Lecture notes on Stochastic Calculus and Quantitative Methods

Instructor: Jie LIU[1]

The lecture notes are based on chapters from the following books

- G. H. Choe, "Stochastic Analysis for Finance with Simulations", Springer, 2016. (electronic book available online from NUS library).
- Lishang Jiang, "Mathematical Modeling and Methods of Option Pricing", Translated by Canguo Li, World Scientific, 2005.
- B. Oksendal, "Stochastic Differential Equations: an Introduction with Applications", 6th edition, Springer, 2005.
- S. E. Shreve, "Stochastic Calculus for Finance I: The Binomial Asset Pricing Model", Springer, 2004.
- S. E. Shreve, "Stochastic Calculus for Finance II: Continuous-Time Models", Springer, 2004.

Approximate Schedule

	Date	Topics
1.	Aug 14	Option pricing and arbitrage-free principle
2.	Aug 19	Binomial tree method and Black-Scholes-Merton formula
3.	Aug 26	Finite probability space
4.	Sep 2	Brownian motion
5.	Sep 9	The Itô integral
6.	Sep 16	Midterm
7.	Sep 30	The Itô formula
8.	Oct 7	Stochastic differential equations
9.	Oct 14	Black-Scholes-Merton equation
10.	Oct 21	Risk neutral pricing
11.	Nov 4	Monte Carlo method
12.	Nov 11	Importance sampling
13.	Nov 30	Final

[1]email: rmiv71@nus.edu.sg

Grading policy:

- In-class quizzes 13% (The lowest grade will be dropped.)
- Homework 12% (4 homework assignments. No late homework will be accepted.)
- Midterm 25%, Final 50%. (Both are in-class, closed-book exams. Some of the exam problems may be taken from the homework or quizzes. The exams cannot be taken at any time other than the regularly scheduled time. Only scientific calculator is allowed in the exams.)

Guildline for the letter grades:

z-score ($= \frac{x-\mu}{\sigma}$)	grade
≥ 0.90	A,A+
$[0.40, 0.90)$	A-
$[-1.0, 0.40)$	B+
$[-1.5, -1.0)$	B
$[-2.0, -1.5)$	B-
$[-2.5, -2.0)$	C+
$[-3.0, -2.5)$	C
$[-3.5, -3.0)$	D+
$[-4.0, -3.5)$	D
< -4.0	F

A tentative schedule for the 4 homework assignments are:

1) Assignment 1: Due Aug 31 (Saturday), 11:55 pm. Homework I and Homework II (cover Chapters 1 and 2).

2) Assignment 2: Due Sep 14 (Saturday), 11:55 pm. Homework III and Homework IV (cover Chapters 3 and 4). Since I will upload the solution in the early morning of Sep 15 so that one can use it to prepare for one's midterm on Sep 16, no late homework is accepted.

3) Assignment 3: Due Oct 21 (Monday), 11:55 pm. Homework V, VI, VII (cover Chapters 5, 6, 7).

4) Assignment 4: Due Nov 16 (Saturday), 11:55 pm. Homework VIII, IX (cover Chapters 8 and 9).

Many homework problems already contain solutions. Those are for your own practice and you do not need to submit them.

Please solve homework problems that originally do not contain solutions. And put your solutions (photos, scanned pages, or other types of electronic files) into a **single** pdf file and name it as StudentID_Yourname_AssignmentNumber.pdf, before you submit it through LumiNUS.

Contents

1 Option pricing and arbitrage-free principle (1 lecture)

Chapter 1 of the lecture notes mainily follows Chapter 2 of "Mathematical Modeling and Methods of Option Pricing" by Lishang Jiang. Some table and examples are taken from "Options, Futures, and Other Derivatives" (10th edition) by John Hull and "Principles of Corporate Finance" (12th edition) by Brealey, Myers, and Allen.

1.1 Options and option pricing

An option is a contact that the holder can buy from, or sell to, the seller of the option a certain amount of underlying asset at a specified price (strike price) at any time on or before a given date (expiration date).

A call option gives the holder the right to buy the underlying asset. A put option gives the holder the right to sell.

The action to perform the buying or selling of the underlying asset according to the option contract is called exercise. European options can be exercised only at the expiration date. American options can be exercised on or prior to the expiration date. As a matter of fact, most equity options are American style while many index options are European style.

Let K be the strike price and T be the expiration date. Then an option's payoff (value) $ⓟ_T$ at expiration date is

$$ⓟ_T = (S_T - K)^+, \qquad \text{(call option)} \tag{1.1}$$

$$ⓟ_T = (K - S_T)^+, \qquad \text{(put option)} \tag{1.2}$$

where S_T denotes the price of the underlying asset at maturity time $t = T$ and $a^+ = \max(a, 0)$.

Example 1.1 *Suppose the price of certain stock is 66.6 USD on April 30, and the stock may go up in August. The investor pays 39,000 USD to purchase a call option to buy 10,000 shares at the strike price 68.0 USD per share on August 22.*

We consider two scenarios

A) The stock goes up to 73.0 USD on August 22. The investor exercises the option to receive a payoff $= (73 - 68) \times 10,000 = 50,000$ *USD. The return is*

$$return = \frac{50,000 - 39,000}{39,000} = 28.2\%.$$

B) The stock goes down to 66.0 USD on August 22. The investor lost the entire invested 39,000 USD.

As a derived security, the price of an option, denoted by $ⓟ_t$, varies with the price of its underlying asset (S_t). We want to find the function $V(t, S)$ of two variables such that

$$ⓟ_t = V(t, S_t) \qquad \text{for } t \in [0, T].$$

We know the value of $ⓟ_t$ when $t = T$:

$$ⓟ_T = V(T, S_T) = \begin{cases} (S_T - K)^+, & \text{(call option)} \\ (K - S_T)^+, & \text{(put option)} \end{cases} \tag{1.3}$$

The problem of option pricing is hence a backward problem.

Example 1.2 *A bull call spread can be made by buying a call option with a certain exercise price and selling a call option on the same stock with a higher exercise price. Both call options have the same expiration date. Consider a European call with an exercise price of* K_1 *and a second European call with an exercise price of* $K_2 > K_1$*. The following is the payoff table for this strategy if we ignore the premium (=option price) paid.*

strategy	$S_T \le K_1$	$K_1 < S_T \le K_2$	$K_2 < S_T$
A long call at K_1	*0*	$S_T - K_1$	$S_T - K_1$
A short call at K_2	*0*	0	$K_2 - S_T$
Total	*0*	$S_T - K_1$	$K_2 - K_1$

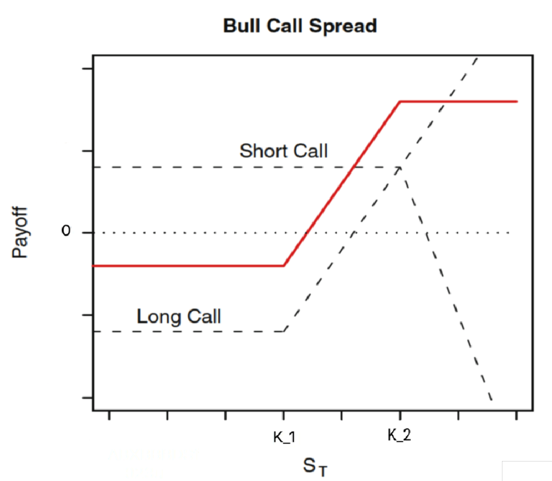


Figure 1.1: Payoff v.s. S_T. Note that the payoff is lowered by the option price paid. We will study option pricing later.

Please read "Options Markets" by John Cox and Mark Rubinstein if you want to learn more about how to use options as building blocks to create interesting payoff structures.

In the table in Example 1.2, longing an option (or stock) simply means that you have purchased the option (or stock). You can also short a stock by borrowing shares of stock (in most cases from a brokerage house), and sells these shares. Later, you will cover your short position by purchasing the same number of shares you have shorted and deliver these shares to your brokerage house to replace the shares you borrowed. If you establish a short option position you may be required to margin this position. This means that you must provide funds or securities as a guarantee since a short option position may entail an obligation and you must show that you are financially able to do so.

Options are traded both on exchanges and in the over-the-counter market. The following table gives the bid and ask quotes for some of the call options trading on Google (ticker symbol: GOOG) from Chicago Board Options Exchange website `http://www.cboe.com/delayedquote/quote-table`

GOOG (GOOGLE INC.)

Aug 08, 2019 @ 13:40 ET (Delayed) BID 1196.46008 ASK 1197.24011 VOL 726,635 LAST 1196.85669 CHANGE +22.8667 (+1.95%)

Filters by:

Volume: All Expiration Type: All Options Range: Near the Money Expiration: 2019 September View Chain

Options Chain

Total Records: 16

Calls — 09/06/2019 — Puts

Last	Net	Bid	Ask	Vol	IV	Delta	Gamma	Int	Strike	Last	Net	Bid	Ask	Vol	IV	Delta	Gamma	Int
0	0	31.8	36.5	0	0.2322	0.5418	0.0051	0	GOOG 1192.500	0	0	25.3	28.5	0	0.2207	-0.4564	0.0053	0
25.6	0	31.7	33	0	0.2279	0.5297	0.0052	5	GOOG 1195.000	19.1	0	25.6	30.4	0	0.2201	-0.4693	0.0053	3
32	+7.3	30.4	31.5	1	0.2267	0.5167	0.0052	1	GOOG 1197.500	31.1	0	26.7	32.4	0	0.2228	-0.4819	0.0053	5
29.42	+5.82	28.7	30.2	9	0.2248	0.5032	0.0052	14	GOOG 1200.000	31.4	-14.7	29.9	32.4	2	0.2245	-0.4958	0.0053	19

Calls — 09/13/2019 — Puts

Last	Net	Bid	Ask	Vol	IV	Delta	Gamma	Int	Strike	Last	Net	Bid	Ask	Vol	IV	Delta	Gamma	Int
0	0	35	40.5	0	0.2296	0.5446	0.0046	0	GOOG 1192.500	0	0	28	31.9	0	0.2213	-0.454	0.0048	0
0	0	33	38	0	0.2226	0.5338	0.0047	0	GOOG 1195.000	34.2	0	28.8	32.4	0	0.2175	-0.4659	0.0049	4
27.6	0	32.1	37.5	0	0.2261	0.5222	0.0047	4	GOOG 1197.500	0	0	30.2	35.9	0	0.226	-0.477	0.0047	0
32.15	+5.1	32.8	34	3	0.2249	0.5104	0.0047	23	GOOG 1200.000	34.58	-13.77	33.7	37.2	1	0.2329	-0.4891	0.0045	0

The column Net reports the difference from the current business day's last reported sales price, and a previous business day's last reported sales price. Since the Google's stock price $\uparrow$, the call price $\uparrow$, while the put price $\downarrow$. Since the options are American style which allows early exercise, it is obvious that as $T \uparrow$, both prices should $\uparrow$ which is also confirmed from the above table.

The Bid and Ask prices are the prices for you to sell or buy an option respectively. The trading Volume (Vol) is the number of contacts (usually 100 shares per contract) traded in that day. Open Interest (Int) is the number of contact outstanding, that is, the number of long positions or, equivalently, the number of short positions. Open interest decreases as open trades are closed. The higher the volume and the open interest, the more liquid the option is thought to be. We will learn the meanings of Implied Volatility (IV), Delta, and Gamma later.

1.2 Financial market and arbitrage-free principle

There are two basic laws of economics: (1) there can be no reward without risk, and (2) gaining an advantage over skilled and knowledgeable competitors in a free market is extraordinarily difficult.

Consider a financial **market** consisted of a risk-free asset B (e.g. bond or money market account) and n risky assets (stocks, options, ...) S_i, $i = 1, ..., n$. Let

$$V_t(B) = B_t, \qquad V_t(S_i) = S_{it}$$

be the values of the bound and risky asset i at time t ($i = 1, ..., n$). In time period $[t_0, t_1]$. The corresponding payoffs are $B_{t_1} - B_{t_0}$ and $S_{it_1} - S_{it_0}$, and returns are $\frac{B_{t_1} - B_{t_0}}{B_{t_0}}$ and $\frac{S_{it_1} - S_{it_0}}{S_{it_0}}$, $(i = 1, ..., n)$.

A portfolio Φ is

$$\Phi = \alpha B + \sum_{i=1}^{n} \phi_i S_i, \tag{1.4}$$

where $(\alpha, \phi_1, ..., \phi_n) \in \mathbb{R}^{n+1}$ and is called the investment strategy. In general, $\alpha, \phi_1, ..., \phi_n$ are functions of time t, as the investor may adjust the investment strategy over the time. However, $\alpha, \phi_1, ..., \phi_n$ are not to be changed between two adjacent transaction times $t = t_m$, $t = t_{m+1}$.

The value of the portfolio Φ at time t, which is denoted by $V_t(\Phi)$ or Φ_t, is

$$\Phi_t = V_t(\Phi) = \alpha_t B_t + \sum_{i=1}^{n} \phi_{it} S_{it}. \tag{1.5}$$

$V_{t_{m+1}}(\Phi) - V_{t_m}(\Phi)$ is the portfolio Φ's profit over time period $[t_m, t_{m+1}]$ under the investment strategy $(\alpha_{t_m}, \phi_{1t_m}, ..., \phi_{nt_m})$:

$$V_{t_{m+1}}(\Phi) - V_{t_m}(\Phi) = \alpha_{t_m}[B_{t_{m+1}} - B_{t_m}] + \sum_{i=1}^{n} \phi_{it_m}[S_{it_{m+1}} - S_{it_m}]. \tag{1.6}$$

If during the entire transaction period $[0, T]$, the investor does not add or withdraw fund, then the entire transaction process is said to be self-financing. It means the current portfolio value is precisely the initial investment plus the any trading gains.

Definition 1.1 *A self-financing investment strategy Φ is said to have arbitrage opportunity in $[0, T]$, if there exists $T^* \in [0, T)$, such that*

$$V_{T^*}(\Phi) = 0$$

and

$$V_T(\Phi) \geq 0,$$

and

$$\mathbb{P}(V_T(\Phi) > 0) > 0,$$

where $\mathbb{P}(A)$ denotes the probability of event A.

Definition 1.2 *If there exists no arbitrage opportunity for any self-financing investment strategy Φ in $[0, T]$, then the market is said to be arbitrage-free in the time period $[0, T]$.*

Theorem 1.1 *If the market is arbitrage-free in a time period $[0, T]$, and Φ_1 and Φ_2 are portfolios satisfying*

$$V_T(\Phi_1) \geq V_T(\Phi_2), \tag{1.7}$$

and

$$\mathbb{P}(V_T(\Phi_1) > V_T(\Phi_2)) > 0, \tag{1.8}$$

then

$$V_t(\Phi_1) > V_t(\Phi_2) \qquad \textit{for any } t \in [0, T). \tag{1.9}$$

Remark: In particular, if $V_T(\Phi_1) > V_T(\Phi_2)$, then $V_t(\Phi_1) > V_t(\Phi_2)$ for any $t \in [0, T]$.
Proof: We prove by contradiction. If (1.9) is false, then there exists a $t^* \in [0, T)$ such that

$$V_{t^*}(\Phi_1) \leq V_{t^*}(\Phi_2).$$

Now, we define

$$E = V_{t^*}(\Phi_2) - V_{t^*}(\Phi_1) \geq 0 \tag{1.10}$$

and construct a portfolio Φ_c at $t = t^*$

$$\Phi_c = \Phi_1 - \Phi_2 + B$$

where B is the risk-free asset (bond) of the market that satisfies $B_{t^*} = E$. We now claim that Φ_c has arbitrage opportunity in $[t^*, T]$. This leads to contradiction since the market is assumed to be arbitrage-free. Hence our initial assumption is not correct and we proves (1.9) by contradiction.

We are left to prove the claim that Φ_c has arbitrage opportunity. Note that

$$V_{t^*}(\Phi_c) = V_{t^*}(\Phi_1) - V_{t^*}(\Phi_2) + B_{t^*} \overset{(1.10) \text{ and } B_{t^*}=E}{=} 0$$

$$V_T(\Phi_c) = V_T(\Phi_1) - V_T(\Phi_2) + B_T \overset{(1.7)}{\geq} 0$$

"if event $V_T(\Phi_1) > V_T(\Phi_2)$ happens, then event $V_T(\Phi_c) > 0$ happens"

$$\overset{(1.8)}{\Rightarrow} \mathbb{P}(V_T(\Phi_c) > 0) > 0. \qquad \square$$

By a similar argument, we can prove (Question 3 of Homework I) the following theorem

Theorem 1.2 *If the market is arbitrage-free in time period* $[0, T]$*, and* Φ_1 *and* Φ_2 *are portfolios satisfying*

$$V_T(\Phi_1) \geq V_T(\Phi_2), \tag{1.11}$$

then

$$V_t(\Phi_1) \geq V_t(\Phi_2) \qquad \textit{for any } t \in [0, T). \tag{1.12}$$

Corollary 1.1 *If the market is arbitrage-free in time period* $[0, T]$*, and at time* T

$$V_T(\Phi_1) = V_T(\Phi_2),$$

then

$$V_t(\Phi_1) = V_t(\Phi_2) \qquad \textit{for any } t \in [0, T). \tag{1.13}$$

Proof: Apply Theorem 1.2 to $V_T(\Phi_1) \geq V_T(\Phi_2)$ to get $V_t(\Phi_1) \geq V_t(\Phi_2)$. Apply Theorem 1.2 to $V_T(\Phi_2) \geq V_T(\Phi_1)$ to get $V_t(\Phi_2) \geq V_t(\Phi_1)$. Hence we have (1.13). □

1.3 European option pricing and put-call parity

We now assume

1. The market is arbitrage-free.
2. All transactions are free of charge.
3. The risk-free interest rate r is a constant.
4. The underlying asset pays no dividends.

Consider the following notation

- S_t: the risky asset price.
- c_t: European call option price [2].
- p_t: European put option price.
- C_t: American call option price.
- P_t: American put option price.
- K: the option's strike price.
- T: the option's expiration date.

[2]Here, the option price is the price to buy or sell 1 share of stock.

- r: the risk-free interest rate [3].

For a zero-coupon bond with face value $K_0 = Ke^{-rT}$ and compound interest rate r, its value, denoted by B_t, satisfies $\frac{dB_t}{dt} = rB_t$ [4], or

$$B_t = K_0 e^{rt} = Ke^{-r(T-t)}. \tag{1.14}$$

Theorem 1.3 *For European option pricing, the following valuations are true when $t \in [0, T)$:*

$$(S_t - Ke^{-r(T-t)})^+ < c_t < S_t, \tag{1.15}$$

$$(Ke^{-r(T-t)} - S_t)^+ < p_t < Ke^{-r(T-t)}. \tag{1.16}$$

Example 1.3 *($c_0 > (S_0 - Ke^{-rT})^+$). Suppose that $S_0 = \$20$, $K = \$18$, $r = 10\%$ per annum, and $T = 1$ year. If the European call price is \$3, an arbitrageur can short the stock and long the call. His cash inflow is $\$20 - \$3 = \$17$. If invested for 1 year at 10% per annum, the \$17 grows to $17e^{0.1} = \$18.79$. At the end of the year, the option expires. If the stock price is greater than \$18, the arbitrageur exercises the option for \$18, closes the short position in the stock, and makes a profit of*

$$\$18.79 - \$18 = \$0.79.$$

If the stock price is less than \$18, say, \$17, a stock is bought in the market and the short position is closed. The arbitrageur's profit is

$$\$18.79 - \$17 = \$1.79.$$

One can use a portfolio to describe the above process: the arbitrageur build a portfolio $\Phi = c - S + B$ with $B_0 = 17$ which ensures $\Phi_0 = 0$. Then one can show that $\Phi_T > 0$ no matter $S_T \geq 18$ or $S_T < 18$.

Proof of Theorem 1.3: Let us first prove $\underbrace{S_t}_{\Phi_2} < \underbrace{c_t + Ke^{-r(T-t)}}_{\Phi_1}$. We consider two portfolios at $t = 0$:

$$\Phi_1 = c + B, \qquad \Phi_2 = S$$

[3] Instead of the interest rates implies by Treasury bills, traditionally derivatives dealers use LIBOR (London Interbank Offered Rate) rates as risk-free interest rate. But LIBOR rates are not totally risk-free. Following the credit crisis that started in 2007, many dealers switched to using overnight indexed swap (OIS) rates as risk-free rates, at least for collateralized transactions. An OIS allows overnight borrowing or lending between financial institutions for a period to be swapped for borrowing or lending at a fixed rate for the period. The fixed rate in an OIS is referred to as the OIS rate. See John Hull "Options, Futures and Other Derivatives", 10th edition, §4.2, 4.3 for more details.

[4] which can be written as $\frac{dB_t}{B_t} = rdt$, saying that the return $\frac{B_{t+\Delta t} - B_t}{B_t}$ is proportional to Δt with a constant coefficient r.

where the value of the bond B at time t is $B_t = Ke^{-r(T-t)}$. Then

$$V_T(\Phi_1) = V_T(c) + V_T(B) = (S_T - K)^+ + K = \left\{ \begin{array}{ll} S_T, & \text{if } S_T \geq K \\ K, & \text{if } S_T < K \end{array} \right\} \geq S_T = V_T(\Phi_2).$$

Moreover,

$$\mathbb{P}(V_T(\Phi_1) > V_T(\Phi_2)) = \mathbb{P}(S_T < K) > 0.$$

Hence by Theorem 1.1, for all $t \in [0, T)$,

$$V_t(\Phi_1) > V_t(\Phi_2),$$

i.e., $c_t + Ke^{-r(T-t)} > S_t$ or $c_t > S_t - Ke^{-r(T-t)}$. But it is obvious that $c_t > 0$, hence

$$c_t > \max(S_t - Ke^{-r(T-t)}, 0) = (S_t - Ke^{-r(T-t)})^+.$$

This proves the first half of (1.15). The rest can be proved similarly and is skipped. □

Theorem 1.4 *(put-call parity)*

$$c_t + Ke^{-r(T-t)} = p_t + S_t. \tag{1.17}$$

Proof: Construct two portfolios at $t = 0$,

$$\Phi_1 = c + Ke^{-rT}, \qquad \Phi_2 = p + S,$$

and then consider their values at $t = T$:

$$\begin{aligned} V_T(\Phi_1) &= V_T(c) + V_T(Ke^{-rT}) = (S_T - K)^+ + K = \max(K, S_T), \\ V_T(\Phi_2) &= V_T(p) + V_T(S) = (K - S_T)^+ + S_T = \max(K, S_T). \end{aligned}$$

By Corollary 1.1, we have

$$V_t(\Phi_1) = V_t(\Phi_2) \qquad \forall t \leq T.$$

This proves (1.17). □

Example 1.4 *(put-call parity and capital structure) (This part is taken from §11.5 of John Hull's "Options, Futures and Other Derivatives", 10th edition.) Fischer Black, Myron Scholes*[5]*, and Robert Merton were the pioneers of option pricing. In the early 1970s, they also showed that options can be used to characterize the capital structure of a company. Today this analysis is widely used by financial institutions to assess a company's credit risk.*

To illustrate the analysis, consider a simple situation where a company has assets that are financed with zero-coupon bonds and equity. The bonds mature in five years at which time a principal payment of K is required. The company pays no dividends. If the assets are

[5] The Pricing of Options and Corporate Liabilities, The Journal of Political Economy, Vol. 81, 1973, 637–654.

worth more than K in five years, the equity holders choose to repay the bond holders. If the assets are worth less than K, the equity holders choose to declare bankruptcy and the bond holders end up owning the company.

$$\textit{The value of the equity in five years is therefore } \max(A_T - K, 0),$$

where A_T is the value of the company's assets at that time. This shows that the equity holders have a five-year European call option on the assets of the company with a strike price of K. What about the bond holders? They get $\min(A_T, K)$ in five years. Simple mathematics shows that $\min(A_T, K) = K - \max(K - A_T, 0)$. Hence we can consider the value of the bonds in five years as $V_T(\Phi)$ with $\Phi =$ a risk-free asset$-$ a European put option p. This shows that **today** *the bonds are worth $V_0(\Phi)$ which is the present value of K minus the value of a five-year European put option on the assets with a strike price of K.*

To summarize, if c and p are the values, respectively, of five-year call and put options on the company's assets with strike price K, then

$$\begin{aligned}
&\textit{Value of company's equity} = c \\
&\textit{Value of company's debt} = \textit{PresentValue}(K) - p = Ke^{-rT} - p.
\end{aligned}$$

Denote the value of the assets of the company today by A_0. The value of the assets must equal the total value of the instruments used to finance the assets. This means that it must equal the sum of the value of the equity and the value of the debt, so that

$$A_0 = c + [Ke^{-rT} - p].$$

Rearranging this equation, we have

$$c + Ke^{-rT} = p + A_0.$$

This is the put-call parity result for call and put options on the assets of the company.

1.4 American option pricing and early exercise

Theorem 1.5 *If the market is arbitrage-free, then for all $t \in [0, T]$, there must be*

$$C_t \geq (S_t - K)^+, \tag{1.18}$$
$$P_t \geq (K - S_t)^+. \tag{1.19}$$

Proof: Because the proofs are similar, we only prove (1.18). Suppose (1.18) is not true. Then, there is a $t \in [0, T]$ such that

$$\underbrace{0 < C_t}_{\text{easy to prove}} < \max(S_t - K, 0).$$

From $\max(S_t - K, 0) > 0$, we conclude $S_t - K > 0$. Hence

$$C_t < S_t - K.$$

Then a trader can borrow money C_t from the bank to buy the American call option at time t, and then immediately exercise the option. The cash flow at time t is $S_t - K - C_t > 0$. Thus the trader gains a riskless profit immediately. This contradict with the assumption that the market is arbitrage-free. Hence (1.18) is true. □

Theorem 1.6 *If a stock S does not pay dividend, then*

$$C_t = c_t \qquad \forall t \in [0, T], \tag{1.20}$$

i.e., the "early exercise" term is of no use for American call option on a non-dividend-paying stock.

Proof: Since American option can be early exercised, its gaining opportunity must be no less than that of the European option. Therefore, $C_t \geq c_t$. [6] By the first half of (1.15), if $t \in [0, T)$,

$$C_t \geq c_t > (S_t - Ke^{-r(T-t)})^+ \geq \underbrace{(S_t - K)^+}_{\text{profit if exercise } C \text{ at } t}$$

(recall $-Ke^{-r(T-t)} \geq -K$). This indicates that it is unwise to **early** exercise the American call option C. □

Remark: For the put option, we do not have $P_t = p_t$. For example, if at time t, $S_t < K(1 - e^{-r(T-t)})$, then the holder should exercise the put option immediately. The immediate gain is

$$K - S_t > K - K(1 - e^{-r(T-t)}) = Ke^{-r(T-t)},$$

and by depositing the gain in a saving's account, the total payoff will exceed K at $t = T$. This should be compared with the payoff at the option's expiration date T, which will never exceed K in any case.

Recall the put-call parity (1.17) which says that $c_t - p_t = S_t - Ke^{-r(T-t)}$. We have the following relation for C_t and P_t (Theorem 2.6 of Jiang, and (11.7) of John Hull, 10th editon):

Theorem 1.7 *If C, P are non-dividend-paying American call option and put option respectively, then*

$$S_t - K \leq C_t - P_t \leq S_t - Ke^{-r(T-t)} \tag{1.21}$$

Proof: The right hand side of (1.21) is easy to prove:

$$\underbrace{P_t \geq p_t}_{\text{obvious}} = c_t + Ke^{-r(T-t)} - S_t \overset{(1.20)}{=} C_t + Ke^{-r(T-t)} - S_t. \tag{1.22}$$

To prove the left hand side of (1.21), we construct two portfolios at time t

$$\Phi_1 = C + K, \qquad \Phi_2 = P + S.$$

[6] For the same reason, we have $P_t \geq p_t$.

If during the period $[t,T)$, the American put option P is not early exercised, then

$$V_T(\Phi_1) = (S_T - K)^+ + Ke^{r(T-t)},$$
$$V_T(\Phi_2) = (K - S_T)^+ + S_T.$$

Therefore,

$$\begin{cases} \text{when } K \geq S_T, & V_T(\Phi_1) = Ke^{r(T-t)} > K = V_T(\Phi_2) \\ \text{when } K < S_T, & V_T(\Phi_1) = S_T + K(e^{r(T-t)} - 1) > S_T = V_T(\Phi_2) \end{cases}$$

If the American put option P is early exercised at some time $\tau \in [t,T)$, then

$$V_\tau(\Phi_1) = C_\tau + Ke^{r(\tau-t)} \overset{(1.18)}{\geq} (S_\tau - K)^+ + Ke^{r(\tau-t)} = \begin{cases} S_\tau + K(e^{r(\tau-t)} - 1) & \text{if } S_\tau > K \\ Ke^{r(\tau-t)} & \text{if } S_\tau < K \end{cases},$$

$$V_\tau(\Phi_2) = \underbrace{(K - S_\tau)^+}_{\text{by early exercise assumption}} + S_\tau = \begin{cases} S_\tau & \text{if } S_\tau > K \\ K & \text{if } S_\tau < K \end{cases}.$$

Hence $V_\tau(\Phi_1) \geq V_\tau(\Phi_2)$.

In any case, by Theorem 1.1,

$$V_t(\Phi_1) \geq V_t(\Phi_2).$$

This proves the left hand side of (1.21). □

Example 1.5 *An American call option on a non-dividend-paying stock with strike price \$20.00 and maturity in 5 months is worth \$1.50. Suppose that the current stock price is \$19.00 and the risk-free interest rate is 10% per annum. From (1.21) with $t = 0$, we have*

$$19 - 20 \leq C_0 - P_0 \leq 19 - 20e^{-0.1\times 5/12}$$

or $1 \geq P_0 - C_0 \geq 0.18$. With C_0 at \$1.50, P_0 must lie between \$1.68 and \$2.50.

Now we mention, without proving, two theorems that show the dependence of option pricing on the strike price. They are stated for your information only and will not be tested.

Theorem 1.8 *If $K_1 > K_2$, $\theta \in [0,1]$ and $\alpha > 0$, then*

$$0 \leq c_t(K_2) - c_t(K_1) \leq K_1 - K_2, \tag{1.23}$$
$$0 \leq p_t(K_1) - p_t(K_2) \leq K_1 - K_2, \tag{1.24}$$
$$c_t(\theta K_1 + (1-\theta)K_2) \leq \theta c_t(K_1) + (1-\theta)c_t(K_2), \tag{1.25}$$
$$p_t(\theta K_1 + (1-\theta)K_2) \leq \theta p_t(K_1) + (1-\theta)p_t(K_2). \tag{1.26}$$

The same is true with c_t replaced by C_t and p_t replaced by P_t.

Theorem 1.9

$$c_t(\alpha S_t, \alpha K) = \alpha c_t(S_t, K), \tag{1.27}$$
$$p_t(\alpha S_t, \alpha K) = \alpha p_t(S_t, K). \tag{1.28}$$

The same is true with c_t replaced by C_t and p_t replaced by P_t.

Remark: A stock split occurs when the existing shares are "split" into more shares. For example, in a 3-for-1 stock split, three new shares are issued to replace each existing share. The 3-for-1 stock split should cause the stock price to go down to 1/3 of its previous value. By (1.27) and (1.28) with $\alpha = 1/3$, we should reduce the strike price to 1/3 of its previous value, and the number of shares covered by one contract should be increased to 3 times of its previous amount.

In practice, after an n-for-m stock split, the stock price as well as the strike price are reduced to m/n of their previous values ($\alpha = m/n$), and the number of shares covered by one contract is increased to n/m of its previous value.

1.5 Real options and why do we need a pricing formula

Material in this section is taken from §22-1, The Value of Follow-On Investment Opportunities, from "Principles of Corporate Finance" 12th edition by Brealey, Myers, and Allen. It won't be tested. It is presented merely to convince you why we need a pricing formula even though the price of a stock option is actually determined by the supply and demand in the market. It is 1982. You are assistant to the chief financial officer (CFO) of Blitzen Computers, an established computer manufacturer casting a profit-hungry eye on the rapidly developing personal computer market. You are helping the CFO evaluate the proposed introduction of the Blitzen Mark I Micro.

Firms can best help their shareholders by accepting all projects that are worth more than they cost. In other words, they need to seek out projects with positive net present values (NPV). To find net present value we have to calculate present value of future cash flows. Just discount future cash flows by an appropriate rate r, usually called the discount rate, hurdle rate, or opportunity cost of capital. It is really an opportunity cost of capital because it depends on the investment opportunities available to investors in financial markets. Whenever a corporation invests cash in a new project, its shareholders lose the opportunity to invest the cash on their own.

$$NPV = C_0 + \frac{C_1}{1+r} + \frac{C_2}{(1+r)^2} + \frac{C_3}{(1+r)^3} + \ldots\ldots$$

Here, $(C_0, C_1, C_2, \cdots)$ is the stream of cash flows at (year 0, year 1, year 2, $\cdots$).

The discount rate r is determined by rates of return prevailing in financial markets. If the future cash flow is absolutely safe, then the discount rate is the interest rate on safe securities such as U.S. government debt. If the future cash flow is uncertain, then the expected cash flow should be discounted at the expected rate of return offered by equivalent-risk securities.

TABLE 22.1 Summary of cash flows and financial analysis of the Mark I microcomputer ($ millions).

	Year					
	1982	1983	1984	1985	1986	1987
After-tax operating cash flow (1)		+110	+159	+295	+185	0
Capital investment (2)	450	0	0	0	0	0
Increase in working capital (3)	0	50	100	100	–125	–125
Net cash flow (1) – (2) – (3)	–450	+60	+59	+195	+310	+125
NPV at 20% = –$46.45, or about –$46 million						

Cash flows are discounted for two simple reasons: because (1) a dollar today is worth more than a dollar tomorrow and (2) a safe dollar is worth more than a risky one.

For our example,

$$NPV = -450 + \frac{60}{1.2} + \frac{59}{1.2^2} + \frac{195}{1.2^3} + \frac{310}{1.2^4} + \frac{125}{1.2^5} \approx -46.45.$$

TABLE 22.2 Valuing the option to invest in the Mark II microcomputer.

Assumptions

1. The decision to invest in the Mark II must be made after three years, in 1985.
2. The Mark II investment is double the scale of the Mark I (note the expected rapid growth of the industry). Investment required is $900 million (the exercise price), which is taken as fixed.
3. Forecasted cash inflows of the Mark II are also double those of the Mark I, with present value of $807 million in 1985 and $807/(1.2)^3 = \$467$ million in 1982.
4. The future value of the Mark II cash flows is highly uncertain. This value evolves as a stock price does with a standard deviation of 35% per year. (Many high-technology stocks have standard deviations higher than 35%.)
5. The annual interest rate is 10%.

Interpretation

The opportunity to invest in the Mark II is a three-year call option on an asset worth $467 million with a $900 million exercise price.

Valuation

$$\text{PV(exercise price)} = \frac{900}{(1.1)^3} = 676$$

$$\text{Call value} = [N(d_1) \times P] - [N(d_2) \times \text{PV(EX)}]$$

$$d_1 = \log[P/\text{PV(EX)}]/\sigma\sqrt{t} + \sigma\sqrt{t}/2$$

$$= \log[.691]/.606 + .606/2 = -.3072$$

$$d_2 = d_1 - \sigma\sqrt{t} = -.3072 - .606 = -.9134$$

$$N(d_1) = .3793,\ N(d_2) = .1805$$

$$\text{Call value} = [.3793 \times 467] - [.1805 \times 676] = \$55.1 \text{ million}$$

Here the option is to buy a nontraded real asset, the Mark II. We cannot observe the Mark II's value; we have to compute it. The Mark II's forecasted cash flows are set out in Table 22.3. The project involves an initial outlay of $900 million in 1985. The cash inflows start in the following year and have a present value of \$807 $(= \frac{120}{1.2} + \frac{118}{1.2^2} + \frac{390}{1.2^3} + \frac{620}{1.2^4} + \frac{250}{1.2^5})$ million in 1985, equivalent to \$467 $(= \frac{807}{1.2^3}$ or equivalently $= \frac{120}{1.2^4} + \frac{118}{1.2^5} + \frac{390}{1.2^6} + \frac{620}{1.2^7} + \frac{250}{1.2^8})$

million in 1982 as shown in Table 22.3. So the real option to invest in the Mark II amounts to a three-year call on an underlying asset worth \$467 million, with a \$900 million exercise price. We evaluate it in Table 22.2 using the Black-Scholes formula that we will learn in Chapter 2 and then in Chapter 6 again.

Table 22.2 uses a standard deviation of 35% per year. Where does that number come from? We recommend you look for comparables, that is, traded stocks with business risks similar to the investment opportunity. For the Mark II, the ideal comparables would be growth stocks in the personal computer business, or perhaps a broader sample of high-tech growth stocks. Use the average standard deviation of the comparable companies' returns as the benchmark for judging the risk of the investment opportunity.

The NPV of the Mark I project is −\$46 million, but it creates the expansion option for the Mark II. The expansion option is worth \$55 million, so

$$\text{Adjusted Present Value} = -46 + 55 = \$9 \text{ million.}$$

	Year						
	1982		1985	1986	1987	1988	1989	1990
After-tax operating cash flow				+220	+318	+590	+370	0
Increase in working capital				100	200	200	−250	−250
Net cash flow				+120	+118	+390	+620	+250
Present value at 20%	+467	←	+807					
Investment, PV at 10%	676	←	900					
	(PV in 1982)							
Forecasted NPV in 1985			−93					

TABLE 22.3 Cash flows of the Mark II microcomputer, as forecasted from 1982 (\$ millions).

1.6 Homework I

(Only submit solutions to Questions 1,4,5,9.)

1. A straddle is a portfolio with long positions in a European call and a European put with the same strike price, maturity, and underlying. The straddle is seen to benefit from a movement in either direction away from the strike price. Show that the payoff of a straddle is $|S_T - K|$ if we ignore the premium by constructing the payoff table as in Example 1.2.

2. Suppose that an amount A is invested for n years at an interest rate of R per annum. If the rate is compounded m times per annum, the terminal value of the investment is

$$A\left(1+\frac{R}{m}\right)^{mn}.$$

If the interest rate is 10% and is measured with semiannual compounding, What is the value of \$100 at end of 1 year?

Determine the limit

$$\lim_{m\to\infty} A\left(1+\frac{R}{m}\right)^{mn}.$$

Solution:

$$\$100 \times (1+0.05)^2 = \$110.25.$$

Recall $\lim_{x\to\infty}(1+\frac{1}{x})^x = e$ and $\lim_{x\to\infty}(f(x))^k = (\lim_{x\to\infty} f(x))^k$. Hence

$$\lim_{m\to\infty} A\left(1+\frac{R}{m}\right)^{mn} = A\lim_{\frac{m}{R}\to\infty}\left(1+\frac{R}{m}\right)^{\frac{m}{R}Rn} = A\left(\lim_{\frac{m}{R}\to\infty}\left(1+\frac{R}{m}\right)^{\frac{m}{R}}\right)^{Rn} = Ae^{Rn}.$$

3. Prove Theorem 1.2. [Hint: Prove by contradiction. If the conclusion is false, one can construct a portfolio $\Phi_c = \Phi_1 - \Phi_2 + B$ which has arbitrage opportunity. You need to specify the value of B_t at some $t = t^*$.]

Proof: We prove by contradiction. If the conclusion is false, then there exists a $t^* \in [0, T)$ such that

$$V_{t^*}(\Phi_1) < V_{t^*}(\Phi_2).$$

Now, we define

$$E = V_{t^*}(\Phi_2) - V_{t^*}(\Phi_1) > 0 \tag{1.29}$$

and construct a portfolio Φ_c at $t = t^*$

$$\Phi_c = \Phi_1 - \Phi_2 + B$$

where B is the risk-free asset (bond) of the market that satisfies $B_{t^*} = E$. We now claim that Φ_c has arbitrage opportunity in $[t^*, T]$. This proves (1.12).

To prove the claim, note that

$$V_{t^*}(\Phi_c) = V_{t^*}(\Phi_1) - V_{t^*}(\Phi_2) + B_{t^*} = 0 \text{ by (1.29) and } B_{t^*} = E$$
$$V_T(\Phi_c) = V_T(\Phi_1) - V_T(\Phi_2) + B_T > 0 \text{ by (1.11)}. \qquad \square$$

4. Consider a European put option on a non-dividend-paying stock when the stock price is \$38, the strike price is \$40, the time to maturity is 3 months, and the risk-free rate of interest is 10% per annum. Find a lower bound for the option price.

5. The price of a non-dividend paying stock is \$19 and the price of a three-month European call option on the stock with a strike price of \$20 is \$1. The risk-free rate is 4% per annum. What is the price of a three-month European put option with a strike price \$20?

6. We want to prove (1.25): For any $K_0, K_1 \geq 0$, if $\theta \in [0,1]$, $c_t(\theta K_1 + (1-\theta)K_0) \leq \theta c_t(K_1) + (1-\theta)c_t(K_0)$. To do that, we construct two portfolios at $t = 0$:

$$\Phi_1 = \theta c(K_1) + (1-\theta)c(K_0), \qquad \Phi_2 = c(\theta K_1 + (1-\theta)K_0).$$

   Prove that $V_T(\Phi_1) \geq V_T(\Phi_2)$. Then (1.25) follows from Theorem 1.2.

   Proof: On the expiration date $t = T$,

$$V_T(\Phi_1) = \theta(S_T - K_1)^+ + (1-\theta)(S_T - K_0)^+, \tag{1.30}$$
$$V_T(\Phi_2) = (S_T - \theta K_1 - (1-\theta)K_0)^+. \tag{1.31}$$

   Without loss of generality, we can assume $K_1 \geq K_0$. Let $K_\theta = \theta K_1 + (1-\theta)K_0$. $K_1 \geq K_\theta \geq K_0$. There are 4 cases:

   1. $S_T \geq K_1$: $V_T(\Phi_1) = S_T - K_\theta = V_2(\Phi_2)$.
   2. $K_\theta \leq S_T < K_1$: $V_T(\Phi_1) = (1-\theta)(S_T - K_2)$. $V_T(\Phi_2) = (S_T - K_\theta) = \theta(S_T - K_1) + (1-\theta)(S_T - K_2) \leq V_T(\Phi_1)$.
   3. $K_0 \leq S_T < K_\theta$: $V_T(\Phi_1) = (1-\theta)(S_T - K_2)$. $V_T(\Phi_2) = 0 \leq V_T(\Phi_1)$.
   4. $S_T < K_0$: $V_T(\Phi_1) = 0 = V_2(\Phi_2)$.

   Hence $V_T(\Phi_1) \geq V_T(\Phi_2)$. $\square$

7. A 1-month European put option on a non-dividend-paying stock is currently selling for \$2.50. The stock price is \$47, the strike price is \$50, and the risk-free interest rate is 6% per annum. What opportunities are there for an arbitrageur?

   Solution: For European put option, if the market is arbitrage-free, we should have (1.16)

$$(Ke^{-rT} - S_0)^+ < p_0 < Ke^{-rT}$$

   which implies

$$(50e^{-0.06\times(1/12)} - 47)^+ < p_0 < 50e^{-0.06\times(1/12)},$$

$$49.75 - 47 < p_0 < 49.75 \quad \Rightarrow \quad 2.75 < p_0 < 49.75.$$

Hence $(Ke^{-rT} - S_0)^+ < p_0$ is violated. Since the European put option is undervalued, the arbitrageur should buy the put option. Then he also need to buy the stock so that he can sell the stock to the option seller to close out. Hence an arbitrageur can build a portfolio

$$\Phi = p + S - B$$

by buying the put option, buying the stock, and borrowing B_0 dollars from the money market at $t = 0$. (This portfolio can also be seen from the condition $p_0 + S_0 - Ke^{-rT} < 0$.) The value of B_0 is determined by the requirement that

$$V_0(\Phi) = 0 = p_0 + S_0 - B_0.$$

Then

$$V_T(\Phi) = (K - S_T)^+ + S_T - B_0e^{rT} = (K - S_T)^+ + S_T - (p_0 + S_0)e^{rT}$$

Right now, $p_0 < (Ke^{-rT} - S_0)$. $-p_0e^{rT} > -(Ke^{-rT} - S_0)e^{rT} = -(K - S_0e^{rT})$. Hence

$$V_T(\Phi) > (K - S_T)^+ + S_T - (K - \cancel{S_0e^{rT}}) - \cancel{S_0e^{rT}} \geq 0.$$

In the last step, we have used the fact that $a^+ - a = \max(a, 0) - a \geq 0$.

If one uses $p_0 = (Ke^{-rT} - S_0) - \varepsilon$ with $\varepsilon = \$0.25$ for this problem, then $V_T(\Phi) \geq \varepsilon e^{rT}$.

8. A European call option and put option on a stock both have strike price of \$20 and an expiration date in three months. Both sell for \$3. The risk-free interest rate is 10% per annum, the current stock price is \$18. Identity the arbitrage opportunity open to a trader.

Solution: By (1.17),

$$c_0 + Ke^{-rT} = p_0 + S_0. \tag{1.32}$$

$$3 + 20e^{-0.1\times(3/12)} = p_0 + 18 \quad \Rightarrow \quad p_0 = 4.51$$

The put is then undervalued relative to the call. So, the arbitrageur should buy the put and sell the call. He also need to buy a stock so that he can sell it to the the buyer of the call if the stock price is high at the end of three months. This portfolio can also be seen from the violation of (1.32):

$$p_0 + S_0 - c_0 - Ke^{-rT} < 0. \tag{1.33}$$

So, the arbitrageur should build a portfolio

$$\Phi = p + S - c - B$$

by borrowing money B from the money market to buy the put, buy the stock, and sell the call. B_0 is determined by

$$0 = V_0(\Phi) = p_0 + S_0 - c_0 - B_0.$$

Since $p_0 + S_0 - c_0 < Ke^{-rT}$ by (1.33), we able to choose an $\varepsilon > 0$ so that $p_0 + S_0 - c_0 = Ke^{-rT} - \varepsilon$. For this problem, $\varepsilon = 20e^{-0.1\times 0.25} - 18 = 1.51$.

$$\begin{aligned}
V_T(\Phi) =&(K - S_T)^+ + S_T - (S_T - K)^+ - (p_0 + S_0 - c_0)e^{rT} \\
=&(K - S_T)^+ + S_T - (S_T - K)^+ - (Ke^{-rT} - \varepsilon)e^{rT} \\
=&\left\{ \begin{array}{ll} K & \text{if } S_T \le K \\ K & \text{if } S_T > K \end{array} \right\} - K + \varepsilon e^{rT} \\
=&\varepsilon e^{rT} > 0.
\end{aligned}$$

We know exactly the gain of the arbitrageur.

Remark: The above solution indeed proves why the put-call parity $p_0 + S_0 - c_0 - Ke^{-rT} = 0$ should be valid in an arbitrage-free market.

9. (Dividend Put-Call Parity Formula) Note that by the arbitrage-free principle, it is easy to show that when a stock pays a dividend D_1 at t_1, the stock's value is immediately reduced by the amount of the dividend. In other words, $\lim_{t\uparrow t_1} S_t - D_1 = \lim_{t\downarrow t_1} S_t$.

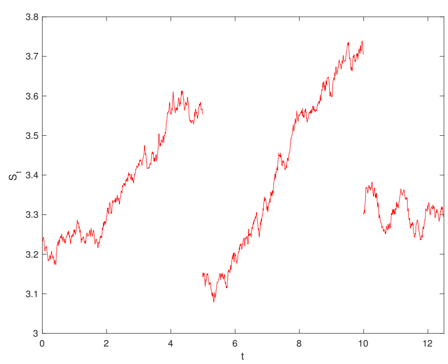


Figure 1.2: $t_1 = 5$, $t_2 = 10$, $D_1 = 0.4$, $D_2 = 0.4$.

Assume that a dividend D_j is paid at time t_j, where $0 < t_1 < t_2 < \cdots < t_n \le T$. Let D denote the present value of the dividend stream:

$$D = e^{-rt_1}D_1 + e^{-rt_2}D_2 + \cdots e^{-rt_n}D_n. \tag{1.34}$$

Consider an European call option and an European put option, each with strike price K, maturity T, and underlying one share of S, assumed to be dividend-paying. We want to prove

$$c_0 + Ke^{-rT} + D = p_0 + S_0. \tag{1.35}$$

The idea is to consider two portfolios at $t = 0$:

$$\begin{aligned}
\Phi_1 &= c + Ke^{-rT} + D, \\
\Phi_2 &= p + S.
\end{aligned}$$

It means that Φ_1 consists of a call option, a zero-coupon bond with face value (par value) K and maturity T, and n zero-coupon bonds with face values D_k and maturity times t_k, $k = 1, 2, \cdots, n$. Φ_2 consists of a put option and the stock which will pay dividend D_j at t_j.

Now, you are asked to **prove** $V_T(\Phi_1) = V_T(\Phi_2)$. Then we can conclude $V_0(\Phi_1) = V_0(\Phi_2)$ which is precisely (1.35) by Corollary 1.1.

10. (No arbitrage delivery price of a forward) A forward contract is an agreement to buy or sell an asset at a certain future time (expiration date) for a certain price (delivery price). One of the parties to a forward contract assumes a long position and agrees to buy the underlying asset on expiration date for certain delivery price F. The other party assumes a short position and agrees to sell the asset at the expiration date for the same price. Using arbitrage-free principle to show that the delivery price F on a non-dividend-paying asset with spot price S_0 is given by

$$K = S_0 e^{rT}$$

where r is the risk-free interest rate and T is the time to expiry of the forward contract.

Solution: Let F denote a forward contract and construct a portfolio

$$\Phi = -B + S - F,$$

with $V_0(B) = S_0$. It means that at $t = 0$, one borrows S_0 dollars from the bank to buy the asset S from the market, and one also sell a forward contract. This means he/she agrees to sell S at the expiration date for the delivery price K. So, $V_0(F) = 0$ and $V_T(-F) = K - S_T$. The latter equation says that since he/she sold a forward contract, by T, he/she has to sell S (whose price is S_T) for price K to the buyer of the forward contract. Hence

$$V_0(\Phi) = -S_0 + S_0 + 0 = 0,$$
$$V_T(\Phi) = -S_0 e^{rT} + S_T + K - S_T = K - S_0 e^{rT}.$$

So, if $K > S_0 e^{rT}$, $V_0(\Phi) = 0$ and $V_T(\Phi) > 0$, one has an arbitrage opportunity. On the other hand, if $K < S_0 e^{rT}$, $V_0(-\Phi) = 0$ and $V_T(-\Phi) > 0$, one has an arbitrage opportunity by building a portfolio $-\Phi$ which is $B - S + F$. Since we assume the market is arbitrage-free, we must have $K = S_0 e^{rT}$.

11. Consider a European call c_1 with a strike price of K_1 and a second European call c_2 on the same stock with a strike price of $K_2 > K_1$. Both call options have the same expiration date. Let $c(t, K)$ denote the price of the European call option at time t with strike price K. So, $c(t, K_i) = V_t(c_i)$ for $i = 1, 2$. Prove that

$$-e^{-r(T-t)}(K_2 - K_1) < c(t, K_2) - c(t, K_1) < 0.$$

Furthermore, deduce that

$$-e^{-r(T-t)} \leq \frac{\partial c}{\partial K}(t, K) \leq 0. \tag{1.36}$$

In other words, suppose the call price can be expressed as a differentiable function of the strike price, then the derivative must be non-positive and no greater in absolute value than the price of a zero-coupon bond with face value of unity and the same maturity.

Solution: Consider the following portfolio at $t = 0$:

$$\Phi = c_2 - c_1 + e^{-rT}(K_2 - K_1).$$

Then

$$V_T(\Phi) = (S_T - K_2)^+ - (S_T - K_1)^+ + (K_2 - K_1) = \begin{cases} (K_2 - K_1) & \text{if } S_T \le K_1 \\ -S_T + K_2 & \text{if } K_1 < S_T \le K_2 \\ 0 & \text{if } K_2 < S_T \end{cases}$$

Hence $V_T(\Phi) \ge 0$ and $\mathbb{P}(V_T(\Phi) > 0) > 0$. By Definition 1.1, we must have

$$V_t(\Phi) > 0, \qquad \forall t \in [0, T). \tag{1.37}$$

(If not, then $V_{t^*}(\Phi) \le 0$ for some $t^* \in [0, T)$, which means Φ has arbitrage opportunity by Definition 1.1.) Equation (1.37) means

$$V_t(\Phi_1) = c(t, K_2) - c(t, K_1) + e^{-r(T-t)}(K_2 - K_1) > 0.$$

Next, we consider $\Phi = c_1 - c_2$, Then

$$V_T(\Phi) = (S_T - K_1)^+ - (S_T - K_2)^+ = \begin{cases} 0 & \text{if } S_T \le K_1 \\ S_T - K_1 & \text{if } K_1 < S_T \le K_2 \\ K_2 - K_1 & \text{if } K_2 < S_T \end{cases}$$

Hence $V_T(\Phi) \ge 0$ and $\mathbb{P}(V_T(\Phi) > 0) > 0$. By Definition 1.1, we must have

$$V_t(\Phi) > 0, \qquad \forall t \in [0, T).$$

This means

$$V_t(\Phi_1) = c(t, K_1) - c(t, K_2) > 0.$$

Finally, by the definition of derivative,

$$\frac{\partial c}{\partial K}(t, K) = \lim_{h \downarrow 0} \frac{c(t, K+h) - c(t, K)}{h}.$$

Since $h > 0$, $-e^{-r(T-t)} < \frac{c(t,K+h)-c(t,K)}{h} < 0$, we get (1.36) by letting $h \downarrow 0$.

2 The binomial tree methods (2 lectures)

The analysis made on option pricing in the previous lecture is solely based in the assumption that the market is arbitrage-free, without referring to any price model of the underlying asset. Without a price model of the underlying asset, only qualitative discussions on option pricing are possible. Quantitative pricing of the derivatives requires specific model on price movement of the underlying asset.

In 1900, Louis Bachelier, a young French mathematician, completed a thesis called “the theory of speculation”. Bachelier developed the now universally used concept of stochastic process and proposed a model on how the stock price change with respect to time. This enable him to make the first theoretical attempt to value options [7].

Bachelier’s thesis was not well received. The pioneering nature of his work was recognized only after several decades, first by Kolmogorov[8] who pointed out his work to Lévy, then by L. J. Savage who brought the work of Bachelier to the attention of Paul Samuelson in 1950s (no economist at the time had ever heard of Bachelier).

An intensive period of development in financial economics followed, first at MIT, where Samuelson worked, leading to the Nobel Prize-winning solution of the option pricing problem by Fischer Black, Myron Scholes and Robert Merton in 1973. In the same year, the world’s first listed options exchange opened its doors in Chicago.

In 1979, Cox, Ross and Rubinstein introduced the N-period binomial tree model for stock price and N-period binomial method for option pricing [9]. They are easy to understand and implement in practice. We will also show that resulting option price converges to the Black-Scholes-Merton formula for option pricing as $N \to \infty$.

We will start with the 1-period binomial tree model for the stock price. Then we move to N-period model, and finally the geometric Brownian motion model proposed by Samuelson and adapted by Black-Scholes-Merton.

2.1 An example

$$
\begin{tikzcd}
 & S_{\delta t}(\$45) & & & c_{\delta t} = (S_{\delta t} - K)^+ = \$5 \\
S_0(\$40) \arrow[ur] \arrow[dr] & & & c_0 = ? \arrow[ur] \arrow[dr] & \\
 & S_{\delta t}(\$35) & & & c_{\delta t} = (S_{\delta t} - K)^+ = 0
\end{tikzcd}
$$

[7] A 1688 treatise on the workings of the Amsterdam stock exchange (established in 1602 by the Dutch East India Company) reveals that options were already dominating trading activities at the time. Amsterdam was the most sophisticated and important financial center of the seventeenth century.

[8] Born in 1903. In 1933, Kolmogorov published his book, Foundations of the Theory of Probability, laying the modern axiomatic foundations of probability theory.

[9] J. C. Cox, S. A. Ross, and M. Rubinstein, Option pricing: A simplified approach, Journal of Financial Economics 7 (1979) 229–263.

Let the price of a stock be \$40 at $t = 0$, and suppose a month later ($t = \delta t$) the stock price will either goes up to \$45 or down to \$35. Now consider buying a European call option of the stock at $t = 0$ with strike price \$40 and 1 month maturity. If the risk-free annual interest rate is 12%, how much should the price for the call option be?

Now, construct two portfolios

$$\Phi_1 = S - 2c, \qquad \Phi_2 = B \text{ with } B_0 = \frac{35}{e^{0.01}} = 34.652.$$

At $t = \delta t$,

$$V_{\delta t}(\Phi_1) = S_{\delta t} - 2(S_{\delta t} - K)^+ = \begin{cases} 45 - 2 \times (45 - 40)^+ = 35 & \text{if } S \text{ goes up} \\ 35 - 2 \times (35 - 40)^+ = 35 & \text{if } S \text{ goes up} \end{cases}.$$

$$V_{\delta t}(\Phi_2) = \frac{35}{e^{0.01}} e^{0.12 \times \frac{1}{12}} = 35$$

$V_{\delta t}(\Phi_1) = V_{\delta t}(\Phi_2)$. Hence

$$V_0(\Phi_1) = V_0(\Phi_2).$$

So, $S_0 - 2c_0 = B_0$ which leads to $c_0 = \frac{40 - 34.652}{2} \approx \$2.67$. We conclude that the investor should pay \$2.67 for the stock option.

This example reveals the idea of hedging: it is possible to construct a risk-free investment portfolio Φ with c and its underlying asset S.

2.2 The one period binomial tree method

Now we give a more systematic analysis.

- Let T be the length of the time interval we are considering.
- Let r denote the risk-free interest rate and let $\rho = e^{rT}$ so that \$1 deposit in the bank becomes \$$\rho$ after T units of time. Here we assume that the interest is compounded continuously. Otherwise, we can change the formula of ρ accordingly.
- Assume that the underlying Stock S, whose value is S_0 at $t = 0$, can have two values $S^u = uS_0$ and $S^d = dS_0$ at time T with $d < u$. Under the no arbitrage principle, we must have

  $$d < \rho < u. \tag{2.1}$$

  This is because if $\rho \leq d$ then no investor would deposit his/her money in the bank (or buy treasury bills), and if $u \leq \rho$ then no investor would invest in the stock market.
- Option price $Ⓟ_t$ has two values $Ⓟ^u$ and $Ⓟ^d$ at time T depending on whether S goes up or goes down [10].

[10] If T is the length between now and the expiration date, and if it is a call option, then $Ⓟ^u = \max(uS_0 - K, 0)$ and $Ⓟ^d = \max(dS_0 - K, 0)$. If it is a put option, then $Ⓟ^u = \max(K - uS_0, 0)$ and $Ⓟ^d = \max(K - dS_0, 0)$.

Now, suppose we know S_0, u, d, $ⓟ^u$, $ⓟ^d$, ρ, and we want to determine $ⓟ_0$, which is the option price at $t = 0$.

To do that, let us build a portfolio Φ as follows: buy an option ⓟ, short sell Δ shares of the underlying stock S[11]:

$$\Phi = ⓟ - \Delta S. \tag{2.2}$$

Δ is determined so that Φ takes the same value at T no matter whether S is going up or going down. Hence we find Δ by solving

$$ⓟ^u - \Delta S_0 u = ⓟ^d - \Delta S_0 d \qquad \Rightarrow \qquad \Delta = \frac{ⓟ^u - ⓟ^d}{S_0(u-d)}. \tag{2.3}$$

Hence $V_T(\Phi) = ⓟ^u - \frac{ⓟ^u - ⓟ^d}{u-d}u = \frac{-d}{u-d}ⓟ^u + \frac{u}{u-d}ⓟ^d$.

Suppose there is a bank deposit B_0 made at $t = 0$. We choose B_0 so that

$$V_T(\Phi) = V_T(B) = \rho B_0.$$

Thus

$$B_0 = \frac{1}{\rho}\left(\frac{-d}{u-d}ⓟ^u + \frac{u}{u-d}ⓟ^d\right). \tag{2.4}$$

By Corollary 1.1, $V_0(\Phi) = V_0(B)$, which means $ⓟ_0 - \Delta S_0 = B_0$, i.e.,

$$ⓟ_0 = B_0 + \Delta S_0 = \frac{1}{\rho}\left(\frac{\rho - d}{u-d}ⓟ^u + \frac{u-\rho}{u-d}ⓟ^d\right) \stackrel{\text{def}}{=} \frac{1}{\rho}\left(q_u ⓟ^u + q_d ⓟ^d\right). \tag{2.5}$$

Please note that $q_u + q_d = 1$. Hence (2.5) says that the option price is a weighted average of the value of the call option in the up state and the down state. Please also note that the definitions of q_u and q_d imply that

$$\frac{1}{\rho}\left(q_u S^u + q_d S^d\right) = \frac{1}{\rho}\left(\frac{\rho - d}{u-d}u + \frac{u-\rho}{u-d}d\right) S_0 = S_0. \tag{2.6}$$

Example 2.1 *For the example at the beginning of this section, we get* $q_u = \frac{e^{0.01} - \frac{35}{40}}{\frac{10}{40}}$ *and* $q_d = \frac{\frac{45}{40} - e^{0.01}}{\frac{10}{40}}$. $ⓟ_0 = \frac{1 - \frac{35}{40}e^{-0.01}}{\frac{10}{40}} 5 = \$2.67$.

The formula (2.5) is astonishing. It is astonishing because what it says is that as long as we agree on the value of u and d, even if we disagree on the probability of how the stock is going up or down tomorrow, yet we still are going to agree on what the value of a call option on that stock is.

For example, you may think that the price is more likely to go up, therefore, you want to have that kind of a call option bet, but I may think that the price is more likely to go down, so I'm happy to sell it to you. As long as we agree on the value of u and d, we will agree on the price of the option. That's what drives the market.

[11] A negative Δ from (2.3) implies that we should buy.

2.3 The multiperiod binomial tree method

In the previous one period model, we assume that after say, 5 minutes, or 5 days, or 5 months, the stock price can either be uS_0 or vS_0. If we can't agree on that there are only two prices, let's agree that between now and 5 minutes from now, there are 100 possible outcomes for the stock price. Do you agree on that? If you agree, then all we need to do is to have enough steps between now and 5 minutes from now to have 100 possible prices.

The multiperiod method is simply that we now have a bunch of possibilities, and you are figuring out what the price of the option is at date 0 when it pays off at date N.

Given an option with expiration date T, we consider a multiperiod binomial tree of length N, obtained by stringing together single period binomial trees, where the length of the time interval for each single period binomial tree is $\delta t = T/N$.

See Figure 2.1 for an illustration.

Image we toss a coin repeatedly (we do not need to worry whether the coin is fair or biased). Whenever we get a head (H), the stock price moves up by a factor u, and whenever we get a tail (T), the stock price moves down by a factor d. In addition to this stock, there is a money market asset with an interest rate r. We assume $d < \rho = e^{r\delta t} < u$ as in (2.1).

Let S_0 be the initial stock price. We denote the stock price at time δt by $S_1(H) = uS_0$ if the first toss result in head, and by $S_1(T) = dS_0$ if the first toss result in tail. After the second toss, the stock price will be one of

$$S_2(HH) = uS_1(H) = u^2S_0, \quad S_2(HT) = dS_1(H) = duS_0,$$
$$S_2(TH) = uS_1(T) = udS_0, \quad S_2(TT) = dS_1(T) = d^2S_0.$$

After 3 tosses, there are 8 possible coin sequences, and 4 different stock price at $t = 3\delta t$. This process will repeat until we reach $t = N\delta t = T$. See Figure 2.1 for an illustration.

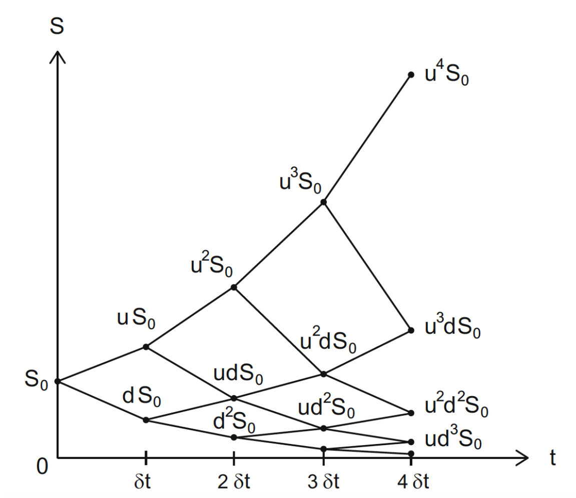


Figure 2.1: A multiperiod binomial tree of length $N = 4$ with $u = 3/2$ and $d = 1/2$. At time $t = N\delta t$, the stock price can take one of $N + 1$ possible values.

To price an option, we start from the nodes at $t = N\delta t$ and travel backward, repeatedly using the formula (2.5)

$$\text{ⓟ}_0 = \frac{1}{\rho}\left(q_u \text{ⓟ}^u + q_d \text{ⓟ}^d\right)$$

or more precisely, using our current notation,

$$\text{ⓟ}_n(\omega_1\omega_2\cdots\omega_n) = \frac{1}{\rho}\left(q_u \text{ⓟ}_{n+1}(\omega_1\omega_2\cdots\omega_n H) + q_d \text{ⓟ}_{n+1}(\omega_1\omega_2\cdots\omega_n T)\right). \tag{2.7}$$

Here $\rho \stackrel{\text{def}}{=} e^{r\delta t}$, ω_i is either H or T, $\text{ⓟ}_n(\omega_1\omega_2\cdots\omega_n)$ denotes the option price at $t = n\delta t$ after observing the head-tail sequence $(\omega_1\omega_2\cdots\omega_n)$ (i.e., knowing the underlying stock price up to $t = n\delta t$).

Example 2.2 *We now compute the price of a European call option with the strike price $K = 30$ and $S_0 = 32$ using a three-step binomial tree. $u = \frac{3}{2}$ and $d = \frac{1}{2}$. For the sake of computational convenience, we assume $r = 0$ (i.e., $\rho = 1$). Note that $q_u = \frac{\rho-d}{u-d} = \frac{1}{2}$ and $q_d = \frac{1}{2}$. The solution is shown by Figures 2.2 to 2.4.*

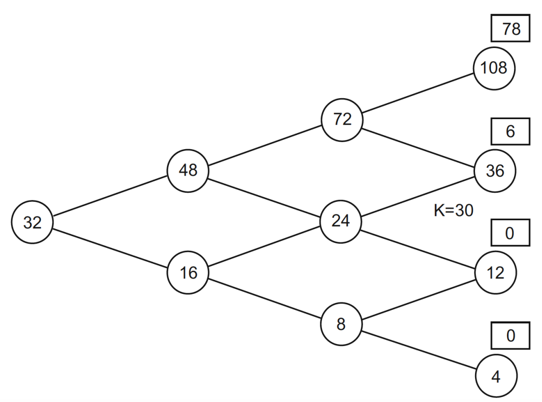


Figure 2.2: Payoff of a European call option in a binomial tree of length $N = 3$. $K = 30$. $\boxed{\text{option price}}$. stock price

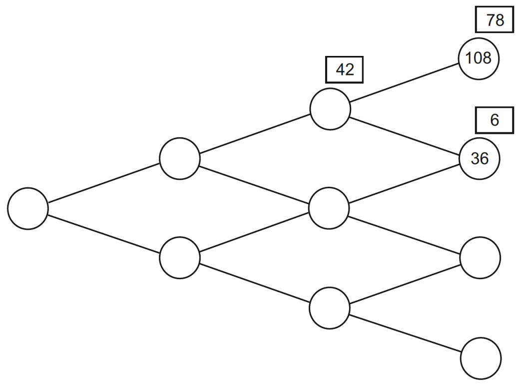


Figure 2.3: The first step in pricing of a European call option. $K = 30$. $q_u = \frac{1}{2} = q_d$. $\rho = 1$. $\boxed{\text{option price}}$. stock price . The number 42 in $\boxed{42}$ is derived using $\textcircled{p}_0 = \frac{1}{\rho}\left(q_u \textcircled{p}^u + q_d \textcircled{p}^d\right)$ in (2.5) (or (2.7)) with $\textcircled{p}^u = 78$, $\textcircled{p}^d = 6$.

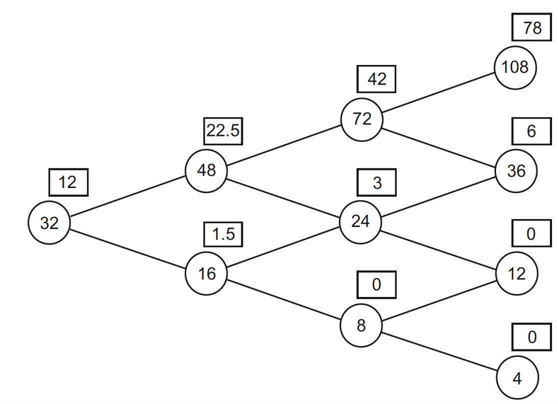


Figure 2.4: Pricing of a European call option in a binomial tree of length $N = 3$. $K = 30$. The result is 12. $\boxed{\text{option price}}$. stock price

Now, we want to derive the general formula for the option price. It means that we want to derive a formula of $ⓟ_0$ in terms of $\{ⓟ_N(\omega_1\omega_2\cdots\omega_N), \omega_i \in \{H, T\}\}$ using the backward recursive relation

$$ⓟ_n(\omega_1\omega_2\cdots\omega_n) = \frac{1}{\rho}\left(q_u ⓟ_{n+1}(\omega_1\omega_2\cdots\omega_n H) + q_d ⓟ_{n+1}(\omega_1\omega_2\cdots\omega_n T)\right). \tag{2.8}$$

If it is 2-period binomial tree,

$$\begin{aligned}
ⓟ_0 =& \frac{1}{\rho}\left(q_u ⓟ_1(H) + q_d ⓟ_1(T)\right) \\
=& \frac{1}{\rho^2}\left(q_u\left[q_u ⓟ_2(HH) + q_d ⓟ_2(HT)\right] + q_d\left[q_u ⓟ_2(TH) + q_d ⓟ_2(TT)\right]\right) \\
=& \frac{1}{\rho^2}\left(q_u^2 ⓟ_2(HH) + 2q_u q_d ⓟ_2(HT) + q_d^2 ⓟ_2(TT)\right).
\end{aligned}$$

We have used $ⓟ_2(HT) = ⓟ_2(TH)$ in the last step.

For 3-period binomial tree, we continue and get

$$\begin{aligned}
ⓟ_0 =& \frac{1}{\rho^2}\left(q_u^2 ⓟ_2(HH) + 2q_u q_d ⓟ_2(HT) + q_d^2 ⓟ_2(TT)\right) \\
=& \frac{1}{\rho^3}\left(q_u^2\left[q_u ⓟ_3(HHH) + q_d ⓟ_3(HHT)\right] + 2q_u q_d\left[q_u ⓟ_3(HTH) + q_d ⓟ_3(HTT)\right]\right. \\
& \left. + q_d^2\left[q_u ⓟ_3(TTH) + q_d ⓟ_3(TTT)\right]\right) \\
=& \frac{1}{\rho^3}\left(q_u^3 ⓟ_3(HHH) + 3q_u^2 q_d ⓟ_3(HHT) + 3q_u q_d^2 ⓟ_3(HTT) + q_d^3 ⓟ_3(TTT)\right).
\end{aligned}$$

We have used $ⓟ_3(HHT) = ⓟ_3(HTH)$ etc. in the last step.

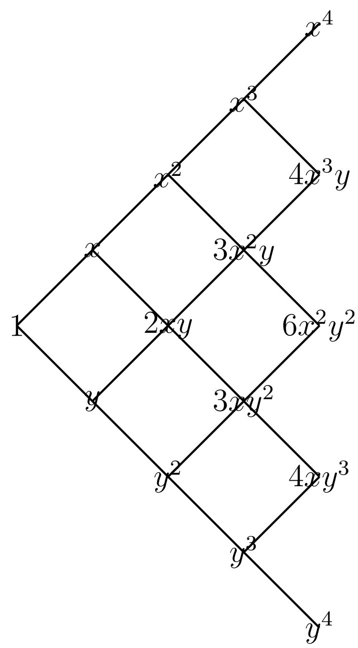
For general N-period binomial tree, each $ⓟ_N(\omega_1\omega_2\cdots\omega_N)$ contributes to $ⓟ_0$. By (2.8), if $(\omega_1\omega_2\cdots\omega_N)$ contains a H, it contributes $\frac{q_u}{\rho}$. If it contains a T, it contributes $\frac{q_d}{\rho}$. We should keeping in mind that, say, $ⓟ_4(HTTT) = ⓟ_4(THTT) = \cdots = ⓟ_4(TTTH) = \max(ud^3 S_0 - K, 0)$ if we are considering a 4 period European call option.

Recall $\binom{N}{j} = \frac{N!}{(N-j)!j!}$ gives the number of ways, disregarding order, that j objects can be chosen from among N objects. Using it, we can have the binomial expansion formula

$$(x+y)^N = \underbrace{(x+y)\,(x+y)\,(x+y)\,\cdots(x+y)}_{N\text{terms}} = \sum_{j=0}^{N}\binom{N}{j}x^j y^{N-j} \tag{2.9}$$

and the corresponding Pascal's triangle

The stock price paths, moving from left to right, containing $N-j$ down's and j up's, always go to the same terminal point. The number of such paths is $\binom{N}{j}$. Hence we find that the contribution of $ⓟ_N(\omega$ contains j H's and $N-j$ T's) to $ⓟ_0$ is precisely $\frac{1}{\rho^N}\binom{N}{j}q_u^j q_d^{N-j}$. So we have proved the following theorem (which can also be proved by induction)



$n \backslash j$	0	1	2	3	4	5	6	...
0	1							
1	1	1						
2	1	2	1					
3	1	3	3	1				
4	1	4	6	4	1			
5	1	5	10	10	5	1		
6	1	6	15	20	15	6	1	...
⋮			⋮					

Figure 2.5: Pascal's triangle

Theorem 2.1 *The price of an European call option with strike price K, expiration date T, interests rate r, and stock price S_0 at $t = 0$, is*

$$\begin{aligned}
\textcircled{p}_0 =& \frac{1}{\rho^N} \sum_{i=0}^{N} \binom{N}{j} q_u^j q_d^{N-j} \textcircled{p}_N(\omega \textit{ contains } j \textit{ } H\textit{'s and } N-j \textit{ } T\textit{'s}) \\
=& e^{-rT} \sum_{j=0}^{N} \binom{N}{j} q_u^j q_d^{N-j} \left(S_0 u^j d^{N-j} - K\right)^+ . \qquad (2.10)
\end{aligned}$$

2.4 Convergence to the Black-Scholes-Merton formula

In this section we show that the European call option price (2.10) converges to the solution from the Black-Scholes-Merton partial differential equation (which will be discussed later) as $N \to \infty$.

Note that as $N \to \infty$, $\delta t = T/N \to 0$. In (2.10), we need to decide when $\left(S_0 u^i d^{N-i} - K\right)^+ \neq 0$.

So, let m be the smallest integer such that

$$S_0 u^m d^{N-m} > K.$$

Then

$$\textcircled{p}_0 = e^{-rT} \sum_{i=m}^{N} \binom{N}{i} q_u^i q_d^{N-i} \left(S_0 u^i d^{N-i} - K\right).$$

Let $A = e^{-rT}\sum_{i=m}^{N}\binom{N}{i}q_u^i q_d^{N-i}u^i d^{N-i}$ and $B = \sum_{i=m}^{N}\binom{N}{i}q_u^i q_d^{N-i}$. Then

$$ⓟ_0 = AS_0 - Ke^{-rT}B.$$

One can then manage to prove that (see Page 204 of "Option markets" by of Cox and Rubinstein or Section 14.4 the book of G. H. Choe)

$$\lim_{N\to\infty} ⓟ_0 = N(d_1)S_0 - Ke^{-rT}N(d_2), \tag{2.11}$$

where $N(x) = \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy = \int_{-x}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy$,

$$d_1 = \lim_{N\to\infty}\frac{\frac{\log\frac{S_0}{K}+N\log d}{\log\frac{u}{d}}+Nq^*}{\sqrt{Nq^*(1-q^*)}} \quad d_2 = \lim_{N\to\infty}\frac{\frac{\log\frac{S_0}{K}+N\log d}{\log\frac{u}{d}}+Nq_u}{\sqrt{Nq_u(1-q_u)}} \tag{2.12}$$

with $q^* = e^{-r\delta t}q_u u$.

If we set

$$u = e^{\sigma\sqrt{\delta t}}, \qquad d = e^{-\sigma\sqrt{\delta t}},$$

Then

$$d_1 = \frac{\log\frac{S_0}{K}+(r+\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}, \qquad d_2 = \frac{\log\frac{S_0}{K}+(r-\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}. \tag{2.13}$$

The right hand side of (2.11) is then exactly the Black-Scholes-Merton (1973) option pricing formula.

Remark: For your information, $u, d = e^{\pm\sigma\sqrt{\delta t}}$ and $q_u = \frac{e^{r\delta t}-d}{u-1}$ is proposed by Cox, Ross and Rubinstein in their 1979 paper. Another option is to choose $u = e^{(r-\frac{1}{2}\sigma^2)\delta t+\sigma\sqrt{\delta t}}$, $d = e^{(r-\frac{1}{2}\sigma^2)\delta t-\sigma\sqrt{\delta t}}$, $q_u = 1/2$ is proposed by Jarrow and Rudd in 1980 [12]. In both case, the σ is the volatility of the stock price which will be discussed later.

2.4.1 Limit distribution of the stock price in the risk-neutral world as $\delta t \to 0$ [13]

In the multiperiod binomial tree model, the stock price satisfies $S_n = S_0 u^i d^{n-i}$ if the first n coin toss contains i heads and $n-i$ tails. Here we take $\delta t = 1/N$ (i.e., N steps per unit time), $t_n = n\delta t$, $u = e^{\sigma\sqrt{\delta t}}$, $d = e^{-\sigma\sqrt{dt}}$. I like to explain why as $N\to\infty$, S_{Nt} converges to the distribution of

$$S(t) = S(0)e^{\sigma W(t)+(r-\frac{1}{2}\sigma^2)t} \tag{2.14}$$

[12] R. Jarrow and A. Rudd, Approximate option valuation for arbitrary stochastic process, Journal of Financial Economics 10 (1982) 347–369.

[13] For your information only. It won't be tested. Please skip in the first reading if you have not learned probability yet.

with $W(t) \sim N(0,t)$. [14] By Taylor expansion, $q_u = \frac{e^{r\delta t} - e^{-\sigma\sqrt{\delta t}}}{e^{\sigma\sqrt{\delta t}} - e^{-\sigma\sqrt{\delta t}}} \approx \frac{1 + r\delta t - (1 - \sigma\sqrt{\delta t} + \frac{1}{2}\sigma^2\delta t)}{2\sigma\sqrt{\delta t}} = \frac{(r - \frac{1}{2}\sigma^2)\delta t}{2\sigma\sqrt{\delta t}} + \frac{1}{2}$. From $S_n = S_0 u^i d^{n-i} = S_0 e^{\xi_1} e^{\xi_2} \cdots e^{\xi_n}$, we see that

$$\log(S_{Nt}/S_0) = i \log u + (Nt - i) \log d = \sum_{k=1}^{Nt} \xi_k.$$

Here i is a binomial distributed variable $\sim B(Nt, q_u)$ (i.e. among the Nt trials, one obtains i heads with q_u probability to get a head in each trial). $e^{\xi_k} = u$ if the kth coin toss is a head. $e^{\xi_k} = d$ if the kth coin toss is a tail. In other words, $\mathbb{P}(\xi_k = \log u) = q_u$, $\mathbb{P}(\xi_k = \log d) = q_d$.

$$\begin{aligned}\mathbb{E}\xi_k =& q_u \log u + q_d \log d \approx \left(\frac{(r - \frac{1}{2}\sigma^2)\delta t}{2\sigma\sqrt{\delta t}} + \frac{1}{2}\right)\sigma\sqrt{\delta t} + \left(-\frac{(r - \frac{1}{2}\sigma^2)\delta t}{2\sigma\sqrt{\delta t}} + \frac{1}{2}\right)(-\sigma\sqrt{\delta t}) \\ =& (r - \frac{1}{2}\sigma^2)\delta t.\end{aligned}$$

Denote $\mathbb{E}\xi_k \stackrel{\text{def}}{=} a$. Then

$$\begin{aligned}\mathrm{Var}(\xi_k) =& \mathbb{E}(\xi_k^2) - (\mathbb{E}\xi_k)^2 \\ \approx& \left(\frac{(r - \frac{1}{2}\sigma^2)\delta t}{2\sigma\sqrt{\delta t}} + \frac{1}{2}\right)\sigma^2\delta t + \left(-\frac{(r - \frac{1}{2}\sigma^2)\delta t}{2\sigma\sqrt{\delta t}} + \frac{1}{2}\right)\sigma^2\delta t - a^2 \\ =& \sigma^2\delta t + \text{ higher order terms w.r.t. } \delta t.\end{aligned}$$

By the central limit theorem[15], we know $\frac{\sum_{k=1}^{Nt}(\xi_k - \mathbb{E}\xi_k)}{\sqrt{Nt}\sqrt{\mathrm{Var}(\xi_k)}} \to N(0,1)$, or

$$\frac{\sum_{k=1}^{Nt}\left(\xi_k - (r - \frac{1}{2}\sigma^2)\delta t\right)}{\sqrt{Nt}\sigma\sqrt{\delta t}} \to W(1)$$

where $W(1) \sim N(0,1)$. Note that $N\delta t = 1$. $Nt\delta t = t$. Hence

$$\frac{\left(\sum_{k=1}^{Nt}\xi_k\right) - (r - \frac{1}{2}\sigma^2)t}{\sigma\sqrt{t}} \to W(1).$$

Since $\sqrt{t}W(1) \sim N(0,t)$ and can be denoted as $W(t)$,

$$\log(S_{Nt}/S_0) = \sum_{k=1}^{Nt}\xi_k \to \sigma W(t) + (r - \frac{1}{2}\sigma^2)t.$$

This proves (2.14). The resulting $S(t)$ is said to satisfy log-normal distribution (which means that its log has normal distribution).

[14] It is not a fully rigorous proof as I use $\approx$ instead of $=$ in the calculation. But it is enough for you to get the idea. Please pay attention to where $r - \frac{1}{2}\sigma^2$ comes from. More rigorous proofs can be found in Shreve II, Section 3.2.7 for $r = 0$ and Exercise 3.8 for general r.

[15] Suppose $\{X_1, \cdots\}$ is a sequence of independent identically distributed random variables with $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2 < \infty$. Then as $n \to \infty$, the random variable $\frac{\sum_{i=1}^n (X_i - \mu)}{\sqrt{n}\sigma}$ converge in the distribution to a standard normal random variable $N(0,1)$.

2.5 Computer experiments

The Matlab code in this section is taken from Choe's book but with slight modification.

Example 2.3 *Now consider buying a European call option of the stock at $t = 0$ with strike price \$110 and 1 year maturity. If the risk-free annual interest rate is 5% and $u = e^{\sigma\sqrt{dt}}$ and $d = e^{-\sigma\sqrt{dt}}$ with $\sigma = 0.3$, how much should the price for the call option be?*

```
S0 = 100;
K = 110;
T = 1;
r = 0.05;
sigma = 0.3;
M = 10; % number of time steps
dt = T/M;
u = exp(sigma*sqrt(dt));
d = exp(-sigma*sqrt(dt));
q = (exp(r*dt)-d)/(u-d);
% We compute asset prices.
for i = 0:1:M
    fprintf('time = %i\n', i)
    S = S0*u.^([i:-1:0]).*d.^([0:1:i])
end
fprintf('Payoff at expiry\n')
Call = max(S0*u.^([M:-1:0]).*d.^([0:1:M])- K,0)
% We proceed backward to compute option value at time 0.
for i = M:-1:1
    fprintf('time = %i\n', i-1)
    Call = exp(-r*dt)*(q*Call(1:i) + (1-q)*Call(2:i+1))
end
```

If we increase M, the option priced obtained by the above code converges to the result given by the classic Black-Scholes-Merton formula.

```
S0 = 100;
K = 110;
T = 1;
r = 0.05;
sigma = 0.3;

M_values = [50:1:1000];
Call_prices = zeros(length(M_values),1);

for j = 1:length(M_values)
```

```
    M = M_values(j);
    dt = T/M;
    u = exp(sigma*sqrt(dt));
    d = exp(-sigma*sqrt(dt));
    q = (exp(r*dt)-d)/(u-d);
    Call = max(S0 *u.^([M:-1:0]) .* d.^([0:1:M]) - K,0);

    for i = M:-1:1
        Call = exp(-r*dt)*(q*Call(1:i)+(1-q)*Call(2:i+1));
    end

    Call_prices(j) = Call;
end

plot(M_values,Call_prices,'.');
hold on;
% the Black-Scholes-Merton formula.
d1 = (log(S0/K)+(r+0.5*sigma^2)*T)/(sigma*sqrt(T));
d2 = d1-sigma*sqrt(T);
Call_BSM = S0*normcdf(d1) - K*exp(-r*T)*normcdf(d2);
x=0:1000;
plot(x,Call_BSM*ones(size(x)),'r');
```

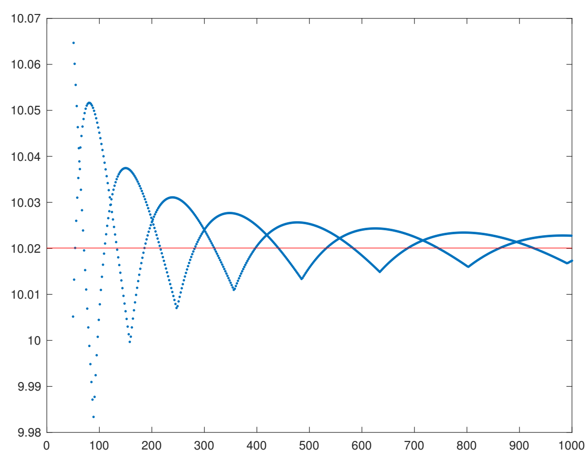


Figure 2.6: Convergence to the Black-Scholes-Merton price.

2.6 Finite probability space

So far, we have used little probability. Let us now put what we have discussed so far into the standard probability framework. Besides Shreve I and II which we are going to follow, another standard textbook for probability is “A First Course in Probability” by Sheldon Ross.

Recall that in our multiperiod binomial model, at each $t_n = n\delta t$, a coin is tossed, and the outcome of the coin toss determine how S_{n-1} should change to S_n. We are interested in the stock price S_n. But there is a correspondence between the stock price and the sequence of the coin tosses. Let Ω be the set of all possible outcomes of the coin tosses. For example, if we toss the coin three times, the set of all possible outcome is

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}. \tag{2.15}$$

Suppose that on each toss, the probability of a head is p_u and the probability of a tail is p_d. We assume the tosses are independent. So the probabilities of the individual elements ω in Ω are

$$\mathbb{P}(HHH) = p_u^3, \mathbb{P}(HHT) = p_u^2 p_d, \mathbb{P}(HTH) = p_u^2 p_d, \mathbb{P}(HTT) = p_u p_d^2, \ldots\ldots$$

In probability, the subset of Ω are called events. For example, the event

$$\text{“the first toss is a head”} = \{\omega \in \Omega; \omega_1 = H\} = \{HHH, HHT, HTH, HTT\}$$

and we can compute

$$\begin{aligned}\mathbb{P}(\text{“the first toss is a head”}) =& \mathbb{P}(HHH) + \mathbb{P}(HHT) + \mathbb{P}(HTH) + \mathbb{P}(HTT) \\ =& (p_u^3 + p_u^2 p_d) + (p_u^2 p_d + p_u p_d^2) = p_u^2 + p_u p_d = p_u.\end{aligned}$$

Definition 2.1 *(Shreve I, Defintion 2.1.1) A finite probability space consists of a sample space Ω and a probability measure $\mathbb{P}$. The sample space Ω is a nonempty finite set and the probability measure $\mathbb{P}$ is a function that assigns to each element ω of Ω a number in $[0,1]$ so that*

$$\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1. \tag{2.16}$$

An event is a subset of Ω, and we define the probability of an event A to be

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega). \tag{2.17}$$

Remark: By definition, $\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1$. If A and B are disjoint subsets of Ω, then

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \tag{2.18}$$

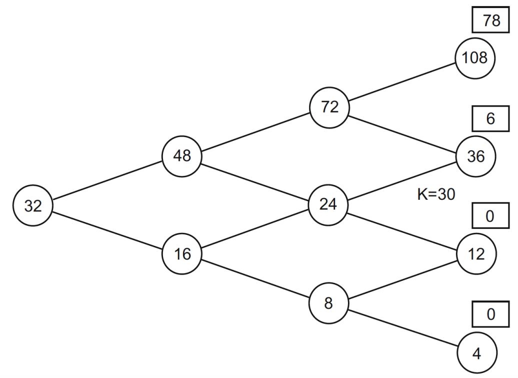
Definition 2.2 *Let $(\Omega, \mathbb{P})$ be a finite probability space. A random variable X is a real-valued function defined on Ω. The expectation of X is defined to be*

$$\mathbb{E}X = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega). \tag{2.19}$$

The variance of X is

$$Var(X) = \mathbb{E}[(X - \mathbb{E}X)^2] = \mathbb{E}[X^2] - (\mathbb{E}X)^2. \tag{2.20}$$

Example 2.4 *(Stock prices) Recall the space Ω in (2.15) which consists of sequence of 3 independent coin tosses. As in Figure 2.2, let us define the Stock prices by the formula*



$$S_0(w_1w_2w_3) = 32 \quad \textit{for all } \omega = \omega_1\omega_2\omega_3 \in \Omega,$$

$$S_1(w_1w_2w_3) = \begin{cases} 48 & \textit{if } \omega_1 = H \\ 16 & \textit{if } \omega_1 = T \end{cases}$$

$$S_2(w_1w_2w_3) = \begin{cases} 72 & \textit{if } \omega_1\omega_2 = HH \\ 24 & \textit{if } \omega_1\omega_2 = HT \textit{ or } TH \\ 8 & \textit{if } \omega_1\omega_2 = TT \end{cases}$$

$$S_3(w_1w_2w_3) = \begin{cases} 108 & \textit{if } \omega_1\omega_2\omega_3 = HHH \\ 36 & \textit{if there are two } H\textit{'s and one } T \\ 12 & \textit{if there are one } H \textit{ and two } T\textit{'s} \\ 4 & \textit{if } \omega_1\omega_2\omega_3 = TTT \end{cases}$$

Here we have written the arguments of S_0, S_1, S_2, and S_3 as $\omega_1\omega_2\omega_3$, even though some of these random variables do not depend on all the coin tosses.

Example 2.5 *Recall that in the one period binomial model. Define $\Omega = \{H, T\}$ and a probability measure $\mathbb{P}$ with $\mathbb{P}(H) = p_u$, $\mathbb{P}(T) = p_d$. They satisfy $p_u + p_d = 1$, $q_u, q_d > 0$. Then the option price at time T, denoted by $ⓟ_T$, can be $ⓟ^u$ or $ⓟ^d$ after one period, depending on whether event H or event T happen.*

$$\mathbb{E}[ⓟ_T] = p_u ⓟ^u + p_d ⓟ^d$$

This $\mathbb{P}$ is the real-world probability. $\mathbb{E}$ is the expectation with respect to $\mathbb{P}$.

Example 2.6 *Recall that in the one period binomial model, we have (see (2.5)):*

$$ⓟ_0 = \frac{1}{e^{rT}} \left(q_u ⓟ^u + q_d ⓟ^d \right).$$

Now define $\Omega = \{H, T\}$ and a probability measure $\tilde{\mathbb{P}}$ with $\tilde{\mathbb{P}}(H) = q_u$, $\tilde{\mathbb{P}}(T) = q_d$. They satisfy $q_u + q_d = 1$, $q_u, q_d > 0$. Then

$$ⓟ_0 = e^{-rT} \tilde{\mathbb{E}}[ⓟ_T] \tag{2.21}$$

This $\tilde{\mathbb{P}}$ is called a risk neutral probability. $\tilde{\mathbb{E}}$ is the expectation with respect to $\tilde{\mathbb{P}}$.

Example 2.7 *Recall that in the N-step binomial model, we have (see (2.10) with $\rho = e^{r\delta t}$):*

$$\begin{aligned}
ⓟ_0 =& \frac{1}{\rho^N}\sum_{i=0}^{N}\binom{N}{i} q_u^i q_d^{N-i} ⓟ_N(\omega \text{ contains } i \text{ } H\text{'s and } N-i \text{ } T\text{'s}) \\
=& e^{-rT}\sum_{i=0}^{N}\binom{N}{i} q_u^i q_d^{N-i}\left(S_0 u^i d^{N-i} - K\right)^+ .
\end{aligned}$$

Now define $\Omega = \{\omega : \omega = \omega_1 \cdots \omega_N, \omega_i = H \text{ or } T\}$ and a probability measure $\tilde{\mathbb{P}}$ with $\tilde{\mathbb{P}}(\omega) = q_u^{H(\omega)} q_d^{N-H(\omega)}$ where $H(\omega) =$ the number of H's in ω. Define the set $A_i = \{\omega : \omega$ contains i H's and $N-i$ T's$\}$. Then A_i contains $\binom{N}{i}$ elements and on those elements, $ⓟ_N$ takes the same value $\left(S_0 u^i d^{N-i} - K\right)^+$. Hence

$$\tilde{\mathbb{P}}(A_i) = \binom{N}{i} q_u^i q_d^{N-i}.$$

and

$$ⓟ_0 = e^{-rT}\sum_{i=0}^{N}\tilde{\mathbb{P}}(A_i) ⓟ_N(A_i) = e^{-rT}\tilde{\mathbb{E}}[ⓟ_N].$$

This $\tilde{\mathbb{P}}$ is called a risk neutral probability. $\tilde{\mathbb{E}}$ is the expectation with respect to $\tilde{\mathbb{P}}$.

Remark: Recall that the capital asset pricing model (CAPM) proposed by Sharpe in 1964 (see for example, the books of John Hull or Brealey-Myers-Allen) says that

$$\text{Expected return on an asset } = r_f + \beta(r_m - r_f), \tag{2.22}$$

where r_m is the return on the market, and r_f is the return on a risk-free investment. r_m is usually approximated as the return on a well-diversified stock index such as S&P 500. β is a parameter measuring systematic risk and depends on the asset.

Once we have introduced probability and expectation, let's look at what we have found very early in Section 2.2 . If we live in a "risk neutral world" which is an imaginary world with q_u, q_d being the probabilities of the price of the stock to go up and down respectively, and if we want to compute the expected values of stock and option in this imaginary world after one time interval δt, we would get

$$\tilde{\mathbb{E}}[ⓟ_{\delta t}] = q_u ⓟ^u + q_d ⓟ^d \overset{(2.5)}{=} \rho ⓟ_0$$

and

$$\tilde{\mathbb{E}}[S_{\delta t}] = q_u u S_0 + q_d d S_0 \overset{(2.6)}{=} \rho S_0$$

with $\rho = e^{r\delta t}$ where r is the risk-free interest rate. So, the expected return for the option and the stock are the same as that of the risk-free bank account. In other words, in this "risk neutral world", people do not get extra pay or reward for taking the risk to invest in option and stock. The introduction of such an imaginary world makes the option pricing easier. In the real world, the probability is different from the risk-neutral probability, investors do get a risk premium which is the $\beta(r_m - r_f)$ part in (2.22).

Example 2.8 *Recall (2.3)*

$$\Delta = \frac{\textcircled{p}^u - \textcircled{p}^d}{S_0(u-d)}$$

is the hedge ratio in a one period binomial tree model so that $\Phi = \textcircled{p} - \Delta S$ is risk-free. It is the change in the value of the option relative to the change of the value of the stock. If we wish to make this comparison in terms of percentage change, we get the option's elasticity

$$\Omega = \frac{(\textcircled{p}^u - \textcircled{p}^d)/\textcircled{p}_0}{(uS_0 - dS_0)/S_0} = \Delta \frac{S_0}{\textcircled{p}_0}. \tag{2.23}$$

The rate of the return of an option is $\frac{\textcircled{p}_{\delta t} - \textcircled{p}_0}{\textcircled{p}_0}$, which could be $\frac{\textcircled{p}^u - \textcircled{p}_0}{\textcircled{p}_0}$ or $\frac{\textcircled{p}^d - \textcircled{p}_0}{\textcircled{p}_0}$, depending on S_t going up or down from $t = 0$ to $t = \delta t$. In the risk neutral world, the mean (denoted by $m_{\textcircled{p}}$) and the standard deviation or volatility (which is the square root of the variance, and is denoted by $\sigma_{\textcircled{p}}$) of the rate of return of an option are therefore

$$m_{\textcircled{p}} = q_u \frac{\textcircled{p}^u - \textcircled{p}_0}{\textcircled{p}_0} + (1-q_u)\frac{\textcircled{p}^d - \textcircled{p}_0}{\textcircled{p}_0}$$

and

$$\sigma_{\textcircled{p}} = \left[q_u \left(\frac{\textcircled{p}^u - \textcircled{p}_0}{\textcircled{p}_0} - m_{\textcircled{p}} \right)^2 + (1-q_u) \left(\frac{\textcircled{p}^d - \textcircled{p}_0}{\textcircled{p}_0} - m_{\textcircled{p}} \right)^2 \right]^{1/2}.$$

Similarly, we can define the mean m_S and standard diviation σ_S of the rate of return of the stock. **Show** *that*

$$\sigma_{\textcircled{p}} = \sqrt{q_u(1-q_u)} \left| \frac{\textcircled{p}^u - \textcircled{p}^d}{\textcircled{p}_0} \right| = |\Omega| \sigma_S. \tag{2.24}$$

Here, $|\Omega|$ is the absolute value of Ω. This equation relates the risk of a call to the risk of the underlying stock: The risk of a call (the standard deviation of its rate of return) equals its elasticity times its underlying stock volatility.

Solution: By Question 8 of Homework II, if $X = \begin{cases} X_a & \text{with probability } q \\ X_b & \text{with probability } 1-q. \end{cases}$ Then $E[X] = qX_a + (1-q)X_b$ and

$$\sigma(X) = \sqrt{q(1-q)}|X_a - X_b|.$$

Hence $\sigma_{\textcircled{p}} = \sqrt{q_u(1-q_u)} \left| \frac{\textcircled{p}^u - \textcircled{p}^d}{\textcircled{p}_0} \right|$ and

$$\sigma_S = \sqrt{q_u(1-q_u)}\, |u-d|\,, \qquad \sigma_{\textcircled{p}}/\sigma_S = |\Omega|.$$

2.7 Conditional expectations

Recall that in the binomial pricing model, we have $S_0 = \frac{1}{\rho}\left(q_u S^u + q_d S^d\right)$ ((2.6)) which then implies

$$S_n(\omega_1 \cdots \omega_n) = e^{-r\delta t}\left(q_u S_{n+1}(\omega_1 \cdots \omega_n H) + q_d S_{n+1}(\omega_1 \cdots \omega_n T)\right). \tag{2.25}$$

Define

$$\tilde{\mathbb{E}}_n[S_{n+1}](\omega_1 \cdots \omega_n) = q_u S_{n+1}(\omega_1 \cdots \omega_n H) + q_d S_{n+1}(\omega_1 \cdots \omega_n T). \tag{2.26}$$

Then

$$S_n(\omega_1 \cdots \omega_n) = e^{-r\delta t}\tilde{\mathbb{E}}_n[S_{n+1}](\omega_1 \cdots \omega_n). \tag{2.27}$$

We call $\tilde{\mathbb{E}}_n[S_{n+1}]$ the conditional expectation of S_{n+1} based on the information at time n. It is still a random variable depending on $(\omega_1 \cdots \omega_n)$. It is an estimate of the value of S_{n+1} based on the information of the first n coin tosses.

Definition 2.3 *Consider an N-period binomial model. Let n satisfy $1 \le n \le N$, and let $\omega_1, ..., \omega_n$ be given and, for the moment, fixed. There are 2^{N-n} possible continuations $\omega_{n+1} \cdots \omega_N$ of the sequence $\omega_1 \cdots \omega_n$. Denote by $\#H(\omega_{n+1} \cdots \omega_N)$ the number of H's in the continuation $\omega_{n+1} \cdots \omega_N$ and by $\#T(\omega_{n+1} \cdots \omega_N)$ the number of T's. Define*

$$\tilde{\mathbb{E}}_n[X](\omega_1 \cdots \omega_n) = \sum_{\omega_{n+1} \cdots \omega_N} q_u^{\#H(\omega_{n+1} \cdots \omega_N)} q_d^{\#T(\omega_{n+1} \cdots \omega_N)} X(\omega_1 \cdots \omega_n \omega_{n+1} \cdots \omega_N) \tag{2.28}$$

and call $\tilde{\mathbb{E}}_n[X]$ the conditional expectation of X based on the information at time $n\delta t$.

The two extreme cases of conditioning are $\tilde{\mathbb{E}}_0[X]$, the conditional expectation of X based on no information, which we define by

$$\tilde{\mathbb{E}}_0[X] = \tilde{\mathbb{E}}[X],$$

and $\tilde{\mathbb{E}}_N[X]$, the conditional expectation of X based on information of the full trajectory, which we define by

$$\tilde{\mathbb{E}}_N[X] = X.$$

Example 2.9 *Consider Example 2.4 with $q_u = 1/2$,*

$$\begin{aligned}
\tilde{\mathbb{E}}_1[S_3](H) =& \frac{1}{4} \times 108 + \frac{1}{4} \times 36 + \frac{1}{4} \times 36 + \frac{1}{4} \times 12 = 48, \\
\tilde{\mathbb{E}}_1[S_3](T) =& \frac{1}{4} \times 36 + \frac{1}{4} \times 12 + \frac{1}{4} \times 12 + \frac{1}{4} \times 4 = 16.
\end{aligned}$$

But if $q_u = 1/3$,

$$\begin{aligned}
\tilde{\mathbb{E}}_1[S_3](H) =& \frac{1}{9} \times 108 + \frac{2}{9} \times 36 + \frac{2}{9} \times 36 + \frac{4}{9} \times 12 = \frac{100}{3}, \\
\tilde{\mathbb{E}}_1[S_3](T) =& \frac{1}{9} \times 36 + \frac{2}{9} \times 12 + \frac{2}{9} \times 12 + \frac{4}{9} \times 4 = \frac{100}{9}.
\end{aligned}$$

Theorem 2.2 *(See Theorem 2.3.2 of Shreve, volume I.) Let N be a positive integer, and let X and Y be random variables depending on the first N steps of the trajectory. Let $0 \le n \le N$ be given. The following properties hold*

i) (Linearity of conditional expectations) For all constants c_1 and c_2, we have

$$\tilde{\mathbb{E}}_n[c_1X + c_2Y] = c_1\tilde{\mathbb{E}}_n[X] + c_2\tilde{\mathbb{E}}_n[Y]. \tag{2.29}$$

ii) (Taking out what is known) If X actually depends only on the first n coin toss, then

$$\tilde{\mathbb{E}}_n[XY] = X\tilde{\mathbb{E}}_n[Y]. \tag{2.30}$$

iii) (Iterated conditioning) If $0 \le n \le m \le N$, then

$$\tilde{\mathbb{E}}_n[\tilde{\mathbb{E}}_m[X]] = \tilde{\mathbb{E}}_n[X]. \tag{2.31}$$

In particular, $\tilde{\mathbb{E}}[\tilde{\mathbb{E}}_m[X]] = \tilde{\mathbb{E}}[X]$.

iv) (Independence) If X depends only on ω_{n+1} through ω_N, then

$$\tilde{\mathbb{E}}_nX = \tilde{\mathbb{E}}X. \tag{2.32}$$

v) (Conditional Jensen's inequality) If $\varphi(x)$ is a convex function of the dummy variable x (e.g. $\varphi(x) = x^2$), then

$$\tilde{\mathbb{E}}_n[\varphi(X)] \ge \varphi(\tilde{\mathbb{E}}_n[X]). \tag{2.33}$$

Example 2.10 *Consider Example 2.4 with $q_u = 1/3$, calculate $\tilde{\mathbb{E}}_2[S_3]$ and $\tilde{\mathbb{E}}_1[\tilde{\mathbb{E}}_2[S_3]]$ by definition.*

Solution:

$$\begin{aligned}
\tilde{\mathbb{E}}_2[S_3](HH) &= \frac{1}{3}\times 108 + \frac{2}{3}\times 36 = 60\\
\tilde{\mathbb{E}}_2[S_3](HT) &= \frac{1}{3}\times 36 + \frac{2}{3}\times 12 = 20\\
\tilde{\mathbb{E}}_2[S_3](TH) &= \frac{1}{3}\times 36 + \frac{2}{3}\times 12 = 20\\
\tilde{\mathbb{E}}_2[S_3](TT) &= \frac{1}{3}\times 12 + \frac{2}{3}\times 4 = \frac{20}{3}.
\end{aligned}
\qquad
\begin{aligned}
\tilde{\mathbb{E}}_1[\tilde{\mathbb{E}}_2[S_3]](H) &= \frac{1}{3}\times 60 + \frac{2}{3}\times 20 = \frac{100}{3}\\
\tilde{\mathbb{E}}_1[\tilde{\mathbb{E}}_2[S_3]](T) &= \frac{1}{3}\times 20 + \frac{2}{3}\times \frac{20}{3} = \frac{100}{9}.
\end{aligned}$$

Comparing with the second part of Example 2.9, we have verified that $\tilde{\mathbb{E}}_1[\tilde{\mathbb{E}}_2[S_3]] = \tilde{\mathbb{E}}_1[S_3]$.

Proof of Theorem 2.2: We only show the proof of iii). The proof won't be tested at all. It is presented so that one can get a feeling of why iii) makes sense.

Keep in mind that $n \le m \le N$. Denote $Z = \tilde{\mathbb{E}}_m[X]$. Then Z actually depends on $\omega_1 \cdots \omega_m$ only.

$$
\begin{aligned}
&\tilde{\mathbb{E}}_n[\tilde{\mathbb{E}}_m[X]](\omega_1 \cdots \omega_n) = \tilde{\mathbb{E}}_n[Z](\omega_1 \cdots \omega_n) \\
=& \sum_{\omega_{n+1},\cdots,\omega_N} q_u^{\#H(\omega_{n+1}\cdots\omega_N)} q_d^{\#T(\omega_{n+1}\cdots\omega_N)} Z(\omega_1 \cdots \omega_n \omega_{n+1} \cdots \omega_N) \\
=& \sum_{\omega_{n+1},\cdots,\omega_N} q_u^{\#H(\omega_{n+1}\cdots\omega_N)} q_d^{\#T(\omega_{n+1}\cdots\omega_N)} Z(\omega_1 \cdots \omega_n \omega_{n+1} \cdots \omega_m) \\
=& \sum_{\omega_{n+1},\cdots,\omega_N} \left(q_u^{\#H(\omega_{n+1}\cdots\omega_m)} q_d^{\#T(\omega_{n+1}\cdots\omega_m)} Z(\omega_1 \cdots \omega_m) \times q_u^{\#H(\omega_{m+1}\cdots\omega_N)} q_d^{\#T(\omega_{m+1}\cdots\omega_N)} \right) \\
=& \left(\sum_{\omega_{n+1},\cdots,\omega_m} q_u^{\#H(\omega_{n+1}\cdots\omega_m)} q_d^{\#T(\omega_{n+1}\cdots\omega_m)} Z(\omega_1 \cdots \omega_m) \right) \times \left(\sum_{\omega_{m+1},\cdots,\omega_N} q_u^{\#H(\omega_{m+1}\cdots\omega_N)} q_d^{\#T(\omega_{m+1}\cdots\omega_N)} \right) \\
=& \sum_{\omega_{n+1},\cdots,\omega_m} q_u^{\#H(\omega_{n+1}\cdots\omega_m)} q_d^{\#T(\omega_{n+1}\cdots\omega_m)} Z(\omega_1 \cdots \omega_m).
\end{aligned}
$$

In the last step, we have used $\sum_{\omega_{m+1},\cdots,\omega_N} q_u^{\#H(\omega_{m+1}\cdots\omega_N)} q_d^{\#T(\omega_{m+1}\cdots\omega_N)} = (q_u + q_d)^{N-m} = 1$ which is true because of (2.9) or the Pascal's triangle in Figure 2.5.

Then we continue to get

$$
\begin{aligned}
&\tilde{\mathbb{E}}_n[\tilde{\mathbb{E}}_m[X]](\omega_1 \cdots \omega_n) \\
=& \sum_{\omega_{n+1},\cdots,\omega_m} q_u^{\#H(\omega_{n+1}\cdots\omega_m)} q_d^{\#T(\omega_{n+1}\cdots\omega_m)} \times \left(\sum_{\omega_{m+1}\cdots\omega_N} q_u^{\#H(\omega_{m+1}\cdots\omega_N)} q_d^{\#T(\omega_{m+1}\cdots\omega_N)} X(\omega_1 \cdots \omega_N) \right) \\
=& \sum_{\omega_{n+1},\cdots,\omega_N} q_u^{\#H(\omega_{n+1}\cdots\omega_N)} q_d^{\#T(\omega_{n+1}\cdots\omega_N)} X(\omega_1 \cdots \omega_N) \\
=& \tilde{\mathbb{E}}_n[X](\omega_1 \cdots \omega_n). \qquad \square
\end{aligned}
$$

Remark:

Conditional expectation of X is the average of X over certain subset of Ω. One can consider X as the 1st image of Figure 2.7 which is the graph of a function g defined on a rectangle Ω. It contains all the information about X. But $\tilde{\mathbb{E}}_k[X]$ contains less information about X. For example, $\tilde{\mathbb{E}}_4[X]$ could be an image of resolution 64×32, $\tilde{\mathbb{E}}_2[X]$ could be an image of resolution 16×8, and so on.

For example, (2.31) with $n = 2$, $m = 3$, or $\tilde{\mathbb{E}}_2[\tilde{\mathbb{E}}_3[X]] = \tilde{\mathbb{E}}_2[X]$, says that image-4 in Figure 2.7 can be obtained in two different ways. (1) $\tilde{\mathbb{E}}_2[X]$: we can directly average image-1 to obtain image-4; (2) $\tilde{\mathbb{E}}_2[\tilde{\mathbb{E}}_3[X]]$: we can first average image-1 to image-3, and then we average image-3 to obtain image-4.

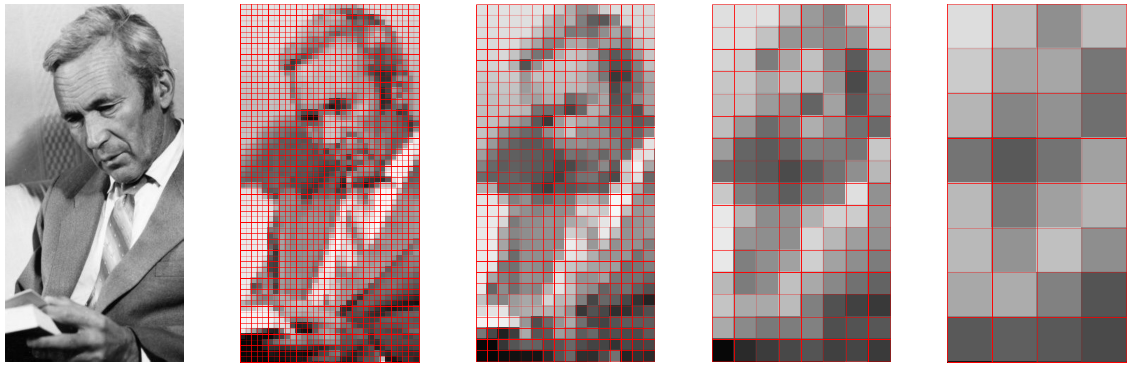


Figure 2.7: $X \leftarrow \tilde{\mathbb{E}}_4[X] \leftarrow \tilde{\mathbb{E}}_3[X] \leftarrow \tilde{\mathbb{E}}_2[X] \leftarrow \tilde{\mathbb{E}}_1[X]$.

2.8 Martingales

Definition 2.4 *Consider the binomial asset-pricing model. Let M_0, M_1, ... , M_N be a sequence of random variables, with each M_n depending only on the first n coin tosses (and M_0 constant).* [16]

i) If

$$M_n = \tilde{\mathbb{E}}_n[M_{n+1}], \qquad n = 0, 1, ..., N-1,$$

we say that this process is a *<u>martingale</u>*.

ii) If

$$M_n \leq \tilde{\mathbb{E}}_n[M_{n+1}], \qquad n = 0, 1, ..., N-1,$$

we say that this process is a *<u>submartingale</u>*.

iii) If

$$M_n \geq \tilde{\mathbb{E}}_n[M_{n+1}], \qquad n = 0, 1, ..., N-1,$$

we say that this process is a *<u>supermartingale</u>*.

Remark: The martingale property is a "one-step-ahead" condition. It implies a similar condition for any number of steps:

$$M_n = \tilde{\mathbb{E}}_n[M_{n+1}] = \tilde{\mathbb{E}}_n[\tilde{\mathbb{E}}_{n+1}[M_{n+2}]] \overset{(2.31)}{=} \tilde{\mathbb{E}}_n[M_{n+2}].$$

Iterating this argument, we can show that whenever $0 \leq n \leq m \leq N$,

$$M_n = \tilde{\mathbb{E}}_n[M_m]. \tag{2.34}$$

[16]Such a sequence of random variables is called an adapted stochastic process.

Theorem 2.3 *Consider the general binomial model. Let the risk-neutral probabilities be q_u, q_d. Then, in the risk-neutral world, the discounted stock price and the discounted option price are martingale, i.e.,*

$$\tilde{\mathbb{E}}_n[\frac{S_{n+1}}{e^{r(n+1)\delta t}}] = \frac{S_n}{e^{rn\delta t}}, \tag{2.35}$$

$$\tilde{\mathbb{E}}_n[\frac{ⓟ_{n+1}}{e^{r(n+1)\delta t}}] = \frac{ⓟ_n}{e^{rn\delta t}}, \tag{2.36}$$

Proof: (2.35) follows from (2.6) or (2.27) . (2.36) follows from (2.7) . □

Example 2.11 *Toss a coin repeatedly. Assume the probability of head on each toss is $\frac{1}{2}$, so is the probability of tail. Let $X_j = 1$ if the jth toss results in a head and $X_j = -1$ if the jth toss results in a tail. Consider $M_1, M_1, M_2, \cdots$ (which is an example of stochastic process) defined by $M_0 = 0$ and*

$$M_n = \sum_{j=1}^{n} X_j, \qquad n \geq 1.$$

This is called a symmetric random walk; with each head, it steps up one, and with each tails, it steps down one. Define $I_0 = 0$ and

$$I_n = \sum_{j=0}^{n-1} M_j(M_{j+1} - M_j), \quad n = 1, 2, \cdots . \tag{2.37}$$

Show that

$$I_n = \frac{1}{2}M_n^2 - \frac{n}{2}. \tag{2.38}$$

Let n be an arbitrary nonnegative integer, and let $f(y)$ be an arbitrary function of a variable y. In terms of n and f, find the function $g(x)$ satisfying

$$\tilde{\mathbb{E}}_n[f(I_{n+1})] = g(I_n). \tag{2.39}$$

Solution: $M_j(M_{j+1}-M_j) = \frac{1}{2}\left(M_{j+1}^2 - M_j^2 - (M_{j+1} - M_j)^2\right) = \frac{1}{2}\left(M_{j+1}^2 - M_j^2 - (X_{n+1})^2\right) = \frac{1}{2}\left(M_{j+1}^2 - M_j^2 - 1\right)$. Hence

$$I_n = \sum_{j=0}^{n-1} M_j(M_{j+1} - M_j) = \frac{1}{2}M_n^2 - \frac{n}{2}.$$

$$\begin{aligned}
&\tilde{\mathbb{E}}_n[f(I_{n+1})](\omega_1 \cdots \omega_n) \\
=&q_u f(I_{n+1})(\omega_1 \cdots \omega_n H) + q_d f(I_{n+1})(\omega_1 \cdots \omega_n T) \\
=&\frac{1}{2} f(I_n + M_n X_{n+1}(H))(\omega_1 \cdots \omega_n) + \frac{1}{2} f(I_n + M_n X_{n+1}(T))(\omega_1 \cdots \omega_n) \\
=&\frac{1}{2} f(I_n + M_n)(\omega_1 \cdots \omega_n) + \frac{1}{2} f(I_n - M_n)(\omega_1 \cdots \omega_n) \\
=&\frac{1}{2} f(I_n + \sqrt{2I_n + n})(\omega_1 \cdots \omega_n) + \frac{1}{2} f(I_n - \sqrt{2I_n + n})(\omega_1 \cdots \omega_n).
\end{aligned}$$

Hence $g(I_n) = \frac{1}{2}\left(f(I_n + \sqrt{2I_n + n}) + f(I_n - \sqrt{2I_n + n})\right)$
or $g(x) = \frac{1}{2}\left(f(x + \sqrt{2x + n}) + f(x - \sqrt{2x + n})\right)$.

2.9 Information, σ-algebra, and filtration

(From Chapter 2 of Shreve II.) Image that some random experiment is performed and the outcome is a particular ω in the set of all possible outcomes Ω. We might be given some information – not enough to know the precise value of ω, but enough to narrow down the possibilities. For example, the true ω might be the result of three coin tosses, and we are told only the first one. Or we are only told the price of the stock S at time $t = 2$ without being told any of the coin toss. In such a situation, we can make a list of sets that are sure to contain the true ω and other sets that are sure not to contain the true ω. These are the sets that are **resolved** by the information. Please note that the word "resolve" does not mean solve or make a decision here, but means you can separate the whole thing (the sample space Ω) into two or more parts and you are able to distinguish between them. More precisely, a set of outcomes is resolved by certain information (e.g. result of ω_1) if and only if we can tell whether ω belongs to the set or not once we know that information (e.g. result of ω_1).

Suppose Ω is the possible outcomes of three coin tosses. If we are told the outcome of the first toss only, the following sets of outcomes are resolved

$$A_H = \{HHH, HHT, HTH, HTT\}, \quad A_T = \{THH, THT, TTH, TTT\}.$$

The empty set $\emptyset$ and the whole sample space Ω (see (2.15)) are always resolved, even without any information; the true ω does not belong to $\emptyset$ and does belong to Ω. The four sets that are resolved by the first coin toss from a set of sets (whose elements are themselves all sets, it will be called a σ-algebra later in Definition 2.5)[17]

$$\mathcal{F}_1 = \{\emptyset, \Omega, A_H, A_T\}.$$

We shall think of $\mathcal{F}_1$ as containing the information learned by observing the first coin toss.

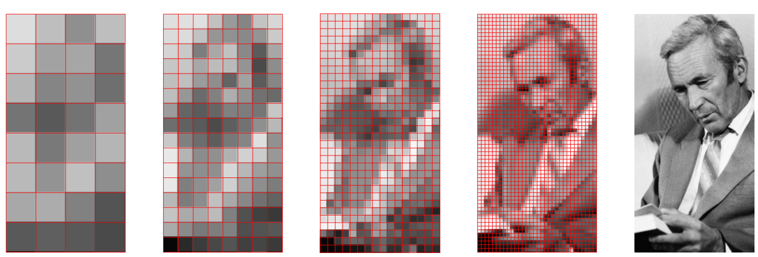


Figure 2.8: More is resolved as you move to the right.

[17] If set A is resolved, so is its complement A^c. If sets A and B are resolved, so are $A \cup B$.

If we are told the first two coin tosses, four additional sets

$$\begin{aligned} &A_{HH} = \{HHH, HHT\},\ A_{HT} = \{HTH, HTT\}, \\ &A_{TH} = \{THH, THT\},\ A_{TT} = \{TTH, TTT\} \end{aligned} \tag{2.40}$$

are resolved. Of course, the sets in $\mathcal{F}_1$ are still resolved. But we now have a higher resolution. Whenever a set is resolved, so is its complement, which means A_{HH}^c, A_{HT}^c, A_{TH}^c, and A_{TT}^c are resolved. Whenever two sets are resolved, so is their union. Finally, we have 16 resolved sets that together form a set of sets called $\mathcal{F}_2$; i.e.,

$$\mathcal{F}_2 = \left\{ \begin{array}{l} \emptyset, \Omega, A_H, A_T, A_{HH}, A_{HT}, A_{TH}, A_{TT}, A_{HH}^c, A_{HT}^c, A_{TH}^c, A_{TT}^c \\ A_{HH} \cup A_{TH}, A_{HH} \cup A_{TT}, A_{HT} \cup A_{TH}, A_{HT} \cup A_{TT} \end{array} \right\} \tag{2.41}$$

We shall think of $\mathcal{F}_2$ as containing the information learned by observing the first two coin tosses. In other words, (I) if we know the first two coin tosses, we can tell for each set in $\mathcal{F}_2$, whether the true ω belongs to it; (II) $\mathcal{F}_2$ contains all such sets.

Example 2.12 *Question: does the set $\{HHH\}$ belongs to $\mathcal{F}_2$? The answer is no because if the first two coin tosses is HH, we cannot tell whether the true ω belongs to $\{HHH\}$ or not. We are only able to tell the true $\omega \in \{HHH, HHT\}$ which is A_{HH}. To be able to tell whether the true $\omega \in \{HHH\}$ or not (to get higher resolution), we need to observe all three coin tosses.*

The way we construct $\mathcal{F}_2$ motivates the following definition:

Definition 2.5 *Let Ω be a nonempty set, and let $\mathcal{F}$ be a collection of subsets of Ω. We say that $\mathcal{F}$ is a* **σ-algebra** *(or* **tribe***) if the following three conditions are satisfied*

i) the empty set $\emptyset$ belongs to $\mathcal{F}$.

ii) whenever a set A belongs to $\mathcal{F}$, its complement A^c also belongs to $\mathcal{F}$.

iii) whenever a sequence of sets $A_1, A_2, \cdots$ belongs to $\mathcal{F}$, their union $\cup_{n=1}^{\infty} A_n$ also belongs to $\mathcal{F}$.

Remark: From *ii*) and *iii*), we also have that

iv) whenever a sequence of sets $A_1, A_2, \cdots$ belongs to $\mathcal{F}$, their intersection $\cap_{n=1}^{\infty} A_n = (\cup_{n=1}^{\infty} A_n^c)^c$ also belongs to $\mathcal{F}$.

Hence the resolved sets form a σ-algebra. For your information, a set is an "algebra" if the set is closed under certain operation or operations. It means that if one take any or some elements, do the operation, the result is still in the set. This is a way to generate new elements from known elements. This is a standard way to enlarge the subjects we are studying. For example, starting from number 1, by addition and subtraction, we can generate all integers. $\{1; \quad +, -\} \to \mathbb{Z}$ (learned in kindergarten). If we also include multiplication and division, we generate all rational numbers $\{1; \quad +, -, /, \times\} \to \mathbb{Q}$ (learned in primary school). If we consider taking limits, we obtain all real numbers $\{1; \quad +, -, /, \times, \text{limits}\} \to \mathbb{R}$ (learned in high school or unversity). For the same reason, if we add in an additional element $\sqrt{-1}$ at the beginning, $\{1, \sqrt{-1}; \quad +, -, /, \times, \text{limits}\} \to \mathbb{C}$. Now, we are saying $\{\text{some sets } A; \quad \text{complement}, \cup\} \to \sigma\text{-algebra generated by } A$ (learned in graduate school).

If we are told all three coin tosses, we know the true ω and every subset of Ω is resolved. There are $2^{2^3} = 256$ subsets of Ω, and taken all together, they constitute the σ-algebra $\mathcal{F}_3$:

$$\mathcal{F}_3 = \text{ The set of all subsets of } \Omega.$$

If we are told nothing about the coin tosses, the only resolved sets are $\emptyset$ and Ω. We form the so called trivial σ-field $\mathcal{F}_0$ with these two sets:

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

We have then four σ-algebra, $\mathcal{F}_0$, $\mathcal{F}_1$, $\mathcal{F}_2$, and $\mathcal{F}_3$, indexed by time. As time moves forward, we obtain finer resolution. In other words, of $n < m$, $\mathcal{F}_n \subset \mathcal{F}_m$. This means $\mathcal{F}_m$ contains more information than $\mathcal{F}_n$. The collection of σ-algebras $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ is an example of a filtration.

Even though we have only discussed the discrete-time version, we now directly give a continuous-time formulation of filtration.

Definition 2.6 *(Definition 2.1.1 of Shreve II) Let Ω be a nonempty set. Let T be a fixed positive number, and assume that for each $t \in [0, T]$ there is a σ-algebra $\mathcal{F}_t$ of subset of Ω. Assume further that if $s \leq t$, then every set in $\mathcal{F}_s$ is also in $\mathcal{F}_t$. Then we call the collection of σ-algebras $\mathcal{F}_t$, $0 \leq t \leq T$, a filtration.*

A filtration tells us the information we will have at future times. More precisely, when we get to time t, we will know for each set in $\mathcal{F}_t$ whether the true ω lies in that set.

Example 2.13 *Suppose our sample space is $\Omega = C_0[0, T]$, the set of continuous functions defined on $[0, T]$ taking the value zero at time zero. The sets that are resolved by time t are just those sets that can be described in terms of the path of ω up to time t. For example,*

- *the set $\{\omega \in \Omega : \max_{0 \leq s \leq t} \omega(s) \leq 1\}$ is resolved and belongs to $\mathcal{F}_t$.*
- *if $t < T$, the set $\{\omega \in \Omega : \omega(T) > 0\}$ is not resolved by time t and does not belong to $\mathcal{F}_t$.*

Besides observing the evolution of an economy over time, there is a second way we might acquire information about the value of ω. Let X be a random variable that depends on ω. Suppose that rather than being told the value of ω, we are told only the value of $X(\omega)$, say, 5, then we know whether ω is in a set, say, $\{\omega : X(\omega) \leq 1\}$. For example, when we know $X(\omega) = 5$, we know ω is not in $\{\omega : X(\omega) \leq 1\}$.

Definition 2.7 *Let X be a random variable defined on a nonempty sample space Ω. The σ-algebra generated by X, denoted by $\sigma(X)$, is the collection of all subsets of Ω of the form $\{\omega : X(\omega) \in B\}$ where B ranges over the Borel subset of $\mathbb{R}^n$.*

For technical reasons, in the above definition, we require B to be a Borel set[18]

Example 2.14 *We return to Example 2.4 for the three-period model. Ω is the set of eight possible outcomes of three coin tosses and S_2 is defined in Example 2.4. Let $B = \{72\}$, then $\{\omega : S_2 \in B\} = A_{HH} = \{HHH, HHT\}$. Let $B = \{72, 24\}$, then $\{\omega : S_2 \in B\} = A_{HH} \cup A_{HT} \cup A_{TH}$. If we let B range over the Borel sets of $\mathbb{R}$, we will obtain the list of the set*

$$\emptyset, \Omega, A_{HH}, A_{TT}, A_{HT} \cup A_{TH}$$

and all unions and complements of them. This is the σ-algebra $\sigma(S_2)$. One can think of $\sigma(S_2)$ as the information on $\omega_1\omega_2\omega_3$ contained in knowing the value of S_2. Note that it is not precisely the value of $\omega_1\omega_2$ since we can not distinguish HT and TH by knowing the value of S_2.

Recall the $\mathcal{F}_2$ defined in (2.41). It represent the information contained in the first two coin tosses. Note that $\sigma(S_2) \subset \mathcal{F}_2$. This means that there is enough information in $\mathcal{F}_2$ to determine the value of S_2 [19] and even more. We say that S_2 is $\mathcal{F}_2$ measurable.

By the way, in this example, $\mathcal{F}_2$ is strictly larger than $\sigma(S_2)$. For example, $A_{HT} \in \mathcal{F}_2$ but $\sigma(S_2)$. This means that knowing the value of S_2 does not tell us everything about the first two coin tosses. In fact, we cannot tell HT from TH if all we know is the value of S_2.

Example 2.15 *In Question 5 of Homework II, $\Phi_n - \Delta_n S_n$ is the amount of money invested in the money market at $t_n = n\delta t$, while Δ_n is the number of stocks held in the time interval $[t_n, t_{n+1}]$. Note that the Δ_n defined by (2.45) is a function of $\omega_1 \cdots \omega_n$ and is $\mathcal{F}_n$ measurable.*

Remark: Note that S_2 is not $\mathcal{F}_1$ measurable as $\mathcal{F}_1 = \{\emptyset, \Omega, A_H, A_T\}$ is too coarse to distinguish the different values S_2 can take.

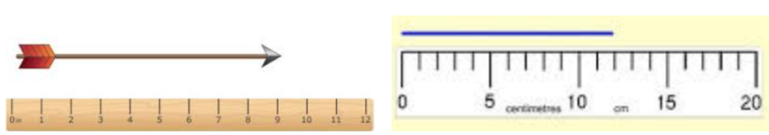


Figure 2.9: Pictures from the internet. Left: the arrow is not measurable by the ruler. To be able to measure the arrow (X), the ruler ($\mathcal{F}$ is the collection of the ticks on the ruler) needs to have finer scale ticks. Right: measurable.

[18]The σ-algebra obtained by beginning with cross products of open intervals $(a_1, b_1) \times (a_2, b_2) \cdots \times (a_n, b_n)$ and then adding everything else necessary in order to form a σ-algebra is called the Borel σ-algebra of $\mathbb{R}^n$. (In most case, we just need $n = 1$.) If B is a set in the Borel σ-algebra, it is called a Borel set.

[19]If for every subset (also called event) A in $\mathcal{F}_2$ (or $\sigma(S_2)$), I know whether ω is in it or not (meaning I know whether A happens or not), then I can determine the value of S_2.

Definition 2.8 *Let X be a random variable (i.e. a function) defined on a nonempty sample space Ω. Let $\mathcal{G}$ be a σ-algebra of subsets of Ω. If every set in $\sigma(X)$ is also in $\mathcal{G}$, we say that X is $\mathcal{G}$-measurable.*

Remark: Combining Definitions 2.7, 2.8 and the definition of Boreal set, we can say that random variable X, which is a function that maps a $\omega \in \Omega$ to a number in $\mathbb{R}$, is $\mathcal{G}$-measurable if

$$X^{-1}(U) \stackrel{\text{def}}{=} \{\omega \in \Omega : X(\omega) \in U\} \in \mathcal{G} \tag{2.42}$$

for all closed interval $U \in \mathbb{R}$. Equivalently, we can also require U be to all closed interval of $\mathbb{R}$.

We shall give another interpretation of measurability at the beginning of Chapter 4.

A random variable X is $\mathcal{G}$-measurable if and only if the information in $\mathcal{G}$ is sufficient to determine the value of X. If X is $\mathcal{G}$-measurable, then $f(X)$ is also $\mathcal{G}$-measurable for any Borel-measurable function f; if the information in $\mathcal{G}$ is sufficient to determine the value of X, it will also determine the value of $f(X)$. If X and Y are $\mathcal{G}$-measurable, then $f(X, Y)$ is $\mathcal{G}$-measurable for any Borel-measurable function $f(x, y)$ of two variables.

Why should investors know about measurability? A portfolio Φ at time t must be $\mathcal{F}_t$-measurable, as investors must depend solely on information available to the investor at time t to adjust the investment strategy. Think about the last term in (1.6) (with $n = 1$): $\phi_{t_m}[S_{t_{m+1}} - S_{t_m}]$. At time t_m, the invested decided to hold ϕ_{t_m} shares of stock. Then by t_{m+1}, his wealth increases by $\phi_{t_m}[S_{t_{m+1}} - S_{t_m}]$. We cannot use $\phi_{t_{m+1}}[S_{t_{m+1}} - S_{t_m}]$ as the investment strategy cannot depend on information not yet available. Later we will introduce Itô integral $\int_0^T \phi_t dS_t$ as a limiting process of $\sum_{m=0}^{N/\Delta t-1} \phi_{it_m}[S_{it_{m+1}} - S_{it_m}]$ where we require ϕ_t to be $\mathcal{F}_t$-measurable.

See (2.46) of Question 5 Homework II for an example of how to adjust the investment strategy.

Definition 2.9 *(Definition 2.1.6 of Shreve II) Let Ω be a nonempty sample space equipped with a filtration $\mathcal{F}_t$, $0 \le t \le T$. Let $\{X_t, 0 \le t \le T\}$ be a collection of random variables indexed by t. We say this collection of random variables is an $\mathcal{F}_t$-adapted stochastic process if, for each t, the random variable X_t is $\mathcal{F}_t$-measurable.*

Remark: See the first half of Definition 2.4 for the definition of discrete adapted stochastic process.

In continuous-time finance models, asset prices, portfolio processes will all be adapted to a filtration that we regard as a model of the flow of public information.

2.10 Homework II

(Only submit solutions to Questions 1,2,3,6,10,15.)

1. Using the following three-step binomial tree to compute the price of an European call option with strike price \$90 and $T = 6$ months. The initial price for the underlying stock is \$80. $r = 0.05$ per annum. $\delta t = 2$ months. $u = \frac{5}{4}$. $d = \frac{4}{5}$.

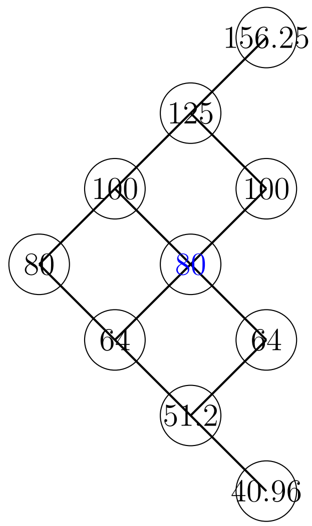


Figure 2.10: A binomial tree

2. (stochastic volatility, random interest rate) Consider a binomial pricing model, but at each time $n \geq 1$, the "up factor" $u_n(\omega_1 \cdots \omega_n)$, the "down factor" $d_n(\omega_1 \cdots \omega_n)$, and the interest rate $r_n(\omega_1 \cdots \omega_n)$ are allowed to depend on n and on the first n coin tosses $\omega_1 \cdots \omega_n$. The initial up factor u_0, the initial down factor d_0, and the initial interest rate r_0 are not random. More specifically, the stock price at time one is given by

$$S_1(\omega_1) = \begin{cases} u_0 S_0 & \text{if } \omega_1 = H, \\ d_0 S_0 & \text{if } \omega_1 = T, \end{cases}$$

and, for $n \geq 1$, the stock price at time $n+1$ is given by

$$S_{n+1}(\omega_1 \cdots \omega_{n+1}) = \begin{cases} u_n(\omega_1 \cdots \omega_n) S_n(\omega_1 \cdots \omega_n) & \text{if } \omega_{n+1} = H, \\ d_n(\omega_1 \cdots \omega_n) S_n(\omega_1 \cdots \omega_n) & \text{if } \omega_{n+1} = T. \end{cases}$$

One dollar invested in or borrowed from the money market at time $t = 0$ grows to an invest or debt of $\rho = e^{r_0 \delta t}$ at time $t_1 = \delta t$, and, for $n \geq 1$, one dollar invested in or borrowed from the money market at time t_n grows to an investment or debt of $e^{r_n(\omega_1 \cdots \omega_n)\delta t}$ at time t_{n+1}. We assume that for each n and for all $\omega_1 \cdots \omega_n$, the no-arbitrage condition

$$0 < d_n(\omega_1 \cdots \omega_n) < \rho_n = e^{r_n(\omega_1 \cdots \omega_n)\delta t} < u_n(\omega_1 \cdots \omega_n)$$

holds. We also assume that $0 < d_0 < \rho_0 = e^{r_0 \delta t} < u_0$.

i) Let N be a positive integer and let U_n be the price at time $t_n = n\delta t$ of a derivative security. Derive the formula that relates $U_n(\omega_1 \cdots \omega_n)$ to random variables $U_{n+1}(\omega_1 \cdots \omega_n H)$, $U_{n+1}(\omega_1 \cdots \omega_n T)$, $q_{u,n}(\omega_1 \cdots \omega_n) = \frac{\rho_n(\omega_1 \cdots \omega_n) - d_n(\omega_1 \cdots \omega_n)}{u_n(\omega_1 \cdots \omega_n) - d_n(\omega_1 \cdots \omega_n)}$, and $q_{d,n}(\omega_1 \cdots \omega_n) = \frac{u_n(\omega_1 \cdots \omega_n) - \rho_n(\omega_1 \cdots \omega_n)}{u_n(\omega_1 \cdots \omega_n) - d_n(\omega_1 \cdots \omega_n)}$.

ii) Suppose the initial stock price is $S_0 = 80$, with each head the stock price increase by 10, and with each tail the stock price decrease by 10. In another words, $S_1(H) = 90$, $S_1(T) = 70$, $S_2(HH) = 100$. Assume that the interest rate is always zero. Consider a European call with strike price 80, expiring at $t_3 = 3\delta t$. What is the price of this call at time $t = 0$?

3. Consider a European call option on an underlying stock with its present price S_0 =\$50 per share. Suppose that at the expiry date T the stock has only two possible values $S^u = \$80$ and $S^d = \$40$. Assume the strike price $K = \$60$ and the risk-free interest rate is $r = 0$.

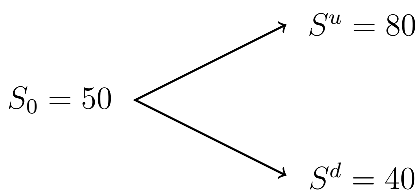


Figure 2.11: A single period binomial tree

Here is how to determine the option price by replication: If we construct a portfolio at $t = 0$ which consists of a debt of \$20 and $\frac{1}{2}$ shares of the stock

$$\Phi = -20 + \frac{1}{2}S.$$

Show that $V_T(\Phi) = V_T(\textcircled{p})$ no matter whether the stock price will go up or go down at $t = T$. We say that such a portfolio replicates the given option. Then use Corollary 1.1 to determine $\textcircled{p}_0 = V_0(\textcircled{p})$. Is this value the same as the one determined by (2.5)?

4. (A generalization of the last problem: Replication in the one period binomial model) Given an option, consider a portfolio Φ consisting of a risk-free asset with interest rate r and an underlying stock:

$$\Phi = B + \Delta S \tag{2.43}$$

where Δ is the number of shares of the underlying stock. We would like to choose $B_0 = V_0(B)$ and Δ so that $V_T(\Phi) = V_T(\textcircled{p}) = \begin{cases} \textcircled{p}^u & \text{if } V_T(S) = S^u, \\ \textcircled{p}^d & \text{if } V_T(S) = S^d, \end{cases}$. Find the formula for B_0 and Δ using r, T, $\textcircled{p}^u$, $\textcircled{p}^d$, S_0, S^u, and S^d. Can you determine $\textcircled{p}_0$ using Corollary 1.1?

Solution:

$$\begin{cases} e^{rT}B_0 + \Delta S^u = \textcircled{p}^u \\ e^{rT}B_0 + \Delta S^d = \textcircled{p}^d \end{cases}$$

or equivalently

$$\begin{bmatrix} e^{rT} & S^u \\ e^{rT} & S^d \end{bmatrix} \begin{bmatrix} B_0 \\ \Delta \end{bmatrix} = \begin{bmatrix} \textcircled{p}^u \\ \textcircled{p}^d \end{bmatrix}.$$

Hence

$$\begin{bmatrix} B_0 \\ \Delta \end{bmatrix} = \frac{1}{e^{rT}(S^d - S^u)} \begin{bmatrix} S^d & -S^u \\ -e^{rT} & e^{rT} \end{bmatrix} \begin{bmatrix} \textcircled{p}^u \\ \textcircled{p}^d \end{bmatrix}.$$

Thus

$$\begin{aligned} B_0 &= \frac{1}{e^{rT}} \left(\frac{-S^d}{S^u - S^d} \textcircled{p}^u + \frac{S^u}{S^u - S^d} \textcircled{p}^d \right), \\ \Delta &= \frac{\textcircled{p}^u - \textcircled{p}^d}{S^u - S^d}. \end{aligned}$$

We mention in passing that the Δ used to build the portfolio (2.43) that replicate a option equals to the Δ used to build a risk-free portfolio in (2.2).

$$\textcircled{p}_0 = V_0(\Phi) = B_0 + \Delta S_0 = \frac{1}{\rho} \left(\frac{\rho S_0 - S^d}{S^u - S^d} \textcircled{p}^u + \frac{S^u - \rho S_0}{S^u - S^d} \textcircled{p}^d \right)$$

with $\rho = e^{rT}$. The above formula for $\textcircled{p}_0$ is the same as the formula we derived in (2.5).

5. (Replication in the multiperiod binomial model) Consider the multiperiod binomial model introduced in Section 2.3. Suppose we define

$$\textcircled{p}_n(\omega_1 \cdots \omega_n) = e^{-r\delta t} \left(q_u \textcircled{p}_{n+1}(\omega_1 \cdots \omega_n H) + q_d \textcircled{p}_{n+1}(\omega_1 \cdots \omega_n T) \right) \tag{2.44}$$

and

$$\Delta_n(\omega_1 \cdots \omega_n) = \frac{\textcircled{p}_{n+1}(\omega_1 \cdots \omega_n H) - \textcircled{p}_{n+1}(\omega_1 \cdots \omega_n T)}{S_{n+1}(\omega_1 \cdots \omega_n H) - S_{n+1}(\omega_1 \cdots \omega_n T)}, \tag{2.45}$$

with $n = N-1, \cdots, 0$. Prove that if we set $\Phi_0 = V_0$ and define recursively forward in time the portfolio values $\Phi_1, \Phi_2, \cdots \Phi_N$ by

$$\Phi_{n+1} = e^{r\delta t} \left(\Phi_n - \Delta_n S_n \right) + \Delta_n S_{n+1}, \tag{2.46}$$

then

$$\Phi_N(\omega_1 \cdots \omega_N) = \textcircled{p}_N(\omega_1 \cdots \omega_N) \qquad \text{for all } \omega_1 \cdots \omega_N. \tag{2.47}$$

Proof: We prove by induction in n that

$$\Phi_n(\omega_1 \cdots \omega_n) = \textcircled{p}_n(\omega_1 \cdots \omega_n) \qquad \text{for all } \omega_1 \cdots \omega_n \tag{2.48}$$

for $n = 0, ..., N$.

We know (2.48) is true when $n = 0$. We assume (2.48) is true for n and show that it is true for $n+1$.

By (2.46),

$$\Phi_{n+1}(\omega_1\cdots\omega_n H) = e^{r\delta t}\left(\Phi_n(\omega_1\cdots\omega_n) - \Delta_n(\omega_1\cdots\omega_n)S_n(\omega_1\cdots\omega_n)\right) + \Delta_n(\omega_1\cdots\omega_n)uS_n(\omega_1\cdots\omega_n).$$

To simplify the notation, we suppress $\omega_1\cdots\omega_n$ and write the equation simply as

$$\Phi_{n+1}(H) = e^{r\delta t}\left(\Phi_n - \Delta_n S_n\right) + \Delta_n u S_n. \tag{2.49}$$

With $\omega_1\cdots\omega_n$ suppressed, (2.45) can be written as

$$\Delta_n = \frac{ⓟ_{n+1}(H) - ⓟ_{n+1}(T)}{S_{n+1}(H) - S_{n+1}(T)} = \frac{ⓟ_{n+1}(H) - ⓟ_{n+1}(T)}{(u-d)S_n}.$$

Plugging the above equation into (2.49), we get

$$\begin{aligned}
\Phi_{n+1}(H) &= e^{r\delta t}\Phi_n + \left(u - e^{r\delta t}\right)\Delta_n S_n = e^{r\delta t}\Phi_n + \left(u - e^{r\delta t}\right)\frac{ⓟ_{n+1}(H) - ⓟ_{n+1}(T)}{u-d} \\
&\overset{\text{induction assumption and def of } q_d}{=} e^{r\delta t}ⓟ_n + q_d\left(ⓟ_{n+1}(H) - ⓟ_{n+1}(T)\right) \\
&\overset{(2.44)}{=} q_u ⓟ_{n+1}(H) + q_d ⓟ_{n+1}(T) + q_d\left(ⓟ_{n+1}(H) - ⓟ_{n+1}(T)\right) \\
&= ⓟ_{n+1}(H).
\end{aligned}$$

A similar argument shows that $\Phi_{n+1}(T) = ⓟ_{n+1}(T)$.

6. (The call can never be less risky than the underlying stock) With the same setup as in Example 2.8, consider a one period binomial tree model. Show that for put option, $\Omega \le 0$, and for call option $\Omega \ge 1$.

   [Hint: For call option, you need to prove $ⓟ^u - ⓟ^d \ge (u-d)ⓟ_0 > 0$. Use (2.5) to show that you only need to prove $dⓟ^u - uⓟ^d \ge 0$. Now, use $ⓟ^u = \max(uS_0 - K, 0)$ and $ⓟ^d = \max(dS_0 - K, 0)$ for the one period model.] As a matter of fact, this result is also true for the multi-period model. See Page 187 of "Options Markets" by Cox and Rubinstein.

7. (Bernoulli Random Variable) An experiment, whose outcome can be classified as either a success or a failure is performed. Let $X = 1$ when the outcome is a success, and $X = 0$ if the outcome is a failure. Then the probability mass function of X is given by

$$\mathbb{P}(X=0) = 1-p$$
$$\mathbb{P}(X=1) = p$$

   where $p \in [0,1]$ is the probability that the trial is a success. A random variable X is said to be a **Bernoulli random variable** if its probability mass function is given as above. Prove that $E[X] = p$ and $\mathrm{Var}[X] = p(1-p)$.

   Solution:

$$E[X] = 0 \times \mathbb{P}(X=0) + 1 \times \mathbb{P}(X=1) = p.$$
$$E[X^2] = 0^2 \times \mathbb{P}(X=0) + 1^2 \times \mathbb{P}(X=1) = p.$$
$$\mathrm{Var}[X] = E[X^2] - (E[X])^2 = p - p^2 = p(1-p).$$

8. If X is a discrete random variable with

$$\mathbb{P}(X = a) = p$$
$$\mathbb{P}(X = b) = 1 - p$$

where $p \in [0, 1]$. Prove that $E[X] = ap + b(1 - p)$ and $\mathrm{Var}[X] = (a - b)^2 p(1 - p)$.

Solution:

$$E[X] = a \times \mathbb{P}(X = a) + b \times \mathbb{P}(X = b) = ap + b(1 - p).$$
$$E[X^2] = a^2 p + b^2(1 - p).$$

$$\begin{aligned}\mathrm{Var}(X) = E[X^2] - (E[X])^2 &= a^2 p + b^2(1 - p) - (ap + b(1 - p))^2 \\ &= (a - b)^2 p(1 - p).\end{aligned}$$

9. (Binomial Random Variable) Suppose n independent trials, each results in a success with probability p or in a failure with probability $1 - p$, are to be performed. If Y represents the number of successes occur in the n trials, then Y is said to be **binomial random variable** with parameters (n, p), and denoted as $Y \sim B(n, p)$. Prove that $E[Y] = np$ and $\mathrm{Var}[Y] = np(1 - p)$.

Solution: Note that

$$P(Y = i) = \binom{n}{k} p^i (1 - p)^{n-i} \quad \text{with } i = 0, 1, 2, ..., n.$$

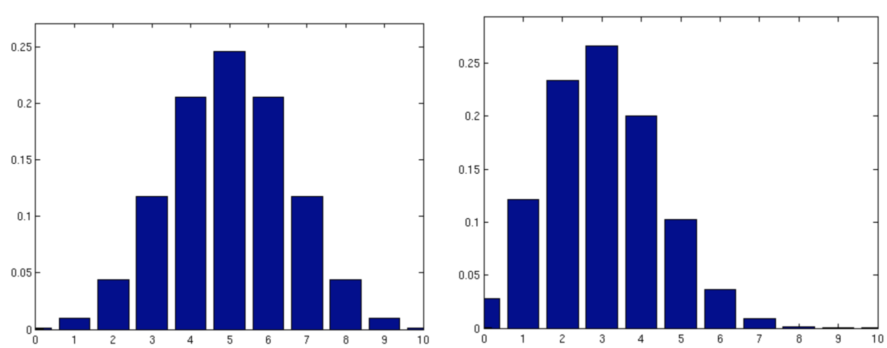


Figure 2.12: $\binom{n}{i} p^i (1 - p)^{n-i}$ with $n = 10$, $i = 0, 1, 2, \ldots, n$. Left: $p = 0.5$. Right: $p = 0.3$

Then, we can prove the statements using definition. We start with computing the kth order moment

$$
\begin{aligned}
E[Y^k] &= \sum_{i=0}^{n} i^k \mathbb{P}(Y=i) = \sum_{i=0}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i} \\
&= \sum_{i=0}^{n} i^{k-1} np \binom{n-1}{i-1} p^{i-1}(1-p)^{(n-1)-(i-1)} \\
&= \sum_{i=1}^{n} i^{k-1} np \binom{n-1}{i-1} p^{i-1}(1-p)^{(n-1)-(i-1)} \\
&\overset{j=i-1}{=} np \sum_{j=0}^{n-1} (j+1)^{k-1} \binom{n-1}{j} p^j (1-p)^{(n-1)-j} \\
&= npE[(Z+1)^{k-1}] \qquad (2.50)
\end{aligned}
$$

with $Z \sim B(n-1,p)$. Hence

$$E[Y] = npE[(Z+1)^0] = npE[1] = np. \qquad (2.51)$$

$$
\begin{aligned}
E[Y^2] &= npE[(Z+1)] = np\,(E[Z]+1) \overset{(2.51)}{=} np\,((n-1)p+1)\,. \qquad (2.52) \\
\mathrm{Var}[Y] &= E[Y^2] - (E[Y])^2 = np\,((n-1)p+1) - (np)^2 = np(1-p).
\end{aligned}
$$

For later reference, I would like to present another (easier) proof using a decomposition:

$$Y = X_1 + X_2 + \cdots + X_n$$

where $X_i = \begin{cases} 1 & \text{if the } i\text{th trial is a success} \\ 0 & \text{if the } i\text{th trial is a failure} \end{cases}$ and X_i's are independent. The definition of independence will be introduced in the next Chapter where one can show that if U and V are independent, then $\mathrm{Var}[U+V] = \mathrm{Var}[U] + \mathrm{Var}[V]$. Note that we already have $E[U+V] = E[U] + E[V]$ no matter U, V are independent or not. Since each X_i is a Bernoulli random variable, $E[X] = p$ and $\mathrm{Var}[X] = p(1-p)$. Hence

$$E[Y] = E[X_1] + E[X_2] + \cdots + E[X_n] = np,$$

$$\mathrm{Var}[Y] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2] + \cdots + \mathrm{Var}[X_n] = np(1-p).$$

10. (Skewness) The skewness of a random variable X is defined to be

$$Sk = E\left[\left(\frac{X - E[X]}{\sigma}\right)^3\right]$$

where $\sigma = \sqrt{\mathrm{Var}[X]}$. Prove that the skewness of $B(n,p)$ distribution is

$$Sk(n,p) = \frac{1-2p}{\sqrt{np(1-p)}}.$$

[Hint: Use (2.50) and $(x+y)^3 = x^3 + 3x^2y + 3xy^2 + y^3$.] [20]

11. Suppose X is a random variable. Define $M_k = \tilde{\mathbb{E}}_k[X]$. Prove that $\{M_k\}$ is a martingale.

    Proof: $\tilde{\mathbb{E}}_n[M_{n+1}] = \tilde{\mathbb{E}}_n[\tilde{\mathbb{E}}_{n+1}[X]] \overset{(2.31)}{=} \tilde{\mathbb{E}}_n[X] = M_n$.

12. (random walk) Toss a coin repeatedly. Assume the probability of head on each toss is $\frac{1}{2}$, so is the probability of tail. Let $X_j = 1$ if the jth toss results in a head and $X_j = -1$ if the jth toss results in a tail. Consider $M_1, M_1, M_2, \cdots$ (which is an example of stochastic process) defined by $M_0 = 0$ and

$$M_n = \sum_{j=1}^{n} X_j, \qquad n \geq 1.$$

    This is called a symmetric random walk; with each head, it steps up one, and with each tails, it steps down one. Using Theorem 2.2 to show that $M_1, M_1, M_2, \cdots, M_n, \cdots$ is a martingale.

    Proof: M_n only depends on $\omega_1 \cdots \omega_n$.

$$\tilde{\mathbb{E}}_n[M_{n+1}] = \tilde{\mathbb{E}}_n[M_n + X_{n+1}] = \tilde{\mathbb{E}}_n[M_n] + \tilde{\mathbb{E}}_n[X_{n+1}] = M_n + \mathbb{E}[X_{n+1}] = M_n.$$

13. (discrete-time stochastic integral) Suppose $M_0, M_1, \cdots, M_N$ is a martingale, and let $\Delta_0, \Delta_1, \cdots, \Delta_{N-1}$ be an adapted process (see definition 2.4). Define the discrete-time stochastic integral $I_0, I_1, \cdots, I_N$ by setting $I_0 = 0$ and

$$I_n = \sum_{j=0}^{n-1} \Delta_j (M_{j+1} - M_j), \qquad n = 1, ..., N. \tag{2.53}$$

    Show that $I_0, I_1, \cdots, I_N$ is a martingale.

    Proof: I_n only depends on $\omega_1 \cdots \omega_n$ and $I_0, I_1, \cdots, I_N$ is therefore an adapted stochastic process.

$$\begin{aligned}\tilde{\mathbb{E}}_n[I_{n+1}] =& \tilde{\mathbb{E}}_n[I_n + \Delta_n(M_{n+1} - M_n)] \overset{(2.29)}{=} \tilde{\mathbb{E}}_n[I_n] + \tilde{\mathbb{E}}_n[\Delta_n(M_{n+1} - M_n)] \\ \overset{(2.30)}{=}& I_n + \Delta_n \tilde{\mathbb{E}}_n[M_{n+1} - M_n] \\ =& I_n.\end{aligned}$$

[20] For your information, the kurtosis of a random variable X is defined to be

$$K = E\left[\left(\frac{X - E[X]}{\sigma}\right)^4\right]$$

where $\sigma = \sqrt{\mathrm{Var}[X]}$. Using the same idea as in the computation of $Sk(n,p)$, one can prove that the kurtosis of $B(n,p)$ distribution is

$$K(n,p) = 3 + \frac{1 - 6p(1-p)}{np(1-p)}.$$

You are not asked to prove the above formula of $K(n,p)$ in the homework.

In the last step, we have used $\tilde{\mathbb{E}}_n[M_{n+1} - M_n] = \tilde{\mathbb{E}}_n[M_{n+1}] - M_n \overset{M_n \text{ is martingale}}{=} 0$.

14. (i) Consider the dice-toss space similar to the coin-toss space. A typical element ω in this space is an infinite sequence $\omega = \omega_1\omega_2\omega_3\cdots$ with $\omega_i \in \{1, 2, \cdots, 6\}$. Define a random variable

$$X(\omega) = \omega_1$$

and a function

$$f(x) = \begin{cases} 1 & \text{if } x \geq 4 \\ 0 & \text{if } x < 4 \end{cases}.$$

Recall the definition of $\sigma(X)$ (Defintion 2.7). Let $\Omega = \{\omega : \omega = \omega_1\omega_2\omega_3\cdots\}$ and $A_i = \{\omega : \omega_1 = i\}$. It is not hard to see that $\sigma(X) = \{\emptyset, \Omega, A_i,\ A_i \cup A_j,\ A_i \cup A_j \cup A_k,\ A_i \cup A_j \cup A_k \cup A_l,\ A_i \cup A_j \cup A_k \cup A_l \cup A_m,\ i, j, k, l, m$ are not equal pairwisely, $i, j, k, l, m = 1, \cdots, 6\}$.

Since $f(X)$ is also a random variable defined on Ω, we can also define $\sigma(f(X))$. What is $\sigma(f(X))$?

(ii) In general, if X is a random variable, can the σ-algebra generated by $f(X)$ ever be strictly larger than the σ-algebra generated by X?

Solution: (i) $\sigma(f(X)) = \{\emptyset, \Omega, \{\omega : \omega_1 = 1, \text{or } 2, \text{or } 3\}, \{\omega : \omega_1 = 4, \text{or } 5, \text{or } 6\}\}$.

(ii) No. $\sigma(f(X))$ is always a subset of $\sigma(X)$.

15. Consider the symmetric random walk $M_1, M_1, M_2, \cdots$ defined in Example 2.11. Let $\sigma > 0$ be a constant.

a) Define $J_0 = 0$ and

$$J_n = \sum_{j=0}^{n-1} e^{\sigma M_j}(M_{j+1} - M_j), \quad n = 1, 2, \cdots. \tag{2.54}$$

Show that $J_0, J_1, \cdots, J_N$ is a martingale, which means, you need to prove that $\mathbb{E}_n[J_{n+1}] = J_n$.

b) Define

$$K_n = \sum_{j=0}^{n-1} M_{j+1}(M_{j+1} - M_j), \quad n = 1, 2, \cdots. \tag{2.55}$$

Show that

$$K_n = \frac{1}{2}M_n^2 + \frac{n}{2}.$$

3 Brownian motion (1 lecture)

We mainly follow Oksendal's book: Stochastic Differential Equations: an Introduction with Applications (6th edition) in this section. The computer code in the first subsection is from D. J. Higham, An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations, SIAM Review, 43 (2001) 525–546. See the appendix of the book of Choe for an introduction to Matlab.

3.1 Computer experiments

```
randn('state',100)           % set the state of randn
T = 1; N = 500; dt = T/N;
dW = zeros(1,N);             % preallocate arrays for efficiency
W = zeros(1,N);
dW(1) = sqrt(dt)*randn;      % prepare the iteration W(j) = W(j-1) + dW(j)
W(1) = dW(1);
for j = 2:N                  % start the iteration
    dW(j) = sqrt(dt)*randn; % general increment
    W(j) = W(j-1) + dW(j);
end
plot([0:dt:T],[0,W],'r-')    % plot W against t
xlabel('t','FontSize',16)
ylabel('W(t)','FontSize',16,'Rotation',0)
```

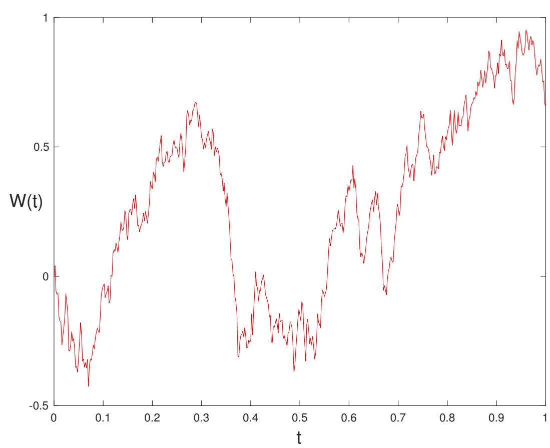


Figure 3.1: Discretized Brownian path

3.2 Probability space

Probability theory studies random experiments. The set of all possible outcomes of an experiment is called the sample space and is denoted by Ω. Any subset of Ω is called an event. The most basic operations on sets are union and intersection. Correspondingly, in practice, we need to study the intersection or union of different events. These needs motivate the introduction of σ-algebra (also called tribe) $\mathcal{F}$ that we have introduced before in Definition 2.5.

Definition 3.1 *The pair $(\Omega, \mathcal{F})$ we mentioned above is called a measurable space. A probability measure $\mathbb{P}$ on a measurable space $(\Omega, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ such that*

i) $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$.

ii) whenever a sequence of sets $A_1, A_2, \cdots$ belongs to $\mathcal{F}$ and $\{A_i\}_{i=1}^{\infty}$ is disjoint (i.e. $A_i \cap A_j = \emptyset$ if $i \neq j$), then

$$\mathbb{P}\left(\cup_{n=1}^{\infty} A_n\right) \overset{def}{=} \mathbb{P}\left(A_1 \cup A_2 \cup A_3 \cup \cdots\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space.

Remark: Note that a measure or a probability measure is a mapping that maps a set to a number. You can compare that with the definition of a scalar function which is a mapping that maps a number to a number. A scalar random variable X is a mapping that maps an element in Ω to a number.

For practical reason, we often need to study $\mathbb{P}(X \in [a, b])$ or $\mathbb{P}(X \in (a, b])$ or $\mathbb{P}(X \in [a, b))$. Let $I = [a, b]$ or $(a, b]$ or $[a, b)$. I is the so called Borel set. To study its probability, the event

$$\{\omega \in \Omega : X(\omega) \in I\} \overset{\text{def}}{=} X^{-1}(I)$$

must belongs to $\mathcal{F}$ because only set that belongs to $\mathcal{F}$ has a $\mathbb{P}$ value. By Definition 2.8, this means X is $\mathcal{F}$-measurable. So, finally, we **define** a scalar random variable Y as an $\mathcal{F}$-measurable function $Y : \Omega \to \mathbb{R}$.

If there is a function $\rho_X(x)$ so that

$$\mathbb{P}(X \in I) = \int_I \rho_X(x) dx \tag{3.1}$$

for every Borel set $I \subset \mathbb{R}$ (you can think of I as an interval), we say $\rho_X(x)$ is the probability density function (pdf) of X and X is then called a continuous random variable. For example, if

$$\mathbb{P}(X \leq a) = \int_{-\infty}^{a} \underbrace{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}_{\text{pdf } \rho_X(x)} dx = \underbrace{F_X(a)}_{cdf}, \tag{3.2}$$

then we say that X is normally distributed with mean μ and standard deviation σ which can be written as $X \sim N(\mu, \sigma^2)$.

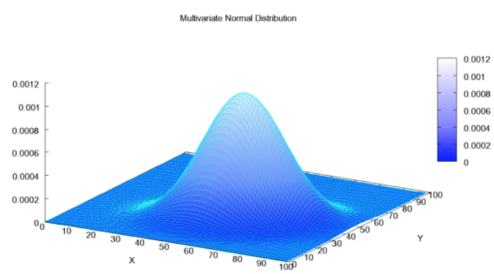
X can also be a vector in $\mathbb{R}^n$. For example, the prices of n different stocks form a n-dimensional vector. A random variable $X \in \mathbb{R}^n$ is called a continuous random variable, if we can find ρ_X, called the joint probability density function of X, such that

$$\mathbb{P}(X \in B) = \int_B \rho_X(x)dx \tag{3.3}$$

for any Borel set $B \subset \mathbb{R}^n$. You can safely think of B as a rectangular box. Note that $dx = dx_1 dx_2 \cdots dx_n$ and $\rho_X(x)$ maps $x \in \mathbb{R}^n$ to a value in $\mathbb{R}$. For example, if an n-dimensional vector X satisfies (3.3) with

$$\rho_X(x) = \frac{1}{(2\pi)^{n/2}|\det\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right), \tag{3.4}$$

then we say X is a n-dimensional normally distributed random variable, and write $X \sim N(\mu, \Sigma)$. Its components X_1, X_2, $\cdots$, X_n are called jointly normally distributed. In (3.4), $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$ are column vectors and $\Sigma \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix. Σ^{-1} is the inverse of Σ. The superscript $^\top$ denotes the transpose of a column vector or a matrix.



Let us briefly review some basic properties of normally distributed random variables. First, by the standard change of variable formula in multivariable integration, we can prove that if $X \sim N(\mu, \Sigma)$, then

$$Y = AX + \beta \sim N(A\mu + \beta, A\Sigma A^\top) \tag{3.5}$$

where A is an invertible $n \times n$ matrix and β is an n-dimensional column vector. In particular, Y, and also each component of it, are still normally distributed. Hence, by letting the first row of B be $(1, 1, ..., 1)$, we see that the summation of jointly normally distributed random variables are still normally distributed.

Using (3.5), one can manage to prove that

$$\mathbb{E}[(X-\mu)(X-\mu)^\top] = \Sigma, \quad i.e., \quad \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] = \Sigma_{ij}, \tag{3.6}$$

where Σ_{ij} is the (i, j) entry of Σ. Hence Σ is the covariance matrix. The leads to a nice property about multivariate normally distributed random variable: once we know its mean and its variance, we know its pdf and hence we know everything about it.

One can prove that for discrete random variable Y, the expectation of $g(Y)$ (defined in Definition 2.19) can be computed in another two ways

$$\mathbb{E}[g(Y)] \overset{\text{def}}{=} \sum_{\omega\in\Omega} g(Y(\omega))\mathbb{P}(\omega) \overset{(a)}{=} \sum_{y_i} g(y_i)\mathbb{P}(Y=y_i) \overset{(b)}{=} \sum_{g_j} g_j\mathbb{P}(g(Y)=g_j), \tag{3.7}$$

with ω running through all possible outcomes in Ω, y_i running through all possible values of Y, g_j running through all possible values of g. The equalities are obviously since they are just three different ways to count when taking a weighted average of all the values $g(Y(\omega))$: (def) for different ω, (a) for different $Y(\omega)$, (b) for different $g(Y(\omega))$. (See for example, Ross, First course in Probability, 9th edition, Pages 167, 133.)

Example 3.1 *Suppose that two independent flips of a coin that comes up heads with probability p are made, and let Y denote the number of heads obtained. Let $g(y)=y^2$. We can compute $\mathbb{E}[g(Y)]$ in three different ways.*

$$\begin{aligned}
&\sum_{\omega\in\Omega} g(Y(\omega))\mathbb{P}(\omega)\\
=&(Y(HH))^2\mathbb{P}(HH)+(Y(HT))^2\mathbb{P}(HT)+(Y(TH))^2\mathbb{P}(TH)+(Y(TT))^2\mathbb{P}(TT)\\
=&2^2p^2+1^2p(1-p)+1^2(1-p)p+0(1-p)^2=2p^2+2p.\\
&\sum_{y_i\in\{0,1,2\}} g(y_i)\mathbb{P}(Y=y_i)\\
=&0^2\mathbb{P}(Y=0)+1^2\mathbb{P}(Y=1)+2^2\mathbb{P}(Y=2)\\
=&0(1-p)^2+2p(1-p)+4p^2=2p^2+2p.\\
&\sum_{g_j\in\{0,1,4\}} g_j\mathbb{P}(g(Y)=g_j)\\
=&0\mathbb{P}(g(Y)=0)+1\mathbb{P}(g(Y)=1)+4\mathbb{P}(g(Y)=4)\\
=&0(1-p)^2+2p(1-p)+4p^2=2p^2+2p.
\end{aligned}$$

Normally, in an undergraduate textbook on probability, for continuous random variable Y with pdf $\rho_Y(y)$, the expectation $\mathbb{E}[g(Y)]$ can be defined as (See for example, Ross, First course in Probability, 9th edition, Pages 196)

$$\mathbb{E}[g(Y)] = \int_{\mathbb{R}} g(y)\rho_Y(y)dy. \tag{3.8}$$

As the right hand side is the limit of the Riemann sum $\sum_{i=1}^N g(y_i)\underbrace{\rho_Y(y_i)(y_{i+1}-y_i)}_{\approx\mathbb{P}(y_i\le Y<y_{i+1})}$, (3.8) is a generalization of $\mathbb{E}[g(Y)]=\sum_{y_i} g(y_i)\mathbb{P}(Y=y_i)$ in (3.7). To use (3.8), one needs to know $\rho_Y(y)$. You may have wondered whether we can generalize $\mathbb{E}[g(Y)]=\sum_{\omega\in\Omega} g(Y(\omega))\mathbb{P}(\omega)$ in (3.7) to the continuous random variable case so that we do not need to know precisely the pdf of Y when computing $\mathbb{E}[g(Y)]$. The answer is the following definition

Definition 3.2 *If $\int_\Omega |g(Y(\omega))| d\mathbb{P}(\Omega) < \infty$, then the number*

$$\mathbb{E}[g(Y)] = \int_\Omega g(Y(\omega)) d\mathbb{P}(\omega) \tag{3.9}$$

is called the expectation of $g(Y)$ w.r.t. $\mathbb{P}$. See Section 1.1.3 of Shreve II for the precise definition of $\int_\Omega g(Y(\omega)) d\mathbb{P}(\omega)$.

Example: Let $\Omega = \{\text{any numbers between 0 and 1}\} = [0,1]$ and let $X(\omega) = \sin(\omega)$ with $\omega \in [0,1]$. X is a random variable. The randomness lies in what is the probability that you can choose a particular ω. $\mathcal{F}$ is now the Borel set (generated by the σ-algebra of all closed or open intervals). You can think of $\mathcal{F}$ as collection of all intervals (open, closed, half open half closed intervals). Suppose we define $\mathbb{P}(\{\omega \in (a,b)\}) = \mathbb{P}(\{\omega \in [a,b]\}) = b - a$ for $a, b \in [0,1]$, which means each ω is equally likely to be selected, then $d\mathbb{P}(\omega)$ is precisely $d\omega$.

$$\mathbb{E}[X] = \int_\Omega X(\omega) d\mathbb{P}(\omega) = \int_0^1 \sin(\omega) d\omega = 1 - \cos(1).$$

Example: Let $\Omega = \{\text{all continuous functions defined on [0,1] whose value at 0 is } x_0\} = C^0([0,1])$. We can fix $t \in [0,1]$ and define $X_t(\omega) = \omega(t)$ with $\omega \in C^0([0,1])$. X_t maps any $\omega \in C^0([0,1])$ to a number $\omega(t)$ and is a random variable. The randomness lies in what is the probability that you can choose a particular ω.

You can define $I^{t_1,t_2,\cdots,t_n}_{\Gamma_1 \times \Gamma_2 \cdots \times \Gamma_n} = \{\omega : \omega \in C^0([0,1]), \omega(t_1) \in \Gamma_1, \omega(t_2) \in \Gamma_2, \cdots, \omega(t_n) \in \Gamma_n\}$ and define

$$\mathbb{P}(\omega \in I^{t_1,t_2,\cdots,t_n}_{\Gamma_1 \times \Gamma_2 \cdots \times \Gamma_n}) = \int_{\Gamma_1 \times \cdots \times \Gamma_n} p(t_1, x_0, x_1) p(t_2 - t_1, x_1, x_2) \cdots p(t_n - t_{n-1}, x_{n-1}, x_n) dx_1 \cdots dx_n$$

with $p(t,x,y) = (2\pi t)^{-1/2} e^{-\frac{|y-x|^2}{2t}}$. It turns out we can extend this $\mathbb{P}$ to the σ-algebra generated by $I^{t_1,t_2,\cdots,t_n}_{\Gamma_1 \times \Gamma_2 \cdots \times \Gamma_n}$'s (It is then called the Wiener measure). Then we can talk about $d\mathbb{P}(\omega)$, but this time it is not the integration you have learned in Calculus before. However, the $X_t(\omega) = \omega(t)$ is precisely the Brownian motion that we will introduce later.

One can prove that (Theorem 1.5.2 of Shreve II)

$$\mathbb{E}[g(Y)] = \underbrace{\int_\Omega g(Y(\omega)) d\mathbb{P}(\omega)}_{\text{integrate/sum on } \omega} = \underbrace{\int_{\mathbb{R}^n} g(y) \rho_Y(y) dy}_{\text{integrate/sum on } y}. \tag{3.10}$$

From (3.9), one immediately obtains

$$\mathbb{E}[f(X) + g(Y)] = \mathbb{E}[f(X)] + \mathbb{E}[g(Y)] \tag{3.11}$$

for any random variables X and Y.

Example 3.2 *Later on, we will learn Brownian motion* $W_t \sim N(x_0, t)$*,*

$$\mathbb{E}[g(W_t)] = \int_\Omega g(W_t(\omega))d\mathbb{P}(\omega) = \int_{\mathbb{R}} g(y)\rho_{W_t}(y)dy = \int_{\mathbb{R}} g(y)dF_{W_t}(y),$$

where $\rho_{W_t}(y) = \frac{1}{\sqrt{2\pi t}}e^{-(y-x_0)^2/(2t)}$, $F_{W_t}(y) = \mathbb{P}(W_t \le y) = \int_{-\infty}^y \rho_{W_t}(s)ds$, $dF_{W_t}(y) = \frac{dF_{W_t}(y)}{dy}dy = \rho_{W_t}(y)dy$*. (see (3.2)).*

Example 3.3 *Let* r *and* σ *be constants. Define*

$$S(t) = S_0 e^{(r-\frac{1}{2}\sigma^2)t+\sigma W(t)} \tag{3.12}$$

with $W(t) \sim N(0,t)$*. Let* K *be a positive constant. Show that*

$$\mathbb{E}\left[e^{-rT}(S(T)-K)^+\right] = S_0 N(d_+) - Ke^{-rT}N(d_-) \tag{3.13}$$

with

$$d_\pm = \frac{1}{\sigma\sqrt{T}}\left[\log\frac{S_0}{K} + \left(r \pm \frac{\sigma^2}{2}\right)T\right], \tag{3.14}$$

$$N(y) = \frac{1}{\sqrt{2\pi}}\int_{-y}^{\infty} e^{-\frac{1}{2}z^2}dz = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{1}{2}z^2}dz. \tag{3.15}$$

The above formula has appeared before in (2.11).

Proof*:* $W(T) \sim N(0,T)$ *and its pdf is* $\frac{e^{-\frac{x^2}{2T}}}{\sqrt{2\pi T}}$*. By (3.10),*

$$\begin{aligned}
&\mathbb{E}\left[e^{-rT}(S_0 e^{(r-\frac{1}{2}\sigma^2)T+\sigma W(T)} - K)^+\right] \\
=&e^{-rT}\int_{-\infty}^{\infty}\left(S_0 e^{(r-\frac{1}{2}\sigma^2)T+\sigma x} - K\right)^+ \frac{e^{-\frac{x^2}{2T}}}{\sqrt{2\pi T}}dx.
\end{aligned}$$

The integrand is zero if $S_0 e^{(r-\frac{1}{2}\sigma^2)T+\sigma x} - K \le 0$*.* $S_0 e^{(r-\frac{1}{2}\sigma^2)T+\sigma x} - K > 0$ *if and only if* $x > \frac{1}{\sigma}\left(\log\frac{K}{S_0} - (r-\frac{1}{2}\sigma^2)T\right) = \frac{1}{\sigma}\left(-\log\frac{S_0}{K} - (r-\frac{1}{2}\sigma^2)T\right) = -\sqrt{T}d_-$*. Hence*

$$\begin{aligned}
\mathbb{E}\left[e^{-rT}(S(T)-K)^+\right] =& e^{-rT}\int_{-\sqrt{T}d_-}^{\infty}\left(S_0 e^{(r-\frac{1}{2}\sigma^2)T+\sigma x} - K\right)\frac{e^{-\frac{x^2}{2T}}}{\sqrt{2\pi T}}dx \\
\overset{y=x/\sqrt{T}}{=}& e^{-rT}\int_{-d_-}^{\infty}\left(S_0 e^{(r-\frac{1}{2}\sigma^2)T+\sigma\sqrt{T}y} - K\right)\frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}dy.
\end{aligned}$$

In the last step, we have used the change of variable $y = x/\sqrt{T}$*. After removing the bracket, we get two terms*

$$\begin{aligned}
&\mathbb{E}\left[e^{-rT}(S(T)-K)^+\right] \\
=&S_0\int_{-d_-}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\sigma^2 T+\sigma\sqrt{T}y-\frac{1}{2}y^2}dy - Ke^{-rT}\int_{-d_-}^{\infty}\frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}dy \\
=&S_0\int_{-d_-}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(y-\sigma\sqrt{T})^2}dy - Ke^{-rT}N\left(d_-\right).
\end{aligned}$$

In the last step, we have used the function N introduced in (3.15). To handle the first term, we do a change of variable $z = y - \sigma\sqrt{T}$ and notice $-d_- - \sigma\sqrt{T} = -d_+$:

$$\begin{aligned}\mathbb{E}\left[e^{-rT}(S(T)-K)^+\right] =& S_0 \int_{-d_- - \sigma\sqrt{T}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz - Ke^{-rT}N(d_-) \\ =& S_0 N(d_+) - Ke^{-rT}N(d_-). \qquad \square\end{aligned}$$

Remark: By the same calculation as above[21], one can show that if $S = e^X$ with $X \sim N(\mu, \gamma^2)$, then

$$\mathbb{E}[(S-K)^+] = e^{\mu + \frac{1}{2}\gamma^2} N(d_+) - KN(d_-) \tag{3.16}$$

with $d_- = \frac{\log \frac{1}{K} + \mu}{\gamma}$, $d_+ = d_- + \gamma$.

We say that two random variable X and Y are independent if for any intervals A and B on $\mathbb{R}$,

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B). \tag{3.17}$$

If a continuous 2-dimensional random variable (X, Y) has joint pdf $\rho_{X,Y}(x, y)$, one can introduce the marginal pdf

$$\rho_X(x) = \int_{-\infty}^{\infty} \rho_{X,Y}(x,y)dy \quad \text{and} \quad \rho_Y(y) = \int_{-\infty}^{\infty} \rho_{X,Y}(x,y)dx. \tag{3.18}$$

Then one can manage to show that X and Y are independent if and only if

$$\rho_{X,Y}(x,y) = \rho_X(x)\rho_Y(y). \tag{3.19}$$

Indeed, as long as $\rho_{X,Y}(x, y)$ is separable, namely, there are functions $f(x)$ and $g(y)$ so that

$$\rho_{X,Y}(x,y) = f(x)g(y), \tag{3.20}$$

then X and Y are independent. Moreover, there is a constant C so that $f(x) = C\rho_X(x)$ and $g(y) = \frac{1}{C}\rho_Y(y)$.

If X and Y are independent, for any function f and g

$$\begin{aligned}\mathbb{E}[f(X)g(Y)] &= \int_{\mathbb{R}^2} f(x)g(y)\rho_{X,Y}(x,y)dxdy \\ &= \int_{\mathbb{R}} f(x)\rho_X(x)dx \times \int_{\mathbb{R}} g(y)\rho_Y(y)dy = \mathbb{E}[f(X)]\mathbb{E}[g(Y)].\end{aligned} \tag{3.21}$$

Criteria (3.20) for independence can be generalized to n-dimensional random variables. For example, if $(X_1, \cdots, X_n) \stackrel{\text{def}}{=} X \sim N(\mu, \Sigma)$ and Σ is a diagonal matrix, i.e, $\Sigma =$

[21] Set $S_0 = 1$, $\sigma\sqrt{T} = \gamma$, $rT - \frac{\sigma^2 T}{2} = \mu$. Then $rT + \frac{\sigma^2 T}{2} = \mu + \gamma^2$, $rT = \mu + \frac{1}{2}\gamma^2$.

$\mathrm{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2)$, it is easy to see that $\det\Sigma = \Pi_{i=1}^n \sigma_i^2$, $\Sigma^{-1} = \mathrm{diag}(\sigma_1^{-2}, \sigma_2^{-2}, \cdots, \sigma_n^{-2})$. Then

$$\begin{aligned} \rho_X(x_1, \cdots, x_n) =& \frac{1}{(2\pi)^{n/2}|\det\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right) \\ =& \Pi_{i=1}^n \rho_{X_i}(x_i), \end{aligned} \tag{3.22}$$

with $\rho_{X_i} = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-(x_i-\mu_i)^2/(2\sigma_i^2)}$. We then conclude that X_i's are independent of each other.

Definition 3.3 *(Definition 2.1.4 of Oksendal) A stochastic process is a parameterized collection of random variables*

$$\{X_t\}_{t\in\mathcal{T}}$$

defined on a prabability space $(\Omega, \mathcal{F}, \mathbb{P})$ and assuming values in $\mathbb{R}^n$. [22]

The parameter space $\mathcal{T}$ is usually the half space $[0, \infty)$ or an interval $[0, T]$. Note that for each $t \in \mathcal{T}$ fixed, we have a random variable

$$\omega \to X_t(\omega), \qquad \omega \in \Omega.$$

On the other hand, fixing $\omega \in \Omega$, we can consider the function

$$t \to X_t(\omega), \qquad t \in \mathcal{T},$$

which is called a path of X_t.

It may be useful for the intuition to think of t as "time" and each ω as an individual "particle" [23] or "the result of the experiment which randomly selects a particle". With this picture, $X_t(\omega)$ would represent the position (or result) at time t of the particle (experiment) ω.

Given an m-dimensional stochastic process $\{X_t\}_{t\in T}$, we can introduce the finite-dimensional distribution of $\{X_t\}_{t\in T}$ as the collection of the measures $\mu_{t_1,\cdots,t_n}$ defined on $\underbrace{\mathbb{R}^m \times \mathbb{R}^m \cdots \times \mathbb{R}^m}_{n} = \mathbb{R}^{mn}$, $n = 1, 2, \cdots$, such that

$$\mu_{t_1,\cdots,t_n}(\Gamma_1 \times \Gamma_2 \cdots \times \Gamma_n) = \mathbb{P}\left(X_{t_1} \in \Gamma_1, \cdots, X_{t_n} \in \Gamma_n\right), \qquad t_i \in \mathcal{T}. \tag{3.23}$$

Here $\Gamma_1, \cdots, \Gamma_n$ denote Borel sets (think of them as intervals when $m = 1$ or boxes when $m > 1$) in $\mathbb{R}^m$.

The family of all finite-dimensional distributions determine many (but not all) important properties of the process $\{X_t\}$. On the other hand, given a family $\{\nu_{t_1,\cdots,t_n} : n \in \mathbb{Z}_+, t_i \in \mathcal{T}\}$ of probability measures on $\mathbb{R}^{mn}$, can we construct a stochastic process $\{Y_t\}_{t\in\mathcal{T}}$ having $\{\nu_{t_1,\cdots,t_n} : n \in \mathbb{Z}_+, t_i \in \mathcal{T}\}$ as its finite-dimensional distribution? Here $\mathbb{Z}_+ = \{1, 2, 3, 4, \cdots\}$. One of Kolmogorov's famous theorems states that this can be done provided $\{\nu_{t_1,\cdots,t_n} : n \in \mathbb{Z}_+, t_i \in \mathcal{T}\}$ satisfies two natural consistency conditions:

[22] In this lecture notes, X_t and $X(t)$ denote the same thing. $X_t(\omega)$ can also be written as $X(t, \omega)$.

[23] think about Brownian motion as random motion of particles suspended in a fluid.

Theorem 3.1 *(Kolmogorov's extension theorem. Theorem 2.1.5 of Oksendal's book) Suppose we have a collection of probability measures $\{\nu_{t_1,\cdots,t_n}\}$ that satisfies for any $n \in \mathbb{Z}_+$, for any $t_1, \cdots, t_n \in \mathcal{T}$, $\nu_{t_1,\cdots,t_n}$ is a probability measure on $\mathbb{R}^{mn}$ such that*

$$\nu_{t_{\sigma(1)},\cdots,t_{\sigma(n)}}(F_1 \times \cdots \times F_n) = \nu_{t_1,\cdots,t_n}(F_{\sigma^{-1}(1)} \times \cdots \times F_{\sigma^{-1}(n)}) \tag{3.24}$$

for all permutation σ on $\{1, 2, \cdots, n\}$ [24] *and*

$$\nu_{t_1,\cdots,t_n}(F_1 \times \cdots \times F_n) = \nu_{t_1,\cdots,t_n,t_{n+1},\cdots,t_{n+k}}(F_1 \times \cdots \times F_n \times \underbrace{\mathbb{R}^m \times \cdots \times \mathbb{R}^m}_{k}) \tag{3.25}$$

for all $k \in \mathbb{Z}_+$. Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an m-dimensional (meaning $X_t \in \mathbb{R}^m$) stochastic process $\{X_t\}$ on Ω such that

$$\nu_{t_1,\cdots,t_n}(F_1 \times F_2 \cdots \times F_n) = \mathbb{P}\left(X_{t_1} \in F_1, \cdots, X_{t_n} \in F_n\right), \tag{3.26}$$

for all $t_i \in \mathcal{T}$, $n \in \mathbb{Z}_+$ and all Borel sets $F_i \subset \mathbb{R}^m$.

3.3 Brownian motion in $\mathbb{R}^m$

In 1828 the Scottish botanist Robert Brown observed the pollen grains suspended in liquid performed an irregular motion. We will introduce a stochastic process $\{W_t(\omega)\}$, interpreted as the position at time t of the pollen grain ω. Since all our later studies are based on this $\{W_t\}$ which has some nice but sometimes quite weird properties, we'd better confirm that such a nice (or weird) thing does exist.

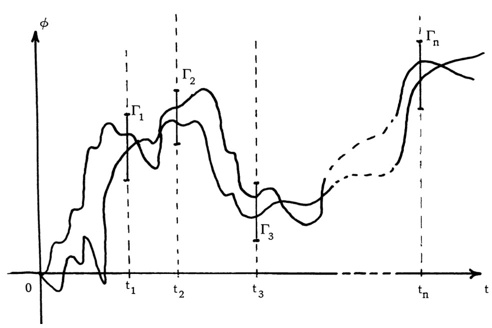


Figure 3.2: Picture taken from "Functional integration and partial differential equations" by Mark Freidlin. $x_0 = 0$. $m = 1$.

To construct $\{W_t\}_{t\geq 0}$, it suffices, by the Kolmogorov's extension theorem (Theorem 3.1), to specify a family $\nu_{t_1,\cdots,t_n}$ of probability measures satisfying (3.24) and (3.25). Let $W =$

[24] Note that we do not require $t_1 < t_2 < t_3 < \cdots$. (3.24) is a natural requirement. For example, $n = 3$, $\sigma : (1, 2, 3) \mapsto (2, 3, 1)$, which means $\sigma(1) = 2$, $\sigma(2) = 3$, $\sigma(3) = 1$. $\nu_{t_{\sigma(1)},t_{\sigma(2)},t_{\sigma(3)}}(F_1, F_2, F_3) \overset{\text{def of } \sigma(i)}{=} \nu_{t_2,t_3,t_1}(F_1, F_2, F_3) \overset{\text{since we want (3.26)}}{=} \mathbb{P}(X_{t_2} \in F_1, X_{t_3} \in F_2, X_{t_1} \in F_3) = \mathbb{P}(X_{t_1} \in F_3, X_{t_2} \in F_1, X_{t_3} \in F_2) = \nu_{t_1,t_2,t_3}(F_3, F_1, F_2) = \nu_{t_1,t_2,t_3}(F_{\sigma^{-1}(1)}, F_{\sigma^{-1}(2)}, F_{\sigma^{-1}(3)},).$

$(W_{t_1}, W_{t_2}, \cdots, W_{t_n})$. If we are able to do so, we obtain (3.26), which says

$$\nu_{t_1,\cdots,t_n}(F_1 \times F_2 \cdots \times F_n) = \mathbb{P}(W_{t_1} \in F_1, \cdots, W_{t_n} \in F_n)$$
$$= \int_{F_1 \times F_2 \times \cdots \times F_n} \rho_W(w_1, w_2, \cdots, w_n) dw_1 dw_2 \cdots dw_n.$$

In the last step, we have used (3.3). So, when we specify $\{\nu_{t_1,\cdots,t_n}\}$, what we are really doing is indeed specifing ρ_W, the joint pdf of W.

What kind of joint pdf of W do we want to obtain? We take $m = 1$ and $k = 2$ as an example.

(1) We want W_{t_1} to be normally distributed with mean x_0 (starting point) and variance t_1. So the pdf of W_{t_1} is $\rho_{W_{t_1}}(x_1) = \frac{1}{\sqrt{2\pi t_1}} e^{-\frac{(x_1-x_0)^2}{2t_1}}$.

(2) Once we know $W_{t_1} = x_1$, we hope W_{t_2} is normally distributed around x_1, with variance $= t_2 - t_1$.

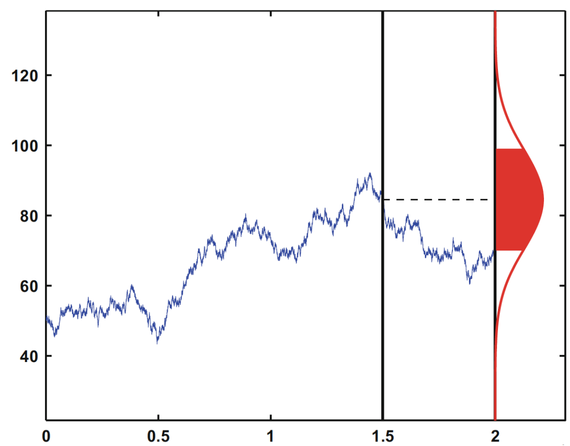


Figure 3.3: $t_1 = 1.5$. $t_2 = 2$. Given the value of W_{t_1}, W_{t_2} is normally distributed around the value of W_{t_1}.

Hence the pdf of W_{t_2}, given $W_{t_1} = x_1$, is $\rho_{W_{t_2}|W_{t_1}}(x_2|x_1) = \frac{1}{\sqrt{2\pi(t_2-t_1)}} e^{-\frac{(x_2-x_1)^2}{2(t_2-t_1)}}$. See Question 8 of Homework III for the notation.

Combining (1) and (2) and using the standard formula for conditional pdf

$$\rho_{W_{t_2}|W_{t_1}}(x_2|x_1) = \frac{\rho_{W_{t_1},W_{t_2}}(x_1, x_2)}{\rho_{W_{t_1}}(x_1)}$$

(see (3.52) in Question 8 of Homework III or Page 270 of A First Course in Probability by Ross), we get

$$\rho_{W_{t_1},W_{t_2}}(x_1, x_2) = \frac{1}{\sqrt{2\pi t_1}} e^{-\frac{(x_1-x_0)^2}{2t_1}} \frac{1}{\sqrt{2\pi(t_2 - t_1)}} e^{-\frac{(x_2-x_1)^2}{2(t_2-t_1)}}. \tag{3.27}$$

So, if we introduce (now, consider general $m \in \mathbb{Z}_+$ and I_m is $m \times m$ identity matrix)

$$p(t,x,y) = (2\pi t)^{-m/2} e^{-\frac{|y-x|^2}{2t}} \sim \text{pdf of } N(x, tI_m), \qquad \text{for } x, y \in \mathbb{R}^m, \quad t > 0, \tag{3.28}$$

we see that $\rho_{W_{t_1}, W_{t_2}}(x_1, x_2) = p(t_1, x_0, x_1) p(t_2 - t_1, x_1, x_2)$.

Now, we are ready to propose the $\nu_{t_1, \cdots, t_n}$ that we want: If $0 \le t_1 \le t_2 \le \cdots \le t_n$, define a measure $\nu_{t_1, \cdots, t_n}$ on $\mathbb{R}^{mn}$ by (Here Γ_i is any Borel set or m-dimensional box in $\mathbb{R}^m$. $x_0 \in \mathbb{R}^m$ is the starting point. $x_1, \cdots, x_n \in \mathbb{R}^m$ are the integrating/dummy variables.)

$$\begin{aligned} &\nu_{t_1, \cdots, t_n}(\Gamma_1 \times \cdots \times \Gamma_n) \\ &= \int_{\Gamma_1 \times \cdots \times \Gamma_n} p(t_1, x_0, x_1) p(t_2 - t_1, x_1, x_2) \cdots p(t_n - t_{n-1}, x_{n-1}, x_n) dx_1 \cdots dx_n. \end{aligned} \tag{3.29}$$

Extend this definition to all finite sequences of t_i's by using (3.24). Since $\int_{\mathbb{R}^n} p(t,x,y)dy = 1$ for all $t > 0$, (3.25) holds, so by Theorem 3.1 there exist a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a stochastic process $\{W_t\}_{t \ge 0}$ on Ω such that the finite-dimensional distributions of W_t are given by (3.29), i.e.,

$$\mathbb{P}(W_{t_1} \in \Gamma_1, \cdots, W_{t_n} \in \Gamma_n) = \int_{\Gamma_1 \times \cdots \times \Gamma_n} p(t_1, x_0, x_1) \cdots p(t_n - t_{n-1}, x_{n-1}, x_n) dx_1 \cdots dx_n. \tag{3.30}$$

Definition 3.4 *(Definition 2.2.1 of Oksendal) W_t is called Brownian motion starting at x_0.*

So, $p(t_1, x_0, x_1) \cdots p(t_n - t_{n-1}, x_{n-1}, x_n)$, is the joint pdf of the n-dimensional random variable $X = (W_{t_1}, \cdots, W_{t_n})$. If we introduce $t_0 = 0$ and let $\delta t_i = t_i - t_{i-1}$, then

$$\rho_X(x_1, x_2, \cdots, x_n) = (2^n \pi^n \delta t_1 \cdots \delta t_n)^{-m/2} e^{-\sum_{i=1}^n \frac{1}{2\delta t_i} |x_i - x_{i-1}|^2}. \tag{3.31}$$

Let's summarize it as a theorem,

Theorem 3.2 *If $t_0 = 0$, given the initial position $W_{t_0} = x_0 \in \mathbb{R}^m$, the joint probability density function of $(W_{t_1}, \cdots, W_{t_n})$ is*

$$\frac{1}{(2\pi)^{mn/2}[(t_1 - t_0)(t_2 - t_1) \cdots (t_n - t_{n-1})]^{m/2}} \exp\left(-\frac{|x_1 - x_0|^2}{2(t_1 - t_0)} - \frac{|x_2 - x_1|^2}{2(t_2 - t_1)} - \cdots \frac{|x_n - x_{n-1}|^2}{2(t_n - t_{n-1})} \right). \tag{3.32}$$

So, the joint pdf indeed takes a rather simple form. In particular, it can be written as the product of pdfs of the increments. This reminds us of the criteria (3.20) for independence and motivates the following change of variable from X to Y:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix} W_{t_1} \\ W_{t_2} - W_{t_1} \\ W_{t_3} - W_{t_2} \\ \vdots \\ W_{t_n} - W_{t_{n-1}} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} W_{t_1} \\ W_{t_2} \\ W_{t_3} \\ \vdots \\ W_{t_n} \end{pmatrix} \stackrel{\text{def}}{=} AX \tag{3.33}$$

In fact, using the change of variable formula in multivariate calculus/probability [25], we see that

$$\begin{aligned}\rho_Y(y_1, y_2, \cdots, y_n) =& (2^n \pi^n \delta t_1 \cdots \delta t_n)^{-m/2} e^{-\sum_{i=1}^n \frac{1}{2\delta t_i}|x_i - x_{i-1}|^2} \frac{1}{|\det A|} \\ =& (2^n \pi^n \delta t_1 \cdots \delta t_n)^{-m/2} e^{-\sum_{i=1}^n \frac{1}{2\delta t_i}|y_i|^2}.\end{aligned}$$

We have used the fact that $\det A = 1$. $(2\pi\delta t_i)^{-m/2} e^{-\frac{1}{2\delta t_i}|y_i|^2}$ is the pdf of a $N(0, \delta t_i)$ random variable. ρ_Y is the production of them. By (3.3), we recognize that

Theorem 3.3 *The increments, $W_{t_i} - W_{t_{i-1}}$ and $W_{t_j} - W_{t_{j-1}}$, are independent as long as $i \neq j$. $W_{t_i} - W_{t_{i-1}} \sim N(0, t_i - t_{i-1})$.*

The proof that we have just presented is FYI only, and won't be tested. I just want to show you how the independence of the increments is hidden in our construction at the beginning. You should be able to apply the theorem, for example, to perform the calculation in (3.36) later on.

From the increments, we can recover W_t by the telescoping sum

$$W_t = (W_t - W_{t_{n-1}}) + (W_{t_{n-1}} - W_{t_{n-2}}) + (W_{t_{n-2}} - W_{t_{n-3}}) + \cdots + (W_{t_1} - W_{t_0}) + x_0 \quad (3.35)$$

with $W_{t_0} = x_0$ being the initial position. This is precisely the inverse of the mapping (3.33) with

$$A^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix}.$$

(See also Question 16 of Homework III for another interesting construction.)

[25](See for example Page 283 of A First Course in Probability by Ross) If $(X_1, X_2, \ldots, X_n)$ is an n-dimensional continuous random variable with probability density function $\rho_{X_1, X_2, \ldots, X_n}$, and

$$Y_i = g_i(X_1, X_2, \ldots, X_n), \quad 1 \le i \le n$$

$$J(x_1, x_2, \ldots, x_n) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \frac{\partial g_n}{\partial x_2} & \cdots & \frac{\partial g_n}{\partial x_n} \end{vmatrix} \neq 0,$$

then

$$\rho_{Y_1, Y_2, \ldots, Y_n}(y_1, y_2, \ldots, y_n) = \rho_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) \left|J(x_1, x_2, \ldots, x_n)\right|^{-1}. \quad (3.34)$$

Remark: In particular, by (3.5), the increments are normal implies $W_{t_1}, W_{t_2}, \cdots, W_{t_n}$ are normal, even though the later can also be derived directly from (3.32). To see how things are connected, let us perform some direct calculation to verify it:

Example 3.4 *(The matrix-vector multiplication way to write the joint pdf) Let $t_2 > t_1$. If (W_{t_1}, W_{t_2}) is multivariate normal distribution with pdf*

$$f(x_1, x_2) = \frac{1}{(2\pi)(det\ \Sigma)^{1/2}} e^{-\frac{1}{2}(x_1-x_0, x_2-x_0)\Sigma^{-1}(x_1-x_0,x_2-x_0)^\top},$$

where $\Sigma = \begin{pmatrix} t_1 & t_1 \\ t_1 & t_2 \end{pmatrix}$, *then* $\Sigma^{-1} = \frac{1}{t_1 t_2 - t_1^2}\begin{pmatrix} t_2 & -t_1 \\ -t_1 & t_1 \end{pmatrix}$ *and*

$$f(x_1, x_2) = \frac{1}{(2\pi)(t_1 t_2 - t_1^2)^{1/2}} e^{-\frac{1}{2}\frac{(x_1-x_0,x_2-x_0)\begin{pmatrix} t_2 & -t_1 \\ -t_1 & t_1 \end{pmatrix}\begin{pmatrix} x_1 - x_0 \\ x_2 - x_0 \end{pmatrix}}{t_1(t_2-t_1)}}.$$

After the matrix vector multiplication, we obtain

$$\begin{aligned} f(x_1, x_2) =& \frac{1}{(2\pi t_1)^{1/2}} \frac{1}{(2\pi(t_2 - t_1))^{1/2}} e^{-\frac{t_2(x_1-x_0)^2 + t_1(x_2-x_0)^2 - 2t_1(x_1-x_0)(x_2-x_0)}{2t_1(t_2-t_1)}} \\ =& \frac{1}{(2\pi t_1)^{1/2}} \frac{1}{(2\pi(t_2 - t_1))^{1/2}} e^{-\frac{(t_2-t_1)(x_1-x_0)^2 + t_1(x_2-x_1)^2}{2t_1(t_2-t_1)}} \\ =& \frac{1}{(2\pi t_1)^{1/2}} e^{-\frac{(x_1-x_0)^2}{2t_1}} \frac{1}{(2\pi(t_2 - t_1))^{1/2}} e^{-\frac{(x_2-x_1)^2}{2(t_2-t_1)}}. \end{aligned}$$

Hence

$$\mathbb{P}(W_{t_1} \in \Gamma_1, W_{t_2} \in \Gamma_2) = \mathbb{P}((W_{t_1}, W_{t_2}) \in \Gamma_1 \times \Gamma_2) = \int_{\Gamma_1 \times \Gamma_2} f(x_1, x_2) dx_1 dx_2$$
$$= \int_{\Gamma_1 \times \Gamma_2} \frac{1}{(2\pi t_1)^{1/2}} e^{-\frac{(x_1-x_0)^2}{2(t_1-t_0)}} \frac{1}{(2\pi(t_2 - t_1))^{1/2}} e^{-\frac{(x_2-x_1)^2}{2(t_2-t_1)}} dx_1 dx_2.$$

The right hand side is exactly (3.29) with the integrand being (3.31).

Remark: By (3.6),

$$\begin{aligned} \begin{pmatrix} t_1 & t_1 \\ t_1 & t_2 \end{pmatrix} = \Sigma = \mathbb{E}\left[\begin{pmatrix} W_{t_1} - x_0 \\ W_{t_2} - x_0 \end{pmatrix} \begin{pmatrix} W_{t_1} - x_0 & W_{t_2} - x_0 \end{pmatrix}\right] \\ = \mathbb{E}\left[\begin{pmatrix} (W_{t_1} - x_0)^2 & (W_{t_1} - x_0)(W_{t_2} - x_0) \\ (W_{t_1} - x_0)(W_{t_2} - x_0) & (W_{t_2} - x_0)^2 \end{pmatrix}\right]. \end{aligned}$$

We hence obtain $\mathbb{E}[(W_{t_i} - x_0)^2] = t_i$ (for $i = 1, 2$) and $\mathbb{E}[(W_{t_1} - x_0)(W_{t_2} - x_0)] = t_1 = \min(t_2, t_1)$ as by-products of the above calculation.

To see how things are connected, let us also use the independence and mean zero properties of the increments to compute:

$$
\begin{aligned}
&\mathbb{E}[(W_{t_1}-x_0)(W_{t_2}-x_0)] \\
=&\mathbb{E}[(W_{t_1}-x_0)(W_{t_2}-W_{t_1}+W_{t_1}-x_0)] \\
=&\mathbb{E}[(W_{t_1}-x_0)(W_{t_1}-x_0)]+\mathbb{E}[(W_{t_1}-W_{t_0})(W_{t_2}-W_{t_1})] \\
\overset{(3.21)}{=}&\mathrm{Var}[W_{t_1}]+\mathbb{E}[(W_{t_1}-W_{t_0})]\mathbb{E}[(W_{t_2}-W_{t_1})]=t_1+0=t_1.
\end{aligned} \tag{3.36}
$$

Then using (3.6) and the fact that $(W_{t_1}, W_{t_2}, \cdots, W_{t_n})$ is multi-dimensional normal (Remark after Theorem 3.3), we obtain $\Sigma_{1,2} = \Sigma_{2,1} = t_1$.

Indeed, by simple calculations like in (3.36), we can obtain

$$
\mathbb{E}[(X-M)(X-M)^{\top}] = \begin{pmatrix} t_1 & t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & t_2 & \cdots & t_2 \\ t_1 & t_2 & t_3 & \cdots & t_3 \\ \vdots & \vdots & \vdots & & \vdots \\ t_1 & t_2 & t_3 & \cdots & t_n \end{pmatrix}. \tag{3.37}
$$

where $X = (W_{t_1}, \cdots, W_{t_n})^{\top} \in \mathbb{R}^{n\times 1}$ and $M = \mathbb{E}[X] = (x_1, \cdots, x_0)^{\top}$. Again, we have taken $m = 1$ for simplicity and required $t_1 < t_2 < \cdots < t_n$. Note that the (i, j) entry of the above matrix is $\min(t_i, t_j) = t_{\min(i,j)}$.

The Brownian motion defined in Defintion 3.4 is not unique and we indeed can ask for more about it: By the following Kolmogorov's continuity theorem and Question 2 of Homework III, for almost all $\omega \in \Omega$, we can have a continuous path $t \to W_t(\omega)$ from $[0, \infty)$ to $\mathbb{R}^n$ which satisfies (3.30).

Theorem 3.4 *(Kolmogorov's continuity theorem) Suppose that the process $X = \{X_t\}_{t\geq 0}$ satisfies the following condition: For all $T > 0$, there exist positive constants α, β, D such that*

$$
E[|X_t - X_s|^{\alpha}] \leq D|t-s|^{1+\beta}, \qquad 0 \leq t, s \leq T. \tag{3.38}
$$

Then there exists a process $\tilde{X} = \{\tilde{X}_t\}_{t\geq 0}$ that is continuous with respect to t such that

$$
\mathbb{P}(\{\omega : X_t(\omega) = \tilde{X}_t(\omega)\}) = 1 \qquad \forall\, t. \tag{3.39}
$$

To summarize, we have proved the following three basic properties of Brownian motion. They sometimes are used as the definition of Brownian motion [26] and it has already been used in the Matlab code in Section 3.1 .

[26] See for example Chapter 7.1 of "Probability: Theory and Examples" by Durrett. Indeed, Durrett then needs to answer "Is there a process with these properties". The way to find such a process is by defining measure (3.29) and then use Kolmogorov's Theorems 3.1 and 3.4.

1) $t \to W_t(\omega)$ is continuous and $\mathbb{P}(\{\omega : W_0(\omega) = x_0\}) = 1$.

2) For $0 \le s < t \le T$, the random variable given by the increment $W(t) - W(s)$ is normally distributed with mean zero and variance $t - s$; equivalently, $W(t) - W(s) \sim \sqrt{t-s}N(0,1)$ where $N(0,1)$ denotes a normally distributed random variable with zero mean and unit variance.

3) For $0 \le t_1 < t_2 < \cdots < t_k$,

$$W_{t_1}, W_{t_2} - W_{t_1}, \cdots, W_{t_k} - W_{t_{k-1}} \text{ are independent.} \tag{3.40}$$

In particular, $W_{t_k} - W_{t_{k-1}}$ is independent of $W_{t_{k-1}}$ since the latter is $(W_{t_{k-1}} - W_{t_{k-2}}) + \cdots + (W_{t_2} - W_{t_1}) + W_{t_1}$

A third equivalent definition for $\{W_t\}_{t\ge 0}$ to be an m-dimensional Brownian motion is the following (see for example the book "Functional integration and partial differential equations" by Freidlin):

1) $t \to W_t(\omega)$ is continuous and $\mathbb{P}(\{\omega : W_0(\omega) = x_0\}) = 1$.

2) W_t is a Gaussian process, which means for all $0 \le t_1 \le \cdots \le t_n$, the random variable $Z = (W_{t_1}, \cdots, W_{t_n})$ has a multi normal distribution.

3) $\mathbb{E}W_s = x_0$ and $\mathbb{E}[(W_s - x_0)(W_t - x_0)^\top] = \min(s,t)I_m \overset{\text{def}}{=} (s \wedge t)I_m$.

Please note that our previous Example 3.4 indeed says that the third definition implies Theorem 3.2 which is the construction behind Defintion 3.4, which is the first definition of Brownian motion.

Even though we have not yet introduced "continuous-time martingale" [27], we'd like to document a fourth way to tell if a stochastic process is a Brownian motion: In Question 6 of Homework III, we have seen that $[W,W](t) \overset{\text{def}}{=} \lim_{\max_j |t_{j+1}-t_j| \to 0} \sum_{j=0}^{n-1} |M_{t_{j+1}} - M_{t_j}|^2 = t$ if W is a Brownian motion. The opposite is also true:

Theorem 3.5 *(Levy) Let M_t, $t \ge 0$, be a martingale with respect to a filtration* [28] *$\mathcal{F}_t$, $t \ge 0$. Assume that $M_0 = 0$, M_t has continuous path, and*

$$[M,M](t) \overset{def}{=} \lim_{\max_j |t_{j+1}-t_j| \to 0} \sum_{j=0}^{n-1} |M_{t_{j+1}} - M_{t_j}|^2 = t \quad \forall t \ge 0$$

where $0 = t_0 < t_1 \cdots < t_n = t$ is a partition of the interval $[0,t]$. Then M_t is a Brownian motion.

[27] We have already introduced discrete-time martingale in Defintion 2.4

[28] Definition 2.6

See Theorem 4.6.4 of Shreve II for the proof.

Example 3.5 *(Paley-Wiener representation)* [29] *This is FYI only and won't be tested. It is the example from Page 21 of the book "Diffusion processes and their sample paths" by Ito and McKean. See also Page 29 of the book "Functional integration and partial differential equations" by Freidlin. Given a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and a sequence of i.i.d. random variables* $\xi_i \sim N(0,1)$*. Define*

$$\phi_t(\omega) = t\xi_0 + \frac{\sqrt{2}}{\pi} \sum_{k=1}^{\infty} \frac{\sin(k\pi t)}{k} \xi_k, \qquad 0 \le t \le 1. \tag{3.41}$$

One can check that this series converges uniformly on $[0,1]$ *with probability 1. Since the sum of independent Gaussian (i.e. normally distributed) random variables is also a Gaussian random variable , and since the limit of Gaussian random variables is also a Gaussian random variable, we conclude that the process* $\phi_t(\omega)$ *is also Gaussian. From (3.41), we know*

$$\mathbb{E}\phi_s\phi_t = ts + \sum_{k=1}^{\infty} \frac{2}{\pi^2} \frac{\sin(k\pi s)\sin(k\pi t)}{k^2} = s \wedge t, \qquad s,t \in [0,1].$$

Hence $\phi_t(\omega)$ *is a 1-dimensional Brownian motion. The last identity can be proved using the identity* $\sum_{k=1}^{\infty} \frac{\cos(kx)}{k^2} = \frac{\pi^2}{6} - \frac{\pi x}{2} + \frac{x^2}{4}$ *for* $0 \le x \le 2\pi$ [30]*. More general construction can be found in Section 1.5 of Itô and McKean's book.*

Example 3.6 *Let* $\{W_t, t \ge 0\}$ *be Brownian motion with* $W_0 = 0$*. What is the distribution of* $W_1 + W_3$*?*

Solution*: Recall that the sum of jointly normally distributed random variables is still normally distributed.* $\mathbb{E}[W_1 + W_3] = 0 + 0 = 0$*. To determine the variance, note that*

$$W_1 + W_3 = W_3 - W_1 + 2W_1.$$

$W_3 - W_1$ *and* $2W_1$ *are independent.* $W_3 - W_1 \sim N(0,2)$*.* $Var(W_3 - W_1 + 2W_1) = Var(W_3 - W_1) + Var(2W_1) = 2 + 4\,Var(W_1) = 2 + 4 = 6$*. Hence* $W_1 + W_3 \sim N(0,6)$*.*

Example 3.7 *Let* $\{W_t, t \ge 0\}$ *be Brownian motion with* $W_0 = 0$*. Recall we have learned in (3.37) that* $\mathbb{E}[W_t W_s] = \min(t,s)$*. Find* $Cov(W_2 + W_4, W_1 + W_3)$ *where* $Cov(X,Y) \stackrel{def}{=} \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$*. Then show that* $W_2 + W_4 - \frac{7}{10}(W_1 + W_3)$ *is independent of* $W_1 + W_3$*.*

[29] If the summation in (3.41) is finite $\sum_{k=1}^{N}$ instead of $\sum_{k=1}^{\infty}$, then you can take any times derivative of $\phi_t(\omega)$ with respect to t. But once we let $N \to \infty$, this $\phi_t(\omega)$ becomes not differentiable in t.

[30] See for example Page 47 of "Table of Integrals, Series, and Products" 7th edition by Gradshteyn and Ryzhik.

Solution:

$$\begin{aligned}
Cov(W_2+W_4, W_1+W_3) =& \mathbb{E}[(W_2+W_4-\mathbb{E}[W_2+W_4])(W_1+W_3-\mathbb{E}[W_1+W_3])] \\
=& \mathbb{E}[(W_2+W_4)(W_1+W_3)] \\
=& \mathbb{E}[W_2W_1]+\mathbb{E}[W_2W_3]+\mathbb{E}[W_4W_1]+\mathbb{E}[W_4W_3] \\
=& 1+2+1+3=7.
\end{aligned}$$

$$\begin{aligned}
& Cov(W_2+W_4-\frac{7}{6}(W_1+W_3), W_1+W_3) \\
=& Cov(W_2+W_4, W_1+W_3)-\frac{7}{6}Cov(W_1+W_3, W_1+W_3) \\
=& 7-\frac{7}{6}6=0.
\end{aligned}$$

Let $X = W_2+W_4-\frac{7}{6}(W_1+W_3)$ and $Y = W_1+W_3$. Then X and Y are jointly normally distributed random variables with zero covariance. Hence they are independent.

Example 3.8 *Let W_t be a 2-dimensional Brownian motion which means that its pdf is $\frac{1}{2\pi t}e^{-\frac{x_1^2+x_2^2}{2t}}$. Let $D_\rho = \{x \in \mathbb{R}^2 : \|x\| = \sqrt{x_1^2+x_2^2} < \rho\}$. Compute*

$$\mathbb{P}^0[W_t \in D_\rho].$$

Solution:

$$\mathbb{P}^0(W_t \in D_\rho) = \iint_{D_\rho} \frac{1}{2\pi t}e^{-\frac{x_1^2+x_2^2}{2t}}dx_1dx_2.$$

Let $x_1 = r\cos\theta$ and $x_2 = r\sin\theta$ with $r \in [0,\rho]$, $\theta \in [0,2\pi]$.

$$det\begin{pmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \theta} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \theta} \end{pmatrix} = r.$$

By the change of variable formula from multivariable calculus or simply the polar coordinates,

$$\mathbb{P}^0(W_t \in D_\rho) = \frac{1}{2\pi t}\int_0^{2\pi}\int_0^\rho e^{-\frac{r^2}{2t}}rdrd\theta = \frac{1}{2\pi}2\pi\left(-e^{-\frac{r^2}{2t}}\right)\Big|_{r=0}^{r=\rho} = 1-e^{-\frac{\rho^2}{2t}}.$$

3.4 Homework III

(Only submit solutions to Questions 1,4,6,9,14. Unless otherwise specified, a Brownian motion is assume to be 1-dimensional with zero initial value.)

1. a) Let $X \sim N(0,t)$ which means its pdf is $f_X = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$. Define its moment generating function $\varphi(u) = \mathbb{E}[e^{uX}] = \int_{\mathbb{R}} e^{ux} f_X(x)dx$. Show that $\varphi(u) = e^{\frac{1}{2}u^2 t}$.

   b) If $X \sim N(\mu, \sigma^2)$, show that the so called log-normally distributed random variable $S = e^X$ has mean $e^{\mu + \frac{1}{2}\sigma^2}$. [Hint: To simplify the calculation, consider $\mathbb{E}[e^X] = e^\mu \mathbb{E}[e^{X-\mu}]$.]

   c) What is the variance of e^X if $X \sim N(\mu, \sigma^2)$?

2. The kurtosis of a random variable is defined to be the ratio of its fourth central moment to the square of its variance. Prove that for a normal random variable $X \sim N(\mu, \sigma^2)$, the kurtosis $\frac{\mathbb{E}[(X-\mu)^4]}{(\mathbb{E}[(X-\mu)^2])^2} = 3$. (Hint: If $\mu = 0$, we can use the moment generating function $\varphi(u) = \mathbb{E}[e^{uX}]$ defined in Question 1. Taking 2 derivatives with respect to u, we get $\varphi''(u) = \mathbb{E}[X^2 e^{\mu X}]$, which is $(t + tu)e^{\frac{1}{2}u^2 t}$ since $\varphi(u) = e^{\frac{1}{2}u^2 t}$. By setting $u = 0$, we obtain $\mathbb{E}[X^2] = t$. By taking 4 derivatives, one can calculate $\mathbb{E}[X^4]$.) Then show that

   $$E[|W_t - W_s|^4] = 3|t-s|^2, \qquad 0 \le t, s \le T.$$

   Hence (3.38) is true with $D = 3$, $\alpha = 4$, $\beta = 1$ and "$\le$" is indeed "$=$" for the Brownian motion case.

   Solution: Define $\varphi(u) = \mathbb{E}[e^{u(X-\mu)}]$. Then $\varphi(u) = e^{\frac{1}{2}u^2\sigma^2}$ by Question 1.

   $$\varphi^{(4)}(u) = \mathbb{E}[(X-\mu)^4 e^{u(X-\mu)}] = \left(3\sigma^4 + 6u^2\sigma^6 + u^4\sigma^8\right) e^{\frac{1}{2}u^2\sigma^2}.$$

   Letting $u = 0$, we get

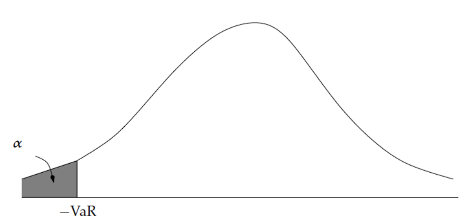
   $$\mathbb{E}[(X-\mu)^4] = 3\sigma^4.$$

   Since $W_t - W_s \sim N(0, t-s)$, $E[|W_t - W_s|^4] = 3|t-s|^2$.

3. Value-at-risk (VaR) denotes, within a confidence level, the maximum loss a portfolio could suffer. To be more precise, denote by X the change in the market value of a portfolio during a given time period. Then the VaR with confidence level $1-\alpha$ is defined to be the value VaR in

   $$\mathbb{P}(X \le -\mathrm{VaR}) = \alpha. \tag{3.42}$$

   In other words, with probability $1-\alpha$, the maximum loss will not exceed VaR.



Show that if X is assumed to be normally distributed with mean μ and variance σ^2, the VaR with confidence level $1-\alpha$ is given by

$$\mathrm{VaR} = z_\alpha \sigma - \mu \tag{3.43}$$

where z_α satisfies $N(-z_\alpha) = \alpha$ with function N defined in (3.15). For example, $z_{0.1} = 1.2816$, $z_{0.05} = 1.6449$, $z_{0.01} = 2.3263$, $z_{0.005} = 2.5758$.

Proof: $X \sim N(\mu, \sigma^2)$ implies that $\frac{X-\mu}{\sigma} \sim N(0,1)$. Hence

$$\begin{aligned} N(-z_\alpha) = \alpha = \mathbb{P}(X \le -\mathrm{VaR}) &= \mathbb{P}\left(\frac{X-\mu}{\sigma} \le \frac{-\mathrm{VaR}-\mu}{\sigma}\right) \\ &= \int_{-\infty}^{-\frac{\mathrm{VaR}+\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = N\left(-\frac{\mathrm{VaR}+\mu}{\sigma}\right). \end{aligned}$$

So $z_\alpha = \frac{\mathrm{VaR}+\mu}{\sigma}$ which proved (3.43). $\square$

4. A binary call option with maturity T pays one dollar when the stock price at time T is at or above a certain level K and pays nothing otherwise. The payoff can be written in the form of an indicator function

$$X = 1_{\{S_T \ge K\}}.$$

Suppose in the risk neutral world $\log S_T$ is normally distributed with mean $\log S_0 + (r - \frac{\sigma^2}{2})T$ and variance $\sigma^2 T$. Show that the discounted expected payoff $e^{-rT}\mathbb{E}X$ is

$$e^{-rT}\mathbb{E}X = e^{-rT} N(d_-), \tag{3.44}$$

where the function N is defined in (3.15) and the d_- is defined in (3.14). [Hint: Note that $\mathbb{E}X = \mathbb{P}(\log S_T \ge \log K)$.]

5. Let $f(t)$ be a function defined for $0 \le t \le T$. The quadratic variation of f up to time T is defined to be

$$[f,f](T) = \lim_{\max_j |t_{j+1}-t_j| \to 0} \sum_{j=0}^{n-1} |f(t_{j+1}) - f(t_j)|^2 \tag{3.45}$$

where $0 = t_0 < t_1 < \cdots < t_n = T$ is a partition of the interval $[0,T]$. Recall that the mean value theorem says that if f has continuous derivative, for any $[t_j, t_{j+1}]$, there is a $t_j^* \in (t_j, t_{j+1})$ so that $f(t_{j+1}) - f(t_j) = f'(t_j^*)(t_{j+1} - t_j)$.

Prove that if f has continuous derivative on $[0,T]$ with $T < \infty$, then

$$[f,f](T) = 0.$$

Here you can use the fact that f has continuous derivative on $[0,T]$ implies that $M \stackrel{\text{def}}{=} \sup_{s\in[0,T]} |f'(s)| < \infty$.

Proof:

$$
\begin{aligned}
0 \le [f,f](T) &= \lim_{\max_j |t_{j+1}-t_j| \to 0} \sum_{j=0}^{n-1} |f(t_{j+1}) - f(t_j)|^2 \\
&= \lim_{\max_j |t_{j+1}-t_j| \to 0} \sum_{j=0}^{n-1} |f'(t_j^*)|^2 |t_{j+1} - t_j|^2 \\
&\le \lim_{\max_j |t_{j+1}-t_j| \to 0} \sum_{j=0}^{n-1} \left(M^2 \max_j |t_{j+1} - t_j| \right) |t_{j+1} - t_j| \\
&= \lim_{\max_j |t_{j+1}-t_j| \to 0} \left(M^2 \max_j |t_{j+1} - t_j| \right) \sum_{j=0}^{n-1} |t_{j+1} - t_j| \\
&= \lim_{\max_j |t_{j+1}-t_j| \to 0} \left(M^2 \max_j |t_{j+1} - t_j| \right) T = 0.
\end{aligned}
$$

6. By definition, we say that $X_n(\omega) \to X(\omega)$ in mean square sense if $\mathbb{E}[(X_n - X)^2] \to 0$ as $n \to \infty$. Given a Brownian motion W_t, define a random variable $[W,W]_{\Pi,T}$ by

$$[W,W]_{\Pi,T}(\omega) = \sum_{j=0}^{n-1} |W_{t_{j+1}}(\omega) - W_{t_j}(\omega)|^2$$

where $\Pi = \{0 = t_0, t_1, \cdots, t_n = T\}$ is the partition of $[0,T]$. This question asks you to show that the random variable $[W,W]_{\Pi,T} \to T$ in the mean square sense when $\|\Pi\| \stackrel{\text{def}}{=} \max_j |t_{j+1} - t_j| \to 0$ [31]. The proof is split into two steps:

a) Show that
$$\mathbb{E}[[W,W]_{\Pi,T}] = T. \tag{3.47}$$

b) By (3.47) and $\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$, the variance of $[W,W]_{\Pi,T}$ is $\mathrm{Var}[[W,W]_{\Pi,T}] = \mathbb{E}[([W,W]_{\Pi,T} - T)^2]$. Show that
$$\mathrm{Var}[[W,W]_{\Pi,T}] = \sum_{j=0}^{n-1} \mathrm{Var}[|W_{t_{j+1}} - W_{t_j}|^2] \to 0$$
as $\|\Pi\| \to 0$.

7. We now continue with Example 2.11 of Chapter 2 to introduce Markov property:

[31] Theorem 3.4.3 of Shreve II says that
$$[W,W]_{\Pi,T} \to T \text{ almost surely when } \max_j |t_{j+1} - t_j| \to 0. \tag{3.46}$$
Note that we say that a statement A that depends on ω is true almost surely if $\mathbb{P}(\{\omega : A(\omega) \text{ is true}\}) = 1$. Hence what we will prove in this question is that the convergence in (3.46) happens in mean square sense.

Definition 3.5 *(Definition 2.5.1 of Shreve I) Consider the binomial asset-pricing model. Let X_0, X_1, ... , X_N be a sequence of random variables, with each X_n depending only on the first n coin tosses (and X_0 constant). If, for every n between 0 and $N-1$ and for every function $f(x)$, there is another function $g(x)$ (depending on f and n) such that*

$$\tilde{\mathbb{E}}_n[f(X_{n+1})] = g(X_n), \tag{3.48}$$

we say that $X_0, X_1, \cdots, X_n$ is a Markov process.

The Markov property says that the dependence of $\tilde{\mathbb{E}}_n[f(X_{n+1})]$ on the first n coin tosses occurs through X_n (i.e., the information about the coin tosses one needs in order to evaluate $\tilde{\mathbb{E}}[f(X_{n+1})]$ is summarized by X_n).

It also implies the "two-step ahead" property:

$$\tilde{\mathbb{E}}_n[h(X_{n+2})] \overset{(2.31)}{=} \tilde{\mathbb{E}}_n[\tilde{\mathbb{E}}_{n+1}[h(X_{n+2})]] = \tilde{\mathbb{E}}_n[f(X_{n+1})] = g(X_n), \tag{3.49}$$

or "multi-step ahead" property: For any function f and any $0 \leq n \leq m \leq N$, there is a function g so that

$$\tilde{\mathbb{E}}_n[h(X_m)] = g(X_n). \tag{3.50}$$

Now comes the question: Consider the stock price model

$$S_{n+1}(\omega_1 \cdots \omega_n \omega_{n+1}) = \begin{cases} uS_n(\omega_1 \cdots \omega_n) & \text{if } \omega_{n+1} = H, \\ dS_n(\omega_1 \cdots \omega_n) & \text{if } \omega_{n+1} = T. \end{cases}$$

Find the function g so that

$$\tilde{\mathbb{E}}_n[f(S_{n+1})](\omega_1 \cdots \omega_n) = g(S_n).$$

Solution: $g(x) = q_u f(ux) + q_d f(dx)$.

8. (Conditional distribution) The conditional probability of event A happens under the condition that event B already happens is denoted as $\mathbb{P}(A|B)$ and we have the relation

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)} \tag{3.51}$$

if $\mathbb{P}(B) > 0$. If we have joint density function $\rho_{X,Y}$ for two continuous random variables X and Y, we define marginal density function $\rho_X(x)$ and $\rho_Y(y)$ as in (3.18). Then we can calculate conditional distribution

$$\begin{aligned}
\mathbb{P}(X \in A|Y = y) &\overset{\text{def}}{=} \lim_{a \downarrow 0} \mathbb{P}(X \in A|Y \in [y-a, y+a]) \overset{(3.51)}{=} \lim_{a \downarrow 0} \frac{\mathbb{P}(X \in A, Y \in [y-a, y+a])}{\mathbb{P}(Y \in [y-\Delta, y+\Delta])} \\
&= \lim_{a \downarrow 0} \frac{\int_{[y-a,y+a]} \left(\int_A \rho_{X,Y}(x,v)dx\right) dv}{\int_{[y-a,y+a]} \rho_Y(v)dv} \\
&= \lim_{a \downarrow 0} \frac{\left(\int_A \rho_{X,Y}(x,y)dx\right) \times (2a) + \text{ higher order term of } a}{\rho_Y(y) \times (2a) + \text{ higher order term of } a}.
\end{aligned}$$

In the last step, we used again the fact that $\int_a^b f(v)dv = f(\frac{a+b}{2})(b-a)+$ higher order term of $(b-a)$. Now, dividing $2a$ from both the numerator and denomonator, we get

$$\mathbb{P}(X \in A|Y = y) = \lim_{a \downarrow 0} \frac{\int_A \rho_{X,Y}(x,y)dx + \frac{\text{higher order term of } a}{2a}}{\rho_Y(y) + \frac{\text{higher order term of } a}{2a}}$$
$$= \frac{\int_A \rho_{X,Y}(x,y)dx + \lim_{a \downarrow 0} \frac{\text{higher order term of } a}{2a}}{\rho_Y(y) + \lim_{a \downarrow 0} \frac{\text{higher order term of } a}{2a}}.$$

In the last step, we have used $\lim_{a \to b} \frac{f(a)}{g(a)} = \frac{\lim_{a \to b} f(a)}{\lim_{a \to b} g(a)}$. Continue, we get

$$\mathbb{P}(X \in A|Y = y) = \frac{\int_A \rho_{X,Y}(x,y)dx + 0}{\rho_Y(y) + 0} = \int_A \frac{\rho_{X,Y}(x,y)}{\rho_Y(y)} dx.$$

Hence by (3.1), the conditional density of X given $Y = y$ is

$$\frac{\rho_{X,Y}(x,y)}{\rho_Y(y)} \stackrel{\text{def}}{=} \rho_{X|Y}(x|y), \tag{3.52}$$

which, by the way, also depends on y. Since we now have the density function, by (3.10), we know

$$\mathbb{E}[h(X)|Y = y] = \int_{\mathbb{R}} h(x)\rho_{X|Y}(x|y)dx \tag{3.53}$$

which is a function of y. By the way, if X and Y are independent, by (3.20), $\rho_{X|Y}(x|y) = \frac{\rho_{X,Y}(x,y)}{\rho_Y(y)} = \rho_X(x)$, then

$$\mathbb{E}[h(X)|Y = y] = \int_{\mathbb{R}} h(x)\rho_{X|Y}(x|y)dx = \int_{\mathbb{R}} h(x)\rho_X(x)dx = \mathbb{E}[h(X)]. \tag{3.54}$$

Go back to (3.53). Note that $\mathbb{E}[h(X)|Y] \stackrel{\text{def}}{=} \mathbb{E}[h(X)|Y = Y]$ is a function of Y which is hence a random variable by itself. As we already know the density function of Y, by (3.10), we can compute

$$\mathbb{E}[\mathbb{E}[h(X)|Y]] = \int_{\mathbb{R}} \mathbb{E}[h(X)|Y = y]\ \rho_Y(y)dy.$$

Here comes the question: Prove the iterated conditioning property

$$\mathbb{E}[\mathbb{E}[h(X)|Y]] = \mathbb{E}[h(X)]. \tag{3.55}$$

Remark: You should compare $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$ with

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}_m[X]] = \mathbb{E}[\mathbb{E}[X|\text{given } \omega_1 \cdots \omega_m]] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_m]], \tag{3.56}$$

which we have learned before in Theorem 2.2 (we used $\tilde{\mathbb{E}}$ there to stress that we were using the risk-neutral probability. But apparently, it is true for any probability). The precise meaning of the last term, $\mathbb{E}[\mathbb{E}[X|\mathcal{F}_m]]$, will be clarified in Theorem 4.3 of the next chapter.

Proof:

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[h(X)|Y]] &\overset{(3.10)}{=} \int_{\mathbb{R}} \mathbb{E}[h(X)|Y=y]\rho_Y(y)dy \\
&\overset{(3.53)}{=} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x)\rho_{X|Y}(x|y)dx \right) \rho_Y(y)dy \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x)\frac{\rho_{X,Y}(x,y)}{\cancel{\rho_Y(y)}}dx \right) \cancel{\rho_Y(y)}dy \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x)\rho_{X,Y}(x,y)dx \right) dy \\
&= \mathbb{E}[h(X)].
\end{aligned}$$

9. (Continue with Question 8.) Let S_1 and S_2 be the prices of two assets. Assume that $X = \log S_1$ and $Y = \log S_2$ have a joint density function

$$\rho_{X,Y} = \frac{\sqrt{3}}{4\pi}e^{-\frac{1}{2}(x^2-xy+y^2)}, \qquad \text{for } x, y \in \mathbb{R}.$$

Determine $\mathbb{E}[X|Y]$.

10. (This problem is rather difficult and won't be tested. But I hope you can read it so that you have a better understanding of the density function of Brownian motion. This example is taken from Question 2.12 of Hui Wang, "Monte Carlo Simulations with Applications to Finance". Please bear with me for the complicated computations. If your probability/finance intuition already tells you that the conclusion is obvious, you do not need to read the proof.) Given an arbitrary constant θ, let $B = \{B_t : t \geq 0\}$ be a Brownian motion with drift θ, i.e.,

$$B_t = W_t + \theta t, \qquad t \geq 0, \tag{3.57}$$

where $W = \{W_t : t \geq 0\}$ is a Brownian motion. Given $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = T$, show that the conditional distribution of the $(n-1)$-dimensional random variable $X = (B_{t_1}, \cdots, B_{t_{n-1}})$ given $Y = B_T = y$ does not depend on θ. In particular, letting $\theta = 0$, we conclude that the conditional distribution of $\{B_t : 0 \leq t \leq T\}$ given $B_T = y$ is the same as the conditional distribution of $\{W_t : 0 \leq t \leq T\}$ given $W_T = y$.

Proof: Recall (3.32)

$$\rho_{W_{t_1},\cdots,W_{t_n}}(x_1,\cdots x_n) = \frac{1}{(2\pi)^{n/2}\sqrt{\prod_{i=1}^n (t_i - t_{i-1})}} \exp\left(-\sum_{i=1}^n \frac{(x_i - x_{i-1})^2}{2(t_i - t_{i-1})} \right)$$

and $W_{t_n} \sim N(x_0, t_n)$ which means

$$\rho_{W_{t_n}}(x_n) = \frac{1}{\sqrt{2\pi(t_n - t_0)}} \exp(-\frac{(x_n - x_0)^2}{2t_n}).$$

Since $B_{t_i} = W_{t_i} + \theta t_i \sim N(x_0 + \theta t_i, t_i)$, by the change of variable formula in multivariable calculus or probability, which is (3.34), we get

$$\rho_{B_{t_1},\cdots,B_{t_n}}(x_1, \cdots x_n) = \frac{1}{(2\pi)^{n/2}\sqrt{\prod_{i=1}^{n}(t_i - t_{i-1})}} \exp\left(-\sum_{i=1}^{n} \frac{[(x_i - \theta t_i) - (x_{i-1} - \theta t_{i-1})]^2}{2(t_i - t_{i-1})}\right)$$

and

$$\rho_{B_{t_n}}(x_n) = \frac{1}{\sqrt{2\pi(t_n - t_0)}} \exp\left(-\frac{(x_n - \theta t_n - x_0)^2}{2t_n}\right).$$

To calculate $\frac{\rho_{B_{t_1},\cdots,B_{t_n}}(x_1,\cdots x_n)}{\rho_{B_{t_n}}(x_n)}$, we need to compute

$$\begin{aligned}
&-\sum_{i=1}^{n} \frac{[(x_i - x_{i-1}) - \theta(t_i - t_{i-1})]^2}{2(t_i - t_{i-1})} + \frac{[(x_n - x_0) - \theta t_n]^2}{2t_n} \\
&= -\sum_{i=1}^{n} \frac{(x_i - x_{i-1})^2}{2(t_i - t_{i-1})} + \cancel{\sum_{i=1}^{n} \theta(x_i - x_{i-1})} - \cancel{\frac{\theta^2}{2}\sum_{i=1}^{n}(t_i - t_{i-1})} - \frac{(x_n - x_0)^2}{2t_n} + \cancel{\theta(x_n - x_0)} + \cancel{\frac{\theta^2}{2}t^n} \\
&= -\sum_{i=1}^{n} \frac{(x_i - x_{i-1})^2}{2(t_i - t_{i-1})} + \frac{(x_n - x_0)^2}{2t_n}.
\end{aligned}$$

Hence the conditional density function of $(B_{t_1}, \cdots, B_{t_{n-1}})$ given $B_{t_n} = x_n$ is

$$\frac{\rho_{B_{t_1},\cdots,B_{t_n}}(x_1, \cdots x_n)}{\rho_{B_{t_n}}(x_n)} = \frac{\sqrt{t_n - t_0}}{(2\pi)^{(n-1)/2}\sqrt{\prod_{i=1}^{n}(t_i - t_{i-1})}} \exp\left(-\sum_{i=1}^{n} \frac{(x_i - x_{i-1})^2}{2(t_i - t_{i-1})} + \frac{(x_n - x_0)^2}{2t_n}\right). \tag{3.58}$$

The parameter θ has disappeared from the above formula. This proves the result. □

11. (This problem is rather difficult and won't be tested. This example is taken from Page 40 of Hui Wang, "Monte Carlo Simulations with Applications to Finance". It shows an application of conditional expectation. The result is useful for pricing, say, a look back call option whose payoff is $(\max_{0 \le t \le T} S_t - K)^+$ and is path-dependent. See Example 2.6 of "Monte Carlo Simulations with Applications to Finance" by Hui Wang for more details.) We say a function is path-dependent if it depends on the sample paths of the relevant process. For example, we can define

$$h(W_{[0,T]}) = \max_{0 \le t \le T} W_t - \min_{0 \le t \le T} W_t - W_T \tag{3.59}$$

whose value depends on the entire sample path $W_{[0,T]} = \{W_t : 0 \le t \le T\}$.

Introduce B_t and W_t as in Question 10 with $W_0 = 0$. Prove that for any path dependent function h,

$$\mathbb{E}\left[h(B_{[0,T]})\right] = \mathbb{E}\left[e^{\theta W_T - \frac{1}{2}\theta^2 T} h(W_{[0,T]})\right]. \tag{3.60}$$

Proof: Since $W_T \sim N(0,T)$, by (3.55),

$$\begin{aligned} RHS &= \int_{\mathbb{R}} \mathbb{E}\left[e^{\theta W_T - \frac{1}{2}\theta^2 T} h(W_{[0,T]}) \middle| W_T = x\right] \frac{1}{\sqrt{2\pi T}} e^{-\frac{x^2}{2T}} dx \\ &= \int_{\mathbb{R}} \mathbb{E}\left[e^{\theta x - \frac{1}{2}\theta^2 T} h(W_{[0,T]}) \middle| W_T = x\right] \frac{1}{\sqrt{2\pi T}} e^{-\frac{x^2}{2T}} dx \\ &= \int_{\mathbb{R}} \mathbb{E}\left[h(W_{[0,T]}) \middle| W_T = x\right] \frac{1}{\sqrt{2\pi T}} e^{\theta x - \frac{1}{2}\theta^2 T - \frac{x^2}{2T}} dx \\ &= \int_{\mathbb{R}} \mathbb{E}\left[h(W_{[0,T]}) \middle| W_T = x\right] \frac{1}{\sqrt{2\pi T}} e^{-\frac{1}{2T}(x-\theta T)^2} dx. \end{aligned}$$

By Question 10, $\mathbb{E}\left[h(W_{[0,T]}) \middle| W_T = x\right] = \mathbb{E}\left[h(B_{[0,T]}) \middle| B_T = x\right]$. Hence

$$RHS = \int_{\mathbb{R}} \mathbb{E}\left[h(B_{[0,T]}) \middle| B_T = x\right] \frac{1}{\sqrt{2\pi T}} e^{-\frac{1}{2T}(x-\theta T)^2} dx$$

and $\frac{1}{\sqrt{2\pi T}} e^{-\frac{1}{2T}(x-\theta T)^2}$ is pdf of B_T. Hence

$$RHS = \int_{\mathbb{R}} \mathbb{E}\left[h(B_{[0,T]}) \middle| B_T = x\right] \rho_{B_T}(x) dx = \mathbb{E}\left[\mathbb{E}\left[h(B_{[0,T]}) \middle| B_T\right]\right] = \mathbb{E}[h(B_{[0,T]})]$$

by (3.55). This finishes the proof.

12. Let $\{W_t, t \geq 0\}$ be a Brownian motion. Find $\mathbb{P}(W_3 \leq 1 | W_2 = \frac{1}{2})$.

    Solution: Note that W_2 and $W_3 - W_2$ are independent.

    $$\mathbb{P}(W_3 \leq 1 | W_2 = \frac{1}{2}) = \mathbb{P}(W_3 - W_2 \leq \frac{1}{2} | W_2 = \frac{1}{2}) = \mathbb{P}(W_3 - W_2 \leq \frac{1}{2})$$

    Since $W_3 - W_2 \sim N(0,1)$, $\mathbb{P}(W_3 - W_2 \leq \frac{1}{2}) = \int_{-\infty}^{1/2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \approx 0.6915$.

13. (To compare with Question 12.) Let $\{W_t, t \geq 0\}$ be a Brownian motion with $W_0 = 0$. Find $\mathbb{P}(W_2 \leq 1 | W_3 = \frac{1}{2})$.

    Solution: By (3.52), if $t_1 < t_2$,

    $$\rho_{W_{t_1}|W_{t_2}}(x_1|x_2) = \frac{\rho_{W_{t_1},W_{t_2}}(x_1,x_2)}{\rho_{W_{t_2}}(x_2)} = \frac{\sqrt{t_2}}{\sqrt{2\pi}\sqrt{t_1(t_2-t_1)}} e^{-\frac{x_1^2}{2t_1} - \frac{(x_2-x_1)^2}{2(t_2-t_1)} + \frac{x_2^2}{2t_2}}$$

    $$= \frac{1}{\sqrt{2\pi \frac{t_1(t_2-t_1)}{t_2}}} e^{-\frac{(t_2 x_1 - t_1 x_2)^2}{2t_1 t_2 (t_2-t_1)}} = \frac{1}{\sqrt{2\pi \frac{t_1(t_2-t_1)}{t_2}}} e^{-\frac{(x_1 - \frac{t_1}{t_2}x_2)^2}{2\frac{t_1(t_2-t_1)}{t_2}}} \sim N(\frac{t_1}{t_2}x_2, \frac{t_1(t_2-t_1)}{t_2}). \tag{3.61}$$

Hence $W_2|W_3 = \frac{1}{2} \sim N(\frac{2}{3}\frac{1}{2}, \frac{2}{3}) = N(\frac{1}{3}, \frac{2}{3})$ with pdf $\frac{1}{\sqrt{2\pi\frac{2}{3}}}e^{-\frac{(x_2-\frac{1}{3})^2}{2\times\frac{2}{3}}}$,

$$\mathbb{P}(W_2 \le 1 | W_3 = \frac{1}{2}) = \int_{-\infty}^{1} \frac{1}{\sqrt{2\pi\frac{2}{3}}}e^{-\frac{(x_2-\frac{1}{3})^2}{2\times\frac{2}{3}}}dx_2 \overset{u=\frac{x_2-\frac{1}{3}}{\sqrt{\frac{2}{3}}}}{=} \int_{-\infty}^{\sqrt{2/3}} \frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du \approx 0.7929.$$

14. Let $X_t = e^{W_t - \frac{1}{2}t}$ be the price of a stock where W_t is a Brownian motion with $W_0 = 0$. Find $\mathbb{P}(X_3 \le 3)$. [Hint: Event $X_3 \le 3$ and event $\log X_3 \le \log 3$ are equivalent.]

15. Stock prices are sometimes modeled by distributions other than lognormal in order to fit the empirical data more accurately. For instance, soon after his Black-Scholes-Merton work, Merton introduces a jump diffusion model for stock prices. A special case of Merton's model assumes that the underlying stock price S satisfies

$$S = e^Y, \qquad Y = X_1 + \sum_{i=1}^{X_2} Z_i \tag{3.62}$$

where X_1 is $N(\mu, \sigma^2)$, X_2 is Poisson with parameter λ[32], Z_i is $N(0, \nu^2)$, and X_1, X_2, $\{Z_i\}$ are all independent. The evaluation of call options involves expected values such as

$$\mathbb{E}[(S-K)^+], \tag{3.63}$$

where K is some positive constant. Compute this expected value.

Solution: For every $n \ge 0$, we can compute the conditional expected value

$$v_n = \mathbb{E}[(S-K)^+ | X_2 = n].$$

Recall that the sum of jointly normally distributed random variables are still normally distributed. So, conditional on $X_2 = n$, $Y = X_1 + Z_1 + \cdots + Z_n$ is normally distributed as $N(\mu, \sigma^2 + n\nu^2)$ since $\mathbb{E}[Y] = \mu$ and $\mathrm{Var}[Y] = \sigma^2 + n\nu^2$ by the independence assumption. By (3.16), we get

$$v_n = \mathbb{E}[(S-K)^+ | X_2 = n] = e^{\mu + \frac{1}{2}(\sigma^2 + n\nu^2)}N(d_{n,+}) - KN(d_{n,-}) \tag{3.64}$$

with $d_{n,-} = \frac{\log\frac{1}{K} + \mu}{\sqrt{\sigma^2 + n\nu^2}}$ and $d_{n,+} = d_{n,-} + \sqrt{\sigma^2 + n\nu^2}$. By (3.55) of Question 8,

$$\begin{aligned}\mathbb{E}[(S-K)^+] = \mathbb{E}[\mathbb{E}[(S-K)^+ | X_2]] &= \sum_{n=0}^{\infty} \mathbb{E}[(S-K)^+ | X_2 = n]\mathbb{P}(X_2 = n) \\ &= \sum_{n=0}^{\infty} v_n \mathbb{P}(X_2 = n) = e^{-\lambda}\sum_{n=0}^{\infty} v_n \frac{\lambda^n}{n!}.\end{aligned}$$

The last expression can be evaluated numerically.

[32] which means $\mathbb{P}(X_2 = n) = e^{-\lambda}\frac{\lambda^n}{n!}$ for $n = 0, 1, 2, 3, \ldots$.

16. (Cholesky factorization and independent increment) Show that the right hand side of (3.37), called Σ, has a decomposition

$$\Sigma = \mathbb{E}[(Z-M)(Z-M)^\top] = \begin{pmatrix} t_1 & t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & t_2 & \cdots & t_2 \\ t_1 & t_2 & t_3 & \cdots & t_3 \\ \vdots & \vdots & \vdots & & \vdots \\ t_1 & t_2 & t_3 & \cdots & t_n \end{pmatrix} = AA^\top \tag{3.65}$$

where

$$A = \begin{pmatrix} \sqrt{t_1} & 0 & 0 & \cdots & 0 \\ \sqrt{t_1} & \sqrt{t_2 - t_1} & 0 & \cdots & 0 \\ \sqrt{t_1} & \sqrt{t_2 - t_1} & \sqrt{t_3 - t_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \sqrt{t_1} & \sqrt{t_2 - t_1} & \sqrt{t_3 - t_2} & \cdots & \sqrt{t_n - t_{n-1}} \end{pmatrix}. \tag{3.66}$$

Let $X = (X_1, \cdots, X_n)^\top \sim N(\mathbf{0}, I_n)$, which means $X_i \in N(0,1)$ and X_i's are independent. **Show** that $AX \sim N(\mathbf{0}, \Sigma)$. By the way, as an application of the result, one can generate $Z = (W_{t_1}, \cdots, W_{t_n})^\top$ as AX. This is precisely the telescoping sum (3.35) and is also the method used in the Matlab code at the beginning of this Chapter:

```
for j = 2:N                   % start the iteration
    dW(j) = sqrt(dt)*randn; % general increment
    W(j) = W(j-1) + dW(j);
end
```

Proof: By direct calculation, one can easily check that $AA^\top = \Sigma$. By (3.5), we know $AX \sim N(\mathbf{0}, AI_nA^\top) = N(\mathbf{0}, AA^\top) = N(\mathbf{0}, \Sigma)$.

4 The Itô integral (1 lecture)

Recall the $B_t = K_0 e^{rt}$ defined in (1.14). B_t satisfies the following ordinary differential equation

$$dB_t = rB_t dt, \qquad B_0 = K_0 \tag{4.1}$$

which can also be written into integral form

$$B_t = B_0 + \int_0^t rB_s ds.$$

The last term is a standard integral which is defined by

$$\int_0^t f(x)dx = \lim_{\max_i \Delta s_i \to 0} \sum_{i=1}^N f(s_i^*)\Delta s_i$$

where $0 = s_0 < s_1 < \cdots s_N = t$ form a partition of interval $[0, t]$, $\Delta s_i = s_i - s_{i-1}$, s_i^* is any point in the i-th interval $[s_{i-1}, s_i]$. The value of the integral does not depend on how we choose s_i^*.

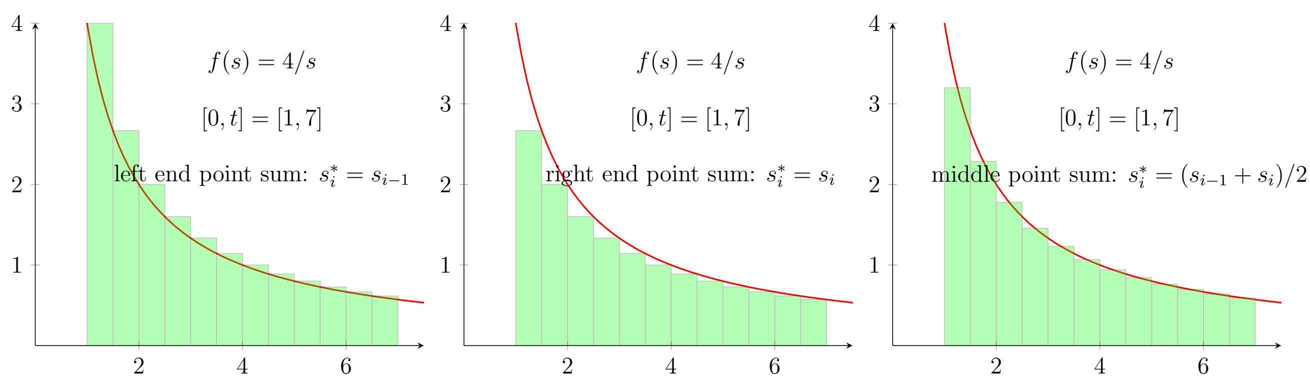


Figure 4.1: $\int_1^7 f(s)ds$ (area under the red line) and $\sum_{i=1}^{12} f(s_i^*)\Delta s_i$ (green area)

We want to introduce Itô integral so that the stochastic process $S_t \in \mathbb{R}^m$ satisfying

$$S_t(\omega) = S_0(\omega) + \int_0^t b(t, S_s)ds + \underbrace{\int_0^t \sigma(s, S_s)dW_s}_{\text{Itô integral}}$$

can be used as a model for prices of m different assets. Here $S_t, b \in \mathbb{R}^m$, $\sigma \in \mathbb{R}^{m\times n}$, and $W_s \in \mathbb{R}^n$. (You won't miss anything important if you think $m = 1 = n$.) We will know (Question 16 of Homework IV) that W_s is not differentiable. Hence we cannot write $\frac{dW_s}{ds}$ and

$$\int_0^t \sigma(s, S_s)dW_s \neq \int_0^t \sigma(s, S_s)\frac{dW_s}{ds}ds.$$

It turns out under some mild condition on f,

$$\int_0^t f(s,\omega)dW_s = \lim_{\max_i \Delta s_i \to 0} \sum_{i=1}^{N} f(s_{i-1},\omega)(W_{s_i}(\omega) - W_{s_{i-1}}(\omega)), \tag{4.2}$$

where $0 = s_0 < s_1 < \cdots < s_N = t$ form a partition of $[0,t]$. We should always choose s_i^* to be the left end point of the interval $[s_{i-1}, s_i]$. See (1.6). The issue is how you pass to the limit and how you justify the existence of such a limit.

Example 4.1 *We generate a single sample path of $\int_0^t W_s dW_s$ and compare it with the theoretical result $\frac{1}{2}W_t^2 - \frac{1}{2}t$ which will be proved later. We have to stress again that in the definition of the Itô integral, we take the left endpoint from the subinterval $[s_{i-1}, s_i]$*

```
T= 3.0; N = 300;
dt = T/N;
t = 0:dt:T;
dW = sqrt(dt)*randn(1,N);
W = zeros(1,N+1);
Integral = zeros(1,N+1);
Exact = zeros(1,N+1);

for i = 1:N
    W(i+1) = W(i) + dW(i);
    Integral(i+1) = Integral(i) + W(i)*dW(i); % Take the left endpoint.
    Exact(i+1)=W(i+1)^2/2 - i*dt/2;
end
plot(t,Integral,'r-',t,Exact,'k-.');
xlabel('t');
hlegend=legend('approx','exact');
```

Definition 4.1 *(Definition 3.1.2 of Oksendal) Let $W_t(\omega)$ be n-dimensional Brownian motion. Then we define $\mathcal{F}_t$ to be the smallest σ-algebra (Definition 2.5) containing all sets of the form*

$$\{\omega : W_{t_1}(\omega) \in \Gamma_1, \cdots, W_{t_k}(\omega) \in \Gamma_k\}, \tag{4.3}$$

where $t_j \le t$ and $\Gamma_j \subset \mathbb{R}^n$ are Borel sets (you can think of them as intervals for $n = 1$ and rectangular boxes for $n \ge 2$), k can be any positive integer. For technical reasons, we also assume that all sets of measure zero are included in $\mathcal{F}_t$.

Note that $\mathcal{F}_s \subset \mathcal{F}_t$ for $s \le t$. By Definition 2.6, $\{\mathcal{F}_t : 0 \le t \le T\}$ is a filtration. In addition to $\{\mathcal{F}_t\}$, we also have an $\mathcal{F}$ which is the σ-algebra on Ω. The existence of Ω and $\mathcal{F}$ is guaranteed by Theorem 3.1. Recall (2.15) and Example 2.4 to see what Ω looks like.

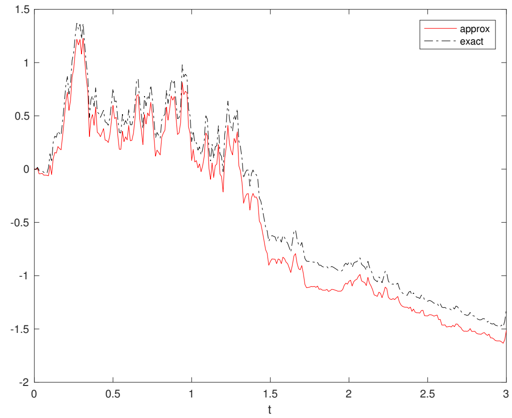


Figure 4.2: Simulation of the Itô integral $\int_0^t W_s dW_s$ and the exact answer $\frac{1}{2}W_t^2 - \frac{1}{2}t$.

Since Ω is larger than any set (representing information about ω) that generates $\mathcal{F}_t$, $\mathcal{F} \supset \mathcal{F}_t$ for all t. You should now recall the $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ that we have introduced in Section 2.9 and see that $\{\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3\}$ is just a discrete in time version of $\{\mathcal{F}_t : 0 \le t \le T\}$.

Now we follow the presentation of Oksendal. One often thinks of $\mathcal{F}_t$ as the history of W_s up to time t. Recall Definition 2.8 on measurability. It turns out that (which follows from (4.3) in the definition of $\mathcal{F}_t$. But don't worry about its proof.) a function $X(\omega)$ is $\mathcal{F}_t$ measurable if and only if X can be written as the pointwise limit (for almost all ω) of summations of functions of the form

$$g_1(W_{t_1})g_2(W_{t_2})\cdots g_k(W_{t_k})$$

where $g_1, \cdots, g_k$ are bounded continuous functions and $t_j \le t$ for $j \le k$, k can be any positive integer. Intuitively, X is $\mathcal{F}_t$ measurable means that to know the value of $X(\omega)$, we do not have to know the full ω, we just need to know $W_s(\omega)$ for $s \le t$. The value of $X(\omega)$ can be derived from the value of $W_s(\omega)$ for $s \le t$. For example, the random variable $X(\omega) = (W_{t/2}(\omega))^3 + 7W_t(\omega)$ is $\mathcal{F}_t$-measurable, while $Y(\omega) = \sqrt{\sin(W_{2t}(\omega)) + 5}$ is not $\mathcal{F}_t$-measurable.

Let $\{\mathcal{F}_t\}_{t\ge 0}$ be an increasing family of σ-algebras of subsets of Ω. Recall Definition 2.9 on adapted stochastic process, we see that a process $g(t,\omega) : [0,\infty) \times \Omega \to \mathbb{R}^n$ is called $\mathcal{F}_t$-adapted if for each $t \ge 0$ the function

$$\omega \to g(t,\omega)$$

is $\mathcal{F}_t$-measureable. Sometimes we write $g(t,\omega)$ as $g_t(\omega)$. In the lecture notes, X_t and $X(t)$ are interchangeable.

Example 4.2 *The process* $X(t,\omega) = (W_{t/2}(\omega))^3 + 7W_t(\omega)$ *is* $\mathcal{F}_t$*-adapated while* $Y(t,\omega) = \sqrt{\sin(W_{2t}(\omega)) + 5}$ *is not.*

We have mentioned near the end of Section 2.9 why we need a stochastic process to be $\mathcal{F}_t$-adapted. It is not surprising that Itô integral is based on this kind of stochastic processes (Shreve II Theorem 4.3.1, Oksendal Definition 3.1.4, and Appendix D of Duffie's "Dynamic Asset Pricing Theory"):

Definition 4.2 *Let $\mathcal{V} = \mathcal{V}(0, T)$ be the class of functions*

$$f(t, \omega) : [0, \infty) \times \Omega \to \mathbb{R}$$

(meaning a stochastic process) such that

i) $(t, \omega) \to f(t, \omega)$ *is* $\mathcal{B} \times \mathcal{F}$*-measurable, where* $\mathcal{B}$ *denotes the Borel* σ*-algebra on* $[0, \infty)$.

ii) $f(t, \omega)$ *is* $\mathcal{F}_t$*-adapted.*

iii) $\mathbb{E}[\int_0^T f(t, \omega)^2 dt] < \infty$.

For functions $f \in \mathcal{V}$, we can define its Itô integral

$$\mathcal{I}[f](\omega) = \int_0^T f(t, \omega) dW_t(\omega)$$

where W_t is 1-dimensional Brownian motion.

The idea is natural: First we define $\mathcal{I}[\phi]$ for simple class of functions ϕ (elementary functions). Then we show that each $f \in \mathcal{V}$ can be approximated by such ϕ_k's and we use this to define $\int_0^T f dW_t$ as the limit of $\int_0^T \phi_k dW_t$ for any $\phi_k \to f$.

Definition 4.3 *A function $\phi \in \mathcal{V}(0, T)$ is called* <u>*elementary*</u> *if it has the form*

$$\phi(t, \omega) = \sum_{j=0}^{N-1} e_j(\omega) \chi_{[t_j, t_{j+1})}(t) \tag{4.4}$$

where $0 = t_0 < t_1 < \cdots < t_N = T$ form a partition of $[0, T]$.

Remark: Note that since $\phi \in \mathcal{V}$, each function $e_j = \phi(t_j, \cdot)$ must be $\mathcal{F}_{t_j}$-measurable. This means that the value of e_j only depends on $\{W_s, s \le t_j\}$ and is hence independent of $W_{t_{j+1}} - W_{t_j}$ (recall (3.40)).

For elementary function ϕ in (4.4), <u>define</u>

$$\int_0^T \phi(t, \omega) dW_t(\omega) = \sum_{j=0}^{N-1} e_j(\omega) [W_{t_{j+1}} - W_{t_j}](\omega). \tag{4.5}$$

Now, we have the following important observation:

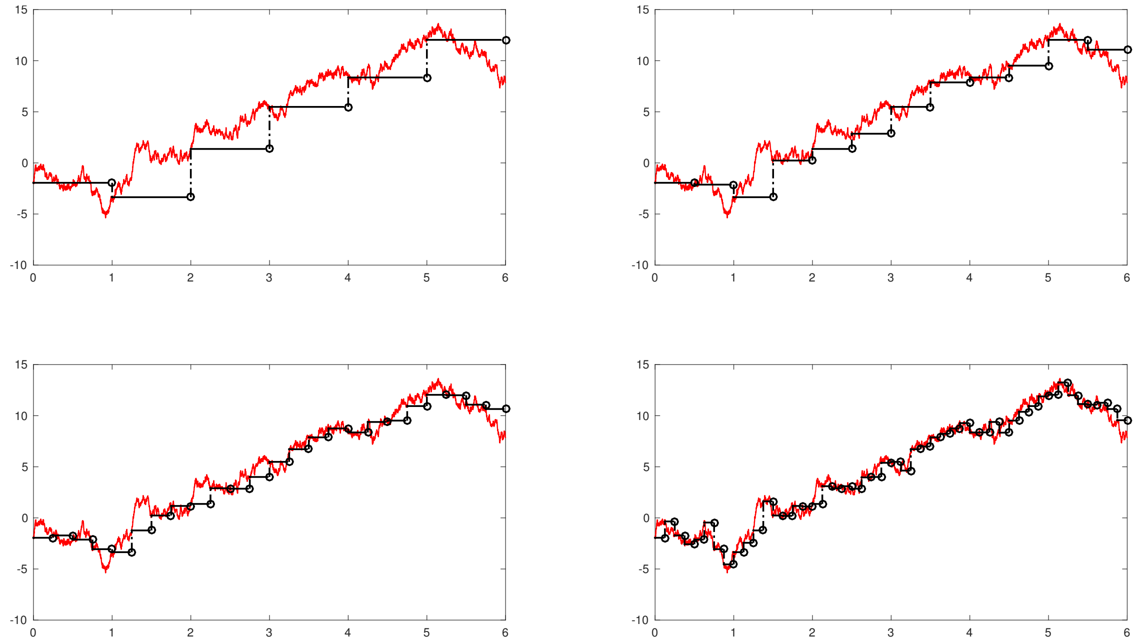


Figure 4.3: A function f (red curve) in $\mathcal{V}$ and its elementary function approximation ϕ_k (black curve) for a fixed ω. x-axis is for the t-variable.

Lemma 4.1 *(The Itô isometry. Theorem 4.2.2 of Shreve II, Lemma 3.1.5 of Oksendal) If $\phi(t,\omega) \in \mathcal{V}$ is bounded and elementary, then*

$$\mathbb{E}\left[\left(\int_0^T \phi(t,\omega)dW_t(\omega)\right)^2\right] = \mathbb{E}\left[\int_0^T \phi(t,\omega)^2 dt\right]. \tag{4.6}$$

Proof: Let $\Delta W_{t_j} = W_{t_{j+1}} - W_{t_j}$. Recall the remark after Defintion 4.3 and note two things: (i) $e_i e_j \Delta W_{t_i}$ and ΔW_{t_j} are independent if $i < j$; (ii) e_i and ΔW_{t_i} are independent. Hence

$$\mathbb{E}[e_i e_j \Delta W_{t_i} \Delta W_{t_j}] = \begin{cases} \mathbb{E}[e_i e_j \Delta W_{t_i}]\mathbb{E}[\Delta W_{t_j}] = 0 & \text{if } i \neq j, \text{ say, } i < j, \\ \mathbb{E}[e_j^2]\mathbb{E}[\Delta W_{t_j}^2] = \mathbb{E}[e_j^2](t_{j+1} - t_j) & \text{if } i = j. \end{cases}$$

Thus

$$\begin{aligned}
\mathbb{E}\left[\left(\int_0^T \phi dW_t\right)^2\right] &\overset{(4.5)}{=} \mathbb{E}\left[\left(\sum_i e_i \Delta W_{t_i}\right)\left(\sum_i e_j \Delta W_{t_j}\right)\right] \\
&= \sum_{i,j} \mathbb{E}[e_i e_j \Delta W_{t_i} \Delta W_{t_j}] = \sum_j \mathbb{E}[e_j^2](t_{j+1} - t_j) \\
&= \mathbb{E}[\sum_j e_j^2 (t_{j+1} - t_j)] \overset{(4.4)}{=} \mathbb{E}\left[\int_0^T \phi^2 dt\right]. \qquad \square
\end{aligned}$$

Following Oksendal, Page 27, or Duffie, Page 335, we have the following technical results whose proof is skipped for simplicity (won't be tested. But see Figure 4.3 for an illustration) [33]:

Lemma 4.2 *(Oksendal, Page 28 or Duffie, equation (D.1)) Let $f \in \mathcal{V}$. Then there exist elementary functions $\phi_n \in \mathcal{V}$ such that*

$$\mathbb{E}\left[\int_0^T (f-\phi_n)^2 dt\right] \to 0 \qquad \text{as } n \to \infty. \tag{4.7}$$

The following discussion up to Defintion 4.4 are rather technical. Feel free to skip all of them and they won't be tested. By inequality $\mathbb{E}[(X+Y)^2] \le 2\mathbb{E}[X^2] + 2\mathbb{E}[Y^2]$, we know that as long as $\{\phi_n\}$ satisfies $\mathbb{E}\left[\int_0^T (f-\phi_n)^2 dt\right] \to 0$,

$$\mathbb{E}\left[\int_0^T (\phi_m-\phi_n)^2 dt\right] \le 2\mathbb{E}\left[\int_0^T (\phi_m-f)^2 dt\right] + 2\mathbb{E}\left[\int_0^T (f-\phi_n)^2 dt\right] \to 0 \tag{4.8}$$

when $m, n \to \infty$.

By the isometry (4.6), we know the left hand side of (4.8) is

$$\mathbb{E}\left[\left(\int_0^T [\phi_m(t,\omega)-\phi_n(t,\omega)]dW_t(\omega)\right)^2\right], \tag{4.9}$$

which is also $\mathbb{E}[(Z_m - Z_n)^2]$ where $Z_n(\omega) \stackrel{\text{def}}{=} \int_0^T \phi_n(t,\omega)dW_t(\omega)$ is a random variable by (4.5). Hence $\lim_{m,n\to 0} \mathbb{E}[(Z_m-Z_n)^2] = 0$ and therefore $\{Z_n\}$ forms a Cauchy sequence in $L^2(\mathbb{P}) \stackrel{\text{def}}{=} \{X(\omega) : \int_\Omega X^2(\omega)d\mathbb{P}(\omega) \stackrel{(3.9)}{=} \mathbb{E}[X^2] < \infty\}$ where the distance between X and Y is defined to be $\sqrt{\mathbb{E}[(X-Y)^2]} \stackrel{\text{denoted as}}{=} \|X-Y\|$ [34].

One can learn from an advanced probability course that $L^2(\mathbb{P})$ is a complete space which means that as long as $\{Z_n : n = 1,2,3\cdots\}$ is a Cauchy sequence in $L^2(\mathbb{P})$ (meaning $\lim_{n,m\to 0}\|Z_n - Z_n\| = 0$), then there exists an $Z \in L(\mathbb{P})$ so that $\lim_{n\to\infty}\|Z_n - Z\| = 0$. Hence $\{Z_n\}$ has a limit in $L^2(\mathbb{P})$. This limit, which is in $L^2(\mathbb{P})$ (meaning its a random variable with finite mean and variance), is defined to be $\int_0^T f(t,\omega)dW_t(\omega)$:

[33] If you read Oksendal, you will see that his proof contains two preliminary steps before he can prove (4.7):

a) Let $g \in \mathcal{V}$ be bounded and $g(\cdot,\omega)$ be continuous (in t) for each ω. Then the elementary function $\phi_n = \sum_j g(t_j,\omega)\chi_{[t_j,t_{j+1})}(t)$ is in $\mathcal{V}$ and $\mathbb{E}\left[\int_0^T (g-\phi_n)^2 dt\right] \to 0$ as $n \to \infty$.

b) Let $h \in \mathcal{V}$ be bounded. Then there exist bounded functions $g_n \in \mathcal{V}$ such that $g_n(\cdot,\omega)$ is continuous (in t) for any ω and n, and $\mathbb{E}\left[\int_0^T (h-g_n)^2 dt\right] \to 0$ as $n \to \infty$.

[34] Recall that a sequence of real numbers $\{a_n\}$ is a Cauchy sequence if $\lim_{m,n\to\infty}|a_n - a_m| = 0$. It is a result of advanced calculus that a sequence of real numbers has a limit if and only of it is a Cauchy sequence. This result can be extended to sequences of random variables in $L^2(\mathbb{P})$ which, according to its definition, is the collection of all random variables having finite mean and finite variance since $\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

Definition 4.4 *(The Itô integral) Let $f \in \mathcal{V}(0,T)$. Then the Itô integral of f from 0 to T is defined by*

$$\int_0^T f(t,\omega)dW_t(\omega) = \lim_{n\to\infty} \int_0^T \phi_n(t,\omega)dW_t(\omega) \qquad (\textit{limit in the mean square sense}^{35}) \tag{4.10}$$

where $\{\phi_n\}$ is any[36] sequence of elementary functions such that

$$\mathbb{E}\left[\int_0^T (f-\phi_n)^2 dt\right] \to 0 \qquad \textit{as } n \to \infty. \tag{4.11}$$

Remark: It turns out, we can choose f from a larger class and the limit in (4.10) can be a limit in probability sense. See Section 3.3 of Oksendal or Appendix D of Duffie.

Lemma 4.1 immediately leads to

Corollary 4.1 *(The Itô isometry)*

$$\mathbb{E}\left[\left(\int_0^T f(t,\omega)dW_t(\omega)\right)^2\right] = \mathbb{E}\left[\int_0^T f(t,\omega)^2 dt\right] \qquad \textit{for all } f \in \mathcal{V}(0,T). \tag{4.12}$$

Proof: (This proof is not required for the exam. But the conclusion is.) We can simply let $n \to \infty$ in $\mathbb{E}\left[\left(\int_0^T \phi_n(t,\omega)dW_t(\omega)\right)^2\right] = \mathbb{E}\left[\int_0^T \phi_n(t,\omega)^2 dt\right]$ and notice that (4.10) implies[37] $\lim_{n\to\infty} \mathbb{E}\left[\left(\int_0^T \phi_n(t,\omega)dW_t(\omega)\right)^2\right] = \mathbb{E}\left[\left(\int_0^T f(t,\omega)dW_t(\omega)\right)^2\right]$ while (4.11) implies $\lim_{n\to\infty} \mathbb{E}\left[\int_0^T \phi_n(t,\omega)^2 dt\right] = \mathbb{E}\left[\int_0^T f(t,\omega)^2 dt\right]$.

Corollary 4.2 *For any $f, g \in \mathcal{V}(0,T)$,*

$$\mathbb{E}\left[\left(\int_0^T f(t,\omega)dW_t(\omega)\right)\left(\int_0^T g(t,\omega)dW_t(\omega)\right)\right] = \mathbb{E}\left[\int_0^T f(t,\omega)g(t,\omega)dt\right]. \tag{4.13}$$

Proof: (This proof is not required for the exam. But the conclusion is.) Simply consider the identity $\mathbb{E}\left[\left(\int_0^T (f(t,\omega)+g(t,\omega))dW_t(\omega)\right)^2\right] \overset{(4.12)}{=} \mathbb{E}\left[\int_0^T (f(t,\omega)+g(t,\omega))^2 dt\right]$ and then expand the square terms from both sides. □

[35] $X_n \to X$ in the mean square sense if and only if $\lim_{n\to\infty} \mathbb{E}[(X_n - X)^2] = 0$.

[36] If you choose another sequence, say, $\tilde{\phi}_n$, that also satisfies (4.11), then we can form a new sequence $\{\int_0^T \phi_1 dW_t, \int_0^T \tilde{\phi}_1 dW_t, \int_0^T \phi_2 dW_t, \int_0^T \tilde{\phi}_2 dW_t, \cdots\}$ which is again a Cauchy sequence in $L^2(\mathbb{P})$ because of (4.8) and (4.9). Since subsequences of a Cauchy sequence have the same limit, $\lim_{n\to\infty} \int_0^T \phi_n(t,\omega)dW_t(\omega) = \lim_{n\to\infty} \int_0^T \tilde{\phi}_n(t,\omega)dW_t(\omega)$.

[37] Let X be in $L^2(\mathbb{P})$. Suppose $\lim_{n\to\infty} \mathbb{E}[(X_n - X)^2] = 0$, then $\mathbb{E}[X_n^2] \le 2\mathbb{E}[(X_n - X)^2] + 2\mathbb{E}[X^2] < C_1$ for some constant C_1. $\mathbb{E}[(X_n)^2] - \mathbb{E}[X^2] = \mathbb{E}[(X_n - X)(X_n + X)] \le \sqrt{\mathbb{E}[(X_n - X)^2]}\sqrt{\mathbb{E}[(X_n + X)^2]} \le C_2\sqrt{\mathbb{E}[(X_n - X)^2]} \to 0$ for some constant C_2. So, $\lim_{n\to\infty} \mathbb{E}[(X_n)^2] = \mathbb{E}[X^2]$.

Example 4.3 *Assume $W_0 = 0$. Then*

$$\int_0^t W_s dW_s = \frac{1}{2}W_t^2 - \frac{1}{2}t. \tag{4.14}$$

Proof: *Put $\phi_n(s,\omega) = \sum_{j=0}^{N-1} W_{t_j}(\omega)\chi_{[t_j,t_{j+1})}(s)$ where $t_j = j\delta t$, $\delta t = t/N$. Then*

$$\mathbb{E}\left[\int_0^t (\phi_n - W_s)^2 ds\right] = \mathbb{E}\left[\sum_j \int_{t_j}^{t_{j+1}} (W_{t_j} - W_s)^2 ds\right]$$
$$= \sum_j \int_{t_j}^{t_{j+1}} (s - t_j) ds = \sum_j \frac{1}{2}(t_{j+1} - t_j)^2 = \frac{1}{2}T\delta t \to 0$$

as $\delta t \to 0$. So, by definition,

$$\int_0^t W_s dW_s = \lim_{\delta t \to 0} \int_0^t \phi_n dW_s = \lim_{\delta t \to 0} \sum_j W_{t_j}(W_{t_{j+1}} - W_{t_j})$$
$$= \lim_{\delta t \to 0} \sum_j \frac{1}{2}(W_{t_{j+1}}^2 - W_{t_j}^2) - \frac{1}{2}(W_{t_{j+1}} - W_{t_j})^2 = \frac{1}{2}W_t^2 - \frac{1}{2}t.$$

In the last step, we have used Question 6 of Homework III. One should compare (4.14) with Question 2.11 of Homework II.

Remark: If f and g are differentiable, we have the product rule $(fg)' = f'g + g'f$. We can then derive the integration by parts formula using the fundamental theorem of calculus:

$$f(t)g(t) - f(0)g(0) = \int_0^t (fg)' ds = \int_0^t f'g ds + \int_0^t fg' ds = \int_0^t g df + \int_0^t f dg.$$

Hence

$$\int_0^t f df = \frac{1}{2}\left(f^2(t) - f^2(0)\right).$$

However, because W_t is not differentiable on t, we cannot apply the above formula to conclude

$$\int_0^t W_s dW_s = \frac{1}{2}W_t^2 - \frac{1}{2}W_0^2 = \frac{1}{2}W_t^2.$$

Indeed, there is an extra $-\frac{1}{2}t$ term on the right hand side.

Example 4.4 *We are ready to verify the Itô isometry $\mathbb{E}\left[\left(\int_0^T f(t,\omega)dW_t(\omega)\right)^2\right] = \mathbb{E}\left[\int_0^T f(t,\omega)^2 dt\right]$ for $f(t,\omega) = W_t(\omega)$:*

$$
\begin{aligned}
LHS &= \mathbb{E}\left[\left(\int_0^T W_t(\omega)dW_t(\omega)\right)^2\right] \overset{(4.14)}{=} \mathbb{E}\left[\left(\frac{W_T^2}{2} - \frac{T}{2}\right)^2\right] \\
&= \frac{1}{4}\left(\mathbb{E}[W_T^4] - 2T\mathbb{E}[W_T^2] + T^2\right) \overset{Q2\ of\ HW\text{-}III}{=} \frac{3T^2 - 2T^2 + T^2}{4} = \frac{T^2}{2}. \\
RHS &= \mathbb{E}\left[\int_0^T (W_t(\omega))^2 dt\right] = \int_0^T \mathbb{E}\left[(W_t(\omega))^2\right] dt = \int_0^T t dt = \frac{T^2}{2}.
\end{aligned}
$$

4.1 Properties of the Itô integral

The following properties can be proved easily for elementary functions. By taking limits, we see that they are true for all $f, g \in \mathcal{V}(0, T)$:

Theorem 4.1 *(Theorem 3.2.1 of Oksendal) Let $f, g \in \mathcal{V}(0,T)$ and let $0 < U < T$. Then*

i) $\int_0^T f dW_t = \int_0^U f dW_t + \int_U^T f dW_t$ *for a.a.* ω.

ii) $\int_0^T (c_1 f + c_2 g) dW_t = c_1 \int_0^T f dW_t + c_2 \int_0^T g dW_t$ *for a.a.* ω.

iii)

$$
\mathbb{E}\left[\int_0^T f dW_t\right] = 0. \tag{4.15}
$$

iv) $\int_0^T f dW_t$ *is* $\mathcal{F}_T$ *measurable.*

Here is the proof of (4.15) (the proof of (4.15) won't be tested): It is a one-line-proof: (4.15) holds for f being elementary functions. When you pass to the limit, it remains true for $f \in \mathcal{V}(0,T)$.

Here are the details: We know that for elementary function ϕ, we have $\phi(t,\omega) = \sum_{j=0}^{N-1} e_j(\omega)\chi_{[t_j, t_{j+1})}(t)$ with e_j only depends on $\{W_s, s \le t_j\}$. Hence

$$
e_j \text{ and } W_{t_{j+1}} - W_{t_j} \text{ are independent.} \tag{4.16}
$$

Recall that the Itô integral $\int_0^T \phi_n(t,\omega)dW_t(\omega)$ is defined to be $\sum_{j=0}^{N-1} e_j(\omega)(W_{t_{j+1}} - W_{t_j})$. So

$$
\begin{aligned}
\mathbb{E}\int_0^T \phi_n(t,\omega)dW_t(\omega) &= \mathbb{E}\sum_{j=0}^{N-1} e_j(\omega)(W_{t_{j+1}} - W_{t_j}) = \sum_{j=0}^{N-1} \mathbb{E}\left(e_j(\omega)(W_{t_{j+1}} - W_{t_j})\right) \\
&\overset{(4.16)}{=} \sum_{j=0}^{N-1} \mathbb{E}\left(e_j(\omega)\right)\mathbb{E}\left((W_{t_{j+1}} - W_{t_j})\right) = \sum_{j=0}^{N-1} \mathbb{E}\left(e_j(\omega)\right) \times 0 = 0.
\end{aligned} \tag{4.17}
$$

Now, by definition (4.10), when $f \in \mathcal{V}(0,T)$,

$$\int_0^T f(t,\omega)dW_t(\omega) = \lim_{n\to\infty}\int_0^T \phi_n(t,\omega)dW_t(\omega), \tag{4.18}$$

for any elementary function sequence $\{\phi_n\}$ that satisfies (4.11). Taking expectation on both sides of (4.18), and switching the order of $\mathbb{E}$ and $\lim_{n\to\infty}$ [38], we get

$$\mathbb{E}\left(\int_0^T f(t,\omega)dW_t(\omega)\right) = \lim_{n\to\infty}\left(\mathbb{E}\int_0^T \phi_n(t,\omega)dW_t(\omega)\right) \overset{(4.17)}{=} 0. \tag{4.19}$$

This finishes the proof of (4.15).

Example 4.5 *Since a constant function* 1 *on* $[0,T)$ *is automatically an elementary function of the form* $\phi = 1\chi_{[0,T)}(t)$ *where* $\chi_{[0,T)}(t) = \begin{cases} 1 & t\in[0,T) \\ 0 & otherwise \end{cases}$, *by definition (4.5),*

$$\int_0^T dW_t = \int_0^T 1\chi_{[0,T)}dW_t = W_T - W_0.$$

To see that this definition makes sense, show by Itô isometry and (4.15) that $\int_0^T dW_t$ *and* $W_T - W_0$ *have the same mean and variance.*

Solution: By (4.15), $\mathbb{E}\int_0^T dW_t = 0$. By Itô isometry,

$$\mathrm{Var}\left[\int_0^T dW_t\right] = \mathbb{E}\left[\left(\int_0^T dW_t - 0\right)^2\right] = \int_0^T 1^2 dt = T.$$

At the same time, $\mathbb{E}[W_T - W_0] = 0$, $\mathrm{Var}[W_T - W_0] = T$.

Example 4.6 *Find the probability density of* $Z = \exp(\int_0^T t dW_t)$.
Solution: *Let* $X = \int_0^T t dW_t = \lim_{N\to\infty}\sum_{i=0}^{N-1} t_i(W_{t_{i+1}} - W_{t_i})$. *Since the sum of jointly normally distributed random variables are still normally distributed* [39], *and since the limit of normally distributed random variables, if exists, is still normally distributed* [40], *X has normal distribution.* $\mathbb{E}X = 0$, $Var[X] = \mathbb{E}[X^2] = \int_0^T t^2 dt = \frac{T^3}{3}$. $X \sim N(0, \frac{T^3}{3})$. *For* $a \geq 0$,

$$\mathbb{P}(Z \leq z) = \mathbb{P}(e^X \leq z) = \mathbb{P}(X \leq \log z) = \int_{-\infty}^{\log z} \frac{1}{\sqrt{2\pi}(T^3/3)^{1/2}} e^{-\frac{x^2}{2T^3/3}}dx.$$

[38] If you worry why we can switch the order, here is the explanation: In the lecture, I have explained to you that the limit in (4.18) is in the mean square sense, which means that if $Z_n(\omega) \overset{\text{def}}{=} \int_0^T \phi_n(t,\omega)dW_t(\omega)$ and $Z \overset{\text{def}}{=} \int_0^T f(t,\omega)dW_t(\omega)$, then $\mathbb{E}[|Z_n - Z|^2] \to 0$. Since $(\mathbb{E}[|X|])^2 \leq \mathbb{E}[X^2]$, we conclude $|\mathbb{E}Z_n - \mathbb{E}Z| \leq \sqrt{\mathbb{E}[|Z_n - Z|^2]} \to 0$. Hence $\lim_{n\to\infty}\mathbb{E}[Z_n] = \mathbb{E}[Z]$ which is the first equation in (4.19).

[39] See the discussion after (3.5)

[40] The proof may need to use characteristic function (a little bit like the moment generating function you have seen in Question 1 of Homework III) and is not required for this course.

We know that the pdf is the derivative of cdf, hence

$$\begin{aligned}\rho_Z(z) =&\frac{d}{dz}\mathbb{P}(Z\le z)=\frac{d}{dz}\left(\int_{-\infty}^{\log z}\frac{1}{\sqrt{2\pi}(T^3/3)^{1/2}}e^{-\frac{x^2}{2T^3/3}}dx\right)\\ =&\frac{1}{\sqrt{2\pi}(T^3/3)^{1/2}}e^{-\frac{(\log z)^2}{2T^3/3}}\frac{d}{dz}\log z=\frac{\sqrt{3}}{z\sqrt{2\pi}T^{\frac{3}{2}}}e^{-\frac{3(\log z)^2}{2T^3}}.\end{aligned}$$

Example 4.7 *Consider $Z=\int_0^T e^{-rt}dW_t$ where T and r are positive constants. Find the pdf of Z.*

Solution: By the same observation as in Example 4.6, we know Z has normal distribution. By (4.15), $\mathbb{E}Z=0$. By Itô isometry,

$$\mathrm{Var}[Z]=\mathbb{E}[(Z-0)^2]=\mathbb{E}\left[\left(\int_0^T e^{-rt}dW_t\right)^2\right]=\mathbb{E}\int_0^T\left(e^{-rt}\right)^2dt=\int_0^T e^{-2rt}dt=\left.\frac{e^{-2rt}}{-2r}\right|_{t=0}^{t=T}=\frac{1-e^{-2rT}}{2r}$$

Hence $Z\sim N(0,\frac{1-e^{-2rT}}{2r})$ and its pdf is

$$\frac{1}{\sqrt{2\pi}\sigma}e^{-z^2/(2\sigma^2)}$$

with $\sigma=\sqrt{\frac{1-e^{-2rT}}{2r}}$.

Example 4.8 *Consider $Z=\int_0^T e^{\sigma W_t}dW_t$, where σ and T are positive constants and we require $W_0=0$ for simplicity. Find the mean and variance of Z. [Hint: In Question 1 of Homework III, you have proved/will prove that if $X\sim N(\mu,\sigma^2)$, $\mathbb{E}[e^X]=e^{\mu+\frac{1}{2}\sigma^2}$.] (If similar problem appears in the exam, similar hint will be given.)*

Solution: By (4.15), $\mathbb{E}Z=0$. $X=2\sigma W_t\sim N(0,4\sigma^2 t)$

$$\begin{aligned}\mathrm{Var}[Z]=&\mathbb{E}[(Z-0)^2]=\mathbb{E}\left[\left(\int_0^T e^{\sigma W_t}dW_t\right)^2\right]=\mathbb{E}\int_0^T\left(e^{\sigma W_t}\right)^2dt\\ =&\int_0^T\mathbb{E}\left(e^{2\sigma W_t}\right)dt=\int_0^T e^{2\sigma^2 t}dt=\frac{e^{2\sigma^2 T}-1}{2\sigma^2}.\end{aligned}$$

Remark: Please note that unlike Example 4.5 and Example 4.7,

$$Z=\int_0^T e^{\sigma W_t}dW_t\overset{(4.2)}{=}\lim_{N\to\infty}\sum_{i=1}^N e^{\sigma W_{t_{i-1}}}(W_{t_i}-W_{t_{i-1}})$$

(with $\delta t=T/N$, $t_i=i\delta t$) is not the limit of summation of jointly normally distributed random variables (unless we change $e^{\sigma W_{t_{i-1}}}$ to some deterministic function). Hence Z in

Example 4.8 may not be normally distributed. Knowing its mean and variance may not be enough to determine its pdf.

We state without prove the following technical result. (Don't worry about those technical points in your first study. It just says that $\int_0^t f(s,\omega)dW_s$ is continuous in t, which means that if you chance t slightly, the random variable $\int_0^t f(s,\omega)dW_s$ also changes slightly.)

Theorem 4.2 *(Theorem 3.2.5 of Oksendal) Let $f \in \mathcal{V}(0,T)$. Then there exists a t-continuous version of*

$$\int_0^t f(s,\omega)dW_s(\omega), \qquad 0 \le t \le T,$$

i.e., there exists a t-continuous stochastic process J_t on $(\Omega, \mathcal{F}, \mathbb{P})$ such that

$$\mathbb{P}\left[J_t = \int_0^t f dB\right] = 1 \qquad \forall t \in [0,T].$$

One of the most important property of Itô integral is that it is a (cotinuous-time) Martingale.

So, recall what is a discrete-time martingale. For finite probability space $\Omega = \{$ sequence of coin tosses $\}$, a discrete-time stochastic process M_n is called a martingale if $M_n = \tilde{\mathbb{E}}_n[M_{n+1}]$ (Definition 2.4). See Defintion 2.3 for the definition of conditional expectation and Section 2.9 for $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2\, \mathcal{F}_3$. In our later discussion, we can denote $\tilde{\mathbb{E}}_n[X]$ as $\tilde{\mathbb{E}}[X|\mathcal{F}_n]$ (we have seen it in (3.56)). Please work out Question 5 of Homework IV and read its footnote.

For more general Ω, we have to generalize conditional expectation (in particular, its averaging propeorty) to the following. To gain a better understanding, you should read or work out Questions 8 and 9 of Homework III.

Definition 4.5 *(Defintion 2.3.1 of Shreve II) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$ [41] and let X be a random variable that is either nonnegative or have finite mean. The conditional expectation of X given $\mathcal{G}$, denoted $\mathbb{E}[X|\mathcal{G}]$, is any random variable that satisfies*

i) (measurability) $\mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable (see Definition 2.8),

ii) (partial averaging property)

$$\int_A \mathbb{E}[X|\mathcal{G}](\omega)d\mathbb{P}(\omega) = \int_A X(\omega)d\mathbb{P}(\omega) \quad \text{for all } A \in \mathcal{G}. \tag{4.20}$$

In this more general setting, conditional expectation still have the 5 properties stated in Theorem 2.2.

[41] meaning $\mathcal{G}$ is a σ-algebra and $\mathcal{G} \subset \mathcal{F}$.

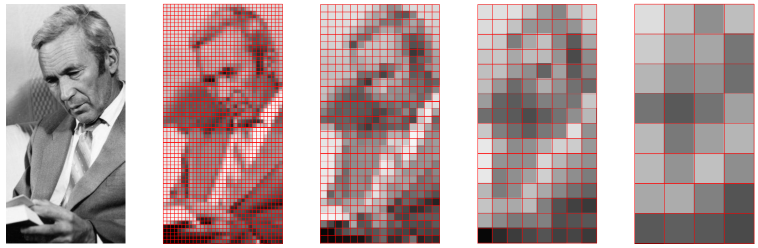


Figure 4.4: $X \leftarrow \mathbb{E}[X|\mathcal{F}_4] \leftarrow \mathbb{E}[X|\mathcal{F}_3] \leftarrow \mathbb{E}[X|\mathcal{F}_2] \leftarrow \mathbb{E}[X|\mathcal{F}_1]$.

Theorem 4.3 *(Theorem 2.3.2 of Shreve II) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Assume X, Y are random variables with finite mean.*

i) (Linearity of conditional expectations) For all constants c_1 and c_2, we have

$$\mathbb{E}[c_1X + c_2Y|\mathcal{G}] = c_1\mathbb{E}[X|\mathcal{G}] + c_2\mathbb{E}[Y|\mathcal{G}]. \tag{4.21}$$

ii) (Taking out what is known) If Z is $\mathcal{G}$-measurable, then

$$\mathbb{E}[ZX|\mathcal{G}] = Z\mathbb{E}[X|\mathcal{G}]. \tag{4.22}$$

iii) (Iterated conditioning) If $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$ (which means $\mathcal{H}$ is coarser and contains less information than $\mathcal{G}$), then

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]. \tag{4.23}$$

In particular (see also (3.55)),

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]. \tag{4.24}$$

iv) (Independence) If X is independent of $\mathcal{G}$, then (see also (3.54))

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}X. \tag{4.25}$$

v) (Conditional Jensen's inequality) If $\varphi(x)$ is a convex function of the dummy variable x (e.g. $\varphi(x) = x^2$), then

$$\mathbb{E}[\varphi(X)|\mathcal{G}] \geq \varphi(\mathbb{E}[X|\mathcal{G}]). \tag{4.26}$$

Definition 4.6 *(Page 312 of Oksendal, Page 74 of Shreve II, Page 332 of Duffie) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\{\mathcal{F}_t : t \geq 0\}$ be an increasing family of σ-algebra (meaning $\mathcal{F}_t \subset \mathcal{F}_s$ when $t < s$). A stochastic process $\{M_t(\omega)\}$ is a martingale with respect to $\mathcal{F}_t$ if M_t is $\mathcal{F}_t$-adapted*[42]*, $\mathbb{E}[|M_t|] < \infty$ for all t, and*

$$\mathbb{E}[M_s|\mathcal{F}_t] = M_t \qquad \text{for all } s \geq t. \tag{4.27}$$

[42]Definition 2.9.

Theorem 4.4 *(Corollary 3.2.6 of Oksendal) Let $f(t,\omega) \in \mathcal{V}(0,T)$ for all T. Then*

$$M_t(\omega) = \int_0^t f(s,\omega)dW_s \tag{4.28}$$

is a martingale with respect to $\mathcal{F}_t$[43]*. Furthermore,*

$$\mathbb{P}\left[\sup_{0\le t\le T} |M_t| \ge \lambda\right] \le \frac{1}{\lambda^2}\mathbb{E}\left[\int_0^T f(s,\omega)^2 ds\right], \qquad \textit{for any } \lambda, T > 0. \tag{4.29}$$

Remark: In $\mathbb{E}[M_s|\mathcal{F}_t]$, we are given the value of W_τ for $\tau \le t$. $f(\tau,\omega)$ is $\mathcal{F}_\tau$ measurable means that the value of $f(\tau,\omega)$ depends on $\{W_u, u \le \tau\}$ only. So, we know the value of $\int_0^t f(\tau,\omega)dW_\tau(\omega)$. Then $\mathbb{E}[M_s|\mathcal{F}_t]$ asks, if we have already known that much of W_τ, what the average value of M_s is, after considering all possible values of the "tail" $\{W_v, v > t\}$. So, the conditional expectation $\mathbb{E}[M_s|\mathcal{F}_t]$ will average out all the "tail" part of the Brownian motion $\{W_v, v > t\}$. This is precisely (2.28).
Remark: You may compare (4.29) with Chebyshev inequality

$$\mathbb{P}\left[|X - \mathbb{E}X| \ge \lambda\right] \le \frac{1}{\lambda^2}\mathbb{E}[(X-\mathbb{E}X)^2], \qquad \text{for any } \lambda > 0 \tag{4.30}$$

which immediately implies

$$\mathbb{P}\left[|M_t| \ge \lambda\right] \le \frac{1}{\lambda^2}\mathbb{E}\left[\int_0^t f(s,\omega)^2 ds\right], \qquad \text{for any } \lambda > 0. \tag{4.31}$$

Hence (4.29) says more than what Chebyshev tells us.

Example 4.9 *(Section 3.5 of Shreve II) Let $W(t)$ be a Brownian motion and let $\mathcal{F}(t)$, $t \ge 0$, be an associated filtration (Definition 4.1). For $\alpha \in \mathbb{R}$, consider the Brownian motion with drift α:*

$$X(t) = \alpha t + W(t). \tag{4.32}$$

Show that for any function $f(y)$, and for any $0 \le s < t$, the function

$$g(x) = \frac{1}{\sqrt{2\pi(t-s)}}\int_{-\infty}^{\infty} f(y)e^{-\frac{(y-\alpha(t-s)-x)^2}{2(t-s)}}dy$$

satisfies

$$\mathbb{E}[f(X(t))|\mathcal{F}(s)] = g(X(s)). \tag{4.33}$$

[43]This is because $M_s = M_t + \int_t^s f(\tau,\omega)dW_\tau$ and $\mathbb{E}[\int_t^s f(\tau,\omega)dW_\tau|\mathcal{F}_t]$=0.

Proof: *Let $a = \alpha t + W(s)$. a is $\mathcal{F}(s)$-measurable, meaning that a can be treated as given when evaluating $\mathbb{E}[\cdot|\mathcal{F}(s)]$. By (3.40), $W(t) - W(s) \sim N(0, t-s)$ with pdf $\frac{1}{\sqrt{2\pi(t-s)}}e^{-\frac{x^2}{2(t-s)}}$ and $W(t) - W(s)$ is independent of $\mathcal{F}(s)$.*

$$\begin{aligned}
\mathbb{E}[f(X(t))|\mathcal{F}(s)] =& \mathbb{E}[f(W(t) - W(s) + a)|\mathcal{F}(s)] \\
\overset{(3.10)}{=} & \left.\int_{-\infty}^{\infty} f(x+a)\frac{1}{\sqrt{2\pi(t-s)}}e^{-\frac{x^2}{2(t-s)}}dx\right|_{a=W(s)+\alpha t \text{ is given}} \qquad (4.34)\\
\overset{y=x+a}{=} & \int_{-\infty}^{\infty} f(y)\frac{1}{\sqrt{2\pi(t-s)}}e^{-\frac{(y-a)^2}{2(t-s)}}dy \\
=& g(a - \alpha(t-s)) = g(\alpha s + W(s)).
\end{aligned}$$

Remark: (4.33) implies $X(t)$ is a Markov process according to the definition in Question 7 of Homework III, or more generally, Definition 2.3.6 Shreve II. In particular, we know Brownian motion (by setting $\alpha = 0$) is a Markov process.

Take $\alpha = 0$ for simplicity and introduce transition density

$$p(\tau, x, y) = \frac{1}{\sqrt{2\pi\tau}}e^{-\frac{(y-x)^2}{2\tau}}. \qquad (4.35)$$

(We have seen it before in (3.28). So, the integrand in (3.29) is the production of transition densities.) Then

$$g(x) = \int_{-\infty}^{\infty} f(y)p(t-s, x, y)dy \qquad (4.36)$$

and (4.33) can be rewritten as (with $t > s$)

$$\mathbb{E}[f(W(t))|\mathcal{F}(s)] = \int_{-\infty}^{\infty} f(y)p(t-s, W(s), y)dy. \qquad (4.37)$$

This equation has the following interpretation. Conditioned on the information in $\mathcal{F}(s)$ (which contains all the information obtained by observing the Brownian motion up to and including time s), the conditional density of $W(t)$ is $p(t-s, W(s), y)$. This is a density in the variable y. This density is normal with mean $W(s)$ and variance $t-s$. In particular, the only information from $\mathcal{F}(s)$ that is relevant is the value of $W(s)$. The fact that only $W(s)$ is relevant is the essence of the Markov property. See Question 7 of Homework III for the finite probability space case.

4.2 Homework IV

(Only submit solutions to Questions 5,6,9,10,17.)

1. (The weak law of large number) Suppose $\{Y_i\}$ are i.i.d. (independent identically distributed) random variables with $\mathbb{E}Y_i = a$ and $\text{Var}Y_i = b < \infty$. Show that $\frac{1}{N}\sum_{i=1}^N Y_i \to a$ (which is equivalent to $\frac{1}{N}\sum_{i=1}^N (Y_i - a) \to 0$) in the mean square sense as $N \to \infty$.

   Proof: $\mathbb{E}[(\frac{1}{N}\sum_{i=1}^N Y_i - a)^2] = \frac{1}{N^2}\mathbb{E}[(\sum_{i=1}^N Y_i - Na)^2] = \frac{1}{N^2}\sum_{i=1}^N \mathbb{E}[(Y_i - a)^2] = \frac{b}{N} \to 0$ as $N \to \infty$.

2. (Another proof of Question 6 of Homework III) Let $\{W_s : 0 \le s \le T\}$ be a 1-dimensional Brownian motion with $W_{t_0} = W_0 = 0$. Given $t > 0$ and $N \in \mathbb{Z}_+$, let $\delta t = t/N$ and $t_j = j\delta t$. Use the weak law of large number to prove that

$$\lim_{\delta t \to 0} \sum_{j=0}^{N-1} (W_{t_{j+1}} - W_{t_j})^2 = t \tag{4.38}$$

   in the mean square sense. [44]

   Proof: $\mathbb{E}\left(\frac{(W_{t_{j+1}} - W_{t_j})^2}{\delta t}\right) = 1$.

$$\lim_{\delta t \to 0} \sum_j (W_{t_{j+1}} - W_{t_j})^2 = t \lim_{N\to\infty} \frac{\sum_{j=0}^{N-1} \frac{(W_{t_{j+1}} - W_{t_j})^2}{\delta t}}{N} = t \times 1 = 1. \tag{4.39}$$

3. Use the same notation as in Question 2. Show that W_t has unbounded first variation, which means

$$\lim_{N\to\infty} \sum_{j=0}^{N-1} |W_{t_{j+1}} - W_{t_j}| = \infty. \tag{4.40}$$

   Proof: Since

$$\sum_{j=1}^{N-1} (W_{t_{j+1}} - W_{t_j})^2 \le \left(\max_{0\le k \le N-1} |W_{t_{k+1}} - W_{t_k}|\right)\left(\sum_{j=1}^{N-1} |W_{t_{j+1}} - W_{t_j}|\right),$$

   we know

$$\infty > \sum_{j=1}^{N-1} |W_{t_{j+1}} - W_{t_j}| \ge \frac{\sum_{j=1}^{N-1} (W_{t_{j+1}} - W_{t_j})^2}{\max_{0\le k\le N-1} |W_{t_{k+1}} - W_{t_k}|}. \tag{4.41}$$

[44] FYI: Let Y_i, $i = 1, 2, \cdots$, be a sequence of independent and identically distributed random variables with mean $E[Y_i] = \mu$. The strong law of large number says that $\frac{\sum_{i=1}^N (Y_i - \mu)}{N} \to 0$ with probability 1. It means that for almost all ω, $\frac{\sum_{i=1}^N (Y_i(\omega) - \mu)}{N} \to 0$. So, if we use the strong law of large number, we can conclude that (4.38) happens with probability 1.

As W_t is continuous, $\lim_{N\to\infty} \max_{0\le k\le N-1} |W_{t_{k+1}} - W_{t_k}| = 0$. Using (4.38) and letting $N \to \infty$ in (4.41), we obtain

$$\lim_{N\to\infty} \sum_{j=0}^{N-1} |W_{t_{j+1}} - W_{t_j}| = \infty.$$

4. Let Y_i, $i = 1, 2, \cdots$, be a sequence of independent and identically distributed random variables with mean $E[Y_i] = \mu$ and $\mathrm{Var}[Y_i] = \sigma^2$. The central limit theorem says that

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n}(Y_i - \mu)}{\sqrt{n}} \to Z \sim N(0, \sigma^2) \tag{4.42}$$

in the sense of distribution. It means that as n increases, the distribution of the random variable $\frac{\sum_{i=1}^{n}(Y_i-\mu)}{\sqrt{n}}$ becomes closer and closer to that of $N(0, \sigma^2)$ random variable Z.

Now, consider the symmetric random walk (defined in Example 2.11)

$$M_k = \sum_{i=1}^{k} Z_i$$

where $M_0 = 0$, $\mathbb{P}(Z_i = 1) = \mathbb{P}(Z_i = -1) = \frac{1}{2}$. Fix t, define $W_t^{(n)} = \sqrt{\frac{t}{n}} M_n$. Show that

$$\lim_{n\to\infty} W_t^{(n)} \to Z \sim N(0, t) \tag{4.43}$$

in the sense of distribution.

Solution: $\mathbb{E}[Z_i] = 1 \times \frac{1}{2} + (-1) \times \frac{1}{2} = 0$. $\mathrm{Var}[Z_i] = \mathbb{E}[Z_0^2] - (\mathbb{E}[Z_i])^2 = 1$. By the central limit theorem

$$\frac{\sum_{i=1}^{k}(Z_i - \mathbb{E}[Z_i])}{\sqrt{n}} \to Z \sim N(0, 1)$$

in the sense of distribution. Hence

$$W_t^{(n)} = \sqrt{t}\, \frac{\sum_{i=1}^{k}(Z_i - \mathbb{E}[Z_i])}{\sqrt{n}} \to \sqrt{t} Z \sim N(0, t)$$

in the sense of distribution.

5. Recall (2.28)

$$\tilde{\mathbb{E}}_n[X](\omega_1 \cdots \omega_n) = \sum_{\omega_{n+1}\cdots\omega_N} q_u^{\#H(\omega_{n+1}\cdots\omega_N)} q_d^{\#T(\omega_{n+1}\cdots\omega_N)} X(\omega_1 \cdots \omega_n \omega_{n+1} \cdots \omega_N)$$

and the S_3 defined in Example 2.4. **Show** that

$$\tilde{\mathbb{E}}_2[S_3](HH)\mathbb{P}(A_{HH}) = \sum_{\omega \in A_{HH}} S_3(\omega)\mathbb{P}(\omega), \tag{4.44}$$

where $A_{HH} = \{HHH, HHT\}$ is defined in (2.40). Since $\tilde{\mathbb{E}}_2[S_3](\omega)$ does not change value on A_{HH}, (4.44) can be rewritten as

$$\int_{A_{HH}} \tilde{\mathbb{E}}_2[S_3](\omega)d\mathbb{P}(\omega) = \int_{A_{HH}} S_3(\omega)d\mathbb{P}(\omega). \tag{4.45}$$

Similarly, **prove**

$$\int_{A_{HT}} \tilde{\mathbb{E}}_2[S_3](\omega)d\mathbb{P}(\omega) = \int_{A_{HT}} S_3(\omega)d\mathbb{P}(\omega).\text{[45]} \tag{4.47}$$

6. Let W_t be 1-dimensional Brownian motion with $W_0 = 0$ and let $X_t = e^{W_t^2}$. Show that

$$\mathbb{E}[X_t^2] = \frac{1}{\sqrt{1-4t}}, \qquad t \in [0, 1/4).$$

7. Let $W(t)$, $t \geq 0$, be a Brownian motion, and let $\mathcal{F}(t)$, $t \geq 0$, be a filtration for this Brownian motion. Show that $W^2(t) - t$ is a martingale, i.e., $\mathbb{E}[W^2(t) - t|\mathcal{F}(s)] = W^2(s) - s$ or equivalently $\mathbb{E}[W^2(t) - W^2(s)|\mathcal{F}(s)] = t - s$ for $0 \leq s \leq t$.

   Proof:

$$\begin{aligned} &\mathbb{E}[W^2(t) - W^2(s)|\mathcal{F}(s)] \\ =&\mathbb{E}[(W(t) - W(s))^2 + 2W(t)W(s) - 2W^2(s)|\mathcal{F}(s)] \\ \overset{(4.22)}{=}&\mathbb{E}[(W(t) - W(s))^2|\mathcal{F}(s)] + 2W(s)\mathbb{E}[W(t) - W(s)|\mathcal{F}(s)] = t - s. \end{aligned}$$

8. Use the Itô isometry which is Corollary 4.1 and property (iii) of Theorem 4.1 to answer the following questions. Let W_t be 1-dimensional Brownian motion with $W_0 = 0$. Define

$$X = \int_0^T t dW_t, \qquad \text{and} \qquad Y = \int_0^T (T-t) dW_t.$$

   Determine $\mathbb{E}[X]$, $\mathrm{Var}[X]$, $\mathbb{E}[Y]$, $\mathrm{Var}[Y]$. Note that

$$X + Y = \int_0^T T dW_t = TW_T.$$

   Determine $\mathrm{Var}[X+Y]$ and then determine $\mathrm{Cov}(X, Y) \overset{\text{def}}{=} \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$.

   Solution: $\mathbb{E}X = \mathbb{E}Y = 0$. $\mathrm{Var}X = \mathbb{E}[X^2] = \mathbb{E}[\left(\int_0^T t dW_t\right)^2] = \mathbb{E}\int_0^T t^2 dt = \frac{T^3}{3}$.

[45] By the same method, one can prove that

$$\int_A \tilde{\mathbb{E}}_2[S_3](\omega)d\mathbb{P}(\omega) = \int_A S_3(\omega)d\mathbb{P}(\omega) \tag{4.47}$$

for $A \in \{A_{HH}, A_{HT}, A_{TH}, A_{TT}\}$ or more generally for any $A \in \mathcal{F}_2$ defined by (2.41). Hence if one recalls the standard definition of conditional expectation of $\tilde{\mathbb{E}}[S_3|\mathcal{F}_2]$ in (4.20), we have $\mathbb{E}_2[S_3] = \tilde{\mathbb{E}}[S_3|\mathcal{F}_2]$.

$\mathrm{Var}Y = \mathbb{E}[Y^2] = \mathbb{E}[\left(\int_0^T (T-t)dW_t\right)^2] = \mathbb{E}\int_0^T (T-t)^2 dt = \frac{T^3}{3}$.
$\mathrm{Var}[X+Y] = \mathbb{E}[(X+Y)^2] = T^2 T = T^3$.
Since $\mathbb{E}[(X+Y)^2] = \mathbb{E}[X^2+Y^2+2XY]$,

$$\mathbb{E}[XY] = \frac{1}{2}\left(\mathbb{E}[(X+Y)^2] - \mathbb{E}[X^2] - \mathbb{E}[Y^2]\right) = \frac{T^3}{6}.$$

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \frac{T^3}{6}.$$

9. For the M_t defined in (4.28), prove that

$$\mathbb{E}[M_t] = 0.$$

(Hint: Use (4.24) with $\mathcal{G} = \mathcal{F}_0$.) Then find the variance of

$$M_t = \int_0^t e^{-\alpha s} dW_s$$

for $\alpha > 0$.

10. Let

$$Y_t = \int_0^t \sqrt{|W_s|}dW_s,$$

where $|W_s|$ denotes the absolute value of W_s. Determine $\mathrm{Var}[Y]$.

11. Let W_t be one-dimensional Brownian motion, $\sigma \in \mathbb{R}$ be constant and $s \geq t \geq 0$. (1) Use Question 1 of Homework III to prove that

$$\mathbb{E}[e^{\sigma(W_s - W_t)}] = e^{\frac{1}{2}\sigma^2(s-t)}. \tag{4.48}$$

(2) Prove directly from the definition that

$$M_t = e^{\sigma W_t - \frac{1}{2}\sigma^2 t}, \qquad t \geq 0 \tag{4.49}$$

is a martingale with respect to $\mathcal{F}_t$ ((4.27)). Then use this result to prove that $\mathbb{E}[M_t] = 1$ for all $t \geq 0$ if $W_0 = 0$. (Hint: If $s \geq t$, then $\mathbb{E}[M_s|\mathcal{F}_t] =$
$\mathbb{E}[M_t e^{\sigma(W_s - W_t) - \frac{1}{2}\sigma^2(s-t)}|\mathcal{F}_t] \overset{(4.22)}{=} M_t e^{-\frac{1}{2}\sigma^2(s-t)}\mathbb{E}[e^{\sigma(W_s-W_t)}|\mathcal{F}_t]$.)

Proof: (1) $W_s - W_t \sim N(0, s-t)$. $\mathbb{E}[e^{\sigma(W_s-W_t)}] = e^{\frac{1}{2}\sigma^2(s-t)}$.
(2)

$$\begin{aligned}\mathbb{E}[M_s|\mathcal{F}_t] =& M_t e^{-\frac{1}{2}\sigma^2(s-t)}\mathbb{E}[e^{\sigma(W_s-W_t)}|\mathcal{F}_t] \\ \overset{(4.25),(3.12)}{=}& M_t e^{-\frac{1}{2}\sigma^2(s-t)}\mathbb{E}[e^{\sigma(W_s-W_t)}] \\ =& M_t e^{-\frac{1}{2}\sigma^2(s-t)} e^{\frac{1}{2}\sigma^2(s-t)} = M_t.\end{aligned}$$

$\mathbb{E}[M_t] \overset{(4.24)}{=} \mathbb{E}[\mathbb{E}[M_t|\mathcal{F}_0]] = \mathbb{E}[M_0] = \mathbb{E}[1] = 1.$

12. (Page 324 of "Dynamic Asset Pricing Theory", 3rd edition, by Duffie. An equivalent definition of conditional expectation. This is for your information only, in case you will read Duffie or other books later in your career. It won't be tested.) For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with Ω finite, if $\mathcal{G}$ is a sub-σ-algebra[46], then $\mathcal{G}$ represents in some sense "less information". The conditional expectation of X given a sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$ is defined as any $\mathcal{G}$-measurable random variable denoted by $\mathbb{E}[X|\mathcal{G}]$, satisfying the property that

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]Z] = \int_\Omega \mathbb{E}[X|\mathcal{G}](\omega)Z(\omega)d\mathbb{P}(\omega) = \int_\Omega X(\omega)Z(\omega)d\mathbb{P}(\omega) = \mathbb{E}[XZ] \tag{4.50}$$

for any $\mathcal{G}$-measurable random variable Z. Please compare it with (4.20) and show that (4.50) implies (4.20). Then use (4.22) to show that (4.20) implies (4.50).

Proof: "(4.50)⇒(4.20)": For any set $A \in \mathcal{G}$, define $I_A = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{otherwise.} \end{cases}$ Then 1_A is $\mathcal{G}$-measurable since $A \in \mathcal{G}$. So, we can let $Z = 1_A$ in (4.50). This leads to

$$\int_\Omega \mathbb{E}[X|\mathcal{G}](\omega)1_A(\omega)d\mathbb{P}(\omega) = \int_\Omega X(\omega)1_A(\omega)d\mathbb{P}(\omega)$$

or

$$\int_A \mathbb{E}[X|\mathcal{G}](\omega)d\mathbb{P}(\omega) = \int_A X(\omega)d\mathbb{P}(\omega)$$

which is precisely (4.20).

"(4.20)⇒(4.50)": (4.20) implies (4.22) $\mathbb{E}[ZX|\mathcal{G}] = Z\mathbb{E}[X|\mathcal{G}]$. Taking $\mathbb{E}$ on both sides, we obtain

$$\mathbb{E}[\mathbb{E}[ZX|\mathcal{G}]] = \mathbb{E}[Z\mathbb{E}[X|\mathcal{G}]].$$

Since the left hand side is $\mathbb{E}[ZX]$ by (4.24), we get (4.50).

13. (Continue with Question 12. Page 324 of "Dynamic Asset Pricing Theory", 3rd edition, by Duffie. This is for your information only. It won't be tested.) If Y is a nonnegative random variable with $\mathbb{E}Y = 1$, then we can create a new probability measure $\tilde{\mathbb{P}}$ from the old probability measure $\mathbb{P}$ by defining

$$\tilde{\mathbb{P}}(A) = \mathbb{E}(1_A Y) \tag{4.51}$$

for any event A, where $1_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{otherwise.} \end{cases}$ Note that (4.51) can also be written as $\int_A d\tilde{\mathbb{P}}(\omega) = \int_\Omega Y(\omega)1_A(\omega)d\mathbb{P}(\omega) = \int_A Y(\omega)d\mathbb{P}(\omega)$. So, we write $Y = \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}$ and call Y the Radon-Nikodym derivative of $\tilde{\mathbb{P}}$ with respect to $\mathbb{P}$. (4.51) can also be written as $\tilde{\mathbb{E}}(1_A) = \mathbb{E}(Y1_A)$ for any set $A \in \Omega$, where $\tilde{\mathbb{E}}$ denotes the expectation of under $\tilde{\mathbb{P}}$ and $\mathbb{E}$ denotes the expectation of under $\mathbb{P}$. With some standard mathematics/probability

[46] meaning $\mathcal{G}$ is a σ-algebra and is also a subset of $\mathcal{F}$.

technics which you do not need to know the details, the last equation implies that for any random variable X,

$$\tilde{\mathbb{E}}(X) = \mathbb{E}(YX). \tag{4.52}$$

Definition 4.7 *If $\tilde{\mathbb{P}}(A) > 0$ whenever $\mathbb{P}(A) > 0$, and vice verse, then $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are said to be equivalent measures; they have the same events of probability zero.*

Prove that if $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$ and $\tilde{\mathbb{P}}$ is equivalent to $\mathbb{P}$, then

$$\tilde{\mathbb{E}}(Z|\mathcal{G}) = \frac{1}{\mathbb{E}(\xi|\mathcal{G})}\mathbb{E}(\xi Z|\mathcal{G}), \tag{4.53}$$

where $\xi = \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}$.

Proof: By definition (4.50), we need to prove that for any random variable Y being $\mathcal{G}$-measurable,

$$\tilde{\mathbb{E}}[ZY] = \tilde{\mathbb{E}}[\frac{1}{\mathbb{E}(\xi|\mathcal{G})}\mathbb{E}(\xi Z|\mathcal{G})Y], \tag{4.54}$$

which, by the definition of $\tilde{\mathbb{E}}$, is equivalent to

$$\mathbb{E}[\xi ZY] = \mathbb{E}[\xi\frac{1}{\mathbb{E}(\xi|\mathcal{G})}\mathbb{E}(\xi Z|\mathcal{G})Y]. \tag{4.55}$$

But by (4.24) $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]]$, the right hand side of (4.55) is

$$\mathbb{E}\left[\mathbb{E}\left[\xi\frac{1}{\mathbb{E}(\xi|\mathcal{G})}\mathbb{E}(\xi Z|\mathcal{G})Y\middle|\mathcal{G}\right]\right] \overset{(4.22)}{=} \mathbb{E}\left[\mathbb{E}[\xi|\mathcal{G}]\frac{1}{\mathbb{E}(\xi|\mathcal{G})}\mathbb{E}(\xi Z|\mathcal{G})Y\right]$$

since Y and $\frac{1}{\mathbb{E}(\xi|\mathcal{G})}\mathbb{E}(\xi Z|\mathcal{G})$ are already $\mathcal{G}$-measurable. Hence the right hand side of (4.55) becomes $\mathbb{E}[\mathbb{E}(\xi Z|\mathcal{G})Y] \overset{(4.22)}{=} \mathbb{E}[\mathbb{E}[\xi ZY|\mathcal{G}]] \overset{(4.24)}{=} \mathbb{E}[\xi ZY]$. This proves (4.55).

14. (Generalization of Question 9 of Homework III) Let X and Y be a pair of jointly normal random variables with joint density

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right) \\ &= \frac{1}{2\pi\det(\Sigma)^{1/2}}\exp\left(-\frac{1}{2}(x-\mu_1, y-\mu_2)\Sigma^{-1}\begin{pmatrix} x-\mu_1 \\ y-\mu_2 \end{pmatrix}\right) \end{aligned} \tag{4.56}$$

where σ_1, $\sigma_2 > 0$, $|\rho| < 1$, μ_1, μ_2 are real numbers, and $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$. Define $W = Y - \frac{\rho\sigma_2}{\sigma_1}X$. Prove W and X are independent and show that

$$\mathbb{E}[Y|X] = \frac{\rho\sigma_2}{\sigma_1}(X-\mu_1) + \mu_2. \tag{4.57}$$

Proof: Since we know Σ is the covariance matrix, we know $\mathrm{Var}(X) = \sigma_1^2$, $\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mu_1)(Y - \mu_2)] = \rho\sigma_1\sigma_2$. Then

$$\mathrm{Cov}(X, W) = \mathrm{Cov}(X, Y) - \frac{\rho\sigma_2}{\sigma_1}\mathrm{Cov}(X, X) = \rho\sigma_1\sigma_2 - \rho\sigma_1\sigma_2 = 0.$$

Since X and W are jointly normal distributed and are un-correlated, X and W are independent.

Because $Y = \frac{\rho\sigma_2}{\sigma_1}X + W$ and X and W are indenpendent,

$$\mathbb{E}[Y|X] = \mathbb{E}[\frac{\rho\sigma_2}{\sigma_1}X + W|X] \overset{(4.21)}{=} \frac{\rho\sigma_2}{\sigma_1}X + \mathbb{E}[W|X] = \frac{\rho\sigma_2}{\sigma_1}(X - \mu_1) + \mu_2.$$

In the last step, we have used $\mathbb{E}[W|X] \overset{(4.25)}{=} \mathbb{E}[W] = \mu_2 - \frac{\rho\sigma_2}{\sigma_1}\mu_1$.

15. Let $W(t)$, $t \geq 0$, be a Brownian motion, and let $\mathcal{F}(t)$, $t \geq 0$, be a filtration for this Brownian motion. Show that $W^2(t) - t$ is a martingale, i.e., $\mathbb{E}[W^2(t) - t|\mathcal{F}(s)] = W^2(s) - s$ or equivalently $\mathbb{E}[W^2(t) - W^2(s)|\mathcal{F}(s)] = t - s$ for $0 \leq s \leq t$. (Hint: Write $W^2(t)$ as $(W(t) - W(s))^2 + 2W(t)W(s) - W^2(s)$.)

    Proof:

$$\begin{aligned}\mathbb{E}[W^2(t) - W^2(s)|\mathcal{F}(s)] =& \mathbb{E}[(W(t) - W(s))^2 + 2W(t)W(s) - 2W^2(s)|\mathcal{F}(s)] \\ =& \mathbb{E}[(W(t) - W(s))^2|\mathcal{F}(s)] + 2W(s)\mathbb{E}[W(t) - W(s)|\mathcal{F}(s)] = t - s.\end{aligned}$$

16. Let $\{W_s : s \geq 0\}$ be a 1-dimensional Brownian motion with $W_0 = 0$. Show that the covariance of $\int_0^s W_u du$ and $\int_0^t W_v dv$ is

$$\mathrm{Cov}\left(\int_0^s W_u du, \int_0^t W_v dv\right) = \frac{1}{3}\min\{s^3, t^3\} + \frac{1}{2}|t - s|\min\{s^2, t^2\}. \tag{4.58}$$

Proof: By definition, when $t \geq s$,

$$
\begin{aligned}
&\mathrm{Cov}\left(\int_0^s W_u du, \int_0^t W_v dv\right)\\
=&\mathbb{E}\left[\left(\int_0^s W_u du\right)\left(\int_0^t W_v dv)\right)\right] - \mathbb{E}\left[\left(\int_0^s W_u du\right)\right]\mathbb{E}\left[\left(\int_0^t W_v dv)\right)\right]\\
=&\mathbb{E}\left[\int_0^s W_u \left(\int_0^t W_v dv\right) du\right] = \mathbb{E}\left[\int_0^s \left(\int_0^t W_u W_v dv\right) du\right]\\
=&\left[\int_0^s \left(\int_0^t \mathbb{E}[W_u W_v] dv\right) du\right] = \left[\int_0^s \left(\int_0^t \min(u,v) dv\right) du\right]\\
=&\left[\int_0^s \left(\int_u^t \min(u,v) dv\right) du\right] + \left[\int_0^s \left(\int_0^u \min(u,v) dv\right) du\right] \qquad \text{use } t \geq s\\
=&\left[\int_0^s \left(\int_u^t u dv\right) du\right] + \left[\int_0^s \left(\int_0^u v dv\right) du\right]\\
=&\left[\int_0^s u(t-u) du\right] + \left[\int_0^s u^2/2 du\right]\\
=&ts^2/2 - s^3/6 = \frac{1}{3}s^3 + \frac{1}{2}(t-s)s^2.
\end{aligned}
$$

When $t \leq s$, we switch s and t in the above computation and get

$$\mathrm{Cov}\left(\int_0^s W_u du, \int_0^t W_v dv\right) = \mathrm{Cov}\left(\int_0^t W_v dv, \int_0^s W_u du\right) = st^2/2 - t^3/6 = \frac{1}{3}t^3 + \frac{1}{2}(s-t)t^2.$$

Combining them together, we have proved the desired result.

17. Let $\{W_s : s \geq 0\}$ be a 1-dimensional Brownian motion with $W_0 = 0$. Show that $X_t = W_t^3 - 3tW_t$ is a martingale, i.e., show that for $s \leq t$

$$\mathbb{E}[W_t^3 - 3tW_t | \mathcal{F}_s] = W_s^3 - 3sW_s.$$

[Hint: Rewrite $W_t^3 - 3tW_t$ in terms of the increment $W_t - W_s$ and derive $W_t^3 - 3tW_t = (W_t - W_s)^3 + 3(W_t - W_s)^2 W_s + 3(W_t - W_s)W_s^2 + W_s^3 - 3t(W_t - W_s) - 3tW_s$.]