t dX_t + X_t dY_t + d\langle X, Y \rangle_t.$$ 

**Example 9.5 (Example 21/ page 109)**

**10 Radon-Nikodim Derivative**

Let $W_t$ be SBM under $\mathbb{P}$, that is

$$\mathbb{E}_\mathbb{P}[e^{sW_t}] = e^{st/2}.$$ 

If we let $B_t = W_t + t\gamma$, $\gamma \neq 0$, then $B_t$ is not SBM because $B_t \not\sim N(0, t)$ is not satisfied. There exists, however, a probability law $\mathbb{Q}$ such that $B_t$ is SMB under $\mathbb{Q}$. Obviously $\text{Var}_\mathbb{P}(B_t) = \text{Var}_\mathbb{P}(W_t)$ but $\mathbb{E}_\mathbb{P}(B_t) = t\gamma \neq 0$, hence under $\mathbb{Q}$ we require $\mathbb{E}_\mathbb{Q}[B_t] = 0$ or $\mathbb{E}_\mathbb{Q}[W_t] = -\gamma t$. We thus wish to have:

$$\mathbb{E}_\mathbb{Q}[B_t] = e^{st/2} = e^{st\gamma} \mathbb{E}_\mathbb{Q}[W_t]$$

Hence under $\mathbb{Q}$, $W_t$ is $N(-t\gamma, t)$ distributed and in such a case $B_t$ is $N(0, t)$ distributed. Hence:

1. $B_t$ is SBM under $\mathbb{Q}$, but not $\mathbb{P}$
2. $W_t$ is SBM under $\mathbb{P}$, but not $\mathbb{Q}$.

We can compute $\mathbb{E}_\mathbb{P}[W_t] = \int_\mathcal{R} N(0, t) wdw = \int_\mathcal{R} N(-t\gamma, t) N(0,t) wdw = \mathbb{E}_\mathbb{Q}[\Lambda_t W_t]$, where $\Lambda = \frac{N(0,t)}{N(-t\gamma,t)} = e^{w_t-\frac{1}{2}\gamma^2 t}$ (with slight abuse of notation) is the so-called Likelihood Ratio and is usually denoted by $\frac{d\mathbb{P}}{d\mathbb{Q}}$ so that $\mathbb{E}_\mathbb{Q}[\frac{d\mathbb{P}}{d\mathbb{Q}} W_t] = \int_\mathcal{W} \frac{d\mathbb{P}}{d\mathbb{Q}} W(\omega) d\mathbb{Q} = \mathbb{E}_\mathbb{P}[W_t]$. Let $\mathbb{P}$ and $\mathbb{Q}$ be defined on the same space, then

**Definition 10.1** $\mathbb{Q}$ is called absolutely continuous with respect to $\mathbb{P}$, written $\mathbb{Q} \ll \mathbb{P}$ if for all $A \in \Omega$ $\mathbb{P}(A) = 0 \Rightarrow \mathbb{Q}(A) = 0$. $\mathbb{P}$ and $\mathbb{Q}$ are said to be equivalent if $\mathbb{Q} \ll \mathbb{P}$ and $\mathbb{Q} \gg \mathbb{P}$.

**Theorem 10.1 (Radon-Nikodym)** Let $\mathbb{Q} \ll \mathbb{P}$, then there exists a random variable $\Lambda \geq 0$, such that $\mathbb{E}_\mathbb{P}[\Lambda] = 1$ and

$$\mathbb{Q}(A) = \mathbb{E}_\mathbb{P}[\Lambda I_A] = \int_A \Lambda d\mathbb{P}$$

for any measurable set $A$. 

34
The random variable $\Lambda$ is usually denoted as $\frac{dQ}{dP}$ and is called the Radon-Nikodym derivative of the measure $Q$ with respect to $P$. Thus we can compute expectations under a different measure

$$E_Q[X] = E_P[\Lambda X]$$

**Example 10.1**

Note that for time-dependent processes $\Lambda = \frac{dQ}{dP}$ depends on $t$. In our example we have

$$\Lambda(t) = e^{\gamma W_t - \frac{1}{2} \gamma^2 t}$$

Note that $\Lambda(t)$ is an $(\mathcal{F}_t, P)$ martingale:

$$\zeta_t = E_E[\Lambda_s | \mathcal{F}_t, s \geq t] = \Lambda_t$$

This is true in general.

For $X_t$ adapted $(\mathcal{F}_t, t \geq 0)$ it can be shown that

$$E_P[X_t | \mathcal{F}_s] = E_Q\left[ X_t \frac{\zeta_t}{\zeta_s} | \mathcal{F}_s \right], \quad t \geq s,$$

where $\zeta_t = E_P[\frac{dQ}{dP} | \mathcal{F}_t]$.

**Example 10.2 (Verification)** In our case we have $E_P[W_t | \mathcal{F}_s] = W_s$ since $W_t$ is SMB under $P$ and hence an $(\mathcal{F}_t, P)$ martingale. Now, $E_Q\left[ W_t e^{(W_t - W_s)} | \mathcal{F}_s \right] = e^{-\frac{1}{2} \gamma^2(t-s)} E_Q\left[ (W_t - W_s) e^{(W_t - W_s)} \right] + e^{-\frac{1}{2} \gamma^2(t-s)} W_s E_Q\left[ e^{(W_t - W_s)} \right] = 0 + W_s$. So, the identity holds for our particular example.

# 11 Girsanov’s theorem

Changes of measure are useful when working with martingales (and hence with the risk-neutral pricing formula).

**Theorem 11.1 (Girsanov’s theorem)** Let $(B_t, T \geq t \geq 0)$ be SBM under $\mathbb{P}$ and hence an $(\mathcal{F}_t, \mathbb{P})$ martingale. Define

$$\Lambda_T = e^{\int_0^T \theta_s dB_s - \frac{1}{2} \int_0^T \theta_s^2 ds},$$

then

$$W_t = B_t - \int_0^t \theta_s ds$$

is $Q$ Brownian motion, where the probability measure

$$Q(A) = \int_A \Lambda_T d\mathbb{P},$$

and we assume $E_E\left[ e^{\frac{1}{2} \int_0^T \theta_s^2 ds} \right] < \infty$ (Novikov’s condition).

35
For example,

\[ \mathbb{E}_P[\Lambda_T X_T] = \mathbb{E}_Q[X_T] \]

and in general

\[ \mathbb{E}_P\left[ \frac{\Lambda_t}{\Lambda_s} X_t \bigg| \mathcal{F}_s \right] = \mathbb{E}_Q[X_t \big| \mathcal{F}_s], \quad t > s. \]

**Example 11.1 (Verification)** Suppose \( \theta \) is deterministic, then

\[ \mathbb{E}_P[\Lambda_T | \mathcal{F}_s] = e^{\int_0^s \theta_u du - \frac{1}{2} \int_0^s \theta_u^2 du} \mathbb{E}_P \left[ e^{\int_0^t \theta_u dB_u} \bigg| \mathcal{F}_s \right] = e^{\int_t^s \theta_u du - \frac{1}{2} \int_t^s \theta_u^2 du} e^{\frac{1}{2} \int_t^s \theta_u^2 du} e^{\frac{1}{2} \int_t^s \theta_u^2 du} = e \]

\[ = \Lambda_s. \] Hence, from Novikov’s condition, \( \Lambda_s \) is a martingale.

Example 22/page 119

**12 Stochastic Differential Equations**

**12.1 Strong solution of SDE’s**

Consider the SDE

\[ dX_t = \mu(t, X_t) dt + b(t, X_t) dB_t \]

**Definition 12.1 (Strong Solution of SDEs)** \( X_t \) is called a strong solution if:

1. \( X_t \) is a function of \( t \) and the given Brownian motion \( (B_s, s \leq t) \), that is, it can be written as \( X_t = F(t, (B_s, s \leq t)) \) for some function \( F \);

2. the integrals \( \int_0^t \mu_s ds \) and \( \int_0^t \sigma_s dB_s \) exist, that is \( \sigma_t \) is adapted to the underlying brownian filtration and \( \int_0^t \sigma_s^2 ds < \infty \).

3. \( X_t \) satisfies the integral equation:

\[ X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dB_s. \]

**Example 12.1** Consider the SDE:

\[ dX_t = X_t dB_t, \quad X_0 = 1. \]

The solution is

\[ X_t = e^{B_t - t/2} = F(t, (B_s, s \leq t)), \]

where \( F(t, x) = e^{x - t/2} \). It is a strong solution.
Example 12.2 The solution of the SDE
\[ dX_t = rX_t dt + \sigma X_t dB_t, \quad X_0 = 1 \]
is
\[ X_t = e^{(r-\sigma^2/2)t + \sigma B_t}. \]
Here again \( F(t,x) = e^{(r-\sigma^2/2)t + \sigma x} \) and we evaluate \( F \) at \( B_t \in (B_s, 0 \leq s \leq t) \).

Example 12.3 The Langevin equation
\[ dX_t = -\alpha X_t dt + \sigma dB_t, \]
where \( \alpha \) and \( \sigma \) are non-negative constants. The solution can be written as:
\[ X_t = F(t, (B_s, s \leq t)), \]
where
\[ F(t, (x(s), 0 \leq s \leq t)) = e^{-\alpha t}x_0 + \sigma x(t) - \sigma \alpha \int_0^t e^{-\alpha(t-s)}x(s)ds. \]

12.2 Existence and uniqueness of strong solutions
We now discuss existence and uniqueness issues and some examples. We consider the SDE
\[ dX_t = a(t,X_t) dt + b(t,X_t) dB_t, \quad X_0 \text{ given}. \quad (3) \]

Theorem 12.1 (Uniqueness) Suppose the following a satisfied:
1. \(|a(t,x)| + |b(t,x)| \leq c(1 + |x|) \) (linear growth condition)
2. \(|a(t,x) - a(t,y)| \leq D|x-y| \) (Lipschitz continuity in \( x \) and uniformly in \( t \))
3. \(|b(t,x) - b(t,y)| \leq D|x-y| \)

where \( t \in [0,T], x,y \in \mathbb{R}, c,D \) positive constants. Then given \( X_0 \) such that \( \mathbb{E}[X_0^2] < \infty \) and \( X_0 \in \mathcal{F}_0^B \), (3) satisfies the following:
- has a unique strong solution on \([0,T]\);
- \( t \rightarrow X_t \) is continuous with probability 1;
- the solution \( X_t \) is adapted to \((\mathcal{F}_t^B, t \geq 0)\) and \( \int_0^T \mathbb{E}[X_s^2]ds < \infty \). (also see page 126 of Klebaner)

Example 12.4
\[ dX_t = X_t^2 dt, \quad X_0 = x_0 > 0. \]
Here \( a = x^2 \) and \( b = 0 \), hence \(|x^2| \leq c|x| \) is not true for all \( x \in \mathbb{R} \).
In fact, \( X_t = \frac{x_0}{1-x_0t} \) will blow up at \( t^* = \frac{1}{x_0} \).
Example 12.5 If \( a(t, x) = |x|^\lambda, \ 0 < \lambda < 1, \ b = 0 \) then the Lipschitz condition 2. fails to hold and with \( x_0 = 0 \), we have \( X_t = 0 \) and \( X_t = ((1 + \lambda) t)^{1/\lambda} \) are both solutions of

\[
X_t = \int_0^t |X_s|^\lambda ds.
\]

Example 12.6 If \( a(t, x) = 0 \) and \( b(t, x) = |x|^\lambda, \ X_0 = 0, \) then Girsanov (1962) showed that

\[
X_t = \int_0^t |X_s|^\lambda dB_s
\]

has

1. a strong unique solution for \( \lambda \geq 1/2 \). I.e., condition 2. is too restrictive and can be substituted for

\[
|b(t, x) - b(t, y)| \leq K|x - y|\nu
\]

for some \( \nu \geq 1/2 \). (this is Holder continuity condition of order \( \nu \).)

2. an infinite number of solutions for \( \lambda < 1/2 \).

Anderson (1969 PhD) and Yamada (1976) generalized condition 2. further to

\[
|b(t, x) - b(t, y)| \leq \varrho(|x - y|),
\]

where \( \varrho \) satisfies

\[
\int_{(0, \varepsilon]} \frac{1}{\varrho^2(u)} du < \infty, \ (\varrho(0) = 0)
\]

Example 12.7 Consider the Cox-Ross interest rate model

\[
dr_t = K(r_0 - r_t)dt + \varphi \sqrt{r_t} dB_t.
\]

Here \( a(t, x) = K(r_0 - x) \) and \( b(t, x) = \varphi \sqrt{x} \) and \( |b(t, x) - b(t, y)| = \varphi \sqrt{x} - \sqrt{y} \leq \varphi \sqrt{|x - y|} \) and so a unique solution exists.

Example 12.8 The Lipschitz continuity conditions may be weakened to

\[
|a(t, x) - a(t, y)| \leq K_N|x - y| \text{ for all } N \geq 0 \text{ and } |x|, |y| \leq N,
\]

where we assume there exists a constant \( K_N \). For example for

\[
dx_t = \sin(X_t^2)dt, \quad X_0 = x_0
\]

we have

\[
|\sin(x^2) - \sin(y^2)| \leq K_N|x - y|
\]

where \( K_N = 2N \) and \( |x|, |y| \leq N \). (see Arnold’s “SDE” (1973), page 112)
Example 12.9 If the Lipschitz conditions fail and the growth condition 1. fails, then with $a, b \in C^1$ there exists a unique $(\mathcal{F}_t^B, t \geq 0)$ adapted solution up to an explosion time $\eta$. That is, the solution exists in $[0, \eta)$. If $\eta < \infty$, then $X_t \to \pm \infty$ as $t \to \eta^-$. For example,

$$dX_t = \frac{1}{2} e^{-2X_t} dt + e^{-X_t} dB_t, \quad X_0 = c$$

has solution

$$X_t = \ln(B_t + e^c)$$

up to the explosion time $\eta$:

$$\eta(\omega) = \inf \{ t : B_t(\omega) = -e^c \} > 0$$

for each $\omega$. (see Arnold page 113)

12.3 Weak solutions of SDE’s

The two notions of solutions of SDEs can informally be summarized as:

**Strong solution:** Given $\mu, \sigma$, $(B_s, s \leq t)$, find $X_t$. By uniqueness here we mean that if $\tilde{X}_t$ is another solution then

$$P(X_t = \tilde{X}_t) = 1, \quad t \in [0, T].$$

**Weak solution:** Given $\mu, \sigma$, find $(B_s, s \leq t)$, find $X_t$. By uniqueness here we mean that if $\tilde{X}_t$ is another solution then for a bounded continuous $g(\cdot)$:

$$E[g(X_t)] = E[g(\tilde{X}_t)], \quad t \in [0, T].$$

**Definition 12.2 (Weak solution of SDEs)** A process $\hat{X}_t$ is called a weak solution of the SDE with initial distribution $F_0$:

$$dX_t = \mu(t, X_t) dt + \sigma(t, X_t) dB_t$$

if:

1. there exists a on a probability space with filtration;
2. Brownian motion $\hat{B}_t$ and process $\hat{X}_t$ adapted to that filtration,
3. $\hat{X}_0$ has a given distribution $F_0$,
4. the integrals $\int_0^t \mu_s ds$ and $\int_0^t \sigma_s dB_s$ exist,
5. and $\hat{X}_t$ satisfies the equation:

$$\hat{X}_t = \hat{X}_0 + \int_0^t \mu(s, \hat{X}_s)ds + \int_0^t \sigma(s, \hat{X}_s)d\hat{B}_s.$$ 

A weak solution $X_t$ is unique if any other solution $\hat{X}_t$ has the same distributional properties.

Example 12.10 (Tanaka’s SDE) If

$$dX_t = (2I_{\{X_t > 0\}} - 1)dB_t,$$

then $\sigma(x) = 2I_{\{x > 0\}} - 1$ is not Lipschitz continuous and therefore there does not exist a strong solution. It can be shown that Brownian motion is the unique weak solution of the SDE.

12.4 Existence and uniqueness of weak solutions

Theorem 12.2 (Existence of weak solutions) The SDE has at least one weak solution if $\sigma(t, x)$ and $\mu(t, x)$ are bounded and continuous. The SDE has a unique weak solution if either of these holds:

1. $\sigma(t, x)$ and $\mu(t, x)$ and their partial derivatives up to order 2 with respect to $x$ are bounded and continuous;

2. For any $T > 0$, there exists $K_T$ such that

$$|\sigma(t, x)| + |\mu(t, x)| \leq K_T(1 + |x|), \quad \forall x \in \mathbb{R}.$$

13 Ito’s formula for vector processes

Let $X_t = (X_t^{(1)}, \ldots, X_t^{(n)})^T$ be an $n$-dimensional Ito process:

$$dX_t^{(1)} = \mu_1 dt + \sum_{j=1}^m \sigma_{1j} dW_t^{(j)}$$

$$\vdots \quad \vdots$$

$$dX_t^{(n)} = \mu_n dt + \sum_{j=1}^m \sigma_{nj} dW_t^{(j)}.$$

or in matrix notation:

$$dX_t = \mu_t dt + \sigma_t dW_t,$$

40
where
\[
\mu_t = \begin{pmatrix} 
\mu_1 \\
\vdots \\
\mu_n 
\end{pmatrix}, \quad \sigma_t = \begin{pmatrix} 
\sigma_{11} & \cdots & \sigma_{1m} \\
\vdots & \ddots & \vdots \\
\sigma_{n1} & \cdots & \sigma_{nm} 
\end{pmatrix}, \quad W_t = \begin{pmatrix} 
W_t^{(1)} \\
\vdots \\
W_t^{(m)} 
\end{pmatrix}.
\]

Suppose \( g : \mathbb{R}^n \rightarrow \mathbb{R} \), then by Taylor’s theorem, we have:
\[
g(\mathbf{X}_{t+\delta t}) - g(\mathbf{X}_t) = \nabla g(\mathbf{X}_t)^T (\mathbf{X}_{t+\delta t} - \mathbf{X}_t) + \frac{1}{2} \nabla^2 g(\mathbf{X}_t) (\mathbf{X}_{t+\delta t} - \mathbf{X}_t)(\mathbf{X}_{t+\delta t} - \mathbf{X}_t) + \text{HOT},
\]
where \( \nabla^2 g(\mathbf{X}_t) = \left[ \frac{\partial^2 g(\mathbf{X}_t)}{\partial x_i \partial x_j} \right]_{ij} \) is the Hessian matrix of partial derivatives. This can be written as:
\[
d(g(\mathbf{X}_t)) = \sum_{i=1}^{m} \frac{\partial g}{\partial x_i^{(i)}} (\mathbf{X}_{t+\delta t} - \mathbf{X}_t^{(i)}) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 g}{\partial x_i^{(i)} \partial x_j^{(j)}} (\mathbf{X}_{t+\delta t} - \mathbf{X}_t^{(i)})(\mathbf{X}_{t+\delta t} - \mathbf{X}_t^{(j)}) + \text{HOT}.
\]

Therefore, it can be argued that the multidimensional Ito formula is:
\[
d(g(\mathbf{X}_t)) = \sum_{i=1}^{m} \frac{\partial g}{\partial x_i^{(i)}} d\mathbf{X}_t^{(i)} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 g}{\partial x_i^{(i)} \partial x_j^{(j)}} d\mathbf{X}_t^{(i)} d\mathbf{X}_t^{(j)},
\]
where, using the rules \( dt^2 = 0 = dt dW_t, \ dW_t^{(i)} dW_t^{(j)} = \delta_{ij} dt \), we can show:
\[
d\mathbf{X}_t^{(i)} d\mathbf{X}_t^{(j)} = d(\mathbf{X}_t^{(i)}, \mathbf{X}_t^{(j)}) = \left[ \sum_{k=1}^{m} \sigma_{ik} \sigma_{jk} \right] dt.
\]

### 14 Martingale Representation Theorems

We know that Ito integrals are martingales. It turns out that any square integrable martingale can also be written in terms of an Ito integral. This follows from the following theorem:

**Theorem 14.1 (Martingale Representation Theorem)** Let \( M_t, \ t \in [0,T] \) be a square integrable martingale adapted to \( (\mathcal{F}_t, 0 \leq t \leq T) \). Then there exists a unique \( (\mathcal{F}_t, 0 \leq t \leq T) \) adapted process \( \phi_t \) such that:

1. \( \mathbb{E} \int_0^T \phi_t^2 dt < \infty \)
2. \( M_t = M_0 + \int_0^t \phi_s dB_s, \ t \in [0,T] \).
Example 14.1 Recall that if $Y$ is any $\mathcal{F}_T$ measurable process with $\mathbb{E}[Y^2] < \infty$, then we can construct the $(\mathcal{F}_t, \mathbb{P})$ martingale

$$M_t = \mathbb{E}[Y | \mathcal{F}_t], \quad t \in [0,T].$$

Say $Y = B^2_T$, then $M_t = T + B^2_t - t$ and we have

$$M_t = T + 2 \int_0^t B_s dB_s,$$

that is, $\phi_t = B_t \in \mathcal{V}[0,T]$ in the martingale representation theorem.

Example 14.2 (previous example continued) If $Y = e^{\sigma B_T}$, then $M_t = e^{\sigma B_t + \frac{1}{2} \sigma^2 (T-t)}$ and the martingale representation theorem is confirmed by writing

$$M_t = e^{\frac{1}{2} T \sigma^2} + \int_0^t \sigma^2 e^{\sigma B_s + \frac{1}{2} \sigma^2 (T-s)} dB_s.$$

The examples suggest the a square integrable random variable can be represented as an Ito integral.

Theorem 14.2 (The Ito representation theorem) Let $Y \in L^2(\mathcal{F}_T, \mathbb{P})$, that is, it is a square integrable random variable. Then there exists an $(\mathcal{F}_t, T \geq t \geq 0)$-adapted process $f(t, \omega) \in \mathcal{V}[0,T]$ such that

$$Y = \mathbb{E}_\mathbb{P}[Y] + \int_0^T f(t, \omega) dB_t(\omega),$$

where $B_t$ is $(\mathcal{F}_t, \mathbb{P})$ Brownian motion.

Proof: Let $M_t = \mathbb{E}_\mathbb{P}[Y | \mathcal{F}_t]$, then $M_t$ is a square integrable martingale and by the representation we have that

$$M_t = M_0 + \int_0^t f(s, \omega)dB_s,$$

therefore by setting $t = T$, we get:

$$Y = M_T = \mathbb{E}[Y | \mathcal{F}_0] + \int_0^T f(s, \omega)dB_s.$$

We know that if $f$ is deterministic, then the Ito integral has Gaussian distribution (as per assignment). The converse is also true!

Lemma 14.1 (The Ito representation of Gaussian process) If $Y$ has a Gaussian distribution, then $f(t, \omega) = f(t)$ is deterministic.
Example 14.3 (Bonus Question (iv) hint) Turn the random variables into martingales by taking an expectation conditional on \( F_t \). For example, consider

\[
M_t = \mathbb{E}[\sin(B_T) \mid F_t] = e^{-T/2} \mathbb{E}[e^{T/2} \sin(B_T) \mid F_t] = e^{-(T-t)/2} \sin(B_t).
\]

Hence,

\[
M_t = M_0 + \int_0^t e^{-(T-s)/2} \cos(B_s) dB_s
\]

and evaluating this at \( t = T \), we obtain

\[
\sin(B_T) = 0 + \int_0^T e^{-(T-s)/2} \cos(B_s) dB_s.
\]

Example 14.4 (Bonus Question Hint for (ii)) Consider

\[
M_t = \mathbb{E}[\int_0^T B_s ds \mid F_t] = \mathbb{E}[\int_0^T B_s ds - TB_T + TB_T \mid F_t] = \int_0^t B_s ds + (T-t)B_t.
\]

Hence, we have

\[
M_t = M_0 + \int_0^t (T-s) dB_s,
\]

from where

\[
M_T = \int_0^T B_s ds = 0 + \int_0^T (T-s) dB_s.
\]

15 Martingale Construction and Dynkins’ formula

Suppose the process \( X_t \) evolves according to the SDE:

\[
dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dB_t, \quad t \geq 0.
\]

By Ito’s formula, we have for \( f(x, t) \in C^{2,1} \):

\[
df(X_t) = (f_t + \mu_t f_x + \frac{1}{2} \sigma_t^2 f_{xx}) dt + \sigma_t f_x(X_t, t) dB_t.
\]

For notational convenience, we use the differential operator:

\[
L_t[.] = \mu_t \frac{\partial}{\partial x} [.] + \frac{1}{2} \sigma_t^2 \frac{\partial^2}{\partial x^2} [].
\]

Then Ito’s Lemma becomes

\[
df(X_t) = (L_t[f(x, t)] + f_t(X_t, t)) dt + \sigma_t f_x(X_t, t) dB_t.
\]

We now have the following theorem:

Theorem 15.1 (Construction of a Martingale from a process \( X_t \)) Suppose:
1. \( X_t \) satisfies \( dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dB_t, \quad t \geq 0. \)

2. The SDE coefficient \( \mu \) and \( \sigma \) are Lipschitz continuous in the state variable and uniformly in \( t \), that is:
\[
|\mu(x, t) - \mu(y, t)| + |\sigma(x, t) - \sigma(y, t)| \leq D|x - y|, \quad \forall x, y \in \mathbb{R}.
\]
The linear growth condition is satisfied:
\[
|\mu(x, t)| + |\sigma(x, t)| \leq K(1 + |x|).
\]

3. \( f(x, t) \in C^{2,1} \), then the process
\[
M_t = f(X_t, t) - \int_0^t \left( L_u[f] + \frac{\partial f}{\partial t}(X_u, u) \right) du
\]
is a martingale (with respect to the natural filtration and probability measure of the underlying Brownian motion). Moreover, we obtain Dynkin’s formula:
\[
\mathbb{E}[f(X_t, t) | \mathcal{F}_0^B] = f(X_0, 0) + \mathbb{E} \int_0^t \left( L_u[f] + \frac{\partial f}{\partial t}(X_u, u) \right) du.
\]

Proof:
By Itô’s formula:
\[
M_t = M_0 + \int_0^t \sigma(X_u, u)f_x(X_u, u)dB_u,
\]
thus we only need to establish the square integrability. Since \( f_x(x, t) < C_1 \) for all \( x \) and \( t \), and by the existence and uniqueness theorem, under the specified conditions on \( \mu \) and \( \sigma \), the solution of the SDE is square integrable (\( \int_0^t \mathbb{E}[X_u^2]du < C_2 \)), then:
\[
\mathbb{E}[M_t^2] = \int_0^t \mathbb{E}[\sigma^2(X_u, u)f_x^2(X_u, u)]du < C_1^2K^2(1 + 2 \int_0^t \mathbb{E}[[X_u]] + \mathbb{E}[X_u^2]du) < \infty.
\]
This completes the proof.

This suggests that given the Itô process \( X_t \) and any \( C^{2,1} \) function \( f(x, t) \) such that \( f \) solves the pde
\[
L_t[f] + \frac{\partial f}{\partial t}(x, t) = 0, \quad \forall t > 0, x \in \mathbb{R},
\]
then \( M_t = f(X_t, t) \) is a martingale.
Example 15.1 (Is $f(B_t,t) = e^{B_t-t/2}$ a martingale?) Yes, since $f(x,t)$ is smooth and satisfies:

$$L_t[f] + f_t = \frac{1}{2}f_{xx} + f_t = 0.$$ 

Example 15.2 (Assignment question) Let $X_t = \int_0^t g(u)dB_u$, where $g(u)$ is deterministic and smooth, then $dX_t = g(t)dB_t$ and we have:

$$h(t) = E[e^{uX_t}] = 1 + \frac{1}{2}E \int_0^t u^2 g^2(u)e^{uX_u}du.$$ 

It thus follows that

$$h'(t) = \frac{u^2 g^2(t)}{2} h(t), \quad h(0) = 1.$$ 

Hence, by separation of variables, $h(t) = e^{\int_0^t g^2(s)ds}$ which is the MGF of the $\mathcal{N}\left(0, \int_0^T g^2(s)ds\right)$ distribution.

16 Markov property of Brownian motion

Definition 16.1 (Markov Process) A process $X_t$ is called a Markov process if

$$\mathbb{P}(X_{t+s} < x \mid \mathcal{F}_t) = \mathbb{P}(X_{t+s} < x \mid X_t) \quad \text{a.s.},$$

where $\mathcal{F}_t = \sigma\{X_s, 0 \leq s \leq t\}$. In other words, the conditional distribution given the whole past is identical to the conditional distribution given the present state.

An alternative and equivalent definition states that $X_t$ is a Markov process if given the present, the future behavior of the process is independent of the past, that is,

$$\mathbb{P}(X_{t+s} < x, X_{t-u} < y \mid X_t) = \mathbb{P}(X_{t+s} < x \mid X_t)\mathbb{P}(X_{t-u} < y)$$

for all $s > 0, u \in [0,t]$.

Theorem 16.1 (Brownian motion is a Markov process) Proof: It is easy to show that the MGF of $B_{t+s}$ given $\mathcal{F}_t$ is the same as the MGF of $B_{t+s}$ given $B_t$. Alternatively we need to show that which combined with the Gaussian assumption implies that given $B_t$ the statistical behavior of the future $B_{t+s}$ is independent of the past $B_{t-u}$ given $B_t$.

The Markov property can be used to establish a connection between PDE and stochastic processes. Suppose $f(x,t)$ solves the so-called backward equation:

$$L_t[f] + f_t = 0, \quad x \in \mathbb{R}, \quad t \in [0,T].$$
with BC \( f(x, T) = g(x) \). Then by theorem we have that \( M_t = f(X_t, t) - 0, \quad t \in [0, T] \) is a martingale. Hence,
\[
\mathbb{E}[f(X_T, T) \mid \mathcal{F}_t] = f(X_t, t)
\]
or using the Markov property, we have
\[
f(X_t, t) = \mathbb{E}[g(X_T) \mid X_t],
\]
that is,
\[
f(x, t) = \mathbb{E}[g(X_T) \mid X_t = x].
\]
We thus have a strikingly elegant method of computing the solution of the linear PDE \( L_t[f] + f_t = 0 \) through stochastic simulation. A generalization of this result gives

**Theorem 16.2 (Feynmann-Kac formula)** Assume that a solution to the PDE
\[
L_t[f] + \frac{\partial f}{\partial t}(x, t) = r(x, t)f(x, t), \quad t \in [0, T], x \in \mathbb{R},
\]
with BC \( f(x, T) = g(x) \) exists and is given by \( C(x, t) \). Then
\[
C(x, t) = \mathbb{E}_P \left[ e^{-\int_0^T r(X_u, u) du} g(X_T) \mid X_t = x \right]
\]
where \( B_t \) is \( (\mathcal{F}_t, \mathbb{P}) \) Brownian motion.

(Richard Feynmann got the Nobel Prize for applying this discovery in quantum physics) Sketch of proof: Let \( Y_t = e^{-\int_0^t r(X_u, u) du} f(X_t, t) \), then by Ito’s Lemma, we have:
\[
dY_t = e^{-\int_0^t r(X_u, u) du} \left[ (f_t - r(X_t, t)f(X_t, t) + \mu_t f_x + \frac{1}{2} \sigma_t^2 f_{xx}) dt + \sigma_t f_x dB_t \right].
\]
Hence, \( Y_t \) is martingale such that:
\[
Y_t = Y_0 + \int_0^t \sigma(X_u, u) f_x(X_u, u) dB_u.
\]
It then follows that
\[
\mathbb{E}[Y_T \mid \mathcal{F}_t] = Y_t = e^{-\int_0^t r(X_u, u) du} f(X_t, t),
\]
then rearranging, and using the Markov property and the boundary condition, we get:
\[
f(X_t, t) = \mathbb{E}[e^{-\int_t^T r(X_u, u) du} f(X_T, T) \mid \mathcal{F}_t] = \mathbb{E}[e^{-\int_t^T r(X_u, u) du} g(X_T) \mid X_t],
\]
from whence
\[
C(x, t) = \mathbb{E}[e^{-\int_T^T r(X_u, u) du} g(X_T) \mid X_t = x],
\]
as desired.
Example 16.1 (Computing the Risk-Neutral Formula) Suppose we wish to compute the price of a contract $C$ at time $t$ that depends on an underlying stock price process governed by

$$dS_t = \mu(S_t, t)dt + \sigma(S_t, t)dB_t.$$ 

According to the risk-neutral pricing formula, we have:

$$C_t = e^{-r(T-t)}E[C(S_T) | \mathcal{F}_t],$$

where $C(S_T)$ is computable as a payoff function. How can we compute this conditional expectation? One possible method is to solve the linear non-homogenous, isotropic PDE:

$$L_t[f] + f_t(x, t) = rf(x, t), \quad x > 0, t \in [0, T]$$

with $BC \ f(x, T) = C(S_T), \ f(0, t) = 0, \ f(\infty, t) = S_T - K$ for all $t$, then $C_t = f(S_t, t)$.

17 Market Model Using Stochastic Calculus

Elements of the market model:

1. **Stock price model (risky assets)** The stock market now contains $d$ risky assets $(S^{(1)}_t, S^{(2)}_t, \ldots, S^{(d)}_t)$, which are assumed to evolve as per:

$$dS^{(i)}_t = \mu^{(i)} dt + \sum_{j=1}^{m} \sigma^{(i,j)}_t dW^{(j)}_t, \quad \text{for } i \in \{1, \ldots, d\}$$

$$dS_t = \mu_t + \sigma_t dW_t, \quad \text{where}$$

$$dS_t = [dS^{(1)}_t, dS^{(2)}_t, \ldots, dS^{(d)}_t]^T$$

$$m_t = [\mu^{(1)}_t, \mu^{(2)}_t, \ldots, \mu^{(d)}_t]^T$$

$$\sigma_t = \begin{pmatrix} \sigma^{(1,1)}_t & \cdots & \sigma^{(1,m)}_t \\ \vdots & \ddots & \vdots \\ \sigma^{(d,1)}_t & \cdots & \sigma^{(d,m)}_t \end{pmatrix}$$

$$dW_t = [dW^{(1)}_t, dW^{(2)}_t, \ldots, dW^{(m)}_t]^T,$$

all $\{W^{(i)}_t\}$ are independent and the initial condition $S^{(i)}_0 = s_i$ for all $i$.

2. **Bonds (riskless assets)**

$$B'_t = r_t B_t, \quad B_0 = 1.$$
3. Self-financing assumption. Call \((\psi_t, \phi_t)\) a trading strategy if it is \(\mathcal{F}_t\) measurable for all time. The portfolio which we hold now contains \(\psi_t\) units of bonds and \(\phi_t\) units of each stock at a given time \(t\):

\[
V_t = B_t \psi_t + \sum_{i=1}^{d} \phi_t^{(i)} S_t^{(i)} = B_t \psi_t + \phi_t^T S_t.
\]

By the product rule, we have:

\[
dV_t = \left( B_t d\psi_t + \sum_{i=1}^{d} d\phi_t^{(i)} S_t^{(i)} \right) + \psi_t dB_t + \sum_{i=1}^{d} \phi_t^{(i)} dS_t^{(i)}
\]

A portfolio is self-financing if \(B_t d\psi_t + \sum_{i=1}^{d} d\phi_t^{(i)} S_t^{(i)} = 0\), that is:

\[
V_{t+h} - V_t = \psi_t (B_{t+h} - B_t) + \sum_{i=1}^{d} \phi_t^{(i)} (S_{t+h}^{(i)} - S_t^{(i)})
\]

\[
dV_t = \psi_t dB_t + \sum_{i=1}^{d} \phi_t^{(i)} dS_t^{(i)}
\]

\[
V_T = V_0 + \int_{0}^{T} \psi_t r_t B_t dt + \sum_{i=1}^{d} \int_{0}^{T} \phi_t^{(i)} dS_t^{(i)}.
\]

In other words, changes in the value of the portfolio are solely due to changes in the value of the stock prices and bonds.

4. Arbitrage free assumption By definition, the market is arbitrage free if for every self-financing trading strategy \((\psi, \phi)\), \(V_0 = 0 \Rightarrow P(V_T > 0) = 0\). We know, however, that the market is arbitrage free if there exists a martingale measure \(Q\) for the discounted stock price process

\[
d\tilde{S}_t = d(B_t^{-1} S_t) = B_t^{-1} dS_t - r B_t^{-1} S_t dt = B_t^{-1} (\mu_t dt + \sigma_t d\mathcal{W}_t) - r B_t^{-1} S_t dt.
\]

\[
= B_t^{-1} (\mu_t - r S_t) dt + B_t^{-1} \sigma_t d\mathcal{W}_t = B_t^{-1} (\mu_t - r S_t) dt + B_t^{-1} \sigma_t (d\tilde{\mathcal{W}}_t - u_t dt)
\]

\[
= B_t^{-1} (\mu_t - r S_t - \sigma_t u_t) dt + B_t^{-1} \sigma_t d\mathcal{W}_t
\]

where \(\mathcal{W}_t\) is \((\mathcal{F}, \mathbb{P})\) \(d\)-dimensional BM,

\[
\tilde{\mathcal{W}}_t = \mathcal{W}_t + \int_{0}^{t} u_s ds.
\]
is \((\mathcal{F}, \mathbb{Q})\) \(d\)-dimensional BM with \(\mathbb{Q}\) defined via 
\[
\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp \left\{ -\frac{1}{2} \int_0^T u'_t u_t dt - \int_0^T u'_t dW_t \right\}.
\]
(We also assume \(\mathbb{E}[e^{\frac{1}{2} \int_0^T u'_t u_t dt}] < \infty\).)

So we need to check that there exists a solution \(u_t\) (market price of risk)
\[
\mu_t - r_t S_t = \sigma_t u_t.
\]

We can thus write:
\[
d\tilde{S}_t = B_t^{-1} \sigma_t d\tilde{W}_t
\]
let
\[
\tilde{V}_t = B_t^{-1} V_t = \psi_t + \phi'_t B_t^{-1} S_t = \psi_t + \phi'_t \tilde{S}_t
\]
\[
\Rightarrow d\tilde{V}_t = \phi'_t d\tilde{S}_t = B_t^{-1} \phi'_t \sigma_t d\tilde{W}_t.
\]
So \(\mathbb{E}_\mathbb{Q}[\tilde{S}_t | \mathcal{F}_s] = \tilde{S}_s\) for \(t \geq s\) and \(\mathbb{E}_\mathbb{Q}[\tilde{V}_t | \mathcal{F}_s] = \tilde{V}_s\) or \(V_t = B_t \mathbb{E}_\mathbb{Q}[B_T^{-1} V_T | \mathcal{F}_t]\).

5. Complete market. If the solution \(u_t\) is unique, then the market is complete (every claim can be replicated). A market is complete if every contingent claim can be replicated using a self-financing trading strategy on the underlying securities (such as portfolio \(V_t\)). So for every \(C_t = C(S_t, T - t, K)\) we should be able to write:
\[
C_T = V_T
\]
\[
= V_0 + \int_0^T d\tilde{V}_t
\]
\[
= V_0 + \int_0^T \psi_t dB_t + \sum_{i=1}^d \int_0^T \phi^{(i)}_t dS^{(i)}
\]
\[
B_t^{-1} V_t = \mathbb{E}_\mathbb{Q}[B_T^{-1} V_T | \mathcal{F}_t]
\]
\(\Rightarrow B_t^{-1} C_t = \mathbb{E}_\mathbb{Q}[B_T^{-1} C_T | \mathcal{F}_t]
\)
\[
\tilde{V}_t = \mathbb{E}_\mathbb{Q}[\tilde{V}_T | \mathcal{F}_t]
\]
\[
\tilde{V}_t = B_t^{-1} C_t = \tilde{C}_t
\]
\[
d\tilde{V}_t = d\tilde{C}_t \Rightarrow \]
\[
B_t^{-1} \phi'_t \sigma_t d\tilde{W}_t = B_t^{-1} \theta'_t d\tilde{W}_t, \quad \text{by martingale representation theorem}
\]
\[
\phi'_t \sigma_t = \theta'_t
\]
where \(C_t = B_t \tilde{C}_t = C_0 + B_t \int_t^T B_s^{-1} \theta'_t d\tilde{W}_s\)
should have solution if we are to price the option. In other words, varying the claim \( C \) is equivalent to varying \( \theta_t \) due to the one-to-oneness relation between \( C \) and \( \theta_t \). Thus, we must be able to solve \( \phi_t' \sigma_t = \theta_t' \) for every \( \theta_t \), which implies that \( \sigma_t \) must be a full rank matrix.

**Example 17.1**

Deriving the Black-Sholes Formula

Given:

\[
B_t = e^{rt} \\
dS_t = \mu S_t dt + \sigma S_t dW_t
\]

For this market model we have unique solution to:

\[
(\sigma S_t)u = (\mu S_t) - rS_t
\]

Since \( u = (\mu - r)/\sigma \), the market is arbitrage free and complete. Hence we can price any contingent claim using a self-financing trading strategy using the underlying securities. The price of EU Call can be computed from:

\[
C_0 = E_Q [e^{-rT} C_T | S_0] = E_Q [B_T^{-1} (S_T - K)^+ | S_0] \\
= \int_{e^{-rT}K}^{\infty} (x - e^{-rT}K) \ p(x) \ dx
\]

where \( p(x) \) is the pdf of \( \tilde{S}_T = e^{-rT} S_T \) under the measure \( Q \). Note that under \( Q \) we have \( d\tilde{S}_t = \sigma \tilde{S}_t dW_t \). Hence the solution is \( \tilde{S}_t = S_0 e^{\sigma W_t - \frac{1}{2} \sigma^2 t} \) and here \( \ln \tilde{S}_T \sim N(\ln(S_0) - \frac{1}{2} \sigma^2 T; \sigma^2 T) \equiv p(x) \). From this follows the Black-Sholes formula.

**Appendix: Hermite functions**

**Definition 17.1 (Hermite functions)**

\[
H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}, \quad n = 0, 1, 2, \ldots
\]

For example, \( H_0(x) = 1, \ H_1(x) = x, \ H_2(x) = x^2 - 1, \ H_3(x) = x^3 - 3x \) etc. Properties of the Hermite functions include:

1. \( H_{n+1}(x) = xH_n(x) - nH_{n-1}(x) \)

2. \( H_n'(x) = nH_{n-1}(x) \)
3. \( \sum_{n=0}^{\infty} \frac{t^n}{n!} H_n(x) = e^{xt - t^2/2} \)

4. \( \int_{\mathbb{R}} H_n(x)H_m(x)e^{-x^2/2}dx = \sqrt{2\pi n!}\delta_{nm}, \)

where \( \delta_{mn} \) is the kronecker delta function.

**Definition 17.2 (Normalized Hermite functions)**

\( h_n(x) = (2\pi)^{-1/4} ((n - 1)!)^{-1/2} e^{-x^2/2}H_{n-1}(\sqrt{2}x), \quad n = 1, 2, \ldots, \)

so that

\( \int_{\mathbb{R}} h_n(x)h_m(x)dx = \delta_{nm}. \)