ST401: Stochastic Methods for Finance

Gechun Liang
Department of Statistics
University of Warwick

2020-2021

Chapter 1: Information and conditioning

Gechun Liang

1 Information and σ-algebra

Financial intuition of filtration. The no-arbitrage theory of option pricing is based on contingency plans. In order to price an option, we would need to set up a hedging position. The hedge must specify what position we will take in the underlying security at each future time contingent on how the uncertainty between the present time and that future time is resolved. In order to make these contingency plans, we need a way to mathematically model the information on which our future decisions can be based, which is called filtration.

Definition 1. Let Ω be the sample space consisting of all possible outcomes. Let T be a fixed positive number representing the maturity, and assume that for each $t \in [0,T]$ there is a σ-algebra $\mathscr{F}(t)$. Assume further that if $s \leq t$, then $\mathscr{F}(s) \subseteq \mathscr{F}(t)$. Then we call the collection of σ-algebras $\mathscr{F}(t), 0 \leq t \leq T$, a filtration. □

Example 1. Tossing a coin three times. Then, Ω is the set of eight possible outcomes. If we are told the outcome of the first coin toss only, the sets

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

are resolved by the information of the first toss. Hence,

$$\mathscr{F}(1) = \{\emptyset, \Omega, A_H, A_T\}$$

Note that the empty set $\emptyset$ and the whole space Ω are always resolved, even without any information.

If we are told the first two coin tosses, we obtain a finer resolution. In particular, the four sets

---

Gechun Liang
Department of Statistics, University of Warwick, U.K. e-mail: g.liang@warwick.ac.uk

$$A_{HH} = \{HHH, HHT\}, \quad A_{HT} = \{HTH, HTT\}$$
$$A_{TH} = \{THH, THT\}, \quad A_{TT} = \{TTH, TTT\}$$

are resolved by the information of the first two tosses. Of course, the sets in $\mathscr{F}_1$ are still resolved. In all, we have 16 resolved sets that together form a σ-algebra we call $\mathscr{F}(2)$; i.e.

$$\mathscr{F}(2) = \left\{ \begin{array}{l} \emptyset, \Omega, A_H, A_T, A_{HH}, A_{HT}, A_{TH}, A_{TT}, A^c_{HH}, A^c_{HT}, A^c_{TH}, A^c_{TT}, \\ A_{HH} \cup A_{TH}, A_{HH} \cup A_{TT}, A_{HT} \cup A_{TH}, A_{HT} \cup A_{TT} \end{array} \right\}.$$

Finally, $\mathscr{F}(3)$ consists of $2^8 = 256$ outcomes (why?) □

Definition 2. Let Ω be a nonempty sample space equipped with a filtration $\mathscr{F}(t), 0 \leq t \leq T$. Let $X(t)$ be a collection of random variables indexed by $t \in [0,T]$. We say this collection of random variables is an adapted stochastic process if, for each t, the random variable $X(t)$ is $\mathscr{F}(t)$-measurable. □

We often associate a stochastic process with its *natural filtration*. Let X be a random variable defined on a nonempty sample space Ω. The σ-algebra generated by X, denoted $\sigma(X)$, is the collection of all subsets of Ω of the form $\{X \in B\}$, where B ranges over the Borel subsets of $\mathbb{R}$. Then, the natural filtration generated by a stochastic process $X(t)$ is $\mathscr{F}(t) = \sigma(X(s), s \leq t)$, $t \in [0,T]$.

2 Independence

Two extremes of information sets. When a random variable X is measurable with respect to a σ-algebra $\mathscr{F}$, the information contained in $\mathscr{F}$ is sufficient to determine the value of the random variable X. The other extreme is when a random variable X is independent of a σ-algebra $\mathscr{G}$. In this case, the information contained in the σ-algebra $\mathscr{G}$ gives no clue about the value of the random variable, which motivates us to introduce the concept of independence.

Definition 3. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, and let $\mathscr{G}, \mathscr{H} \subseteq \mathscr{F}$. We say these two σ-algebras are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) \text{ for all } A \in \mathscr{G}, B \in \mathscr{H}.$$

Similarly, let X and Y be random variables on $(\Omega, \mathscr{F}, \mathbb{P})$. We say these two random variables are independent if the σ-algebras they generate, $\sigma(X)$ and $\sigma(Y)$, are independent.

Finally, we say that the random variable X is independent of the σ-algebra $\mathscr{G}$ if $\sigma(X)$ and $\mathscr{G}$ are independent. □

Recall that $\sigma(X)$ is the collection of all sets of the form $\{X \in C\}$, where C ranges over the Borel subsets of $\mathbb{R}$. Similarly, every set in $\sigma(Y)$ is of the form $\{Y \in D\}$. Then, X and Y are independent if and only if

$$\mathbb{P}\{X \in C \text{ and } Y \in D\} = \mathbb{P}\{X \in C\} \cdot \mathbb{P}\{Y \in D\}$$

for all Borel subsets C and D of $\mathbb{R}$.

Exercise 1. In an analogous way, write down the definition of independence for more than two random variables.

Proposition 1. *Let X and Y be independent random variables, and let f and g be Borel-measurable functions on $\mathbb{R}$. Then $f(X)$ and $g(Y)$ are independent random variables.*

Proof. Let A be in the σ-algebra generated by $f(X)$, i.e. $A \in \sigma(f(X))$. Note that $\sigma(f(X)) \subseteq \sigma(X)$, so $A \in \sigma(X)$. Similarly, let B be in the σ-algebra generated by $g(Y)$, i.e. $B \in \sigma(g(Y))$. Then, we also have $B \in \sigma(Y)$. since X and Y are independent, we conclude

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B). \square$$

Exercise 2. Prove that σ-algebra $\sigma(f(X))$ is a sub-σ-algebra of $\sigma(X)$, i.e.

$$\sigma(f(X)) \subseteq \sigma(X).$$

Proposition 2. *Let X and Y be random variables. The following are equivalent (TFAE).*

1. *X and Y are <u>independent</u>.*
2. *The joint distribution measure factors:*

$$\mu_{X,Y}(A \times B) = \mu_X(A) \cdot \mu_Y(B) \tag{1}$$

   for all Borel subsets $A \subset \mathbb{R}, B \subset \mathbb{R}$
3. *The joint cumulative distribution function factors:*

$$F_{X,Y}(a,b) = F_X(a) \cdot F_Y(b) \tag{2}$$

   for all $a \in \mathbb{R}, b \in \mathbb{R}$.
4. *The joint moment-generating function factors:*

$$\mathbb{E}e^{uX+vY} = \mathbb{E}e^{uX} \cdot \mathbb{E}e^{vY} \tag{3}$$

   for all $u \in \mathbb{R}, v \in \mathbb{R}$ for which the expectations are finite.
   If there is a joint density, each of the conditions above is equivalent to the following.
5. *The joint density factors:*

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) \tag{4}$$

   for almost every $x \in \mathbb{R}, y \in \mathbb{R}$.
   The conditions above imply but are not equivalent to the following.

6. *The expectation factors:*
$$\mathbb{E}[XY] = \mathbb{E}X \cdot \mathbb{E}Y, \tag{5}$$
provided $\mathbb{E}[XY] < \infty$. *If (5) holds, X and Y is also called uncorrelated.*

Exercise 3. Prove Proposition 2.

Example 2. (Uncorrelated, dependent normal random variables). Let X be a standard normal random variable and let Z be independent of X and satisfy

$$\mathbb{P}\{Z = 1\} = \frac{1}{2} \text{ and } \mathbb{P}\{Z = -1\} = \frac{1}{2}.$$

Define $Y = ZX$. Then, the following results hold.

1. The random variable Y is standard normal. To see this, we compute
$$\begin{aligned} F_Y(b) &= \mathbb{P}\{Y \le b\} \\ &= \mathbb{P}\{Y \le b \text{ and } Z = 1\} + \mathbb{P}\{Y \le b \text{ and } Z = -1\} \\ &= \mathbb{P}\{X \le b \text{ and } Z = 1\} + \mathbb{P}\{-X \le b \text{ and } Z = -1\} \end{aligned}$$
Because X and Z are independent, we have
$$\begin{aligned} &\mathbb{P}\{X \le b \text{ and } Z = 1\} + \mathbb{P}\{-X \le b \text{ and } Z = -1\} \\ &= \mathbb{P}\{Z = 1\} \cdot \mathbb{P}\{X \le b\} + \mathbb{P}\{Z = -1\} \cdot \mathbb{P}\{-X \le b\} \\ &= \frac{1}{2} \cdot \mathbb{P}\{X \le b\} + \frac{1}{2} \cdot \mathbb{P}\{-X \le b\} \end{aligned}$$
Because X is a standard normal random variable, so is $-X$. Therefore, $\mathbb{P}\{X \le b\} = \mathbb{P}\{-X \le b\} = N(b)$. It follows that $F_Y(b) = N(b)$;
2. *The random variables X and Y are uncorrelated*. Indeed, since $\mathbb{E}X = \mathbb{E}Y = 0$, the covariance of X and Y is
$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] = \mathbb{E}\left[ZX^2\right]$$
Because Z and X are independent, so are Z and X^2. Hence,
$$\mathbb{E}\left[ZX^2\right] = \mathbb{E}Z \cdot \mathbb{E}\left[X^2\right] = 0 \cdot 1 = 0$$
3. *The random variables X and Y are not independent*. If they were, then $|X|$ and $|Y|$ would also be independent (why?) . But $|X| = |Y|$. In particular,
$$\mathbb{P}\{|X| \le 1, |Y| \le 1\} = \mathbb{P}\{|X| \le 1\} = N(1) - N(-1)$$
and
$$\mathbb{P}\{|X| \le 1\} \cdot \mathbb{P}\{|Y| \le 1\} = (N(1) - N(-1))^2.$$
4. *The pair* (X,Y) *does not have a joint density*. Note that $|X| = |Y|$, the pair (X,Y) takes values only in the set

$$C = \{(x,y); x = \pm y\}$$

which has measure 0. However, $\mu_{X,Y}(C) = 1$ and $\mu_{X,Y}(C^c) = 0$. If the density $f_{X,Y}$ were existing, then

$$1 = \mu_{X,Y}(C) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{I}_C(x,y) f_{X,Y}(x,y)\,dy\,dx = 0,$$

which is absurd. □

However, if X and Y are jointly normal, i.e. they have the joint density

$$\begin{aligned} &f_{X,Y}(x,y) \\ &= \tfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\tfrac{1}{2\left(1-\rho^2\right)}\left[\tfrac{(x-\mu_1)^2}{\sigma_1^2} - \tfrac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right.\right. \\ &\left.\left.+\tfrac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\} \end{aligned}$$

where $\sigma_1 > 0, \sigma_2 > 0, |\rho| < 1$, and μ_1, μ_2 are real numbers, then X and Y are independent iff they are uncorrelated. Indeed, the *only if* direction follows from part (6) in Proposition 2. To show the *if* direction, note that X and Y are uncorrelated means

$$\rho = \rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}} = 0$$

so the density factors which is equivalent to the independence of X and Y according to part (5) in Proposition 2.

Exercise 4. Let (X,Y) be a pair of random variables with joint density function

$$f_{X,Y}(x,y) = \begin{cases} \frac{2|x|+y}{\sqrt{2\pi}} \exp\left\{-\frac{(2|x|+y)^2}{2}\right\} & \text{if } y \geq -|x| \\ 0 & \text{if } y < -|x| \end{cases}$$

Show that X and Y are standard normal random variables and that they are *uncorrelated but not independent*.

3 Conditional expectation

Intuition of conditional expectations. Usually, the random variable X is neither measurable with respect to the σ-algebra $\mathscr{G}$ nor independent of $\mathscr{G}$, so the information contained in $\mathscr{G}$ is insufficient to evaluate X, but we can estimate X based on $\mathscr{G}$, which leads to conditional expectation.

Definition 4. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, let $\mathscr{G}$ be a sub-σ-algebra of $\mathscr{F}$, and let X be a random variable that is either nonnegative or integrable. The conditional expectation of X given $\mathscr{G}$, denoted $\mathbb{E}[X \mid \mathscr{G}]$, is any random variable that satisfies

1. (Measurability) $\mathbb{E}[X \mid \mathscr{G}]$ is $\mathscr{G}$-measurable,
2. (Partial averaging)

$$\int_A \mathbb{E}[X \mid \mathscr{G}](\omega) d\mathbb{P}(\omega) = \int_A X(\omega) d\mathbb{P}(\omega) \text{ for all } A \in \mathscr{G}$$

If $\mathscr{G}$ is the σ-algebra generated by some other random variable W (i.e., $\mathscr{G} = \sigma(W)$), we generally write $\mathbb{E}[X \mid W]$ rather than $\mathbb{E}[X \mid \sigma(W)]$.

Interpretations of the two properties of conditional expectations. Condition (1) captures the fact that the estimate $\mathbb{E}[X|\mathscr{G}]$ of X is based on the information in $\mathscr{G}$. Condition (2) ensures that $\mathbb{E}[X \mid \mathscr{G}]$ is indeed an estimate of X. It gives the same averages as X over all the sets in $\mathscr{G}$. If $\mathscr{G}$ has many sets, which provide a fine resolution of the uncertainty inherent in ω, then this partial-averaging property over the "small" sets in $\mathscr{G}$ says that $\mathbb{E}[X \mid \mathscr{G}]$ is a good estimator of X. If $\mathscr{G}$ has only a few sets, this partial-averaging property guarantees only that $\mathbb{E}[X \mid \mathscr{G}]$ is a crude estimate of X.

Condition expectations admit the following five fundamental properties.

Theorem 1. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $\mathscr{G}$ be a sub-σ-algebra of $\mathscr{F}$.*

1. *(Linearity of conditional expectations) If X and Y are integrable random variables and c_1 and ca are constants, then*

$$\mathbb{E}\left[c_1 X + c_2 Y \mid \mathscr{G}\right] = c_1 \mathbb{E}[X \mid \mathscr{G}] + c_2 \mathbb{E}[Y \mid \mathscr{G}] \tag{6}$$

   This equation also holds if we assume that X and Y are nonnegative (rather than integrable) and c_1 and c_2 are positive, although both sides may be $+\infty$.
2. *(Taking out what is known) If X and Y are integrable random variables, and X is $\mathscr{G}$-measurable, then*

$$\mathbb{E}[XY \mid \mathscr{G}] = X \mathbb{E}[Y \mid \mathscr{G}] \tag{7}$$

   This equation also holds if we assume that X is positive and Y is nonnegative (rather than integrable), although both sides may be $+\infty$.
3. *(Iterated conditioning) If $\mathscr{H}$ is a sub-σ-algebra of $\mathscr{G}$ (H contains less information than $\mathscr{G}$) and X is an integrable random variable, then*

$$\mathbb{E}[\mathbb{E}[X \mid \mathscr{G}] \mid \mathscr{H}] = \mathbb{E}[X \mid \mathscr{H}] \tag{8}$$

   This equation also holds if we assume that X is nonnegative (rather than integrable), although both sides may be $+\infty$.
4. *(Independence) If X is integrable and independent of $\mathscr{G}$, then*

$$\mathbb{E}[X \mid \mathscr{G}] = \mathbb{E}X. \tag{9}$$

   This equation also holds if we assume that X is nonnegative (rather than integrable), although both sides may be $+\infty$.

5. *(Conditional Jensen's inequality) If $\varphi(x)$ is a convex function of a dummy variable x and X is integrable, then*

$$\mathbb{E}[\varphi(X) \mid \mathscr{G}] \geq \varphi(E[X \mid \mathscr{G}]). \tag{10}$$

Proof. We only verify the partial averaging property, as the measurability property is clear from its definition.

Part (1). For any $A \in \mathscr{G}$,

$$\begin{aligned}
&\textstyle\int_A \left(c_1 \mathbb{E}[X \mid \mathscr{G}](\omega) + c_2 \mathbb{E}[Y \mid \mathscr{G}](\omega)\right) d\mathbb{P}(\omega) \\
&= c_1 \textstyle\int_A \mathbb{E}[X \mid \mathscr{G}](\omega) d\mathbb{P}(\omega) + c_2 \int_A \mathbb{E}[Y \mid \mathscr{G}](\omega) d\mathbb{P}(\omega) \\
&= c_1 \textstyle\int_A X(\omega) d\mathbb{P}(\omega) + c_2 \int_A Y(\omega) d\mathbb{P}(\omega) \\
&= \textstyle\int_A \left(c_1 X(\omega) + c_2 Y(\omega)\right) d\mathbb{P}(\omega),
\end{aligned}$$

which shows the partial averaging property.

Part (2). We only consider the case when X is a $\mathscr{G}$-measurable indicator random variable (i.e., $X = \mathbb{I}_B$, where B is a set in $\mathscr{G}$). The general case follows from the standard approximation procedure. We have for every set $A \in \mathscr{G}$ that

$$\begin{aligned}
\int_A X(\omega)\mathbb{E}[Y \mid \mathscr{G}](\omega) d\mathbb{P}(\omega) &= \int_{A\cap B} \mathbb{E}[Y \mid \mathscr{G}](\omega) d\mathbb{P}(\omega) \\
&= \int_{A\cap B} Y(\omega) d\mathbb{P}(\omega) \\
&= \int_A X(\omega) Y(\omega) d\mathbb{P}(\omega),
\end{aligned}$$

which shows the partial averaging property.

Part (3). For any $A \in \mathscr{H} \subseteq \mathscr{G}$, we have

$$\int_A \mathbb{E}[\mathbb{E}[X \mid \mathscr{G}] \mid \mathscr{H}](\omega) d\mathbb{P}(\omega) = \int_A \mathbb{E}[X \mid \mathscr{G}](\omega)\mathbb{P}(\omega),$$

and

$$\int_A X(\omega) d\mathbb{P}(\omega) = \int_A \mathbb{E}[X \mid \mathscr{G}](\omega) d\mathbb{P}(\omega),$$

which shows the partial averaging property.

Part (4). We only consider the case when X is an indicator random variable independent of $\mathscr{G}$ (i.e., $X = \mathbb{I}_B$, where the set B is independent of $\mathscr{G}$). For all $A \in \mathscr{G}$, we have then

$$\int_A X(\omega) d\mathbb{P}(\omega) = \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) = \mathbb{P}(A)\mathbb{E}X = \int_A \mathbb{E}X d\mathbb{P}(\omega).$$

We then complete the proof using the standard approximation procedure.

Part (5). since φ is convex, it admits the representation

$$\varphi(x) = \max\nolimits_{h \leq \varphi} h(x)$$

with $h(x) = ax + b$ below φ. Then,

$$\begin{aligned}\mathbb{E}[\varphi(X) \mid \mathscr{G}] &\geq \mathbb{E}[aX + b \mid \mathscr{G}] \\ &= a\mathbb{E}[X \mid \mathscr{G}] + b \\ &= h(\mathbb{E}[X \mid \mathscr{G}])\end{aligned}$$

This implies

$$\mathbb{E}[\varphi(X) \mid \mathscr{G}] \geq \max_{h \leq \varphi} h(\mathbb{E}[X \mid \mathscr{G}]) = \varphi(\mathbb{E}[X \mid \mathscr{G}]),$$

which completes the proof. □

Example 3. Let (X,Y) be jointly normal with the density

$$\begin{aligned}&f_{X,Y}(x,y) \\ &= \tfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\tfrac{1}{2\left(1-\rho^2\right)}\left[\tfrac{(x-\mu_1)^2}{\sigma_1^2} - \tfrac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right.\right. . \\ &\left.\left.+\tfrac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}\end{aligned}$$

Let us take the conditioning σ-algebra to be $\mathscr{G} = \sigma(X)$. *Our aim is estimate Y based on X, i.e. compute* $\mathbb{E}[Y|X]$.

To this end, we first decompose Y by defining

$$W = Y - \frac{\rho\sigma_2}{\sigma_1}X.$$

Then, X and W are independent. It suffices to show that X and W have covariance zero since they are jointly normal. We compute

$$\begin{aligned}\mathrm{Cov}(X,W) &= \mathbb{E}[(X - \mathbb{E}X)(W - \mathbb{E}W)] \\ &= \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] - \mathbb{E}\left[\frac{\rho\sigma_2}{\sigma_1}(X - \mathbb{E}X)^2\right] \\ &= \mathrm{Cov}(X,Y) - \frac{\rho\sigma_2}{\sigma_1}\sigma_1^2 \\ &= 0\end{aligned}$$

The expectation of W is $\mu_3 = \mu_2 - \frac{\rho\sigma_2\mu_1}{\sigma_1}$, and the variance is

$$\begin{aligned}\sigma_3^2 &= \mathbb{E}\left[(W - \mathbb{E}W)^2\right] \\ &= \mathbb{E}\left[(Y - \mathbb{E}Y)^2\right] - \tfrac{2\rho\sigma_2}{\sigma_1}\mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] + \tfrac{\rho^2\sigma_2^2}{\sigma_1^2}\mathbb{E}\left[(X - \mathbb{E}X)^2\right] \\ &= \left(1 - \rho^2\right)\sigma_2^2.\end{aligned}$$

Hence, the joint density of X and W is

$$f_{X,W}(x,w) = \frac{1}{2\pi\sigma_1\sigma_3}\exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(w-\mu_3)^2}{2\sigma_3^2}\right\}$$

Hence, we have decomposed Y into the linear combination

$$Y = \frac{\rho\sigma_2}{\sigma_1}X + W$$

of a pair of independent normal random variables X and W.

Thus, we obtain the *linear regression equation*

$$\mathbb{E}[Y \mid X] = \frac{\rho\sigma_2}{\sigma_1}X + \mathbb{E}W = \frac{\rho\sigma_2}{\sigma_1}(X - \mu_1) + \mu_2$$

Note that the right-hand side is random but is $\sigma(X)$-measurable (i.e., if we know the information in $\sigma(X)$, which is the same as knowing the value of X, then we can evaluate $\mathbb{E}[Y \mid X]$).

The error made by the estimator is

$$Err := Y - \mathbb{E}[Y \mid X] = W - \mathbb{E}W \tag{11}$$

The error is random, with expected value zero (*so the estimator is unbiased*), and is independent of the estimate $\mathbb{E}[Y \mid X]$ (why?). The independence between the error and the conditioning random variable X is a consequence of the joint normality of (W, X). In general, *the error and the conditioning random variable are uncorrelated, but not necessarily independent.* □

Exercise 5. *(Estimation of Y based on X, general case.)*

Let X and Y be integrable random variables on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Then $Y = \mathbb{E}[Y \mid X] + Err$ according to (11).

1. Show that Err and X are uncorrelated. More generally, show that Err is uncorrelated with every $\sigma(X)$-measurable random variable.
2. Let X' be some other $\sigma(X)$-measurable random variable, which we can regard as another estimate of Y. Show that

$$\mathrm{Var}(\mathrm{Err}) \leq \mathrm{Var}(Y - X')$$

   In other words, the estimate $\mathbb{E}[Y \mid X]$ minimizes the variance of the error among all estimates based on the information in $\sigma(X)$. □

Finally, we introduce *martingales* and *Markov processes* in a continuous-time framework. Their examples will be given after we construct Brownian motion and Itô integrals in the next chapters.

Definition 5. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, let T be a fixed positive number, and let $\mathscr{F}(t), 0 \leq t \leq T$, be a filtration of sub-σ-algebras of $\mathscr{F}$. Consider an adapted stochastic process $M(t), 0 \leq t \leq T$.

1. If

$$\mathbb{E}[M(t) \mid \mathscr{F}(s)] = M(s) \text{ for all } 0 \leq s \leq t \leq T,$$

   we say this process is a <u>martingale</u>. It has no tendency to rise or fall.

2. If
$$\mathbb{E}[M(t) \mid \mathscr{F}(s)] \geq M(s) \text{ for all } 0 \leq s \leq t \leq T,$$
we say this process is a submartingale. It has no tendency to fall; it may have a tendency to rise.
3. If
$$\mathbb{E}[M(t) \mid \mathscr{F}(s)] \leq M(s) \text{ for all } 0 \leq s \leq t \leq T,$$
we say this process is a supermartingale. It has no tendency to rise; it may have a tendency to fall.

Definition 6. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, let T be a fixed positive number, and let $\mathscr{F}(t), 0 \leq t \leq T$, be a filtration of sub-σ-algebras of $\mathscr{F}$. Consider an adapted stochastic process $X(t), 0 \leq t \leq T$. Assume that for all $0 \leq s \leq t \leq T$ and for every nonnegative, Borel-measurable function f, there is another Borel-measurable function g such that
$$\mathbb{E}[f(X(t)) \mid \mathscr{F}(s)] = g(X(s))$$
Then we say that the X is a Markov process.

Chapter 2: Brownian motion

Gechun Liang

1 Scaled random walks

Symmetric random walk and its basic properties. Consider tossing a fair coin repeatedly. The successive outcomes of the tosses is $\omega = \omega_1\omega_2\omega_3\ldots$. In other words, ω is the infinite sequence of tosses, and ω_n is the outcome of the nth toss. Let

$$X_j = \begin{cases} 1 \text{ if } \omega_j = H \\ -1 \text{ if } \omega_j = T \end{cases}$$

and define $M_0 = 0$

$$M_k = \sum_{j=1}^{k} X_j, k = 1, 2, \ldots$$

The process $M_k, k = 0, 1, 2, \ldots$ is called a symmetric random walk. With each toss, it either steps up one unit or down one unit, and each of the two possibilities is equally likely. It has the following properties.

1. *The random walk M has independent increments.* For any nonnegative integers $0 = k_0 < k_1 < k_2 \ldots,$

$$M_{k_{i+1}} - M_{k_i} = \sum_{j=k_i+1}^{k_{i+1}} X_j$$

   is called an increment of the random walk. It is the change in the position of the random walk between times k_i and k_{i+1}. Increments over nonoverlapping time intervals are independent because they depend on different coin tosses.
2. *The random walk M is a martingale.* For any nonnegative integers $k < \ell$, we compute

Gechun Liang
Department of Statistics, University of Warwick, U.K. e-mail: g.liang@warwick.ac.uk

$$
\begin{aligned}
\mathbb{E}[M_\ell \mid \mathscr{F}_k] &= \mathbb{E}[(M_\ell - M_k) + M_k \mid \mathscr{F}_k] \\
&= \mathbb{E}[M_\ell - M_k \mid \mathscr{F}_k] + \mathbb{E}[M_k \mid \mathscr{F}_k] \\
&= \mathbb{E}[M_\ell - M_k \mid \mathscr{F}_k] + M_k \\
&= \mathbb{E}[M_\ell - M_k] + M_k = M_k,
\end{aligned}
$$

where the last step follows from the zero mean of the symmetric random walk M.

3. *Quadratic variation and variance of the random walk M.* The quadratic variation up to time k is defined to be

$$[M,M]_k = \sum_{j=1}^{k} \left(M_j - M_{j-1}\right)^2 = k$$

Note that this is computed *path-by-path*. The quadratic variation up to time k along a path is computed by taking all the one-step increments $M_j - M_{j-1}$ along that path (these are equal to X_j, which is either 1 or -1, depending on the path), squaring these increments, and then summing them. since $\left(M_j - \ M_{j-1}\right)^2 = 1$, regardless of whether $M_j - M_{j-1}$ is 1 or -1, the sum in is equal to $\sum_{i=1}^{k} 1 = k$, as reported in that equation.
On the other hand, the variance is computed as

$$Var(M_k) = Var(\sum_{j=1}^{k} X_j) = k.$$

We note that $[M,M]_k$ is the same as $\mathrm{Var}(M_k)$, but the computations of these two quantities are quite different. $\mathrm{Var}(M_k)$ is computed by taking an average over all paths, taking their probabilities into account (so probabilities matter). By contrast, $[M,M]_k$ is computed along a single path, and the probabilities of up and down steps do not enter the computation.

Scaled symmetric random walk. To approximate a Brownian motion, we *speed up time* and *scale down the step size* of a symmetric random walk. For a fixed positive integer n, define the scaled symmetric random walk

$$W^{(n)}(t) = \frac{1}{\sqrt{n}} M_{nt} \tag{1}$$

provided nt is itself an integer. Hence, if M takes 1 step per unit time with step size 1, then its scaled version $W^{(n)}$ takes n steps per unit time with step size $1/\sqrt{n}$.

If nt is not an integer, we define $W^{(n)}(t)$ by linear interpolation. For example, if s and u are the nearest points to the left and right of t for which ns and nu are integers, then

$$W^{(n)}(t) = \frac{u-t}{u-s} W^{(n)}(s) + \frac{t-s}{u-s} W^{(n)}(u).$$

Like the random walk, the scaled random walk also has the properties.

1. *Independent increments*. If $0 = t_0 < t_1 < t_2 < \cdots$ are such that each nt_j is an integer, then

$$\left(W^{(n)}(t_1) - W^{(n)}(t_0)\right), \left(W^{(n)}(t_2) - W^{(n)}(t_1)\right), \ldots,$$

   are independent. Moreover, for $s \leq t$ such that ns and nt are integers,

$$\mathbb{E}\left(W^{(n)}(t) - W^{(n)}(s)\right) = 0, \quad \mathrm{Var}\left(W^{(n)}(t) - W^{(n)}(s)\right) = t - s.$$

   This is because $W^{(n)}(t) - W^{(n)}(s)$ is the sum of $n(t-s)$ independent random variables, each with expected value zero and variance $\frac{1}{n}$.
2. *Martingale property*. We may prove the martingale property for the scaled random walk as we did for the random walk.

$$\mathbb{E}\left[W^{(n)}(t) \mid \mathscr{F}(s)\right] = W^{(n)}(s)$$

   for $0 \leq s \leq t$ such that ns and nt are integers.
3. *Quadratic variation*. For $t \geq 0$ such that nt is an integer,

$$\begin{aligned}\left[W^{(n)}, W^{(n)}\right](t) &= \sum_{j=1}^{nt}\left[W^{(n)}\left(\frac{j}{n}\right) - W^{(n)}\left(\frac{j-1}{n}\right)\right]^2 \\ &= \sum_{j=1}^{nt}\left[\frac{1}{\sqrt{n}}X_j\right]^2 = \sum_{j=1}^{nt}\frac{1}{n} = t.\end{aligned}$$

The main result of this section is the following central limit theorem.

Theorem 1. (Central limit). *Fix $t \geq 0$. As $n \to \infty$, the distribution of the scaled random walk $W^{(n)}(t)$ defined in (1) evaluated at time t converges to the normal distribution with mean zero and variance t, i.e. $W^n(t)$ converges weakly to $N(0,t)$.*

Proof. It is sufficient to prove the corresponding moment-generating functions converge (why?) For the normal density

$$f(x) = \frac{1}{\sqrt{2\pi t}}e^{-\frac{x^2}{2t}},$$

its moment-generating function is

$$\varphi(u) = \int_{-\infty}^{\infty} e^{ux} f(x)dx = e^{\frac{1}{2}u^2 t}.$$

If t is such that nt is an integer, then the moment-generating function for $W^{(n)}(t)$ is

$$\varphi_n(u) = \mathbb{E}e^{uW^{(n)}(t)} = \mathbb{E}\exp\left\{\frac{u}{\sqrt{n}}M_{nt}\right\}$$
$$= \mathbb{E}\exp\left\{\frac{u}{\sqrt{n}}\sum_{j=1}^{nt}X_j\right\} = \mathbb{E}\prod_{j=1}^{nt}\exp\left\{\frac{u}{\sqrt{n}}X_j\right\}$$

Because the random variables are independent, the RHS may be written as

$$\prod_{j=1}^{nt}\mathbb{E}\exp\left\{\frac{u}{\sqrt{n}}X_j\right\} = \prod_{j=1}^{nt}\left(\frac{1}{2}e^{\frac{u}{\sqrt{n}}} + \frac{1}{2}e^{-\frac{u}{\sqrt{n}}}\right) = \left(\frac{1}{2}e^{\frac{u}{\sqrt{n}}} + \frac{1}{2}e^{-\frac{u}{\sqrt{n}}}\right)^{nt}$$

We need to show that, as $n \to \infty$

$$\varphi_n(u) = \left(\frac{1}{2}e^{\frac{u}{\sqrt{n}}} + \frac{1}{2}e^{-\frac{u}{\sqrt{n}}}\right)^{nt}$$

converges to the moment-generating function $\varphi(u) = e^{\frac{1}{2}u^2t}$. To do this, it suffices to consider the logarithm of $\varphi_n(u)$ and show that

$$\log\varphi_n(u) = nt\log\left(\tfrac{1}{2}e^{\frac{u}{\sqrt{n}}} + \tfrac{1}{2}e^{-\frac{u}{\sqrt{n}}}\right) \to \log\varphi(u) = \tfrac{1}{2}u^2t,$$

which follows from L'Hopital rule (left as an exercise).□

Example 1. (Log-Normal Distribution as the Limit of the Binomial Model)
Using the CLT, we show that the binomial model is a discrete-time version of the geometric Brownian motion model, which is the basis for the Black-Scholes-Merton option-pricing formula.

Let us build a model for a stock price on the time interval from 0 to t by choosing an integer n (such that nt is an integer), and constructing a binomial model for the stock price that takes n steps per unit time. We take the up factor to be $u_n = 1 + \frac{\sigma}{\sqrt{n}}$ and the down factor to be $d_n = 1 - \frac{\sigma}{\sqrt{n}}$. The risk-free factor is taken to be $\rho_n = 1 + \frac{r}{n}$. Here σ is a positive constant that will turn out to be the volatility of the limiting stock price process, and r is an interest rate assumed to be $r = 0$. Then, the risk-neutral probabilities are

$$\tilde{p} = \frac{\rho_n - d_n}{u_n - d_n} = \frac{\sigma/\sqrt{n}}{2\sigma/\sqrt{n}} = \frac{1}{2}, \quad \tilde{q} = \frac{u_n - \rho_n}{u_n - d_n} = \frac{\sigma/\sqrt{n}}{2\sigma/\sqrt{n}} = \frac{1}{2}.$$

The stock price at time t is determined by the initial stock price $S(0)$ and the result of the first nt coin tosses. The sum of the number of heads H_{nt} and number of tails T_{nt} in the first nt coin tosses is nt, a fact that we write as

$$nt = H_{nt} + T_{nt}$$

The random walk M_{nt} is the number of heads minus the number of tails in these nt coin tosses:

$$M_{nt} = H_{nt} - T_{nt}$$

In turn,

$$H_{nt} = \frac{1}{2}\left(nt + M_{nt}\right); \quad T_{nt} = \frac{1}{2}\left(nt - M_{nt}\right).$$

In the model with up factor u_n and down factor d_n, the stock price at time t is

$$S_n(t) = S(0) u_n^{H_{nt}} d_n^{T_{nt}} = S(0)\left(1 + \frac{\sigma}{\sqrt{n}}\right)^{\frac{1}{2}(nt+M_{nt})}\left(1 - \frac{\sigma}{\sqrt{n}}\right)^{\frac{1}{2}(nt-M_{nt})}$$

As $n \to \infty$, the distribution of $S_n(t)$ converges to the log-normal distribution of

$$S(t) = S(0)\exp\left\{\sigma W(t) - \frac{1}{2}\sigma^2 t\right\}$$

where $W(t)$ is a normal random variable with mean zero and variance t. Indeed,

$$\log S_n(t) = \log S(0) + \frac{1}{2}\left(nt + M_{nt}\right)\log\left(1 + \frac{\sigma}{\sqrt{n}}\right) + \frac{1}{2}\left(nt - M_{nt}\right)\log\left(1 - \frac{\sigma}{\sqrt{n}}\right).$$

According to Taylor's approximation, $\log(1+x) = x - \frac{1}{2}x^2 + O(x^3)$, where $O(x^3)$ indicates a term of order x^3. Thus,

$$\begin{aligned}
\log S_n(t) =& \log S(0) + \frac{1}{2}\left(nt + M_{nt}\right)\left(\frac{\sigma}{\sqrt{n}} - \frac{\sigma^2}{2n} + O\left(n^{-\frac{3}{2}}\right)\right) \\
&+ \frac{1}{2}\left(nt - M_{nt}\right)\left(-\frac{\sigma}{\sqrt{n}} - \frac{\sigma^2}{2n} + O\left(n^{-\frac{3}{2}}\right)\right) \\
=& \log S(0) + nt\left(-\frac{\sigma^2}{2n} + O\left(n^{-\frac{3}{2}}\right)\right) + M_{nt}\left(\frac{\sigma}{\sqrt{n}} + O\left(n^{-\frac{3}{2}}\right)\right) \\
=& \log S(0) - \frac{1}{2}\sigma^2 t + O\left(n^{-\frac{1}{2}}\right) + \sigma W^{(n)}(t) + O\left(n^{-1}\right) W^{(n)}(t)
\end{aligned}$$

and we conclude that as $n \to \infty$ the distribution of $\log S(t)$ approaches the distribution of $\log S(0) - \frac{1}{2}\sigma^2 t + \sigma W(t)$. □

Exercise 1. *(The general case with r not necessarily 0)*

1. Show that

$$\ln S_n(t) = \ln S(0) + \sum_{j=1}^{nt} \ln\left(\frac{S_n(\frac{j}{n}) - S_n(\frac{j-1}{n})}{S_n(\frac{j-1}{n})} + 1\right).$$

2. Let $\bar{X}_j := \frac{S_n(\frac{j}{n}) - S_n(\frac{j-1}{n})}{S_n(\frac{j-1}{n})}$. Then show that for μ_n, d_n and ρ_n given in Example 1 (with $r = 0$), it holds that $\bar{X}_j$ has the same distribution as $\frac{\sigma}{\sqrt{n}}X_j$. Re-prove the convergence of $S_n(t)$ to $S(t)$ based on this result.
3. For $r \neq 0$, we take the up factor $\mu_n = \exp\{\frac{\sigma}{\sqrt{n}}\}$, the down factor $d_n = \exp\{-\frac{\sigma}{\sqrt{n}}\}$ and the risk-free factor $\rho_n = \exp\{\frac{r}{n}\}$. Define the discounted stock price (under

the binomial model) at time t as $S_n^*(t) = \frac{S_n(t)}{\rho^{nt}}$. Show that under the risk-neutral probability measures $(\tilde{p}, \tilde{q})$, the random variable defined by

$$\bar{X}_j^* := \frac{S_n^*(\frac{j}{n}) - S_n^*(\frac{j-1}{n})}{S_n^*(\frac{j-1}{n})}$$

has mean 0 and variance $\frac{\sigma^2}{n}$. Hence, we may approximate $\bar{X}_j^*$ by $\frac{\sigma}{\sqrt{n}} X_j$ (as they have the same mean and variance). Based on this result, prove that $S_n(t)$ converges weakly to

$$S(t) = S(0) \exp\left\{ \sigma W(t) - \frac{1}{2}\sigma^2 t + rt \right\}.$$

2 Brownian motion

Brownian motion is obtained as the limit of the scaled random walks $W^{(n)}(t)$ as $n \longrightarrow \infty$, and inherits its properties from these random walks.

Definition 1. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. For each $\omega \in \Omega$, suppose there is a continuous function $W(t)$ of $t \geq 0$ that satisfies $W(0) = 0$ and that depends on ω. Then $W(t), t \geq 0$, is a Brownian motion if for all $0 = t_0 < t_1 < t_2 < \cdots$ the increments

$$W(t_1) = W(t_1) - W(t_0), W(t_2) - W(t_1), \ldots,$$

are independent and each of these increments is normally distributed with

$$W(t_{i+1}) - W(t_i) \sim N(0, t_{i+1} - t_i).$$

Definition 2. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space on which is defined a Brownian motion $W(t), t \geq 0$. A filtration for the Brownian motion is a collection of σ-algebras $\mathscr{F}(t), t \geq 0$, satisfying:

1. *(Information accumulates)* For $0 \leq s < t$, every set in $\mathscr{F}(s)$ is also in $\mathscr{F}(t)$. In other words, there is at least as much information available at the later time $\mathscr{F}(t)$ as there is at the earlier time $\mathscr{F}(s)$. .
2. *(Adaptivity)* For each $t \geq 0$, the Brownian motion $W(t)$ at time t is $\mathscr{F}(t)-$measurable. In other words, the information available at time t is sufficient to evaluate the Brownian motion $W(t)$ at that time.
3. *(Independence of future increments)* For $0 \leq t < u$, the increment $W(u) - W(t)$ is independent of $\mathscr{F}(t)$. In other words, any increment of the Brownian motion after time t is independent of the information available at time t.

Exercise 2. Show that W admits the following properties.

1. For any $t, s \geq 0$, the covariance of $W(s)$ and $W(t)$ is given by

$$\mathbb{E}[W(t)W(s)] = t \wedge s. \tag{2}$$

2. The moment-generating function for W (i.e., for the m-dimensional random vector $(W(t_1), W(t_2), \ldots, W(t_m)))$ is

$$\begin{aligned}\varphi(u_1, u_2, \ldots, u_m) =& \mathbb{E}\exp\{u_m W(t_m) + u_{m-1} W(t_{m-1}) + \cdots + u_1 W(t_1)\} \\ =& \exp\left\{\frac{1}{2}(u_1 + u_2 + \cdots + u_m)^2 t_1 + \frac{1}{2}(u_2 + u_3 + \cdots + u_m)^2 (t_2 - t_1) + \right. \\ & \left. \cdots + \frac{1}{2}(u_{m-1} + u_m)^2 (t_{m-1} - t_{m-2}) + \frac{1}{2} u_m^2 (t_m - t_{m-1})\right\}\end{aligned} \tag{3}$$

3. *(Martingale property)* For any $t \geq s \geq 0$,

$$\mathbb{E}[W(t)|\mathscr{F}(s)] = W(s). \tag{4}$$

Remark 1. If W is continuous, then (2) and (3) can be shown equivalent to the normal independent increment property and, therefore, can both be used as equivalent definitions of BM.

3 Quadratic variation

For Brownian motion, there is no natural step size. If we are given $T > 0$ we could simply choose a step size, say $\frac{T}{n}$ for some large n, and compute the quadratic variation up to time T with this step size. In other words, we could compute the sample quadratic variation along Π

$$[W, W; \Pi](T) := \sum_{j=1}^{n} \left[W(t_j) - W(t_{j-1})\right]^2,$$

where $\Pi := \{t_0, t_1, \ldots, t_n\}$ with $t_j = \frac{jT}{n}$ is a partition of $[0, T]$. We are interested in this quantity for small step sizes, and so as a last step we could evaluate the limit as $n \to \infty$. If we do this, we will get T, the same final answer as for the scaled random walk in $W^{(n)}$.

Definition 3. Let $f(t)$ be a function defined for $0 \leq t \leq T$. The quadratic variation of f up to time T is

$$[f, f](T) = \lim_{\|\Pi\| \to 0} [f, f; \Pi](T) = \lim_{\|\Pi\| \to 0} \sum_{j=1}^{n} \left[f(t_j) - f(t_{j-1})\right]^2$$

where $\Pi = \{t_0, t_1, \ldots, t_n\}$ is a (nested) partition of $[0, T]$ such that $0 = t_0 < t_1 < \cdots < t_n = T$ and $||\Pi|| = \max_{1 \leq j \leq n}(t_j - t_{j-1})$.

Theorem 2. *Let W be a BM. Then, $[W,W](T)=T$ in L^2, i.e.*

$$\lim_{\|\Pi\|\to 0} \mathbb{E}([W,W;\Pi](T)-T)^2 = 0. \tag{5}$$

Moreover, if the partition Π is dyadic, i.e. $t_j = \frac{jT}{2^n}$, then $[W,W](T)=T$, a.s.

Proof. We first show the convergence in (5). To this end, we shall prove that $\mathbb{E}([W,W;\Pi]) = T$ and its variance $Var([W,W;\Pi])$ converges to zero. Hence, (5) is established (why?).

The sampled quadratic variation is the sum of independent random variables. Therefore, its mean and variance are the sums of the means and variances of these random variables. We have

$$\mathbb{E}\left[\left(W\left(t_{j+1}\right)-W\left(t_j\right)\right)^2\right] = \mathrm{Var}\left[W\left(t_{j+1}\right)-W\left(t_j\right)\right] = t_{j+1}-t_j$$

which implies

$$\mathbb{E}([W,W,;\Pi](T)) = \sum_{j=0}^{n-1}\mathbb{E}\left[\left(W\left(t_{j+1}\right)-W\left(t_j\right)\right)^2\right] = \sum_{j=0}^{n-1}\left(t_{j+1}-t_j\right) = T$$

as desired. Moreover, The fourth moment of a normal random variable with zero mean is three times its variance squared. That is, if $\xi \sim N(0,\sigma^2)$, then $\mathbb{E}(|\xi|^4) = 3\sigma^2$. Therefore,

$$\begin{aligned}
\mathbb{E}\left[\left(W\left(t_{j+1}\right)-W\left(t_j\right)\right)^4\right] &= 3\left(t_{j+1}-t_j\right)^2, \\
\mathrm{Var}\left[\left(W\left(t_{j+1}\right)-W\left(t_j\right)\right)^2\right] &= \mathbb{E}\left[\left(W\left(t_{j+1}\right)-W\left(t_j\right)\right)^4\right] - \left(\mathbb{E}\left[\left(W\left(t_{j+1}\right)-W\left(t_j\right)\right)^2\right]\right)^2 \\
&= 2\left(t_{j+1}-t_j\right)^2,
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}\left(([W,W,;\Pi](T))\right) &= \sum_{j=0}^{n-1}\mathrm{Var}\left[\left(W\left(t_{j+1}\right)-W\left(t_j\right)\right)^2\right] = \sum_{j=0}^{n-1} 2\left(t_{j+1}-t_j\right)^2 \\
&\le \sum_{j=0}^{n-1} 2\|\Pi\|\left(t_{j+1}-t_j\right) = 2\|\Pi\| T
\end{aligned}$$

In particular, $\lim_{\|\Pi\|\to 0}\mathrm{Var}\left(([W,W,;\Pi])\right) = 0$.

For the second part, using Markov inequality, for any $\varepsilon > 0$,

$$\mathbb{P}(|[W,W,;\Pi](T)-T|^2 > \varepsilon) \le \frac{1}{\varepsilon}\mathbb{E}\left(|[W,W,;\Pi](T)-T|^2\right) \le \frac{2||\Pi||T}{\varepsilon}.$$

For the dyadic partition, $||\Pi|| = \frac{1}{2^n}$. In turn,

$$\sum_n \mathbb{P}(|[W,W,;\Pi](T)-T|^2 > \varepsilon) \le \frac{2T}{\varepsilon}\sum_n \frac{1}{2^n} < \infty.$$

Then, Borel-Cantelli lemma implies that

$$\mathbb{P}\left(\limsup_n \{|[W,W,;\Pi](T)-T|^2 > \varepsilon\}\right) = \mathbb{P}\left(\cap_n \cup_{k\ge n} \{|[W,W,;\Pi](T)-T|^2 > \varepsilon\}\right) = 0.$$

Hence,

$$\mathbb{P}\left(\cup_n \cap_{k\ge n} \{|[W,W,;\Pi](T)-T|^2 < \varepsilon\}\right) = 1,$$

which further implies that $[W,W,;\Pi](T) \to T$, $a.s.$ □

Exercise 3. *(Other variations of BM)*

1. Show that as the number n of partition points approaches infinity and the length of the longest subinterval approaches zero, the sample first variation

$$\sum_{j=0}^{n-1} \left|W\left(t_{j+1}\right) - W\left(t_j\right)\right|$$

approaches ∞ for almost every path of the Brownian motion W.

2. Show that as the number n of partition points approaches infinity and the length of the longest subinterval approaches zero, the sample cubic variation

$$\sum_{j=0}^{n-1} \left|W\left(t_{j+1}\right) - W\left(t_j\right)\right|^3$$

approaches zero for almost every path of the Brownian motion W.

4 Markov property

Brownian motion is a Markov process.

Theorem 3. *Let $W(t), t \ge 0$, be a Brownian motion and let $\mathscr{F}(t), t \ge 0$, be a filtration for this Brownian motion. Then $W(t)$, $t \ge 0$, is a Markov process.*

Proof. we must show that whenever $0 \le s \le t$ and f is a Borel-measurable function, there is another Borel-measurable function g such that

$$\mathbb{E}[f(W(t)) \mid \mathscr{F}(s)] = g(W(s)) \tag{6}$$

To do this, we write

$$\mathbb{E}[f(W(t)) \mid \mathscr{F}(s)] = \mathbb{E}[f((W(t)-W(s)) + W(s)) \mid \mathscr{F}(s)]$$

The random variable $W(t)-W(s)$ is independent of $\mathscr{F}(s)$, and the random variable $W(s)$ is $\mathscr{F}(s)$-measurable. In order to compute the expectation on the RHS, we replace $W(s)$ by a dummy variable x to hold it constant and then take the unconditional expectation of the remaining random variable (i.e., we define $g(x)=\mathbb{E}f(W(t)-W(s)+x)$). But $W(t)-W(s)\sim N(0,t-s)$. Therefore,

$$g(x)=\frac{1}{\sqrt{2\pi(t-s)}}\int_{-\infty}^{\infty} f(w+x)e^{-\frac{w^2}{2(t-s)}}dw \tag{7}$$

Finally, if we now take the function $g(x)$ defined by (7) and replace the dummy variable x by the random variable $W(s)$ then equation (6) holds. $\square$

Remark 2. We may make the change of variable $\tau=t-s$ and $y=w+x$ in (7) to obtain

$$g(x)=\frac{1}{\sqrt{2\pi\tau}}\int_{-\infty}^{\infty} f(y)e^{-\frac{(y-x)^2}{2\tau}}dy$$

We define the transition density $p(\tau,x,y)$ for Brownian motion to be

$$p(\tau,x,y)=\frac{1}{\sqrt{2\pi\tau}}e^{-\frac{(y-x)^2}{2\tau}}$$

so that we may further rewrite (7) as

$$g(x)=\int_{-\infty}^{\infty} f(y)p(\tau,x,y)dy,$$

and (6) as

$$\mathbb{E}[f(W(t))\mid\mathscr{F}(s)]=\int_{-\infty}^{\infty} f(y)p(\tau,W(s),y)dy.$$

This equation has the following interpretation. Conditioned on the information in $\mathscr{F}(s)$(which contains all the information obtained by observing the Brownian motion up to and including time s), the conditional density of $W(t)$ is $p(\tau,W(s),y)$. This is a density in the variable y. This density is normal with mean $W(s)$ and variance $\tau=t-s$. In particular, the only information from $\mathscr{F}(s)$ that is relevant is the value of $W(s)$. The fact that only $W(s)$ is relevant is the essence of the Markov property.

Exercise 4. *(BM with drift)* For $\mu\in\mathbb{R}$, consider the Brownian motion with drift μ :

$$X(t)=\mu t+W(t)$$

1. Show that for any Borel-measurable function $f(y)$, and for any $0\leq s<t$ the function

$$g(x)=\frac{1}{\sqrt{2\pi(t-s)}}\int_{-\infty}^{\infty} f(y)\exp\left\{-\frac{(y-x-\mu(t-s))^2}{2(t-s)}\right\}dy$$

satisfies $\mathbb{E}[f(X(t)) \mid \mathscr{F}(s)] = g(X(s))$, and hence X has the Markov property. We may rewrite $g(x)$ as $g(x) = \int_{-\infty}^{\infty} f(y)p(\tau,x,y)dy$, where $\tau = t-s$ and

$$p(\tau,x,y) = \frac{1}{\sqrt{2\pi\tau}} \exp\left\{-\frac{(y-x-\mu\tau)^2}{2\tau}\right\}$$

is the transition density for Brownian motion with drift μ.

2. For $\nu \in \mathbb{R}$ and $\sigma > 0$, consider the geometric Brownian motion

$$S(t) = S(0)e^{\sigma W(t)+\nu t}$$

Set $\tau = t-s$ and

$$p(\tau,x,y) = \frac{1}{\sigma y\sqrt{2\pi\tau}} \exp\left\{-\frac{\left(\log\frac{y}{x} - \nu\tau\right)^2}{2\sigma^2\tau}\right\}$$

Show that for any Borel-measurable function $f(y)$ and for any $0 \leq s < t$, the function $g(x) = \int_0^{\infty} h(y)p(\tau,x,y)dy$ satisfies $\mathbb{E}[f(S(t)) \mid \mathscr{F}(s)] = g(S(s))$ and hence S has the Markov property and $p(\tau,x,y)$ is its transition density.

5 First passage time distribution

Let $m > 0$ be a positive number, and define the first passage time of BM to level m

$$\tau_m = \min\{t \geq 0; W(t) = m\}$$

This is the first time the Brownian motion W reaches the level m. If the Brownian motion never reaches the level m, we set $\tau_m = \infty$.

Theorem 4. *The first passage time τ_m admits the following properties.*

1. *τ_m is finite: $\mathbb{P}(\tau_m < \infty) = 1$.*
2. *τ_m has the Laplace transform: For all $\alpha > 0$,*

$$\mathbb{E}e^{-\alpha\tau_m} = e^{-m\sqrt{2\alpha}}. \tag{8}$$

3. *τ_m is not integrable: $\mathbb{E}[\tau_m] = \infty$.*

Proof. We study τ_m using the exponential martingale

$$Z(t) := \exp\{\sigma W(t) - \frac{\sigma^2}{2}t^2\},$$

where $\sigma > 0$. *(Verify that Z is a martingale.)*

A martingale that is stopped at a stopping time is still a martingale (why?), and thus must have constant expectation. Hence, $Z(t \wedge \tau_m)$, $t \geq 0$, is a martingale:

$$1 = Z(0) = \mathbb{E}Z(t \wedge \tau_m) = \mathbb{E}\left[\exp\left\{\sigma W(t \wedge \tau_m) - \frac{1}{2}\sigma^2 (t \wedge \tau_m)\right\}\right]. \tag{9}$$

Note that the Brownian motion is always at or below level m for $t \le \tau_m$ and so

$$0 \le \exp\{\sigma W(t \wedge \tau_m)\} \le e^{\sigma m}.$$

1. *For the term* $\exp\left\{-\frac{1}{2}\sigma^2 (t \wedge \tau_m)\right\}$, if $\tau_m < \infty$, it is equal to $\exp\left\{-\frac{1}{2}\sigma^2 \tau_m\right\}$ for large enough t. On the other hand, if $\tau_m = \infty$, then the term $\exp\left\{-\frac{1}{2}\sigma^2 (t \wedge \tau_m)\right\}$ is equal to $\exp\left\{-\frac{1}{2}\sigma^2 t\right\}$, and as $t \to \infty$, this converges to zero. We capture these two cases by writing

$$\lim_{t\to\infty} \exp\left\{-\frac{1}{2}\sigma^2 (t \wedge \tau_m)\right\} = \mathbb{I}_{\{\tau_m < \infty\}} \exp\left\{-\frac{1}{2}\sigma^2 \tau_m\right\}.$$

2. *For the term* $\exp\{\sigma W(t \wedge \tau_m)\}$, if $\tau_m < \infty$, then

$$\exp\{\sigma W(t \wedge \tau_m)\} = \exp\{\sigma W(\tau_m)\} = e^{\sigma m},$$

when t becomes large enough. If $\tau_m = \infty$, then we do not know what happens to $\exp\{\sigma W(t \wedge \tau_m)\}$ as $t \to \infty$ but we at least know that this term is bounded by $e^{\sigma m}$. That is enough to ensure that the product of $\exp\{\sigma W(t \wedge \tau_m)\}$ and $\exp\left\{-\frac{1}{2}\sigma^2 \tau_m\right\}$ has limit zero in this case.

In conclusion, we have

$$\lim_{t\to\infty} \exp\left\{\sigma W(t \wedge \tau_m) - \frac{1}{2}\sigma^2 (t \wedge \tau_m)\right\} = \mathbb{I}_{\{\tau_m < \infty\}} \exp\left\{\sigma m - \frac{1}{2}\sigma^2 \tau_m\right\}$$

Taking the limit in (9) and using DCT yield

$$1 = \mathbb{E}\left[\mathbb{I}_{\{\tau_m < \infty\}} \exp\left\{\sigma m - \frac{1}{2}\sigma^2 \tau_m\right\}\right]$$

or, equivalently,

$$\mathbb{E}\left[\mathbb{I}_{\{\tau_m < \infty\}} \exp\left\{-\frac{1}{2}\sigma^2 \tau_m\right\}\right] = e^{-\sigma m} \tag{10}$$

Equation (10) holds when $\sigma > 0$. We may not substitute $\sigma = 0$ into this equation, but since it holds for every positive σ, we may take the limit on both sides as $\sigma \downarrow 0$(Why?). This yields $\mathbb{E}\left[\mathbb{I}_{\{\tau_m < \infty\}}\right] = 1$ or, equivalently, part (1) holds.

Because $\tau_m < \infty$, *a.s.*, we may drop the indicator of this event in (10) to obtain

$$\mathbb{E}\left[\exp\left\{-\frac{1}{2}\sigma^2 \tau_m\right\}\right] = e^{-\sigma m}.$$

Then part (2) follows by setting $\sigma = \sqrt{2\alpha}$.

Differentiation of (8) with respect to α results in

$$\mathbb{E}\left[\tau_m e^{-\alpha\tau_m}\right] = \frac{m}{\sqrt{2\alpha}} e^{-m\sqrt{2\alpha}}, \text{ for all } \alpha > 0$$

Letting $\alpha \downarrow 0$, we obtain $\mathbb{E}\tau_m = \infty$, which is part (3). □

6 Reflection principle

The properties of τ_m can also be derived using reflection principle. For $m > 0$ and $w \le m$, each Brownian motion path that reaches level m prior to time t but is at a level w at time t, there is a *reflected path* that is at level $2m - w$ at time t. This reflected path is constructed by switching the up and down moves of the Brownian motion from time τ_m onward. This leads to the key reflection equality often called the reflection principle:

$$\mathbb{P}\{\tau_m \le t, W(t) \le w\} = \mathbb{P}\{W(t) \ge 2m - w\}, \quad m > 0, w \le m.$$

Theorem 5. *For all $m > 0$, the random variable τ_m has cumulative distribution function*

$$\mathbb{P}\{\tau_m \le t\} = \frac{2}{\sqrt{2\pi}} \int_{\frac{m}{\sqrt{t}}}^{\infty} e^{-\frac{y^2}{2}} dy, \quad t \ge 0 \tag{11}$$

and density

$$f_{\tau_m}(t) = \frac{d}{dt}\mathbb{P}\{\tau_m \le t\} = \frac{m}{t\sqrt{2\pi t}} e^{-\frac{m^2}{2t}}, \quad t \ge 0 \tag{12}$$

Proof. We substitute $w = m$ into the reflection principle to obtain

$$\mathbb{P}\{\tau_m \le t, W(t) \le m\} = \mathbb{P}\{W(t) \ge m\}$$

On the other hand, if $W(t) \ge m$, then we are guaranteed that $\tau_m \le t$. In other words,

$$\mathbb{P}\{\tau_m \le t, W(t) \ge m\} = \mathbb{P}\{W(t) \ge m\}$$

Adding these two equations, we obtain the cumulative distribution function for τ_m

$$\begin{aligned}\mathbb{P}\{\tau_m \le t\} &= \mathbb{P}\{\tau_m \le t, W(t) \le m\} + \mathbb{P}\{\tau_m \le t, W(t) \ge m\} \\ &= 2\mathbb{P}\{W(t) \ge m\} = \frac{2}{\sqrt{2\pi t}} \int_m^{\infty} e^{-\frac{x^2}{2t}} dx\end{aligned}$$

We make the change of variable $y = \frac{x}{\sqrt{t}}$ in the integral, and this leads to the cdf (11) for τ_m. Finally, the density (12) is obtained by differentiating (11) with respect to t. □

Remark 3. From (12), we see that, for all $\alpha > 0$,

$$\mathbb{E}e^{-\alpha\tau_m} = \int_0^{\infty} e^{-\alpha m} f_{\tau_m}(t) dt = \int_0^{\infty} \frac{m}{t\sqrt{2\pi t}} e^{-\alpha m - \frac{m^2}{2t}} dt.$$

(Verify the above formula gives the same expression as in part (2) of Theorem 4.)

Exercise 5. Let $W(t), t \geq 0$ be a BM. Define $\bar{W}(t) = \max_{s\in[0,t]} W(s)$. For $b > 0$ and $b \geq a$, prove that $(\bar{W}(t), W(t))$ has the joint cdf

$$\mathbb{P}(\bar{W}(t) \leq b, W(t) \leq a) = N\left(\frac{a}{\sqrt{t}}\right) - N\left(\frac{a-2b}{\sqrt{t}}\right)$$

with the joint density

$$f_{\bar{W}(t),W(t)}(b,a) = \frac{2(2b-a)}{t\sqrt{2\pi t}} e^{-\frac{(2b-a)^2}{2t}}$$

and $\bar{W}(t)$ has the cdf

$$\mathbb{P}(\bar{W}(t) \leq b) = N\left(\frac{b}{\sqrt{t}}\right) - N\left(\frac{-b}{\sqrt{t}}\right)$$

with the density

$$f_{W(t)}(b) = \frac{2}{\sqrt{2\pi t}} e^{-\frac{b^2}{2t}}.$$

Chapter 3: Stochastic calculus

Gechun Liang

Itô's integrals are used to model the value of a portfolio that results from trading assets in continuous time. The calculus used to manipulate these integrals is based on Itô's formula and differs from ordinary calculus. This difference can be traced to the fact that Brownian motion has a nonzero quadratic variation and is the source of the volatility term in the Black-Scholes PDE.

1 Itô's integral for simple integrands

We fix a positive number T and seek to make sense of

$$\int_0^T \Delta(t)dW(t),$$

where $\Delta(t)$ is the position we take in the asset at time t, which is assumed to be adapted to the Brownian filtration. The problem we face when trying to assign meaning to the above integral is that Brownian motion paths cannot be differentiated with respect to time. If $g(t)$ is a differentiable function, then we can define

$$\int_0^T \Delta(t)dg(t) = \int_0^T \Delta(t)g'(t)dt$$

where the right-hand side is an ordinary (Lebesgue) integral with respect to time. This will not work for Brownian motion.

To define Itô's integrals, we start with simple integrands. Let $\Pi = \{t_0, t_1, \dots, t_n\}$ be a partition of $[0,T]$; i.e.

$$0 = t_0 \le t_1 \le \cdots \le t_n = T$$

Gechun Liang
Department of Statistics, University of Warwick, U.K. e-mail: g.liang@warwick.ac.uk

Assume that $\Delta(t)$ is constant in t on each subinterval $\left[t_j, t_{j+1}\right)$. Such a process $\Delta(t)$ is a simple process.[1]

We shall think of the interplay between the simple process $\Delta(t)$ and the Brownian motion $W(t)$ in the following way. Regard $W(t)$ as the price per share of an asset at time t. Think of $t_0, t_1, \ldots, t_{n-1}$ as the trading dates in the asset, and think of $\Delta\left(t_0\right), \Delta\left(t_1\right), \ldots, \Delta\left(t_{n-1}\right)$ as the position (number of shares) taken in the asset at each trading date and held to the next trading date. The gain from trading at each time t is given by

$$I(t)=\sum_{j=0}^{k-1} \Delta\left(t_j\right)\left[W\left(t_{j+1}\right)-W\left(t_j\right)\right]+\Delta\left(t_k\right)\left[W(t)-W\left(t_k\right)\right] \tag{1}$$

for $t \in\left[t_k, t_{k+1}\right]$. The process $I(t)$ is the Itô's integral of the simple process $\Delta(t)$, a fact that we write as

$$\int_0^t \Delta(t) d W(t):=I(t). \tag{2}$$

Theorem 1. *The Itô's integral defined by (1) is a martingale.*

Proof. Let $0 \leq s \leq t \leq T$ be given. We shall assume that s and t are in different subintervals of the partition Π (i.e., there are partition points t_ℓ and t_k such that $t_\ell<t_k, s \in\left[t_\ell, t_{\ell+1}\right)$, and $\left.t \in\left[t_k, t_{k+1}\right)\right)$. If s and t are in the same subinterval, the following proof simplifies (try it!). Equation (1) may be rewritten as

$$\begin{aligned} I(t)= & \sum_{j=0}^{\ell-1} \Delta\left(t_j\right)\left[W\left(t_{j+1}\right)-W\left(t_j\right)\right]+\Delta\left(t_\ell\right)\left[W\left(t_{\ell+1}\right)-W\left(t_\ell\right)\right] \\ & +\sum_{j=\ell+1}^{k-1} \Delta\left(t_j\right)\left[W\left(t_{j+1}\right)-W\left(t_j\right)\right]+\Delta\left(t_k\right)\left[W(t)-W\left(t_k\right)\right] . \end{aligned} \tag{3}$$

We must show that $\mathbb{E}[I(t) \mid \mathscr{F}(s)]=I(s)$.

For the first term, we have

$$\mathbb{E}\left[\sum_{j=0}^{\ell-1} \Delta\left(t_j\right)\left[W\left(t_{j+1}\right)-W\left(t_j\right)\right] \mid \mathscr{F}(s)\right]=\sum_{j=0}^{\ell-1} \Delta\left(t_j\right)\left[W\left(t_{j+1}\right)-W\left(t_j\right)\right]$$

because the latest time appearing in this sum is t_ℓ and $t_\ell \leq s$.

For the second term, by the martingale property of W, we have

$$\begin{aligned} \mathbb{E}\left[\Delta\left(t_\ell\right)\left(W\left(t_{\ell+1}\right)-W\left(t_\ell\right)\right) \mid \mathscr{F}(s)\right] & =\Delta\left(t_\ell\right)\left(\mathbb{E}\left[W\left(t_{\ell+1}\right) \mid \mathscr{F}(s)\right]-W\left(t_\ell\right)\right) \\ & =\Delta\left(t_\ell\right)\left(W(s)-W\left(t_\ell\right)\right) \end{aligned}$$

[1] Note that the simple process $\Delta(t)$ is right continuous, so it is not *predictable*. However, if we replace BM by any jump processes (say a *compensated Poisson martingale*), then we need to restrict integrands/strategies to predictable processes. From the BM case, changing the strategy at one point does not have any effect on the resulting gain of the investor, so one could allow in that case for right-continuous strategies.

Therefore, adding the first and second terms yields $I(s)$. It remains to show that the conditional expectations of the third and fourth terms on the right-hand side of (3) are zero. We will then have $\mathbb{E}[I(t) \mid \mathscr{F}(s)] = I(s)$

The summands in *the third term* are of the form $\Delta(t_j)\left[W\left(t_{j+1}\right) - W(t_j)\right]$, where $t_j \geq t_{\ell+1} > s$. Hence,

$$\begin{aligned}
&\mathbb{E}\left\{\Delta(t_j)\ \left(W\left(t_{j+1}\right) - W(t_j)\right) \mid \mathscr{F}(s)\right\} \\
&\qquad = \mathbb{E}\left\{\mathbb{E}\left[\Delta(t_j)\left(W\left(t_{j+1}\right) - W(t_j)\right) \mid \mathscr{F}(t_j)\right] \mid \mathscr{F}(s)\right\} \\
&\qquad = \mathbb{E}\left\{\Delta(t_j)\left(\mathbb{E}\left[W\left(t_{j+1}\right) \mid \mathscr{F}(t_j)\right] - W(t_j)\right) \mid \mathscr{F}(s)\right\} \\
&\qquad = \mathbb{E}\left\{\Delta(t_j)\left(W(t_j) - W(t_j)\right) \mid \mathscr{F}(s)\right\} = 0
\end{aligned}$$

Because the conditional expectation of each of the summands in the third term on the right-hand side of (3) is zero, the conditional expectation of the whole term is zero:

$$\mathbb{E}\left\{\sum_{j=\ell+1}^{k-1} \Delta(t_j)\left[W\left(t_{j+1}\right) - W(t_j)\right] \mid \mathscr{F}(s)\right\} = 0.$$

Finally, *for the fourth term,*

$$\begin{aligned}
&\mathbb{E}\{\Delta(t_k)\ (W(t) - W(t_k)) \mid \mathscr{F}(s)\} \\
&\qquad = \mathbb{E}\{\mathbb{E}[\Delta(t_k)(W(t) - W(t_k)) \mid \mathscr{F}(t_k)] \mid \mathscr{F}(s)\} \\
&\qquad = \mathbb{E}\{\Delta(t_k)(\mathbb{E}[W(t) \mid \mathscr{F}(t_k)] - W(t_k)) \mid \mathscr{F}(s)\} \\
&\qquad = \mathbb{E}\{\Delta(t_k)(W(t_k) - W(t_k)) \mid \mathscr{F}(s)\} = 0
\end{aligned}$$

This concludes the proof. □

Because $I(t)$ is a martingale and $I(0) = 0$, we have $\mathbb{E}I(t) = 0$ for all $t \geq 0$ It follows that $\operatorname{Var} I(t) = \mathbb{E}I^2(t)$, a quantity that can be evaluated by the Itô's isometry below.

Theorem 2. *(Itô's isometry). The Itô's integral defined by (1) satisfies*

$$\mathbb{E}I^2(t) = \mathbb{E}\int_0^t \Delta^2(u)\,du$$

Proof. To simplify the notation, we set the Brownian increment $D_j = W\left(t_{j+1}\right) - W(t_j)$ for $j = 0, \ldots, k-1$ and $D_k = W(t) - W(t_k)$ so that (1) may be written as $I(t) = \sum_{j=0}^{k} \Delta(t_j) D_j$ and

$$I^2(t) = \sum_{j=0}^{k} \Delta^2(t_j) D_j^2 + 2 \sum_{0 \leq i < j \leq k} \Delta(t_i) \Delta(t_j) D_i D_j$$

We first show that the expected value of each of the cross terms is zero. For $i < j$, the random variable $\Delta(t_i)\Delta(t_j) D_i$ is $\mathscr{F}(t_j)$-measurable, while the Brownian increment D_j is independent of $\mathscr{F}(t_j)$. Furthermore, $\mathbb{E}D_j = 0$. Therefore,

$$\mathbb{E}\left[\Delta(t_i)\Delta(t_j)D_iD_j\right]=\mathbb{E}\left[\Delta(t_i)\Delta(t_j)D_i\right]\cdot\mathbb{E}D_j=\mathbb{E}\left[\Delta(t_i)\Delta(t_j)D_i\right]\cdot 0=0$$

We next consider the square terms $\Delta^2(t_j)D_j^2$. The random variable $\Delta^2(t_j)$ is $\mathscr{F}(t_j)$-measurable, and the squared Brownian increment D_j^2 is independent of $\mathscr{F}(t_j)$. Furthermore, $\mathbb{E}D_j^2=t_{j+1}-t_j$ for $j=0,\dots,k-1$ and $\mathbb{E}D_k^2=t-t_k$. Therefore,

$$\begin{aligned}\mathbb{E}I^2(t)&=\sum_{j=0}^{k}\mathbb{E}\left[\Delta^2(t_j)D_j^2\right]=\sum_{j=1}^{k}\mathbb{E}\Delta^2(t_j)\cdot\mathbb{E}D_j^2\\&=\sum_{j=0}^{k-1}\mathbb{E}\Delta^2(t_j)\left(t_{j+1}-t_j\right)+\mathbb{E}\Delta^2(t_k)(t-t_k)\end{aligned}\tag{4}$$

But $\Delta(t_j)$ is constant on the interval $\left[t_j,t_{j+1}\right)$, and hence $\Delta^2(t_j)\left(t_{j+1}-t_j\right)=\int_{t_j}^{t_{j+1}}\Delta^2(u)du$. Similarly, $\Delta^2(t_k)(t-t_k)=\int_{t_k}^{t}\Delta^2(u)du$. We may thus continue (4) to obtain

$$\begin{aligned}\mathbb{E}I^2(t)&=\sum_{j=0}^{k-1}\mathbb{E}\int_{t_j}^{t_{j+1}}\Delta^2(u)du+\mathbb{E}\int_{t_k}^{t}\Delta^2(u)du\\&=\mathbb{E}\left[\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}}\Delta^2(u)du+\int_{t_k}^{t}\Delta^2(u)du\right]=\mathbb{E}\int_0^t\Delta^2(u)du.\ \square\end{aligned}$$

For the quadratic variation of the Itô's integral $I(t)$, although Brownian motion accumulates quadratic variation at rate one per unit time, it is scaled in a time- and path-dependent way by the integrand $\Delta(u)$ as it enters the Itô's integral $I(t)=\int_0^t\Delta(u)dW(u)$. Because increments are squared in the computation of quadratic variation, the quadratic variation of Brownian motion will be scaled by $\Delta^2(u)$ as it enters the Itô's integral. The following theorem gives the precise statement.

Theorem 3. *(Quadratic variation) The quadratic variation accumulated up to time t by the Itô's integral (1) is*

$$[I,I](t)=\int_0^t\Delta^2(u)du$$

Proof. We first compute the quadratic variation accumulated by the Itô's integral on one of the subintervals $\left[t_j,t_{j+1}\right]$ on which $\Delta(u)$ is constant. For this, we choose partition points

$$t_j=s_0<s_1<\cdots<s_m=t_{j+1}$$

and consider

$$\begin{aligned}\sum_{i=0}^{m-1}\left[I\left(s_{i+1}\right)-I\left(s_{i}\right)\right]^{2} &= \sum_{i=0}^{m-1}\left[\Delta\left(t_{j}\right)\left(W\left(s_{i+1}\right)-W\left(s_{i}\right)\right)\right]^{2} \\ &= \Delta^{2}\left(t_{j}\right)\sum_{i=0}^{m-1}\left(W\left(s_{i+1}\right)-W\left(s_{i}\right)\right)^{2} \qquad (5)\end{aligned}$$

As $m \to \infty$ and the step size $\max_{i=0,\ldots,m-1}\left(s_{i+1}-s_{i}\right)$ approaches zero, the term $\sum_{i=0}^{m-1}\left(W\left(s_{i+1}\right)-W\left(s_{i}\right)\right)^{2}$ converges to the quadratic variation accumulated by Brownian motion between times t_j and t_{j+1}, which is $t_{j+1}-t_j$. Therefore, the limit of (5), which is the quadratic variation accumulated by the Itô's integral between times t_j and t_{j+1}, is

$$\Delta^{2}\left(t_{j}\right)\left(t_{j+1}-t_{j}\right)=\int_{t_{j}}^{t_{j+1}} \Delta^{2}(u) du$$

where again we have used the fact that $\Delta(u)$ is constant for $t_j \leq u < t_{j+1}$. Analogously, the quadratic variation accumulated by the Itô's integral between times t_k and t is $\int_{t_k}^{t} \Delta^{2}(u) du$. Adding up all these pieces, we obtain the desired result. □

Remark 1. In the rest of the course, we will use the following notations synonymously,

$$I(t)=\int_{0}^{t} \Delta(u) dW(u),$$

or in differential form

$$dI(t)=\Delta(t) dW(t).$$

Exercise 1. Let $W(t), 0 \leq t \leq T$, be a Brownian motion, and let $\mathscr{F}(t)$ $0 \leq t \leq T$, be an associated filtration. Let $\Delta(t), 0 \leq t \leq T$, be a *non-random simple process* (i.e., there is a partition $\Pi=\left\{t_{0}, t_{1}, \ldots, t_{n}\right\}$ of $[0, T]$ such that for every $j, \Delta\left(t_{j}\right)$ is a non-random quantity and $\Delta(t)=\Delta\left(t_{j}\right)$ is constant in t on the subinterval $\left[t_{j}, t_{j+1}\right)\big)$. For $t \in\left[t_{k}, t_{k+1}\right]$, define the stochastic integral

$$I(t)=\sum_{j=0}^{k-1} \Delta\left(t_{j}\right)\left[W\left(t_{j+1}\right)-W\left(t_{j}\right)\right]+\Delta\left(t_{k}\right)\left[W(t)-W\left(t_{k}\right)\right]$$

1. Show that whenever $0 \leq s<t \leq T$, the increment $I(t)-I(s)$ is independent of $\mathscr{F}(s)$.
2. Show that whenever $0 \leq s<t \leq T$, the increment $I(t)-I(s)$ is a normally distributed random variable with mean zero and variance $\int_{s}^{t} \Delta^{2}(u) du$.
3. Use (i) and (ii) to show that $I(t), 0 \leq t \leq T$, is a martingale.
4. Show that $I^{2}(t)-\int_{0}^{t} \Delta^{2}(u) du, 0 \leq t \leq T$, is a martingale.

2 Itô's integral for general integrands

In this section, we define the Itô's integral $\int_0^T \Delta(t)dW(t)$ for integrands $\Delta(t)$ that are allowed to *vary continuously with time and also to jump*. In particular, we no longer assume that $\Delta(t)$ is a simple process.

Assumption 1 *The integrand process $\Delta(t), t \geq 0$, is adapted to the filtration $\mathscr{F}(t), t \geq 0$, and also square integrable*

$$\mathbb{E}\int_0^T \Delta^2(t)dt < \infty.$$

The intuitive idea of how to construct a general Itô's integral. The idea is to choose a sequence $\Delta_n(t)$ of simple processes such that as $n \to \infty$, these processes converge to the continuously varying $\Delta(t)$ in L^2:

$$\lim_{n\to\infty} \mathbb{E}\int_0^T |\Delta_n(t) - \Delta(t)|^2 dt = 0 \tag{6}$$

For each $\Delta_n(t)$, the Itô's integral $\int_0^t \Delta_n(u)dW(u)$ has already been defined for $0 \leq t \leq T$. Note that $I_n(t) = \int_0^t \Delta_n(u)dW(u)$ is a Cauchy sequence in L^2. This is because of Itô's isometry in Theorem 2, which yields

$$\mathbb{E}\left(I_n(t) - I_m(t)\right)^2 = \mathbb{E}\int_0^t |\Delta_n(u) - \Delta_m(u)|^2 du.$$

As a consequence of (6), the right-hand side has limit zero as n and m approach infinity. Hence, we may define the Itô's integral for the continuously varying integrand $\Delta(t)$ as the limit of the simple Itô's integral in L^2:

$$\int_0^t \Delta(u)dW(u) := \lim_{n\to\infty}\int_0^t \Delta_n(u)dW(u), \quad 0 \leq t \leq T. \tag{7}$$

Fortunately, this integral inherits the properties of Itô's integrals of simple processes. We summarize these in the next theorem, whose proof is omitted. See Chapter 3 of *Brownian Motion and Stochastic Calculus* by Karatzas and Shreve for the proof.

Theorem 4. *Let T be a positive constant and let $\Delta(t), 0 \leq t \leq T$, satisfy Assumption 1. Then $I(t) = \int_0^t \Delta(u)dW(u)$ defined by (7) has the following properties.*

1. *<u>(Continuity)</u> As a function of the upper limit of integration the paths of $I(t)$ are continuous.*
2. *<u>(Adaptivity)</u> For each $t, I(t)$ is $\mathscr{F}(t)$-measurable.*
3. *<u>(Linearity)</u> If $I(t) = \int_0^t \Delta(u)dW(u)$ and $J(t) = \int_0^t \Gamma(u)dW(u)$, then*

$$I(t) \pm J(t) = \int_0^t (\Delta(u) \pm \Gamma(u))dW(u);$$

furthermore, for every constant c,

$$cI(t) = \int_0^t c\Delta(u)dW(u).$$

4. *(Martingale) $I(t)$ is a martingale.*
5. *(Itô's isometry)*

$$\mathbb{E}I^2(t) = \mathbb{E}\int_0^t \Delta^2(u)du$$

6. *(Quadratic variation)*

$$[I,I](t) = \int_0^t \Delta^2(u)du.$$

Remark 2. We have defined the Itô's integral $\int_0^T \Delta(t)dW(t)$ under the condition

$$\mathbb{E}\int_0^T \Delta^2(t)dt < \infty$$

The integral can be defined under the weaker condition

$$\int_0^T \Delta^2(t) < \infty \quad \text{a.s.}$$

but then is not guaranteed to be a martingale. It is still a local martingale satisfying the properties (1)-(3) and (6) in Theorem 4. We again refer to the textbook by Karatzas and Shreve.

Exercise 2. *(Stratonovich integral).* Let $W(t), t \geq 0$, be a Brownian motion. Let T be a fixed positive number and let $\Pi = \{t_0, t_1, \ldots, t_n\}$ be a partition of $[0,T]$ (i.e., $0 = t_0 < t_1 < \cdots < t_n = T$). For each j, define $t_j^* = \frac{t_j + t_{j+1}}{2}$ to be the midpoint of the interval $\left[t_j, t_{j+1}\right]$.

1. Define the half-sample quadratic variation corresponding to Π to be

$$[W,W;\Pi/2](T) = \sum_{j=0}^{n-1}\left(W\left(t_j^*\right) - W\left(t_j\right)\right)^2$$

   Show that $[W,W;\Pi/2](T)$ has limit $\frac{1}{2}T$ as $\|\Pi\| \to 0$. (Hint: It suffices to show that $\mathbb{E}Q_{\Pi/2} = \frac{1}{2}T$ and $\lim_{\|\Pi\|\to 0} \operatorname{Var}\left(Q_{\Pi/2}\right) = 0$.)
2. Define the Stratonovich integral of $W(t)$ with respect to $W(t)$ to be

$$\int_0^T W(t) \circ dW(t) = \lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} W\left(t_j^*\right)\left(W\left(t_{j+1}\right) - W\left(t_j\right)\right)$$

   In contrast to the Itô's integral $\int_0^T W(t)dW(t) = \frac{1}{2}W^2(T) - \frac{1}{2}T$ of which evaluates the integrand at the left endpoint of each subinterval $\left[t_j, t_{j+1}\right]$, here we evaluate the integrand at the midpoint t_j^*. Show that

$$\int_0^T W(t) \circ dW(t) = \frac{1}{2} W^2(T)$$

3 Itô's formula

In this section, we provide an intuitive proof for Itô's formula/chain rule[2]. *We first establish a chain rule for $f(t,W(t))$.*

Theorem 5. *(Itô's formula for Brownian motion). Let $f(t,x)$ be a function for which the partial derivatives $f_t(t,x), f_x(t,x)$, and $f_{xx}(t,x)$ are defined and continuous, and let $W(t)$ be a Brownian motion. Then, for every $T \geq 0$,*

$$\begin{aligned} f(T,W(T)) &= f(0,W(0)) + \int_0^T f_t(t,W(t))dt \\ &\quad + \int_0^T f_x(t,W(t))dW(t) + \frac{1}{2}\int_0^T f_{xx}(t,W(t))dt \end{aligned} \tag{8}$$

Proof. We first show why (8) holds when $f(x) = \frac{1}{2}x^2$. In this case, $f'(x) = x$ and $f''(x) = 1$. Let x_{j+1} and x_j be numbers. Taylor's formula implies

$$f\left(x_{j+1}\right) - f\left(x_j\right) = f'\left(x_j\right)\left(x_{j+1} - x_j\right) + \frac{1}{2} f''\left(x_j\right)\left(x_{j+1} - x_j\right)^2 \tag{9}$$

Fix $T > 0$, and let $\Pi = \{t_0, t_1, \ldots, t_n\}$ be a partition of $[0,T]$, i.e.,

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

We are interested in the difference between $f(W(0))$ and $f(W(T))$. This change in $f(W(t))$ between times $t = 0$ and $t = T$ can be written as the sum of the changes in $f(W(t))$ over each of the subintervals $\left[t_j, t_{j+1}\right]$. We do this and then use Taylor's formula (9) with $x_j = W\left(t_j\right)$ and $x_{j+1} = W\left(t_{j+1}\right)$ to obtain

$$\begin{aligned} f(W(T)) - f(W(0)) &= \sum_{j=0}^{n-1}\left[f\left(W\left(t_{j+1}\right)\right) - f\left(W\left(t_j\right)\right)\right] \\ &= \sum_{j=0}^{n-1} f'\left(W\left(t_j\right)\right)\left[W\left(t_{j+1}\right) - W\left(t_j\right)\right] \\ &\quad + \frac{1}{2}\sum_{j=0}^{n-1} f''\left(W\left(t_j\right)\right)\left[W\left(t_{j+1}\right) - W\left(t_j\right)\right]^2 \end{aligned} \tag{10}$$

For the function $f(x) = \frac{1}{2}x^2$, the right-hand side of the above equation (10) is

[2] For a more rigorous treatment, please refer to the textbook *Brownian Motion and Stochastic Calculus* by Karatzas and Shreve. See Chapter 3 therein.

$$\sum_{j=0}^{n-1} W(t_j)\left[W\left(t_{j+1}\right)-W(t_j)\right]+\frac{1}{2}\sum_{j=0}^{n-1}\left[W\left(t_{j+1}\right)-W(t_j)\right]^2$$

Sending $\|\Pi\| \to 0$ yields the Itô's formula for $f(x)=\frac{1}{2}x^2$, i.e.

$$\begin{aligned}
&f(W(T))-f(W(0)) \\
=&\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1} W(t_j)\left[W\left(\iota_{j+1}\right)-W(t_j)\right]+\lim_{\|\Pi\|\to 0}\frac{1}{2}\sum_{j=0}^{n-1}\left[W\left(t_{j+1}\right)-W(t_j)\right]^2 \\
=&\int_0^T W(t)dW(t)+\frac{1}{2}T \\
=&\int_0^T f'(W(t))dW(t)+\frac{1}{2}\int_0^T f''(W(t))dt.
\end{aligned} \tag{11}$$

In general, if we take a function $f(t,x)$ of both the time variable t and the variable x, then Taylor's Theorem says that

$$\begin{aligned}
&f\left(t_{j+1},x_{j+1}\right)-f(t_j,x_j) \\
=&f_t(t_j,x_j)\left(t_{j+1}-t_j\right)+f_x(t_j,x_j)\left(x_{j+1}-x_j\right) \\
&+\frac{1}{2}f_{xx}(t_j,x_j)\left(x_{j+1}-x_j\right)^2+f_{tx}(t_j,x_j)\left(t_{j+1}-t_j\right)\left(x_{j+1}-x_j\right) \\
&+\frac{1}{2}f_{tt}(t_j,x_j)\left(t_{j+1}-t_j\right)^2+\text{ higher-order terms.}
\end{aligned} \tag{12}$$

We replace x_j by $W(t_j)$, replace x_{j+1} by $W\left(t_{j+1}\right)$, and sum:

$$\begin{aligned}
&f(T,W(T))-f(0,W(0)) \\
=&\sum_{j=0}^{n-1}\left[f\left(t_{j+1},W\left(t_{j+1}\right)\right)-f(t_j,W(t_j))\right] \\
=&\sum_{j=0}^{n-1} f_t(t_j,W(t_j))\left(t_{j+1}-t_j\right)+\sum_{j=0}^{n-1} f_x(t_j,W(t_j))\left(W\left(t_{j+1}\right)-W(t_j)\right) \\
&+\frac{1}{2}\sum_{j=0}^{n-1} f_{xx}(t_j,W(t_j))\left(W\left(t_{j+1}\right)-W(t_j)\right)^2 \\
&+\sum_{j=0}^{n-1} f_{tx}(t_j,W(t_j))\left(t_{j+1}-t_j\right)\left(W\left(t_{j+1}\right)-W(t_j)\right) \\
&+\frac{1}{2}\sum_{j=0}^{n-1} f_{tt}(t_j,W(t_j))\left(t_{j+1}-t_j\right)^2+\text{ higher-order terms.}
\end{aligned} \tag{13}$$

When we take the limit as $\|\Pi\| \to 0$, the *first term* on the right-hand side of (13) contributes the ordinary (Lebesgue) integral

$$\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1} f_t\left(t_j, W\left(t_j\right)\right)\left(t_{j+1}-t_j\right)=\int_0^T f_t(t,W(t))dt.$$

As $\|\Pi\|\to 0$, the *second term* contributes the Itô's integral $\int_0^T f_x(t,W(t))dW(t)$.

The *third term* contributes another Lebesgue integral, $\frac{1}{2}\int_0^T f_{xx}(t,W(t))dt$, similar to the way we obtained this integral in (11). In other words, in the third term we can replace $\left(W\left(t_{j+1}\right)-W\left(t_j\right)\right)^2$ by $t_{j+1}-t_j$. This is not an exact substitution, but when we sum the terms this substitution gives the correct limit as $\|\Pi\|\to 0$. With this substitution, the third term on the right-hand side of (13) contributes $\frac{1}{2}\int_0^T f_{xx}(t,W(t))dt$.

For the *fourth term*, we observe that

$$\begin{aligned}
&\lim_{\|\Pi\|\to 0}\left|\sum_{j=0}^{n-1} f_{tx}\left(t_j, W\left(t_j\right)\right)\left(t_{j+1}-t_j\right)\left(W\left(t_{j+1}\right)-W\left(t_j\right)\right)\right| \\
\leq &\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}\left|f_{tx}\left(t_j, W\left(t_j\right)\right)\right|\cdot\left(t_{j+1}-t_j\right)\cdot\left|W\left(t_{j+1}\right)-W\left(t_j\right)\right| \\
\leq &\lim_{\|\Pi\|\to 0}\max_k\left|W\left(t_{k+1}\right)-W\left(t_k\right)\right|\cdot\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}\left|f_{tx}\left(t_j, W\left(t_j\right)\right)\right|\left(t_{j+1}-t_j\right) \\
=&0\cdot\int_0^T \mid f_{tx}(t,W(t))dt=0
\end{aligned}$$

The *fifth term* is treated similarly:

$$\begin{aligned}
&\lim_{\|\Pi\|\to 0}\left|\frac{1}{2}\sum_{j=0}^{n-1} f_{tt}\left(t_j, W\left(t_j\right)\right)\left(t_{j+1}-t_j\right)^2\right| \\
\leq &\lim_{\|\Pi\|\to 0}\frac{1}{2}\sum_{j=0}^{n-1}\left|f_{tt}\left(t_j, W\left(t_j\right)\right)\right|\cdot\left(t_{j+1}-t_j\right)^2 \\
\leq &\frac{1}{2}\lim_{\|\Pi\|\to 0}\max_k(t_{k+1}-t_k)\sum_{j=0}^{n-1}\left|f_{tt}\left(t_j, W\left(t_j\right)\right)\right|\left(t_{j+1}-t_j\right) \\
=&\frac{1}{2}\cdot 0\cdot\int_0^T f_{tt}(t,W(t))dt=0.
\end{aligned}$$

The higher-order terms likewise contribute zero to the final answer. □

Remark 3. A convenient way to write Itô's formula is to write it in differential form

$$df(t,W(t))=f_t(t,W(t))dt+f_x(t,W(t))dW(t)+\frac{1}{2}f_{xx}(t,W(t))dt.$$

and informally, we also write

$$dW(t)dW(t)=dt,\quad dtdW(t)=dW(t)dt=0,\quad dtdt=0.$$

Exercise 3. In this exercise, we develop a correct explanation for the equation

$$\lim_{\|\Pi\| \to 0} \sum_{j=0}^{n-1} f''(W(t_j)) \left[W(t_{j+1}) - W(t_j)\right]^2 = \int_0^T f''(W(t))dt, \tag{14}$$

i.e. the third term in (13).

Define

$$Z_j = f''(W(t_j)) \left[\left(W(t_{j+1}) - W(t_j)\right)^2 - \left(t_{j+1} - t_j\right) \right]$$

so that

$$\sum_{j=0}^{n-1} f''(W(t_j)) \left[W(t_{j+1}) - W(t_j)\right]^2 = \sum_{j=0}^{n-1} Z_j + \sum_{j=0}^{n-1} f''(W(t_j)) \left(t_{j+1} - t_j\right) \tag{15}$$

1. Show that Z_j is $\mathscr{F}(t_{j+1})$-measurable and

$$\mathbb{E}[Z_j \mid \mathscr{F}(t_j)] = 0, \quad \mathbb{E}\left[Z_j^2 \mid \mathscr{F}(t_j)\right] = 2\left[f''(W(t_j))\right]^2 \left(t_{j+1} - t_j\right)^2$$

It remains to show that

$$\lim_{\|\Pi\| \to 0} \sum_{j=0}^{n-1} Z_j = 0 \tag{16}$$

This will cause us to obtain (14) when we take the limit in (15). Prove (16) in the following steps.

2. Show that $\mathbb{E} \sum_{j=0}^{n-1} Z_j = 0$
3. Under the assumption that $\mathbb{E} \int_0^T [f''(W(t))]^2 dt$ is finite, show that

$$\lim_{\|\Pi\| \to 0} \operatorname{Var}\left[\sum_{j=0}^{n-1} Z_j\right] = 0$$

From part (3), we conclude that $\sum_{j=0}^{n-1} Z_j$ converges to its mean, which by part (2) is zero. This establishes (16).

Next, we extend Itô's formula/chain rule to stochastic processes more general than Brownian motion.

Definition 1. Let $W(t), t \geq 0$, be a Brownian motion, and let $\mathscr{F}(t)$ $t \geq 0$, be an associated filtration. An Itô's process is a stochastic process of the form

$$X(t) = X(0) + \int_0^t \Delta(u) dW(u) + \int_0^t \Theta(u) du \tag{17}$$

where $X(0)$ is non-random and $\Delta(u)$ and $\Theta(u)$ are adapted stochastic processes satisfying integrability conditions

$$\mathbb{E}[\int_0^t |\Delta(u)|^2 du] < \infty, \quad \int_0^t |\Theta(u)| du < \infty$$

for any $t \geq 0$.

Lemma 1. *The quadratic variation of the Itô's process (17) is*

$$[X,X](t)=\int_0^t \Delta^2(u)du$$

Proof. We introduce the notation $I(t)=\int_0^t \Delta(u)dW(u), R(t)=\int_0^t \Theta(u)du$ Both these processes are continuous in their upper limit of integration t. To determine the quadratic variation of X on $[0,t]$, we choose a partition $\Pi=\{t_0,t_1,\ldots,t_n\}$ of $[0,t]$ (i.e., $0=t_0<t_1<\cdots<t_n=t$) and we write the sampled quadratic variation

$$\begin{aligned}\sum_{j=0}^{n-1}\left[X\left(t_{j+1}\right)-X\left(t_j\right)\right]^2=&\sum_{j=0}^{n-1}\left[I\left(t_{j+1}\right)-I\left(t_j\right)\right]^2+\sum_{j=0}^{n-1}\left[R\left(t_{j+1}\right)-R\left(t_j\right)\right]^2\\&+2\sum_{j=0}^{n-1}\left[I\left(t_{j+1}\right)-I\left(t_j\right)\right]\left[R\left(t_{j+1}\right)-R\left(t_j\right)\right]\end{aligned}$$

As $\|\Pi\|\to 0$, the first term on the right-hand side, $\sum_{j=0}^{n-1}\left[I\left(t_{j+1}\right)-I\left(t_j\right)\right]^2$ converges to the quadratic variation of I on $[0,t]$, which according to Theorem 4 part (6), is $[I,I](t)=\int_0^t \Delta^2(u)du$. The absolute value of the second term is bounded above by

$$\begin{aligned}&\max_{0\leq k\leq n-1}\left|R\left(t_{k+1}\right)-R\left(t_k\right)\right|\cdot\sum_{j=0}^{n-1}\left|R\left(t_{j+1}\right)-R\left(t_j\right)\right|\\&=\max_{0\leq k\leq n-1}\left|R\left(t_{k+1}\right)-R\left(t_k\right)\right|\cdot\sum_{j=0}^{n-1}\left|\int_{t_j}^{t_{j+1}}\Theta(u)du\right|\\&\leq\max_{0\leq k\leq n-1}\left|R\left(t_{k+1}\right)-R\left(t_k\right)\right|\cdot\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}|\Theta(u)|du\\&=\max_{0\leq k\leq n-1}\left|R\left(t_{k+1}\right)-R\left(t_k\right)\right|\cdot\int_0^t|\Theta(u)|du\end{aligned}$$

and as $\|\Pi\|\to 0$, this has limit $0\cdot\int_0^t|\Theta(u)|du=0$ because $R(t)$ is continuous. The absolute value of the third term is bounded above by

$$\begin{aligned}&2\max_{0\leq k\leq n-1}\left|I\left(t_{k+1}\right)-I\left(t_k\right)\right|\cdot\sum_{j=0}^{n-1}\left|R\left(t_{j+1}\right)-R\left(t_j\right)\right|\\&\leq 2\max_{0\leq k\leq n-1}\left|I\left(t_{k+1}\right)-I\left(t_k\right)\right|\cdot\int_0^t|\Theta(u)|du\end{aligned}$$

and this has limit $0\cdot\int_0^t|\Theta(u)|^2du=0$ as $\|\Pi\|\to 0$ because $I(t)$ is continuous. We conclude that $[X,X](t)=[I,I](t)=\int_0^t \Delta^2(u)du$. □

Theorem 6. *(Itô's formula for an Itô's process). Let $X(t)$, $t\geq 0$, be an Itô's process as described in Definition 1, and let $f(t,x)$ be a function for which the partial* *derivatives $f_t(t,x), f_x(t,x)$, and $f_{xx}(t,x)$ are defined and continuous. Then, for every $T \geq 0$,*

$$\begin{aligned} f(T,X(T)) =& f(0,X(0)) + \int_0^T f_t(t,X(t))dt + \int_0^T f_x(t,X(t))dX(t) \\ &+ \frac{1}{2}\int_0^T f_{xx}(t,X(t))d[X,X](t) \\ =& f(0,X(0)) + \int_0^T f_t(t,X(t))dt + \int_0^T f_x(t,X(t))\Delta(t)dW(t) \\ &+ \int_0^T f_x(t,X(t))\Theta(t)dt + \frac{1}{2}\int_0^T f_{xx}(t,X(t))\Delta^2(t)dt. \end{aligned} \tag{18}$$

Proof. Analogous to the proof of Theorem 5, we have the decomposition

$$\begin{aligned} & f(T,X(T)) - f(0,X(0)) \\ =& \sum_{j=0}^{n-1} f_t\left(t_j, X\left(t_j\right)\right)\left(t_{j+1}-t_j\right) + \sum_{j=0}^{n-1} f_x\left(t_j, X\left(t_j\right)\right)\left(X\left(t_{j+1}\right) - X\left(t_j\right)\right) \\ &+ \frac{1}{2}\sum_{j=0}^{n-1} f_{xx}\left(t_j, X\left(t_j\right)\right)\left(X\left(t_{j+1}\right) - X\left(t_j\right)\right)^2 \\ &+ \sum_{j=0}^{n-1} f_{tx}\left(t_j, X\left(t_j\right)\right)\left(t_{j+1}-t_j\right)\left(X\left(t_{j+1}\right) - X\left(t_j\right)\right) \\ &+ \frac{1}{2}\sum_{j=0}^{n-1} f_{tt}\left(t_j, X\left(t_j\right)\right)\left(t_{j+1}-t_j\right)^2 + \text{ higher-order terms.} \end{aligned} \tag{19}$$

The last two sums on the right-hand side have zero limits as $\|\Pi\| \to 0$ for the same reasons the analogous terms have zero limits in the proof of Theorem 5. The higher-order terms likewise have limit zero. The limit of the first term on the right-hand side of (19) is $\int_0^T f_t(t,X(t))dt$. The limit of the second term is

$$\int_0^T f_x(t,X(t))dX(t) = \int_0^T f_x(t,X(t))\Delta(t)dW(t) + \int_0^T f_x(t,X(t))\Theta(t)dt$$

Finally, the limit of the third term on the right-hand side of (19) is

$$\frac{1}{2}\int_0^T f_{xx}(t,X(t))d[X,X](t) = \frac{1}{2}\int_0^T f_{xx}(t,X(t))\Delta^2(t)dt$$

because the Itô's process $X(t)$ accumulates quadratic variation at rate $\Delta^2(t)$ per unit time (Lemma 1). □

Example 1. (Generalized geometric Brownian motion). Let $W(t), t \geq 0$ be a Brownian motion, let $\mathscr{F}(t), t \geq 0$, be an associated filtration, and let $\alpha(t)$ and $\sigma(t)$ be adapted processes. Define the Itô's process

$$X(t) = \int_0^t \sigma(s)dW(s) + \int_0^t \left(\alpha(s) - \frac{1}{2}\sigma^2(s)\right) ds$$

Consider an asset price process given by

$$S(t) = S(0)e^{X(t)} = S(0)\exp\left\{\int_0^t \sigma(s)dW(s) + \int_0^t \left(\alpha(s) - \frac{1}{2}\sigma^2(s)\right) ds\right\}$$

where $S(0)$ is non-random and positive. We may write $S(t) = f(X(t))$, where $f(x) = S(0)e^x, f'(x) = S(0)e^x$, and $f''(x) = S(0)e^x$. According to Itô's formula

$$\begin{aligned} dS(t) = df(X(t)) &= f'(X(t))dX(t) + \frac{1}{2}f''(X(t))d[X,X](t) \\ &= S(0)e^{X(t)}dX(t) + \frac{1}{2}S(0)e^{X(t)}d[X,X](t) \\ &= S(t)dX(t) + \frac{1}{2}S(t)d[X,X](t) \\ &= \alpha(t)S(t)dt + \sigma(t)S(t)dW(t) \end{aligned}$$

The asset price $S(t)$ has instantaneous mean rate of return $\alpha(t)$ and volatility $\sigma(t)$. Both the instantaneous mean rate of return and the volatility are allowed to be time-varying and random.

Remark 4. This example includes all possible models of an asset price process that is always positive, has no jumps, and is driven by a single Brownian motion. Although the model is driven by a Brownian motion, the distribution of $S(t)$ does not need to be *log-normal* because $\alpha(t)$ and $\sigma(t)$ are allowed to be time varying and random. If α and σ are constant, we have the usual geometric Brownian motion model, and the distribution of $S(t)$ is log-normal.

Exercise 4. (Itô's integral of a deterministic integrand). Let $W(s)$ $s \geq 0$, be a Brownian motion, and let $\Delta(s)$ be a *non-random* function of time. Define $I(t) = \int_0^t \Delta(s)dW(s)$. For each $t \geq 0$, prove that the random variable $I(t)$ is *normally distributed* with expected value zero and variance $\int_0^t \Delta^2(s)ds$.

4 Multivariable stochastic calculus

Definition 2. A d-dimensional Brownian motion is a process

$$W(t) = (W_1(t), \ldots, W_d(t))$$

with the following properties:

1. Each $W_i(t)$ is a one-dimensional Brownian motion.
2. If $i \neq j$, then the processes $W_i(t)$ and $W_j(t)$ are independent.

Associated with a d-dimensional Brownian motion, we have a Brownian filtration $\mathscr{F}(t)$, $t \geq 0$, such that the following holds:

1. (Information accumulates) For $0 \leq s < t$, every set in $\mathscr{F}(s)$ is also in $\mathscr{F}(t)$.
2. (Adaptivity) For each $t \geq 0$, the random vector $W(t)$ is $\mathscr{F}(t)$-measurable.
3. (Independence of future increments) For $0 \leq t < u$, the vector of increments $W(u) - W(t)$ is independent of $\mathscr{F}(t)$.

Let $\Pi = \{t_0, \ldots, t_n\}$ be a partition of $[0,T]$. For $i \neq j$, define the *sampled cross variation* of W_i and W_j on $[0,T]$ to be

$$[W_i, W_j; \Pi] = \sum_{k=0}^{n-1} [W_i(t_{k+1}) - W_i(t_k)][W_j(t_{k+1}) - W_j(t_k)]$$

For simplicity, we write $C_\Pi = [W_i, W_j; \Pi]$.

The increments appearing on the right-hand side of the equation above are all independent of one another and all have mean zero. Therefore, $\mathbb{E}C_\Pi = 0$.

Next, we compute $\operatorname{Var}(C_\Pi)$. Note first that

$$\begin{aligned} C_\Pi^2 = &\sum_{k=0}^{n-1} [W_i(t_{k+1}) - W_i(t_k)]^2 [W_j(t_{k+1}) - W_j(t_k)]^2 \\ &+ 2 \sum_{\ell<k}^{n-1} [W_i(t_{\ell+1}) - W_i(t_\ell)][W_j(t_{\ell+1}) - W_j(t_\ell)] \\ &\cdot [W_i(t_{k+1}) - W_i(t_k)][W_j(t_{k+1}) - W_j(t_k)] \end{aligned}$$

All the increments appearing in the sum of cross-terms are independent of one another and all have mean zero. Therefore,

$$\operatorname{Var}(C_\Pi) = \mathbb{E}C_\Pi^2 = \mathbb{E}\sum_{k=0}^{n-1} [W_i(t_{k+1}) - W_i(t_k)]^2 [W_j(t_{k+1}) - W_j(t_k)]^2$$

But $[W_i(t_{k+1}) - W_i(t_k)]^2$ and $[W_j(t_{k+1}) - W_j(t_k)]^2$ are independent of one another, and each has expectation $(t_{k+1} - t_k)$. It follows that

$$\operatorname{Var}(C_\Pi) = \sum_{k=0}^{n-1} (t_{k+1} - t_k)^2 \leq \|\Pi\| \cdot \sum_{k=0}^{n-1} (t_{k+1} - t_k) = \|\Pi\| \cdot T$$

As $\|\Pi\| \to 0$, we have $\operatorname{Var}(C_\Pi) \to 0$, so C_Π converges to the constant $\mathbb{E}C_\Pi = 0$ in L^2 (and also almost surely, why?). From this, we conclude that the cross variation of W_i and W_j is

$$[W_i, W_j](T) = 0 \tag{20}$$

for $i \neq j$.

Next, we present Itô's formula for multiple processes. Let $X(t)$ and $Y(t)$ be Itô's processes, which means they are processes of the form

$$X(t) = X(0) + \int_0^t \Theta_1(u)du + \int_0^t \sigma_{11}(u)dW_1(u) + \int_0^t \sigma_{12}(u)dW_2(u)$$
$$Y(t) = Y(0) + \int_0^t \Theta_2(u)du + \int_0^t \sigma_{21}(u)dW_1(u) + \int_0^t \sigma_{22}(u)dW_2(u)$$

The integrands $\Theta_i(u)$ and $\sigma_{ij}(u)$ are assumed to be adapted processes. In differential notation, we also write

$$dX(t) = \Theta_1(t)dt + \sigma_{11}(t)dW_1(t) + \sigma_{12}(t)dW_2(t)$$
$$dY(t) = \Theta_2(t)dt + \sigma_{21}(t)dW_1(t) + \sigma_{22}(t)dW_2(t)$$

Then, we have

$$[X,X](T) = \int_0^T \left(\sigma_{11}^2(t) + \sigma_{12}^2(t)\right) dt$$

and

$$[X,Y](T) = \int_0^T \left(\sigma_{11}(t)\sigma_{21}(t) + \sigma_{12}(t)\sigma_{22}(t)\right) dt.$$

Theorem 7. *(Two-dimensional Itô's formula). Let $f(t,x,y)$ be a function whose partial derivatives $f_t, f_x, f_y, f_{xx}, f_{xy}, f_{yx}$, and f_{yy} are defined and are continuous. Let $X(t)$ and $Y(t)$ be Itô's processes as discussed above. Then,*

$$\begin{aligned}
&f(t,X(t),Y(t)) - f(0,X(0),Y(0)) \\
=&\int_0^t \left[\sigma_{11}(u) f_x(u,X(u),Y(u)) + \sigma_{21}(u) f_y(u,X(u),Y(u))\right] dW_1(u) \\
&+ \int_0^t \left[\sigma_{12}(u) f_x(u,X(u),Y(u)) + \sigma_{22}(u) f_y(u,X(u),Y(u))\right] dW_2(u) \\
&+ \int_0^t \Big[f_t(u,X(u),Y(u)) \\
&\quad + \Theta_1(u) f_x(u,X(u),Y(u)) + \Theta_2(u) f_y(u,X(u),Y(u)) \\
&\quad + \frac{1}{2}\left(\sigma_{11}^2(u) + \sigma_{12}^2(u)\right) f_{xx}(u,X(u),Y(u)) \\
&\quad + \left(\sigma_{11}(u)\sigma_{21}(u) + \sigma_{12}(u)\sigma_{22}(u)\right) f_{xy}(u,X(u),Y(u)) \\
&\quad + \frac{1}{2}\left(\sigma_{21}^2(u) + \sigma_{22}^2(u)\right) f_{yy}(u,X(u),Y(u)) \Big] du
\end{aligned}$$

Remark 5. If we do not write down the specific dynamics of X and Y, then the above two-dimensional Itô's formula simplifies to (in differential form)

$$\begin{aligned}
df(t,X,Y) = &f_t dt + f_x dX + f_y dY \\
&+ \frac{1}{2} f_{xx} d[X,X] + f_{xy} d[X,Y] + \frac{1}{2} f_{yy} d[Y,Y].
\end{aligned} \tag{21}$$

In general, for $X = (X_1, X_2, \ldots, X_n)$ and $f(t,X)$, we have

$$df(t,X) = f_t dt + \sum_{i=1}^n f_{x_i} dX_i + \frac{1}{2} \sum_{i,j=1}^n f_{x_i x_j} d[X_i, X_j]. \tag{22}$$

Example 2. (Product rule) Let $f(t,x,y) = xy$, then $f_t = 0$, $f_x = y$, $f_y = x$, $f_{xy} = 1$ and $f_{xx} = f_{yy} = 0$. Itô's formula implies that

$$d(X(t)Y(t)) = X(t)dY(t) + Y(t)dX(t) + d[X,Y](t).$$

5 Lévy's characterization of Brownian motion

A Brownian motion $W(t)$ is a martingale with continuous paths whose quadratic variation is $[W,W](t) = t$. It turns out that these conditions characterize Brownian motion in the sense of the following theorem.

Theorem 8. *(Lévy, one dimension). Let $M(t), t \geq 0$, be a martingale relative to a filtration $\mathscr{F}(t), t \geq 0$. Assume that $M(0) = 0, M(t)$ has continuous paths, and $[M,M](t) = t$ for all $t \geq 0$. Then $M(t)$ is a Brownian motion.*

Proof. Since $[M,M](t) = t$, Itô's formula yields

$$\begin{aligned} f(t,M(t)) = f(0,M(0)) &+ \int_0^t \left[f_t(s,M(s)) + \frac{1}{2} f_{xx}(s,M(s)) \right] ds \\ &+ \int_0^t f_x(s,M(s)) dM(s) \end{aligned}$$

Because $M(t)$ is a martingale, the stochastic integral $\int_0^t f_x(s,M(s))dM(s)$ is also(why?). At $t = 0$, this stochastic integral takes the value zero, and so its expectation is always zero. Taking expectations, we obtain

$$\mathbb{E} f(t,M(t)) = f(0,M(0)) + \mathbb{E} \int_0^t \left[f_t(s,M(s)) + \frac{1}{2} f_{xx}(s,M(s)) \right] ds \tag{23}$$

We fix a number u and define

$$f(t,x) = \exp\left\{ ux - \frac{1}{2}u^2 t \right\}$$

Then $f_t(t,x) = -\frac{1}{2}u^2 f(t,x), f_x(t,x) = uf(t,x)$, and $f_{xx}(t,x) = u^2 f(t,x)$ particular,

$$f_t(t,x) + \frac{1}{2} f_{xx}(t,x) = 0$$

For this function $f(t,x)$, the second term on the right-hand side of (23) is zero, and that equation becomes

$$\mathbb{E} \exp\left\{ uM(t) - \frac{1}{2}u^2 t \right\} = 1$$

In other words, we have the moment-generating function formula

$$\mathbb{E} e^{uM(t)} = e^{\frac{1}{2}u^2 t}$$

This is the moment-generating function for the normal distribution with mean zero and variance t. Hence, that is the distribution that $M(t)$ must have. □

Theorem 9. *(Lévy, two dimensions). Let $M_1(t)$ and $M_2(t), t \geq 0$ be martingales relative to a filtration $\mathscr{F}(t), t \geq 0$. Assume that for $i = 1,2$, we have $M_i(0) = 0, M_i(t)$ has continuous paths, and $[M_i, M_i](t) = t$ for all $t \geq 0$. If, in addition, $[M_1, M_2](t) = 0$ for all $t \geq 0$, then $M_1(t)$ and $M_2(t)$ are independent Brownian motions.*

Proof. By Itô's formula (21), we have

$$\begin{aligned}
&f(t, M_1(t), M_2(t)) \\
&= f(0, M_1(0), M_2(0)) + \int_0^t \left[f_t(s, M_1(s), M_2(s)) + \tfrac{1}{2} f_{xx}(s, M_1(s), M_2(s)) \right. \\
&\quad \left. + \tfrac{1}{2} f_{yy}(s, M_1(s), M_2(s)) \right] ds \\
&\quad + \int_0^t f_x(s, M_1(s), M_2(s))\, dM_1(s) + \int_0^t f_y(x, M_1(s), M_2(s))\, dM_2(s)
\end{aligned}$$

The last two terms on the right-hand side are martingales, starting at zero at time zero, and hence having expectation zero. Therefore,

$$\begin{aligned}
&\mathbb{E} f(t, M_1(t), M_2(t)) \\
&= f(0, M_1(0), M_2(0)) + \mathbb{E} \int_i^t \left[f_t(s, M_1(s), M_2(s)) + \tfrac{1}{2} f_{xx}(s, M_1(s), M_2(s)) \right. \\
&\quad \left. + \tfrac{1}{2} f_{yy}(s, M_1(s), M_2(s)) \right] ds
\end{aligned}$$

We now fix numbers u_1 and u_2 and define

$$f(t,x,y) = \exp\left\{ u_1 x + u_2 y - \frac{1}{2}\left(u_1^2 + u_2^2\right) t \right\}$$

Then

$$\begin{aligned}
f_t(t,x,y) &= -\frac{1}{2}\left(u_1^2 + u_2^2\right) f(t,x,y) \\
f_x(t,x,y) &= u_1 f(t,x,y) \\
f_y(t,x,y) &= u_2 f(t,x,y) \\
f_{xx}(t,x,y) &= u_1^2 f(t,x,y) \\
f_{yy}(t,x,y) &= u_2^2 f(t,x,y)
\end{aligned}$$

In turn, we conclude that

$$\mathbb{E} \exp\left\{ u_1 M_1(t) + u_2 M_2(t) - \frac{1}{2}\left(u_1^2 + u_2^2\right) t \right\} = 1$$

which gives us the moment-generating function formula

$$\mathbb{E} e^{u_1 M_1(t) + u_2 M_2(t)} = e^{\frac{1}{2} u_1^2 t} \cdot e^{\frac{1}{2} u_2^2 t}$$

Because the joint moment-generating function factors into the product of moment-generating functions, $M_1(t)$ and $M_2(t)$ must be independent. □

Remark 6. In general, a d-dimensional continuous (local) martingale

$$M = (M_1, M_2, \cdots, M_d)$$

is a d-dimensional Brownian motion if $[M_i, M_i](t) = t$ and $[M_i, M_j](t) = 0$ for $i \neq j$.

Exercise 5. *(Correlated Brownian motions from independent ones).*

Let $(W_1(t), \ldots, W_d(t))$ be a d -dimensional Brownian motion. In particular, these Brownian motions are independent of one another. Let $(\sigma_{ij}(t))_{i=1,\ldots,m; j=1,\ldots,d}$ be an $m \times d$ matrix-valued process adapted to the filtration associated with the d -dimensional Brownian motion. For $i = 1, \ldots, m$ define

$$\sigma_i(t) = \left[\sum_{j=1}^{d} \sigma_{ij}^2(t) \right]^{\frac{1}{2}}$$

and assume this is never zero. Define also

$$B_i(t) = \sum_{j=1}^{d} \int_0^t \frac{\sigma_{ij}(u)}{\sigma_i(u)} dW_j(u)$$

1. Show that, for each i, B_i is a Brownian motion.
2. Show that $dB_i(t) dB_k(t) = \rho_{ik}(t)$, where

$$\rho_{ik}(t) = \frac{1}{\sigma_i(t)\sigma_k(t)} \sum_{j=1}^{d} \sigma_{ij}(t)\sigma_{kj}(t)$$

Exercise 6. *(Independent Brownian motions from correlated ones).*

Let $B_1(t), \ldots, B_m(t)$ be m one-dimensional Brownian motions with

$$dB_i(t) dB_k(t) = \rho_{ik}(t) dt \text{ for all } i, k = 1, \ldots, m$$

where $\rho_{ik}(t)$ are adapted processes taking values in (-1,1) for $i \neq k$ and $\rho_{ik}(t) = 1$ for $i = k$. Assume that the symmetric matrix

$$C(t) = \begin{bmatrix} \rho_{11}(t) & \rho_{12}(t) & \cdots & \rho_{1m}(t) \\ \rho_{21}(t) & \rho_{22}(t) & \cdots & \rho_{2m}(t) \\ \vdots & \vdots & & \vdots \\ \rho_{m1}(t) & \rho_{m2}(t) & \cdots & \rho_{mm}(t) \end{bmatrix}$$

is *positive definite* for all t almost surely. Because the matrix $C(t)$ is symmetric and positive definite, it has a *matrix square root*. In other words, there is a matrix

$$A(t) = \begin{bmatrix} a_{11}(t) & a_{12}(t) & \cdots & a_{1m}(t) \\ a_{21}(t) & a_{22}(t) & \cdots & a_{2m}(t) \\ \vdots & \vdots & & \vdots \\ a_{m1}(t) & a_{m2}(t) & \cdots & a_{mm}(t) \end{bmatrix}$$

such that $C(t) = A(t)A^{\mathrm{tr}}(t)$, which when written componentwise is

$$\rho_{ik}(t) = \sum_{j=1}^{m} a_{ij}(t)a_{kj}(t) \text{ for all } i,k = 1,\dots,m$$

This matrix can be chosen so that its components $a_{ik}(t)$ are adapted processes. Furthermore, the matrix $A(t)$ has an inverse

$$A^{-1}(t) = \begin{bmatrix} \alpha_{11}(t) & \alpha_{12}(t) & \cdots & \alpha_{1m}(t) \\ \alpha_{21}(t) & \alpha_{22}(t) & \cdots & \alpha_{2m}(t) \\ \vdots & \vdots & & \vdots \\ \alpha_{m1}(t) & \alpha_{m2}(t) & \cdots & \alpha_{mm}(t) \end{bmatrix}$$

which means that

$$\sum_{j=1}^{m} a_{ij}(t)\alpha_{jk}(t) = \sum_{j=1}^{m} \alpha_{ij}(t)a_{jk}(t) = \delta_{ik}$$

where we define

$$\delta_{ik} = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{if } i \neq k \end{cases}$$

to be the so-called *Kronecker delta*. Show that there exist m independent Brownian motions $W_1(t),\dots,W_m(t)$ such that

$$B_i(t) = \sum_{j=1}^{m} \int_0^t a_{ij}(u)dW_j(u) \text{ for all } i = 1,\dots,m$$

Chapter 4: Risk-neutral pricing

Gechun Liang

Risk-neutral pricing is a powerful method for computing prices of derivative securities, but it is fully justified only when it is accompanied by a hedge for a short position in the security being priced. To determine the value of a European call, we determine the initial capital required to set up a portfolio that with probability one hedges a short position in the derivative security.

1 Risk-neutral measure

1.1 Girsanov's theorem for a single Brownian motion

First, we perform change of measure in order to change a mean for a whole process. To set the stage, suppose we have a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a filtration $\mathscr{F}(t)$, defined for $0 \le t \le T$, where T is a fixed final time. Suppose further that Z is an almost surely positive random variable satisfying $\mathbb{E}Z = 1$, and we define $\widetilde{\mathbb{P}}$ by given by

$$\widetilde{\mathbb{P}}(A) = \int_A Z(\omega) dP(\omega), \text{ for all } A \in \mathscr{F}. \tag{1}$$

We can then define the Radon-Nikodym derivative process

$$Z(t) = \mathbb{E}[Z \mid \mathscr{F}(t)], \quad 0 \le t \le T$$

The Radon-Nikodym derivative process is a martingale because of iterated conditioning: for $0 \le s \le t \le T$

$$\mathbb{E}[Z(t) \mid \mathscr{F}(s)] = \mathbb{E}[\mathbb{E}[Z \mid \mathscr{F}(t)] \mid \mathscr{F}(s)] = \mathbb{E}[Z \mid \mathscr{F}(s)] = Z(s).$$

Gechun Liang
Department of Statistics, University of Warwick, U.K. e-mail: g.liang@warwick.ac.uk

Lemma 1. *Let t satisfying $0 \le t \le T$ be given and let Y be an $\mathscr{F}(t)$-measurable random variable. Then*

$$\widetilde{\mathbb{E}} Y = \mathbb{E}[YZ(t)]$$

Proof. We use the property "taking out what is known" for conditional expectation and the definition of $Z(t)$ to write

$$\widetilde{\mathbb{E}} Y = \mathbb{E}[YZ] = \mathbb{E}[\mathbb{E}[YZ \mid \mathscr{F}(t)]] = \mathbb{E}[Y\mathbb{E}[Z \mid \mathscr{F}(t)]] = \mathbb{E}[YZ(t)]. \square$$

Lemma 2. *Let s and t satisfying $0 \le s \le t \le T$ be given and let Y be an $\mathscr{F}(t)$-measurable random variable. Then*

$$\widetilde{\mathbb{E}}[Y \mid \mathscr{F}(s)] = \frac{1}{Z(s)}\mathbb{E}[YZ(t) \mid \mathscr{F}(s)]. \tag{2}$$

Proof. It is clear that $\frac{1}{Z(s)}\mathbb{E}[YZ(t) \mid \mathscr{F}(s)]$ is $\mathscr{F}(s)$-measurable. We must check the partial-averaging property for conditional expectation which in this case is

$$\int_A \frac{1}{Z(s)}\mathbb{E}[YZ(t) \mid \mathscr{F}(s)]d\widetilde{\mathbb{P}} = \int_A Y d\widetilde{\mathbb{P}} \text{ for all } A \in \mathscr{F}(s) \tag{3}$$

Note that because we are claiming that the right-hand side of (2) is the conditional expectation of Y under the $\widetilde{\mathbb{P}}$ probability measure, we must integrate with respect to the measure $\widetilde{\mathbb{P}}$ in the statement of the partial-averaging property (3). We may write the left-hand side of (3) as

$$\widetilde{\mathbb{E}}\left[\mathbb{I}_A \frac{1}{Z(s)}\mathbb{E}[YZ(t) \mid \mathscr{F}(s)]\right]$$

and then use Lemma 1 to write

$$\begin{aligned}
\widetilde{\mathbb{E}}\left[\mathbb{I}_A \frac{1}{Z(s)}\mathbb{E}[YZ(t) \mid \mathscr{F}(s)]\right] &= \mathbb{E}\left[\mathbb{I}_A \mathbb{E}[YZ(t) \mid \mathscr{F}(s)]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbb{I}_A YZ(t) \mid \mathscr{F}(s)\right]\right] \\
&= \mathbb{E}\left[\mathbb{I}_A YZ(t)\right] \\
&= \widetilde{\mathbb{E}}\left[\mathbb{I}_A Y\right] \\
&= \int_A Y d\mathbb{P}
\end{aligned}$$

This verifies (3) which in turn proves (2). $\square$

Theorem 1. *(Girsanov, one dimension)*

Let $W(t), 0 \le t \le T$, be a Brownian motion on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, and let $\mathscr{F}(t), 0 \le t \le T$ be a filtration for this Brownian motion. Let $\Theta(t), 0 \le t \le T$, be an adapted process. Define

$$Z(t) = \exp\left\{-\int_0^t \Theta(u)dW(u) - \frac{1}{2}\int_0^t \Theta^2(u)du\right\} \quad (\underline{stochastic exponential})$$
$$\widetilde{W}(t) = W(t) + \int_0^t \Theta(u)du,$$

and assume that[1]

$$\mathbb{E}\int_0^T \Theta^2(u)Z^2(u)du < \infty$$

Set $Z = Z(T)$. *Then* $\mathbb{E}Z = 1$ *and under the probability measure* $\widetilde{\mathbb{P}}$ *the process* $\widetilde{W}(t), 0 \le t \le T$, *is a Brownian motion.*

Proof. We use Lévy's Theorem, Theorem 8 in Chapter 3, which says that a martingale starting at zero at time zero, with continuous paths and with quadratic variation equal to t at each time t, is a Brownian motion. Note that

$$d\widetilde{W}(t)d\widetilde{W}(t) = (dW(t) + \Theta(t)dt)^2 = dW(t)dW(t) = dt$$

It remains to show that $\widetilde{W}(t)$ is a martingale under $\widetilde{\mathbb{P}}$.

Step 1. We first observe that $Z(t)$ is a martingale under $\mathbb{P}$. Indeed, With

$$X(t) = -\int_0^t \Theta(u)dW(u) - \frac{1}{2}\int_0^t \Theta^2(u)du$$

and $f(x) = e^x$ so that $f'(x) = e^x$ and $f''(x) = e^x$, we have

$$\begin{aligned}
dZ(t) &= df(X(t)) \\
&= f'(X(t))dX(t) + \frac{1}{2}f''(X(t))dX(t)dX(t) \\
&= e^{X(t)}\left(-\Theta(t)dW(t) - \frac{1}{2}\Theta^2(t)dt\right) + \frac{1}{2}e^{X(t)}\Theta^2(t)dt \\
&= -\Theta(t)Z(t)dW(t)
\end{aligned}$$

Integrating both sides of the equation above, we see that

$$Z(t) = Z(0) - \int_0^t \Theta(u)Z(u)dW(u)$$

Because Itô's integrals are martingales, $Z(t)$ is a martingale. In particular, $\mathbb{E}Z = \mathbb{E}Z(T) = Z(0) = 1$ Because $Z(t)$ is a martingale and $Z = Z(T)$, we have

$$Z(t) = \mathbb{E}[Z(T) \mid \mathscr{F}(t)] = \mathbb{E}[Z \mid \mathscr{F}(t)], \quad 0 \le t \le T$$

This shows that $Z(t), 0 \le t \le T$, is a Radon-Nikodym derivative process.

Step 2. We next show that $\widetilde{W}(t)Z(t)$ is a martingale under $\mathbb{P}$. To see this, note that

[1] A sufficient condition to ensure that Z is a martingale is called Novikov's condition: $\mathbb{E}[e^{\frac{1}{2}\int_0^T |\Theta(t)|^2 dt}] < \infty$.

$$
\begin{aligned}
d(\widetilde{W}(t)Z(t)) =& \widetilde{W}(t)dZ(t) + Z(t)d\widetilde{W}(t) + d\widetilde{W}(t)dZ(t) \\
=& -\widetilde{W}(t)\Theta(t)Z(t)dW(t) + Z(t)dW(t) + Z(t)\Theta(t)dt \\
& + (dW(t) + \Theta(t)dt)(-\Theta(t)Z(t)dW(t)) \\
=& (-\widetilde{W}(t)\Theta(t) + 1)Z(t)dW(t)
\end{aligned}
$$

Because the final expression has no dt term, the process $\widetilde{W}(t)Z(t)$ is a martingale under $\mathbb{P}$.

Step 3. Let $0 \leq s \leq t \leq T$ be given. Lemma 2 and the martingale property for $\widetilde{W}(t)Z(t)$ under $\mathbb{P}$ imply

$$
\widetilde{\mathbb{E}}[\widetilde{W}(t) \mid \mathscr{F}(s)] = \frac{1}{Z(s)}\mathbb{E}[\widetilde{W}(t)Z(t) \mid \mathscr{F}(s)] = \frac{1}{Z(s)}\widetilde{W}(s)Z(s) = \widetilde{W}(s)
$$

This shows that $\widetilde{W}(t)$ is a martingale under $\widetilde{\mathbb{P}}$. The proof is complete. □

The probability measures $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ in Girsanov's Theorem are equivalent (i.e., they agree about which sets have probability zero and hence about which sets have probability one). This is because $\mathbb{P}\{Z > 0\} = 1$. In the remainder of this section, we set up an asset price model in which $\mathbb{P}$ is the *actual probability measure* and $\widetilde{\mathbb{P}}$ is the *risk-neutral measure*.

1.2 Stock under the risk-neutral measure

Let $W(t), 0 \leq t \leq T$, be a Brownian motion on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and let $\mathscr{F}(t), 0 \leq t \leq T$, be a filtration for this Brownian motion. Here T is a fixed final time. Consider a stock price process whose differential is

$$
dS(t) = \alpha(t)S(t)dt + \sigma(t)S(t)dW(t), \quad 0 \leq t \leq T. \tag{4}
$$

The mean rate of return $\alpha(t)$ and the volatility $\sigma(t)$ are allowed to be adapted processes. We assume that, for all $t \in [0,T], \sigma(t)$ is almost surely not zero. This stock price is a generalized geometric Brownian motion (see Example 1 in Chapter 3), and an equivalent way of writing (4) is

$$
S(t) = S(0)\exp\left\{\int_0^t \sigma(s)dW(s) + \int_0^t \left(\alpha(s) - \frac{1}{2}\sigma^2(s)\right)ds\right\}
$$

In addition, suppose we have an adapted interest rate process $R(t)$. We define the discount process

$$
D(t) = e^{-\int_0^t R(s)ds} \tag{5}
$$

and note that (verify it!)

$$
dD(t) = -R(t)D(t)dt. \tag{6}
$$

Observe that although $D(t)$ is random, it has *zero quadratic variation*. This is because it is "smooth." It has a derivative, namely $D'(t) = -R(t)D(t)$, and one does not need stochastic calculus to do this computation. The stock price $S(t)$ is random and has nonzero quadratic variation. It is "more random" than $D(t)$.

If we invest in the stock, we have no way of knowing whether the next move of the driving Brownian motion will be up or down, and this move directly affects the stock price. Hence, we face a high degree of uncertainty. On the other hand, consider a money market account with variable interest rate $R(t)$, where money is rolled over at this interest rate. If the price of a share of this money market account at time zero is 1, then the price of a share of this money market account at time t is $e^{\int_0^t R(s)ds} = \frac{1}{D(t)}$. If we invest in this account, we know the interest rate at the time of the investment and hence have a high degree of certainty about what the return will be over a short period of time. Over longer periods, we are less certain because the interest rate is variable, and at the time of investment, we do not know the future interest rates that will be applied. However, the randomness in the model affects the money market account only indirectly by affecting the interest rate. Changes in the interest rate do not affect the money market account instantaneously but only when they act over time. Unlike the price of the money market account, the stock price is susceptible to instantaneous unpredictable changes and is, in this sense, "more random" than $D(t)$. Our mathematical model captures this effect because $S(t)$ has *nonzero quadratic variation*, while $D(t)$ has *zero quadratic variation*.

Next, we introduce *the risk-neutral measure*. The *discounted stock price process* is

$$D(t)S(t) = S(0)\exp\left\{\int_0^t \sigma(s)dW(s) + \int_0^t \left(\alpha(s) - R(s) - \frac{1}{2}\sigma^2(s)\right)ds\right\}$$

and its differential is

$$\begin{aligned} d(D(t)S(t)) &= (\alpha(t) - R(t))D(t)S(t)dt + \sigma(t)D(t)S(t)dW(t) \\ &= \sigma(t)D(t)S(t)[\Theta(t)dt + dW(t)] \end{aligned} \tag{7}$$

where we define the market price of risk to be

$$\Theta(t) = \frac{\alpha(t) - R(t)}{\sigma(t)}. \tag{8}$$

We then introduce the probability measure $\widetilde{\mathbb{P}}$ defined in Girsanov's Theorem, Theorem 1, which uses the market price of risk $\Theta(t)$. In terms of the Brownian motion $\widetilde{W}(t)$ of that theorem, we may rewrite $D(t)S(t)$ as

$$d(D(t)S(t)) = \sigma(t)D(t)S(t)d\widetilde{W}(t)$$

We call $\widetilde{\mathbb{P}}$, the measure defined in Girsanov's Theorem, the risk-neutral measure because *it is equivalent to the original measure $\mathbb{P}$ and it renders the discounted* *stock price $D(t)S(t)$ into a martingale.* See Definition 1 later on. Equivalently, we have

$$D(t)S(t) = S(0)\exp\left\{\int_0^t \sigma(u)d\widetilde{W}(u) - \int_0^t \frac{1}{2}\sigma^2(u)du\right\}. \tag{9}$$

The undiscounted stock price $S(t)$ has mean rate of return equal to the interest rate under $\widetilde{\mathbb{P}}$, as one can verify by making the replacement $dW(t) = -\Theta(t)dt + d\widetilde{W}(t)$ in (4). With this substitution, we have

$$dS(t) = R(t)S(t)dt + \sigma(t)S(t)d\widetilde{W}(t),$$

or equivalently,

$$S(t) = S(0)\exp\left\{\int_0^t \sigma(s)d\widetilde{W}(s) + \int_0^t \left(R(s) - \frac{1}{2}\sigma^2(s)\right)ds\right\}. \tag{10}$$

1.3 Value of portfolio process under the risk-neutral measure

Consider an agent who begins with initial capital $X(0)$ and at each time t, $0 \le t \le T$, holds $\Delta(t)$ shares of stock, investing or borrowing at the interest rate $R(t)$ as necessary to finance this. Then, the wealth process satisfies

$$X(t) = \Delta(t)S(t) + \frac{X(t) - \Delta(t)S(t)}{B(t)}B(t)$$

where $B(t) = 1/D(t)$ represents the money market account (bank account) satisfying (verify it!)

$$dB(t) = R(t)B(t)dt.$$

Then, the self-financing condition yields that

$$\begin{aligned}
dX(t) &= \Delta(t)dS(t) + R(t)(X(t) - \Delta(t)S(t))dt \\
&= \Delta(t)(\alpha(t)S(t)dt + \sigma(t)S(t)dW(t)) + R(t)(X(t) - \Delta(t)S(t))dt \\
&= R(t)X(t)dt + \Delta(t)(\alpha(t) - R(t))S(t)dt + \Delta(t)\sigma(t)S(t)dW(t)
\end{aligned}$$

The three terms appearing in the last line can be understood as follows:

1. an average underlying rate of return $R(t)$ on the portfolio, which is reflected by the term $R(t)X(t)dt$;
2. a risk premium $\alpha(t) - R(t)$ for investing in the stock, which is reflected by the term $\Delta(t)(\alpha(t) - R(t))S(t)dt$; and
3. a volatility term proportional to the size of the stock investment, which is the term $\Delta(t)\sigma(t)S(t)dW(t)$.

In turn, using the market price of risk $\Theta(t)$, we derive

$$dX(t) = R(t)X(t)dt + \Delta(t)\sigma(t)S(t)[\Theta(t)dt + dW(t)]$$

and

$$\begin{aligned} d(D(t)X(t)) &= \Delta(t)\sigma(t)D(t)S(t)[\Theta(t)dt + dW(t)] \\ &= \Delta(t)d(D(t)S(t)) \end{aligned} \tag{11}$$

Changes in the discounted value of an agent's portfolio are entirely due to fluctuations in the discounted stock price. We may use (9) to rewrite $D(t)X(t)$ as

$$d(D(t)X(t)) = \Delta(t)\sigma(t)D(t)S(t)d\widetilde{W}(t) \tag{12}$$

Our agent has two investment options: (1) a money market account B with rate of return $R(t)$, and (2) a stock S with mean rate of return $R(t)$ under $\widetilde{\mathbb{P}}$. Regardless of how the agent invests, the mean rate of return for his portfolio will be $R(t)$ under $\widetilde{\mathbb{P}}$, and hence the discounted value of his portfolio, $D(t)X(t)$ will be a martingale. This is the content of (12).

1.4 Pricing under the risk-neutral measure

Let $V(T)$ be an $\mathscr{F}(T)$-measurable random variable. This represents the payoff at time T of a derivative security. We allow this payoff to be path-dependent (i.e., to depend on anything that occurs between times 0 and T), which is what $\mathscr{F}(T)$-measurability means.

We wish to know what initial capital $X(0)$ and portfolio process $\Delta(t), 0 \le t \le T$, an agent would need in order to hedge a short position in this derivative security, i.e., in order to have

$$X(T) = V(T), a.s. \tag{13}$$

Our agent wishes to choose initial capital $X(0)$ and portfolio strategy $\Delta(t)$, $0 \le t \le T$, such that (13) holds. We shall see in the next section that this can be done via *martingale representation theorem*. Once it has been done, the fact that $D(t)X(t)$ is a martingale under $\widetilde{\mathbb{P}}$ implies

$$D(t)X(t) = \widetilde{\mathbb{E}}[D(T)X(T) \mid \mathscr{F}(t)] = \widetilde{\mathbb{E}}[D(T)V(T) \mid \mathscr{F}(t)] \tag{14}$$

The value $X(t)$ of the hedging portfolio in (14) is the capital needed at time t in order to successfully complete the hedge of the short position in the derivative security with payoff $V(T)$. Hence, we can call this the price $V(t)$ of the derivative security at time t, and (14) becomes

$$D(t)V(t) = \widetilde{\mathbb{E}}[D(T)V(T) \mid \mathscr{F}(t)],\ 0 \le t \le T \tag{15}$$

This is the risk-neutral pricing formula. Using the discounted process $D(t)$ given in (5), it may be written as

$$V(t) = \widetilde{\mathbb{E}}\left[e^{-\int_t^T R(u)du} V(T) \mid \mathscr{F}(t)\right],\ 0 \le t \le T \tag{16}$$

which is also referred to as the *risk-neutral pricing formula for the continuous-time model*. Note that (16) is valid *only if the hedging portfolio $\Delta(t)$ exists.*

1.5 Black-Scholes formula

To obtain the Black-Scholes price of a European call, we assume a constant volatility σ, constant interest rate r, and take the derivative security payoff to be $V(T) = (S(T) - K)^+$. The right-hand side of (16) becomes

$$\widetilde{\mathbb{E}}\left[e^{-r(T-t)}(S(T) - K)^+ \mid \mathscr{F}(t)\right]$$

Because geometric Brownian motion is a Markov process, this expression depends on the stock price $S(t)$ and of course on the time t at which the conditional expectation is computed, but not on the stock price prior to time t. In other words, there is a function $c(t,x)$ such that

$$c(t, S(t)) = \widetilde{\mathbb{E}}\left[e^{-r(T-t)}(S(T) - K)^+ \mid \mathscr{F}(t)\right] \tag{17}$$

We compute $c(t,x)$ as follows. With constant σ and r, equation (10) becomes

$$S(t) = S(0)\exp\left\{\sigma\widetilde{W}(t) + \left(r - \frac{1}{2}\sigma^2\right)t\right\}$$

and we may thus write

$$\begin{aligned} S(T) &= S(t)\exp\left\{\sigma(\widetilde{W}(T) - \widetilde{W}(t)) + \left(r - \frac{1}{2}\sigma^2\right)\tau\right\} \\ &= S(t)\exp\left\{-\sigma\sqrt{\tau}Y + \left(r - \frac{1}{2}\sigma^2\right)\tau\right\} \end{aligned}$$

where Y is the standard normal random variable

$$Y = -\frac{\widetilde{W}(T) - \widetilde{W}(t)}{\sqrt{T-t}}$$

and τ is the "time to expiration" $T - t$. We see that $S(T)$ is the product of the $\mathscr{F}(t)$-measurable random variable $S(t)$ and the random variable

$$\exp\left\{-\sigma\sqrt{\tau}Y + \left(r - \frac{1}{2}\sigma^2\right)\tau\right\}$$

which is independent of $\mathscr{F}(t)$. Therefore, (17) holds with

$$\begin{aligned}
c(t,x) &= \widetilde{\mathbb{E}}\left[e^{-r\tau}\left(x\exp\left\{-\sigma\sqrt{\tau}Y + \left(r - \frac{1}{2}\sigma^2\right)\tau\right\} - K\right)^+\right] \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-r\tau}\left(x\exp\left\{-\sigma\sqrt{\tau}y + \left(r - \frac{1}{2}\sigma^2\right)\tau\right\} - K\right)^+ e^{-\frac{1}{2}y^2}dy
\end{aligned}$$

The integrand

$$\left(x\exp\left\{-\sigma\sqrt{\tau}y + \left(r - \frac{1}{2}\sigma^2\right)\tau\right\} - K\right)^+$$

is positive if and only if

$$y < d_-(\tau,x) = \frac{1}{\sigma\sqrt{\tau}}\left[\log\frac{x}{K} + \left(r - \frac{1}{2}\sigma^2\right)\tau\right]$$

Therefore,

$$\begin{aligned}
c(t,x) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d-(\tau,x)} e^{-r\tau}\left(x\exp\left\{-\sigma\sqrt{\tau}y + \left(r - \frac{1}{2}\sigma^2\right)\tau\right\} - K\right) e^{-\frac{1}{2}y^2}dy \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d-(\tau,x)} x\exp\left\{-\frac{y^2}{2} - \sigma\sqrt{\tau}y - \frac{\sigma^2\tau}{2}\right\}dy \\
&\quad - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d-(\tau,x)} e^{-r\tau}Ke^{-\frac{1}{2}y^2}dy \\
&= \frac{x}{\sqrt{2\pi}}\int_{-\infty}^{d-(\tau,x)} \exp\left\{-\frac{1}{2}(y+\sigma\sqrt{\tau})^2\right\}dy - e^{-r\tau}KN\left(d_-(\tau,x)\right) \\
&= \frac{x}{\sqrt{2\pi}}\int_{-\infty}^{d-(\tau,x)+\sigma\sqrt{\tau}} \exp\left\{-\frac{z^2}{2}\right\}dz - e^{-r\tau}KN\left(d_-(\tau,x)\right) \\
&= xN\left(d_+(\tau,x)\right) - e^{-r\tau}KN\left(d_-(\tau,x)\right)
\end{aligned}$$

where

$$d_+(\tau,x) = d_-(\tau,x) + \sigma\sqrt{\tau} = \frac{1}{\sigma\sqrt{\tau}}\left[\log\frac{x}{K} + \left(r + \frac{1}{2}\sigma^2\right)\tau\right]$$

For future reference, we introduce the notation

$$\mathrm{BSM}(\tau,x;K,r,\sigma) = \widetilde{\mathbb{E}}\left[e^{-r\tau}\left(x\exp\left\{-\sigma\sqrt{\tau}Y + \left(r - \frac{1}{2}\sigma^2\right)\tau\right\} - K\right)^+\right]$$

where Y is a standard normal random variable under $\tilde{\mathbb{P}}$. We have just shown that

$$\mathrm{BSM}(\tau,x;K,r,\sigma) = xN\left(d_+(\tau,x)\right) - e^{-r\tau}KN\left(d_-(\tau,x)\right). \tag{18}$$

Exercise 1. *(Black-Scholes formula for time-varying, nonrandom interest rate and volatility).* Consider a stock whose price differential is

$$dS(t) = r(t)S(t)dt + \sigma(t)S(t)d\widetilde{W}(t)$$

where $r(t)$ and $\sigma(t)$ are nonrandom functions of t and $\widetilde{W}$ is a Brownian motion under the risk-neutral measure $\widetilde{\mathbb{P}}$. Let $T > 0$ be given, and consider a European call, whose value at time zero is

$$c(0,S(0)) = \widetilde{\mathbb{E}}\left[\exp\left\{-\int_0^T r(t)dt\right\}(S(T)-K)^+\right]$$

1. Show that $S(T)$ is of the form $S(0)e^X$, where X is a normal random variable, and determine the mean and variance of X.
2. Let

$$\begin{aligned}\mathrm{BSM}(T,x;K,R,\Sigma) = x\,N&\left(\tfrac{1}{\Sigma\sqrt{T}}\left[\log\tfrac{x}{K} + \left(R+\Sigma^2/2\right)T\right]\right)\\ -e^{-RT}KN&\left(\tfrac{1}{\Sigma\sqrt{T}}\left[\log\tfrac{x}{K} + \left(R-\Sigma^2/2\right)T\right]\right)\end{aligned}$$

   denote the value at time zero of a European call expiring at time T wher the underlying stock has constant volatility Σ and the interest rate R constant. Show that

$$c(0,S(0)) = \mathrm{BSM}\left(T,S(0);K,\frac{1}{T}\int_0^T r(t)dt, \sqrt{\frac{1}{T}\int_0^T \sigma^2(t)dt}\right)$$

2 Martingale representation theorem

In this section, in the model with one stock driven by one Brownian motion, we verify the assumption on which the risk-neutral pricing formula (16) is based.

2.1 Martingale representation with one Brownian motion

The existence of a hedging portfolio in the model with one stock and one Brownian motion depends on the following martingale representation theorem. See Chapter 3 of *Brownian Motion and Stochastic Calculus* by Karatzas and Shreve for the proof.

Theorem 2. *(Martingale representation, one dimension).*
Let $W(t)$ $\overline{0 \le t \le T}$, be a Brownian motion on a probability space $(\Omega,\mathscr{F},\mathbb{P})$, and let $\mathscr{F}(t), 0 \le t \le T$, be the filtration generated by this Brownian motion. Let

$M(t), 0 \leq t \leq T$, be a martingale with respect to this filtration. Then, there is an adapted process $\Gamma(u), 0 \leq u \leq T$, such that

$$M(t) = M(0) + \int_0^t \Gamma(u) dW(u),\ 0 \leq t \leq T.$$

The Martingale Representation Theorem asserts that when the filtration is the one generated by a Brownian motion, then every martingale with respect to this filtration is an initial condition plus an Itô's integral with respect to the Brownian motion. The relevance to hedging of this is that the only source of uncertainty in the model is the Brownian motion appearing in Theorem 2 and hence there is only one source of uncertainty to be removed by hedging.

The assumption that the filtration in Theorem 2 is the one generated by the Brownian motion is more restrictive than the assumption of Girsanov's Theorem, Theorem 1, in which the filtration can be larger than the one generated by the Brownian motion. If we include this extra restriction in Girsanov's Theorem, then we obtain the following corollary. The first paragraph of this corollary is just a repeat of Girsanov's Theorem; the second part contains the new assertion.

Corollary 1. *Let $W(t), 0 \leq t \leq T$, be a Brownian motion on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, and let $\mathscr{F}(t), 0 \leq t \leq T$, be the filtration generated by this Brownian motion. Let $\Theta(t), 0 \leq t \leq T$, be an adapted process, define*

$$Z(t) = \exp\left\{ -\int_0^t \Theta(u) dW(u) - \frac{1}{2} \int_0^t \Theta^2(u) du \right\}$$
$$\widetilde{W}(t) = W(t) + \int_0^t \Theta(u) du$$

and assume that $\mathbb{E} \int_0^T \Theta^2(u) Z^2(u) du < \infty$. Set $Z = Z(T)$. Then $\mathbb{E}Z = 1$, and under the probability measure $\widetilde{\mathbb{P}}$ given by (1), the process $\widetilde{W}(t), 0 \leq t \leq T$, is a Brownian motion.

Now let $\widetilde{M}(t), 0 \leq t \leq T$, be a martingale under $\widetilde{\mathbb{P}}$. Then there is an adapted process $\widetilde{\Gamma}(u), 0 \leq u \leq T$, such that

$$\widetilde{M}(t) = \widetilde{M}(0) + \int_0^t \tilde{\Gamma}(u) d\widetilde{W}(u), 0 \leq t \leq T$$

Exercise 2. Prove Corollary 1.

2.2 Hedging with one stock

We now return to the hedging problem. We begin with the stock price process (4) and an interest rate process $R(t)$ that generates the discount process (6). *Recall the assumption that, for all $t \in [0, T]$, the volatility $\sigma(t)$ is almost surely not zero. We* *make the additional assumption that the filtration $\mathscr{F}(t), 0 \leq t \leq T$, is generated by the Brownian motion $W(t), 0 \leq t \leq T$.*

Let $V(T)$ be an $\mathscr{F}(T)$-measurable random variable and, for $0 \leq t \leq T$, define $V(t)$ by the risk-neutral pricing formula (16). Then,

$$D(t)V(t) = \widetilde{\mathbb{E}}[D(T)V(T) \mid \mathscr{F}(t)]$$

This is a $\widetilde{\mathbb{P}}$-martingale(why?). Therefore, $D(t)V(t)$ has a representation as (recall that $D(0)V(0) = V(0)$)

$$D(t)V(t) = V(0) + \int_0^t \tilde{\Gamma}(u) d\widetilde{W}(u),\ 0 \leq t \leq T.$$

On the other hand, for any portfolio process $\Delta(t)$, the differential of the discounted portfolio value is given by (12), and hence

$$D(t)X(t) = X(0) + \int_0^t \Delta(u)\sigma(u)D(u)S(u)d\widetilde{W}(u), 0 \leq t \leq T$$

In order to have $X(t) = V(t)$ for all t, we should choose

$$X(0) = V(0)$$

and choose $\Delta(t)$ to satisfy

$$\Delta(t)\sigma(t)D(t)S(t) = \widetilde{\Gamma}(t), 0 \leq t \leq T \tag{19}$$

which is equivalent to

$$\Delta(t) = \frac{\tilde{\Gamma}(t)}{\sigma(t)D(t)S(t)}, 0 \leq t \leq T$$

With these choices, we have a hedge for a short position in the derivative security with payoff $V(T)$ at time T.

Remark 1. There are two key assumptions that make the hedge possible. *The first is that the volatility $\sigma(t)$ is not zero. The other key assumption is that $\mathscr{F}(t)$ is generated by the underlying Brownian motion*. (i.e., there is no randomness in the derivative security apart from the Brownian motion randomness, which can be hedged by trading the stock). Under these two assumptions, every $\mathscr{F}(T)$-measurable derivative security can be hedged.

3 Fundamental theorems of asset pricing

3.1 Girsanov and martingale representation theorems

We first generalize Theorems 1 and 2. Throughout this section,

$$W(t) = (W_1(t), \dots, W_d(t))$$

is a multidimensional Brownian motion on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. We interpret $\mathbb{P}$ to be the actual probability measure, the one that would be observed from empirical studies of price data. Associated with this Brownian motion, we have a filtration $\mathscr{F}(t)$. We shall have a fixed final time T, and we shall assume that $\mathscr{F} = \mathscr{F}(T)$. We do not always assume that the filtration is the one generated by the Brownian motion. When that is assumed, we say so explicitly.

Theorem 3. *(Girsanov, multiple dimensions)*
Let T be a fixed positive time, and let $\Theta(t) = (\Theta_1(t), \dots, \Theta_d(t))$ be a d-dimensional adapted process. Define

$$Z(t) = \exp\left\{-\int_0^t \Theta(u) \cdot dW(u) - \frac{1}{2}\int_0^t \|\Theta(u)\|^2 du\right\}$$
$$\widetilde{W}(t) = W(t) + \int_0^t \Theta(u) du$$

and assume that

$$\mathbb{E}\int_0^T \|\Theta(u)\|^2 Z^2(u) du < \infty$$

Set $Z = Z(T)$. Then $\mathbb{E}Z = 1$, and under the probability measure $\widetilde{\mathbb{P}}$ given by

$$\widetilde{\mathbb{P}}(A) = \int_A Z(\omega) d\mathbb{P}(\omega) \text{ for all } A \in \mathscr{F}$$

the process $\widetilde{W}(t)$ is a d -dimensional Brownian motion.

The remarkable thing about the conclusion of the multidimensional Girsanov Theorem is that the component processes of $\widetilde{W}(t)$ are independent under $\widetilde{\mathbb{P}}$, but not necessarily under $\mathbb{P}$.

Exercise 3. Use Lévy theorem to prove Theorem 3.

Theorem 4. *(Martingale representation, multiple dimensions).*
Let T be a fixed positive time, and assume that $\mathscr{F}(t), 0 \le t \le T$, is the filtration generated by the d-dimensional Brownian motion $W(t), 0 \le t \le T$. Let $M(t), 0 \le t \le T$, be a martingale with respect to this filtration under $\mathbb{P}$. Then, there is an adapted, d-dimensional process $\Gamma(u) = (\Gamma_1(u), \dots, \Gamma_d(u))$, $0 \le u \le T$, such that

$$M(t) = M(0) + \int_0^t \Gamma(u) \cdot dW(u), 0 \le t \le T$$

If, in addition, we assume the notation and assumptions of Theorem 3 and if $\widetilde{M}(t), 0 \leq t \leq T$, *is a* $\widetilde{\mathbb{P}}$*-martingale, then there is an adapted,* d*-dimensional process* $\tilde{\Gamma}(u) = \left(\tilde{\Gamma}_1(u), \ldots, \tilde{\Gamma}_d(u)\right)$ *such that*

$$\widetilde{M}(t) = \widetilde{M}(0) + \int_0^t \tilde{\Gamma}(u) \cdot d\widetilde{W}(u), 0 \leq t \leq T.$$

We refer to Chapter 3 in *Brownian Motion and Stochastic Calculus* by Karatzas and Shreve for the proof of the above theorem.

3.2 Multidimensional market model

We assume there are m stocks, each with stochastic differential (compared with (4))

$$dS_i(t) = \alpha_i(t)S_i(t)dt + S_i(t)\sum_{j=1}^{d}\sigma_{ij}(t)dW_j(t), \quad i = 1, \ldots, m \tag{20}$$

We assume that the mean rate of return vector $(\alpha_i(t))_{i=1,\ldots,m}$ and the volatility matrix $(\sigma_{ij}(t))_{i=1,\ldots,m;j=1,\ldots,d}$ are adapted processes.

These stocks are typically correlated. To see the nature of this correlation, we set $\sigma_i(t) = \sqrt{\sum_{j=1}^{d}\sigma_{ij}^2(t)}$, which we assume is never zero, and we define processes

$$B_i(t) = \sum_{j=1}^{d}\int_0^t \frac{\sigma_{ij}(u)}{\sigma_i(u)}dW_j(u), \quad i = 1, \ldots, m$$

By Exercise 5 in Chapter 3, we know that (1) for each i, B_i is a Brownian motion; (2) $dB_i(t)dB_k(t) = \rho_{ik}(t)dt$, where

$$\rho_{ik}(t) = \frac{1}{\sigma_i(t)\sigma_k(t)}\sum_{j=1}^{d}\sigma_{ij}(t)\sigma_{kj}(t).$$

Hence, each stock price can be rewritten as

$$dS_i(t) = \alpha_i(t)S_i(t)dt + \sigma_i(t)S_i(t)dB_i(t) \tag{21}$$

with

$$\begin{aligned} dS_i(t)dS_k(t) &= \sigma_i(t)\sigma_k(t)S_i(t)S_k(t)dB_i(t)dB_k(t) \\ &= \rho_{ik}(t)\sigma_i(t)\sigma_k(t)S_i(t)S_k(t)dt. \end{aligned}$$

Rewriting it in terms of ”relative differentials,” we obtain

$$\frac{dS_i(t)}{S_i(t)} \cdot \frac{dS_k(t)}{S_k(t)} = \rho_{ik}(t)\sigma_i(t)\sigma_k(t)dt.$$

Finally, in terms of the discounted stock process $D(t)S_i(t)$, we have (compared with (7))

$$\begin{aligned} d\left(D(t)S_i(t)\right) &= D(t)\left[dS_i(t) - R(t)S_i(t)dt\right] \\ &= D(t)S_i(t)\left[\left(\alpha_i(t) - R(t)\right)dt + \sum_{j=1}^{d}\sigma_{ij}(t)dW_j(t)\right] \\ &= D(t)S_i(t)\left[\left(\alpha_i(t) - R(t)\right)dt + \sigma_i(t)dB_i(t)\right], i = 1,\ldots,m. \end{aligned} \tag{22}$$

3.3 Existence of the risk-neutral measure

Definition 1. A probability measure $\widetilde{\mathbb{P}}$ is said to be risk-neutral if

1. $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are equivalent (i.e., for every $A \in \mathscr{F}, \mathbb{P}(A) = 0$ if and only if $\widetilde{\mathbb{P}}(A) = 0$), and
2. under $\widetilde{\mathbb{P}}$, the discounted stock price $D(t)S_i(t)$ is a martingale for every $i = 1,\ldots,m$

In order to make discounted stock prices be martingales, we would like to rewrite (22) as (compared with (7))

$$d\left(D(t)S_i(t)\right) = D(t)S_i(t)\sum_{j=1}^{d}\sigma_{ij}(t)\left[\Theta_j(t)dt + dW_j(t)\right] \tag{23}$$

If we can find the market price of risk processes $\Theta_j(t)$ that make (23) hold, with one such process for each source of uncertainty $W_j(t)$, we can then use the multidimensional Girsanov Theorem 4 to construct an equivalent probability measure $\widetilde{\mathbb{P}}$ under which $\widetilde{W}(t)$ is a d-dimensional Brownian motion. This permits us to reduce (23) to

$$d\left(D(t)S_i(t)\right) = D(t)S_i(t)\sum_{j=1}^{d}\sigma_{ij}(t)d\widetilde{W}_j(t) \tag{24}$$

and hence $D(t)S_i(t)$ is a martingale under $\widetilde{\mathbb{P}}$.

The problem of finding a risk-neutral measure is simply one of finding processes $\Theta_j(t)$ that make (22) and (23) agree. Since these equations have the same coefficient multiplying each $dW_j(t)$, they agree if and only if the coefficient multiplying dt is the same in both cases, which means that (compared with (8))

$$\alpha_i(t) - R(t) = \sum_{j=1}^{d}\sigma_{ij}(t)\Theta_j(t), \quad i = 1,\ldots,m \tag{25}$$

We call these the market price of risk equations. These are m equations in the d unknown processes $\Theta_1(t),\ldots,\Theta_d(t)$.

If one cannot solve the market price of risk equations, then there is an arbitrage in the model; the model is bad and should not be used for pricing.

Following along the similar arguments in 1.3, we derive the wealth process

$$
\begin{aligned}
dX(t) &= \sum_{i=1}^{m} \Delta_i(t) dS_i(t) + R(t) \left(X(t) - \sum_{i=1}^{m} \Delta_i(t) S_i(t) \right) dt \\
&= R(t)X(t)dt + \sum_{i=1}^{m} \Delta_i(t) \left(dS_i(t) - R(t)S_i(t)dt \right) \\
&= R(t)X(t)dt + \sum_{i=1}^{m} \frac{\Delta_i(t)}{D(t)} d\left(D(t)S_i(t) \right)
\end{aligned}
$$

where we used (22) in the last equality. The differential of the discounted portfolio value is (compared with (11))

$$
\begin{aligned}
d(D(t)X(t)) &= D(t)(dX(t) - R(t)X(t)dt) \\
&= \sum_{i=1}^{m} \Delta_i(t) d\left(D(t)S_i(t) \right)
\end{aligned}
\tag{26}
$$

Definition 2. An arbitrage is a portfolio value process $X(t)$ satisfying $X(0) = 0$ and also satisfying for some time $T > 0$

$$
\mathbb{P}\{X(T) \geq 0\} = 1, \quad \mathbb{P}\{X(T) > 0\} > 0. \tag{27}
$$

An arbitrage is a way of trading so that one starts with zero capital and at some later time T is *sure not to have lost money and furthermore has a positive probability of having made money*.

Exercise 4. Suppose a multidimensional market model as described in Section 3.2 has an arbitrage. In other words, suppose there is a portfolio value process satisfying $X_1(0) = 0$ and

$$
\mathbb{P}\{X_1(T) \geq 0\} = 1, \quad \mathbb{P}\{X_1(T) > 0\} > 0 \tag{28}
$$

for some positive T. Show that if $X_2(0)$ is positive, then there exists a portfolio value process $X_2(t)$ starting at $X_2(0)$ and satisfying

$$
\mathbb{P}\left\{X_2(T) \geq \frac{X_2(0)}{D(T)}\right\} = 1, \quad \mathbb{P}\left\{X_2(T) > \frac{X_2(0)}{D(T)}\right\} > 0 \tag{29}
$$

Conversely, show that if a multidimensional market model has a portfolio value process $X_2(t)$ such that $X_2(0)$ is positive and (29) holds, then the model has a portfolio value process $X_1(0)$ such that $X_1(0) = 0$ and (28) holds.

Theorem 5. *(First fundamental theorem of asset pricing)*
If a market model has a risk-neutral probability measure, then it does not admit arbitrage.

Proof. If a market model has a risk-neutral probability measure $\tilde{\mathbb{P}}$, then every discounted portfolio value process is a martingale under $\widetilde{\mathbb{P}}$. In particular, every portfolio value process satisfies $\widetilde{\mathbb{E}}[D(T)X(T)] = X(0)$. Let $X(t)$ be a portfolio value process with $X(0) = 0$. Then we have

$$\widetilde{\mathbb{E}}[D(T)X(T)] = 0 \tag{30}$$

Suppose $X(T)$ satisfies the first part of (27) (i.e., $\mathbb{P}\{X(T) < 0\} = 0$) since $\widetilde{\mathbb{P}}$ is equivalent to $\mathbb{P}$, we have also $\widetilde{\mathbb{P}}\{X(T) < 0\} = 0$. This, coupled with (30) implies $\widetilde{\mathbb{P}}\{X(T) > 0\} = 0$, for otherwise we would have $\widetilde{\mathbb{P}}\{D(T)X(T) > 0\} > 0$, which would imply $\widetilde{\mathbb{E}}[D(T)X(T)] > 0$. Because $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are equivalent, we have also $\mathbb{P}\{X(T) > 0\} = 0$. Hence, $X(t)$ is not an arbitrage. In fact, there cannot exist an arbitrage since every portfolio value process $X(t)$ satisfying $X(0) = 0$ cannot be an arbitrage. □

3.4 Uniqueness of the risk-neutral measure

Definition 3. A market model is complete if every derivative security can be hedged.

Similar to section 2.2, we would like to be sure we can hedge a short position in the derivative security whose payoff at time T is $V(T)$. We can define $V(t)$ by (16) so that $D(t)V(t)$ satisfies (15). We see that $D(t)V(t)$ is a martingale under $\widetilde{\mathbb{P}}$. According to the Martingale Representation Theorem 4, there are processes $\tilde{\Gamma}_1(u), \ldots, \tilde{\Gamma}_d(u)$ such that

$$D(t)V(t) = V(0) + \sum_{j=1}^{d} \int_0^t \tilde{\Gamma}_j(u)d\widetilde{W}_j(u), \quad 0 \le t \le T$$

On the other hand, consider a portfolio value process that begins at $X(0)$. According to (26) and (24),

$$\begin{aligned} d(D(t)X(t)) &= \sum_{i=1}^{m} \Delta_i(t)d\left(D(t)S_i(t)\right) \\ &= \sum_{j=1}^{d}\sum_{i=1}^{m} \Delta_i(t)D(t)S_i(t)\sigma_{ij}(t)d\widetilde{W}_j(t) \end{aligned}$$

or, equivalently,

$$D(t)X(t) = X(0) + \sum_{j=1}^{d} \int_0^t \sum_{i=1}^{m} \Delta_i(u)D(u)S_i(u)\sigma_{ij}(u)d\widetilde{W}_j(u)$$

In order to have $V(t) = X(t)$ for all t, should take $X(0) = V(0)$ and choose the portfolio processes $\Delta_1(t), \ldots, \Delta_m(t)$ so that the hedging equations (compared with (19))

$$\frac{\widetilde{\Gamma}_j(t)}{D(t)} = \sum_{i=1}^{m} \Delta_i(t)S_i(t)\sigma_{ij}(t), \quad j = 1, \ldots, d \tag{31}$$

are satisfied. These are d equations in m unknown processes $\Delta_1(t),\dots,\Delta_m(t)$.

Theorem 6. *(Second fundamental theorem of asset pricing)*
Consider a market model that has a risk-neutral probability measure. The model is complete if and only if the risk-neutral probability measure is unique.

Proof. Only if part. We first assume that the model is complete. We wish to show that there can be only one risk-neutral measure. Suppose the model has two risk-neutral measures, $\widetilde{\mathbb{P}}_1$ and $\widetilde{\mathbb{P}}_2$. Let A be a set in $\mathscr{F}$, which we assumed at the beginning of this section is the same as $\mathscr{F}(T)$. Consider the derivative security with payoff

$$V(T)=\mathbb{I}_A\frac{1}{D(T)}.$$

Because the model is complete, a short position in this derivative security can be hedged (i.e., there is a portfolio value process with some initial condition $X(0)$ that satisfies $X(T)=V(T)$) since both $\widetilde{\mathbb{P}}_1$ and $\widetilde{\mathbb{P}}_2$ are risk-neutral, the discounted portfolio value process $D(t)X(t)$ is a martingale under both $\widetilde{\mathbb{P}}_1$ and $\widetilde{\mathbb{P}}_2$. It follows that

$$\begin{aligned}\widetilde{\mathbb{P}}_1(A)&=\widetilde{\mathbb{E}}_1[D(T)V(T)]=\widetilde{\mathbb{E}}_1[D(T)X(T)]=X(0)\\&=\widetilde{\mathbb{E}}_2[D(T)X(T)]=\widetilde{\mathbb{E}}_2[D(T)V(T)]=\widetilde{\mathbb{P}}_2(A)\end{aligned}$$

since A is an arbitrary set in $\mathscr{F}$ and $\widetilde{\mathbb{P}}_1(A)=\widetilde{\mathbb{P}}_2(A)$, these two risk-neutral measures are really the same.

If part. For the converse, suppose there is only one risk-neutral measure. This means first of all that the filtration for the model is generated by the d dimensional Brownian motion driving the assets. If that were not the case (i.e., if there were other sources of uncertainty in the model besides the driving Brownian motions), then we could assign arbitrary probabilities to those sources of uncertainty without changing the distributions of the driving Brownian motions and hence without changing the distributions of the assets. This would permit us to create multiple risk-neutral measures. Because the driving Brownian motions are the only sources of uncertainty, the only way multiple risk-neutral measures can arise is via multiple solutions to the market price of risk equations (25).

Hence, uniqueness of the risk-neutral measure implies that the market price of risk equations (25) have only one solution $(\Theta_1(t),\dots,\Theta_d(t))$. For fixed t and ω, these equations are of the form

$$Ax=b \tag{32}$$

where

$$A=\begin{bmatrix}\sigma_{11}(t), & \sigma_{12}(t), & \dots, & \sigma_{1d}(t)\\ \sigma_{21}(t), & \sigma_{22}(t), & \dots, & \sigma_{2d}(t)\\ \vdots & \vdots & \vdots & \vdots\\ \sigma_{m1}(t), & \sigma_{m2}(t), & \dots, & \sigma_{md}(t)\end{bmatrix}$$

$$x = \begin{bmatrix} \Theta_1(t) \\ \Theta_2(t) \\ \vdots \\ \Theta_d(t) \end{bmatrix}$$

and

$$b = \begin{bmatrix} \alpha_1(t) - R(t) \\ \alpha_2(t) - R(t) \\ \vdots \\ \alpha_m(t) - R(t) \end{bmatrix}$$

Our assumption that there is only one risk-neutral measure means that the system of equations (32) has a unique solution x.

In order to be assured that every derivative security can be hedged, we must be able to solve the hedging equations (31) for $\Delta_1(t), \ldots, \Delta_m(t)$ no matter what values of $\frac{\tilde{\Gamma}_j(t)}{D(t)}$ appear on the left-hand side. For fixed t and ω the hedging equations are of the form

$$A^{\mathrm{tr}} y = c \tag{33}$$

where A^{tr} is the transpose of the matrix in (32),

$$y = \begin{bmatrix} \Delta_1(t)S_1(t) \\ \Delta_2(t)S_2(t) \\ \vdots \\ \Delta_m(t)S_m(t) \end{bmatrix}$$

and

$$c = \begin{bmatrix} \frac{\tilde{\Gamma}_1(t)}{D(t)} \\ \frac{\tilde{\Gamma}_2(t)}{D(t)} \\ \vdots \\ \frac{\tilde{\Gamma}_d(t)}{D(t)} \end{bmatrix}$$

In order to be assured that the market is complete, there must be a solution y to the system of equations (33) no matter what vector c appears on the right-hand side. If there is always a solution $y_1, \ldots, y_m$, then there are portfolio processes $\Delta_i(t) = \frac{y_i}{S_i(t)}$ satisfying the hedging equations (31) no matter what processes appear on the left-hand side of those equations. We could then conclude that a short position in an arbitrary derivative security can be hedged.

Finally, note that the uniqueness of the solution x to (32) implies the existence of a solution y to (33) (why?). Consequently, uniqueness of the risk-neutral measure implies that the market model is complete. □

Exercise 5. *(Hedging a cash flow)*. Let $W(t), 0 \le t \le T$, be a Brownian motion on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, and let $\mathscr{F}(t), 0 \le t \le T$, be the filtration generated by this Brownian motion. Let the mean rate of return $\alpha(t)$, the interest rate $R(t)$, and the volatility $\sigma(t)$ be adapted processes, and assume that $\sigma(t)$ is never zero. Consider a stock price process whose differential is given by

$$dS(t) = \alpha(t)S(t)dt + \sigma(t)S(t)dW(t), \quad 0 \leq t \leq T$$

Suppose an agent must pay a cash flow at rate $C(t)$ at each time t, where $C(t), 0 \leq t \leq T$, is an adapted process. If the agent holds $\Delta(t)$ shares of stock at each time t, then the differential of her portfolio value will be

$$dX(t) = \Delta(t)dS(t) + R(t)(X(t) - \Delta(t)S(t))dt - C(t)dt$$

Show that there is a nonrandom value of $X(0)$ and a portfolio process $\Delta(t)$, $0 \leq t \leq T$, such that $X(T) = 0$ almost surely.

Chapter 5: Connection with partial differential equations

Gechun Liang

There are two ways to compute a derivative security price: (1) use *Monte Carlo simulation* to generate paths of the underlying security or securities under the risk-neutral measure and use these paths to estimate the risk-neutral expected discounted payoff; or (2) *numerically solve a partial differential equation*. This chapter addresses the second of these methods by showing how to connect the risk-neutral pricing problem to partial differential equations.

1 Stochastic differential equations

A stochastic differential equation is an equation of the form

$$dX(u) = \beta(u,X(u))du + \gamma(u,X(u))dW(u)$$

Here $\beta(u,x)$ and $\gamma(u,x)$ are given functions, called the drift and diffusion, respectively. In addition to this equation, an initial condition of the form $X(t) = x$, where $t \geq 0$ and $x \in \mathbb{R}$, is specified. The problem is then to find a stochastic process $X(T)$, defined for $T \geq t$, such that

$$\begin{aligned} X(t) &= x \\ X(T) &= X(t) + \int_t^T \beta(u,X(u))du + \int_t^T \gamma(u,X(u))dW(u) \end{aligned} \tag{1}$$

Under mild conditions[1] on the functions $\beta(u,x)$ and $\gamma(u,x)$, there exists a unique process $X(T), T \geq t$, satisfying (1).

Gechun Liang
Department of Statistics, University of Warwick, U.K. e-mail: g.liang@warwick.ac.uk

[1] For example, $\beta(u,x)$ and $\gamma(u,x)$ are Lipschitz continuous in space variable x uniformly in u. See *Brownian Motion and Stochastic Calculus* by Karatzas and Shreve.

Although stochastic differential equations are, in general, difficult to solve, a one-dimensional linear stochastic differential equation can be solved explicitly. This is a stochastic differential equation of the form

$$dX(u) = (a(u)+b(u)X(u))du + (\gamma(u)+\sigma(u)X(u))dW(u)$$

where $a(u), b(u), \sigma(u)$, and $\gamma(u)$ are nonrandom functions of time. Indeed, this equation can even be solved when $a(u), b(u), \gamma(u)$, and $\sigma(u)$ are adapted random processes as shown in the following exercise.

Exercise 1. Consider

$$dX(u) = (a(u)+b(u)X(u))du + (\gamma(u)+\sigma(u)X(u))dW(u) \tag{2}$$

where $W(u)$ is a Brownian motion relative to a filtration $\mathscr{F}(u), u \geq 0$, and we allow $a(u), b(u), \gamma(u)$, and $\sigma(u)$ to be processes adapted to this filtration. Fix an initial time $t \geq 0$ and an initial position $x \in \mathbb{R}$. Define

$$\begin{aligned} Z(u) &= \exp\left\{\int_t^u \sigma(v)dW(v) + \int_t^u \left(b(v) - \tfrac{1}{2}\sigma^2(v)\right)dv\right\} \\ Y(u) &= x + \int_t^u \tfrac{a(v)-\sigma(v)\gamma(v)}{Z(v)}dv + \int_t^u \tfrac{\gamma(v)}{Z(v)}dW(v) \end{aligned}$$

1. Show that $Z(t) = 1$ and

$$dZ(u) = b(u)Z(u)du + \sigma(u)Z(u)dW(u), u \geq t$$

2. By its very definition, $Y(u)$ satisfies $Y(t) = x$ and

$$dY(u) = \frac{a(u)-\sigma(u)\gamma(u)}{Z(u)}du + \frac{\gamma(u)}{Z(u)}dW(u), u \geq t$$

   Show that $X(u) = Y(u)Z(u)$ solves the stochastic differential equation (2) and satisfies the initial condition $X(t) = x$.

Example 1. (Geometric Brownian motion). The stochastic differential equation for geometric Brownian motion is

$$dS(u) = \alpha S(u)du + \sigma S(u)dW(u)$$

In the notation of (2), $a(u) = \gamma(u) = 0$, $b(u) = \alpha$ and $\sigma(u) = \sigma$. Hence, if $S(t) = x$, then

$$S(T) = x\exp\left\{\sigma(W(T)-W(t)) + \left(\alpha - \frac{1}{2}\sigma^2\right)(T-t)\right\}$$

Example 2. (Hull-White interest rate model). Consider the stochastic differential equation

$$dR(u) = (a(u)-b(u)R(u))du + \sigma(u)d\widetilde{W}(u)$$

where $a(u), b(u)$, and $\sigma(u)$ are nonrandom positive functions[2]of the time variable u, and $\widetilde{W}(u)$ is a Brownian motion under a risk-neutral measure $\widetilde{\mathbb{P}}$. From Exercise 1, we obtain

$$R(T) = re^{-\int_t^T b(v)dv} + \int_t^T e^{-\int_u^T b(v)dv}\alpha(u)du + \int_t^T e^{-\int_u^T b(v)dv}\sigma(u)d\widetilde{W}(u)$$

with $R(t) = r$.

Note that the Itô's integral $\int_t^T e^{-\int_u^T b(v)dv}\sigma(u)d\widetilde{W}(u)$ of the nonrandom integrand $e^{-\int_u^T b(v)dv}\sigma(u)$ is normally distributed

$$N\left(0, \int_t^T e^{-2\int_u^T b(v)dv}\sigma^2(u)du\right)$$

(Why?) The other terms appearing in the formula above for $R(T)$ are nonrandom. Therefore, under the risk-neutral measure $\widetilde{\mathbb{P}}, R(T)$ is normally distributed

$$N\left(re^{-\int_t^T b(v)dv} + \int_t^T e^{-\int_u^T b(v)dv}\alpha(u)du, \int_t^T e^{-2\int_u^T b(v)dv}\sigma^2(u)du\right)$$

In particular, there is a positive probability that $R(T)$ is negative. This is one of the principal objections to the Hull-White model.

In the special case of Vasicek model, it follows that $R(T)$ follows (with initial time $t = 0$)

$$N\left(e^{-\beta t}R(0) + \frac{a}{b}\left(1 - e^{-bt}\right), \frac{\sigma^2}{2\beta}\left(1 - e^{-2\beta t}\right)\right)$$

Example 3. (Cox-Ingersoll-Ross interest rate model). In the Cox-Ingersoll-Ross (CIR) model, the interest rate is given by the stochastic differential equation

$$dR(u) = (\alpha - \beta R(u))du + \sigma\sqrt{R(u)}d\widetilde{W}(u)$$

where α, β, and σ are positive constants. Suppose an initial condition $R(0) = r$ is given (with initial time $t = 0$). *Note that in the CIR model, $R(T)$ is always nonnegative.*

Although one cannot derive a closed-form solution for $R(t)$, the distribution of $R(t)$ for each positive t can be determined. That computation would take us too far afield. We instead content ourselves with the derivation of the mean and variance of $R(t)$. To do this, we use the function $f(t,x) = e^{\beta t}x$ and compute

$$\begin{aligned}
&d\left(e^{\beta t}R(t)\right) = df(t, R(t)) \\
&= f_t(t, R(t))dt + f_x(t, R(t))dR(t) + \tfrac{1}{2}f_{xx}(t, R(t))dR(t)dR(t) \\
&= \beta e^{\beta t}R(t)dt + e^{\beta t}(\alpha - \beta R(t))dt + e^{\beta t}\sigma\sqrt{R(t)}d\tilde{W}(t) \\
&= \alpha e^{\beta t}dt + \sigma e^{\beta t}\sqrt{R(t)}d\tilde{W}(t)
\end{aligned}$$

[2] When they are constants: $a(u) = a$, $b(u) = b$ and $\sigma(u) = \sigma$, the model is called Vasicek model.

Integration of both sides yields

$$\begin{aligned} e^{\beta t}R(t) &= R(0) + \alpha \int_0^t e^{\beta u} du + \sigma \int_0^t e^{\beta u} \sqrt{R(u)} d\tilde{W}(u) \\ &= r + \frac{\alpha}{\beta}\left(e^{\beta t} - 1\right) + \sigma \int_0^t e^{\beta u} \sqrt{R(u)} d\tilde{W}(u) \end{aligned}$$

Recalling that the expectation of an Itô's integral is zero, we obtain

$$e^{\beta t}\mathbb{E}R(t) = r + \frac{\alpha}{\beta}\left(e^{\beta t} - 1\right)$$

or, equivalently,

$$\mathbb{E}R(t) = e^{-\beta t} r + \frac{\alpha}{\beta}\left(1 - e^{-\beta t}\right)$$

This is the same expectation as in the Vasicek model.

To compute the variance of $R(t)$, we set $X(t) = e^{\beta t}R(t)$, for which we have

$$dX(t) = \alpha e^{\beta t} dt + \sigma e^{\beta t} \sqrt{R(t)} d\tilde{W}(t) = \alpha e^{\beta t} dt + \sigma e^{\frac{\beta t}{2}} \sqrt{X(t)} d\tilde{W}(t)$$

and $\mathbb{E}X(t) = r + \frac{\alpha}{\beta}\left(e^{\beta t} - 1\right)$. According to the Itô's formula (with $f(x) = x^2, f'(x) = 2x$, and $f''(x) = 2$)

$$\begin{aligned} d\left(X^2(t)\right) &= 2X(t)dX(t) + dX(t)dX(t) \\ &= 2\alpha e^{\beta t} X(t) dt + 2\sigma e^{\frac{\beta t}{2}} X^{\frac{3}{2}}(t) d\tilde{W}(t) + \sigma^2 e^{\beta t} X(t) dt \end{aligned}$$

Integration then yields

$$X^2(t) = X^2(0) + \left(2\alpha + \sigma^2\right) \int_0^t e^{\beta u} X(u) du + 2\sigma \int_0^t e^{\frac{\beta u}{2}} X^{\frac{3}{2}}(u) d\tilde{W}(u)$$

Taking expectations, using the fact that the expectation of an Itô's integral is zero and the formula already derived for $\mathbb{E}X(t)$, we obtain

$$\begin{aligned} \mathbb{E}X^2(t) &= X^2(0) + \left(2\alpha + \sigma^2\right) \int_0^t e^{\beta u} \mathbb{E}X(u) du \\ &= r^2 + \left(2\alpha + \sigma^2\right) \int_0^t e^{\beta u} \left(r + \frac{\alpha}{\beta}\left(e^{\beta u} - 1\right)\right) du \\ &= r^2 + \frac{2\alpha + \sigma^2}{\beta}\left(r - \frac{\alpha}{\beta}\right)\left(e^{\beta t} - 1\right) + \frac{2\alpha + \sigma^2}{2\beta} \cdot \frac{\alpha}{\beta}\left(e^{2\beta t} - 1\right) \end{aligned}$$

Therefore,

$$
\begin{aligned}
\mathbb{E}R^2(t) = &\ e^{-2\beta t}\mathbb{E}X^2(t) \\
= &\ e^{-2\beta t}r^2 + \frac{2\alpha+\sigma^2}{\beta}\left(r-\frac{\alpha}{\beta}\right)\left(e^{-\beta t}-e^{-2\beta t}\right) \\
&+ \frac{\alpha\left(2\alpha+\sigma^2\right)}{2\beta^2}\left(1-e^{-2\beta t}\right)
\end{aligned}
$$

Finally,

$$
\begin{aligned}
\mathrm{Var}(R(t)) &= \mathbb{E}R^2(t) - (\mathbb{E}R(t))^2 \\
&= e^{-2\beta t}r^2 + \frac{2\alpha+\sigma^2}{\beta}\left(r-\frac{\alpha}{\beta}\right)\left(e^{-\beta t}-e^{-2\beta t}\right) \\
&\quad + \frac{\alpha\left(2\alpha+\sigma^2\right)}{2\beta^2}\left(1-e^{-2\beta t}\right) - e^{-2\beta t}r^2 \\
&\quad - \frac{2\alpha}{\beta}r\left(e^{-\beta t}-e^{-2\beta t}\right) - \frac{\alpha^2}{\beta^2}\left(1-e^{-\beta t}\right)^2 \\
&= \frac{\sigma^2}{\beta}r\left(e^{-\beta t}-e^{-2\beta t}\right) + \frac{\alpha\sigma^2}{2\beta^2}\left(1-2e^{-\beta t}+e^{-2\beta t}\right)
\end{aligned}
$$

2 The Markov property

Consider the stochastic differential equation (1). Let $0 \leq t \leq T$ be given, and let $h(y)$ be a Borel-measurable function. Denote by

$$g(t,x) = \mathbb{E}^{t,x}h(X(T)) \tag{3}$$

Theorem 1. *Let $X(u), u \geq 0$, be a solution to the stochastic differential equation (1) with initial condition given at time 0. Then, for $0 \leq t \leq T$*

$$\mathbb{E}[h(X(T)) \mid \mathscr{F}(t)] = g(t, X(t))$$

While the details of the proof of the above theorem are quite technical and will not be given, the intuitive content is clear. Suppose the process $X(u)$ begins at time zero, being generated by the stochastic differential equation (1) and one watches it up to time t. Suppose now that one is asked, based on this information, to compute the conditional expectation of $h(X(T))$, where $T \geq t$. Then one should pretend that the process is starting at time t at its current position, generate the solution to the stochastic differential equation corresponding to this initial condition, and compute the expected value of $h(X(T))$ generated in this way. In other words, replace $X(t)$ by a dummy x in order to hold it constant, compute $g(t,x) = \mathbb{E}^{t,x}h(X(T))$, and after computing this function put the random variable $X(t)$ back in place of the dummy x.

Notice in the discussion above that although one watches the stochastic process $X(u)$ for $0 \le u \le t$, the only relevant piece of information when computing $\mathbb{E}[h(X(T)) \mid \mathscr{F}(t)]$ is the value of $X(t)$. *This means that $X(t)$, as the solution of SDE (1), is a Markov process.*

3 Partial differential equations

Theorem 2. *(Feynman-Kac). Consider X governed by the stochastic differential equation (1) and the function g given by (3). Then, $g(t,x)$ satisfies the partial differential equation*

$$g_t(t,x)+\beta(t,x)g_x(t,x)+\frac{1}{2}\gamma^2(t,x)g_{xx}(t,x)=0 \tag{4}$$

and the terminal condition

$$g(T,x)=h(x) \text{ for all } x$$

Proof. Let $X(t)$ be the solution to the stochastic differential equation (1) starting at time zero. Since $g(t,X(t))$ is a martingale(why?), the dt term in the differential $dg(t,X(t))$ must be zero. If it were positive at any time, then $g(t,X(t))$ would have a tendency to rise at that time; if it were negative, $g(t,X(t))$ would have a tendency to fall. Omitting the argument $(t,X(t))$ in several places below, we compute

$$\begin{aligned} dg(t,X(t)) &= g_t dt + g_x dX + \frac{1}{2} g_{xx} dX dX \\ &= g_t dt + \beta g_x dt + \gamma g_x dW + \frac{1}{2}\gamma^2 g_{xx} dt \\ &= \left[g_t + \beta g_x + \frac{1}{2}\gamma^2 g_{xx} \right] dt + \gamma g_x dW \end{aligned}$$

Setting the dt term to zero and putting back the argument $(t,X(t))$, we obtain

$$g_t(t,X(t))+\beta(t,X(t))g_x(t,X(t))+\frac{1}{2}\gamma^2(t,X(t))g_{xx}(t,X(t))=0$$

along every path of X. Therefore,

$$g_t(t,x)+\beta(t,x)g_x(t,x)+\frac{1}{2}\gamma^2(t,x)g_{xx}(t,x)=0$$

at every point (t,x) that can be reached by $(t,X(t))$. □

Theorem 3. *(Discounted Feynman-Kac). Consider X governed by the stochastic differential equation (1) and define the function*

$$f(t,x) = \mathbb{E}^{t,x}\left[e^{-r(T-t)}h(X(T))\right]$$

Then $f(t,x)$ satisfies the partial differential equation

$$f_t(t,x) + \beta(t,x)f_x(t,x) + \frac{1}{2}\gamma^2(t,x)f_{xx}(t,x) = rf(t,x) \tag{5}$$

and the terminal condition

$$f(T,x) = h(x) \text{ for all } x$$

Proof. Let $X(t)$ be the solution to the stochastic differential equation (1) starting at time zero. Then

$$f(t,X(t)) = \mathbb{E}\left[e^{-r(T-t)}h(X(T)) \mid \mathscr{F}(t)\right]$$

However, it is not the case that $f(t,X(t))$ is a martingale. Indeed, if $0 \le s \le t \le T$, then

$$\begin{aligned}\mathbb{E}[f(t,X(t)) \mid \mathscr{F}(s)] &= \mathbb{E}\left[\mathbb{E}\left[e^{-r(T-t)}h(X(T)) \mid \mathscr{F}(t)\right] \mid \mathscr{F}(s)\right] \\ &= \mathbb{E}\left[e^{-r(T-t)}h(X(T)) \mid \mathscr{F}(s)\right]\end{aligned}$$

which is not the same as

$$f(s,X(s)) = \mathbb{E}\left[e^{-r(T-s)}h(X(T)) \mid \mathscr{F}(s)\right]$$

However, note that

$$e^{-rt}f(t,X(t)) = \mathbb{E}\left[e^{-rT}h(X(T)) \mid \mathscr{F}(t)\right].$$

We may now apply iterated conditioning to show that $e^{-rt}f(t,X(t))$ is a martingale. The differential of this martingale is

$$\begin{aligned}d\left(e^{-rt}f(t,X(t))\right) &= e^{-rt}\left[-rf\,dt + f_t\,dt + f_x\,dX + \frac{1}{2}f_{xx}\,dX\,dX\right] \\ &= e^{-rt}\left[-rf + f_t + \beta f_x + \frac{1}{2}\gamma^2 f_{xx}\right]dt + e^{-rt}\gamma f_x\,dW\end{aligned}$$

Setting the dt term equal to zero, we conclude (5). □

Example 4. (Black-Scholes PDE). Let $h(S(T))$ be the payoff at time T of a derivative security whose underlying asset is the geometric Brownian motion

$$dS(u) = \alpha S(u)du + \sigma S(u)dW(u).$$

We may rewrite this as

$$dS(u) = rS(u)du + \sigma S(u)d\widetilde{W}(u)$$

where $\widetilde{W}(u)$ is a Brownian motion under the risk-neutral probability measure $\widetilde{\mathbb{P}}$. Here we assume that σ and the interest rate r are constant. According to the risk-neutral pricing formula, the price of the derivative security at time t is

$$V(t) = \widetilde{\mathbb{E}}\left[e^{-r(T-t)}h(S(T)) \mid \mathscr{F}(t)\right]$$

Because the stock price is Markov and the payoff is a function of the stock price alone, there is a function $v(t,x)$ such that $V(t) = v(t,S(t))$. Moreover, the function $v(t,x)$ must satisfy the discounted partial differential equation (5). This is the Black-Scholes partial differential equation

$$v_t(t,x) + rxv_x(t,x) + \frac{1}{2}\sigma^2 x^2 v_{xx}(t,x) = rv(t,x)$$

with the terminal condition $v(T,x) = h(x)$. □

It has been observed in markets that if one assumes a constant volatility, the parameter σ that makes the theoretical option price given by Black-Scholes PDE agree with the market price, the so-called implied volatility, is different for options having different strikes. In fact, this implied volatility is generally a convex function of the strike price. One refers to this phenomenon as the volatility smile.

Exercise 2. (Kolmogorov backward equation). Consider the stochastic differential equation

$$dX(u) = \beta(u,X(u))du + \gamma(u,X(u))dW(u)$$

We assume that, just as with a geometric Brownian motion, if we begin a process at an arbitrary initial positive value $X(t) = x$ at an arbitrary initial time t and evolve it forward using this equation, its value at each time $T > t$ could be any positive number but cannot be less than or equal to zero. For $0 \le t < T$, let $p(t,T,x,y)$ be the transition density for the solution to this equation (i.e., if we solve the equation with the initial condition $X(t) = x$ then the random variable $X(T)$ has density $p(t,T,x,y)$ in the y variable). We are assuming that $p(t,T,x,y) = 0$ for $0 \le t < T$ and $y \le 0$ Show that $p(t,T;x,y)$ satisfies the Kolmogorov backward equation

$$-p_t(t,T,x,y) = \beta(t,x)p_x(t,T,x,y) + \frac{1}{2}\gamma^2(t,x)p_{xx}(t,T,x,y)$$

Exercise 3. (Kolmogorov forward equation/Fokker-Planck equation). We begin with the same stochastic differential equation,

$$dX(u) = \beta(u,X(u))du + \gamma(u,X(u))dW(u)$$

as in Exercise 2, use the same notation $p(t,T,x,y)$ for the transition density, and again assume that $p(t,T,x,y) = 0$ for $0 \le t < T$ and $y \le 0$. In this problem, we show that $p(t,T,x,y)$ satisfies the Kolmogorov forward equation

$$\frac{\partial}{\partial T}p(t,T,x,y) = -\frac{\partial}{\partial y}(\beta(t,y)p(t,T,x,y)) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\gamma^2(T,y)p(t,T,x,y)\right)$$

In contrast to the Kolmogorov backward equation, in which T and y were held constant and the variables were t and x, here t and x are held constant and the variables are y and T. The variables t and x are sometimes called the *backward variables*, and T and y are called the *forward variables*.

1. Let b be a positive constant and let $h_b(y)$ be a function with continuous first and second derivatives such that $h_b(x) = 0$ for all $x \le 0, h_b'(x) = 0$ for all $x \ge b$, and $h_b(b) = h_b'(b) = 0$. Let $X(u)$ be the solution to the stochastic differential equation with initial condition $X(t) = x \in (0,b)$ and use Itô's formula to compute $dh_b(X(u))$
2. Let $0 \le t < T$ be given, and integrate the equation you obtained in part (1) from t to T. Take expectations and use the fact that $X(u)$ has density $p(t,u,x,y)$ in the y variable to obtain

$$\begin{aligned}\int_0^b h_b(y)p(t,T,x,y)dy = h_b(x) &+ \int_t^T \int_0^b \beta(u,y)p(t,u,x,y)h_b'(y)dydu \\ &+ \frac{1}{2}\int_t^T \int_0^b \gamma^2(u,y)p(t,u,x,y)h_b''(y)dydu \qquad (6)\end{aligned}$$

3. Integrate the integrals $\int_0^b \cdots dy$ on the right-hand side of the above equation (6) by parts to obtain

$$\begin{aligned}&\int_0^b h_b(y)p(t,T,x,y)dy \\ &= h_b(x) - \int_t^T \int_0^b \frac{\partial}{\partial y}[\beta(u,y)p(t,u,x,y)]h_b(y)dydu \\ &\quad + \frac{1}{2}\int_t^T \int_0^b \frac{\partial^2}{\partial y^2}\left[\gamma^2(u,y)p(t,u,x,y)\right]h_b(y)dydu \qquad (7)\end{aligned}$$

4. Differentiate (7) with respect to T to obtain

$$\begin{aligned}\int_0^b h_b(y)\Bigg[\frac{\partial}{\partial T}p(t,T,x,y) &+ \frac{\partial}{\partial y}(\beta(T,y)p(t,T,x,y)) \\ &- \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\gamma^2(T,y)p(t,T,x,y)\right)\Bigg]dy = 0 \qquad (8)\end{aligned}$$

5. Use (8) to show that there cannot be numbers $0 < y_1 < y_2$ such that

$$\begin{aligned}\frac{\partial}{\partial T}p(t,T,x,y) &+ \frac{\partial}{\partial y}(\beta(T,y)p(t,T,x,y)) \\ &- \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\gamma^2(T,y)p(t,T,x,y)\right) > 0 \text{ for all } y \in (y_1,y_2)\end{aligned}$$

   Similarly, there cannot be numbers $0 < y_1 < y_2$ such that

$$\frac{\partial}{\partial T}p(t,T,x,y)+\frac{\partial}{\partial y}(\beta(T,y)p(t,T,x,y))$$
$$-\frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\gamma^2(T,y)p(t,T,x,y)\right)<0 \text{ for all } y\in[y_1,y_2]$$

This is as much as you need to do for this problem. It is now obvious that if

$$\frac{\partial}{\partial T}p(t,T,x,y)+\frac{\partial}{\partial y}(\beta(T,y)p(t,T,x,y))-\frac{1}{2}\frac{\partial}{\partial y^2}\left(\gamma^2(T,y)p(t,T,x,y)\right)$$

is a continuous function of y, then this expression must be zero for every $y>0$ and hence $p(t,T,x,y)$ satisfies the Kolmogorov forward equation stated at the beginning of this problem.

Exercise 4. (Implying the risk-neutral distribution). Let $S(t)$ be the price of an underlying asset, which is not necessarily a geometric Brownian motion (i.e., does not necessarily have constant volatility). With $S(0)=x$ the risk-neutral pricing formula for the price at time zero of a European call on this asset, paying $(S(T)-K)^+$ at time T, is

$$c(0,T,x,K)=\widetilde{\mathbb{E}}\left[e^{-rT}(S(T)-K)^+\right]$$

(Normally we consider this as a function of the current time 0 and the current stock price x, but in this exercise we shall also treat the expiration time T and the strike price K as variables, and for that reason we include them as arguments of c.) We denote by $\tilde{p}(0,T,x,y)$ the risk-neutral density in the y variable of the distribution of $S(T)$ when $S(0)=x$. Then we may rewrite the risk-neutral pricing formula as

$$c(0,T,x,K)=e^{-rT}\int_K^\infty (y-K)\tilde{p}(0,T,x,y)dy \tag{9}$$

Suppose we know the market prices for calls of all strikes (i.e., we know $c(0,T,x,K)$ for all $K>0$). We can then compute $c_K(0,T,x,K)$ and $c_{KK}(0,T,x,K)$, the first and second derivatives of the option price with respect to the strike. Differentiate (9) twice with respect to K to obtain the equations

$$c_K(0,T,x,K)=-e^{-rT}\int_K^\infty \tilde{p}(0,T,x,y)dy$$
$$c_{KK}(0,T,x,K)=e^{-rT}\tilde{p}(0,T,x,K)$$

Exercise 5. (Implying the volatility surface). Assume that a stock price evolves according to the stochastic differential equation

$$dS(u)=rS(u)dt+\sigma(u,S(u))S(u)d\widetilde{W}(u)$$

where the interest rate r is constant, the volatility $\sigma(u,x)$ is a function of time and the underlying stock price, and $\widetilde{W}$ is a Brownian motion under the riskneutral measure $\widetilde{\mathbb{P}}$. Let $\tilde{p}(t,T,x,y)$ denote the transition density.

According to Exercise 3, the transition density $\tilde{p}(t,T,x,y)$ satisfies the Kolmogorov forward equation

$$\frac{\partial}{\partial T}\tilde{p}(t,T,x,y) = -\frac{\partial}{\partial y}(ry\tilde{p}(t,T,x,y)) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\sigma^2(T,y)y^2\tilde{p}(t,T,x,y)\right) \tag{10}$$

Let

$$c(0,T,x,K) = e^{-rT}\int_K^\infty (y-K)\tilde{p}(0,T,x,y)dy \tag{11}$$

denote the time-zero price of a call expiring at time T, struck at K, when the initial stock price is $S(0)=x$. Note that

$$c_T(0,T,x,K) = -rc(0,T,x,K) + e^{-rT}\int_K^\infty (y-K)\tilde{p}_T(0,T,x,y)dy \tag{12}$$

1. Integrate once by parts to show that

$$-\int_K^\infty (y-K)\frac{\partial}{\partial y}(ry\tilde{p}(0,T,x,y))dy = \int_K^\infty ry\tilde{p}(0,T,x,y)dy \tag{13}$$

   You may assume that

$$\lim_{y\to\infty}(y-K)ry\tilde{p}(0,T,x,y) = 0 \tag{14}$$

2. Integrate by parts and then integrate again to show that

$$\begin{aligned}\frac{1}{2}\int_K^\infty (y-K)\frac{\partial^2}{\partial y^2}\left(\sigma^2\ (T,y)y^2\tilde{p}(0,T,x,y)\right)dy&\\ =\frac{1}{2}\sigma^2(T,K)K^2\tilde{p}(0,T,x,K)&\end{aligned} \tag{15}$$

   You may assume that

$$\begin{aligned}\lim_{y\to\infty}(y-K)\frac{\partial}{\partial y}\left(\sigma^2(T,y)y^2\tilde{p}(0,T,x,y)\right) &= 0\\ \lim_{y\to\infty}\sigma^2(T,y)y^2\tilde{p}(0,T,x,y) &= 0\end{aligned}$$

3. Now use (12) (11), (10), (13) (15) and Exercise 4 in that order to obtain the equation

$$\begin{aligned}&c_T(0,T,x,K)\\ =&e^{-rT}rK\int_K^\infty \tilde{p}(0,T,x,y)dy + \frac{1}{2}e^{-rT}\sigma^2(T,K)K^2\tilde{p}(0,T,x,K)\\ =&-rKc_K(0,T,x,K) + \frac{1}{2}\sigma^2(T,K)K^2c_{KK}(0,T,x,K)\end{aligned} \tag{16}$$

   This is the end of the problem. Note that under the assumption that

$$c_{KK}(0,T,x,K) \neq 0,$$

(16) can be solved for the volatility term $\sigma^2(T,K)$ in terms of the quantities $c_T(0,T,x,K), c_K(0,T,x,K)$, and $c_{KK}(0,T,x,K)$ which can be inferred from market prices.

4 Interest rate models

The simplest models for fixed income markets begin with a stochastic differential equation for the interest rate, e.g.,

$$dR(t) = \beta(t,R(t))dt + \gamma(t,R(t))d\widetilde{W}(t)$$

where $\widetilde{W}(t)$ is a Brownian motion under a risk-neutral probability measure $\widetilde{\mathbb{P}}$.

Models for the interest rate $R(t)$ are sometimes called short-rate models because $R(t)$ is the interest rate for short-term borrowing. When the interest rate is determined by only one stochastic differential equation, as is the case in this section, the model is said to have one factor. *The primary shortcoming of one-factor models is that they cannot capture complicated yield curve behavior; they tend to produce parallel shifts in the yield curve but not changes in its slope or curvature.*

A zero-coupon bond is a contract promising to pay a certain "face" amount, which we take to be 1, at a fixed maturity date T. Prior to that, the bond makes no payments. The risk-neutral pricing formula says that the discounted price of this bond should be a martingale under the risk-neutral measure. In other words, for $0 \leq t \leq T$, the price of the bond $B(t,T)$ should satisfy

$$D(t)B(t,T) = \widetilde{\mathbb{E}}[D(T) \mid \mathscr{F}(t)]$$

(Note that $B(T,T) = 1$.) This gives us the zero-coupon bond pricing formula

$$B(t,T) = \widetilde{\mathbb{E}}\left[e^{-\int_t^T R(s)ds} \mid \mathscr{F}(t)\right]$$

which we take as a definition. Once zero-coupon bond prices have been computed, we can define the yield between times t and T to be

$$Y(t,T) = -\frac{1}{T-t}\log B(t,T)$$

or, equivalently,

$$B(t,T) = e^{-Y(t,T)(T-t)}.$$

Since R is given by a stochastic differential equation, it is a Markov process and we must have

$$B(t,T) = f(t,R(t))$$

for some function $f(t,r)$ of the dummy variables t and r. The discounted Feynman-Kac formula then yields that

$$f_t(t,r)+\beta(t,r)f_r(t,r)+\frac{1}{2}\gamma^2(t,r)f_{rr}(t,r)=rf(t,r)$$

We also have the terminal condition

$$f(T,r)=1 \text{ for all } r$$

because the value of the bond at maturity is its face value 1 .

Example 5. (Hull-White interest rate model). In the Hull-White model, the evolution of the interest rate is given by

$$dR(t)=(a(t)-b(t)R(t))dt+\sigma(t)d\widetilde{W}(t)$$

where $a(t), b(t)$, and $\sigma(t)$ are nonrandom positive functions of time. The partial differential equation for the zero-coupon bond price becomes

$$f_t(t,r)+(a(t)-b(t)r)f_r(t,r)+\frac{1}{2}\sigma^2(t)f_{rr}(t,r)=rf(t,r)$$

We initially guess and subsequently verify that the solution has the form

$$f(t,r)=e^{-rC(t,T)-A(t,T)}$$

for some nonrandom functions $C(t,T)$ and $A(t,T)$ to be determined. These are functions of $t\in[0,T]$; the maturity T is fixed. In this case, the yield

$$Y(t,T)=-\frac{1}{T-t}\log f(t,r)=\frac{1}{T-t}(rC(t,T)+A(t,T))$$

is an *a*ffine function of r. The Hull-White model is a special case of a class of models called affine yield models.

Exercise 6. Verify that the functions $C(t,T)$ and $A(t,T)$ satisfy

$$C'(t,T)=b(t)C(t,T)-1$$

and

$$A'(t,T)=-a(t)C(t,T)+\frac{1}{2}\sigma^2(t)C^2(t,T)$$

The solutions to these equations satisfying the terminal conditions

$$C(T,T)=A(T,T)=0$$

are

$$
\begin{aligned}
C(t,T) &= \int_t^T e^{-\int_t^s b(v)dv} ds \\
A(t,T) &= \int_t^T \left(a(s)C(s,T) - \frac{1}{2}\sigma^2(s)C^2(s,T) \right) ds
\end{aligned}
$$

Example 6. (Cox-Ingersoll-Ross interest rate model). In the CIR model, the evolution of the interest rate is given by

$$dR(t) = (a - bR(t))dt + \sigma\sqrt{R(t)}d\widetilde{W}(t)$$

where a, b, and σ are positive constants. The partial differential equation for the bond price becomes

$$f_t(t,r) + (a - br)f_r(t,r) + \frac{1}{2}\sigma^2 r f_{rr}(t,r) = rf(t,r)$$

Again, we initially guess and subsequently verify that the solution has the form

$$f(t,r) = e^{-rC(t,T) - A(t,T)}$$

The Cox-Ingersoll-Ross model is another example of an affine yield model.

Exercise 7. Verify that the functions $C(t,T)$ and $A(t,T)$ satisfy

$$
\begin{aligned}
C'(t,T) &= bC(t,T) + \frac{1}{2}\sigma^2 C^2(t,T) - 1 \\
A'(t,T) &= -aC(t,T)
\end{aligned}
$$

The solutions to these equations satisfying the terminal conditions

$$C(T,T) = A(T,T) = 0$$

are

$$
\begin{aligned}
C(t,T) &= \frac{\sinh(\gamma(T-t))}{\gamma\cosh(\gamma(T-t)) + \frac{1}{2}b\sinh(\gamma(T-t))} \\
A(t,T) &= -\frac{2a}{\sigma^2}\log\left[\frac{\gamma e^{\frac{1}{2}b(T-t)}}{\gamma\cosh(\gamma(T-t)) + \frac{1}{2}b\sinh(\gamma(T-t))} \right]
\end{aligned}
$$

where $\gamma = \frac{1}{2}\sqrt{b^2 + 2\sigma^2}$, $\sinh u = \frac{e^u - e^{-u}}{2}$, and $\cosh u = \frac{e^u + e^{-u}}{2}$.

5 Multidimensional Feynman-Kac theorems

The Feynman-Kac and discounted Feynman-Kac Theorems, Theorems 2 and 3, have multidimensional versions. The number of differential equations and the number of Brownian motions entering those differential equations can both be larger than one and do not need to be the same. We illustrate the general situation by working out the details for two stochastic differential equations driven by two Brownian motions.

Let $W(t) = (W_1(t), W_2(t))$ be a two-dimensional Brownian motion (i.e., a vector of two independent, one-dimensional Brownian motions). Consider two stochastic differential equations

$$\begin{aligned} dX_1(u) = \beta_1(u, X_1(u), X_2(u))\,du + \gamma_{11}(u, X_1(u), X_2(u))\,dW_1(u) \\ +\gamma_{12}(u, X_1(u), X_2(u))\,dW_2(u) \\ dX_2(u) = \beta_2(u, X_1(u), X_2(u))\,du + \gamma_{21}(u, X_1(u), X_2(u))\,dW_1(u) \\ +\gamma_{22}(u, X_1(u), X_2(u))\,dW_2(u) \end{aligned}$$

The solution to this pair of stochastic differential equations, starting at $X_1(t) = x_1$ and $X_2(t) = x_2$, depends on the specified initial time t and the initial positions x_1 and x_2. Regardless of the initial condition, the solution is a Markov process.

Let a Borel-measurable function $h(y_1, y_2)$ be given. Corresponding to the initial condition t, x_1, x_2, where $0 \le t \le T$, we define

$$\begin{aligned} g(t, x_1, x_2) &= \mathbb{E}^{t,x_1,x_2} h(X_1(T), X_2(T)) \\ f(t, x_1, x_2) &= \mathbb{E}^{t,x_1,x_2}\left[e^{-r(T-t)} h(X_1(T), X_2(T))\right] \end{aligned}$$

Exercise 8. Prove the two-dimensional Feynman-Kac theorems:

$$\begin{aligned} & g_t + \beta_1 g_{x_1} + \beta_2 g_{x_2} \\ & \quad + \tfrac{1}{2}\left(\gamma_{11}^2 + \gamma_{12}^2\right) g_{x_1x_1} + (\gamma_{11}\gamma_{21} + \gamma_{12}\gamma_{22})\, g_{x_1x_2} + \tfrac{1}{2}\left(\gamma_{21}^2 + \gamma_{22}^2\right) g_{x_2x_2} = 0 \\ & f_t + \beta_1 f_{x_1} + \beta_2 f_{x_2} \\ & \quad + \tfrac{1}{2}\left(\gamma_{11}^2 + \gamma_{12}^2\right) f_{x_1x_1} + (\gamma_{11}\gamma_{21} + \gamma_{12}\gamma_{22})\, f_{x_1x_2} + \tfrac{1}{2}\left(\gamma_{21}^2 + \gamma_{22}^2\right) f_{x_2x_2} = rf \end{aligned}$$

satisfy the terminal conditions $g(T, x_1, x_2) = f(T, x_1, x_2) = h(x_1, x_2)$.

Example 7. (Asian option). An Asian option delivers the payoff

$$V(T) = \left(\frac{1}{T}\int_0^T S(u)du - K\right)^+$$

where $S(u)$ is a geometric Brownian motion, the expiration time T is fixed and positive, and K is a positive strike price. In terms of the Brownian motion $\widetilde{W}(u)$ under the risk-neutral measure $\widetilde{\mathbb{P}}$, we may write the stochastic differential equation for $S(u)$ as

$$dS(u) = rS(u)du + \sigma S(u)d\widetilde{W}(u)$$

Because the payoff depends on the whole path of the stock price via its integral, at each time t prior to expiration it is not enough to know just the stock price in order to determine the value of the option. We must also know the integral of the stock price,

$$Y(t) = \int_0^t S(u)du$$

up to the current time t. Similarly, it is not enough to know just the integral $Y(t)$. We must also know the current stock price $S(t)$. Indeed, for the same value of $Y(t)$, the Asian option is worth more for high values of $S(t)$ than for low values because the high values of $S(t)$ make it more likely that the option will have a high payoff. For the process $Y(u)$, we have the stochastic differential equation

$$dY(u) = S(u)du$$

Because the pair of processes $(S(u), Y(u))$ is given by the pair of stochastic differential equations, the pair of processes $(S(u), Y(u))$ is a two-dimensional Markov process.

According to the risk-neutral pricing formula, the value of the Asian option at times prior to expiration is

$$V(t) = \widetilde{\mathbb{E}}\left[e^{-r(T-t)}\left(\frac{1}{T}Y(T) - K\right)^{+} \mid \mathscr{F}(t)\right], \quad 0 \le t \le T$$

Because the pair of processes $(S(u), Y(u))$ is Markov, this can be written as some function of the time variable t and the values at time t of these processes. In other words, there is a function $v(t,x,y)$ such that

$$v(t, S(t), Y(t)) = V(t) = \widetilde{\mathbb{E}}\left[e^{-r(T-t)}\left(\frac{1}{T}Y(T) - K\right)^{+} \mid \mathscr{F}(t)\right]$$

By the two-dimensional Feynman-Kac theorem,

$$v_t(t,x,y) + rxv_x(t,x,y) + xv_y(t,x,y) + \frac{1}{2}\sigma^2 x^2 v_{xx}(t,x,y) = rv(t,x,y)$$

with the terminal condition $v(T,x,y) = (\frac{1}{T}y - K)^+$.

Chapter 6: Exotic barrier option

Gechun Liang

The European calls and puts considered thus far in this text are sometimes called vanilla or even plain vanilla options. Their payoffs depend only on the final value of the underlying asset. Options whose payoffs depend on the path of the underlying asset are called path-dependent or exotic.

1 Maximum of Brownian motion with drift

We derive the joint density for a Brownian motion with drift and its maximum to date. This density is used to obtain explicit pricing formulas for a barrier option. To derive this formula, we begin with a Brownian motion $\widetilde{W}(t), 0 \leq t \leq T$, defined on a probability space $(\Omega, \mathscr{F}, \widetilde{\mathbb{P}})$. Under $\widetilde{\mathbb{P}}$, the Brownian motion $\widetilde{W}(t)$ has zero drift (i.e., it is a martingale). Let α be a given number, and define

$$\widehat{W}(t) = \alpha t + \widetilde{W}(t), \quad 0 \leq t \leq T$$

This Brownian motion $\widehat{W}(t)$ has drift α under $\widetilde{\mathbb{P}}$. We further define

$$\widehat{M}(T) = \max_{0 \leq t \leq T} \widehat{W}(t)$$

Because $\widehat{W}(0) = 0$, we have $\widehat{M}(T) \geq 0$. We also have $\widehat{W}(T) \leq \widehat{M}(T)$ Therefore, the pair of random variables $(\widehat{M}(T), \widehat{W}(T))$ takes values in the set $\{(m,w); w \leq m, m \geq 0\}$.

Theorem 1. *The joint density under* $\widetilde{\mathbb{P}}$ *of the pair* $(\widehat{M}(T), \widehat{W}(T))$ *is*

Gechun Liang
Department of Statistics, University of Warwick, U.K. e-mail: g.liang@warwick.ac.uk

$$\tilde{f}_{\widehat{M}(T),\widehat{W}(T)}(m,w) = \frac{2(2m-w)}{T\sqrt{2\pi T}} e^{\alpha w - \frac{1}{2}\alpha^2 T - \frac{1}{2T}(2m-w)^2}, \quad w \le m, m \ge 0 \tag{1}$$

and is zero for other values of m and w.

Proof. We define the exponential martingale

$$\widehat{Z}(t) = e^{-\alpha \widetilde{W}(t) - \frac{1}{2}\alpha^2 t} = e^{-\alpha \widehat{W}(t) + \frac{1}{2}\alpha^2 t}, \quad 0 \le t \le T$$

and use $\widehat{Z}(T)$ to define a new probability measure $\widehat{\mathbb{P}}$ by

$$\widehat{\mathbb{P}}(A) = \int_A Z(T) d\widetilde{\mathbb{P}} \text{ for all } A \in \mathscr{F}$$

According to Girsanov's Theorem, $\widehat{W}(t)$ is a Brownian motion (with zero drift) under $\widehat{\mathbb{P}}$. Exercise 5 in Chapter 2 gives us the joint density of $(\widehat{M}(T), \widehat{W}(T))$ under $\widehat{\mathbb{P}}$, which is

$$\widehat{f}_{\widehat{M}(T),\widehat{W}(T)}(m,w) = \frac{2(2m-w)}{T\sqrt{2\pi T}} e^{-\frac{1}{2T}(2m-w)^2}, \quad w \le m, m \ge 0$$

and is zero for other values of m and w. To work out the density of $(\widehat{M}(T), \widehat{W}(T))$ under $\widetilde{\mathbb{P}}$, we proceed as follows

$$\begin{aligned}
&\widetilde{\mathbb{P}}\{\widehat{M}(T) \le m, \widehat{W}(T) \le w\} \\
=&\widetilde{\mathbb{E}}\left[\mathbb{I}_{\{\widehat{M}(T) \le m, \widehat{W}(T) \le w\}}\right] \\
=&\widehat{\mathbb{E}}\left[\frac{1}{\widehat{Z}(T)} \mathbb{I}_{\{\widehat{M}(T) \le m, \widehat{W}(T) \le w\}}\right] \\
=&\widehat{\mathbb{E}}\left[e^{\alpha \widehat{W}(T) - \frac{1}{2}\alpha^2 T} \mathbb{I}_{\{\widehat{M}(T) \le m, \widehat{W}(T) \le w\}}\right] \\
=&\int_{-\infty}^{w} \int_{-\infty}^{m} e^{\alpha y - \frac{1}{2}\alpha^2 T} \widehat{f}_{\widehat{M}(T),\widehat{W}(T)}(x,y) dx dy
\end{aligned}$$

Therefore, the density of $(\widehat{M}(T), \widehat{W}(T))$ under $\widetilde{\mathbb{P}}$ is

$$\frac{\partial^2}{\partial m \partial w} \widetilde{\mathbb{P}}\{\widehat{M}(T) \le m, \widehat{W}(T) \le w\} = e^{\alpha w - \frac{1}{2}\alpha^2 T} \widehat{f}_{\widehat{M}(T),\widehat{W}(T)}(m,w)$$

When $w \le m$ and $m \ge 0$, this is formula (1). For other values of m and w, we obtain zero because $\widehat{f}_{\widehat{M}(T),\widehat{W}(T)}(m,w)$ is zero. □

Corollary 1. *The CDF under* $\widetilde{\mathbb{P}}$ *of the random variable* $\widehat{M}(T)$ *is*

$$\widetilde{\mathbb{P}}\{\widehat{M}(T) \le m\} = N\left(\frac{m - \alpha T}{\sqrt{T}}\right) - e^{2\alpha m} N\left(\frac{-m - \alpha T}{\sqrt{T}}\right), \quad m \ge 0$$

and the density under $\widetilde{\mathbb{P}}$ *of the random variable* $\widehat{M}(T)$ *is*

$$\tilde{f}_{\widehat{M}(T)}(m) = \frac{2}{\sqrt{2\pi T}} e^{-\frac{1}{2T}(m-\alpha T)^2} - 2\alpha e^{2\alpha m} N\left(\frac{-m-\alpha T}{\sqrt{T}}\right), m \geq 0$$

and is zero for $m < 0$.

Proof. We integrate the density (1) to compute

$$\begin{aligned}
&\widetilde{\mathbb{P}}\{\widehat{M}(T) \leq m\} \\
=& \int_0^m \int_w^m \frac{2(2\mu - w)}{T\sqrt{2\pi T}} e^{\alpha w - \frac{1}{2}\alpha^2 T - \frac{1}{2T}(2\mu - w)^2} d\mu dw \\
&+ \int_{-\infty}^0 \int_0^m \frac{2(2\mu - w)}{T\sqrt{2\pi T}} e^{\alpha w - \frac{1}{2}\alpha^2 T - \frac{1}{2T}(2\mu - w)^2} d\mu dw \\
=& -\frac{e^{2\alpha m}}{\sqrt{2\pi T}} \int_{-\infty}^m e^{-\frac{1}{2T}(w-2m-\alpha T)^2} dw + \frac{1}{\sqrt{2\pi T}} \int_{-\infty}^m e^{-\frac{1}{2T}(w-\alpha T)^2} dw
\end{aligned}$$

We make the change of variable $y = \frac{w-2m-\alpha T}{\sqrt{T}}$ in the first integral and $y = \frac{w-\alpha T}{\sqrt{T}}$ in the second, thereby obtaining

$$\begin{aligned}
\widetilde{\mathbb{P}}\{\widehat{M}(T) \leq m\} &= -\frac{e^{2\alpha m}}{\sqrt{2\pi}} \int_{-\infty}^{\frac{-m-\alpha T}{\sqrt{T}}} e^{-\frac{1}{2}y^2} dy + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{m-\alpha T}{\sqrt{T}}} e^{-\frac{1}{2}y^2} du \\
&= -e^{2\alpha m} N\left(\frac{-m-\alpha T}{\sqrt{T}}\right) + N\left(\frac{m-\alpha T}{\sqrt{T}}\right)
\end{aligned}$$

This establishes the CDF. To obtain the density we differentiate the CDF with respect to m,

$$\begin{aligned}
&\frac{d}{dm}\widetilde{\mathbb{P}}\{\widehat{M}(T) \leq m\} \\
=& N'\left(\frac{m-\alpha T}{\sqrt{T}}\right)\left(\frac{1}{\sqrt{T}}\right) - 2\alpha e^{2\alpha m} N\left(\frac{-m-\alpha T}{\sqrt{T}}\right) \\
&- e^{2\alpha m} N'\left(\frac{-m-\alpha T}{\sqrt{T}}\right)\left(-\frac{1}{\sqrt{T}}\right) \\
=& \frac{1}{\sqrt{2\pi T}} e^{-\frac{1}{2T}(m-\alpha T)^2} - 2\alpha e^{2\alpha m} N\left(\frac{-m-\alpha T}{\sqrt{T}}\right) + \frac{e^{2\alpha m}}{\sqrt{2\pi T}} e^{-\frac{1}{2T}(-m-\alpha T)^2}
\end{aligned}$$

Combining the first and third terms, we obtain desired density. □

2 Knock-out barrier options

There are several types of barrier options. Some ”knock out” when the underlying asset price crosses a barrier (i.e., they become worthless). If the underlying asset price begins below the barrier and must cross above it to cause the knock-out, the option is said to be up-and-out. A down-and-out option has the barrier below the initial asset price and knocks out if the asset price falls below the barrier. Other options ”knock in” at a barrier (i.e., they pay off zero unless they cross a barrier). Knock-in options also fall into two classes, up-and-in and down-and-in. The payoff at expiration for barrier options is typically either that of a put or a call. More complex barrier options require the asset price to not only cross a barrier but spend a certain amount of time across the barrier in order to knock in or knock out.

In this section, *we treat an up-and-out call on a geometric Brownian motion*. The methodology we develop works equally well for up-and-in, downand-out, and down-and-in puts and calls.

2.1 Up-and-out call

Our underlying risky asset is geometric Brownian motion

$$dS(t) = rS(t)dt + \sigma S(t)d\widetilde{W}(t)$$

where $\widetilde{W}(t), 0 \leq t \leq T$, is a Brownian motion under the risk-neutral measure $\mathbb{P}$. Consider a European call, expiring at time T, with strike price K and up-and-out barrier B. We assume $K < B$; otherwise, the option must knock out in order to be in the money and hence could only pay off zero. The solution to the stochastic differential equation for the asset price is

$$S(t) = S(0)e^{\sigma\widetilde{W}(t)+\left(r-\frac{1}{2}\sigma^2\right)t} = S(0)e^{\sigma\widehat{W}(t)}$$

where $\widehat{W}(t) = \alpha t + \widetilde{W}(t)$, and

$$\alpha = \frac{1}{\sigma}\left(r - \frac{1}{2}\sigma^2\right)$$

We define $\widehat{M}(T) = \max_{0\leq t\leq T}\widehat{W}(t)$, so

$$\max_{0\leq t\leq T} S(t) = S(0)e^{\sigma\widehat{M}(T)}$$

The option knocks out if and only if $S(0)e^{\sigma\widehat{M}(T)} > B$; if $S(0)e^{\sigma\widehat{M}(T)} \leq B$, the option pays off

$$(S(T)-K)^+ = \left(S(0)e^{\sigma\widehat{W}(T)} - K\right)^+$$

In other words, the payoff of the option is

$$\begin{aligned} V(T) &= \left(S(0)e^{\sigma\widehat{W}(T)} - K\right)^+ \mathbb{I}_{\left\{S(0)e^{\sigma\widehat{M}(T)}\leq B\right\}} \\ &= \left(S(0)e^{\sigma\widehat{W}(T)} - K\right) \mathbb{I}_{\left\{S(0)e^{\sigma}\widehat{W}(T)\geq K, S(0)e^{\sigma\widehat{M}(T)}\leq B\right\}} \\ &= \left(S(0)e^{\sigma\widehat{W}(T)} - K\right) \mathbb{I}_{\{\widehat{W}(T)\geq k, \widehat{M}(T)\leq b\}} \end{aligned}$$

where

$$k = \frac{1}{\sigma}\log\frac{K}{S(0)}, \quad b = \frac{1}{\sigma}\log\frac{B}{S(0)}$$

2.2 Pricing PDE

Theorem 2. *Let $v(t,x)$ denote the price at time t of the up-and-out call under the assumption that the call has not knocked out prior to time t and $S(t) = x$. Then $v(t,x)$ satisfies the Black-Scholes partial differential equation*

$$v_t(t,x) + rxv_x(t,x) + \frac{1}{2}\sigma^2 x^2 v_{xx}(t,x) = rv(t,x)$$

in the rectangle $\{(t,x); 0 \leq t < T, 0 \leq x \leq B\}$ and satisfies the boundary conditions

$$\begin{aligned} v(t,B) &= 0,\ 0 \leq t < T \\ v(T,x) &= (x-K)^+,\ 0 \leq x \leq B \end{aligned}$$

We define ρ to be the first time t at which the asset price reaches the barrier B. In other words, ρ is chosen in a path-dependent way so that $S(t) < B$ for $0 \leq t \leq \rho$ and $S(\rho) = B$. since the asset price almost surely exceeds the barrier immediately after reaching it, we may regard ρ as the time of knock-out. If the asset price does not reach the barrier before expiration, we set $\rho = \infty$. If the asset price first reaches the barrier at time T, then $\rho = T$ but knock-out does not occur because there is no time left for the asset price to exceed the barrier. However, the probability that the asset price first reaches the barrier at time T is zero, so this anomaly does not matter.

The random variable ρ is a stopping time because it chooses its value based on the path of the asset price up to time ρ. The *Optional Sampling Theorem* asserts that a martingale stopped at a stopping time is still a martingale. In particular, the process

$$e^{-r(t\wedge\rho)}V(t\wedge\rho) = \begin{cases} e^{-rt}V(t) & \text{if } 0 \leq t \leq \rho \\ e^{-r\rho}V(\rho) & \text{if } \rho < t \leq T \end{cases}$$

is a $\widetilde{\mathbb{P}}$-martingale.

By the Markov property, we have

$$V(t) = v(t, S(t)), \quad 0 \leq t \leq \rho$$

In particular, $e^{-rt}v(t,S(t))$ is a $\widetilde{\mathbb{P}}$-martingale up to time ρ, or, put another way, the stopped process

$$e^{-r(t\wedge\rho)}v(t\wedge\rho, S(t\wedge\rho)), \quad 0\le t\le T$$

is a martingale under $\widetilde{\mathbb{P}}$.

Proof of the theorem. We compute the differential

$$\begin{aligned}
d\left(e^{-rt}v(t,S(t))\right) &= e^{-rt}\Big[-rv(t,S(t))dt + v_t(t,S(t))dt + v_x(t,S(t))dS(t) \\
&\quad + \frac{1}{2}v_{xx}(t,S(t))dS(t)dS(t)\Big] \\
&= e^{-rt}\Big[-rv(t,S(t)) + v_t(t,S(t)) + rS(t)v_x(t,S(t)) \\
&\quad + \frac{1}{2}\sigma^2 S^2(t)v_{xx}(t,S(t))\Big]dt \\
&\quad + e^{-rt}\sigma S(t)v_x(t,S(t))d\widetilde{W}(t)
\end{aligned}$$

The dt term must be zero for $0\le t\le\rho$, (i.e., before the option knocks out). But since $(t,S(t))$ can reach any point in $\{(t,x); 0\le t<T, 0\le x\le B\}$ before the option knocks out, the Black-Scholes PDE must hold for every $t\in[0,T)$ and $x\in[0,B]$. $\square$

2.3 Computation of up-and-out call

The risk-neutral price at time zero of the up-and-out call is

$$V(0) = \widetilde{\mathbb{E}}\left[e^{-rT}V(T)\right].$$

We use the density formula in Theorem 1 to compute this. If $k\ge 0$, we must integrate over the region $\{(m,w); k\le w\le m\le b\}$. On the other hand, if $k<0$, we integrate over the region $\{(m,w); k\le w\le m, 0\le m\le b\}$. In both cases, the region can be described as $\{(m,w); k\le w\le b, w^+\le m\le b\}$.

We assume here that $S(0)\le B$ so that $b>0$. Otherwise, the region over which we integrate has zero area, and the time-zero value of the call is zero rather than the integral computed below. We also assume $S(0)>0$ so that b and k are finite. When $0<S(0)\le B$, the time-zero value of the up-and-out call is

$$
\begin{aligned}
V(0) =& \int_k^b \int_{w^+}^b e^{-rT}\left(S(0)e^{\sigma w} - K\right) \frac{2(2m-w)}{T\sqrt{2\pi T}} e^{\alpha w - \frac{1}{2}\alpha^2 T - \frac{1}{2T}(2m-w)^2} dm dw \\
=& - \int_k^b e^{-rT}\left(S(0)e^{\sigma w} - K\right) \frac{1}{\sqrt{2\pi T}} e^{\alpha w - \frac{1}{2}\alpha^2 T - \frac{1}{2T}(2m-w)^2} \Big|_{m=w^+}^{m=b} dw \\
=& \frac{1}{\sqrt{2\pi T}} \int_k^b \left(S(0)e^{\sigma w} - K\right) e^{-rT + \alpha w - \frac{1}{2}\alpha^2 T - \frac{1}{2T}w^2} dw \\
& - \frac{1}{\sqrt{2\pi T}} \int_k^b \left(S(0)e^{\sigma w} - K\right) e^{-rT + \alpha w - \frac{1}{2}\alpha^2 T - \frac{1}{2T}(2b-w)^2} dw \\
=& S(0)I_1 - KI_2 - S(0)I_3 + KI_4
\end{aligned}
$$

where

$$
\begin{aligned}
I_1 &= \frac{1}{\sqrt{2\pi T}} \int_k^b e^{\sigma w - rT + \alpha w - \frac{1}{2}\alpha^2 T - \frac{1}{2T}w^2} dw \\
I_2 &= \frac{1}{\sqrt{2\pi T}} \int_k^b e^{-rT + \alpha w - \frac{1}{2}\alpha^2 T - \frac{1}{2T}w^2 dw} \\
I_3 &= \frac{1}{\sqrt{2\pi T}} \int_k^b e^{\sigma w - rT + \alpha w - \frac{1}{2}\alpha^2 T - \frac{1}{2T}(2b-w)^2} dw \\
&= \frac{1}{\sqrt{2\pi T}} \int_k^b e^{\sigma w - rT + \alpha w - \frac{1}{2}\alpha^2 T - \frac{2}{T}b^2 + \frac{2}{T}bw - \frac{1}{2T}w^2 dw} \\
I_4 &= \frac{1}{\sqrt{2\pi T}} \int_k^b e^{-rT + \alpha w - \frac{1}{2}\alpha^2 T - \frac{1}{2T}(2b-w)^2} dw \\
&= \frac{1}{\sqrt{2\pi T}} \int_k^b e^{-rT + \alpha w - \frac{1}{2}\alpha^2 T - \frac{2}{T}b^2 + \frac{2}{T}bw - \frac{1}{2T}w^2} dw
\end{aligned}
$$

Each of these integrals is of the form

$$
\begin{aligned}
\frac{1}{\sqrt{2\pi T}} \int_k^b e^{\beta + \gamma w - \frac{1}{2T}w^2} dw &= \frac{1}{\sqrt{2\pi T}} \int_k^b e^{-\frac{1}{2T}(w-\gamma T)^2 + \frac{1}{2}\gamma^2 T + \beta} dw \\
&= e^{\frac{1}{2}\gamma^2 T + \beta} \frac{1}{\sqrt{2\pi}} \int_{\frac{1}{\sqrt{T}}(k-\gamma T)}^{\frac{1}{\sqrt{T}(b-\gamma T)}} e^{-\frac{1}{2}y^2} dy,
\end{aligned}
$$

where we have made the change of variable $y = \frac{w-\gamma T}{\sqrt{T}}$. Using the standard cumulative normal distribution property $N(z) = 1 - N(-z)$, we continue, writing

$$
\begin{aligned}
&\frac{1}{\sqrt{2\pi T}}\int_k^b e^{\beta+\gamma w-\frac{w^2}{2T}}\,dw \\
&= e^{\frac{1}{2}\gamma^2 T+\beta}\left[N\left(\frac{b-\gamma T}{\sqrt{T}}\right)-N\left(\frac{k-\gamma T}{\sqrt{T}}\right)\right] \\
&= e^{\frac{1}{2}\gamma^2 T+\beta}\left[N\left(\frac{-k+\gamma T}{\sqrt{T}}\right)-N\left(\frac{-b+\gamma T}{\sqrt{T}}\right)\right] \\
&= e^{\frac{1}{2}\gamma^2 T+\beta}\left[N\left(\frac{1}{\sigma\sqrt{T}}\left[\log\frac{S(0)}{K}+\gamma\sigma T\right]\right)-N\left(\frac{1}{\sigma\sqrt{T}}\left[\log\frac{S(0)}{B}+\gamma\sigma T\right]\right)\right]
\end{aligned} \tag{2}
$$

Set

$$\delta_{\pm}(\tau,s)=\frac{1}{\sigma\sqrt{\tau}}\left[\log s+\left(r\pm\frac{1}{2}\sigma^2\right)\tau\right]$$

The integral I_1 is of the form (2) with $\beta=-rT-\frac{1}{2}\alpha^2 T$ and $\gamma=\alpha+\sigma$ so $\frac{1}{2}\gamma^2 T+\beta=0$ and $\gamma\sigma=r+\frac{1}{2}\sigma^2$. Therefore

$$I_1=N\left(\delta_+\left(T,\frac{S(0)}{K}\right)\right)-N\left(\delta_+\left(T,\frac{S(0)}{B}\right)\right)$$

The integral I_2 is of the form (2) with $\beta=-rT-\frac{1}{2}\alpha^2 T$ and $\gamma=\alpha$, so $\frac{1}{2}\gamma^2 T+\beta=-rT$ and $\gamma\sigma=r-\frac{1}{2}\sigma^2$. Therefore

$$I_2=e^{-rT}\left[N\left(\delta_-\left(T,\frac{S(0)}{K}\right)\right)-N\left(d_-\left(T,\frac{S(0)}{B}\right)\right)\right]$$

For I_3, we have $\beta=-rT-\frac{1}{2}\alpha^2 T-\frac{2b^2}{T}$ and $\gamma=\alpha+\sigma+\frac{2b}{T}$, so

$$
\begin{aligned}
\frac{1}{2}\gamma^2 T+\beta &= \log\left(\frac{S(0)}{B}\right)^{-\frac{2r}{\sigma^2}-1} \\
\gamma\sigma T &= \left(r+\frac{1}{2}\sigma^2\right)T+\log\left(\frac{B}{S(0)}\right)^2
\end{aligned}
$$

Therefore,

$$I_3=\left(\frac{S(0)}{B}\right)^{-\frac{2r}{\sigma^2}-1}\left[N\left(\delta_+\left(T,\frac{B^2}{KS(0)}\right)\right)-N\left(\delta_+\left(T,\frac{B}{S(0)}\right)\right)\right]$$

Finally, for I_4, we have $\beta=-rT-\frac{1}{2}\alpha^2 T-\frac{2b^2}{T}$ and $\gamma=\alpha+\frac{2b}{T}$, so

$$\frac{1}{2}\gamma^2 T + \beta = -rT + \log\left(\frac{S(0)}{B}\right)^{-\frac{2r}{\sigma^2}+1}$$

$$\gamma\sigma T = \left(r - \frac{1}{2}\sigma^2\right)T + \log\left(\frac{B}{S(0)}\right)^2$$

$$I_4 = e^{-rT}\left(\frac{S(0)}{B}\right)^{-\frac{2r}{\sigma^2}+1}\left[N\left(\delta_-\left(T,\frac{B^2}{KS(0)}\right)\right) - N\left(\delta_-\left(T,\frac{B}{S(0)}\right)\right)\right]$$

Putting all this together, under the assumption $0 < S(0) \leq B$, we have the up-and-out call price formula

$$\begin{aligned} V(0) = &S(0)\left[N\left(\delta_+\left(T,\frac{S(0)}{K}\right)\right) - N\left(\delta_+\left(T,\frac{S(0)}{B}\right)\right)\right] \\ &- e^{-rT}K\left[N\left(\delta_-\left(T,\frac{S(0)}{K}\right)\right) - N\left(\delta_-\left(T,\frac{S(0)}{B}\right)\right)\right] \\ &- B\left(\frac{S(0)}{B}\right)^{-\frac{2r}{\sigma^2}}\left[N\left(\delta_+\left(T,\frac{B^2}{KS(0)}\right)\right) - N\left(\delta_+\left(T,\frac{B}{S(0)}\right)\right)\right] \\ &+ e^{-rT}K\left(\frac{S(0)}{B}\right)^{-\frac{2r}{\sigma^2}+1}\left[N\left(\delta_-\left(T,\frac{B^2}{KS(0)}\right)\right) - N\left(\delta_-\left(T,\frac{B}{S(0)}\right)\right)\right] \end{aligned}$$