Course Syllabus for "Stochastic Analysis and Applications"

Chenxu Li
Guanghua School of Management
Peking University
cxli@caa.columbia.edu

§1 Course objectives

A career in quantitative analysis and trading at financial institutions or academic research concentrating on dynamic and empirical asset pricing is sure to involve decision making based on stochastic models. This course focuses on a quick tutorial on stochastic analysis and then discuss their applications in various fields, especially finance and economic. For students who plan to pursue their career in academia, this course will provide solid fundamental tools for studying derivatives pricing, portfolio planning, the two major topics in financial economics. For students who plan to develop practical quantitative skills in financial industry, this course will serve as an indispensable preparation for both buy-side and sell-side quantitative positions, for example, quant modelers, developers, strategists, and technical traders.

§2 Tentative topics

The course will cover the following key aspects:

- Tentative topics on stochastic analysis may include, but not limited to, conditional expectation-s, stochastic processes, martingales, Markov processes, random walk, Brownian motion, Poisson process, stochastic integrations, stochastic calculus, the Itô formulas, some fundamental theorems in stochastic analysis, for example, time-change, martingale representation, Girsanov change-of-measure, stochastic differential equations, interpretation of solutions to partial differential equa-tions via the Feynman-Kac theorem, etc.
- Based on these mathematical tools, we will further discuss their implementation via Monte Carlo simulation. Detailed topics may include, but not limited to, random number generation, exact

simulation of important distributions and sample paths of stochastic models, variance reduction techniques, discretization methods for simulating solutions to stochastic differential equations. We will discuss the implementation of related algorithms in various softwares, for example, C/C++, Matlab, and Excel.

- With the proliferation of fast and inexpensive information technology, it is important to be able to connect data and models intelligently. We will discuss econometric and statistical inference of continuous-time models based on discretely monitored data. For example, via maximum-likelihood estimation methods, we will provide a convenient tool for academic and industrial empirical studies based on real market trading data.
- For applications, we will target derivatives pricing and portfolio planning of a wide variety of asset classes, such as equity, equity index, fixed-income, credit, commodity, and foreign-exchange. If time allowed, we will discuss the development of the Chinese derivatives market, for example, modeling issues of option pricing on stocks index futures.

§3 Course materials

The course is suitable for both graduate and undergraduate students who are interested in the topics and have ever taken probability and statistics as prerequisites. There is no required text for the course. All material will be covered through lecture notes and selected chapters from the following reference books: Shreve [5], Karatzas and Shreve [4], Duffie [2], Glasserman [3], Campbell et al. [1], and Singleton [6], etc. We will suggest excellent and helpful exercises to enhance understand. In particular, many of them may serve as quantitative job interview questions.

§4 Grading policy

We will have two homework problem sets (taking 50% in grade) and an open-book final exam (taking 50% in grade).

References

[1] Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press.

[2] Duffie, D. (2001). *Dynamic Asset Pricing Theory*. 3rd ed. Princeton University Press.

[3] Glasserman, P. (2004). *Monte Carlo methods in Financial Engineering*, vol. 53 of *Applications of Mathematics (New York)*. Springer-Verlag, New York. Stochastic Modelling and Applied Probability.

[4] Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*, vol. 113 of *Graduate Texts in Mathematics*. 2nd ed. Springer-Verlag, New York.

[5] Shreve, S. E. (2004). *Stochastic Calculus for Finance. II*. Springer Finance, Springer-Verlag, New York.

[6] Singleton, K. J. (2006). *Empirical Dynamic Asset Pricing: Model Specification and Econometric Assessment*. Princeton University Press.

Stochastic Analysis and Applications

An Introduction to Stochastic Analysis

Spring 2017

Agenda

- About the course
- An introduction to stochastic analysis

Course Information

Course information:

- Course title: Stochastic Analysis and Applications
- Instructor: LI, Chenxu (cxli@gsm.pku.edu.cn)
- Office hour: immediately after class or by appointment

Teaching assistants

- Ye Yongxin: 1401211868@pku.edu.cn
- Wang Yaqiong: 1501211708@pku.edu.cn
- Li Chenxu (my PhD student): lichenxu.pku@gmail.com

Course Objectives

- A career in quantitative analysis and trading at financial institutions or academic research concentrating on dynamic and empirical asset pricing is sure to involve decision making based on stochastic models.
- This course focuses on a quick tutorial on stochastic analysis and then discuss their applications in various fields, especially finance and economic.
- For students who plan to pursue their career in academia, this course will provide solid fundamental tools for studying derivatives pricing, portfolio planning, the two major topics in financial economics.
- For students who plan to develop practical quantitative skills in financial industry, this course will serve as an indispensable preparation for both buy-side and sell-side quantitative positions, for example, quant modelers, developers, strategists, and technical traders.

About Financial Engineering/Econometrics

- Financial engineering: use tools from applied probability and stochastic modeling to model real world and business analytic problems.
- Financial econometrics: take the models to data. Use statistics and econometrics tools to estimate the model and put it to applications at an empirical level.
- FE and FE use mathematics/computer sciences/statistics/econometrics, etc.
- But none of these need to be too deep, since we just care about applications

Career in Financial Engineering/Econometrics

- Financial Engineering/Econometrics: a charming academic research field
- Very useful in financial industry, e.g., for investment banks and hedge funds
- In particular, it is booming: schools are lacking of this kind of faculty members; financial institutions need quants modelers/stratgists/traders
- So, do one thing and get two options!

Pave the Way to Research in Financial Engineering/Econometrics

- Model real-world business analytic problem arising from financial market.
- Handle the model using your mathematical skills, e.g., stochastic analysis.
- Implement the model using various numerical methods, e.g., Monte Carlo somulation.
- Take the model to data, e.g., perform econometric analysis of the model to determine coefficients and judge the performance of modeling.

Tentative Topics

- Tentative topics on stochastic analysis: conditional expectations, stochastic processes, martingales, Markov processes, random walk, Brownian motion, Poisson process, stochastic integrations, stochastic calculus, the Itô formulas, some fundamental theorems in stochastic analysis, for example, time-change, martingale representation, Girsanov change-of-measure, stochastic differential equations, interpretation of solutions to partial differential equations via the Feynman-Kac theorem, etc.
- Implementation via Monte Carlo simulation: random number generation, exact simulation of important distributions and sample paths of stochastic models, variance reduction techniques, discretization methods for simulating solutions to stochastic differential equations. We will discuss the implementation of related algorithms in various softwares, for example, C/C++, Matlab, and Excel.

Tentative topics

- With the proliferation of fast and inexpensive information technology, it is important to be able to connect data and models intelligently. We will discuss econometric and statistical inference of continuous-time models based on discretely monitored data. For example, via maximum-likelihood estimation methods, we will provide a convenient tool for academic and industrial empirical studies based on real market trading data.
- For applications, we will target derivatives pricing and portfolio planning of a wide variety of asset classes, such as equity, equity index, fixed-income, credit, commodity, and foreign-exchange. If time allowed, we will discuss the development of the Chinese derivatives market, for example, modeling issues of option pricing on stocks index futures.

Administrative Matters

- During my lectures, I may point out some excellent exercises. We may include them as homework problems. In addition, I will suggest some questions from the references books. Grading will be based on your completion of these exercises.

- The course is suitable for both graduate and undergraduate students who are interested in the topics and have ever taken probability and statistics as prerequisites. There is no required text for the course. All material will be covered through lecture notes and selected chapters from the following reference books. We will suggest excellent and helpful exercises to enhancing understand. In particular, many of them may serve as quantitative job interview questions.

Reference Books

[1] Campbell, J. Y., Lo, A. W. and Mackinlay, A. C. (1997). The Econometrics of Financial Markets. Princeton University Press.
[2] Duffie, D. (2001). Dynamic Asset Pricing Theory. 3rd ed. Princeton University Press.
[3] Glasserman, P. (2004). Monte Carlo methods in Financial Engineering, vol. 53 of Applications of Mathematics (New York). Springer-Verlag, New York. Stochastic Modelling and Applied Probability.
[4] Karatzas, I. and Shreve, S. E. (1991). Brownian Motion and Stochastic Calculus, vol. 113 of Graduate Texts in Mathematics. 2nd ed. Springer-Verlag, New York.
[5] Shreve, S. E. (2004). Stochastic Calculus for Finance. II. Springer Finance, Springer-Verlag, New York.
[6] Singleton, K. J. (2006). Empirical Dynamic Asset Pricing: Model Specification and Econometric Assessment. Princeton University Press.

Foundation of Stochastic Analysis

- conditional expectation
- stochastic process
- martingale
- Brownian motion
- stochastic calculus
- stochastic differential equations
- interpretation of solutions to partial differential equations via the Feynman-Kac theorem
- fundamental theorems in stochastic analysis

Overview

- Conditional expectation

Conditioning: Motivation and Traditional Version

- Motivation: in finance, we usually need to consider conditional behavior (distribution) of asset price given that at an earlier time, e.g. how will the IBM stock price distribute on Oct. 28th assuming that its price at Oct. 1st **were** known.
- Traditional version: $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.
- Consider $A = \{X \leq x\}$ where X is a continuous R.V.
- Conditional CDF: $F(x|B) = \mathbb{P}(X \leq x|B)$
- Conditional PDF: $f(x|B)$ such that $F(x|B) = \int_{-\infty}^{x} f(u|B)du$
- Conditional Expectation: $\mathbb{E}(X|B) = \int_{-\infty}^{+\infty} uf(u|B)du$
- Similarly, we can consider whatever we had for unconditional cases, e.g. conditional variance, etc.
- If X is a discrete R.V., we just change the $\int$ to $\sum$ properly

Conditioning Induced by a Random Variable

- Let $B = \{Y = y\}$ where Y is a R.V.
- If Y is a Discrete R.V. and $\mathbb{P}(B) \neq 0$, just follow previous slides
- If Y is a Continuous R.V., $\mathbb{P}(B) = 0$. Problematic with the division!
- Overcome this by considering $B_\epsilon = \{Y \in (y - \epsilon, y + \epsilon)\}$ and then letting $\epsilon \to 0$
- Conditional CDF:

$$\mathbb{P}(X \leq x | Y = y) = F(x | Y = y) = \int_{-\infty}^{x} f(u | Y = y) du$$

- Conditional PDF:

$$f(u | Y = y) = \frac{f(x, y)}{f_Y(y)}$$

Conditional Expectation: from Classical to Modern Version

Classical definition: a number!

- If X is a continuous R.V., $\mathbb{E}(X|Y = y) = \int_{-\infty}^{+\infty} u f(u|Y = y) du$
- If X is a discrete R.V., $\mathbb{E}(X|Y = y) = \sum_x x\mathbb{P}(x|Y = y)$

Example: Two independent dice are rolled and Y denotes the value of the first roll and X denotes the sum of the two rolls, then we calculate that

$$\mathbb{E}(X|Y = y) = y + \frac{7}{2}.$$

Now, can we consider **random variable** relying on the randomness of the condition in Y (substituting y by Y):

$$\mathbb{E}(X|Y) = Y + \frac{7}{2}.$$

By doing so, we average something random based on something random. (Think about a forward started contract!)

Conditional Expectation: A Further Development

Definition: Given two random variables X and Y, we define the conditional expectation $\mathbb{E}(X|Y)$ as a random variable obtained from

$$\mathbb{E}(X|Y) = g(Y),$$

where $g(y) := \mathbb{E}(X|Y = y)$.

Conditioning on more information (cautious! more abstract):

Let $\sigma(Y)$ be "all the events generated by Y". Formally, $\sigma(Y)$ is the "smallest σ-algebra containing all events like $\{\omega \in \Omega : Y(\omega) \leq y\}$", can we define $\mathbb{E}(X|\sigma(Y))$?

Can we further define a version of conditional expectation in a more informatively way? say, $\mathbb{E}(X|\mathcal{G})$ where $\mathcal{G}$ is the $\sigma-$algebra (just information)?

Conditional Expectation: Formal Mathematical Version

- **Definition**: Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a sub-$\sigma$-algebra $\mathcal{G} \subseteq \mathcal{F}$, and a RV X s.t. either $X \geq 0$ or $|E(X)| < +\infty$. Then the conditional expectation of X given $\mathcal{G}$, denoted by $E(X|\mathcal{G})$, is any RV that satisfies
  - (i) (**Measurability**) $E(X|\mathcal{G})$ is $\mathcal{G}$-measurable, and
  - (ii) (**Partial Averaging**) $\mathbb{E}[1_A E(X|\mathcal{G})] = \mathbb{E}(1_A X)$ for all $A \in \mathcal{G}$. Here, 1_A is a random variable, called indicator, defined as: $1_A(\omega) = 1$, if $\omega \in A$; $1_A(\omega) = 0$, otherwise.
- The mathematical issue of this definition is too abstract to us. We just take it for granted!
- In particular, when $\mathcal{G} = \sigma(Y)$, we can prove that

  $$\mathbb{E}(X|\sigma(Y)) = \mathbb{E}(X|Y).$$

  i.e. the formal mathematical version and the classical version agree!
- $\mathbb{E}(X|Y)$ can be calculated from conditional distributions, i.e.

  $$E(X|Y = y) = g(y) \Longrightarrow E(X|Y) = E(X|Y = y)|_{y=Y} = g(Y)$$

Some Properties

- **(Linearity)** $\mathbb{E}(c_1X + c_2Y|\mathcal{G}) = c_1\mathbb{E}(X|\mathcal{G}) + c_2\mathbb{E}(Y|\mathcal{G})$.
- **(Expectation and conditioning)** $\mathbb{E}[\mathbb{E}(X|\mathcal{G})] = \mathbb{E}X$
- **(Taking out what is known) If $\sigma(X) \subset \mathcal{G}$ (X is $\mathcal{G}$-measurable), then $\mathbb{E}(XY|\mathcal{G}) = X\mathbb{E}(Y|\mathcal{G})$.**
- **(Iterated conditioning) If $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$, then $\mathbb{E}\left[\mathbb{E}(X|\mathcal{G})|\mathcal{H}\right] = \mathbb{E}(X|\mathcal{H})$.**
- **(Independence) If X is independent of $\mathcal{G}$, then $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$.**
- **(Independence) If X is $\mathcal{G}$ measurable (i.e. $\sigma(X) \subset \mathcal{G}$) and Y is independent of $\mathcal{G}$, then $\mathbb{E}[f(X,Y)|\mathcal{G}] = g(X)$, where $g(x) = \mathbb{E}f(x,Y)$.**
- **(Conditional Jensen's Inequality) If $\phi(x)$ is convex, then $\mathbb{E}(\phi(X)|\mathcal{G}) \geq \phi(\mathbb{E}(X|\mathcal{G}))$.**

Understanding the Properties by Simply Taking $\mathcal{G} = \sigma(Y)$

Assume the information is generated by a random variable Y, i.e. $\mathcal{G} = \sigma(Y)$. Classical probability is helpful for us to understand these properties. For example,

$$\begin{aligned} E(E(X|\mathcal{G})) =& E(E(X|\sigma(Y))) \\ =& E(E(X|Y)) = E(g(Y)) \\ =& \int_{-\infty}^{\infty} g(y)f_Y(y)dy \\ =& \int_{-\infty}^{\infty} E(X|Y=y)f_Y(y)dy = EX. \end{aligned} \tag{1}$$

Similarly, try to understand other properties (excellent exercises)!

A rigorous proof by the formal definition

We have

$$\mathbb{E}[\mathbb{E}(X|\mathcal{G})] = \mathbb{E}X$$

According to the definition of conditional expectation, we know that

$$\mathbb{E}[\mathbb{E}(X|\mathcal{G})1_A] = \mathbb{E}(X1_A),$$

for any $A \in \mathcal{G}$. Taking $A = \Omega$, we have

$$\mathbb{E}[\mathbb{E}(X|\mathcal{G})] \equiv \mathbb{E}[\mathbb{E}(X|\mathcal{G})1_\Omega] = \mathbb{E}(X1_\Omega) \equiv \mathbb{E}(X).$$

Conditional Expectation as the Optimal Projection

- Let X be a $\mathcal{F}$-measurable (random variable) R.V..
- For $\mathcal{G} \subset \mathcal{F}$, denote by $L^2(\mathcal{G})$ the collection of $\mathcal{G}$-measurable R.V. satisfying that $\mathbb{E}Z^2 < \infty$.
- Question: for what Z, do we have

$$\min_{Z \in L^2(\mathcal{G})} \mathbb{E}(X - Z)^2?$$

- Answer:

$$\mathbb{E}(X - \mathbb{E}(X|\mathcal{G}))^2 = \min_{Z \in L^2(\mathcal{G})} \mathbb{E}(X - Z)^2?$$

- i.e. the mean square distance is attained by the conditional expectation
- Analogy: the projection of X

Overview

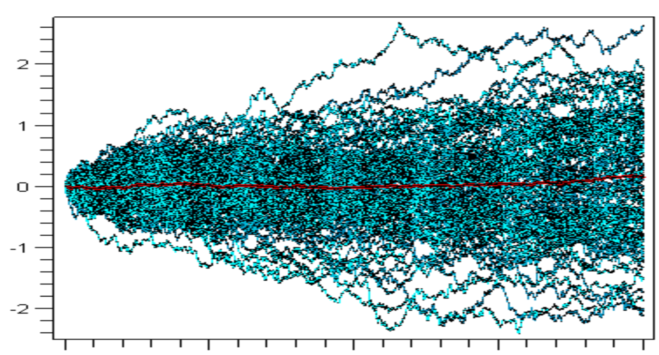
- Stochastic processes, e.g. random walk and Poisson processes, etc.

Stochastic Process

- A collection of random variables:

$$\{X(t), 0 \le t \le T\}$$

- For emphasizing the dependence of both time and random events, we heuristically write $X(t, \omega)$.
- Understanding:
  - A function of two variables $(\omega, t) \in \Omega \times (0, T]$
  - Fix any $t \in (0, T]$, $X(t, \cdot)$ is a random variable
  - Fix any $\omega \in \Omega$, $X(\cdot, \omega)$ is a realization (path) of the process



Stochastic Process: Formal Definition

- **Definition**: Let Ω be a nonempty set and T be a fixed positive number. Assume that for each $t \in (0, T]$, there exists a σ-algebra (information) $\mathcal{F}(t)$, and that for any $s \le t$, $\mathcal{F}(s) \subseteq \mathcal{F}(t)$. Then we call the collection of σ-algebra $\mathcal{F}(t)$, $0 \le t \le T$, a filtration.
- **Definition**: Let Ω be a nonempty set equipped with a filtration $\mathcal{F}(t)$, $0 \le t \le T$. Let $\{X(t); 0 \le t \le T\}$ be a collection of RVs indexed by $t \in [0, T]$. We say this collection of RVs is an adapted (continuous time) stochastic process if for each t, we have $X(t)$ is $\mathcal{F}(t)$-measurable.
- Similarly define filtration (denoted by $\mathcal{F}_n$) and stochastic process (denoted by $\{X_n : n \in \mathbb{N}\}$) on discrete time
- Note: the state space (all possible values) of $X(t)$ (or X_n) is continuous or not is a different concept.

Stochastic Process: Heng Seng Index as an Example

Stochastic Process: Some Concepts and Properties

- Finite dimensional distributions: the distribution of $(X(t_1), X(t_2), ..., X(t_n))$, where $t_1 < t_2 < ... < t_n$
- Covariance function: $c(t, x) = cov(X(t), X(s))$
- A process is stationary: iff the finite dimensional distribution is translate invariant, i.e.

  $$(X(t_1), X(t_2), ..., X(t_n)) =^d (X(t_1 + h), X(t_2 + h), ..., X(t_n + h)).$$

  $A =^d B$ means the distribution of A and B are the same.
- A process has stationary increments: iff

  $$X(t) - X(s) =^d X(t + h) - X(s + h).$$

- A process has independent increments: iff

  $$X(t_2) - X(t_1), ..., X(t_n) - X(t_{n-1})$$

  are all independent.

Markov Property

- **Definition**: Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a filtration $\mathcal{F}(t)(\subseteq \mathcal{F})$, $0 \le t \le T$, and an adapted stochastic process $X(t)$, $0 \le t \le T$. Assume that for all $0 \le s \le t \le T$ and for every function f, there exists another function g such that

  $$\mathbb{E}[f(X(t))|\mathcal{F}(s)] = \mathbb{E}[f(X(t))|X(s)](= g(X(s)))$$

  for some function g and all $0 \le s \le t \le T$, we say the process $X(t)$ enjoys the Markov property.
- A process is Markov if it totally forgets about the past: the future depends on the current state only!
- Markov processes in the discrete case – Markov Chains.
- Note that g depends on $f(\cdot)$, t, and s.

Transition Probability and Transition density

- Describe the transition from one state to another of a Markov process
- For discrete time Markov process with discrete state space, we consider the following transition probability:

$$\mathbb{P}(X_{n+1} = a | X_n = b, X_{n-1} = c, ...) = \mathbb{P}(X_{n+1} = a | X_n = b)$$

- For continuous time Markov process with continuous state space, we consider the transition density: for $0 \leq s < t$,

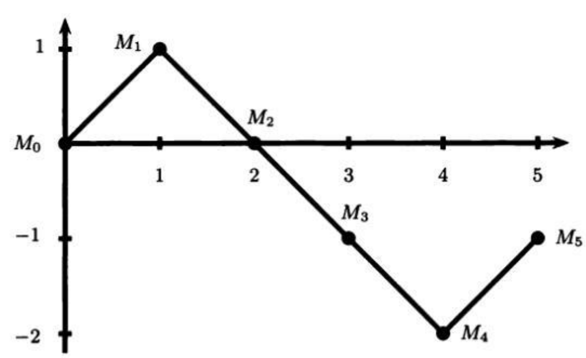
$$p(t, y; s, x) := \frac{d}{dy} P(X(t) \leq y | X(s) = x).$$

An example of discrete time process: Simple Random Walk

A simple symmetric random walk

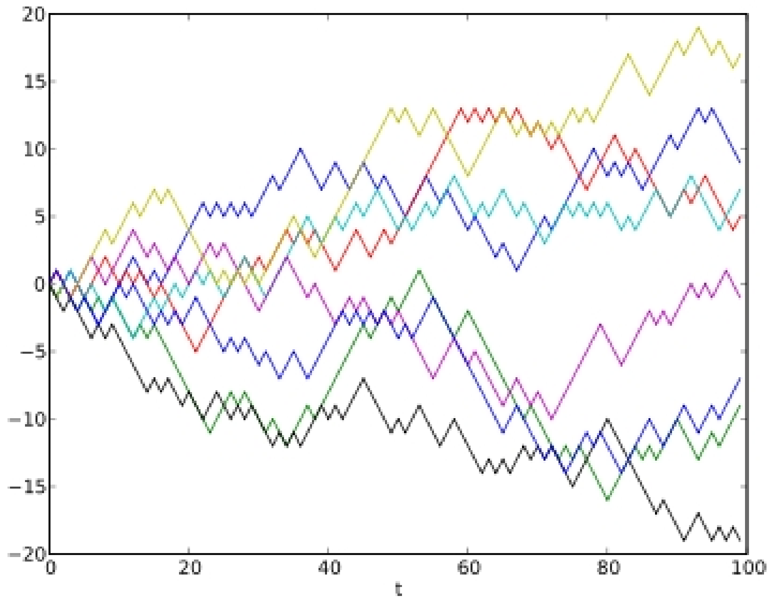
$$W_0 = 0, \quad W_n = \sum_{k=1}^{n} X_k,$$

where $X_k = 1$ or $X_k = -1$ with probability $1/2$.



Five steps of a random walk.

More Sample Paths for a Simple Random Walk



Some Properties of Simple Random Walk

- A simple symmetric random walk

$$W_0 = 0, \quad W_n = \sum_{k=1}^{n} X_k,$$

where $X_k = 1$ or $X_k = -1$ with probability $1/2$.
- A Markov process
- Transition probability:

$$\mathbb{P}(W_{n+1} = s | W_n = r) = \begin{cases} 1/2, & \text{if } s = r + 1; \\ 1/2, & \text{if } s = r - 1. \end{cases}$$

- For $0 < m < n$, conditional expectation: $E(W_n | W_m) = ?$
- Conditional second moments: $E(W_n^2 | \mathcal{F}_m) = ?$ where $\mathcal{F}_m = \sigma(X_1, ..., X_m)$

An example of continuous time process: Poisson Process

- Model arrivals: e.g. a "jump" in financial market
- A Poisson process $\{N(t)\}$ with intensity λ:
  - N(0)=0
  - stationary and independent increment
  - $N(t+\Delta t) - N(t)$ is a Poisson R.V. with parameter $\lambda \Delta t$

$$\mathbb{P}\left(N(t+\Delta t) - N(t) = k\right) = \frac{\lambda^k \Delta t^k}{k!} e^{-\lambda \Delta t} \tag{2}$$

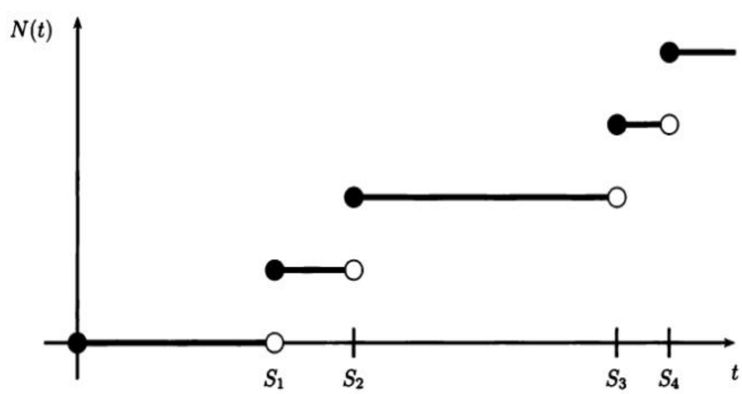
- Constriction from exponential R.V.s,

$$N(t) = \max\left\{n : S_n = \sum_{i=1}^{n} \tau_i \le t\right\},$$

  where τ_i are I.I.D with exponential distribution: $\mathbb{P}(\tau_i \le x) = 1 - e^{-\lambda x}$, for $x \ge 0$.
- (4) can be proved using this construction by computing on the density of $\sum_{i=1}^{n} \tau_i$ (an excellent exercise).

Poisson Process

- The path is right continuous



- Expectation of the increment $\mathbb{E}[N(t+\Delta t) - N(t)] = \lambda \Delta t$; this is why λ is called "intensity": arrival per unit time
- What is the variance of the increment?
- Further reading (if you are interested in financial modeling): compounded Poisson process, non-constant intensity, etc.

Overview

- Martingales
- Applications

Martingales: Intuition and History

- A model of a fair game
- A stochastic process such that the conditional expected value of an observation at some time t, given all the observations up to some earlier time s, is equal to the observation at that earlier time s.
- Referred to a class of betting strategies that was popular in 18th century France
- The strategy had the gambler double his bet after every loss so that the first win would recover all previous losses plus win a profit equal to the original stake. As the gambler's wealth and available time jointly approach infinity, his probability of eventually flipping heads approaches 1, which makes the martingale betting strategy seem like a sure thing. However, the exponential growth of the bets eventually bankrupts its users.
- "Saint Petersburg Paradox"
- e.g. win at the fourth bet: $8-(1+2+4)=1$

Martingales: Formal Definition and Constant Expectation

- Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a filtration $\mathcal{F}(t)(\subseteq \mathcal{F})$, $0 \leq t \leq T$, and an adapted stochastic process $M(t)$, $0 \leq t \leq T$.
  - If $E[M(t)|\mathcal{F}(s)] = M(s)$ for all $0 \leq s \leq t \leq T$, we say this process $M(t)$ is a martingale.
  - If $E[M(t)|\mathcal{F}(s)] \geq M(s)$ for all $0 \leq s \leq t \leq T$, we say this process $M(t)$ is a submartingale.
  - If$E[M(t)|\mathcal{F}(s)] \leq M(s)$ for all $0 \leq s \leq t \leq T$, we say this process $M(t)$ is a supermartingale.
- Discrete-time martingale M_n, $n \geq 1$, iff

$$\mathbb{E}(M_{n+1}|\mathcal{F}_n) = M_n$$

- Martingales have constant expectation over time: for any $T > t > 0$

$$\mathbb{E}M(t) = \mathbb{E}M(0)$$

Example 1: Martingale Betting Strategy (Ancient Finance)

- Betting strategy: keep doubling your bet until you eventually win
- V_n: the winnings/losses up through n trials using this strategy, $n \geq 1$
- When we win, we stop playing: $\mathbb{P}(V_{n+1} = 1|V_n = 1) = 1$
- For first n trials resulting losses, we have $V_n = -(1 + 2 + 4 + ... + 2^{n-1}) = -(2^n - 1)$

$$\begin{aligned} \mathbb{P}(V_{n+1} = 1|V_n = -(2^n - 1)) = 1/2 \\ \mathbb{P}(V_{n+1} = -(2^{n+1} - 1))|V_n = -(2^n - 1)) = 1/2 \end{aligned} \tag{3}$$

- We have $\mathbb{E}(V_{n+1}|\mathcal{F}_n) = V_n$ for $n \geq 1$.
- On average, the winnings/losses doesn't change over time.
- So, we have $\mathbb{E}V_n = \mathbb{E}V_1$. And, $\mathbb{E}V_1 = 0$ implies that $\mathbb{E}(V_n) = 0$.

Blackjack Martingale Betting Strategy

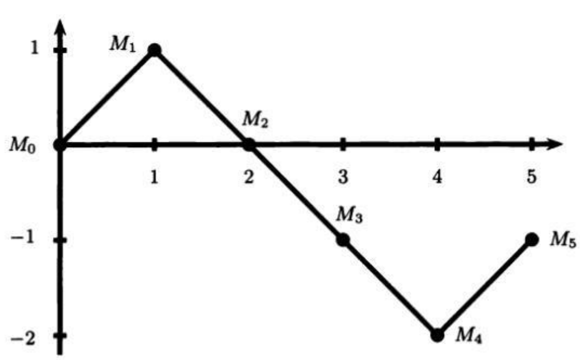
- Blackjack
- MIT Blackjack team
- Martingale betting strategies in Blackjack

Example 2: Simple Random Walk as a Martingale

A simple symmetric random walk

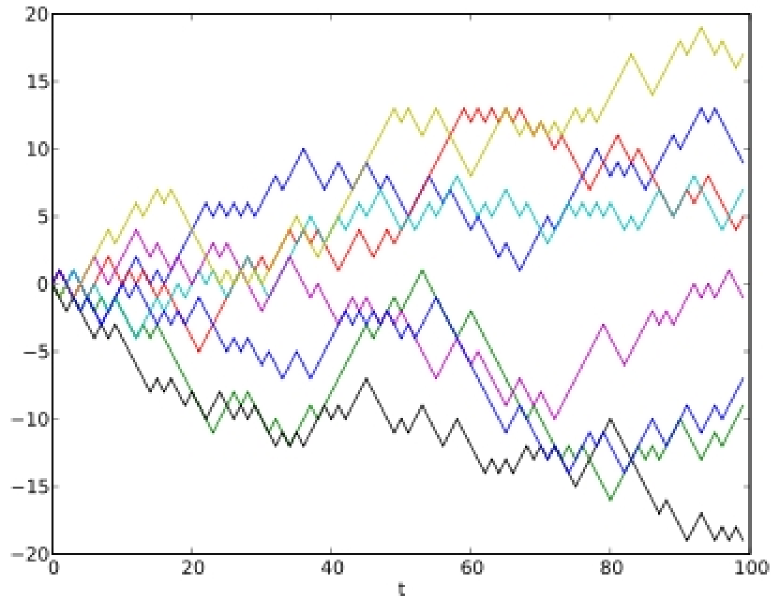
$$W_0 = 0, \quad W_n = \sum_{k=1}^{n} X_k,$$

where $X_k = 1$ or $X_k = -1$ with probability $1/2$.



Five steps of a random walk.

More Sample Paths for a Simple Random Walk



Some Properties of Simple Random Walk

- A simple symmetric random walk

$$W_0 = 0, \quad W_n = \sum_{k=1}^{n} X_k,$$

where $X_k = 1$ or $X_k = -1$ with probability $1/2$.
- A Markov process
- Transition probability:

$$\mathbb{P}(W_{n+1} = s | W_n = r) = \begin{cases} 1/2, & \text{if } s = r + 1; \\ 1/2, & \text{if } s = r - 1. \end{cases}$$

Simple Random Walk as a Martingale

- $\{W_n\}$ is martingale with respect to filtration $\mathcal{F}_n = \sigma(X_1, ..., X_n)$: just prove that
$$\mathbb{E}(W_{n+1}|\mathcal{F}_n) = W_n.$$
- $\{W_n^2 - n\}$ is also a martingale: just need to prove that
$$E(W_{n+1}^2 - (n+1)|\mathcal{F}_n) = W_n^2 - n.$$
- Generalization to asymmetric random walk (an excellent exercise): random walk with a drift (trend)

Example 3: Martingale Transform as a Discrete-time Integral

- Definition: $\{\vartheta_n\}$ is predictable iff ϑ_n is $\mathcal{F}_{n-1}$ measurable, i.e. $\{\vartheta_n\}$ is known up to the information up to $n-1$.
- "an investment decision is made according to the information previously at hand already."
- Given a martingale $\{M_n\}$, we construct
$$T_0 = 0,\ T_n = \sum_{k=1}^{n} \vartheta_k (M_k - M_{k-1})$$
- T_n is an analog to the change of wealth resulted from a dynamic trading strategy ϑ
- $\{T_n\}$ is a martingale (it is called martingale transform)
- T_n resembles an "integration" of ϑ with respect to M
- Later in this course, we will work on the stochastic integral which is a continuous time analog of this.

Recap: Poisson Process

- Model arrivals: e.g. a "jump" in financial market
- A Poisson process $\{N(t)\}$ with intensity λ:
  - N(0)=0
  - stationary and independent increment
  - $N(t+\Delta t) - N(t)$ is a Poisson R.V. with parameter $\lambda \Delta t$

$$\mathbb{P}\left(N(t+\Delta t) - N(t) = k\right) = \frac{\lambda^k \Delta t^k}{k!} e^{-\lambda \Delta t} \tag{4}$$

- Constriction from exponential R.V.s,

$$N(t) = \max\left\{n : S_n = \sum_{i=1}^{n} \tau_i \leq t\right\},$$

  where τ_i are I.I.D with exponential distribution:
  $\mathbb{P}(\tau_i \leq x) = 1 - e^{-\lambda x}$, for $x \geq 0$.

Example 4: Compensated Poisson process as a Martingale

- $\{N(t) - \lambda t; t \geq 0\}$ is martingale. Why?

- We will see more as we introduce Brownian motion and stochastic integral

More Examples (Excellent Exercises)

- Levy martingale:
$$X_n = \mathbb{E}(X|\mathcal{F}_n)$$
- Product:
$$P_0 = 1,\ P_n = \mu^{-n} \prod_{j=1}^{n} X_j,$$
where X_j are I.I.D with mean $\mathbb{E}(X_j) = \mu$
- Wald martingale:
$$W_0 = 1,\ W_n = \frac{e^{\theta \sum_{j=1}^{n} X_j}}{(\phi(\theta))^n},$$
where X_j are I.I.D with moment generating function $\phi(\theta) = \mathbb{E}e^{\theta X_i}$

Some Properties: Martingale with Stopping

- Definition of **stopping times**: a random variable τ such that $\{\tau \le t\} \in \mathcal{F}(t)$ for any $0 < t < T$
- Stopping or not at time t is totally determined by the information up to that time
- Some technical properties:
  - $\sigma-$algebra $\mathcal{F}(\sigma) = \{A \in \mathcal{F} : A \cap \{\tau \le t\} \in \mathcal{F}(t)\}$
  - τ is $\mathcal{F}(\tau)$ measurable
  - $\mathcal{F}(\sigma) \subset \mathcal{F}(\tau)$ if $\sigma \le \tau$
  - $\mathcal{F}(\min(\sigma, \tau)) = \mathcal{F}(\sigma) \cap \mathcal{F}(\tau)$
- An important class of stopping times: first hitting time of a set A:
$$\tau_A = \inf\{t \ge 0 : X(t) \in A\}$$

Application: Gambler's Problem

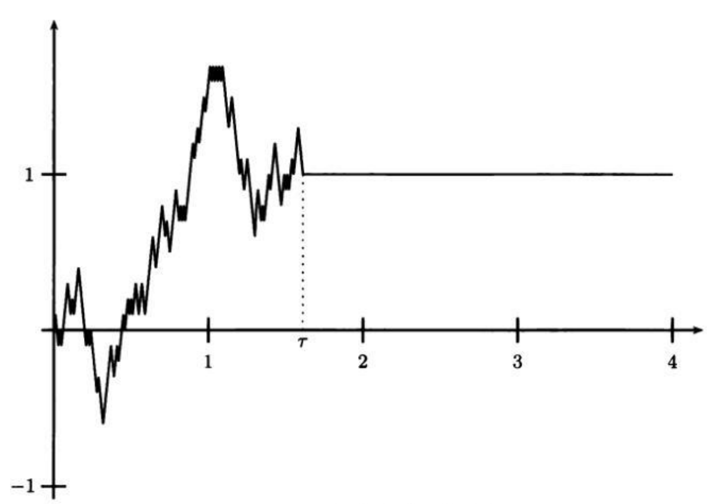
- Suppose you start with n dollars, and make a sequence of bets. For each bet, you win 1 dollar with probability $1/2$, and lose 1 dollar with probability $1/2$. You quit if either you go broke, in which case you lose, or when you reach $n+m$ dollars, what is the probability of your winning?
- The time τ at which he need to quit with a profit (or goes broke and is forced to quit) is a stopping time, i.e.

$$\tau = \min\{s \in \mathbb{N}, W_s = m \text{ or } -n\}.$$

- Previously, we can use R.W. to model and solve it by conditioning.
- Now, we solve this more systematically using "martingale"! How?

Stopped Process and Martingale

- Define: stopped process $X^\tau(t) = X(\min(t,\tau))$ (truncation up to a stopping time)
- Proposition: **stopped martingale is a martingale**, i.e. for martingale $\{X(t)\}$ adapted to filtration $\{\mathcal{F}(t)\}$, $\{X^\tau(t)\}$ is a martingale adapted to the stopped filtration $\{\mathcal{F}^\tau(t) = \mathcal{F}_{\min(t,\tau)}\}$



A stopped process.

Optional Sampling Theorem

- If $\{X_t\}$ is a martingale, under some technical conditions, the constant expectation property can be extended to stopping time τ, i.e.
$$\mathbb{E}X_\tau = \mathbb{E}X_0?$$
- Connection with betting strategies: impossible for a gambler to improve their betting strategies τ to obtain bigger expected profit!
- Interpretation of the Gambler's Ruin Problem
The gambler's fortune over time is a martingale; and the time τ at which he need to quit with a profit (or goes broke and is forced to quit) is a stopping time, i.e.
$$\tau = \min\{s \in \mathbb{N}, W_s = m \text{ or } -n\}.$$
So the optional sampling theorem says that (why?)
$$\mathbb{E}W_\tau = \mathbb{E}W_0 = 0.$$

Optional Sampling Theorem

Some technical conditions are necessary for establishing the optional sampling theorem:

$$\mathbb{E}X_\tau = \mathbb{E}X_0,$$

e.g.

- If τ is a bounded stopping time, i.e. $\mathbb{P}(\tau \le M) = 1$,
- If the following conditions hold
  - $\mathbb{P}(\tau < \infty) = 1$
  - $\mathbb{E}|X_\tau| < \infty$
  - $\lim_{n\to\infty} \mathbb{E}(|X_n| 1\{\tau > n\}) = 0$
- other conditions...

Counter example: consider the random walk W_n and let $\tau = \min\{j \in \mathbb{N}, W_j = 1\}$, we have

$$1 = \mathbb{E}W_\tau \neq EW_0 = 0.$$

Solve Gambler's Problem

By the martingale property of the stopped martingale W_r^τ, we have

$$\mathbb{E}W_{\min\{r,\tau\}} = \mathbb{E}W_0 = 0.$$

By the fact that W_r^τ is bounded, we use the dominated convergence theorem to let $r \to \infty$ and obtain the optional sampling theorem, i.e.

$$\mathbb{E}W_\tau = \mathbb{E}W_0.$$

Therefore,

$$\begin{aligned} &m\mathbb{P}(W_\tau = m) - n\mathbb{P}(W_\tau = -n) = 0, \\ &\mathbb{P}(W_\tau = m) + \mathbb{P}(W_\tau = -n) = 1. \end{aligned} \tag{5}$$

Therefore, the probability of win is

$$\mathbb{P}(W_\tau = m) = \frac{n}{m+n}.$$

Question: Find $\mathbb{E}\tau$.

A Technical Generalization: Local Martingales

- For technical purposes, the literature has developed the notion of local martingale, which generalize the notation of martingale.
- This is because martingale is sometimes too restrictive. By generalization to local martingales, one is still able to get something useful.
- A stochastic process $\{M(t)\}$ adapted to the filtration $\{\mathcal{F}(t)\}$ is called a local martingale if and only if there exists an increasing sequence $\{\tau_n\}_{n=1}^\infty$ of stopping times with $\lim_{n\to\infty} \tau_n = \infty$ such that the stopped process $M(t \wedge \tau_n)$ is a martingale adapted to the filtration $\{\mathcal{F}(t)\}$ for every $n \geq 1$.
- It can be easily shown that every martingale is also a local martingale.
- There exist local martingales which are not martingales.

A Technical Generalization: Local Martingales

To understand the definition, we suggest the following excellent exercise. We can show that every nonnegative local martingale is a supermartingale. Indeed, we have

$$\mathbb{E}[M(t \wedge \tau_n)|\mathcal{F}(s)] = M(s \wedge \tau_n) \qquad \texttt{martingale}$$

for every $0 < s < t$. Taking liminf, we have

$$\lim_{n\to\infty} \inf \mathbb{E}[M(t \wedge \tau_n)|\mathcal{F}(s)] = \lim_{n\to\infty} \inf M(s \wedge \tau_n) \equiv M(s).$$

By the Fatou lemma, we have

$$\lim_{n\to\infty} \inf \mathbb{E}[M(t\wedge\tau_n)|\mathcal{F}(s)] \geq \mathbb{E}[\lim_{n\to\infty} \inf M(t\wedge\tau_n)|\mathcal{F}(s)] \equiv \mathbb{E}[M(t)|\mathcal{F}(s)].$$

Thus, we arrive at $\mathbb{E}[M(t)|\mathcal{F}(s)] \leq M(s)$, which asserts the supermartingale property.

Doob decomposition in discrete-time cases

Definition

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random sequence $\{A_n\}_{n=0}^{\infty}$ adapted to the filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$. $\{A_n\}$ is called **predictable**, if for any $n \geq 1$, A_n is $\mathcal{F}_{n-1}$ – measurable.

Theorem (Doob Decomposition)

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose $\{X_n\}_{n=0}^{\infty}$ is a submartingale adapted to the filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$. Then X_n can be uniquely decomposed into two parts,

$$X_n = M_n + A_n,$$

where $\{M_n\}$ is a martingale adapted to $\{\mathcal{F}_n\}$, and $\{A_n\}$ is a non-decreasing, predictable sequence of random variables with $A_0 = 0$.

Doob decomposition: the solution

Solution:

$$\begin{aligned} A_n &= \sum_{k=1}^{n} (\mathbb{E}[X_k|\mathcal{F}_{k-1}] - X_{k-1}), \\ M_n &= X_n - A_n, \end{aligned}$$

So we only need to verify:

- A_n is a non-decreasing sequence;
- A_n is predictable;
- $M_n = X_n - A_n$ is a martingale.

Examples

1. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$ and an adapted i.i.d. random sequence $X_1, X_2, \cdots$. Suppose

$$S_n = \sum_{k=1}^{n} X_k.$$

If $\mathbb{E}(X_1) = \mu > 0$, then S_n is a submartingale and can be expressed as the following Doob decomposition:

$$S_n = (S_n - n\mu) + n\mu.$$

2. Suppose that S_n is a symmetric random walk, then S_n^2 is a submartingale and can be expressed as the following Doob decomposition:

$$S_n^2 = (S_n^2 - n) + n.$$

Doob-Meyer decomposition in continuous-time cases

1. The Doob decomposition can be similarly developed for continuous-time cases.
2. Under some technical conditions, a continuous-time submartingale can be uniquely decomposed as the sum of a martingale and a non-decreasing "good" process, here "good" means the process has some good properties similar with predictability.
3. This theorem is called the Doob-Meyer decomposition. Since we focus on applications, we will not state the theorem rigorously. Interested readers may refer to the corresponding section in Section 1.5 of Karaztas and Shreve (1991). Instead, we present the following examples of Doob-Meyer decomposition The to be verified by readers.

Doob-Meyer decomposition in continuous-time cases

1. For a Brownian motion $W(t)$, suppose $X(t) = aW(t) + \mu t$ for some $\mu > 0$. Then $X(t)$ is a submartingale and $X(t) = aW(t) + \mu t$ gives its Doob-Meyer decomposition.
2. For a Brownian motion $W(t)$, $W^2(t)$ is a submartingale. Then,
$$W^2(t) = [W^2(t) - t] + t$$
renders its Doob-Meyer decompostion.
3. For a Poisson process $N(t)$ with parameter λ, $N(t)$ is a submartingale. Then, its Doob-Meyer decomposition is given by
$$N(t) = [N(t) - \lambda t] + \lambda t.$$
4. For a Poisson process $N(t)$ with parameter λ, show that $\eta(t) = [N(t) - \lambda t]^2$ is a submartingale and find its Doob-Meyer decomposition. (Hint: $M(t) = N(t) - \lambda t$ is a martingale, and $\langle M \rangle (t) = \lambda t$.)

Quadratic variation

If M_n (resp. $M(t)$) is a martingale, then M_n^2 (or $M^2(t)$) is a submartingale. If M_n (resp. $M(t)$) is a martingale, then

$$\mathbb{E}[M_{n+1}^2|\mathcal{F}_n] - M_n^2 = \mathbb{E}[(M_{n+1} - M_n)^2|\mathcal{F}_n], \tag{6}$$

(resp.

$$\mathbb{E}[M^2(t)|\mathcal{F}(s)] - M^2(s) = \mathbb{E}[(M(t) - M(s))^2|\mathcal{F}(s)].)$$

- We assume that $\mathbb{E}(M_n^2) < \infty$ (resp. $\mathbb{E}M^2(t) < \infty$), i.e., M_n (resp. $M(t)$) are *square-integrable martingales*.
- According to the previous discussions, M_n^2 (resp. $M^2(t)$) is a submartingale, then applying Doob (– Meyer) Decomposition to M_n^2 (resp. $M^2(t)$), we can uniquely decompose them into $M_n^2 = N_n + A_n$ (resp. $M^2(t) = N(t) + A(t)$), i.e., $N_n = M_n^2 - A_n$ (resp. $N(t) = M^2(t) - A(t)$) is a martingale.

Quadratic variation

Definition

For a square-integrable martingale M_n (resp. $M(t)$), the non-decreasing random part in the Doob (-Meyer) decomposition of its square is called the quadratic variation, denoted as $\langle M \rangle_n$ (resp. $\langle M \rangle(t)$).

By the definition of quadratic variation, for a square-integrable martingale M, $M^2 - \langle M \rangle$ is also a martingale.
For discrete-time cases, according to formula of A_n, the quadratic variation has the following explicit form:

$$\langle M \rangle_n = \sum_{k=1}^{n} \left(\mathbb{E}[M_k^2|\mathcal{F}_{k-1}] - M_{k-1}^2\right) = \sum_{k=1}^{n} \mathbb{E}[(M_k - M_{k-1})^2|\mathcal{F}_{k-1}], \tag{7}$$

where the second equality follows from (6).

Quadratic variation

For continuous-time cases, it is natural to conjecture that the quadratic variation of $M(t)$ follows from the continuous-time limit of a discrete analogy.

Definition

Denote by $\Pi: \ 0 = t_0 < t_1 < \cdots < t_m = t$ a partition of the interval $[0, t]$. Then the quadratic variation of M with respect to the partition Π is defined as

$$QV_t^M(\Pi) = \sum_{k=1}^{m} |M(t_k) - M(t_{k-1})|^2.$$

It can be verified that as the partition becomes finer and finer, i.e.,

$$\|\Pi\| = \max_{1 \le k \le m} |M(t_k) - M(t_{k-1})| \to 0,$$

$QV_t^M(\Pi)$ will converge to $\langle M \rangle(t)$ in some mode, i.e.,

$$\lim_{\|\Pi\| \to 0} QV_t^M(\Pi) = \langle M \rangle(t), \text{ in probability.}$$

Cross variation

1. Similarly, given two martingales M and N, is there any stochastic process A, such that $MN - A$ is a martingale?
2. If so, is such stochastic process unique?
3. What is its explicit expression?

Because of

$$MN = \frac{1}{4}(M+N)^2 - \frac{1}{4}(M-N)^2,$$

we have

$$\begin{aligned} & MN - \frac{1}{4}(\langle M+N \rangle - \langle M-N \rangle) \\ = \ & \frac{1}{4}\Big((M+N)^2 - \langle M+N \rangle\Big) - \frac{1}{4}\Big((M-N)^2 - \langle M-N \rangle\Big). \end{aligned}$$

Because the right side of the above equation is a martingale, so is the left side. Therefore, we can find the expression of A such that $MN - A$ is a martingale. Such an A is given by

$$A = \frac{1}{4}(\langle M+N \rangle - \langle M-N \rangle).$$

Cross variation

Definition

For two square-integrable martingales M and N, we define the cross-variation of M and N by

$$\langle M, N \rangle = \frac{1}{4} \left(\langle M + N \rangle - \langle M - N \rangle \right).$$

In discrete-time cases, $\langle M, N \rangle_n$ can be expressed as

$$\begin{aligned} \langle M, N \rangle_n &= \frac{1}{4} \left(\langle M + N \rangle_n - \langle M - N \rangle_n \right) \\ &\equiv \sum_{k=1}^{n} \mathbb{E}[(M_k - M_{k-1})(N_k - N_{k-1}) | \mathcal{F}_{k-1}]. \end{aligned}$$

Moreover, because $\langle M + N \rangle_n$ and $\langle M - N \rangle_n$ are both predictable, by definition, so is $\langle M, N \rangle_n$.

Cross variation

In continuous-time cases, when $\|\Pi\|$ approaches 0, $QV_t^M(\Pi)$ converges to $\langle M \rangle (t)$. Therefore, by the definition of $\langle M, N \rangle (t)$, we can use

$$\begin{aligned} & \frac{1}{4} \left(QV_t^{M+N}(\Pi) - QV_t^{M-N}(\Pi) \right) \\ = & \frac{1}{4} \left(\sum_{k=1}^{m} |M(t_k) + N(t_k) - M(t_{k-1}) - N(t_{k-1})|^2 \right. \\ & \left. - \sum_{k=1}^{m} |M(t_k) - N(t_k) - M(t_{k-1}) + N(t_{k-1})|^2 \right) \\ \equiv & \sum_{k=1}^{m} [M(t_k) - M(t_{k-1})][N(t_k) - N(t_{k-1})] \end{aligned}$$

to approach $\langle M, N \rangle$.

Uniqueness

Now, we move on to discuss the uniqueness of such stochastic process A such that $MN - A$ is a martingale.

- We already have as solution $A = \langle M, N \rangle$ that meets the requirements.
- Though MN is not necessarily a submartingale, as long as A is predictable, we have its uniqueness. Interested readers are suggested to prove it by mimicking our previous proof for the uniqueness of the Doob decomposition of submartingales.
- Besides, because of the similarity between the discrete and the continuous cases, it is reasonable to believe that, under some technical conditions, such decomposition is unique in continuous-time cases. The formulation and proof of such a claim is beyond the scope of these notes. Interested readers may refer to Section 1.5 of Karatzas and Shreve (1991).

Overview

- Definition of Brownian motion and its construction
- Basic properties
- Applications

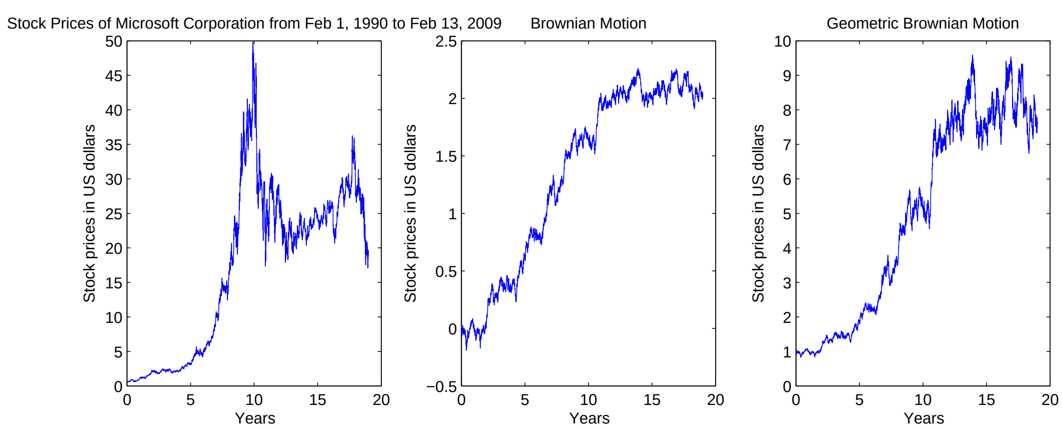
Definition of Standard Brownian Motion

Definition: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Stochastic process $\{W(t)\}$ is a (one-dimensional) standard Brownian motion (BM) if it satisfies that

- $W(0) = 0$;
- for each $\omega \in \Omega$, the realization (path) $W(t)(\omega)$ is a continuous function of $t \geq 0$;
- it has stationary increments with normal distribution $W(t) - W(s) \sim \mathrm{N}(0, t-s)$, and
- it has independent increments. More precisely, for all $0 = t_0 < t_1 < \cdots < t_m$, the increments $W(t_1) - W(t_0)$, $W(t_2) - W(t_1), \cdots, W(t_m) - W(t_{m-1})$ are independent.

Motivation to study Brownian motion

Empirical features of stock prices



Observations: Greatly volatile sample paths

- Brownian motion can serve as important building block
- We can even assume discontinuity or other features

Motivation to study Brownian motion

A Question: asset prices are observed at discrete time, why using Brownian motion (a continuous time stochastic process)?

- As the time increment is usually small, Brownian motion is a proper approximation
- Incorporate high frequency trading data
- Mathematically and numerically tractable (as we shall see)
- Easy to build on Brownian motion to obtain favorable features
- Many other important features as we shall see

Construction of Brownian Motion from Random Walk

Consider a symmetric random walk $M_n := \sum_{j=1}^{n} X_j$ for $n = 1, 2, ...;\ M_0 := 0$, where X_j are i.i.d. random variables such that

$$P(X_j = 1) = P(X_j = -1) = 0.5.$$

- $\{M_n\}$ is a martingale.
- Independent increments: $(M_{k_1} - M_{k_0}), (M_{k_2} - M_{k_1}), \cdots, (M_{k_m} - M_{k_{m-1}})$ are independent where $0 = k_0 < k_1 < \cdots < k_m$. Moreover,

$$\mathrm{Var}(M_{k_{i+1}} - M_{k_i}) = k_{i+1} - k_i.$$

  In particular, we have $\mathrm{Var}(M_k) = k$.

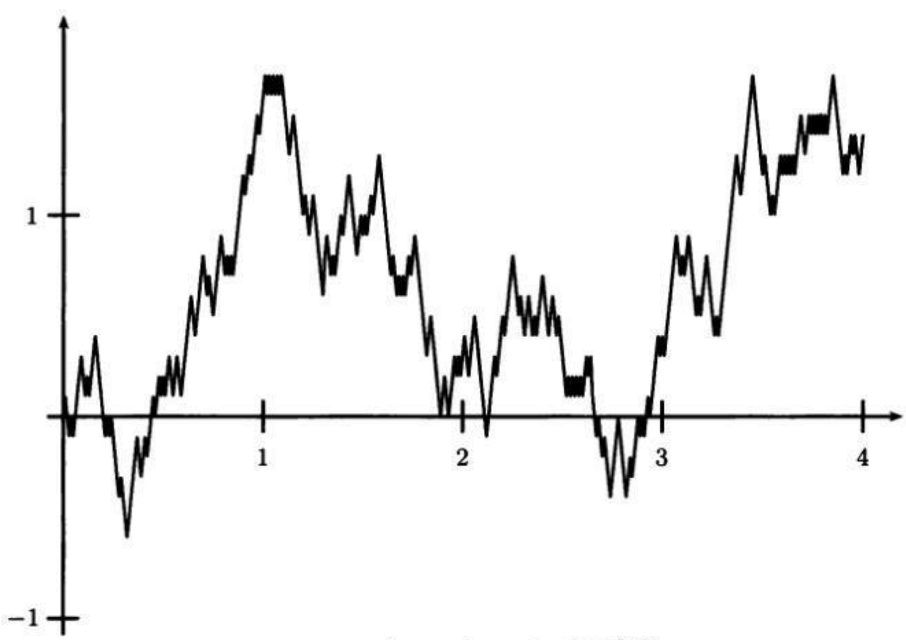
Construction of Brownian Motion from Random Walk

- Divide every unit time into n periods and define the scaled symmetric random walk:

$$W^{(n)}(t) = \frac{M_{nt}}{\sqrt{n}} \quad \text{if } nt \text{ is an integer.}$$

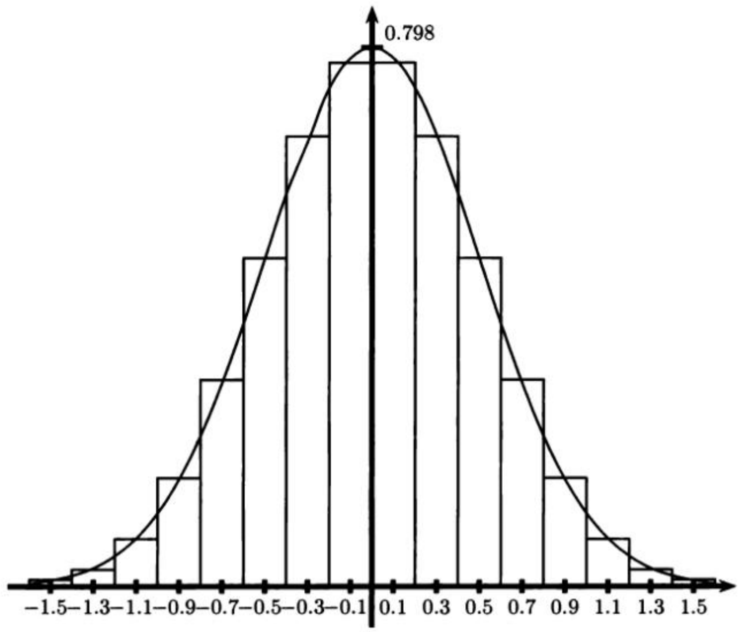
- **Magnify the local behavior** by nt and then **scale** it by $\frac{1}{\sqrt{n}}$.
- If nt and ns are integers, we have: $E(W^{(n)}(t) - W^{(n)}(s)) = 0$ and $Var(W^{(n)}(t) - W^{(n)}(s)) = t - s$
- **Theorem 3.2.1 (Central Limit)** Fix $t \geq 0$. As $n \to +\infty$, the distribution of the scaled random walk $W^{(n)}(t)$ evaluated at time t converges to $\mathrm{N}(0, t)$.

Construction of Brownian Motion from Random Walk



A sample path of $W^{(100)}$.

Construction of Brownian Motion from Random Walk



Distribution of $W^{(100)}(0.25)$ and normal curve $y = \frac{2}{\sqrt{2\pi}} e^{-2x^2}$.

Construction of Brownian Motion from Random Walk

- **Sketch of the Proof:** It suffices to show that the moment generating function $\phi_n(u) := Ee^{uW^{(n)}(t)}$ goes to

$$\phi(u) := Ee^{u\mathrm{N}(0,t)} \equiv e^{\frac{u^2 t}{2}}.$$

- Linear interpolation of $\frac{M_{\lfloor nt \rfloor}}{\sqrt{n}}$ and $\frac{M_{\lfloor nt \rfloor + 1}}{\sqrt{n}}$):

$$W^{(n)}(t) = \frac{M_{\lfloor nt \rfloor}}{\sqrt{n}} + \left(\frac{M_{\lfloor nt \rfloor + 1}}{\sqrt{n}} - \frac{M_{\lfloor nt \rfloor}}{\sqrt{n}} \right) (nt - \lfloor nt \rfloor).$$

- Not hard to calculate $\phi_n(u)$. Then basic algebra yields the results. □

Brownian motion (BM)–distribution

- **Definition:** $X(t)$ is Gaussian process if for any $0 = t_0 < t_1 < \cdots < t_m$ and $m \in \mathcal{N}$, $(X(t_1), X(t_2), \cdots, X(t_m))$ assumes multivariate normal distribution.
- BM is a Gaussian process
- Its mean and covariance function:
  - $EW(t) = 0$.
  - $E[W(t)W(s)] = t \wedge s := \min\{t, s\}$.
- What is the correlation function?
- Characterization of the distribution of $(W(t_1), W(t_2), \cdots, W(t_m))$
  - the moment generating function

$$\phi(u_1, u_2, \cdots, u_n) := \mathbb{E}e^{\sum_{i=1}^{m} u_i W(t_i)}.$$

  - the closed form expression of $\phi(u_1, u_2, \cdots, u_n)$ can be derived using property of independent and stationary increments.

Brownian motion – Filtration

- **Definition:** A filtration for the BM is a collection of σ-algebra $\mathcal{F}(t)$ such that
  - (Information accumulates) $\mathcal{F}(s) \subseteq \mathcal{F}(t)$ if $s < t$.
  - (Adaptivity) For any $t \geq 0$, $W(t)$ is $\mathcal{F}(t)$-measurable.
  - (Independence of future increments) If $u > t \geq 0$, $W(u) - W(t)$ is independent of $\mathcal{F}(t)$.
- Two possibilities of the filtration $\mathcal{F}(t)$.
  - $\mathcal{F}(t)$ contains only the information by observing the BM itself up to time t.
  - $\mathcal{F}(t)$ contains the information by observing the BM as well as other processes up to time t. In this case, the information of other processes cannot give any clues of the future increments of the BM.

Brownian motion – Some Fundamental Properties

- BM is a Markov process.

$$\mathbb{E}(f(W(t))|\mathcal{F}(s)) = \mathbb{E}(f(W(t))|W(s)), \quad \text{for } 0 < s < t.$$

- BM is a **strong Markov** process (generalize the Markov property to stopping times):
  Let τ be a finite stopping time ("known" to the Brownian filtration, i.e. $\{\tau < t\} \in \mathcal{F}(t)$, then

$$\mathbb{E}(f(W(\tau + t))|\mathcal{F}(\tau)) = \mathbb{E}(f(W(\tau + t))|W(\tau)).$$

- Implication of the strong Markov property: **Brownian motion refreshes after a stopping time**!

$$B(t) := W(\tau + t) - W(\tau)$$

  is again a Brownian motion independent of $\mathcal{F}(\tau)$.

Brownian motion – Some Fundamental Properties

- Brownian motion is a **martingale**:

$$\mathbb{E}(W(t)|\mathcal{F}(s)) = W(s), \text{ for } s < t.$$

- Invariance under time translation (a special case of "Brownian motion refreshing after a stopping time"):
  $B(t) = W(t + T) - W(T)$ is a Brownian motion independent of $\mathcal{F}(T)$
- Invariance under scaling: $B(t) = \frac{1}{\sqrt{c}} W(ct)$ is a BM for any given $c > 0$
- Invariance under symmetry: $B(t) = -W(t)$ is a BM
- Invariance under time-reversal: $B(t) = W(T) - W(T - t)$ is a BM for $0 \leq t \leq T$.

Brownian motion – Some Fundamental Properties

- BM is unbounded:

$$\mathbb{P}\left(\sup_{0\le t<\infty} W(t)=\infty\right)=1,\quad \mathbb{P}\left(\inf_{0\le t<\infty} W(t)=-\infty\right)=1.$$

- BM is recurrent; it visits every site on the real line and keeps returning to it **over and over again**. (this can be explained by the strong Markov property)

- The BM path is nowhere differentiable (very zigzag).

- Several related martingales: $W(t)$, $W^2(t)-t$, and $Z(t):=e^{\sigma W(t)-\frac{\sigma^2}{2}t}$.

Brownian motion – Quadratic variation

- **Quadratic variation** (the total variation of the second order) up to time k is

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

- Both the variance and the quadratic variation of the random walk accumulate at rate one per unit time. However, the difference is that the former is deterministic whereas the latter is random.

- Quadratic variation of the scaled random walk: $[W^{(n)},W^{(n)}](t)=\sum_{j=1}^{nt}\left[W^{(n)}\left(\frac{j}{n}\right)-W^{(n)}\left(\frac{j-1}{n}\right)\right]^2=\sum_{j=1}^{nt}\left[\frac{X_j}{\sqrt{n}}\right]^2=\sum_{j=1}^{nt}\frac{1}{n}=t.$

Brownian motion – Quadratic variation

Recall

- $W(t)$ seems to fluctuate very frequently (extreme zig-zagness)
- The scaled random walk $W^{(n)}(t)$ has a quadratic variation t

We anticipate that

- the first-order variation $FV_T(W)$ of the BM $W(t)$ is $+\infty$, i.e., for any $T > 0$,

$$\lim_{||\Pi|| \to 0} \sum_{j=0}^{n-1} |W(t_{j+1}) - W(t_j)| = +\infty,$$

  where $||\Pi|| := \max_{0 \le j \le n-1}(t_{j+1} - t_j)$ and $\Pi := \{t_0, t_1, \cdots, t_n\}$ is a **partition** of $[0, T]$.
- the quadratic variation $[W, W](t)$ of the BM $W(t)$ is t, i.e., it accumulates at rate 1 per unit time, i.e., for any $T > 0$,

$$\lim_{||\Pi|| \to 0} \sum_{j=0}^{n-1} [W(t_{j+1}) - W(t_j)]^2 = T.$$

Brownian motion – Quadratic variation

- **Proposition from Calculus:** If $f(t)$ is continuously differentiable (derivatives exist and smooth enough),
  - its first-order variation: $\sum_{j=0}^{n-1} |f(t_{j+1}) - f(t_j)| = \sum_{j=0}^{n-1} |f'(t_j^*)(t_{j+1} - t_j)| \to \int_0^T |f'(t)|dt < +\infty$, as $||\Pi|| \to 0$.
  - its quadratic variation: $\sum_{j=0}^{n-1} [f(t_{j+1}) - f(t_j)]^2 = \sum_{j=0}^{n-1} f'^2(t_j^*)(t_{j+1} - t_j)^2 \le ||\Pi|| \sum_{j=0}^{n-1} f'^2(t_j^*)(t_{j+1} - t_j) \to 0 \times \int_0^T f'^2(t)dt = 0$, as $||\Pi|| \to 0$.
- However, $W(t)$ is **non-differentiable everywhere** (extremely zigzag).
- **Theorem** For a BM $W(t)$, we have that
  - its first-order variation is: $FV_T(W) = +\infty$
  - its quadratic variation is: $[W, W](T) = T$ for all $T \ge 0$ almost surely

Brownian motion – Quadratic variation

- We write (ii) informally as

$$dW(t)dW(t) = dt,$$

meaning that

$$\lim_{||\Pi||\to 0} \sum_{j=0}^{n-1} [W(t_{j+1}) - W(t_j)]^2 = T.$$

- Similarly, we can obtain that

$$dW(t)dt = 0, \text{ and } dtdt = 0.$$

- Implications of the Theorem: The sample path of the BM must have an **infinite number of ups and downs**, each of which, however, is **infinitesimal**. So the **extreme zig-zagness** of the path implies its non-differentiability.

Brownian motion – First Passage Time Distribution

- The first passage time (FPT) of a process $Y(t)$ to a level m from below is defined to be

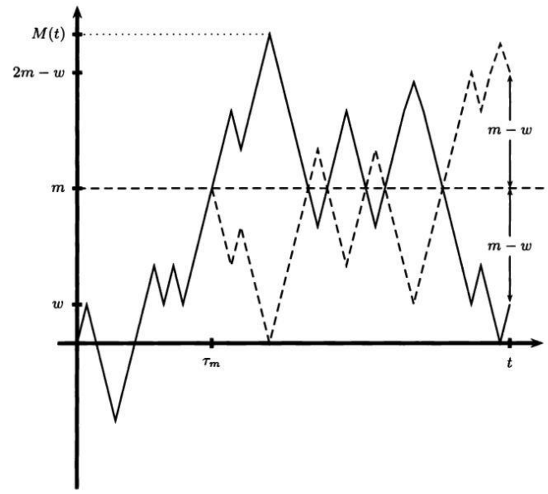
$$\tau_m := \inf\{t \geq 0 : Y(t) \geq m\},$$

where $\inf \emptyset := +\infty$.
- What is the distribution of the first passage time of $W(t)$ to m: τ_m
- Potential application: prediction of the behavior of an asset
- Potential application: modeling the credit default
- Two approaches to find this distribution
  - Calculate the distribution from the Reflection Principle
  - Calculate the Laplace transform of the probability density function (optional for self-reading)

Brownian motion – Reflect Principle and the FPT

Reflection Principle: Assume $m > 0$. If we "reflect the path after τ_m with respect to level m", we get a Brownian motion again! i.e.

$$\begin{aligned} \widetilde{W}(t) =& W(t),\ 0 \le t \le \tau_m; \\ =& 2m - W(t),\ t > \tau_m. \end{aligned} \tag{8}$$



Brownian path and reflected path.

Brownian motion – Reflect Principle and the FPT

Let $w \le m$. We obtain that

$$P(\tau_m \le t, W(t) \le w) = P(W(t) \ge 2m - w).$$

- Let $w = m$,

$$\begin{aligned} P(\tau_m \le t) =& P(\tau_m \le t, W(t) \le m) + P(\tau_m \le t, W(t) \ge m) \\ =& 2P(\tau_m \le t, W(t) \ge m) = 2P(W(t) \ge m) \\ =& \frac{2}{\sqrt{2\pi t}} \int_m^{+\infty} e^{-\frac{x^2}{2t}} dx. \end{aligned} \tag{9}$$

- The Brownian motion goes up or down with the same probability symmetrically
- Taking the derivative w.r.t. t yields the pdf

Brownian motion – The Historical Maximum

- Define the historical maximum $M(t) = \max_{0 \le s \le t} W(s)$, we have

$$P(\tau_m \le t) = P(M(t) \ge m).$$

- Potential application: prediction of the maximum of stock price! How?
-

$$\begin{aligned} P(M(t) \ge m, W(t) \le w) =& P(W(t) \ge 2m - w) \\ =& \frac{2}{\sqrt{2\pi t}} \int_{2m-w}^{+\infty} e^{-\frac{x^2}{2t}} dx. \end{aligned} \tag{10}$$

- Taking the derivative w.r.t. m and w and multiplying the result by -1 yields the joint pdf

$$f_{M(t),W(t)}(m, w) = \frac{2(2m - w)}{t\sqrt{2\pi t}} e^{-\frac{(2m-w)^2}{2t}}.$$

Brownian motion – Laplace Transform of the FPT

- Without loss of generality, we assume $m > 0$
- A common method to study a stopping time is to construct a martingale. Here we use $Z(t) := e^{\sigma W(t) - \frac{\sigma^2}{2} t}$.
- Then $Z(t \wedge \tau_m)$ is also a martingale, as implies that
  $1 = Z(0) = EZ(t \wedge \tau_m) = Ee^{\sigma W(t \wedge \tau_m) - \frac{\sigma^2}{2} t \wedge \tau_m} =$
  $E\left[e^{\sigma W(t \wedge \tau_m) - \frac{\sigma^2}{2} t \wedge \tau_m} I_{\{\tau_m < +\infty\}}\right] +$
  $E\left[e^{\sigma W(t \wedge \tau_m) - \frac{\sigma^2}{2} t \wedge \tau_m} I_{\{\tau_m = +\infty\}}\right] =$
  $E\left[e^{\sigma W(t \wedge \tau_m) - \frac{\sigma^2}{2} t \wedge \tau_m} I_{\{\tau_m < +\infty\}}\right] + E\left[e^{\sigma W(t) - \frac{\sigma^2}{2} t} I_{\{\tau_m = +\infty\}}\right].$
- Letting $t \to +\infty$ and applying the dominated convergence theorem, we obtain that
  $1 = E\left[e^{\sigma W(\tau_m) - \frac{\sigma^2}{2} \tau_m} I_{\{\tau_m < +\infty\}}\right] = E\left[e^{\sigma m - \frac{\sigma^2}{2} \tau_m} I_{\{\tau_m < +\infty\}}\right].$

Brownian motion – Laplace Transform of the FPT

- So we have that $E\left[e^{-\frac{\sigma^2}{2}\tau_m} I_{\{\tau_m<+\infty\}}\right] = e^{-\sigma m}$. Note that $E\left[e^{-\frac{\sigma^2}{2}\tau_m} I_{\{\tau_m=+\infty\}}\right] = 0$. We have $E\left[e^{-\frac{\sigma^2}{2}\tau_m}\right] = e^{-\sigma m}$
- Letting $\sigma \to 0+$ yields

$$P(\tau_m < +\infty) = 1.$$

- For $m \in \mathcal{R}$, the first passage time of the BM to m is finite almost surely, and the Laplace transform of its pdf is given by

$$E\left[e^{-\alpha\tau_m}\right] = e^{-|m|\sqrt{2\alpha}}.$$

- Taking derivative of $E\left[e^{-\alpha\tau_m}\right]$ with respect to α, we get

$$E\left[\tau_m e^{-\alpha\tau_m}\right] = \frac{|m|}{\sqrt{2\alpha}} e^{-|m|\sqrt{2\alpha}}.$$

- Letting $\alpha \to 0+$ leads to $E\tau_m = +\infty$ if $m \neq 0$.

Some Processes Derived from Brownian Motion

Building more processes from Brownian motion towards the goal of modeling financial market!

- Brownian motion with drift:

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

  Allow arbitrary ”volatility” and a ”trend”.
- Geometric Brownian motion (the celebrated Black-Schole-Merton (1973) model):

$$S(t) = \exp\{\sigma W(t) + \alpha t\}.$$

  A fundamental candidate for describing the financial asset price.
- Brownian Bridge: a process equivalent in law to a Brownian motion given a terminal value. example: $B(t) = W(t) - tW(1)$ is a Brownian bridge on $[0, 1]$ with terminal value 0.

Multidimensional Brownian Motion: Standard Case

Intuitively, a standard d-dimensional Brownian motion is d independent copies of standard one-dimensional Brownian motion.

Formal definition: a $d-$dimensional stochastic process $W(t) = (W_1(t), ..., W_d(t))$

- $W(0) = 0$;
- Independent increment
- For any $t > s$, $W(t) - W(s)$ has a joint normal distribution with mean 0 and covariance matrix $(t-s)I$.
- For any $i = 1, 2, ..., d$, $W_i(t)$ is a continuous function of t.

Multidimensional Brownian Motion: Correlated Case

Question: How about the correlated Brownian motions?
Answer: Change the covariance matrix to $(t-s)\Sigma$, where $\Sigma = (\rho_{ij})$. Here

$$\rho_{ij} = \text{Corr}(W_i(t), W_j(t)).$$

Connection with independent Brownian motion:
Cholesky decomposition: We can always find a standard $d-$dimensional Brownian motion $Z(t)$ such that

$$W(t) = AZ(t),$$

where A is sub-triangular matrix satisfying that $AA^T = \Sigma$.
An example:
for $d = 2$,

$$W_1(t) = Z_1(t),\ W_2(t) = \rho_{12}Z_1(t) + \sqrt{1-\rho_{12}^2}Z_2(t).$$

Overview

- Stochastic integral
- Ito's formulae
- Examples

Motivation

- Consider trading in an asset with unit price $W(t)$ (unrealistic, just for simplicity).
  - A partition $\Pi = \{t_0, t_1, \cdots, t_n\}$ s.t. $0 = t_0 < t_1 < \cdots < t_n = T$.
  - In the time period $[t_j, t_{j+1})$, hold Δ_j (adapted process) shares of this asset.
  - Note that the time period is left closed but right open.
- The gain process $I(t)$ at time $t \in [t_k, t_{k+1})$ is given by

$$I(t) = \sum_{j=0}^{k-1} \Delta(t_j)[W(t_{j+1}) - W(t_j)] + \Delta(t_k)[W(t) - W(t_k)]. \quad (11)$$

- Question: When $||\Pi|| := \max_{1 \le j \le n}(t_{j+1} - t_j)$ goes to zero, how to define the associated limiting gain process?
  - A kind of limit of summation (22) as $||\Pi|| \to 0$.
  - A kind of integral written as $\int_0^t \Delta(t)dW(t)$ (recall the definition of Riemann integral in ordinary calculus).

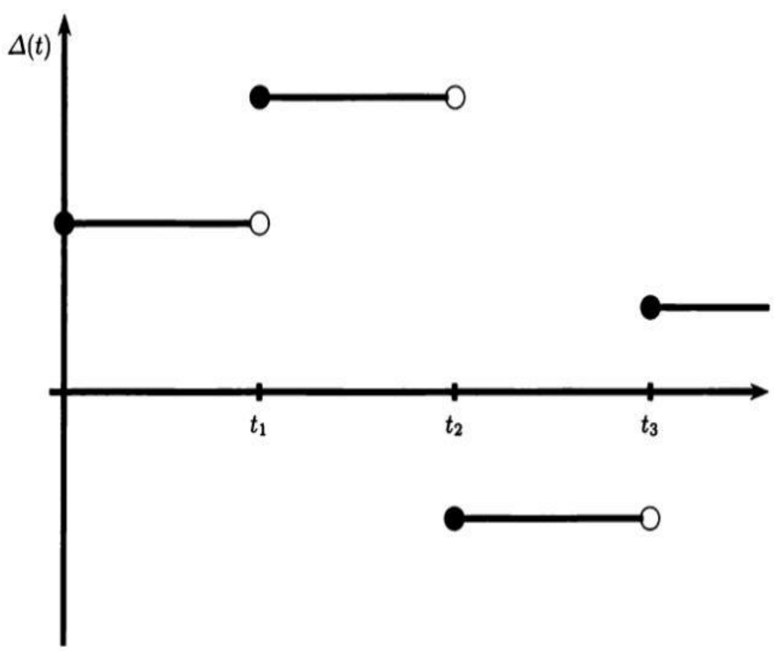
Motivation

- However, it is more complicated.
  - What is the definition of the related limit? In what sense?
  - It is not "traditional" (professionally speaking, Lebesgue or Riemann) integral because $W(t)$ is non-differentiable. It doesn't make sense that

$$\int_0^t \Delta(t)dW(t) = \int_0^t \Delta(t)W'(t)dt.$$

- First, define the integral for a simple process $\Delta(t)$, which is
  - adapted (the investment decisions are made based on the available information up to that time) and $E\int_0^t \Delta^2(u)du < +\infty$
  - equals $\Delta(t_j)$ in the time period $[t_j, t_{j+1})$ for any $j = 0, 1, \cdots, n-1$. (see a graph next page)
- Then the Itô integral at time $t \in [t_k, t_{k+1}]$ is defined to be (22).

Simple Process



A path of a simple process.

Construction of Itô integral

- Properties of Itô integral $I(t)$ for simple processes $\Delta(t)$.
  - (1) $I(t)$ is $\mathcal{F}_t$-measurable; **Linearity**;
  - (2) $I(t)$ is a **martingale**;
  - (3) **(Itô isometry)** $EI^2(t) = E\int_0^t \Delta^2(u)du$;
  - (4) **(Quadratic variation)** $[I, I](t) = \int_0^t \Delta^2(u)du.$
- Another way to express quadratic variation

$$dI(t)dI(t) = \Delta^2(t)dW(t)dW(t) = \Delta^2(t)dt.$$

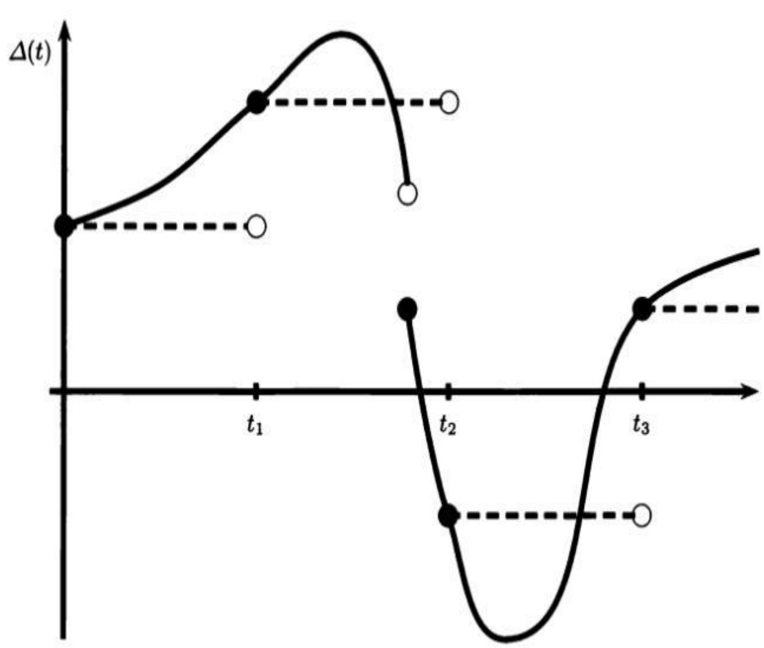
- Another way to express Itô integral

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

  (Differential Form).
- Second, let us construct Itô integral for a general adapted process $\Delta(t)$ that can be approximated by simple processes in some sense.

Construction of Itô integral

Approximate a general adapted process $\Delta(t)$ by simple processes.



Approximating a continuously varying integrand.

Construction of Itô integral

- We can find a sequence of simple processes Δ_n which approximate Δ.
- Note that $\int_0^T \Delta_n(t)dW(t)$ has already been well defined.
- It is natural to define the $\int_0^T \Delta(t)dW(t)$ to be a limit of

$$I_n(t) := \int_0^T \Delta^n(t)dW(t).$$

- Question (1): How do we know a limit exists? What do we mean by “limit"?
- Question (2): Is the limit unique?

Construction of Itô integral

- **Answer**: for an adapted process $\Delta(t) \in L^2[0,T]$, we can define the related Itô integral

$$\int_0^T \Delta(t)dW(t) := \lim_{n\to+\infty} \int_0^T \Delta_n(t)dW(t),$$

where $\{\Delta_n(t) \in L^2[0,T] : n = 0, 1, \cdots\}$ are a sequence of simple processes and the “limit" is unique in some sense.

- The so called “sense” is in $L^2(T)$ (square integrable).

- The meaning of $\int_0^T \Delta(t)dW(t)$: the gain process by holding $\Delta(t)$ shares of asset $W(t)$.

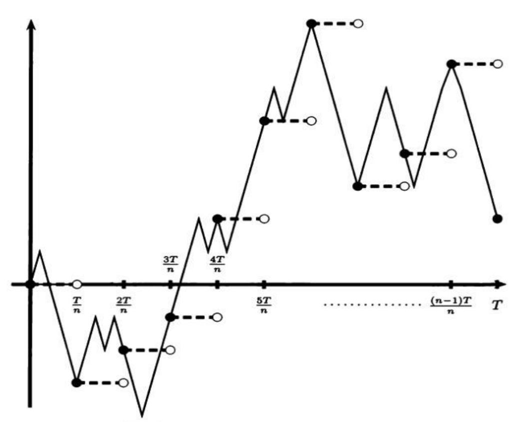
Properties of Itô integral

- **(Continuity)** $I(t)$ is continuous in t;
- **(Adaptivity)** $I(t)$ is $\mathcal{F}_t$-measurable;
- **(Linearity)** If $I(t) = \int_0^t \Delta(u)dW(u)$ and $J(t) = \int_0^t \Gamma(u)dW(u)$, then $I(t) \pm J(t) = \int_0^t (\Delta(u) \pm \Gamma(u))dW(u)$ and $cJ(t) = \int_0^t c\Gamma(u)dW(u)$ for any constant c;
- **(Martingale)** $I(t)$ is a martingale;
- **(Itô Isometry)** $EI^2(t) = E\int_0^t \Delta^2(u)du$;
- **(Quadratic variation)** $(I, I)(t) = \int_0^t \Delta^2(u)du$.

An Example

- **An Example**: Compute $\int_0^T W(t)dW(t)$.
- Select one particular sequence of simple processes as follows:

$$\Delta_n(t) = \sum_{j=0}^{n-1} W\left(\frac{jT}{n}\right) I_{\{t\in[\frac{jT}{n}, \frac{(j+1)T}{n})\}}.$$



Simple process approximating Brownian motion.

An Example

- Next, compute

$$\int_0^T \Delta_n(t)dW(t) := \sum_{j=0}^{n-1} W\left(\frac{jT}{n}\right)\left[W\left(\frac{(j+1)T}{n}\right) - W\left(\frac{jT}{n}\right)\right].$$

- By algebra, we have

$$\begin{aligned}
&2\sum_{j=0}^{n-1} W\left(\frac{jT}{n}\right)\left[W\left(\frac{(j+1)T}{n}\right) - W\left(\frac{jT}{n}\right)\right] \\
=&\sum_{j=0}^{n-1}\left[W^2\left(\frac{(j+1)T}{n}\right) - W^2\left(\frac{jT}{n}\right)\right] - \sum_{j=0}^{n-1}\left[W\left(\frac{(j+1)T}{n}\right) - W\left(\frac{jT}{n}\right)\right]^2 \quad (12)\\
=&W^2(T) - \sum_{j=0}^{n-1}\left[W\left(\frac{(j+1)T}{n}\right) - W\left(\frac{jT}{n}\right)\right]^2.
\end{aligned}$$

An Example

- Therefore,

$$\int_0^T \Delta_n(t)dW(t) = \frac{1}{2}W^2(T) - \frac{1}{2}\sum_{j=0}^{n-1}\left[W\left(\frac{(j+1)T}{n}\right) - W\left(\frac{jT}{n}\right)\right]^2.$$

- Recall the definition of quadratic variation, we have that

$$\sum_{j=0}^{n-1}\left[W\left(\frac{(j+1)T}{n}\right) - W\left(\frac{jT}{n}\right)\right]^2 \to T.$$

- So

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

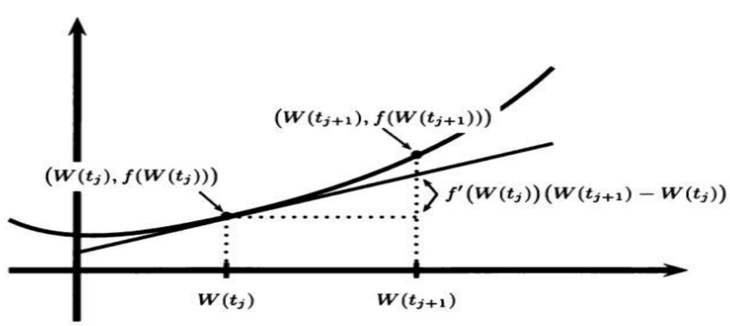
Itô Formula

- Ordinary integral: if $W(t)$ were differentiable, then we have the chain rule
$$df(W(t)) = f'(W(t))dW(t) = f'(W(t))W'(t)dt$$
(differential form) and
$$f(W(T))-f(W(0)) = \int_0^T f'(W(t))dW(t) = \int_0^T f'(W(t))W'(t)dt$$
(integral form).
- Itô integral: however, $W(t)$ is non-differentiable. Then
  - $df(W(t)) = f'(W(t))dW(t)$ is incorrect;
  - $f(W(T)) - f(W(0)) = \int_0^T f'(W(t))dW(t)$ is incorrect, either.
- Question: What is the counterpart of the chain rule for Itô integral?
- We seek to derive a corresponding integral form of $f(W(T)) - f(W(0)) =?$

Itô Formula: A Heuristic Derivation

Note that

$$\begin{aligned}
&f(W(T)) - f(W(0)) \\
=& \sum_{j=0}^{n-1} \left[f(W(t_{j+1})) - f(W(t_j))\right] \\
=& \sum_{j=0}^{n-1} f'(W(t_j))\left[W(t_{j+1}) - W(t_j)\right] + \sum_{j=0}^{n-1} \frac{1}{2} f''(W(t_j))\left[W(t_{j+1}) - W(t_j)\right]^2 \\
&+ \text{higher order smaller error.}
\end{aligned} \tag{13}$$



Taylor approximation to $f(W(t_{j+1})) - f(W(t_j))$.

Itô Formula: A Heuristic Derivation

- Roughly speaking, if $W(t)$ were differentiable, the second term of the RHS goes to 0 as $||\Pi||$ goes to 0;
- If $W(t)$ is non-differentiable, the second term of the RHS roughly goes to $\int_0^T \frac{1}{2} f''(W(t))dt$ as $||\Pi||$ goes to 0 due to finite quadractic variation.
- Higher order small errors vanish
- **Theorem** (An Easiest Version of Itô's Formula)

$$f(W(T)) - f(W(0)) = \int_0^T f'(W(t))dW(t) + \int_0^T \frac{1}{2} f''(W(t))dt.$$

- More general versions...

Itô Formula

- **Theorem (Itô Formula for Brownian Motion)** Let $f(t,x)$ be a function for which $f_t(t,x)$, $f_x(t,x)$, and $f_{xx}(t,x)$ are well defined and continuous. Then

$$\begin{aligned} f(T, W(T)) =& f(0, W(0)) + \int_0^T f_t(t, W(t))dt \\ &+ \int_0^T f_x(t, W(t))dW(t) + \frac{1}{2}\int_0^T f_{xx}(t, W(t))dt. \end{aligned} \tag{14}$$

- Differential form:

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

- **Example**: apply this theorem to $f(x) = \frac{x^2}{2}$.

Itô Formula for Itô Processes

- Motivation: we need more realistic models!
- **Definition:** An **Itô process** is a stochastic process of the form

$$X(t) = X(0) + \int_0^t \Delta(u)dW(u) + \int_0^t \Theta(u)du,$$

where $X(0)$ is non-random, $\Delta(u)$ and $\Theta(u)$ are adapted, and $E\int_0^t \Delta^2(u)du < +\infty$ and $\int_0^t |\Theta(u)|du < +\infty$ for any t.
- Differential form:

$$dX(t) = \Delta(t)dW(t) + \Theta(t)dt.$$

- **Proposition**: The quadratic variation of the Itô process $X(t)$ is

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

Itô Formula for Itô Processes

- **Definition: The integral of an adapted process** $\Gamma(t)$ w.r.t. an Itô process $X(t)$, with $E\int_0^t \Gamma^2(u)\Delta^2(u)du < +\infty$ and $\int_0^t |\Gamma(u)\Theta(u)|du < +\infty$ for any t,

$$\int_0^t \Gamma(u)dX(u) = \int_0^t \Gamma(u)\Delta(u)dW(u) + \int_0^t \Gamma(u)\Theta(u)du.$$

- Itô process is employed to describe a **gain process**!
- Differential form of the Itô Formula for an Itô process $X(t)$.
$df(t, X(t))$
$= f_t(t, X(t))dt + f_x(t, X(t))dX(t) + \frac{1}{2}f_{xx}(t, X(t))d[X, X](t)$
$= f_t(t, X(t))dt + f_x(t, X(t))\Delta(t)dW(t) + f_x(t, X(t))\Theta(t)dt + \frac{1}{2}f_{xx}(t, X(t))\Delta^2(t)dt$.

Generalization: Stochastic Integrals w.r.t. Continuous Local Martingales

Suppose M is a continuous local martingale. Similarly, we define

$$I^M(t) = \int_0^t \Delta(u) dM(u).$$

Under the condition

$$\mathbb{E} \int_0^t \Delta^2(u) d[M, M](u) < \infty;$$

we have:

- $\{I^M(t)\}$ is a true martingale.
- Quadratic variation:

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

Generalization: Stochastic Integrals w.r.t. Continuous Local Martingales

- Itô isometry:

$$\mathbb{E} I^2(t) = \mathbb{E} \left(\int_0^t \Delta^2(u) d[M, M](u) \right).$$

- Indeed, directly working on local martingales, we have the following more general version of the Itô formula. Let X be a semimartingale, i.e., a process of the form

$$X(t) = X(0) + M(t) + V(t),\ 0 \leq t < \infty$$

where M is a local martingale with continuous sample paths, and V a process with continuous sample paths of finite first variation. Then, for every function $f : \mathbb{R} \to \mathbb{R}$ of class C^2, we have

$$f(X(t)) = f(X(0)) + \int_0^t f'(X(s)) dM(s) + \int_0^t f'(X(s)) dV(s) + \frac{1}{2} \int_0^t f''(X(s)) d\langle M \rangle (s).$$

Applications of Itô Formula

Example: Generalized Geometric Brownian Motion for Modeling Stock Process

Consider $S(t) := S(0)e^{X(t)}$, where

$$dX(t) = \sigma(t)dW(t) + \left(\alpha(t) - \frac{1}{2}\sigma(t)^2\right)dt, \quad X(0) = 0$$

- $S(t)$ satisfies the following stochastic differential equation

$$dS(t) = \alpha(t)S(t)dt + \sigma(t)S(t)dW(t) \iff \frac{dS(t)}{S(t)} = \alpha(t)dt + \sigma(t)dW(t).$$

- Modeling issue: $\alpha(t)$ is the instantaneous mean rate of return, and $\sigma(t)$ is the volatility.
- When $\alpha = 0$, we get a martingale

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

- Generalization of the exponential martingale

Applications of Itô Formula: Itô integral of a Deterministic Integrand

- What is the distribution of $I(t) = \int_0^t \Delta(s)dW(s)$, where $\Delta(s)$ is non-random (deterministic!) function of time
- Claim: $I(t)$ is normally distributed $I(t) \sim N\left(0, \int_0^t \Delta(s)^2 ds\right)$.
- We just need to prove that

$$\mathbb{E}e^{uI(t)} = \exp\left\{\frac{1}{2}u^2\int_0^t \Delta(s)^2 ds\right\}, \text{ for all } u \in \mathbf{R}.$$

- Indeed, we observe a fact that the moment generating function of $I(t)$ satisfies

$$\exp\left\{\int_0^t u\Delta(s)dW(s) - \frac{1}{2}\int_0^t (u\Delta(s))^2 ds\right\}$$

is a martingale ($\Leftarrow$ generalized geometric Brownian motion with $\alpha = 0$ and $\sigma(s) = u\Delta(s)$)

Applications of Itô Formula: Characterizing a Brownian motion

- Recall that $W(t)$ satisfies the following three conditions
  - (1) a martingale with $M(0) = 0$;
  - (2) with continuous paths;
  - (3) with quadratic variation $[W, W](t) = t$.
- Surprisingly, conditions (1), (2) and (3) are sufficient to characterize a BM.

Theorem 4.6.4 (Lévy Theorem): Let $M(t)$ be a martingale relative to a filtration $\mathcal{F}(t)$, $t \geq 0$. 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 BM.
A rough proof: Consider a function $f(t, x)$ with partial derivatives f_t, f_x, and f_{xx} continuous. We use the following formula (an Ito formula with respect to martingales):

$$df(t, M(t)) = f_t(t, M(t))dt + f_x(t, M(t))dM(t) + \frac{1}{2} f_{xx}(t, M(t))[M, M](t).$$

Characterizing a Brownian motion

The integral form:
$f(t, M(t)) = f(0, M(0)) + \int_0^t \left[f_t(s, M(s)) + \frac{1}{2} f_{xx}(s, M(s))\right] ds + \int_0^t f_x(s, M(s))dM(s)$. Taking expectations leads to

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

Select $f(t, x) = e^{ux - \frac{1}{2}u^2 t}$. we can verify that

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

Therefore $Ee^{uM(t) - \frac{1}{2}u^2 t} = 1$, i.e., $Ee^{uM(t)} = e^{-\frac{1}{2}u^2 t}$. Thus, $M(t)$ has a normal distribution $N(0, t)$. □
Note: The Lévy Theorem can be extended to the multi-dimensional case.

Multivariable Stochastic Calculus

- Recall: A d-dimensional Brownian motion is a process $W(t) = (W_1(t), W_2(t), \cdots, W_d(t))$ such that
  - Each $W_i(t)$ is a one-dimensional BM;
  - $W_i(t)$ and $W_j(t)$ are independent for any $i \neq j$;
  - Independent increments.
- Some Properties:
  - $[W_i, W_i](t) = t$;
  - $[W_i, W_j](t) = 0$ if $i \neq j$, i.e.,

$$\lim_{||\Pi|| \to 0} E\left\{ \left(\sum_{k=0}^{n-1} [W_i(t_{k+1}) - W_i(t_k)][W_j(t_{k+1}) - W_j(t_k)] \right)^2 \right\} = 0.$$

Multivariable Stochastic Calculus

- Without loss of generality, consider a two-dimensional BM $(W_1(t), W_2(t))$.
- Consider two Itô processes
  $X(t) = X(0) + \int_0^t \Theta_1(u)du + \int_0^t \sigma_{11}(u)dW_1(u) + \int_0^t \sigma_{12}(u)dW_2(u)$
  $Y(t) = Y(0) + \int_0^t \Theta_2(u)du + \int_0^t \sigma_{21}(u)dW_1(u) + \int_0^t \sigma_{22}(u)dW_2(u)$
- The corresponding differential forms
  $dX(t) = \Theta_1(t)dt + \sigma_{11}(t)dW_1(t) + \sigma_{12}(t)dW_2(t)$
  $dY(t) = \Theta_2(t)dt + \sigma_{21}(t)dW_1(t) + \sigma_{22}(t)dW_2(t)$
- Quadratic and cross variations:

$$\begin{aligned} dX(t)dX(t) &= (\sigma_{11}^2(t) + \sigma_{12}^2(t))dt, \\ dY(t)dY(t) &= (\sigma_{21}^2(t) + \sigma_{22}^2(t))dt, \\ dX(t)dY(t) &= (\sigma_{11}(t)\sigma_{21}(t) + \sigma_{12}(t)\sigma_{22}(t))dt. \end{aligned}$$

Multivariable Stochastic Calculus

- **Theorem 4.6.2 (Two-dimensional Itô formula)** Let $f(t,x,y)$ be a function with partial derivatives f_t, f_x, f_y, f_{xx}, f_{xy}, and f_{yy} well defined and continuous. Consider two Itô processes $X(t)$ and $Y(t)$. Then we have

$$\begin{aligned} & df(t,X(t),Y(t)) \qquad (15) \\ =& f_t(t,X(t),Y(t))dt + f_x(t,X(t),Y(t))dX(t) + f_y(t,X(t),Y(t))dY(t) \\ &+ \frac{1}{2}f_{xx}(t,X(t),Y(t))dX(t)dX(t) + f_{xy}(t,X(t),Y(t))dX(t)dY(t) \\ &+ \frac{1}{2}f_{yy}(t,X(t),Y(t))dY(t)dY(t) \end{aligned}$$

- (**Itô product formula**).

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

Itô formula of any arbitrary dimension

Let
$$X(t) = (X_1(t), X_2(t), ..., X_n(t)).$$
Suppose that for $i = 1, 2, ..., n$

$$X_i(t) = X_i(0) + \int_0^t \Theta_i(u)du + \sum_{k=1}^{d}\int_0^t \sigma_{ik}(u)dW_k(u).$$

Following the similar manner, we obtain that

$$dX_i(t)dX_j(t) = d[X_i, X_j](t) = \sum_{k=1}^{d}\int_0^t \sigma_{ik}(u)\sigma_{jk}(u)du. \qquad (16)$$

Now, using (16) as input, we have $df(t, X(t)) =$

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

Generalization: Itô formula for Multidimensional Semimartingales

More generally, let $X = (X^{(1)}, ..., X^{(d)})$ be an $\mathbb{R}^d$- valued process with components

$$X_i(t) = X_i(0) + M_i(t) + V_i(t),$$

where M_i is a local martingale with continuous sample paths, and V_i a process with continuous sample paths of finite first variation for all $i = 1, 2, ..., d$, and $f : \mathbb{R}^d \to \mathbb{R}$ a function of class C^2. We have then

$$\begin{aligned} & f(X(t)) \\ = \; & f(X(0)) + \sum_{i=1}^{d} \int_0^t \frac{\partial f}{\partial x_i}(X(s))dM_i(s) + \sum_{i=1}^{d} \int_0^t \frac{\partial f}{\partial x_i}(X(s))dV_i(s) + \\ & + \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X(s))d\left\langle M_i, M_j \right\rangle(s). \end{aligned}$$

Example: Correlated Stock Prices

- Two asstes with price:

$$\begin{aligned} S_1(t) &= S_1(0) \exp\left\{\sigma_1 W_1(t) + (\alpha_1 - \frac{1}{2}\sigma_1^2)t\right\} \\ S_2(t) &= S_2(0) \exp\left\{\sigma_2[\rho W_1(t) + \sqrt{1-\rho^2} W_2(t)] + (\alpha_2 - \frac{1}{2}\sigma_2^2)t\right\} \end{aligned}$$

  where $W_1(t)$ and $W_2(t)$ are two independent Brownian motions.
- Use Ito's formula, we prove that

$$\begin{aligned} \frac{dS_1(t)}{S_1(t)} &= \alpha_1 dt + \sigma_1 dW_1(t), \\ \frac{dS_2(t)}{S_2(t)} &= \alpha_2 dt + \sigma_2[\rho dW_1(t) + \sqrt{1-\rho^2} dW_2(t)] \end{aligned}$$

Example: Correlated Stock Prices

- Denote $W_3(t) = \rho W_1(t) + \sqrt{1-\rho^2} W_2(t)$. Obviously, $W_3(t)$ is a standard Brownian motion.
- Apply Itô product formula to prove that Corr$(W_1(t), W_3(t)) = \rho$
- Thus, we may spell the joint dynamics as

$$\begin{aligned}\frac{dS_1(t)}{S_1(t)} &= \alpha_1 dt + \sigma_1 dW_1(t),\\ \frac{dS_2(t)}{S_2(t)} &= \alpha_2 dt + \sigma_2 dW_3(t)\end{aligned}$$

- The log-return satisfies that

$$\text{Corr}\left(\log\frac{S_1(t)}{S_1(0)}, \log\frac{S_2(t)}{S_2(0)}\right) = \rho$$

i.e. the correlation btw Brownian motions is exactly that for the returns.

Overview

- Stochastic Differential Equations (SDE)
- Examples in Financial Modeling

Stochastic Differential Equations (SDEs)

- **Definition:** A one-dimensional Stochastic Differential Equation (SDE) is an equation of the form

$$dX(t) = \beta(t, X(t))dt + \gamma(t, X(t))dW(t). \tag{17}$$

  - $\beta(t, x)$: drift;
  - $\gamma(t, x)$: diffusion;
  - $X(0) = x$ for $t \geq 0$ and $x \in \mathbb{R}$: the initial condition.

- Similarly define SDEs with multiple driving Brownian motions

- Similarly define multidimensional SDEs

Multidimensional SDEs

Similar to one-dimensional SDEs, we can define multidimensional SDEs:

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t) \tag{18}$$

where $X(t)$ is an m dimensional stochastic process

$$X(t) = (X_1(t), X_2(t), \cdots, X_m(t))^\top;$$

$\mu(t, x)$ is an m dimensional vector

$$\mu(t, x) = (\mu_1(t, x), \mu_2(t, x), \cdots, \mu_m(t, x))^\top;$$

$\sigma(t, x)$ is an $m \times n$ matrix

$$\sigma(t, x) = \begin{pmatrix} \sigma_{11}(t, x), \sigma_{12}(t, x), \ldots, \sigma_{1d}(t, x) \\ \sigma_{21}(t, x), \sigma_{22}(t, x), \ldots, \sigma_{2d}(t, x) \\ \cdots \\ \sigma_{m1}(t, x), \sigma_{m2}(t, x), \ldots, \sigma_{md}(t, x) \end{pmatrix};$$

$W(t)$ is a standard m dimensional Brownian motions with

$$W(t) = (W_1(t), W_2(t), \cdots, W_m(t))^\top.$$

Two types of solutions

- A strong solution is a process that solves the dynamic (22) on a given probability space (the driving Brownian motion is given as an input);
- A weak solution consists of a probability space and a process on it that solves the dynamic (22).
- strong solution $\Longrightarrow$ weak solution

Existences and Uniqueness of Strong Solution

Suppose that the coefficients of the equation satisfy the Lipschitz and linear growth conditions, i.e.,

$$||\mu(t,x)-\mu(t,y)||+||\sigma(t,x)-\sigma(t,y)|| \leq K||x-y||, \forall x,y \in R^d, \quad (19)$$

and

$$||b(t,x)||+||\sigma(t,x)|| \leq K(1+||x||), \forall x \in R^d, \quad (20)$$

for some real $K>0$. Then, there exists a unique process X that satisfies (18); it has continuous sample paths, is adapted to the filtration $\{\mathcal{F}_W(t)\}$ of the driving Brownian motion W, it is a Markov process. The idea in the proof of the aforementioned existence and uniqueness is to mimic the procedure followed in ordinary differential equations, i.e., to consider the “Picard iterations”

$$X^{(0)} \equiv \eta, X^{(k+1)}(t) = \eta + \int_0^t b(s, X^{(k)}(s))ds + \int_0^t \sigma(s, X^{(k)}(s))dW(s)$$

for $k = 0, 1, 2, ...$

About the Solution

The conditions (19) and (20) then guarantee that the sequence of continuous processes $\{X^{(k)}\}_{k=0}^{\infty}$ converges to a continuous process X, which is the unique solution of the equation (18); they also imply that the sequence $\{X^{(k)}\}_{k=0}^{\infty}$ and the solution X satisfy moment growth conditions of the type

$$E||X(t)||^{2\lambda} \leq C_{\lambda,T}(1 + E||\eta||)^{2\lambda}, \forall 0 \leq t \leq T$$

for any real numbers $\lambda \geq 1$ and $T > 0,$ where $C_{\lambda,T}$ is a positive constant depending only on λ, T and on the constant K of (19) and (20).

- Generally speaking, a SDE is not easy to solve, but sometimes we can solve it explicitly.
- Sometimes, numerical computing (e.g. Monte Carlo simulation) are necessary!

Linear SDEs

SDE:

$$dX(t) = \beta(t, X(t))dt + \gamma(t, X(t))dW(t). \tag{21}$$

with

- Drift $\beta(t, x) = a(t) + b(t)x$;
- Diffusion $\gamma(t, x) = \gamma(t) + \sigma(t)x$; condition.

e.g. One-dimensional linear SDEs:

$$dX(t) = [a(t) + b(t)X(t)]dt + [\gamma(t) + \sigma(t)X(t)]dW(t).$$

Examples in Financial Modeling: GBM

Example 1: (Generalized) Geometric Brownian Motion for Modeling Asset Price

- $S(t)$ satisfies SDE:

$$dS(t) = \alpha(t)S(t)dt + \sigma(t)S(t)dW(t), S(0) = s_0$$

- Modeling issue: $\alpha(t)$ is instantaneous mean rate of return, and $\sigma(t)$ is volatility.
- Both $\alpha(t)$ and $\sigma(t)$ could be very general adapted stochastic processes.
- If $\alpha(t)$ and $\sigma(t)$ are both constants $\Longrightarrow$ Black-Scholes-Merton model (1973)
- Explicit solution:

$$S(t) = s_0 e^{\int_0^t \sigma(u)dW(u) + \int_0^t \left(\alpha(u) - \frac{1}{2}\sigma^2(u)\right)du}.$$

Examples in Financial Modeling: the Vasicek Model

Example 2: Vasicek Model for Interest Rate

$$dR(t) = (\alpha - \beta R(t))dt + \sigma dW(t).$$

- When $\alpha = 0$, $R(t)$ is called an Ornstein-Uhlenbeck
- Equivalently written as

$$dR(t) = \kappa(\theta - R(t))dt + \sigma dW(t).$$

  process.
- κ: mean-reverting speed
- θ: mean-reverting level

How to solve is?

If RHS does not involve $R(t)$, the integral form of $R(t)$ is ready. So our objective is to remove $R(t)$ on the RHS.

Recall from Ordinary Differential Equation

Recall ODE

$$\frac{df(x)}{dx} = -af(x) + g(x),$$

where $g(x)$ is known. We have that

$$df(x) + af(x)dx = g(x)dx,$$

and

$$e^{ax}df(x) + ae^{ax}f(x)dx = e^{ax}g(x)dx,$$

i.e.,

$$d[e^{ax}f(x)] = e^{ax}g(x)dx.$$

Therefore

$$f(x) = e^{-ax}\left[f(0) + \int_0^x e^{as}g(s)ds\right].$$

Examples in Financial Modeling: the Vasicek Model

- Similarly, multiply

$$dR(t) = (\alpha - \beta R(t))dt + \sigma dW(t)$$

by $e^{\beta t}$. Then Itô lemma applies

$$d\left[e^{\beta t}R(t)\right] = e^{\beta t}dR(t) + \beta e^{\beta t}R(t)dt = e^{\beta t}\alpha dt + e^{\beta t}\sigma dW(t)$$

- Integrating both sides yields

$$e^{\beta t}R(t) = R(0) + \frac{\alpha}{\beta}\left(e^{\beta t} - 1\right) + \int_0^t \sigma e^{\beta s}dW(s).$$

- Namely, a closed-form expression for $R(t)$ is given by

$$R(t) = e^{-\beta t}R(0) + \frac{\alpha}{\beta}\left(1 - e^{-\beta t}\right) + \sigma e^{-\beta t}\int_0^t e^{\beta s}dW(s).$$

Several Modeling Issue

- Normal Distribution:

$$R(t) \sim N\left(e^{-\beta t}R(0) + \frac{\alpha}{\beta}\left(1 - e^{-\beta t}\right), \frac{\sigma^2}{2\beta}\left(1 - e^{-2\beta t}\right)\right).$$

- Disadvantage: possibility to be negative.
- Advantage: mean-reverting property.
  - β (speed of mean reversion);
  - $\lim_{t\to+\infty} ER(t) = \frac{\alpha}{\beta}$ (long-term mean level);
  - $\lim_{t\to+\infty} Var(R(t)) = \frac{\sigma^2}{2\beta}$ (long-term variance).

General Linear SDEs

Consider SDE

$$dX(t) = [a(t) + b(t)X(t)]dt + [\gamma(t) + \sigma(t)X(t)]dW(t), X(0) = X_0.$$

Apply Ito's rule to prove that

$$X(t) = Y(t)\left[X_0 + \int_0^t (a(s) - \gamma(s)\sigma(s))Y(s)^{-1}ds + \int_0^t \gamma(s)Y(s)^{-1}dW(s)\right],$$

where

$$Y(t) = \exp\left\{\int_0^t \left(b(s) - \frac{1}{2}\sigma(s)^2\right)ds + \int_0^t \sigma(s)dW(s)\right\}.$$

Question: How to find the expectation and variance of $X(t)$?

Note: Previous examples are both special cases of linear SDEs.

Examples in Financial Modeling: the CIR Model

Example 3: Cox-Ingersoll-Ross (CIR) Model for Interest Rate

$$dR(t) = (\alpha - \beta R(t))dt + \sigma\sqrt{R(t)}dW(t).$$

- So the advantage of CIR over Vasicek is its non-negativity.
- Widely used in modeling interest rate, stochastic volatility, stochastic intensity of credit default and other jumps.
- We cannot derive a closed form formula for $R(t)$.
- However, we know $R(t)$ assumes a noncentral Chi-square distribution.
- Exercise: Compute $\mathbb{E}(R(t))$ and Var$(R(t))$ via Itô formula.

Examples in Financial Modeling: Multidimensional Geometric Brownian Motion

Example 4: Multidimensional Geometric Brownian Motion Model for Multiple Correlated Asset Prices, e.g., for two correlated assets

$$\begin{aligned}\frac{dS_1(t)}{S_1(t)} &= \alpha_1 dt + \sigma_1 dW_1(t),\\ \frac{dS_2(t)}{S_2(t)} &= \alpha_2 dt + \sigma_2[\rho dW_1(t) + \sqrt{1-\rho^2}dW_2(t)],\end{aligned}$$

where $\{(W_1(t), W_2(t))\}$ is a standard two-dimensional Brownian motion.

Equivalent dynamics:

$$\begin{aligned}\frac{dS_1(t)}{S_1(t)} &= \alpha_1 dt + \sigma_1 dW_1(t),\\ \frac{dS_2(t)}{S_2(t)} &= \alpha_2 dt + \sigma_2 dW_3(t).\end{aligned}$$

Example: Correlated Assets

Here $\{(W_1(t), W_3(t))\}$ is a two dimensional Brownian motion with Corr$(W_1(t), W_3(t)) = \rho$.

Apply Ito's formula to $\log S_1(t)$ and $\log S_2(t)$, we find that

$$\begin{aligned} S_1(t) &= S_1(0)\exp\left\{\sigma_1 W_1(t) + \left(\alpha_1 - \frac{1}{2}\sigma_1^2\right)t\right\} \\ S_2(t) &= S_2(0)\exp\left\{\sigma_2[\rho W_1(t) + \sqrt{1-\rho^2}W_2(t)] + \left(\alpha_1 - \frac{1}{2}\sigma_1^2\right)t\right\} \end{aligned}$$

Generalization: multidimensional linear SDEs. Even in linear specifications, not all SDEs are explicitly solvable! This is not as simple as the one-dimensional linear SDEs.

More Examples

SDE provides us a powerful tool to describe the dynamics of financial market. For example, in order to incorporate the "volatility smile", a natural idea is to allow the change of volatility.

- Local volatility models (Dupire, Derman):

$$dS(t) = \mu S(t)dt + \sigma(t, S(t))S(t)dW(t)$$

- The stochastic volatility model (e.g. Heston (1993)):

$$dS(t) = \mu S(t)dt + \sqrt{V(t)}S(t)dW_1(t)$$

$$dV(t) = \kappa(\theta - V(t))dt + \sigma_v\sqrt{V(t)}dW_2(t).$$

- In practice, we may use more advanced models according to the special necessity, e.g. adding jumps, etc.

Overview

- Connect btw SDEs and PDEs: Feynman-Kac Theorem

Connect btw Brownian Motion and Heat Equation

Recall that the heat equation initial value problem

$$\begin{aligned} \frac{\partial u}{\partial t} - \frac{1}{2}\frac{\partial^2 u}{\partial x^2} &= 0, \\ u(0,x) &= f(x), \end{aligned}$$

admits solution

$$u(t,x) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{+\infty} e^{-\frac{(x-y)^2}{2t}} f(y) dy.$$

Obviously, we can express this solution using a standard Brownian motion, i.e.

$$u(t,x) = \mathbb{E} f(W(t) + x),$$

where $\{W(t)\}$ is a standard Brownian motion.

A Heuristic Verification

An alternative heuristic verification that $\mathbb{E}f(W(t)+x)$ solves the heat equation:

The initial condition obviously holds! By Taylor expansion

$$f(b) = f(a) + f'(a)(b-a) + \frac{1}{2}f''(a)(b-a)^2 + o((b-a)^2).$$

Therefore,

$$\begin{aligned}
\frac{\partial u}{\partial t}(t,x) &= \lim_{\Delta t\to 0} \frac{u(t+\Delta t,x)-u(t,x)}{\Delta t} \\
&= \lim_{\Delta t\to 0} \frac{\mathbb{E}f(W(t+\Delta t)+x)-\mathbb{E}f(W(t)+x)}{\Delta t} \\
&= \lim_{\Delta t\to 0} \frac{1}{\Delta t}\mathbb{E}[f'(W(t)+x)(W(t+\Delta t)-W(t)) + \frac{1}{2}f''(W(t)+x) \\
&\quad (W(t+\Delta t)-W(t))^2 + o((W(t+\Delta t)-W(t))^2)] \\
&= \lim_{\Delta t\to 0} \frac{1}{\Delta t}\left[0 + \frac{1}{2}\Delta t u''(t,x) + o(\Delta t)\right] \\
&= \lim_{\Delta t\to 0} 0 + \frac{1}{2}u''(t,x) + o(1) = \frac{1}{2}u''(t,x).
\end{aligned}$$

A Heuristic Verification

Indeed, we have applied

$$\begin{aligned}
& \mathbb{E}f'(W(t)+x)(W(t+\Delta t)-W(t)) \\
=\ & \mathbb{E}\left[\mathbb{E}\left[f'(W(t)+x)(W(t+\Delta t)-W(t))|\mathcal{F}_t\right]\right] \\
=\ & \mathbb{E}\left[f'(W(t)+x)\mathbb{E}\left[W(t+\Delta t)-W(t)|\mathcal{F}_t\right]\right] = 0
\end{aligned}$$

$$\begin{aligned}
& \mathbb{E}f''(W(t)+x)(W(t+\Delta t)-W(t))^2 \\
=\ & \mathbb{E}\left[\mathbb{E}\left[f''(W(t)+x)(W(t+\Delta t)-W(t))^2|\mathcal{F}_t\right]\right] \\
=\ & \mathbb{E}\left[f''(W(t)+x)\mathbb{E}\left[(W(t+\Delta t)-W(t))^2|\mathcal{F}_t\right]\right] \\
=\ & \Delta t\mathbb{E}f''(W(t)+x) = \Delta t u''(t,x)
\end{aligned}$$

Multidimensional Extension

For $x = (x_1, x_2, \cdots, x_d) \in \mathbf{R}^d$ and a known function $g : \mathbf{R}^d \to \mathbf{R}$, an unknown function $u(t,x)$ satisfies that a $d-$dimensional heat equation:

$$\begin{aligned} \frac{\partial u}{\partial t} - \frac{1}{2}\Delta u &= 0, \\ u(0,x) &= g(x), \end{aligned}$$

where

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \cdots + \frac{\partial^2}{\partial x_d^2}.$$

Amazingly, we have

$$u(t,x) = \mathbb{E}g(W(t) + x),$$

where $\{W(t)\}$ is a standard $d-$dimensional Brownian motion.

Connect btw Brownian Motion and Backward Heat Equation

Now, let $v(t,x) = u(T-t,x)$. Calculus yields a Backward heat equation:

$$\begin{aligned} \frac{\partial v}{\partial t} + \frac{1}{2}\frac{\partial^2 v}{\partial x^2} &= 0, \\ v(T,x) &= f(x), \end{aligned}$$

Note that,

$$v(t,x) = \mathbb{E}[f(B(T))|B(t) = x],$$

where $\{B(t)\}$ is a Brownian motion, solves the this equation!

Now, using the fact that $v(t, B(t)) = \mathbb{E}[f(B(T))|B(t)]$ is a martingale (why?), we can give a probabilistic proof! Later we will see something more general!

Connect btw Stochastic Processes and PDEs: Feynman-Kac Theorem

Question: Can we generalize the previous result on the connection btw Brownian motion and heat equation?

- Consider an SDE:

$$dX(u) = \beta(u, X(u))du + \gamma(u, X(u))dW(u). \tag{22}$$

- We assume the existence and uniqueness of its solution
- Can we employ this SDE to express the solution to some certain PDEs as conditional expectation?

Properties of Strong Solutions

- Consider a strong solution of (22) $X(t)$ and a function $h(y)$. Define

$$g(t, x) := E^{t,x}h(X(T)) \equiv E[h(X(T))|X(t) = x] \tag{23}$$

- By the Markov property (let us believe it) of $\{X(t)\}$

$$E[h(X(T))|\mathcal{F}(t)] \equiv E[h(X(T))|X(t)]. \tag{24}$$

- Note that

$$g(t, X(t)) = E^{t,X(t)}h(X(T)) \equiv E[h(X(T))|X(t)].$$

This indicates that $g(t, X(t))$ is a martingale (Levy martingale).

Feynman-Kac Theorem

- **Feynman-Kac Theorem:** Consider the SDE (22), its strong solution $X(t)$, a function $h(y)$, and

$$g(t,x) := E^{t,x}h(X(T))(< +\infty)$$

given by (23). Then $g(t,x)$ satisfies the PDE

$$g_t(t,x) + \beta(t,x)g_x(t,x) + \frac{1}{2}\gamma^2(t,x)g_{xx}(t,x) = 0 \tag{25}$$

and the terminal condition

$$g(T,x) = h(x) \quad \text{for any } x. \tag{26}$$

- **Remarks**:
    - The PDE (25) does not involve $h(\cdot)$.
    - $h(\cdot)$ is only involved in the terminal condition (26).

Feynman-Kac Theorem

- **Proof:** Applying Itô lemma to the process $g(t, X(t))$ and omitting the argument $(t, X(t))$ yield

$$\begin{aligned} dg(t,X(t)) =& g_t dt + g_x dX + \frac{1}{2}g_{xx}dXdX \\ =& \left[g_t + \beta g_x + \frac{1}{2}\gamma^2 g_{xx}\right]dt + \gamma g_x dW. \end{aligned} \tag{27}$$

- Since $g(t, X(t))$ is a martingale, there is no dt term in (27), as results in the PDE (25).
- The Key point to derive a PDE is
    - (1) construct a martingale involving a Markov process $X(t)$ that solves a SDE;
    - (2) apply Itô lemma;
    - (3) Set dt term to be 0.

Feynman-Kac Theorem: A discounted version

- Consider

$$E\left[e^{-r(T-t)}h(X(T))|\mathcal{F}(t)\right] =: f(t, X(t)).$$

- Question: Is there any PDE that $f(t, x)$ solves?
- First, $f(t, X(t))$ is not a martingale because

$$\begin{aligned} E[f(t, X(t))|\mathcal{F}(s)] =& E[E[e^{-r(T-t)}h(X(T))|\mathcal{F}(t)]|\mathcal{F}(s)] \\ =& E[e^{-r(T-t)}h(X(T))|\mathcal{F}(s)], \end{aligned} \tag{28}$$

  where the RHS depends on t.
- However, $e^{-rt}f(t, X(t))$ is a martingale.
- Apply Itô lemma to $e^{-rt}f(t, X(t))$ yields

$$d(e^{-rt}f(t, X(t))) = e^{-rt}\left[-rf + f_t + \beta f_x + \frac{1}{2}\gamma^2 f_{xx}\right]dt + e^{-rt}\gamma f_x dW(t) \tag{29}$$

Feynman-Kac Theorem

- Apply Itô lemma to $e^{-rt}f(t, X(t))$ yields

$$d(e^{-rt}f(t, X(t))) = e^{-rt}\left[-rf + f_t + \beta f_x + \frac{1}{2}\gamma^2 f_{xx}\right]dt + e^{-rt}\gamma f_x dW(t) \tag{30}$$

- Setting dt term to be zero leads to a PDE

$$f_t(t, x) + \beta(t, x)f_x(t, x) + \frac{1}{2}\gamma^2(t, x)f_{xx}(t, x) = rf(t, x) \tag{31}$$

  and the terminal condition

$$f(T, x) = h(x) \quad \text{for any } x. \tag{32}$$

- **Remarks:**
  - The PDE (31) does not depend on $h(\cdot)$ and solely depends on $X(t)$), the Markov process that the payoff relies on.
  - $h(\cdot)$ only affects the terminal condition (32).

Towards a General Case

Suppose we apply the Itô formula on the marginal of an SDE. How to use a very compactly written generator to express the Itô formula? Suppose we have an m dimensional SDE for $\{X(t)\}$ driven by a d dimensional Brownian motion $\{W(t)\}$

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t). \tag{33}$$

Indeed, by the previous discussions, we note that

$$d[X_i, X_j](t) = \sum_{k=1}^{d} \sigma_{ik}(X(t))\sigma_{jk}(X(t))dt = a_{ij}(X(t))dt,$$

where the matrix

$$(a_{ij}(t,x))_{m\times m} = \sigma(t,x)\sigma(t,x)^\top$$

is usually called diffusion matrix.

Infinitesimal Generator

Using the multidimensional Ito formula, we have

$$\begin{aligned} & du(t, X(t)) \\ = \; & \left[u_t(t, X(t)) + \mathcal{A}_t u(t, X(t))\right] dt + \nabla u(t, X(t))\sigma(t, X(t))dW(t), \end{aligned}$$

where the generator $\mathcal{A}_t$ is a second-order differential operator given by

$$\mathcal{A}_t = \frac{1}{2}\sum_{i,j=1}^{m} a_{ij}(t,x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{m} \mu_i(t,x)\frac{\partial}{\partial x_i}. \tag{34}$$

We call it the infinitesimal generator for SDE (33). The terminology "infinitesimal generator" is given because of the following properties, which can be proved as an excellent exercise under some technical conditions (e.g., μ and a are bounded and continuous)

Infinitesimal Generator

$$\lim_{h \to 0} \frac{1}{h} \mathbb{E} \left[X_i(t+h) - x_i | X(t) = x \right] = \mu_i \left(t, x \right)$$

and

$$\lim_{h \to 0} \frac{1}{h} \mathbb{E} \left[\left(X_i(t+h) - x_i \right) \left(X_k(t+h) - x_k \right) | X(t) = x \right] = a_{ik} \left(t, x \right)$$

for $1 \leq i, k \leq d$ hold for every $x \in \mathbb{R}^m$, and, more generally,

$$\lim_{h \downarrow 0} \frac{1}{h} \left[\mathbb{E} f \left(X(t+h) \right) - f \left(x \right) | X(t) = x \right] = \left(\mathcal{A}_t f \right) (x) \, ; \; \forall x \in \mathbb{R}^m$$

hold for every $f \in C^2 \left(\mathbb{R}^m \right)$ which is bounded and has bounded first- and second-order derivatives where the operator $\mathcal{A}_t f$ in (34).

Kolmogorov equations for transition density

- Suppoe that

$$\mathbb{P}(X(t) \in A | X(s) = y] = \int_A p(t, x; s, y) dx.$$

  Thus, $p(t, x; s, y)$ is the transition density of the process $\{X(t)\}$.
- We have the initial conditions as

$$\lim_{t \to s} p(t, x; s, y) = \delta(x - y),$$

  where $\delta(\cdot)$ is the celebrated Dirac Delta function.
- We introduce a differential operator $\mathcal{A}_t^*$, which is the adjoint of the operator of (34), namely

$$A_t^* f(x) \triangleq \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} [a_{ij}(t, x) f(x)] - \sum_{i=1}^{d} \frac{\partial}{\partial x_i} [b_i(t, x) f(x)].$$

Kolmogorov equations for transition density

Theorem

Under appropriate technical conditions, the transition density $p(t, x; s, y)$ satisfies the following two PDEs: the Komogorov backward equation

$$\left(\frac{\partial}{\partial s} + \mathcal{A}_s\right) p(t, x; \cdot, \cdot) = 0;$$

and the Komogorov forward equation (also known as Fokker-Planck equation)

$$\left(\frac{\partial}{\partial t} - \mathcal{A}_t^*\right) p(\cdot, \cdot; s, y) = 0.$$

Feynman-Kac theorem: a general version

Theorem

Assume conditions (19) and (20) hold. Let $f : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ and polynomial growth condition

$$\max_{0 \le t \le T} |f(t, x)| + |g(x)| \le C(1 + ||x||^p), \ \forall x \in R^d \tag{35}$$

for some $C > 0, p \ge 1,$ let $k : [0, T] \times \mathbf{R}^d \to [0, \infty)$ be continuous, and suppose that the Cauchy problem

$$\begin{aligned} \frac{\partial V}{\partial t} + \mathcal{A}_t V + f &= kV, \ \textit{in } [0, T) \times \mathbb{R}^d \\ V(T, \cdot) &= g, \ \textit{in } \mathbb{R}^d \end{aligned}$$

has a solution $V : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ which is continuous on its domain, of class $C^{1,2}$ on $[0, T] \times \mathbb{R}^d$, and satisfies a growth condition of the

Feynman-Kac theorem: a general version

Theorem

type (20). The function V admits then the Feynman-Kac representation

$$V(t,x) = E\left[\int_t^T e^{-\int_t^\theta k(u,X(u))du} f(\theta, X(\theta))d\theta + g(X(T))e^{-\int_t^T k(u,X(u))du}\right]$$

for $0 \le t \le T, x \in R^d$ in terms of the solution X of the stochastic integral equation

$$X(\theta) = x + \int_t^\theta \mu(s, X(s))ds + \int_t^\theta \sigma(s, X(s))dW(s), t \le \theta \le T.$$

Sketch of Proof

The proof of this result relies on the growth condition (35) the following Lemma. If the process X satisfies the equation (33). and $\beta(t) \triangleq \exp(-\int_0^t K(u)du)$ for some measurable, adapted and nonnegative process K; then, the process

$$M^f(t) \triangleq \beta(t)f(t, X(t)) - f(0, X(0)) - \int_0^t \left(\frac{\partial f}{\partial s} + \mathcal{A}_s f - K(s)f\right)(s, X(s))\beta(s)ds,$$

is a local martingale (square-integrable martingale, if f is of compact support) with continuous sample paths, and can be represented actually as

$$\sum_{i=1}^d \sum_{k=1}^n \int_0^t \frac{\partial f(s, X(s))}{\partial x_i}\sigma_{ik}(s, X(s))\beta(s)dW^{(k)}(s).$$

Overview

- The fundamental theorems of stochastic analysis

Levy's characterization of Brownian motions

Theorem

(Levy) Suppose $\{M(t) = (M_1(t), M_2(t), ..., M_d(t))\}$ is a d-dimensional martingale with continuous sample paths adapted to the filtration $\{\mathcal{F}(t)\}$. Assume that $M(0) = 0$ without loss of generality. If the cross variation satisfies the following property

$$[M_i, M_j](t) = \delta_{ij} t$$

with $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise, $\{M(t)\}$ is a standard d-dimensional Brownian motion.

Levy's characterization of Brownian motions

To characterize the distribution, it is sufficies to compute the conditional moment generating function of the increment $M(t) - M(s)$ for $0 \leq s < t$,

$$\mathbb{E}[\exp\left(u(M(t) - M(s))\right) | \mathcal{F}(s)] = \exp\left(\frac{1}{2}\sum_{i=1}^{d} u_i^2 (t-s)\right).$$

For this purpose, we use the Itô formula to find the differential of $\exp\left(u(M(t) - M(s))\right)$ and then make full use of the given condition in the theorem and the martingale property. We suggest this proof as an excellent exercise.

Continuous local martingale as time-changed Brownian motion

Our first result states that "every local martingale with continuous sample paths, is nothing but a Brownian motion, run under a different clock."

Theorem

(Dambis (1965), Dubins and Schwartz (1965)) Suppose $\{M(t)\}$ is a continuous local martingale safisfying $\lim_{t\to\infty} \langle M \rangle (t) = \infty$. There exists then a Brownian motion W, such that:

$$M(t) = W(\langle M \rangle (t)); 0 \leq t < \infty.$$

Continuous local martingale as time-changed Brownian motion

Following the similar spirit, a multidimensional generalization of this theorem can be established as follows.

Theorem

(Knight (1971)) Suppose $\{M(t) = (M_1(t), M_2(t), ..., M_d(t))\}$ is a d-dimensional continuous local martingale safisfying $\lim_{t\to\infty} \langle M_i \rangle (t) = \infty$ for each $i = 1, 2, ..., d$. There exists then a standard d-dimensional Brownian motion W, such that:

$$M_i(t) = W_i(\langle M_i \rangle (t)); 0 \le t < \infty.$$

Sketch of proof to the one-dimensional case

Without loss of essence, we consider the case when $\langle M \rangle$ is strictly increasing. In this case $\langle M \rangle$ has an inverse, say T, which is continuous (as well as strictly increasing). Then it is not hard to see that the process

$$W(s) = M(T(s)),\ 0 \le s < \infty \tag{36}$$

is a local martingale (with respect to the filtration $\{\mathcal{G}(s) = \mathcal{F}(T(s))\}$) with continuous sample paths, as being the composition of the two continuous mappings $T : \mathbb{R}^+ \to \mathbb{R}^+$ and $M : \mathbb{R}^+ \to \mathbb{R}$. On the other hand, by the definition of quadratic variation, it is that

$$\langle W \rangle (s) = \langle M \rangle (T(s)) = s.$$

Thus, by the Levy characterization of Brownian motion, W is a Brownian motion. Furthermore, replacing s by $\langle M \rangle_t$ in the above equation, we obtain

$$W(\langle M \rangle_t) = M(T(\langle M \rangle (t))) \equiv M(t).$$

Theorem

Suppse there is a random variable $F \in L^2$. W is a d-dimensional Brownian motion. If $F \in \mathcal{F}_W(T)$, where $\{\mathcal{F}_W(t)\}$ is the Brownian filtration, we have

$$F = \mathbb{E}[F] + \int_0^T \phi(s)^\top dW(s),$$

where $\{\phi(s)\}$ is a d-dimensional vector process. Denote by

$$M(s) = \mathbb{E}(F|\mathcal{F}_W(t)).$$

Then, we have

$$M(t) = M(0) + \int_0^t \phi(s) dW(s). \tag{37}$$

- The "integrand" ϕ is unique in this representation.
- The differential form of (37) can be written as $dM(t) = \phi(t)dW(t)$.
- Intuitively speaking, any martingale can be represented by increments of Brownian motions.
- A financial interpretation of this theorem can be given as follows. Assume stock prices are Brownian motions. If ϕ is the portfolio, $\phi(t)dW(t)$ is the local return. And thus the integral $\int_0^t \phi(s)dW(s)$ becomes the aggregate return of the portfolio. If one trades Brownian motion $\{W(t)\}$, given any contingent claim with price process $\{M(t)\}$, such contingent claim can be replicated by trading $\{W(t)\}$ according to the portfolio strategy $\{\phi(t)\}$.
- This theorem is about existence. The question is how to find ϕ explicitly? One needs theory of Malliavin calculus.

Change-of-Measure: the Girsanov Theorem

A simple exampls:

How can we use change-of-measure to move the mean of a normal random variable?

- X is a standard normal random variable on a probability space $(\Omega, \mathcal{F}, P)$, θ is a constant.
- Define

$$Z = \exp\left(-\theta X - \frac{1}{2}\theta^2\right) \text{ and } \frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} = Z$$

- Under the probability measure $\widetilde{\mathbb{P}}$, the random variable $Y = X + \theta$ is a standard normal.
- In particular, $\mathbb{E}^{\widetilde{\mathbb{P}}}Y = 0$, whereas $\mathbb{E}^{\mathbb{P}}Y = \mathbb{E}^{\mathbb{P}}X + \theta = \theta$.

Preparation: on a Generalized Local Martingale

Under the probability space $(\Omega, \mathbb{P}, \mathcal{F})$ equipped with the filtration $\{\mathcal{F}(t)\}$, suppose $\{W(t)\}$ is a d-dimensional Brownian motion and $\{\theta(t)\}$ is a stochastic process adapted to the Brownian filtration. Consider

$$\eta(t) := \exp\left(-\sum_{i=1}^{d}\int_0^t \theta_i(s)dW_i(s) - \frac{1}{2}\sum_{i=1}^{d}\int_0^t \theta_i(s)^2 ds\right).$$

By using the Itô formula, it is easy to have (verify it as an excellent exercise)

$$d\eta(t) = -\sum_{i=1}^{d}\theta_i(t)\eta(t)dW_i(t).$$

Thus, $\eta(t)$ can be written as a stochastic integral

$$\eta(t) = -\sum_{i=1}^{d}\int_0^t \theta_i(s)\eta(s)dW_i(s). \tag{38}$$

Local martingale vs true martingale

According to the theory of stochastic integral, without any conditions, $\{\eta(t)\}$ is merely a local martingale, which need not to be a true martingale and is a martingale evaluated at bounded stopping times. Under the following conditions, we get a true martingale:

- a commonly used condition ensuring the stochastic integral defined through (38) is a martingale:

$$\mathbb{E}\left(\int_0^t \theta_i(s)^2\eta(s)^2 ds\right) < \infty.$$

- the Novikov Condition:

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\sum_{i=1}^{d}\int_0^t \theta_i(s)^2 ds\right)\right] < \infty.$$

Girsanov Theorem

Theorem

(Girsanov) When $\{\eta(t)\}$ is a true martingale, we can construct a new probability measure $\mathbb{Q}$ such that

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \eta(T) \Leftrightarrow d\mathbb{Q} = \eta(T)d\mathbb{P},$$

i.e.

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

for any $A \in \Omega$. The probability measures $\mathbb{P}$ and $\mathbb{Q}$ are equivalent. If we definte

$$W^{\mathbb{Q}}(t) = W(t) + \int_0^t \theta(s)ds,$$

$W^{\mathbb{Q}}$ is a standard Brownian motion under the probability measure $\mathbb{Q}$.

A Simple But Useful Case

- Obviously, $W^{\mathbb{Q}}$ is not a standard Brownian motion but a Browian motion with stochastic dirft $\int_0^t \theta(s)ds$ under $\mathbb{P}$.
- A frequenctly used baby version of this theorem is the case when $\theta(t) \equiv \theta$. In this case, we have

$$\eta(t) = \exp\left(-\theta W(t) - \frac{1}{2}\theta^2 t\right),$$

which is exactly the well-known exponential martingale.
- By the Girsanov theorem, under the probability measure $\mathbb{Q}$, $W^{\mathbb{Q}}(t) = W(t) + \theta t$ is a standard Brownian motion.
- However, it is a Brownian motion with drift under the probability measure $\mathbb{P}$.

Sketch of the Proof

To prove the Girsanov theorem, we begin by verifying that $\mathbb{Q}$ is indeed a probability measure. For this purpose, we have

$$\mathbb{Q}(\Omega) = \int_\Omega \eta(T)d\mathbb{P} = \mathbb{E}^{\mathbb{P}}\eta(T) = 1;$$

and, for any two disjoint A and B from the sigma algebra $\mathcal{F}(T)$; we have

$$\mathbb{Q}(A \cap B) = \int_{A\cap B} \eta(T)d\mathbb{P} = \int_A \eta(T)d\mathbb{P} + \int_B \eta(T)d\mathbb{P} = \mathbb{Q}(A) + \mathbb{Q}(B).$$

It's easy to check that the two probability measures $\mathbb{P}$ and $\mathbb{Q}$ are equivalent.

Sketch of the Proof: the Bayes Rule

Theorem

(Bayesian Rule) Suppose Y is a $\mathcal{F}(T)$-measurable random variable satisfying $\mathbb{E}^{\mathbb{P}}|Y| < \infty$ and $\mathbb{E}^{\mathbb{P}}|\eta(t)Y| < \infty$ for all $t \in [0, T]$. Then, we have

$$\mathbb{E}^{\mathbb{Q}} Y = \mathbb{E}^{\mathbb{P}} \left(\eta(T)Y\right),$$

and, more generally,

$$\mathbb{E}^{\mathbb{Q}}[Y|\mathcal{F}(t)] = \mathbb{E}^{\mathbb{P}}\left[\frac{\eta(T)}{\eta(t)}Y|\mathcal{F}(t)\right] = \mathbb{E}^{\mathbb{P}}[\eta(t,T)Y|\mathcal{F}(t)], \tag{39}$$

where

$$\eta(t,T) := \frac{\eta(T)}{\eta(t)}.$$

Obviously, $\eta(t,T)$ plays the role as a normalized version of the Radon-Nikodym derivative for changing measure. Relation (39) reveals how change of measure can be done under conditioning.

Sketch of the Proof: the Main Idea

The main idea for proving the Girsanov theorem is to apply the Levy characterization of Brownian motion.

- Quadratic variation:

$$[W_i^{\mathbb{Q}}, W_j^{\mathbb{Q}}](t) = \left[W_i(\cdot) + \int_0^{\cdot} \theta_i(s)ds, W_j(\cdot) + \int_0^{\cdot} \theta_j(s)ds\right](t) = \delta_{ij}t.$$

- By using the Itô formula, it is straight forward to have (check it as an excellent exercise!)

$$d(W_i^{\mathbb{Q}}(t)\eta(t)) = [-W_i^{\mathbb{Q}}(t)\theta(t) + 1]\eta(t)dW_i(t), \text{ for } i = 1, 2, ..., d,$$

  which leads to the local martingale property of $\{W_i^{\mathbb{Q}}(t)\eta(t)\}$.
- Use the Bayes rule to change measure and then obtain the local martingale property of $\{W^{\mathbb{Q}}(t)\}$.

Supplementary Notes on "An Introduction to Stochastic Analysis"

Chenxu Li *
Department of Business Statistics
and Econometrics
Guanghua School of Management
Peking University

Wednesday 8th February, 2017

Abstract

In this set of notes, we provide some technical details supplementary to our lecture on "Foundation of Stochastic Analysis." As this is a very preliminary version, please keep them for your own use and don't circulate them. We will try to write them into a text book. I appreciate any reports of errors and typos. With this set of notes, you don't need worry about the problem of viewing clustered formula derivations the board from a distant point in our big classroom. However, please note that it is not an exhaustive list of our in-class discussion. A combination of course slides, these supplementary notes, homework problem sets and your notes taken from our in-class discussions would be quite helpful to your study.

*Email: cxli@gsm.pku.edu.cn

Contents

1 Conditional expectations

1.1 Verifications of the conditional-expectation properties

In this section, we provide some verifications of the conditional-expectation properties, based on the rigorous definition.

We have

$$\mathbb{E}[\mathbb{E}(X|\mathcal{G})] = \mathbb{E}X$$

According to the definition of conditional expectation, we know that

$$\mathbb{E}[\mathbb{E}(X|\mathcal{G})1_A] = \mathbb{E}(X1_A), \tag{1.1}$$

for any $A \in \mathcal{G}$. Taking $A = \Omega$, we have

$$\mathbb{E}[\mathbb{E}(X|\mathcal{G})] \equiv \mathbb{E}[\mathbb{E}(X|\mathcal{G})1_\Omega] = \mathbb{E}(X1_\Omega) \equiv \mathbb{E}(X).$$

To prove the other properties, we may need the following fact. Building our way up, from indicators to simple functions to bounded measurable functions, we can show that (1.1) is indeed equivalent to

$$\mathbb{E}[\mathbb{E}(X|\mathcal{G})W] = \mathbb{E}(XW), \tag{1.2}$$

for any $\mathcal{G}$-measurable random variable W.

We prove the "taking-out what is known":

$$\mathbb{E}(XY|\mathcal{G}) = X\mathbb{E}(Y|\mathcal{G}),$$

if X is $\mathcal{G}$-measurable. It is enought to prove

$$\mathbb{E}[X\mathbb{E}(Y|\mathcal{G})1_A] = \mathbb{E}(XY1_A),$$

for any $A \in \mathcal{G}$. This is equivalent to

$$\mathbb{E}[\mathbb{E}(Y|\mathcal{G})X1_A] = \mathbb{E}(YX1_A). \tag{1.3}$$

Obviously, $X1_A$ is a $\mathcal{G}$-measurable random variables. Thus, (1.3) holds according to (1.2).

We prove the "tower property":

$$\mathbb{E}[\mathbb{E}(X|\mathcal{G})|\mathcal{H}] = \mathbb{E}(X|\mathcal{H}).$$

According to the definition, it is enough to show that

$$\mathbb{E}[\mathbb{E}[\mathbb{E}(X|\mathcal{G})|\mathcal{H}]1_A] = \mathbb{E}(X1_A), \tag{1.4}$$

for any $A \in \mathcal{H}$. Indeed, we know from the definition of $\mathbb{E}[\mathbb{E}(X|\mathcal{G})|\mathcal{H}]$ that

$$\mathbb{E}[\mathbb{E}[\mathbb{E}(X|\mathcal{G})|\mathcal{H}]1_A] = \mathbb{E}(\mathbb{E}(X|\mathcal{G})1_A)$$

for any $A \in \mathcal{H}$. Since $A \in \mathcal{H}$ implies $A \in \mathcal{G}$, the definition of $\mathbb{E}(X|\mathcal{G})$ imlies that

$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})1_A) = \mathbb{E}(X1_A).$$

Therefore, we obtained (1.4).

We prove the "independence" property. Without loss of generality, we prove for the simplest case

$$\mathbb{E}(X|\mathcal{G}) = \mathbb{E}X$$

if X is independent of $\mathcal{G}$. It is enough to show that

$$\mathbb{E}[\mathbb{E}X1_A] = \mathbb{E}(X1_A),$$

for any $A \in \mathcal{G}$. Because of the independence, we have

$$\mathbb{E}[\mathbb{E}X1_A] = \mathbb{E}X\mathbb{E}[1_A] = \mathbb{E}X\mathbb{P}(A) = \mathbb{E}(X1_A).$$

The second independence property is left as an exercise.

We prove the "independence" property, i.e., if X is $\mathcal{G}$-measurable (i.e. $\sigma(X) \subset \mathcal{G}$) and Y is independent of $\mathcal{G}$, then $E[f(X,Y)|\mathcal{G}] = g(X)$, where $g(x) = Ef(x,Y)$.According to the definition, it is enough to show that

$$E[g(X)Z] = E[f(X,Y)Z],$$

for any $\mathcal{G}$-measurable random variable Z. The independence between Y and (X,Z) implies that

$$\begin{aligned}
E[f(X,Y)Z] &= \int\int\int f(x,y)zp_{X,Z}(x,z)p_Y(y)dxdydz \\
&= \int\int\left(\int f(x,y)p_Y(y)dy\right)zp_{X,Z}(x,z)dxdz \\
&= \int\int g(x)zp_{X,Z}(x,z)dxdz \\
&= E[g(X)Z].
\end{aligned}$$

where we have interchange the order of multiple integral.

1.2 On the agreement between condition expectations in the classical and modern sense

Let us prove that

$$\mathbb{E}(X|Y) = \mathbb{E}(X|\sigma(Y)), \tag{1.5}$$

where LHS is the classical version of conditional expectation, while RHS is the modern mathematical formal version.

We start from establishing the following result: if Z is a $\mathcal{G}$-measurable r.v., where $\mathcal{G} = \sigma(Y)$, then there exists a Borel-measurable function φ, s.t. $Z = \varphi(Y)$. We prove it as follows.

Proof. First denote by

$$Y^{-1}(\Lambda) = \{\omega : Y(\omega) \in \Lambda\}.$$

Step 1: If $Z = 1_B$, where $B \in \mathcal{G} = \sigma(Y)$, so there exists $\Lambda \in \mathcal{B}_{\mathbb{R}}$, s.t. $B = \{Y \in \Lambda\}$, i.e. $B = Y^{-1}(\Lambda)$ then let $\varphi(x) = 1_{\{x\in\Lambda\}}$, we have

$$\varphi(Y) = 1_{\{Y\in\Lambda\}} = 1_B = Z.;$$

Step 2: If $Z = \sum_{i=1}^{n} a_i 1_{A_i}$, where $A_i \in \sigma(Y)$ are disjoint. So, we have their corresponding disjoint Λ_i, then let $\varphi = \sum_{i=1}^{n} a_i 1_{\Lambda_i}$;

Step 3: If $Z \geq 0$, we have $Z_n \uparrow Z$, where Z_n has the form mentioned in step 2. Then we can get the φ_n s.t. $Z_n = \varphi_n(Y)$. Taking $\varphi = \limsup \varphi_n$, we have

$$Z = \lim Z_n = \lim \varphi_n(Y) = \varphi(Y);$$

Note that it could only be the upper limit, because we have no idea if φ_n has a limit as a function sequence, the thing that we only know is that $\varphi_n(Y)$ has the limit Z.

Step 4: For a r.v. Z, we have the decomposition $Z = Z^+ - Z^-$, then we have $\varphi = \varphi^+ - \varphi^-$. $\square$

Now, we prove (1.5).

Proof. We note that

$$\text{LHS} = g(Y) = \mathbb{E}(X|Y = y)|_{y=Y} = \frac{\int x f(x, Y)\mathrm{d}x}{f_Y(Y)}.$$

And for $\forall A \in \sigma(Y)$, there exists a Borel-measurable function h, s.t. $1_A = h(Y)$. Thus,

$$
\begin{aligned}
\mathbb{E}(1_A g(Y)) &= \mathbb{E}(g(Y)h(Y)) \\
&= \int g(y)h(y)f_Y(y)\mathrm{d}y \\
&= \int \frac{\int xf(x,y)\mathrm{d}x}{f_Y(y)} h(y)f_Y(y)\mathrm{d}y \\
&= \int xh(y)f(x,y)\mathrm{d}x\mathrm{d}y \\
&= \mathbb{E}(Xh(Y)) \\
&= \mathbb{E}(1_A X)
\end{aligned}
$$

which indicates

$$\mathbb{E}(X|\sigma(Y)) = g(Y) = \mathbb{E}(X|Y). \ \square$$

2 Local martingales

For technical purposes, the literature has developed the notion of local martingale, which generalize the notation of martingale. This is because martingale is sometimes too restrictive. By generalization to local martingales, one is still able to get something useful. The definition is given as follows. A stochastic process $\{M(t)\}$ adapted to the filtration $\{\mathcal{F}(t)\}$ is called a local martingale if and only if there exists an increasing sequence $\{\tau_n\}_{n=1}^{\infty}$ of stopping times with $\lim_{n\to\infty} \tau_n = \infty$ such that the stopped process $M(t \wedge \tau_n)$ is a martingale adapted to the filtration $\{\mathcal{F}(t)\}$ for every $n \geq 1$.

It can be easily shown that every martingale is also a local martingale, and that there exist local martingales which are not martingales. To understand the definition, we suggest the following excellent exercise. We can show that every nonnegative local martingale is a supermartingale. Indeed, we have

$$\mathbb{E}[M(t \wedge \tau_n)|\mathcal{F}(s)] = M(s \wedge \tau_n)$$

for every $0 < s < t$. Taking liminf, we have

$$\lim_{n\to\infty}\inf \mathbb{E}[M(t\wedge\tau_n)|\mathcal{F}(s)] = \lim_{n\to\infty}\inf M(s\wedge\tau_n) \equiv M(s).$$

By the Fatou lemma, we have

$$\lim_{n\to\infty}\inf \mathbb{E}[M(t\wedge\tau_n)|\mathcal{F}(s)] \geq \mathbb{E}[\lim_{n\to\infty}\inf M(t\wedge\tau_n)|\mathcal{F}(s)] \equiv \mathbb{E}[M(t)|\mathcal{F}(s)].$$

Thus, we arrive at $\mathbb{E}[M(t)|\mathcal{F}(s)] \leq M(s)$, which asserts the supermartingale property.

3 Doob-Meyer decomposition

3.1 Doob decomposition in discrete-time cases

First, let's introduce some definitions before we move on to the detailed discussions.

DEFINITION 1. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random sequence $\{A_n\}_{n=0}^{\infty}$ adapted to the filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$. $\{A_n\}$ is called **predictable**, if for any $n \geq 1$, A_n is $\mathcal{F}_{n-1}$ − measurable.

THEOREM 1 (Doob Decomposition)**.** *Consider a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*. Suppose* $\{X_n\}_{n=0}^{\infty}$ *is a submartingale adapted to the filtration* $\{\mathcal{F}_n\}_{n=0}^{\infty}$*. Then* X_n *can be uniquely decomposed into two parts,*

$$X_n = M_n + A_n,$$

where $\{M_n\}$ *is a martingale adapted to* $\{\mathcal{F}_n\}$*, and* $\{A_n\}$ *is a non-decreasing, predictable sequence of random variables with* $A_0 = 0$*.*

To prove the theorem, we look at its consequence, which provides us the desirable solution. If there exist $\{A_n\}$ and $\{M_n\}$ satisfying the Doob decomposition, we have

$$X_n - M_n = A_n, \tag{3.1}$$

$$\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n. \tag{3.2}$$

If we change the subscripts of (3.1) into $n+1$, and take conditional expectation given $\mathcal{F}_n$ and subtract (3.1), we have

$$\mathbb{E}[X_{n+1} - M_{n+1}|\mathcal{F}_n] - (X_n - M_n) = \mathbb{E}[A_{n+1}|\mathcal{F}_n] - A_n.$$

By (3.2), the left side of the equation above becomes $\mathbb{E}[X_{n+1}|\mathcal{F}_n] - X_n$. On the other hand, since $\{A_n\}$ is predictable, i.e.,

$$\mathbb{E}[A_{n+1}|\mathcal{F}_n] = A_{n+1},$$

the right hand side of above equation becomes $A_{n+1} - A_n$. Thus we have

$$\mathbb{E}[X_{k+1}|\mathcal{F}_k] - X_k = A_{k+1} - A_k.$$

Taking sum of the above equation from $k = 0$ to $k = n-1$ and plugging in $A_0 = 0$, we obtain the following explicit expression of A_n in terms of $\{X_n\}$ as below:

$$A_n = \sum_{k=1}^{n} (A_k - A_{k-1}) = \sum_{k=1}^{n} (\mathbb{E}[X_k|\mathcal{F}_{k-1}] - X_{k-1}). \tag{3.3}$$

So we only need to verify that A_n defined by (3.3) and $M_n = X_n - A_n$ satisfy the following conditions:

- A_n is a non-decreasing sequence;
- A_n is predictable;
- $M_n = X_n - A_n$ is a martingale.

Proof. (existence) Let

$$A_n := \sum_{k=1}^{n} (\mathbb{E}[X_k|\mathcal{F}_{k-1}] - X_{k-1}) \text{ and } M_n := X_n - A_n.$$

Since X_n is a submartingale, we have $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \geq X_n$. Then, obviously $A_{n+1} \geq A_n$, i.e., A_n is a non-decreasing sequence. Also, since $\mathbb{E}[X_k|\mathcal{F}_{k-1}] - X_{k-1}$ is $\mathcal{F}_{k-1}$-measurable, A_n is also $\mathcal{F}_{n-1}$-measurable, i.e., A_n is predictable. Finally we verify that M_n is a martingale. Indeed, we deduce that

$$\begin{aligned}
M_n &= X_n - A_n \\
&= X_n - \sum_{k=1}^{n} (\mathbb{E}[X_k|\mathcal{F}_{k-1}] - X_{k-1}) \\
&= \sum_{k=1}^{n} (X_k - \mathbb{E}[X_k|\mathcal{F}_{k-1}]) + X_0 \\
&= X_n - \mathbb{E}[X_n|\mathcal{F}_{n-1}] + M_{n-1}
\end{aligned}$$

Taking conditional expectation given $\mathcal{F}_{n-1}$ on both sides, we get

$$\mathbb{E}[M_n|\mathcal{F}_{n-1}] = M_{n-1}.$$

Thus, M_n is a martingale.

(uniqueness) Suppose there is another decomposition $X_n = M'_n + A'_n$, where M'_n is a martingale and A'_n is a non-decreasing sequence with $A'_0 = 0$, i.e.,

$$X_n = M_n + A_n \text{ and } X_n = M'_n + A'_n$$

Subtraction yields

$$M_n - M'_n = A'_n - A_n,$$

denote either side as D_n, i.e.,

$$D_n := M_n - M'_n = A'_n - A_n.$$

$A'_0 = A_0 = 0$ implies that $D_0 = 0$.

Further, both A_n and A'_n being predictable implies that D_n is predictable, and both M_n and M'_n being martingales implies that D_n is a martingale. Thus, inductively, we have

$$D_n = \mathbb{E}[D_n|\mathcal{F}_{n-1}] = D_{n-1} = \cdots = D_0 = 0.$$

Therefore, $D_n = 0$ for all n. It implies that

$$M_n = M'_n \text{ and } \quad A_n = A'_n. \square$$

Next, we show some examples of the Doob decomposition, which can be easily verified by readers.

1. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$ and an adapted i.i.d. random sequence $X_1, X_2, \cdots$. Suppose
$$S_n = \sum_{k=1}^{n} X_k.$$
If $\mathbb{E}(X_1) = \mu > 0$, then S_n is a submartingale and can be expressed as the following Doob decomposition:
$$S_n = (S_n - n\mu) + n\mu.$$

2. Suppose that S_n is a symmetric random walk, then S_n^2 is a submartingale and can be expressed as the following Doob decomposition:
$$S_n^2 = (S_n^2 - n) + n.$$

3.2 Some explanations for the Doob decompostion

We have so far set up and proved Doob Decomposition theorem, and we have given some simple examples. To give readers a better understanding of the therorem, we now give a more concrete example and some explanations.

In a simple game, someone A decides the amount of the next game according to the results of previous games. Assume that she has principal ξ_0 at first, and denote her total principal after the nth game as ξ_n. To express the simple game as a mathematical model, there are two aspects to specify:

1. Let η_n denote whether someone someone will win the nth game, taking values 1 if she wins and -1 if he loses. Here we suppose $\{\eta_n\}$ is an i.i.d. random sequence.

2. Let $\Delta_n(\xi_0, \eta_1, \eta_2, \cdots, \eta_{n-1})$ (or, equivalently, $\tilde{\Delta}_n(\xi_0, \xi_1, \xi_2, \cdots, \xi_{n-1})$) denote the nth game amount according to the results of first n games, where Δ_n should be a nonnegative Borel-measurable function.

Based on the above assumptions, we know the principal of A after the nth game is

$$\begin{aligned} \xi_n &= \xi_{n-1} + \Delta_n(\xi_0, \eta_1, \eta_2, \cdots, \eta_{n-1})\eta_n \\ &= \xi_0 + \sum_{k=1}^{n} \Delta_k(\xi_0, \eta_1, \eta_2, \cdots, \eta_{k-1})\eta_k. \end{aligned}$$

Now assume that

$$\mathbb{P}(\eta_n = 1) = p, \mathbb{P}(\eta_n = -1) = q = 1 - p.$$

Suppose $\mathbb{E}\eta_n = p - q := \mu$ and let $\mathcal{F}_n = \sigma(\eta_1.\eta_2, \cdots, \eta_n)$ denote the filtration. Then, it is easy to verify that, when $\mu > 0$, the sequence $\{\xi_n\}$ is a $\{\mathcal{F}_n\}$ – submartingale. Intuitively the game is beneficial to A, that is, A will win every time, if ruling out ”randomness”, ”uncertainty” or ”risk”.

Thus we can decompose the wealth of A into two parts. The first part is purely random and with the increase of expectation equal to 0. This part represents the uncertainty (martingale) of every game. The second part is known (predictable) to A and ensures that she will gain benefit (nondecreasing) at every game. Written in a mathematical form, we have $\xi_{n+1} = M_{n+1} + A_{n+1}$, where

$$\begin{aligned} M_{n+1} &= \xi_0 + \sum_{k=1}^{n+1} \Delta_k(\xi_0, \eta_1, \eta_2 \cdots, \eta_{k-1})(\eta_k - \mathbb{E}(\eta_k)) \\ A_{n+1} &= \sum_{k=1}^{n+1} \Delta_k(\xi_0, \eta_1, \eta_2 \cdots, \eta_{k-1})\mathbb{E}(\eta_k). \end{aligned}$$

This is the Doob decomposition of ξ_{n+1}, where M_{n+1} is a martingale and A_{n+1} is predictable and nondecreasing sequence.

Indeed, because of

$$\begin{aligned}
\mathbb{E}[\xi_{n+1}|\mathcal{F}_n] &= \mathbb{E}\left[\xi_0 + \sum_{k=1}^{n+1} \Delta_k(\xi_0, \eta_1, \eta_2, \cdots, \eta_{k-1})\eta_k \middle| \mathcal{F}_n\right] \\
&= \mathbb{E}[\xi_n + \Delta_{n+1}(\xi_0, \eta_1, \eta_2, \cdots, \eta_n)\eta_{n+1}|\mathcal{F}_n] \\
&= \xi_n + \Delta_{n+1}(\xi_0, \eta_1, \eta_2, \cdots, \eta_n)\mathbb{E}[\eta_{n+1}|\mathcal{F}_n] \\
&= \xi_n + \Delta_{n+1}(\xi_0, \eta_1, \eta_2, \cdots, \eta_n)\mathbb{E}[\eta_{n+1}] \\
&= \xi_n + \Delta_{n+1}(\xi_0, \eta_1, \eta_2, \cdots, \eta_n)\mu \\
&\geq \xi_n,
\end{aligned}$$

ξ_n is a submartingale. For M_n, its martingale property follows from

$$\begin{aligned}
\mathbb{E}[M_{n+1}|\mathcal{F}_n] &= \mathbb{E}[M_n + \Delta_{n+1}(\xi_0, \eta_1, \eta_2 \cdots, \eta_n)(\eta_{n+1} - \mathbb{E}(\eta_{n+1}))|\mathcal{F}_n] \\
&= M_n + \Delta_{n+1}(\xi_0, \eta_1, \eta_2 \cdots, \eta_n)\mathbb{E}[(\eta_{n+1} - \mathbb{E}(\eta_{n+1}))|\mathcal{F}_n] \\
&= M_n.
\end{aligned}$$

For A_n, we have

$$\begin{aligned}
\mathbb{E}[A_{n+1}|\mathcal{F}_n] &= \mathbb{E}\left[\sum_{k=1}^{n+1} \Delta_k(\xi_0, \eta_1, \eta_2 \cdots, \eta_{k-1})\mathbb{E}(\eta_k) \middle| \mathcal{F}_n\right] \\
&= \sum_{k=1}^{n+1} \Delta_k(\xi_0, \eta_1, \eta_2 \cdots, \eta_{k-1})\mathbb{E}[\mathbb{E}(\eta_k)|\mathcal{F}_n] \\
&= \sum_{k=1}^{n+1} \Delta_k(\xi_0, \eta_1, \eta_2 \cdots, \eta_{k-1})\mathbb{E}(\eta_k) \\
&= A_{n+1}.
\end{aligned}$$

Therefore, A_n is predictable.

3.3 Doob-Meyer decomposition in continuous-time cases

In the previous sections, we have introduced the Doob decomposition in discrete-time cases. The Doob decomposition can be similarly developed for continuous-time cases. Under some technical conditions, a continuous-time submartingale can be uniquely decomposed as the sum of a martingale and a non-decreasing "good" process, here "good" means the process has some good properties similar with predictability. This theorem is called the Doob-Meyer decomposition. Since we focus on applications, we will not state the theorem rigorously. Interested readers may refer to the corresponding section in Section 1.5 of Karatzas and Shreve [1]. Instead, we present the following examples of Doob-Meyer decomposition The to be verified by readers.

1. For a Brownian motion $W(t)$, suppose $X(t) = aW(t) + \mu t$ for some $\mu > 0$. Then $X(t)$ is a submartingale and $X(t) = aW(t) + \mu t$ gives its Doob-Meyer decomposition.

2. For a Brownian motion $W(t)$, $W^2(t)$ is a submartingale. Then,

$$W^2(t) = [W^2(t) - t] + t$$

   renders its Doob-Meyer decompostion.

3. For a Poisson process $N(t)$ with parameter λ, $N(t)$ is a submartingale. Then, its Doob-Meyer decomposition is given by

$$N(t) = [N(t) - \lambda t] + \lambda t.$$

We suggest an excellent exercise for readers. For a Poisson process $N(t)$ with parameter λ, show that $\eta(t) = [N(t) - \lambda t]^2$ is a submartingale and find its Doob-Meyer decomposition. (Hint: $M(t) = N(t) - \lambda t$ is a martingale, and $\langle M \rangle (t) = \lambda t$.)

3.4 Quadratic variation

We have introduced Doob (– Meyer) decomposition theorem for submartingales, and have found out that square of some martingales like S_n^2, $W^2(t)$, and $[N(t) - \lambda t]^2$ are all submartingales. Thus we can conjecture the following proposition:

PROPOSITION 1. *If M_n (resp. $M(t)$) is a martingale, then M_n^2 (or $M^2(t)$) is a submartingale.*

Without loss of generality, we prove the continuous-time case.

Proof. For the convex function $\varphi(x) = x^2$, it follows from the Jensen inequality for conditional expectations that

$$\mathbb{E}[M^2(t)|\mathcal{F}(s)] \geq (\mathbb{E}[M(t)|\mathcal{F}(s)])^2 = M(s)^2.$$

Thus, $M^2(t)$ is a submartingale. □

If M_n (resp. $M(t)$) is a martingale, then

$$\mathbb{E}[M_{n+1}^2|\mathcal{F}_n] - M_n^2 = \mathbb{E}[(M_{n+1} - M_n)^2|\mathcal{F}_n], \tag{3.4}$$

(resp.

$$\mathbb{E}[M^2(t)|\mathcal{F}(s)] - M^2(s) = \mathbb{E}[(M(t) - M(s))^2|\mathcal{F}(s)].)$$

Indeed, this property can be verified by the following derivations:

$$\begin{aligned}
\mathbb{E}[M^2(t)|\mathcal{F}(s)] &= \mathbb{E}[(M(s) + M(t) - M(s))^2|\mathcal{F}(s)] \\
&= M^2(s) + 2M(s)\mathbb{E}[M(t) - M(s)|\mathcal{F}(s)] + \mathbb{E}[(M(t) - M(s))^2|\mathcal{F}(s)] \\
&= M^2(s) + \mathbb{E}[(M(t) - M(s))^2|\mathcal{F}(s)].
\end{aligned}$$

As a by-product, it follows that

$$\mathbb{E}[M^2(t)|\mathcal{F}(s)] = M^2(s) + \mathbb{E}[(M(t) - M(s))^2|\mathcal{F}(s)] \geq M^2(s).$$

We assume that $\mathbb{E}(M_n^2) < \infty$ (resp. $\mathbb{E}M^2(t) < \infty$), i.e., M_n (resp. $M(t)$) are *square-integrable martingales.*

According to the previous discussions, M_n^2 (resp. $M^2(t)$) is a submartingale, then applying Doob (– Meyer) Decomposition to M_n^2 (resp. $M^2(t)$), we can uniquely decompose them into $M_n^2 = N_n + A_n$ (resp. $M^2(t) = N(t) + A(t)$), i.e., $N_n = M_n^2 - A_n$ (resp. $N(t) = M^2(t) - A(t)$) is a martingale.

DEFINITION 2. For a square-integrable martingale M_n (resp. $M(t)$), the non-decreasing random part in the Doob (-Meyer) decomposition of its square is called the quadratic variation, denoted as $\langle M\rangle_n$ (resp. $\langle M\rangle(t)$).

By the definition of quadratic variation, for a square-integrable martingale M, $M^2 - \langle M\rangle$ is also a martingale.

For discrete-time cases, according to formula (3.3) of A_n, the quadratic variation has the following explicit form:

$$\langle M\rangle_n = \sum_{k=1}^{n} \left(\mathbb{E}[M_k^2|\mathcal{F}_{k-1}] - M_{k-1}^2\right) = \sum_{k=1}^{n} \mathbb{E}[(M_k - M_{k-1})^2|\mathcal{F}_{k-1}], \tag{3.5}$$

where the second equality follows from (3.4).

For continuous-time cases, it is natural to conjecture that the quadratic variation of $M(t)$ follows from the continuous-time limit of a discrete analogy.

DEFINITION 3. Denote by $\Pi:\ 0 = t_0 < t_1 < \cdots < t_m = t$ a partition of the interval $[0,t]$. Then the quadratic variation of M with respect to the partition Π is defined as

$$QV_t^M(\Pi) = \sum_{k=1}^{m} |M(t_k) - M(t_{k-1})|^2.$$

It can be verified that as the partition becomes finer and finer, i.e.,

$$\|\Pi\| = \max_{1\le k\le m} |M(t_k) - M(t_{k-1})| \to 0,$$

$QV_t^M(\Pi)$ will converge to $\langle M\rangle(t)$ in some mode, i.e.,

$$\lim_{\|\Pi\|\to 0} QV_t^M(\Pi) = \langle M\rangle(t), \text{ in probability.}$$

3.5 Cross variation

We have already defined the quadratic variation of the square-integrable martingale M, where $M^2 - \langle M\rangle$ is also a martingale. Similarly, given two martingales M and N, is there any stochastic process A, such that $MN - A$ is a martingale? If so, is such stochastic process unique? What is its explicit expression? We now give a detailed discussion about these questions.

First, if M and N are both martingales, it is easy to see that $M+N$ and $M-N$ are also martingales. Therefore, $(M+N)^2 - \langle M+N \rangle$ and $(M-N)^2 - \langle M-N \rangle$ are martingales. Because of

$$(x+y)^2 - (x-y)^2 = 4xy,$$

we have

$$MN = \frac{1}{4}(M+N)^2 - \frac{1}{4}(M-N)^2.$$

Moreover, we have

$$\begin{aligned} & MN - \frac{1}{4}(\langle M+N \rangle - \langle M-N \rangle) \\ = \; & \frac{1}{4}\Big((M+N)^2 - \langle M+N \rangle\Big) - \frac{1}{4}\Big((M-N)^2 - \langle M-N \rangle\Big). \end{aligned}$$

Because the right side of the above equation is a martingale, so is the left side. Therefore, we can find the expression of A such that $MN - A$ is a martingale. Such an A is given by

$$A = \frac{1}{4}(\langle M+N \rangle - \langle M-N \rangle).$$

DEFINITION 4. For two square-integrable martingales M and N, we define the cross-variation of M and N by

$$\langle M, N \rangle = \frac{1}{4}(\langle M+N \rangle - \langle M-N \rangle).$$

In discrete-time cases, $\langle M, N \rangle_n$ can be expressed as

$$\begin{aligned} \langle M, N \rangle_n &= \frac{1}{4}(\langle M+N \rangle_n - \langle M-N \rangle_n) \\ &= \frac{1}{4}\left(\sum_{k=1}^{n} \mathbb{E}[(M_k + N_k - M_{k-1} - N_{k-1})^2 | \mathcal{F}_{k-1}] - \sum_{k=1}^{n} \mathbb{E}[(M_k - N_k - M_{k-1} + N_{k-1})^2 | \mathcal{F}_{k-1}]\right) \\ &= \sum_{k=1}^{n} \mathbb{E}[(M_k - M_{k-1})(N_k - N_{k-1}) | \mathcal{F}_{k-1}]. \end{aligned}$$

Moreover, because $\langle M+N\rangle_n$ and $\langle M-N\rangle_n$ are both predictable, by definition, so is $\langle M,N\rangle_n$. Similarly, to (3.5), we have

$$\langle M,N\rangle_n = \sum_{k=1}^{n} (\mathbb{E}[M_k N_k|\mathcal{F}_{k-1}] - M_{k-1}N_{k-1}).$$

Prove this claim as an exercise.

In continuous-time cases, when $\|\Pi\|$ approaches 0, $QV_t^M(\Pi)$ converges to $\langle M\rangle(t)$. Therefore, by the definition of $\langle M,N\rangle(t)$, we can use

$$\begin{aligned}
& \frac{1}{4}\left(QV_t^{M+N}(\Pi) - QV_t^{M-N}(\Pi)\right) \\
= \ & \frac{1}{4}\left(\sum_{k=1}^{m} |M(t_k)+N(t_k)-M(t_{k-1})-N(t_{k-1})|^2 - \sum_{k=1}^{m} |M(t_k)-N(t_k)-M(t_{k-1})+N(t_{k-1})|^2\right) \\
\equiv \ & \sum_{k=1}^{m} [M(t_k)-M(t_{k-1})][N(t_k)-N(t_{k-1})]
\end{aligned}$$

to approach $\langle M,N\rangle$.

Now, we move on to discuss the uniqueness of such stochastic process A such that $MN-A$ is a martingale. We already have as solution $A=\langle M,N\rangle$ that meets the requirements. Though MN is not necessarily a submartingale, as long as A is predictable, we have its uniqueness. Interested readers are suggested to prove it by mimicking our previous proof for the uniqueness of the Doob decomposition of submartingales. Besides, because of the similarity between the discrete and the continuous cases, it is reasonable to believe that, under some technical conditions, such decomposition is unique in continuous-time cases. The formulation and proof of such a claim is beyond the scope of these notes. Interested readers may refer to Section 1.5 of Karatzas and Shreve [1].

For readers to master the concepts, we suggest the following excellent exercise. Consider the iid random variable sequence $\{(X_n,Y_n)\}$ on the given probability space $(\Omega,\mathcal{F},\mathbb{P})$, where

$$\mathbb{E}X_1=\mu_X,\ \mathbb{E}Y_1=\mu_Y,\ Var(X_1)=\sigma_X^2,\ Var(Y_1)=\sigma_Y^2,\ Cov(X_1,Y_1)=\sigma_{XY}.$$

Denote by

$$S_n=\sum_{k=1}^{n} X_k, \quad T_n=\sum_{k=1}^{n} Y_k,$$

and

$$M_n = S_n - n\mu_X, \quad N_n = T_n - n\mu_Y.$$

1. (a) Prove that M_n and N_n are both square-integrable martingales, and

$$\langle M \rangle_n = n\sigma_X^2, \quad \langle N \rangle_n = n\sigma_Y^2;$$

   (b) Prove that

$$\langle M, N \rangle_n = n\sigma_{XY}.$$

4 Brownian motion

4.1 Variations of a continuously differentiable function

Let $f(t)$ be a second-order differentiable function defined on $0 \le t \le T$. The first-order variation of f up to time T is defined as

$$FV_f(T) := \lim_{||\Pi|| \to 0} \sum_{j=0}^{n-1} \left| f(t_{j+1}) - f(t_j) \right|,$$

where $\Pi = \{t_0, t_1, \cdots, t_n\}$ is a partition with $0 = t_0 < t_1 < \cdots < t_n = T$ and

$$||\Pi|| = \max\{t_{j+1} - t_j,\ j = 0, 1, 2, ..., n-1\}.$$

Then, we have

$$\lim_{||\Pi|| \to 0} \sum_{j=0}^{n-1} |f(t_{j+1}) - f(t_j)| = \lim_{||\Pi|| \to 0} \sum_{j=0}^{n-1} |f'(t_j^*)(t_{j+1} - t_j)| = \int_0^T |f\prime(t)| dt < +\infty.$$

The first-order variation of a continuously differentiable function is finite.

The quadratic variation of f up to time T is defined as

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

We note that

$$\sum_{j=0}^{n-1} [f(t_{j+1}) - f(t_j)]^2 = \sum_{j=0}^{n-1} |f\prime(t_j^*)|^2 (t_{j+1} - t_j)^2 \leq ||\Pi|| \sum_{j=0}^{n-1} |f\prime(t_j^*)|^2 (t_{j+1} - t_j).$$

Thus, we have

$$\begin{aligned} QV_f(T) &\leq \lim_{||\Pi|| \to 0} \left[||\Pi|| \sum_{j=0}^{n-1} |f\prime(t_j^*)|^2 (t_{j+1} - t_j) \right] \\ &= \lim_{||\Pi|| \to 0} ||\Pi|| \lim_{||\Pi|| \to 0} \sum_{j=0}^{n-1} |f\prime(t_j^*)|^2 (t_{j+1} - t_j) \\ &= \lim_{||\Pi|| \to 0} ||\Pi|| \cdot \int_0^T |f\prime(t)|^2 dt = 0. \end{aligned}$$

4.2 Variations of a Brownian Motion

Denote by

$$Q_\Pi := \sum_{j=0}^{n-1} |W(t_{j+1}) - W(t_j)|^2$$

The quadratic variation is thus

$$QV_W(T) = \lim_{||\Pi|| \to 0} Q_\Pi = \lim_{||\Pi|| \to 0} \sum_{j=0}^{n-1} |W(t_{j+1}) - W(t_j)|^2 .$$

We note that Q_Π is a sum of independent random variables. Therefore, its mean and variance are the sums of the means and variances of these random variables, respectively. Because

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

we have

$$\mathbb{E}(Q_\Pi) \equiv \sum_{j=0}^{n-1} \mathbb{E}(W(t_{j+1}) - W(t_j))^2 = \sum_{j=0}^{n-1} (t_{j+1} - t_j) \equiv T.$$

Moreover, we have

$$\begin{aligned}
\operatorname{var}\left[(W(t_{j+1})-W(t_j))^2\right] &= \mathbb{E}\left[\left((W(t_{j+1})-W(t_j))^2-(t_{j+1}-t_j)\right)^2\right] \\
&= \mathbb{E}\left[(W(t_{j+1})-W(t_j))^4\right]-2(t_{j+1}-t_j)\mathbb{E}\left(W(t_{j+1})-W(t_j)\right)^2+(t_{j+1}-t_j)^2.
\end{aligned}$$

It can be easily proved that, for a variable X following normal distribution $\mathcal{N}(0,\sigma^2)$, we have

$$\mathbb{E}\left(X^4\right)=3\sigma^4.$$

Therefore, we have

$$\begin{aligned}
\operatorname{var}\left[(W(t_{j+1})-W(t_j))^2\right] &= 3(t_{j+1}-t_j)^2-2(t_{j+1}-t_j)^2+(t_{j+1}-t_j)^2 \\
&= 2(t_{j+1}-t_j)^2,
\end{aligned}$$

Thus, we have

$$\begin{aligned}
\operatorname{var}(Q_\Pi) &= \operatorname{var}\left[\sum_{j=0}^{n-1}(W(t_{j+1})-W(t_j))^2\right] \\
&= \sum_{j=0}^{n-1}\operatorname{var}\left[(W(t_{j+1})-W(t_j))^2\right] \\
&= \sum_{j=0}^{n-1}2(t_{j+1}-t_j)^2 \\
&\leq \sum_{j=0}^{n-1}2||\Pi||(t_{j+1}-t_j)=2||\Pi||T.
\end{aligned}$$

Now, we provide a heuristic argument. The above inequatlity implies that $\lim_{||\Pi||\to 0}\operatorname{var}(Q_\Pi)\equiv 0$. Hence, if follows that

$$\lim_{||\Pi||\to 0}Q_\Pi\equiv\mathbb{E}(Q_\Pi)=T. \tag{4.1}$$

We provide a rigorous argument in what follows. For $\forall\varepsilon,\delta>0$, when $||\Pi||<\frac{\varepsilon^2\delta}{2T}$, by Chebyshev's

Inequality, we have

$$\begin{aligned}
\mathbb{P}(|Q_\Pi - T| \geq \varepsilon) &\leq \frac{1}{\varepsilon^2}\mathrm{var}(Q_\Pi) \\
&\leq \frac{2||\Pi||T}{\varepsilon^2} \\
&< \delta,
\end{aligned}$$

which implies Q_Π converges to T in probability when $||\Pi||$ goes to 0. In particular, when $\Pi_N : t_j^{(N)} = tj2^{-N}$, we can prove Q_{Π_N} converges to T a.s. when N goes to infinity. This is because

$$Q_{\Pi_n} - T = \sum_{j=0}^{2^N-1} [|W(t_{j+1}) - W(t_j)|^2 - (t_{j+1} - t_j)]$$

is a summation of i.i.d. RV sequence with the expectation of 0 and finite variance. Thus, by SLLN,

$$Q_{\Pi_n} \xrightarrow{a.s.} T.$$

Therefore, Q_Π converges almost surely only when Π take some specific partitions. In general, Q_Π converges in probability.

The first-order variation is defined as

$$FV_W(T) = \lim_{||\Pi||\to 0} \sum_{j=0}^{n-1} |W(t_{j+1}) - W(t_j)|.$$

It is obvious that

$$Q_\Pi \leq \max |W(t_{j+1}) - W(t_j)| \cdot \sum_{j=0}^{n-1} |W(t_{j+1}) - W(t_j)| \tag{4.2}$$

Suppose

$$\lim_{||\Pi||\to 0} \sum_{j=0}^{n-1} |W(t_{j+1}) - W(t_j)| < +\infty.$$

By the continuity property of Brownian motion, we have

$$\lim_{||\Pi||\to 0} \max |W(t_{j+1}) - W(t_j)| = 0.$$

Thus, by taking limit on the both sides of the inequality, we have

$$\lim_{||\Pi||\to 0} Q_\Pi \leq \lim_{||\Pi||\to 0} \max |W(t_{j+1}) - W(t_j)| \cdot \lim_{||\Pi||\to 0} \sum_{j=0}^{n-1} |W(t_{j+1}) - W(t_j)| = 0.$$

This implies that

$$\lim_{||\Pi||\to 0} Q_\Pi = 0,$$

which contradicts to (4.1). Hence, we must have

$$FV_W(T) = \lim_{||\Pi||\to 0} \sum_{j=0}^{n-1} |W(t_{j+1}) - W(t_j)| = +\infty.$$

4.3 On multidimensional Brownian motions

Suppose $\{(B_1(t), B_2(t), \cdots, B_d(t))\}$ is a d dimensional Brownian motion with covarian matrix $(t-s)\Sigma$, where $\Sigma = (\rho_{ij})_{d\times d}$. Denote by

$$B(t) = (B_1(t), B_2(t), \cdots, B_d(t))^\top$$

a column random vector. In this case, we knew that

$$t\Sigma = \mathrm{cov}(B(t), B(t)) \equiv (\mathrm{cov}(B_i(t), B_j(t)))_{d\times d}.$$

Thus, we have the following elementwise form

$$t\rho_{ij} = \mathrm{cov}(B_i(t), B_j(t)) = \mathbb{E}[B_i(t)B_j(t)].$$

So, it is easy to see that

$$\rho_{ij} = \frac{\mathrm{cov}(B_i(t), B_j(t))}{\sqrt{\mathrm{var}(B_i(t))}\sqrt{\mathrm{var}(B_j(t))}} = \frac{\mathbb{E}[B_i(t)B_j(t)]}{\sqrt{\mathrm{var}(B_i(t))}\sqrt{\mathrm{var}(B_j(t))}} = \mathrm{corr}(B_i(t), B_j(t)).$$

That is to say ρ_{ij} represents the correlation between $B_i(t)$ and $B_j(t)$. Now, we claim that ρ_{ij} renders the covariation between $B_i(t)$ and $B_j(t)$.

Suppose we can find matrix A such that

$$AA^\top = \Sigma. \tag{4.3}$$

Then, we can construct $B(t)$ by $B(t) = AZ(t)$, where $\{Z(t)\}$ is a standard d dimensional Brownian motion. Indeed, we just need to check all the points in the definition are satisfied. In particular, to check the covariance matrix, we deduce that

$$\begin{aligned}
\mathrm{cov}(AZ(t), AZ(t)) &= \mathbb{E}[AZ(t)(AZ(t))^\top] - \mathbb{E}(AZ(t))\mathbb{E}(AZ(t))^\top] \\
&= \mathbb{E}[AZ(t)(AZ(t))^\top] \\
&= \mathbb{E}[AZ(t)Z(t)^\top A^\top] \\
&= A\mathbb{E}[Z(t)Z(t)^\top]A^\top \\
&= AIA^\top = AA^\top = \Sigma.
\end{aligned}$$

This can be written in elementwise form (to add in the book.) In linear algebra, (4.3) is called Cholesky decomposition. In particuar the matrix A can be chosen as a lower diagonal matrix.

5 Stochastic calculus

5.1 Properties of Itô integral $I(t)$ for simple processes $\triangle(t)$

First, we add one more point: $I(t)$ **is continuous in** t**;**

Though $I(t)$ is not continuous if we focus on grid points $\{t_k\}$, we have the continuity of $I(t)$ in t. A simple reason is as follows. For any arbitrary $t^* \in [t_k, t_{k+1})$, we have

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

$I(t)$ is $\mathcal{F}(t)$-measurable and $I(t)$ has the linearity for integration: For any $t \in [t_k, t_{k+1})$,

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

We note that $\Delta(t_0), \Delta(t_1), ...,$ and $\Delta(t_k)$ are $\mathcal{F}(t)$-measurable; $W(t_0), W(t_1), ...,$ and $W(t)$ are $\mathcal{F}(t)$-measurable. Thus, $I(t)$ is $\mathcal{F}(t)$-measurable;

$I(t)$ is a martingale: For any $t > s$, we will prove that

$$\mathbb{E}\left[I(t)|\mathcal{F}(s)\right] = I(s). \tag{5.1}$$

Indeed, we can easily work with the grid points as follows:

$$\begin{aligned}
\mathbb{E}\left[I(t_{k+1})|\mathcal{F}(t_k)\right] &= \mathbb{E}\left(\sum_{j=0}^{k} \Delta(t_j)[W(t_{j+1}) - W(t_j)]|\mathcal{F}(t_k) \right) \\
&= I(t_k) + \mathbb{E}[\Delta(t_k)[W(t_{k+1}) - W(t_k)]|\mathcal{F}(t_k)] \\
&= I(t_k) + \Delta(t_k)\mathbb{E}[W(t_{k+1}) - W(t_k)|\mathcal{F}(t_k)] \\
&= I(t_k) + \Delta(t_k)\mathbb{E}[W(t_{k+1}) - W(t_k)] \\
&= I(t_k).
\end{aligned}$$

Similarly, by taking care of some detailed tricks, we can prove the martingale property (5.1) outside of grid points.

Itô Isometry:

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

Without loss of generality, we still prove it for the case when t is a grid point, say $t = t_k$. For simplicity, we denote by $W(t_{j+1}) - W(t_j) = a_j$. Thus, we have

$$I(t) = \sum_{j=0}^{k} \Delta(t_j)a_j$$

and

$$I^2(t) = \sum_{j=0}^{k} \Delta^2(t_j){a_j}^2 + 2\sum_{0\le i\le j\le k} \Delta(t_i)\Delta(t_j)a_ia_j.$$

First, let's prove the expectations of cross terms are 0. Indeed, for each pair $i \le j$, we note that $\Delta(t_i)\Delta(t_j)a_i$ is $\mathcal{F}(t_j)$-measurable, a_j is independent of $\mathcal{F}(t_j)$, and $\mathbb{E}a_j = 0$. Therefore,

$$\begin{aligned}
\mathbb{E}[\Delta(t_i)\Delta(t_j)a_ia_j] &= \mathbb{E}[\mathbb{E}[\Delta(t_i)\Delta(t_j)a_ia_j|\mathcal{F}(t_j)]] \\
&= \mathbb{E}[\Delta(t_i)\Delta(t_j)a_i\mathbb{E}[a_j|\mathcal{F}(t_j)]] \\
&= \mathbb{E}[\Delta(t_i)\Delta(t_j)a_i\mathbb{E}[a_j]] \\
&= 0.
\end{aligned}$$

Then, we consider the squared terms. We have

$$\begin{aligned}
\mathbb{E}[\Delta^2(t_j){a_j}^2] &= \mathbb{E}[\mathbb{E}\left[\Delta(t_j)^2{a_j}^2|\mathcal{F}(t_j)\right]] \\
&= \mathbb{E}[\Delta(t_j)^2\mathbb{E}[{a_j}^2|\mathcal{F}(t_j)] \\
&= \mathbb{E}[\Delta(t_j)^2(t_{j+1} - t_j)].
\end{aligned}$$

Thus, it is easy to have

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

Quadratic Variation:

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

Without loss of generality, we still prove it for the case when t is a grid point, say $t = t_k$. Recall that quadratic variation of I is given by

$$\lim_{\|\Pi\|\to 0} \sum_{i=0}^{m-1} [I(\xi_{i+1}) - I(\xi_i)]^2,$$

where $\Pi = \{\xi_0, \xi_1, \xi_2, , ..., \xi_N\}$ gives a partition of the interval $[0,t]$ even finer than $\{t_0, t_1, t_2, , ..., t_n\}$. We consider partitions of the interval $[t_j, t_{j+1}]$ by taking $t_j = s_0 < s_1 < ... < s_m = t_{j+1}$. Because $\Delta(t_j)$ is constant on $[t_j, t_{j+1})$, we have

$$\sum_{i=0}^{m-1} [I(s_{i+1}) - I(s_i)]^2 = \Delta^2(t_j) \sum_{i=0}^{m-1} [W(s_{i+1}) - W(s_i)]^2.$$

Using the quadratic variation of Brownian motion, we obtain that

$$\lim_{\max\{s_{i+1}-s_i, i=0,1,2,...,m-1\}\to 0} \sum_{i=0}^{m-1} [W(s_{i+1}) - W(s_i)]^2 = t_{j+1} - t_j.$$

Thus, we have

$$\sum_{i=0}^{m-1} [I(s_{i+1}) - I(s_i)]^2 \to \Delta^2(t_j)(t_{j+1} - t_j) = \int_{t_j}^{t_{j+1}} \Delta^2(u)du \tag{5.2}$$

Summing up for all the intervals, by the definition of quadratic variation, we obtain that

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

5.2 Conditions for defining stochastic integrals

We consider two types of conditions. The first one is

$$\mathbb{E} \int_0^t \Delta^2(u)du < \infty; \tag{5.3}$$

the second one is

$$\int_0^t \Delta^2(u)du < \infty, \text{ a.e..} \tag{5.4}$$

By approximating the integrand via a sequence of simple processes, we are able to define stochastic integral under condition (5.3). However, some technical analysis allow us to define stochastic integrals under the more general condition (5.4).

Obviously, condition (5.3) implies condition (5.4). Under condition (5.3), the stochastic integral $I(t)$ is a true martingale. The Itô isometry holds. However, under condition (5.4), the stochastic integral $I(t)$ is a local martingale, which is not necessarily to be a true martingale. In this case, the Itô isometry does not necessarily hold. Under either conditions, the quadratic variation satisfies

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

5.3 Simple applications of the Itô formula

Let $f(x) = \frac{x^2}{2}$. By taking derivatives, we have $f'(x) = x$ and $f''(x) = 1$. Thus, by using the Itô formulas, we have

$$\begin{aligned} f(W(t)) &= \frac{1}{2}W^2(t) \\ &= \frac{1}{2}W^2(0) + \int_0^t W(u)dW(u) + \frac{1}{2}\int_0^t 1du \\ &= \int_0^t W(u)dW(u) + \frac{1}{2}t. \end{aligned}$$

This implies that

$$\int_0^t W(u)dW(u) = \frac{1}{2}(W^2(t) - t). \tag{5.5}$$

We have proved that, using the definition of Brownian motion previously, $\{W^2(t) - t\}$ is a martingale. Since the process of stochastic integrals $\{\int_0^t W(u)dW(u)\}$ is a martingale, the relation (5.5) renders an alternative proof of $\{W^2(t) - t\}$ being a martingale. We suggest readers to consider the following exercise

$$\int_0^t W(u)^n dW(u)$$

for any arbitrary integer n. And further think about the following popular interview question: can you use the Itô formula to find

$$\mathbb{E}W(t)^n = ?$$

Indeed, it is easy to set up an iteration relation. Please solve it!

5.4 Itô processes

Itô processes provide a broad ground for modeling asset returns. In particular, in recent years, the analysis of high-frequency trading date heavily hinges on the analysis and estimation of their quadratic variation. Assume For the Itô process X, we have

$$[X, X](t) = \int_0^t \Delta^2(u) du. \tag{5.6}$$

The proof of this result can be seen in Lemma 4.4.4 in Shreve's book. It follows the definition of quadratic variation, i.e., to show that

$$\sum (X(\xi_{i+1}) - X(\xi_i))^2 \to \int_0^t \Delta^2(u) du \tag{5.7}$$

in probability as the partition becomes finer and finer. Such a result reveals that the dt integrals don't contribute anything to quadratic variation. i.e.,

$$[X, X](t) = [I, I](t) = \int_0^t \Delta^2(u) du, \tag{5.8}$$

where

$$I(t) = \int_0^t \Delta(u) dW(u). \tag{5.9}$$

We note that, when $\Delta \equiv 1$, (5.8) reduces to

$$[I, I](t) = \int_0^t 1 du = t.$$

In this case,

$$I(t) = \int_0^t \Delta(u)dW(u) = \int_0^t 1dW(u) = W(t).$$

Thus, if we view the stochastic integral (5.9) as a generalization of Brownian motion, (5.8) is a generalization of the quadratic variation property of Brownian motion, i.e., $[W,W](t) = t$. We also note that the quadratic variation (5.8) is stochastic in general, not necessarily to be a constant as in the case of Brownian motion. This can be easily seen from (5.7).

5.5 Generalization to stochastic integrals w.r.t continuous local martingales

Before closing this section, we provide a further generalization. It is natural to consider stochastic integrals w.r.t. a general continuous local martingale. Suppose M is a continuous local martingale, we can define

$$I^M(t) = \int_0^t \Delta(u)dM(u). \tag{5.10}$$

The way to define it follows the same spirit for defining stochastic integrals w.r.t. Browian motions.

We consider two types of conditions. The first one is

$$\mathbb{E}\int_0^t \Delta^2(u)d[M,M](u) < \infty; \tag{5.11}$$

the second one is

$$\int_0^t \Delta^2(u)d[M,M](u) < \infty, \text{ a.e..} \tag{5.12}$$

By approximating the integrand via a sequence of simple processes, we are able to define stochastic integral under condition (5.11). However, some technical analysis allow us to define stochastic integrals under the more general condition (5.12).

Obviously, condition (5.11) implies condition (5.12). Under condition (5.11), the stochastic integral $I(t)$ is a true martingale. The Itô isometry holds in the following form

$$\mathbb{E}I^2(t) = \mathbb{E}\left(\int_0^t \Delta^2(u)d[M,M](u)\right), \tag{5.13}$$

However, under condition (5.12), the stochastic integral $I(t)$ is a local martingale, which is not necessarily to be a true martingale. In this case, the Itô isometry does not necessarily hold. Under either conditions, the quadratic variation satisfies

$$[I, I](t) = \int_0^t \Delta^2(u) d[M, M](u). \tag{5.14}$$

Interested readers may wonder why Itô isometry and qudratic variation share the same elements in their expressions.

Indeed, via technical stochastic analysis and under some technical conditions, one is able to show that for any (local-) martingale $\{\mathcal{M}(t)\}$, the stochastic process $[\mathcal{M}, \mathcal{M}](t)$, defined through limit of sum of squres of increments, is the unique process with continuous and nondecreasing paths making the process $\{\mathcal{M}(t)^2 - [\mathcal{M}, \mathcal{M}](t)\}$ is a (local-) martingale. This is obviously out of our reach in this course, which provides mathematical tools for research in finance and economics. Lucikly that we have seen special cases of this general theorem. For example, when $\mathcal{M}(t) = W(t)$, we have a martingale $\{W(t)^2 - [W, W](t)\} = \{W(t)^2 - t\}$ is a martingale. In addition, for the martingale

$$\mathcal{M}(t) = I(t) = \int_0^t \Delta(u) dW(u),$$

we will have a martingale constitued by

$$\mathcal{M}(t)^2 - [\mathcal{M}, \mathcal{M}](t) = I(t)^2 - [I, I](t).$$

Thus, if one has the quadatic variation property

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

by the martingale property (under condition (5.3)), it is easy to have

$$\mathbb{E}[\mathcal{M}(t)^2 - [\mathcal{M}, \mathcal{M}](t)] = \mathbb{E}[I(t)^2 - [I, I](t)] = 0,$$

i.e.,

$$\mathbb{E}I(t)^2 = \mathbb{E}[I,I](t) = \mathbb{E}\left(\int_0^t \Delta^2(u)du\right).$$

Similarly, for the case of (5.10), the aforementioned general property implies that (5.14) leads to (5.13). Thus, the quadratic variation (5.14) plays an important role; it can be understood easily from the following chart:

$$\begin{array}{cc} \sum & (I(\xi_{i+1}) - I(\xi_i))^2 \\ \downarrow & \downarrow \\ \int_0^t & \Delta^2(u)d[M,M](u) \end{array}.$$

Here, the term $\Delta^2(u)d[M,M](u)$ is an infinitesimal analogy of (5.2).

5.6 Multivariate stochastic calculus

5.6.1 Definition of covariation

We first introduce the notion of covariation as a natural generalization of quadratic variation. Suppose we have two stochastic processes $\{X(t)\}$ and $\{Y(t)\}$. The covariation between the two can be defined a stochastic process satisfying

$$[X,Y](t) = \lim_{\|\Pi\|\to 0} \sum_{k=0}^{n-1} (X(t_{k+1}) - X(t_k))(Y(t_{k+1}) - Y(t_k)), \tag{5.15}$$

where we recall that $\Pi = \{t_0, t_1, t_2, ..., t_n\}$ is a partition of the interval $[0,t]$ with $t_0 = 0$ and $t_n = t$. Here, the normal $\|\Pi\| = \max\{t_j - t_{j-1}, j = 1, 2, ..., n\}$. We note that the lim in (5.15) is in the sense of probability. We mention the following useful properties without proofs, which naturally follows from tedious and straightforward applications of the definition (5.15):

1. symmetric property: $[X,Y](t) = [Y,X](t)$;
2. linearity: $[X,Y+Z](t) = [X,Y](t) + [X,Z](t)$ and $[X,cY](t) = c[X,Y](t)$.

5.6.2 On multidimensional Brownian motions

From the slides, we have

$$d[W_i, W_i](t) \equiv dW_i(t)dW_i(t) = dt,$$

and

$$d[W_i, W_j](t) \equiv dW_i(t)dW_j(t) = 0, \text{ for } i \neq j. \tag{5.16}$$

In addition, we recall that

$$d[W_i, \cdot](t) \equiv dW_i(t)dt = 0.$$

Here, to show (5.16), it is equivalent to prove

$$[W_i, W_j](t) = \lim_{\|\Pi\|\to 0} \sum_{k=0}^{n-1} (W_i(t_{k+1}) - W_i(t_k))(W_j(t_{k+1}) - W_j(t_k)) = 0.$$

which can be regarded as an excellent exercise. How? Find it in Shreve? Here is an interesting exercise enhancing our understanding of multidimensional Brownian motions. Suppose $\{(B_1(t), B_2(t), \cdots, B_d(t))\}$ is a d dimensional Brownian motion with covarian matrix $(t-s)\Sigma$, where $\Sigma = (\rho_{ij})_{d\times d}$. Denote by

$$B(t) = (B_1(t), B_2(t), \cdots, B_d(t))^\top$$

a column random vector. In this case, we knew that

$$t\Sigma = \mathrm{cov}(B(t), B(t)) \equiv (\mathrm{cov}(B_i(t), B_j(t)))_{d\times d}.$$

Thus, we have the following elementwise form

$$t\rho_{ij} = \mathrm{cov}(B_i(t), B_j(t)) = \mathbb{E}[B_i(t)B_j(t)].$$

So, it is easy to see that

$$\rho_{ij} = \frac{\mathrm{cov}(B_i(t), B_j(t))}{\sqrt{\mathrm{var}(B_i(t))}\sqrt{\mathrm{var}(B_j(t))}} = \frac{\mathbb{E}[B_i(t)B_j(t)]}{\sqrt{\mathrm{var}(B_i(t))}\sqrt{\mathrm{var}(B_j(t))}} = \mathrm{corr}(B_i(t), B_j(t)).$$

That is to say ρ_{ij} represents the correlation between $B_i(t)$ and $B_j(t)$. Now, we claim that ρ_{ij} renders the covariation between $B_i(t)$ and $B_j(t)$. A simple proof is given as follows. By the Cholesky decomposition, we have the following construction of the two-dimensional Brownian motion $\{(B_i(t), B_j(t))\}$, i.e.,

$$\begin{aligned} B_i(t) &= Z_i(t), \\ B_j(t) &= \rho_{ij} Z_i(t) + \sqrt{1-\rho_{ij}^2} Z_j(t), \end{aligned}$$

where Z_i and Z_j are two independent standard one-dimensional Brownian motions. Thus, we have

$$\begin{aligned} [B_i, B_j](t) &= [Z_i, \rho_{ij} Z_i + \sqrt{1-\rho_{ij}^2} Z_j](t) \\ &= \rho_{ij}[Z_i, Z_i](t) + \sqrt{1-\rho_{ij}^2}[Z_i, Z_j](t) = \rho_{ij} t, \end{aligned}$$

where the second equality follows from the linearity property of covariation. Equivalent notation is

$$d[B_i, B_j](t) = dB_i(t)dB_j(t) = \rho_{ij} t.$$

5.6.3 A quick path to multivariate stochastic calculus

We start from the two-dimensional case. Suppose we have two Itô processes defined by

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

We begin by investigating their quadratic variation and covariation. Without proof, we provide the following useful property as a tool for deriving what we need here. Suppose we have the following two stochastic integrals w.r.t. general maringales (as defined in (5.10)):

$$
\begin{aligned}
I^M(t) &= \int_0^t \Delta_1(u)dM(u), \\
I^N(t) &= \int_0^t \Delta_2(u)dN(u),
\end{aligned}
$$

where M and N are two martingales. For example, we can choose M and N as two Brownian motions, e.g.,

$$M(t) = W_1(t) \text{ and } N(t) = W_2(t).$$

We have their covariation as

$$[I^M, I^N] = \int_0^t \Delta_1(u)\Delta_2(u)d[M,N](u). \tag{5.17}$$

In particualr, when $M = N$ and $\Delta_1 = \Delta_2$, this property reduce to (5.14). The principle behind it follows much the same way as what we have seen before. So, feel free to use it without proof. To summarise, the aforementioned properties will serve as a foundation for us to study multivariate stochastic calculus.

Indeed, similar to the case of (5.6), du integrals contribute nothing to the integrals. Thus, we can use the linearity property of covariation to obtain that

$$
\begin{aligned}
& [X,X](t) \\
= & \left[\int_0^\cdot \sigma_{11}(u)dW_1(u) + \int_0^\cdot \sigma_{12}(u)dW_2(u), \int_0^\cdot \sigma_{11}(u)dW_1(u) + \int_0^\cdot \sigma_{12}(u)dW_2(u)\right](t) \\
= & \left[\int_0^\cdot \sigma_{11}(u)dW_1(u), \int_0^\cdot \sigma_{11}(u)dW_1(u)\right](t) + \left[\int_0^\cdot \sigma_{11}(u)dW_1(u), \int_0^\cdot \sigma_{12}(u)dW_2(u)\right](t) \\
& + \left[\int_0^\cdot \sigma_{12}(u)dW_2(u), \int_0^\cdot \sigma_{11}(u)dW_1(u)\right](t) + \left[\int_0^\cdot \sigma_{12}(u)dW_2(u), \int_0^\cdot \sigma_{12}(u)dW_2(u)\right](t),
\end{aligned}
$$

where using the covariation property (5.17) we obtain that

$$
\begin{aligned}
\left[\int_0^\cdot \sigma_{11}(u)dW_1(u), \int_0^\cdot \sigma_{11}(u)dW_1(u)\right](t) &= \int_0^t \sigma_{11}(u)\sigma_{11}(u)d[W_1,W_1](t) = \int_0^t \sigma_{11}(u)^2dt,\\
\left[\int_0^\cdot \sigma_{11}(u)dW_1(u), \int_0^\cdot \sigma_{12}(u)dW_2(u)\right](t) &= \int_0^t \sigma_{11}(u)\sigma_{12}(u)d[W_1,W_2](t) = 0,\\
\left[\int_0^\cdot \sigma_{12}(u)dW_2(u), \int_0^\cdot \sigma_{11}(u)dW_1(u)\right](t) &= \int_0^t \sigma_{12}(u)\sigma_{11}(u)d[W_1,W_2](t) = 0,\\
\left[\int_0^\cdot \sigma_{12}(u)dW_2(u), \int_0^\cdot \sigma_{12}(u)dW_2(u)\right](t) &= \int_0^t \sigma_{12}(u)\sigma_{12}(u)d[W_2,W_2](t) = \int_0^t \sigma_{12}(u)^2dt.
\end{aligned}
$$

Thus, we have

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

Similarly, we have

$$[Y,Y](t) = \int_0^t \sigma_{21}(u)^2dt + \int_0^t \sigma_{22}(u)^2dt,$$

and

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

The above latter two properties can be regarded as excellent exercises. Using differential notation, we get

$$
\begin{aligned}
dX(t)dX(t) &= d[X,X](t) = [\sigma_{11}(t)^2 + \sigma_{12}(t)^2]dt,\\
dY(t)dY(t) &= d[Y,Y](t) = [\sigma_{21}(t)^2 + \sigma_{22}(t)^2]dt,\\
dX(t)dY(t) &= d[X,Y](t) = [\sigma_{11}(t)\sigma_{21}(t) + \sigma_{21}(t)\sigma_{22}(t)]dt.
\end{aligned}
$$

Now, it is easy to follow the slides to discuss the two-dimensional Itô formula.

Here is an exercise. Assuming we know

$$W_3(t) = \rho W_1(t) + \sqrt{1-\rho^2}W_2(t),$$

we can use the Itô product formula (a straightforward implication of the two-dimensional Itô formula) to show that $\mathrm{corr}(W_1(t), W_3(t)) = \rho$. Indeed, we have

$$dW_1(t)W_3(t) = W_1(t)dW_3(t) + W_3(t)dW_1(t) + d[W_1, W_3](t).$$

Thus, we have

$$\begin{aligned}\mathbb{E}[W_1(t)W_3(t)] &= \mathbb{E}\left(\int_0^t W_1(s)dW_3(s) + \int_0^t W_3(s)dW_1(s) + [W_1, W_3](t)\right)\\ &= \mathbb{E}\left(0 + 0 + [W_1, W_3](t)\right)\\ &= \mathbb{E}[W_1, W_3](t),\end{aligned}$$

where the second equality follows from the martingale property of stochastic integrals. We note that

$$[W_1, W_3](t) = [W_1, \rho W_1 + \sqrt{1-\rho^2}W_2](t) = \rho t.$$

It follows that

$$\mathbb{E}[W_1(t)W_3(t)] = \mathbb{E}[W_1, W_3](t) = \rho t.$$

Thus, it is easy to have $\mathrm{corr}(W_1(t), W_3(t)) = \rho$.

Finally, we can generalize the Itô formula to any arbitrary n dimension. Let

$$X(t) = (X_1(t), X_2(t), ..., X_n(t)).$$

Suppose that for $i = 1, 2, ..., n$

$$X_i(t) = X_i(0) + \int_0^t \Theta_i(u)du + \sum_{k=1}^{d}\int_0^t \sigma_{ik}(u)dW_k(u).$$

Following the similar manner, we obtain that

$$dX_i(t)dX_j(t) = d[X_i, X_j](t) = \sum_{k=1}^{d}\int_0^t \sigma_{ik}(u)\sigma_{jk}(u)du. \tag{5.18}$$

This is an excellent exercise. Now, using (5.18) as input, we have the following Itô formula for n dimension: denote by

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

Indeed, directly working on local martingales, we have the following more general version of the Itô formula. Let X be a semimartingale, i.e., a process of the form

$$X(t) = X(0) + M(t) + V(t),\ 0 \le t < \infty$$

where M is a local martingale with continuous sample paths, and V a process with continuous sample paths of finite first variation. Then, for every function $f : \mathbb{R} \to \mathbb{R}$ of class C^2, we have

$$f(X(t)) = f(X(0)) + \int_0^t f'(X(s))dM(s) + \int_0^t f'(X(s))dV(s) + \frac{1}{2}\int_0^t f''(X(s))d\left\langle M\right\rangle(s),\ 0 \le t < \infty.$$

More generally, let $X = (X^{(1)},...,X^{(d)})$ be an $\mathbb{R}^d$- valued process with components $X_i(t) = X_i(0) + M_i(t) + V_i(t)$ of the above type, and $f : \mathbb{R}^d \to \mathbb{R}$ a function of class C^2. We have then

$$\begin{aligned} f(X(t)) &= f(X(0)) + \sum_{i=1}^{d}\int_0^t \frac{\partial f}{\partial x_i}(X(s))dM_i(s) + \sum_{i=1}^{d}\int_0^t \frac{\partial f}{\partial x_i}(X(s))dV_i(s) + \\ &\quad + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X(s))d\left\langle M_i, M_j\right\rangle(s). \end{aligned}$$

5.7 Multidimensional Brownian motion with stochastic correlations

Multidimension Brownian motion with stochastic correlations plays an important role in financial modeling. Indeed, correlated Brownian motions can be expressed using independent Brownian motions. We start from a simple two-dimensional case to clarify how correlated Brownian motions can be decomposted into independent Brownian motions. Suppose $B_1(t)$ and $B_2(t)$ are Brownian motions and

$$dB_1(t)\, dB_2(t) = \rho(t)\, dt,$$

where ρ is a stochastic process taking values strictly between -1 and 1. Define processes $W_1(t)$ and $W_2(t)$ such that

$$\begin{aligned} B_1(t) &= W_1(t), \\ B_2(t) &= \int_0^t \rho(s)\, dW_1(s) + \int_0^t \sqrt{1-\rho^2(s)}dW_2(s), \end{aligned}$$

As an excellent exercise, using the Levy characterization theorem and properties of stochastic integrals, it is easy to prove that $W_1(t)$ and $W_2(t)$ are independent Brownian motions. One may wonder how this property looks in a multidimensional setting. We can show the following extension. 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

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

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

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

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

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

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

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

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

$$\delta_{ik} := \begin{cases} 1 \; if \; i = k, \\ 0 \; if \; i \neq k, \end{cases}$$

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

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

The above results can be viewed as natural generalizations of the Cholesky decomposition for multidimensional Brownian motion with constant correlation matrix.

On the other hand, we can create 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

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

We can show that $B_1(t), \ldots, B_m(t)$ be m one-dimensional Brownian motions satisfying $dB_i(t)\, dB_k(t) = \rho_{ik}(t)$, where

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

6 Stochastic differential equations

6.1 Definition of multidimensional SDEs

Similar to one-dimensional SDEs, we can define multidimensional SDEs. A multi-dimensional stochastic differential equation is governed by

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t) \tag{6.1}$$

where $X(t)$ is an m dimensional stochastic process

$$X(t) = (X_1(t), X_2(t), \cdots, X_m(t))^\top;$$

$\mu(t,x)$ is an m dimensional vector

$$\mu(t,x) = (\mu_1(t,x), \mu_2(t,x), \cdots, \mu_m(t,x))^\top;$$

$\sigma(t,x)$ is an $m \times n$ matrix

$$\sigma(t,x) = \begin{pmatrix} \sigma_{11}(t,x), \sigma_{12}(t,x), \ldots, \sigma_{1d}(t,x) \\ \sigma_{21}(t,x), \sigma_{22}(t,x), \ldots, \sigma_{2d}(t,x) \\ \cdots \\ \sigma_{m1}(t,x), \sigma_{m2}(t,x), \ldots, \sigma_{md}(t,x) \end{pmatrix};$$

$W(t)$ is a standard m dimensional Brownian motions with

$$W(t) = (W_1(t), W_2(t), \cdots, W_m(t))^\top.$$

Suppose that the coefficients of the equation satisfy the Lipschitz and linear growth conditions, i.e.,

$$||\mu(t,x) - \mu(t,y)|| + ||\sigma(t,x) - \sigma(t,y)|| \leq K||x-y||, \forall x, y \in R^d, \tag{6.2}$$

and

$$||b(t,x)|| + ||\sigma(t,x)|| \leq K(1 + ||x||), \forall x \in R^d, \tag{6.3}$$

for some real $K > 0$. Then, there exists a unique process X that satisfies (6.1); it has continuous sample paths, is adapted to the filtration $\{\mathcal{F}_W(t)\}$ of the driving Brownian motion W, it is a Markov process. The idea in the proof of the aforementioned existence and uniqueness is to mimic the procedure followed in ordinary differential equations, i.e., to consider the "Picard iterations"

$$X^{(0)} \equiv \eta, X^{(k+1)}(t) = \eta + \int_0^t b(s, X^{(k)}(s))ds + \int_0^t \sigma(s, X^{(k)}(s))dW(s)$$

for $k = 0, 1, 2, ...$ The conditions (6.2) and (6.3) then guarantee that the sequence of continuous processes $\{X^{(k)}\}_{k=0}^\infty$ converges to a continuous process X, which is the unique solution of the equation (6.1); they also imply that the sequence $\{X^{(k)}\}_{k=0}^\infty$ and the solution X satisfy moment growth conditions of the type

$$E||X(t)||^{2\lambda} \leq C_{\lambda,T}(1 + E||\eta||)^{2\lambda}, \forall 0 \leq t \leq T$$

for any real numbers $\lambda \geq 1$ and $T > 0$, where $C_{\lambda,T}$ is a positive constant depending only on λ, T and on the constant K of (6.2) and (6.3).

6.2 Explicitly solving the generalized geometric Brownian motion SDE

For $S(t)$ satisfies

$$dS(t) = \alpha(t)S(t)dt + \sigma(t)S(t)dW(t), \tag{6.4}$$

where $\alpha(t)$ and $\sigma(t)$ are processes adapted to the filtration $\{\mathcal{F}(t)\}$ associated with the Brownian motion $\{W(t)\}$. We note that, for the function

$$f(x) = \log x,$$

we calculus to obtain that

$$f'(x) = \frac{1}{x} \text{ and } f''(x) = -\frac{1}{x^2}.$$

Thus, by applying Itô formula, we obtain that

$$\begin{aligned}
d\log S(t) &= \frac{1}{S(t)}dS(t) - \frac{1}{2}\frac{1}{S^2(t)}d[S,S](t) \\
&= \frac{1}{S(t)}\left[\alpha(t)S(t)dt + \sigma(t)S(t)dW(t)\right] - \frac{1}{2}\frac{1}{S^2(t)}\sigma(t)^2 S(t)^2 dt \\
&= \left(\alpha(t) - \frac{1}{2}\sigma^2(t)\right)dt + \sigma(t)dW(t)
\end{aligned}$$

Integration of both sides yields that

$$\log S(t) - \log S(0) = \int_0^t \left(\alpha(u) - \frac{1}{2}\sigma^2(u)\right) dS + \int_0^t \sigma(u)dW(u),$$

which can be written equivalently as

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

6.3 Compute $\mathbb{E}R(t)$ and Var$R(t)$ for the CIR model

The Cox-Ingersoll-Ross model is given by

$$dR(t) = (\alpha - \beta R(t))dt + \sigma\sqrt{R(t)}dW(t), \; R(0) = r_0, \tag{6.5}$$

where α, β, σ are constants. Although we cannot derive an explicit solution for (6.5), we can use the Itô formula and solutions to ODEs to compute $\mathbb{E}R(t)$ and $\mathrm{Var}R(t)$ for the CIR Model.

We solve $\mathbb{E}R(t)$ first. Integretion of the both sides of (6.5) yields

$$R(t) - R(0) = \int_0^t (\alpha - \beta R(s))ds + \int_0^t \sigma\sqrt{R(s)}dW(s).$$

Taking expectation of both sides yields

$$\mathbb{E}R(t) - R(0) = \mathbb{E}\left[\int_0^t (\alpha - \beta R(s))ds\right] + E\left(\int_0^t \sigma\sqrt{R(s)}dW(s)\right) \tag{6.6}$$

Recalling that the Itô integral $\int_0^t \sigma\sqrt{R(s)}dW(s)$ is a martingale, we have

$$\mathbb{E}\left(\int_0^t \sigma\sqrt{R(s)}dW(s)\right) = 0.$$

Thus, we have

$$\mathbb{E}R(t) - R(0) = \mathbb{E}\left[\int_0^t (\alpha - \beta R(s))ds\right].$$

An differential form of this equation reads

$$\begin{aligned} d\mathbb{E}R(t) &= [\alpha - \beta\mathbb{E}R(t)]dt, \\ \mathbb{E}R(0) &= r_0. \end{aligned}$$

Denote by $y(t) = E(R(t))$. We obtain an ODE with initial condition:

$$\begin{aligned} y'(t) &= \alpha - \beta y(t), \\ y(0) &= r_0 \end{aligned} \tag{6.7}$$

Multiplying $e^{\beta t}$ on the both sides of (6.7), we have

$$e^{\beta t}y'(t) + e^{\beta t}\beta y(t) = \alpha e^{\beta t},$$

which is equivalent to

$$[e^{\beta t}y(t)]' = \alpha e^{\beta t}.$$

Integrating the both sides, we have

$$e^{\beta t}y(t) = \int_0^t e^{\beta s}\alpha ds + C,$$

for some constants C. By plugging in $t = 0$, we have

$$y(0) = 0 + C = r_0.$$

Thus, we have

$$y(t) \equiv \mathbb{E}R(t) = e^{-\beta t}\left(r_0 + \int_0^t e^{\beta s}\alpha ds\right) = e^{-\beta t}r_0 + \frac{\alpha}{\beta}(1 - e^{-\beta t}) \tag{6.8}$$

Now, we solve for $\mathrm{Var}R(t)$. We beging by noticing that

$$\mathrm{var}R(t) = \mathbb{E}R^2(t) - (\mathbb{E}R(t))^2$$

As we already obtained $\mathbb{E}R(t)$, we still need $\mathbb{E}R^2(t)$. Applying the Itô Formula, we have

$$\begin{aligned} dR^2(t) &\equiv 2R(t)dR(t) + \frac{1}{2} \times 2d[R,R](t) \\ &\equiv 2R(t)[(\alpha - \beta R(t))dt + \sigma\sqrt{R(t)}dW(t)] + (\sigma\sqrt{R(t)})^2 dt \end{aligned}$$

Integating the both sides and taking expectations yield that

$$\begin{aligned} &\mathbb{E}R^2(t) - R^2(0) \\ = \;& 2\mathbb{E}\left(\int_0^t R(t)(\alpha - \beta R(t))dt\right) + 2\sigma\mathbb{E}\left(\int_0^t R(t)\sqrt{R(t)}dW(t)\right) + \mathbb{E}\left(\int_0^t (\sigma\sqrt{R(t)})^2 dt\right) \\ = \;& 2\int_0^t \mathbb{E}\left[R(t)(\alpha - \beta R(t))\right]dt + 2\sigma\mathbb{E}\left(\int_0^t R(t)\sqrt{R(t)}dW(t)\right) + \sigma^2\int_0^t \mathbb{E}R(t)dt \end{aligned}$$

Because the Itô integral $\int_0^t \sigma R(t)\sqrt{R(t)}dW(t)$ is a martingale, we have

$$\mathbb{E}\left(\int_0^t \sigma R(t)\sqrt{R(t)}dW(t)\right) = 0.$$

Differentiating the both sides, we have equations:

$$d\mathbb{E}R^2(t) = [(2\alpha+\sigma^2)\mathbb{E}R(t) - 2\beta\mathbb{E}R^2(t)]dt,\ \mathbb{E}R^2(0) = r_0^2.$$

Denote by $\mathbb{E}R^2(t) = y(t)$. We have

$$y'(t) = (2\alpha+\sigma^2)\mathbb{E}R(t) - 2\beta y(t),\ y(0) = r_0^2, \tag{6.9}$$

where $\mathbb{E}R(t)$ given in (6.8). By multiplying $e^{2\beta t}$ on each side of (6.9), we obtain that

$$e^{2\beta t}y'(t) + 2\beta e^{2\beta t}y(t) = e^{2\beta t}(2\alpha+\sigma^2)\mathbb{E}(R(t)).$$

Thus,

$$[e^{2\beta t}y(t)]' = e^{2\beta t}(2\alpha+\sigma^2)\mathbb{E}(R(t)).$$

Integrating the both sides, we have

$$e^{\beta t}y(t) = \int_0^t e^{2\beta s}(2\alpha+\sigma^2)\mathbb{E}(R(s))ds + C,$$

for some constant C. Thus, we have that

$$y(t) = \mathbb{E}R^2(t) = e^{-2\beta t}r_0^2 + \frac{2\alpha+\sigma^2}{\beta}\left(r_0 - \frac{\alpha}{\beta}\right)(e^{-\beta t} - e^{-2\beta t}) + \frac{\alpha(2\alpha+\sigma^2)}{2\beta^2}(1 - e^{-2\beta t}).$$

Hence, we have

$$\mathrm{var}R(t) = \mathbb{E}R^2(t) - (\mathbb{E}R(t))^2 = \frac{\sigma^2}{\beta}(e^{-\beta t} - e^{-2\beta t}) + \frac{\alpha\sigma^2}{2\beta^2}(1 - 2e^{-\beta t} + e^{-2\beta t}).$$

6.4 Solving the general one dimensional linear SDE

Consider the general one dimensional linear SDE of the following form

$$dX(t) = [a(t) + b(t)X(t)]dt + [\gamma(t) + \sigma(t)X(t)]dW(t),\ X(0) = X_0. \tag{6.10}$$

where $\{a(t)\}, \{b(t)\}, \{\gamma(t)\}$, and $\{\sigma(t)\}$ are stochastic processes (not necessarily to be constants or deterministic functions) adapted to the filtration $\mathcal{F}(t)$.

Applying the Itô formula, we will show that

$$X(t) = Y(t)Z(t),$$

where

$$Y(t) = \exp\left(\int_0^t \left[b(s) - \frac{1}{2}\sigma(s)^2\right] ds + \int_0^t \sigma(s)dW(s)\right)$$

and

$$Z(t) = X_0 + \int_0^t [a(s) - \gamma(s)\sigma(s)]\, Y(s)^{-1}ds + \int_0^t \gamma(s)Y(s)^{-1}dW(s).$$

We note that

$$\frac{dY(t)}{Y(t)} = b(t)dt + \sigma(t)dW(t).$$

Indeed, by using the Itô product formula, we obtain that

$$\begin{aligned}
dX(t) &= Z(t)dY(t) + Y(t)dZ(t) + dY(t)dZ(t) \\
&= Z(t)\left[b(t)Y(t)dt + \sigma(t)Y(t)dW(t)\right] \\
&\quad + Y(t)\left[[a(t) - \gamma(t)\sigma(t)]\, Y(t)^{-1}dt + \gamma(t)Y(t)^{-1}dW(t)\right] \\
&\quad + \sigma(t)Y(t)\gamma(t)Y(t)^{-1}dt \\
&= b(t)X(t)dt + \sigma(t)X(t)dW(t) \\
&\quad + \left[[a(t) - \gamma(t)\sigma(t)]\, dt + \gamma(t)dW(t)\right] + \sigma(t)\gamma(t)dt \\
&= [a(t) + b(t)X(t)]dt + [\gamma(t) + \sigma(t)X(t)]dW(t).
\end{aligned}$$

Assuming $a(t), b(t), \gamma(t)$, and $\sigma(t)$ are all deterministic functions, we can obtain $\mathbb{E}X(t))$ and $\mathrm{Var}(X(t))$ explicitly. Integration of the both sides of (6.10) yields that

$$X(t) = X_0 + \int_0^t [a(s) + b(s)X(s)]ds + \int_0^t [\gamma(s) + \sigma(s)X(s)]dW(s)$$

Taking expectation on the both sides, we obtain

$$\begin{aligned}\mathbb{E}X(t) &= X_0 + \int_0^t [a(s)ds + b(s)\mathbb{E}X(s)]ds + \mathbb{E}\left(\int_0^t [\gamma(s) + \sigma(s)X(s)]dW(s)\right) \\ &= X_0 + \int_0^t [a(s)ds + b(s)\mathbb{E}X(s)]ds,\end{aligned}$$

which is equivalent to an ODE initial value problem:

$$d\mathbb{E}X(t) = a(t)dt + b(t)\mathbb{E}X(t)dt,\ \mathbb{E}X(0) = X_0.$$

Using the integrating factor $\exp\left(-\int_0^t b(s)ds\right)$, we solve it as

$$\mathbb{E}X(t) = X_0 \exp\left(\int_0^t b(s)ds\right) + \int_0^t a(s)\exp\left(\int_s^t b(u)du\right)ds.$$

Because of

$$\mathrm{var}X(t) = \mathbb{E}X^2(t) - (\mathbb{E}X(t))^2,$$

we just need to calculate $\mathbb{E}X^2(t)$. For this purpose, we use the Itô formula to obtain the dynamics of $dX^2(t)$. Then, by taking expectation and using the previous result, it is straightforward to obtain $\mathbb{E}X^2(t)$ as the solution to an ODE initial value problem. We suggest it as an excellent exercise.

As an excellent exercise, we have an SDE

$$dX(t) = [a(t) + b(t)X(t)]dt + \sum_{j=1}^{d}\left[\sigma_j(t)X(t) + \gamma_j(t)\right]dW_j(t),\ X(0) = X_0, \tag{6.11}$$

where $\{W(t)\}$ is an d-dimensional Brownian motion, and the coefficients a, b, σ_j, σ_j are measurable, $\{\mathcal{F}(t)\}$ −adapted, almost surely locally bounded processes. Note that this SDE is a generalization of

(6.10) by adding multiple driving Brownian motions. Show that the unique solution of this equation is given by

$$X(t) = Y(t)\left[X_0 + \int_0^t Y(u)^{-1}\left(a(u) - \sum_{j=1}^{d}\sigma_j(u)\gamma_j(u)\right)du + \sum_{j=1}^{d}\int_0^t Y(u)^{-1}\gamma_j(u)\,dW_j(u)\right],$$

where we set

$$Y(t) \triangleq \exp\left(\int_0^t b(u)\,du + \sum_{j=1}^{d}\int_0^t \sigma_j(u)\,dW_j(u) - \frac{1}{2}\sum_{j=1}^{d}\int_0^t \sigma_j^2(u)\,du\right).$$

In particular, the solution of the equation

$$dX(t) = b(t)\,X(t)dt + \sum_{j=1}^{r}\sigma_j(t)\,X(t)dW_j(t)$$

is given by

$$X(t) = X_0\exp\left(\int_0^t\left(b(u) - \frac{1}{2}\sum_{j=1}^{d}\sigma_j^2(u)\right)du + \sum_{j=1}^{r}\int_0^t \sigma_j(u)\,dW_j(u)\right).$$

Here is an interesting question: if the one-dimensional linear SDE (6.11) is generalized to a multi-dimensional case, is it possible to give a solution in closed-form? Or, in what case it is possible?

7 Connections with partial differential equations

7.1 Connect Brownian motion and backward heat equation

Consider the heat equation

$$u_\tau(\tau, x) = \frac{1}{2}u_{zz}(\tau, x),$$

for all $\tau \in [0, +\infty)$ and $x \in \mathcal{R}$ with the initial condition

$$u(0, x) = f(x),$$

where $f(x)$ is continuous and uniformly bounded. We have proved that it has a unique solution:

$$u(t,x) = \int_{-\infty}^{+\infty} f(y)G(x,y,t)dy \tag{7.1}$$

where $G(x,y,t)$ is defined as

$$G(x,y,t) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(x-y)^2}{2t}\right). \tag{7.2}$$

If we consider a function $v(t,x) := u(T-t,x)$, it is obvious that

$$\frac{\partial v}{\partial t} = -\frac{\partial u}{\partial t}, \frac{1}{2}\frac{\partial^2 v}{\partial x^2} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2}.$$

Thus, it is straightforward to have

$$\frac{\partial v}{\partial t} + \frac{1}{2}\frac{\partial^2 v}{\partial x^2} = 0, \tag{7.3}$$

with a terminal condition

$$v(T,x) = f(x), \tag{7.4}$$

On the other hand, using (7.1) , we find that $v(t,x)$ admits the following explicit solution

$$v(t,x) = u(T-t,x) = \int_{-\infty}^{+\infty} f(y)\frac{1}{\sqrt{2\pi(T-t)}} \exp\left(-\frac{(x-y)^2}{2(T-t)}\right) dy. \tag{7.5}$$

For a standard one-dimensional Brownian motion $\{B(t)\}$, it is straightforward to have

$$\mathbb{E}[f(B(T))|B(t)=x] = \frac{1}{\sqrt{2\pi(T-t)}} \int_{-\infty}^{+\infty} f(y) \exp\left(-\frac{(y-x)^2}{2(T-t)}\right) dy, \tag{7.6}$$

by using the definition of Brownian motion. Thus, from (7.5) and (7.6), we obtain that

$$v(t,x) = \mathbb{E}[f(B(T))|B(t)=x]. \tag{7.7}$$

The previous discussion is based on calculus. We can obtain that, if a function $v(t,x)$ is defined as

(7.7), by using the martingale property of the Brownian motion, we can show that it solves the backward heat equation (7.3) together with the terminal condition (7.4). Using the definition of conditional expectation and the Markov property of Brownian motion, it is easy to see that

$$v(t, B(t)) = \mathbb{E}[f(B(T))|B(t)] = \mathbb{E}[f(B(T))|\mathcal{F}(t)],$$

where $\{\mathcal{F}(t)\}$ denotes the filtration generated by the Brownian motion. Obviously, $\{\mathbb{E}[f(B(T))|\mathcal{F}(t)]\}$ is a martingale; so is $\{v(t, B(t))\}$. By using the Itô formula, we have

$$\begin{aligned} dv(t, B(t)) &= v_t(t, B(t))dt + v_x(t, B(t))dB(t) + \frac{1}{2}v_{xx}(t, B(t))dt \\ &= \left[v_t(t, B(t)) + \frac{1}{2}v_{xx}(t, B(t))\right] dt + v_x(t, B(t))dB(t). \end{aligned}$$

For the Itô process

$$v(t, B(t)) = v(0, 0) + \int_0^t \left[v_s(s, B(s)) + \frac{1}{2}v_{xx}(s, B(s))\right] ds + \int_0^t v_x(s, B(s))dB(s),$$

it is a martingale if and only if its dt term is zero. (It is a general result that an Itô process is a martingale if and only if its dt term is zero.) Thus, the PDE (7.3) is satisfied and obviously the terminal condition (7.4) is satisfied

$$v(T, x) = \mathbb{E}[f(B(T))|B(T) = x] = f(x).$$

7.2 Itô formula applying to a diffusion

Suppose we apply the Itô formula on the marginal of an SDE. How to use a very compactly written generator to express the Itô formula? Suppose we have an m dimensional SDE for $\{X(t)\}$ driven by a d dimensional Brownian motion $\{W(t)\}$

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t). \tag{7.8}$$

Indeed, by the previous discussions, we note that

$$d[X_i, X_j](t) = \sum_{k=1}^{d} \sigma_{ik}(X(t))\sigma_{jk}(X(t))dt = a_{ij}(X(t))dt,$$

where the matrix

$$(a_{ij}(t,x))_{m\times m} = \sigma(t,x)\sigma(t,x)^{\top}$$

is usually called diffusion matrix. Using the multidimensional Ito formula, we have

$$\begin{aligned}
& du(t, X(t)) \\
= \;& u_t(t, X(t))dt + \sum_{i=1}^{m} u_{x_i}(t, X(t))dX_i(t) + \frac{1}{2}\sum_{i,j=1}^{m} u_{x_i x_j}(t, X(t))d[X_i, X_j](t) \\
= \;& u_t(t, X(t))dt + \nabla u(t, X(t)) \left[\mu(t, X(t))dt + \sigma(t, X(t))dW(t)\right] + \frac{1}{2}\sum_{i,j=1}^{m} u_{x_i x_j}(t, X(t))d[X_i, X_j](t) \\
= \;& u_t(t, X(t))dt + \nabla u(t, X(t)) \left[\mu(t, X(t))dt + \sigma(t, X(t))dW(t)\right] + \frac{1}{2}\sum_{i,j=1}^{m} u_{x_i x_j}(t, X(t))a_{ij}(t, X(t))dt \\
= \;& \left[u_t(t, X(t)) + \mathcal{A}_t u(t, X(t))\right] dt + \nabla u(t, X(t))\sigma(t, X(t))dW(t),
\end{aligned}$$

where the generator $\mathcal{A}_t$ is a second-order differential operator given by

$$\mathcal{A}_t = \frac{1}{2}\sum_{i,j=1}^{m} a_{ij}(t,x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{m} \mu_i(t,x)\frac{\partial}{\partial x_i}. \tag{7.9}$$

We call it the infinitesimal generator for SDE (7.8). The terminology “infinitesimal generator” is given because of the following properties, which can be proved as an excellent exercise under some technical conditions (e.g., μ and a are bounded and continuous)

$$\lim_{h\to 0}\frac{1}{h}\mathbb{E}\left[X_i(t+h) - x_i | X(t) = x\right] = \mu_i\left(t, x\right)$$

and

$$\lim_{h\to 0}\frac{1}{h}\mathbb{E}\left[\left(X_i(t+h) - x_i\right)\left(X_k(t+h) - x_k\right)|X(t) = x\right] = a_{ik}\left(t, x\right)$$

for $1 \le i,\ k \le d$ hold for every $x \in \mathbb{R}^m$, and, more generally,

$$\lim_{h \downarrow 0} \frac{1}{h} \left[\mathbb{E} f\left(X(t+h)\right) - f\left(x\right) | X(t) = x \right] = \left(\mathcal{A}_t f\right)(x)\,;\ \forall x \in \mathbb{R}^m$$

hold for every $f \in C^2\left(\mathbb{R}^m\right)$ which is bounded and has bounded first- and second-order derivatives where the operator $\mathcal{A}_t f$ in (7.9).

Proof. Without loss of generality, we prove for the case $t = 0$. By the Itô formula, we have

$$f(X(h)) - f(x) = \sum_{i=1}^{d} \int_0^h \frac{\partial f(X(s))}{\partial x_i} dX^{(i)}(s) + \frac{1}{2} \sum_{i=1}^{d} \sum_{k=1}^{d} \int_0^h \frac{\partial^2 f(X(s))}{\partial x_i \partial x_k} d\langle X^{(i)}, X^{(k)} \rangle(s),$$

with

$$X^{(i)}(h) = x_i + \int_0^h \mu_i(s, X(s)) ds + \sum_{j=1}^{r} \int_0^h \sigma_{ij}(s, X(s)) dW^{(j)}(s),$$

and

$$\begin{aligned} \langle X^{(i)}, X^{(k)} \rangle(s) &= \sum_{j=1}^{r} \int_0^s \sigma_{ij}(s, X(s)) \sigma_{kj}(s, X(s)) ds \\ &= \int_0^s a_{ik}(s, X(s)) ds. \end{aligned}$$

Then

$$\begin{aligned} f(X(h)) - f(x) &= \sum_{i=1}^{d} \int_0^h \mu_i(s, X(s)) \frac{\partial f(X(s))}{\partial x_i} ds + \sum_{i=1}^{d} \sum_{j=1}^{r} \int_0^h \sigma_{ij}(s, X(s)) \frac{\partial f(X(s))}{\partial x_i} dW^{(j)}(s) \\ &\quad + \frac{1}{2} \sum_{i=1}^{d} \sum_{k=1}^{d} \int_0^h a_{ik}(s, X(s)) \frac{\partial^2 f(X(s))}{\partial x_i \partial x_k} ds \\ &= \int_0^h (\mathcal{A}_s f)(X(s)) ds + \sum_{i=1}^{d} \sum_{j=1}^{r} \int_0^h \sigma_{ij}(s, X(s)) \frac{\partial f(X(s))}{\partial x_i} dW^{(j)}(s). \end{aligned}$$

Taking expectation on both sides of the above equation and we obtain

$$E^x f(X(j)) - f(x) = E \int_0^h (\mathcal{A}_s f)(X(s)) ds,$$

which implies

$$\lim_{t\downarrow 0}\frac{1}{h}[E^x f(X(h)) - f(x)] = (\mathcal{A}_0 f)(x),$$

since $f \in C^2(\mathbb{R}^d)$ implies $\mathcal{A}f \in C(\mathbb{R}^d)$ and X has a continous path.

In particular, if we let $f(y) = y_i$, then $(\mathcal{A}_t f)(y) = \mu_i(t,y)$, and we obtain

$$\lim_{h\downarrow 0}\frac{1}{h}[E^x X^{(i)}(h) - x_i] = \mu_i(0,y).$$

If we let $f(y) = (y_i - x_i)(y_k - x_k)$, then $(\mathcal{A}_t f)(y) = a_{ik}(t,y) + \mu_i(t,y)(y_k - x_k) + \mu_k(t,y)(y_i - x_i)$, and we obtain

$$\lim_{h\downarrow 0}\frac{1}{h}E^x[(X^{(i)}(h) - x_i)(X^{(k)}(h) - x_k)] = a_{ik}(0,x).$$

7.3 Kolmogorov equations for transition density

Suppoe that

$$\mathbb{P}(X(t) \in A | X(s) = y] = \int_A p(t,x;s,y)dx.$$

Thus, $p(t,x;s,y)$ is the transition density of the process $\{X(t)\}$. We have the following property.

THEOREM 2. *Under appropriate technical conditions, the transition density $p(t,x;s,y)$ satisfies the following two PDEs: the Komogorov backward equation*

$$\left(\frac{\partial}{\partial s} + \mathcal{A}_s\right) p(t,x;\cdot,\cdot) = 0;$$

and the Komogorov forward equation (also known as Fokker-Planck equation)

$$\left(\frac{\partial}{\partial t} - \mathcal{A}_t^*\right) p(\cdot,\cdot;s,y) = 0.$$

Here, the differential operator $\mathcal{A}_t^*$ is the adjoint of the operator of (7.9), namely

$$\mathcal{A}_t^* f(x) \triangleq \frac{1}{2}\sum_{i=1}^d\sum_{j=1}^d \frac{\partial^2}{\partial x_i \partial x_j}[a_{ij}(t,x)f(x)] - \sum_{i=1}^d \frac{\partial}{\partial x_i}[b_i(t,x)f(x)].$$

We also have the initial conditions complementary to the equations

$$\lim_{t \to s} p(t, x; s, y) = \delta(x - y),$$

where $\delta(\cdot)$ is the celebrated Dirac Delta function.

The operator is called $\mathcal{A}_t^*$ the adjoint operator of $\mathcal{A}_t$ for the following reason:

$$\langle \mathcal{A}_t f, \phi \rangle_{L^2} = \langle f, \mathcal{A}_t^* \phi \rangle_{L^2}, \tag{7.10}$$

for any smooth-enough functions f and ϕ. Please prove this property as an exercise using integration-by-parts in calculus.

Proof. According to Feynman-Kac theorem, for any function $h : \mathbb{R}^m \to \mathbb{R}$ satisfying condition

$$|g(x)| \leq C(1 + \|x\|^p), \forall x \in \mathbb{R}^m \tag{7.11}$$

for some $C > 0$ and $p \geq 1$, we have

$$g(s, y) = E^{s,y} h(X_t) = \int_R h(x) p(t, x; s, y) dx \tag{7.12}$$

satisfies the PDE

$$\left(\frac{\partial}{\partial s} + \mathcal{A}_s \right) g(s, y) = 0.$$

Plugging in (7.12), we obtain that

$$\int_R h(x) \left(\frac{\partial}{\partial s} + \mathcal{A}_s \right) p(t, x; s, y) dx = 0.$$

This holds for all functions $h(x)$ satisfying condition (7.11). So, we obtain that

$$\left(\frac{\partial}{\partial s} + \mathcal{A}_s \right) p(t, x; s, y) = 0.$$

For any $b > 0$, define a function $h_b(x)$ satisfying conditions as below:

$$h_b(x) = 0 \text{ for } x \le 0,\ h_b'(x) = 0 \text{ for } x \ge b,\ h_b(b) = h_b'(b) = 0, \tag{7.13}$$

thus $h_b(x) = 0$ for $x \le 0$ or $x \ge b$. Let $X(s) = x \in (0, b)$. By the Ito formula, we have

$$h_b(X_t) = h_b(X_s) + \int_s^t \mathcal{A}_u h_b(X_u) du.$$

Then, we have

$$E^{s,x} h_b(X_t) = h_b(x) + \int_s^t E^{s,x} \left[\mathcal{A}_u h_b(X_u)\right] du.$$

Using the transition density, we get that

$$\begin{aligned}
\int_0^b h_b(y) p(t, y; s, x) dy &= h_b(x) + \int_s^t \int_0^b \mathcal{A}_u h_b(y) p(u, y; s, x) dy du \\
&= h_b(x) + \int_s^t \int_0^b \left(\sum_{i=1}^m \mu_i(u, y) \frac{\partial}{\partial y_i} h_b(y) \right) p(u, y; s, x) dy du \\
&\quad + \int_s^t \int_0^b \left(\frac{1}{2} \sum_{i,j=1}^m a_{ij}(u, y) \frac{\partial^2}{\partial y_i \partial y_j} h_b(y) \right) p(u, y; s, x) dy du.
\end{aligned}$$

Via integration-by-parts, we obtain that

$$\begin{aligned}
\int_0^b h_b(y) p(t, y; s, x) dy &= h_b(x) - \int_s^t \int_0^b \sum_{k=1}^m \frac{\partial}{\partial y_k} \left(\mu_k (u, y) p(u, y; s, x)\right) h_b(y) dy du \\
&\quad + \int_s^t \int_0^b \frac{1}{2} \sum_{i,j=1}^m \frac{\partial^2}{\partial y_i \partial y_j} \left(a_{ij} (u, y) p(u, y; s, x)\right) h_b(y) dy du.
\end{aligned}$$

Differentiating w.r.t. time variable t on the both sides, we obtain

$$\int_0^b h_b(y) \left(\frac{\partial}{\partial t} - \mathcal{A}_t^* \right) p(t, y; s, x) dy = 0$$

for all functions $h_b(y)$ under condition (7.13). Thus, we have

$$\left(\frac{\partial}{\partial t} - \mathcal{A}_t^*\right) p(t, y; s, x) = 0. \ \Box$$

Exercise: Seek for an alternative more natural proof for the Komogorov forward equation.

7.4 Feynman-Kac theorem: a general version

In what follows, we propose a general version of the Feynman-Kac theorem.

THEOREM 3. *Assume conditions (6.2) and (6.3) hold. Let $f : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ and polynomial growth condition*

$$\max_{0 \le t \le T} |f(t, x)| + |g(x)| \le C(1 + ||x||^p), \ \forall x \in R^d \tag{7.14}$$

for some $C > 0, p \ge 1$, let $k : [0, T] \times \mathbf{R}^d \to [0, \infty)$ be continuous, and suppose that the Cauchy problem

$$\begin{aligned} \frac{\partial V}{\partial t} + \mathcal{A}_t V + f &= kV, \ \text{in } [0, T) \times \mathbb{R}^d \\ V(T, \cdot) &= g, \ \text{in } \mathbb{R}^d \end{aligned}$$

has a solution $V : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ which is continuous on its domain, of class $C^{1,2}$ on $[0, T] \times \mathbb{R}^d$, and satisfies a growth condition of the type (6.3). The function V admits then the Feynman-Kac representation

$$V(t, x) = E\left[\int_t^T e^{-\int_t^\theta k(u, X(u))du} f(\theta, X(\theta))d\theta + g(X(T))e^{-\int_t^T k(u, X(u))du}\right] \tag{7.15}$$

for $0 \le t \le T, x \in R^d$ in terms of the solution X of the stochastic integral equation

$$X(\theta) = x + \int_t^\theta \mu(s, X(s))ds + \int_t^\theta \sigma(s, X(s))dW(s), t \le \theta \le T.$$

This theorem actually reveals both the existence and uniqueness of the PDE solution. The existence can be easily proved. The techniques are similar to those discussed in Shreve's book. We just need to prove that $V(t, x)$ of the form (7.15) solves the Cauchy problem. Here we apply the very routine method: setting up a Levy martingale, applying the Itô formula and then letting the dt term be zero. Denote by $\mathcal{F}(t)$ the filtration generated by the Brownian motion. First we define

$$\beta(t) := e^{-\int_0^t k(u,X(u))du}. \tag{7.16}$$

Due to the Markov property of $X(t)$, $V(t, X(t))$ can be expressed as

$$V(t, X(t)) = E\left[\int_t^T \frac{\beta(s)}{\beta(t)} f(s, X(s))ds + g(X(T))\frac{\beta(T)}{\beta(t)}|\mathcal{F}(t)\right].$$

So, we have

$$M(t) := \beta(t)V(t, X(t)) + \int_0^t \beta(s)f(s, X(s))ds \equiv E\left[\int_0^T \beta(s)f(s, X(s))ds + g(X(T))\beta(T)|\mathcal{F}(t)\right] \tag{7.17}$$

is a Levy martingale. According to the Itô formula, plugging in $d\beta(t) = -k(t, X_t)\beta(t)dt$, we have

$$\begin{aligned} dM(t) &= \beta(t)\left[-k(t, X(t))V(t, X(t)) + \frac{\partial V}{\partial t}(t, X(t)) + \mathcal{A}_t V(t, X(t)) + f(t, X(t))\right] dt \\ &\quad + \beta(t)\sum_{k=1}^{d}\sum_{i=1}^{m} \frac{\partial V}{\partial x_i}(t, X(t))\sigma_{ik}(t, X(t))dW_t^{(k)}. \end{aligned} \tag{7.18}$$

According to the general result that an Itô process is a martingale if and only if its dt term is zero, we have

$$-kV + \frac{\partial V}{\partial t} + \mathcal{A}_t V + f = 0 \text{ on } [0, T) \times \mathbb{R}^m.$$

Besides, the terminal condition

$$V(T, x) = E\left[g(X_T)|X(T) = x\right] = g(x)$$

is obviously satisfied. Thus the existence is proved.

The proof of uniquesness relies on the growth condition (7.14) the following Lemma. If the process X satisfies the equation (7.8). and $\beta(t) \triangleq \exp(-\int_0^t K(u)du)$ for some measurable, adapted and nonnegative process K; then, the process

$$M^f(t) \triangleq \beta(t)f(t,X(t)) - f(0,X(0)) - \int_0^t \left(\frac{\partial f}{\partial s} + \mathcal{A}_s f - K(s)f\right)(s,X(s))\beta(s)ds, 0 \le t < \infty$$

is a local martingale (square-integrable martingale, if f is of compact support) with continuous sample paths, and can be represented actually as

$$\sum_{i=1}^d \sum_{k=1}^n \int_0^t \frac{\partial f(s,X(s))}{\partial x_i} \sigma_{ik}(s,X(s))\beta(s)dW^{(k)}(s).$$

We will prove that the solution of the Cauchy problem must be expressed as (7.15).

Since the PDE

$$\frac{\partial V}{\partial t} + \mathcal{A}_t V + f = kV$$

holds on $[0,T) \times \mathbb{R}^m$, it follows from (7.18) that

$$dM(t) = \beta(t) \sum_{k=1}^d \sum_{i=1}^m \frac{\partial V}{\partial x_i}(t,X_t)\sigma_{ik}(t,X_t)dW_t^{(k)},$$

i.e., $M(t)$ is an Itô integral:

$$M(t) = \sum_{k=1}^d \sum_{i=1}^m \int_0^t \frac{\partial V}{\partial x_i}(s,X_s)\sigma_{ik}(s,X_s)\beta(s)dW_s^{(k)}.$$

If we always suppose that $M(t)$ is square integrable, then it is a martingale. (For more strict proof, refer to p.366-367 Karatzas and Shreve [1].) According to the definition of martingale, we have

$$E\left[M(T)|\mathcal{F}(t)\right] = M(t).$$

Recall that

$$M(t) = \beta(t)V(t,X(t)) + \int_0^t \beta(s)f(s,X(s))ds$$

Noting the terminal condition $V(T,x) = g(x)$, we can get

$$\beta(t)V(t, X_t(t)) + \int_0^t \beta(s)f(s, X(s))ds = E\left[\int_0^T \beta(s)f(s, X(s))ds + g(X(T))\beta(T)|\mathcal{F}(t)\right].$$

Subtracting $\int_0^t \beta(s)f(s, X(s))ds$ from both sides, and dividing both sides of the equation by $\beta(t)$, we get

$$V(t, X(t)) = E\left[\int_t^T e^{-\int_t^s k(u,X(u))du} f(s, X(s))ds + g(X(T))e^{-\int_t^T k(u,X(u))du}|\mathcal{F}(t)\right].$$

By the Markov property of $X(t)$, we have

$$V(t, X(t)) = E\left[\int_t^T e^{-\int_t^s k(u,X(u))du} f(s, X(s))ds + g(X(T))e^{-\int_t^T k(u,X(u))du}|X(t)\right].$$

Equivalently, we have

$$V(t, x) = E\left[\int_t^T e^{-\int_t^s k(u,X(u))du} f(s, X(s))ds + g(X(T))e^{-\int_t^T k(u,X(u))du}|X(t) = x\right].$$

8 The fundamental theorems of stochastic analysis

8.1 Levy's characterization of Brownian motions

Levy's characterization of Brownian motions plays an important role in stochastic analysis. It asserts the following results.

THEOREM 4. *(Levy) Suppose $\{M(t) = (M_1(t), M_2(t), ..., M_d(t))\}$ is a d-dimensional martingale with continuous sample paths adapted to the filtration $\{\mathcal{F}(t)\}$. Assume that $M(0) = 0$ without loss of generality. If the cross variation satisfies the following property*

$$[M_i, M_j](t) = \delta_{ij} t$$

with $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise, $\{M(t)\}$ is a standard d-dimensional Brownian motion.

To characterize the distribution, it is sufficies to compute the conditional moment generating function of the increment $M(t) - M(s)$ for $0 \leq s < t$,

$$\mathbb{E}[\exp\left(u(M(t) - M(s))\right)|\mathcal{F}(s)] = \exp\left(\frac{1}{2}\sum_{i=1}^{d} u_i^2(t-s)\right).$$

For this purpose, we use the Itô formula to find the differential of $\exp\left(u(M(t) - M(s))\right)$ and then make full use of the given condition in the theorem and the martingale property. We suggest this proof as an excellent exercise.

In the next two theorems, we expound on the theme that *Brownian motion is the fundamental martingale with continuous sample paths.*

8.2 Continuous local martingale as time-changed Brownian motion

Our first result states that "every local martingale with continuous sample paths, is nothing but a Brownian motion, run under a different clock."

THEOREM 5. *(Dambis (1965), Dubins and Schwartz (1965)) Suppose* $\{M(t)\}$ *is a continuous local martingale safisfying* $\lim_{t\to\infty} \langle M \rangle (t) = \infty$. *There exists then a Brownian motion* W, *such that:*

$$M(t) = W(\langle M \rangle (t)); 0 \leq t < \infty.$$

Sketch of proof. Without loss of essence, we consider the case when $\langle M \rangle$ is strictly increasing.* In this case $\langle M \rangle$ has an inverse, say T, which is continuous (as well as strictly increasing). Then it is not hard to see that the process

$$W(s) = M(T(s)),\ 0 \leq s < \infty \tag{8.1}$$

is a local martingale (with respect to the filtration $\{\mathcal{G}(s) = \mathcal{F}(T(s))\}$) with continuous sample paths, as being the composition of the two continuous mappings $T : \mathbb{R}^+ \to \mathbb{R}^+$ and $M : \mathbb{R}^+ \to \mathbb{R}$. On the other

*E.g., if $M(t) = \int_0^t X(s)dB(s)$, where B is Brownian motion and X takes values in $\mathbb{R}\backslash\{0\}$.

hand, by the definition of quadratic variation, it is that

$$\langle W \rangle (s) = \langle M \rangle (T(s)) = s.$$

Thus, by the Levy characterization of Brownian motion, W is a Brownian motion. Furthermore, replacing s by $\langle M \rangle_t$ in the above equation, we obtain

$$W(\langle M \rangle_t) = M(T(\langle M \rangle (t))) \equiv M(t),$$

which is the desired representation.

Following the similar spirit, a multidimensional generalization of this theorem can be established as follows.

THEOREM 6. *(Knight (1971)) Suppose $\{M(t) = (M_1(t), M_2(t), ..., M_d(t))\}$ is a d-dimensional continuous local martingale safisfying $\lim_{t\to\infty} \langle M_i \rangle (t) = \infty$ for each $i = 1, 2, ..., d$. There exists then a standard d-dimensional Brownian motion W, such that:*

$$M_i(t) = W_i(\langle M_i \rangle (t)); 0 \leq t < \infty.$$

Alternative forms of the above two theorems and their rigorous proofs can be found in Section 3.4 of Karatzas and Shreve [1].

8.3 Martingale representation

In what follows, we give the martingale representation theorem.

THEOREM 7. *Suppse there is a random variable $F \in L^2$. W is a d-dimensional Brownian motion. If $F \in \mathcal{F}_W(T)$, where $\{\mathcal{F}_W(t)\}$ is the Brownian filtration, we have*

$$F = \mathbb{E}[F] + \int_0^T \phi(s)^\top dW(s),$$

where $\{\phi(s)\}$ *is a d-dimensional vector process. Denote by*

$$M(s) = \mathbb{E}(F|\mathcal{F}_W(t)).$$

Then, we have

$$M(t) = M(0) + \int_0^t \phi(s)dW(s) \tag{8.2}$$

and, in particular,

$$M(T) = F.$$

The "integrand" ϕ is unique in this representation. The differential form of (8.2) can be written as $dM(t) = \phi(t)dW(t)$. Intuitively speaking, any martingale can be represented by increments of Brownian motions. A financial interpretation of this theorem can be given as follows. Assume stock prices are Brownian motions. If ϕ is the portfolio, $\phi(t)dW(t)$ is the local return. And thus the integral $\int_0^t \phi(s)dW(s)$ becomes the aggregate return of the portfolio. If one trades Brownian motion $\{W(t)\}$, given any contingent claim with price process $\{M(t)\}$, such contingent claim can be replicated by trading $\{W(t)\}$ according to the portfolio strategy $\{\phi(t)\}$. This theorem is about existence. The question is how to find ϕ explicitly? One needs theory of Malliavin calculus.

Alternative forms of the above theorem and their rigorous proofs can be found in Section 3.4 of Karatzas and Shreve [1].

8.4 Girsanov theorem

Before discussing about the formal version of the Girsanov theorem, we propose a baby version to ease understanding.

8.4.1 A simple example for understanding the Girsanov theorem

Assume that we have a standard normal variable X on a probability space $(\Omega, \mathcal{F}, P)$. Now, define

$$Z = \exp\left(-\theta X - \frac{1}{2}\theta^2\right).$$

and create an equivalent probability measure $\widetilde{\mathbb{P}}$ via the Radon-Nykodim derivative:

$$\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} = Z.$$

We just need to show that $\widetilde{\mathbb{P}}$ is indeed a probability measure. For this purpose, we deduce that

$$\begin{aligned}
\widetilde{\mathbb{P}}(\Omega) &= \int_\Omega d\widetilde{\mathbb{P}} = \int_\Omega Z d\mathbb{P} = \mathbb{E}Z \\
&= \mathbb{E}\exp\left(-\theta X - \frac{1}{2}\theta^2\right) = \left[\mathbb{E}\exp\left(-\theta X\right)\right]\exp\left(-\frac{1}{2}\theta^2\right) = 1.
\end{aligned}$$

The above argument can be rigorously interpreted via the perspective of real analysis or measure theory. However, it doesn't matter if you focus on applications and have never learnt anything about those theory. We note that the expectation $\mathbb{E}Z$ can be regarded as an integration of $Z(\omega)$ with respect to ω in the whole set Ω, i.e.,

$$\mathbb{E}Z = \int_\Omega Z d\mathbb{P}.$$

And, just convince yourself the follow analogy between classical integration and expectation considered here

$$\int_R f(x)dx \sim \int_\Omega Z d\mathbb{P}.$$

Under the probability measure $\widetilde{\mathbb{P}}$, it is amazing that $Y = X + \theta$ is a standard normal variable. In other words, $X = Y - \theta$ is a normal variable with distribution $N(-\theta, 1)$ under $\widetilde{\mathbb{P}}$. Indeed, it is sufficient to show that the moment generating function of Y under the probability measure $\widetilde{\mathbb{P}}$ is exactly that of a standard normal variable. We deduce that

$$
\begin{aligned}
\mathbb{E}^{\widetilde{\mathbb{P}}} e^{\lambda Y} &= \int_\Omega e^{\lambda Y} d\widetilde{\mathbb{P}} \\
&= \int_\Omega e^{\lambda Y} \frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} d\mathbb{P} \\
&= \mathbb{E}^{\mathbb{P}} \left[\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} e^{\lambda Y} \right] \\
&= \mathbb{E}^{\mathbb{P}} \left[\exp\left(-\theta X - \frac{1}{2}\theta^2 + \lambda(X+\theta) \right) \right] \\
&= \mathbb{E}^{\mathbb{P}} \left[\exp\left((\lambda-\theta)X - \frac{1}{2}\theta^2 + \lambda\theta \right) \right] \\
&= e^{\frac{1}{2}(\lambda-\theta)^2 - \frac{1}{2}\theta^2 + \lambda\theta} = e^{\frac{1}{2}\lambda^2}.
\end{aligned}
$$

So, since moment generating functions determine distributions, it is enough to see that under the probability measure $\widetilde{\mathbb{P}}$, $Y = X + \theta$ is a standard normal variable.

8.4.2 Girsanov Change of Measure

Under the probability space $(\Omega, \mathbb{P}, \mathcal{F})$ equipped with the filtration $\{\mathcal{F}(t)\}$, suppose $\{W(t)\}$ is a d-dimensional Brownian motion and $\{\theta(t)\}$ is a stochastic process adapted to the Brownian filtration. Consider

$$
\eta(t) := \exp\left(-\sum_{i=1}^d \int_0^t \theta_i(s) dW_i(s) - \frac{1}{2} \sum_{i=1}^d \int_0^t \theta_i(s)^2 ds \right).
$$

By using the Itô formula, it is easy to have (verify it as an excellent exercise)

$$
d\eta(t) = -\sum_{i=1}^d \theta_i(t)\eta(t) dW_i(t).
$$

Thus, $\eta(t)$ can be written as a stochastic integral

$$
\eta(t) = -\sum_{i=1}^d \int_0^t \theta_i(s)\eta(s) dW_i(s). \tag{8.3}
$$

According to the theory of stochastic integral, without any conditions, $\{\eta(t)\}$ is merely a local martingale, which need not to be a true martingale and is a martingale evaluated at bounded stopping times. The following theorem, which is often called as the celebrated Novikov condition, provides a condition under which such a local martingale becomes a true martingale.

THEOREM 8. *(Novikov Condition) If we have*

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\sum_{i=1}^{d}\int_0^t \theta_i(s)^2 ds\right)\right] < \infty,$$

$\{\eta(t)\}$ *is a true martingale and, in particular,*

$$\mathbb{E}\eta(t) = 1, \text{ for any } t \in [0, T].$$

Indeed, one is able to consider other sufficient conditions for making $\{\eta(t)\}$ a true martingale, e.g., the following commonly used condition ensuring the stochastic integral defined through (8.3) is a martingale:

$$\mathbb{E}\left(\int_0^t \theta_i(s)^2 \eta(s)^2 ds\right) < \infty.$$

Now, it is enough for us to propose the Girsanov theorem, one of the most brilliaint mile stone in stochastic analysis.

THEOREM 9. *(Girsanov) When* $\{\eta(t)\}$ *is a true martingale, we can construct a new probability measure* $\mathbb{Q}$ *such that*

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \eta(T) \Leftrightarrow d\mathbb{Q} = \eta(T) d\mathbb{P},$$

i.e.

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

for any $A \in \Omega$*. The probability measures* $\mathbb{P}$ *and* $\mathbb{Q}$ *are equivalent. If we definte*

$$W^{\mathbb{Q}}(t) = W(t) + \int_0^t \theta(s) ds,$$

$W^{\mathbb{Q}}$ *is a standard Brownian motion under the probability measure* $\mathbb{Q}$*.*

Obviously, $W^{\mathbb{Q}}$ is not a standard Brownian motion but a Browian motion with stochastic dirft $\int_0^t \theta(s)ds$ under $\mathbb{P}$. A frequenctly used baby version of this theorem is the case when $\theta(t) \equiv \theta$. In this case, we have

$$\eta(t) = \exp\left(-\theta W(t) - \frac{1}{2}\theta^2 t\right),$$

which is exactly the well-known exponential martingale. By the Girsanov theorem, under the probability measure $\mathbb{Q}$, $W^{\mathbb{Q}}(t) = W(t) + \theta t$ is a standard Brownian motion. However, it is a Brownian motion with drift under the probability measure $\mathbb{P}$.

To prove the Girsanov theorem, we begin by verifying that $\mathbb{Q}$ is indeed a probability measure. For this purpose, we have

$$\mathbb{Q}(\Omega) = \int_\Omega \eta(T)d\mathbb{P} = \mathbb{E}^{\mathbb{P}}\eta(T) = 1;$$

and, for any two disjoint A and B from the sigma algebra $\mathcal{F}(T)$; we have

$$\mathbb{Q}(A \cap B) = \int_{A \cap B} \eta(T)d\mathbb{P} = \int_A \eta(T)d\mathbb{P} + \int_B \eta(T)d\mathbb{P} = \mathbb{Q}(A) + \mathbb{Q}(B).$$

It's easy to check that the two probability measures $\mathbb{P}$ and $\mathbb{Q}$ are equivalent.

THEOREM 10. *(Bayesian Rule) Suppose* Y *is a* $\mathcal{F}(T)$*-measurable random variable satisfying* $\mathbb{E}^{\mathbb{P}}|Y| < \infty$ *and* $\mathbb{E}^{\mathbb{P}}|\eta(t)Y| < \infty$ *for all* $t \in [0, T]$*. Then, we have*

$$\mathbb{E}^{\mathbb{Q}}Y = \mathbb{E}^{\mathbb{P}}(\eta(T)Y),$$

and, more generally,

$$\mathbb{E}^{\mathbb{Q}}[Y|\mathcal{F}(t)] = \mathbb{E}^{\mathbb{P}}\left[\frac{\eta(T)}{\eta(t)}Y|\mathcal{F}(t)\right] = \mathbb{E}^{\mathbb{P}}[\eta(t,T)Y|\mathcal{F}(t)], \tag{8.4}$$

where

$$\eta(t,T) := \frac{\eta(T)}{\eta(t)}.$$

Obviously, $\eta(t,T)$ plays the role as a normalized version of the Radon-Nikodym derivative for changing measure. Relation (8.4) reveals how change of measure can be done under conditioning.

Proof. For the first claim, we know that

$$\mathbb{E}^{\mathbb{Q}}Y = \int_{\Omega} Y d\mathbb{Q} = \int_{\Omega} \eta(T) Y dP = \mathbb{E}^{\mathbb{P}}[\eta(T)Y].$$

Now, we prove the second claim. For simplicity, we write $\mathbb{E}_t^{\mathbb{Q}}(\cdot)$ as $\mathbb{E}^{\mathbb{Q}}[\cdot|\mathcal{F}(t)]$ and write $\mathbb{E}_t^{\mathbb{P}}(\cdot)$ as $\mathbb{E}^{\mathbb{P}}[\cdot|\mathcal{F}(t)]$

For any $A \in \mathcal{F}(t)$, we have on one hand

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}\left(1_A \mathbb{E}_t^{\mathbb{Q}} Y\right) &= \mathbb{E}^{\mathbb{Q}}\left(1_A Y\right) \\ &= \mathbb{E}^{\mathbb{P}}[\eta(T) 1_A Y] \\ &= \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}_t^{\mathbb{P}}[\eta(T) 1_A Y]\right] \\ &= \mathbb{E}^{\mathbb{P}}\left[1_A \mathbb{E}_t^{\mathbb{P}}[\eta(T) Y]\right]; \end{aligned}$$

on the other hand, we have

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}\left(1_A \mathbb{E}_t^{\mathbb{Q}} Y\right) &= \mathbb{E}^{\mathbb{P}}\left(1_A \eta(T) \mathbb{E}_t^{\mathbb{Q}} Y\right) \\ &= \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}_t^{\mathbb{P}}\left(1_A \eta(T) \mathbb{E}_t^{\mathbb{Q}} Y\right)\right] \\ &= \mathbb{E}^{\mathbb{P}}\left[1_A \left(\mathbb{E}_t^{\mathbb{Q}} Y\right)\left(\mathbb{E}_t^{\mathbb{P}} \eta(T)\right)\right] \\ &= \mathbb{E}^{\mathbb{P}}\left[1_A \eta(t) \mathbb{E}_t^{\mathbb{Q}} Y\right]. \end{aligned}$$

Thus, we have

$$\mathbb{E}^{\mathbb{P}}\left[1_A \mathbb{E}_t^{\mathbb{P}}[\eta(T)Y]\right] = \mathbb{E}^{\mathbb{P}}\left[1_A \eta(t) \mathbb{E}_t^{\mathbb{Q}} Y\right].$$

By the definition of conditional expectation, we have

$$\mathbb{E}^{\mathbb{P}}\left[1_A \eta(T) Y\right] \equiv \mathbb{E}^{\mathbb{P}}\left[1_A \mathbb{E}_t^{\mathbb{P}}\left(\eta(T) Y\right)\right].$$

Thus, we have

$$\mathbb{E}^{\mathbb{P}}\left[1_A \eta(T) Y\right] = \mathbb{E}^{\mathbb{P}}\left[1_A \eta(t) \mathbb{E}_t^{\mathbb{Q}} Y\right].$$

Using the definition of conditional expectation again, we obtain

$$\mathbb{E}_t^{\mathbb{P}}\left[\eta(T) Y\right] = \eta(t) \mathbb{E}_t^{\mathbb{Q}} Y,$$

which leads to the claim. □

Proof to Theorem 9. To prove the Girsanove theorem, we rely on the Levy characterization of Brownian motion. First, we show that

$$[W_i^{\mathbb{Q}}, W_j^{\mathbb{Q}}](t) = \delta_{ij} t.$$

Indeed, this is because the drift term $\int_0^t \theta(s)ds$ contributes zero in the calculation of qudratic variation, i.e.,

$$[W_i^{\mathbb{Q}}, W_j^{\mathbb{Q}}](t) = \left[W_i(\cdot) + \int_0^{\cdot} \theta_i(s)ds, W_j(\cdot) + \int_0^{\cdot} \theta_j(s)ds\right](t) = [W_i, W_j](t) = \delta_{ij} t.$$

Second, it remains to show that $\{W^{\mathbb{Q}}(t)\}$ is a local martingale under $\mathbb{Q}$. By using the Itô formula, it is straight forward to have (check it as an excellent exercise!)

$$d(W_i^{\mathbb{Q}}(t)\eta(t)) = [-W_i^{\mathbb{Q}}(t)\theta(t) + 1]\eta(t)dW_i(t), \text{ for } i = 1, 2, ..., d,$$

which leads to the local martingale property of $\{W_i^{\mathbb{Q}}(t)\eta(t)\}$. Standard localization techniques (for details, see p.194 of Karatzas and Shreve [1]), it is enough to consider the case where $\{W_i^{\mathbb{Q}}(t)\eta(t)\}$ becomes a true martingale. Thus, using the Bayes rule (8.4), for any $0 < s < t$, we arrive at

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}[W_i^{\mathbb{Q}}(t)|\mathcal{F}(s)] &= \mathbb{E}^{\mathbb{P}}\left[\frac{\eta(t)}{\eta(s)} W_i^{\mathbb{Q}}(t)|\mathcal{F}(s)\right] \\
&= \frac{1}{\eta(s)} \mathbb{E}^{\mathbb{P}}\left[\eta(t) W_i^{\mathbb{Q}}(t)|\mathcal{F}(s)\right] \\
&= \frac{1}{\eta(s)} \eta(s) W_i^{\mathbb{Q}}(s) \\
&= W_i^{\mathbb{Q}}(s).
\end{aligned}$$

Thus, the above argument leads to the local martingale property of $\{W^{\mathbb{Q}}(t)\}$. $\Box$

References

[1] Karatzas, I., and S. E. Shreve (1991): *Brownian Motion and Stochastic Calculus*, vol. 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edn.

Guanghua School of Management

Stochastic Analysis and Applications

Applications of Stochastic Analysis in Financial Engineering

Spring 2017

Agenda

- applications in derivative pricing
  - option pricing under the Black-Scholes-Merton model
  - beyond Black-Scholes-Merton: stochastic volatility
  - beyond Black-Scholes-Merton: jumps
  - affine jump-diffusions and Fourier pricing methods
- applications in optimal portfolio choice

Overview

- Option pricing under the Black-Scholes-Merton model

Derivatives Valuation and Hedging

- A derivative security is a financial asset whose payoff depends on the value of some underlying variable.
- The underlying variable can be a traded asset (e.g., IBM stock), an index (e.g., Hang Seng Index), other derivatives (e.g., crude oil futures), and so on.
- An option is a derivative security that grants the buyer the right to buy or sell the underlying asset, at or before the maturity date T, for a pre-specified price K, called the strike or exercise price.
- “Right", not “obligation".
- “buy": call options; “sell": put options.
- If it can be exercised only at the maturity date, it is European style; if the exercise can happen at any time before or at the maturity date, it is American style.
  `http://finance.yahoo.com/`
- Excellent leverage and risk-management tools

Derivatives Valuation and Hedging

- An European call option has a payoff: $(S(T) - K)^+$; whereas an European put option: $(K - S(T))^+$, where $S(t)$ denotes the underlying asset price at time t.
- Suppose one attempts to buy an European call option with underlying asset being IBM stock.
- Question: How much should he/she pay? What is the “fair" price? What does “fair" mean?
- The options have value all the time until their maturities.

What is Hedging?

Making an investment to reduce the risk of adverse price movements in an asset.

How to hedge?

- Sell!
- Buy insurance (options, specialized contracts)
- Self-insure (correlation, replication)

Question: If you manufacture a call option (sell it to your customer), what do you need to do? Just wait after you collect the option premium until the maturity?

Self-financing, Arbitrage, and Payoff Replication

- **Definition**: A trading strategy (portfolio) is self-financing if there is no exogenous infusion or withdrawal of money; the purchase of a new asset must be financed by the sale of an old one.
- **Definition**: Arbitrage is a self-financing trading strategy that
  - begins with no money ($X(0) = 0$)
  - has no probability of losing money ($\mathbb{P}(X(T) \geq 0) = 1$)
  - has a positive probability of making money at some future date ($\mathbb{P}(X(T) > 0) > 0$)
- An efficient market should preclude arbitrages.
- If a portfolio starts from some value, at time T, it replicates the option payoff. The value of the option must be that starting value. Otherwise, we are able to construct arbitrage opportunities.

Continuous-Time Finance: Black-Scholes-Merton (1973)

Why?

- More realistic and flexible models is highly needed;
- More mathematical tools;
- etc.

Note: No model is correct! Models should get closer to the reality and capture some main features according to some certain business. Sometimes, even a “wrong” model can do something great!

The Black-Scholes-Merton Model

- Consider a simple financial market with:
  - an asset (stock) $S(t)$, and
  - a money market account with a continuously compounding interest rate r, i.e., Investing 1 dollar in money market becomes e^{rt} at time t.

- We intend to price a European call option that pays $(S(T) - K)^+$ at maturity T.

- We propose a geometric Brownian motion model for the underlying stock:

$$dS(t) = \alpha S(t)dt + \sigma S(t)dW(t). \tag{1}$$

Option Pricing with BSM

- Under this model, assume that the value of the European call option at time $t \in [0, T]$ depends only on the stock price $S(t)$ and the time to expiration $T - t$.
- Denote by $V(t) = c(t, S(t))$ the value of the European call option at time $t \in [0, T]$, where $c(t, x)$ is a deterministic function with two dummy variables t and x.
- $V(t) = c(t, S(t))$ implies that the risk premium is embedded in the underlying asset $S(t)$!
- Moreover, assume that $c_t(t, x)$, $c_x(t, x)$, and $c_{xx}(t, x)$ exist.
- Consider hedging a short position (e.g. you are selling this option to your customer) in the option in the following way:
  - Start from $X(0) := c(0, S(0))$;
  - Construct a self-financing portfolio $X(t)$ s.t. $X(T) = c(T, S(T))$.

Option Pricing with BSM

- In order to rule out arbitrage, we need to have

$$X(t) \equiv V(t) = c(t, S(t))$$

for all $t \in [0, T]$.

- This is equivalent to

$$dX(t) \equiv dV(t) = dc(t, S(t))$$

for any $t \in [0, T)$.

Option Pricing with BSM

- **Step 1:** Calculate $dX(t)$.
- A self-financing strategy to reallocate the time t wealth $X(t)$.
  - buy $\Delta(t)$ shares of stocks;
  - the rest $X(t) - \Delta(t)S(t)$ is invested in money market.
- By the self-financing condition, we have

$$\begin{aligned} dX(t) =& \Delta(t)dS(t) + r(X(t) - \Delta(t)S(t))dt \\ =& \Delta(t)\left[\alpha S(t)dt + \sigma S(t)dW(t)\right] + r(X(t) - \Delta(t)S(t))dt \\ =& rX(t)dt + \Delta(t)(\alpha - r)S(t)dt + \Delta(t)\sigma S(t)dW(t) \end{aligned} \quad (2)$$

- Note! The change of the discounted replicating portfolio value is only due to the change of the discounted stock price.

$$d\left[e^{-rt}X(t)\right] = \Delta(t)d\left[e^{-rt}S(t)\right].$$

(An excellent exercise!)

Option Pricing with BSM

- **Step 2:** Calculate $dV(t) = d\left[c(t, S(t))\right]$.
- Apply Itô formula to $dc(t, S(t))$

$$
\begin{aligned}
dc(t, S(t)) = & \Big[c_t(t, S(t)) + \alpha S(t) c_x(t, S(t)) \\
& + \frac{1}{2}\sigma^2 S^2(t) c_{xx}(t, S(t))\Big] dt + \sigma S(t) c_x(t, S(t)) dW(t)
\end{aligned} \tag{3}
$$

Option Pricing with BSM

- Let

$$dX(t) \equiv dc(t, S(t)) \Longleftrightarrow X(t) \equiv c(t, S(t)).$$

We have

$$
\begin{aligned}
& rX(t)dt + \Delta(t)(\alpha - r)S(t)dt + \sigma\Delta(t)S(t)dW(t) \\
= & \Big[c_t(t, S(t)) + \alpha S(t) c_x(t, S(t)) + \frac{\sigma^2 S^2(t)}{2} c_{xx}(t, S(t))\Big] dt \\
& + \sigma S(t) c_x(t, S(t)) dW(t)
\end{aligned} \tag{4}
$$

- Equate the $dW(t)$ terms $\Longrightarrow \Delta(t) = c_x(t, S(t))$.
- Equate the dt terms and replace $S(t)$ by a dummy variable x. We can obtain a Black-Scholes-Merton equation.

$$c_t(t, x) + rxc_x(t, x) + \frac{1}{2}\sigma^2 x^2 c_{xx}(t, x) = rc(t, x) \quad \text{for all } t \in [0, T), \tag{5}$$

with a terminal condition $c(T, x) = (x - K)^+$.

One question

- One question about the derivation of the BSM formula: How do we know the value of the European call option at time $t \in [0, T]$ depends only on the stock price $S(t)$ and the time to expiration $T - t$?
- Just now, the BSM equation is derived as a necessary condition.
- Now, let us verify it is also sufficient!
- Show that $c(t, S(t))$ is the time t value of the European call option.
- It suffices to construct a self-finance portfolio with time t value $c(t, S(t))$ to replicate the payoff of the European call option.
  - Start from $X(0) = c(0, S(0)) \geq 0$
  - At time t, buy $\Delta(t) = c_x(t, S(t))$ shares of stocks.
  - The rest $X(t) - c_x(t, S(t))S(t)$ is invested in the money market.
- An excellent exercise!

A Puzzle about α

- Observation: in the BSM equation, there is no α!
- The no-arbitrage price has nothing to do with the expected return of the underlying asset. Is this counter intuitive?
- Note that $V(t) = c(t, S(t))$; thus the risk premium is embedded in $S(t)$!
- As $S(t)$ taking larger value with higher probability, $V(t) = c(t, S(t))$ is more valuable!

From BSM to Practice

Assumptions in BSM:

- Constant volatility
- Constant interest rate
- No dividend, tax, transaction cost
- Continuously rebalancing of a perfect hedging portfolio
- etc.

Though ignoring some practical concerns, the BSM successfully opened the door to modern derivative pricing theory. BSM itself can be applied in a smart way through the notion of implied volatility; and models developed based on the idea of BSM are also moving forward the derivative business.

How to solve the Black-Scholes-Merton PDE

- Consider the Black-Scholes-Merton model:

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma dW(t),$$

  and suppose that the interest rate is r.
- Let $C(t) = c(t, S(t))$ be the value of a call option with maturity T with payoff $(s - K)^+$.
- $c(t, x)$ satisfies the Black-Scholes-Merton equation.

$$c_t(t,x) + rxc_x(t,x) + \frac{1}{2}\sigma^2 x^2 c_{xx}(t,x) = rc(t,x) \quad \text{for all } t \in [0,T), \tag{6}$$

  with a terminal condition $c(T, x) = (x - K)^+$.
- Question: How to solve the Black-Scholes-Merton PDE?

An Analytical Method: Change of Variable

Consider the BSM PDE:

$$c_t(t,x) + rxc_x(t,x) + \frac{1}{2}\sigma^2 x^2 c_{xx}(t,x) = rc(t,x) \quad \text{for all } t \in [0,T), \tag{7}$$

Using the change of variables:

$$\begin{aligned} u &= e^{-rt}c, \\ y &= \log x, \tau = (T-t)\sigma^2, \\ z &= y + \frac{1}{\sigma^2}\left(r - \frac{1}{2}\sigma^2\right)\tau, \end{aligned}$$

the Black-Scholes-Merton PDE becomes a heat equation:

$$u_\tau(\tau,z) = \frac{1}{2}u_{zz}(\tau,z),$$

with terminal condition $u(0,z) = e^{-rT}(e^z - K)^+$.

The Black-Scholes-Merton formula

- Solve the heat equation:

$$u(\tau,z) = \int_{-\infty}^{+\infty} u(0,y)G(z,y,\tau)dy,$$

$$G(z,y,\tau) = \frac{1}{\sqrt{2\pi\tau}} \exp\left\{-\frac{(z-y)^2}{2\tau}\right\}.$$

- **The Black-Scholes-Merton formula**: For any $t \in [0,T)$ and $x > 0$,

$$c(t,x) = xN(d_+(T-t,x)) - Ke^{-r(T-t)}N(d_-(T-t,x)), \tag{8}$$

where $N(y)$ is the CDF of standard normal distribution and

$$\begin{aligned} d_+(\tau,x) &= \frac{1}{\sigma\sqrt{\tau}}\left[\log\left(\frac{x}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)\tau\right], \\ d_-(\tau,x) &= \frac{1}{\sigma\sqrt{\tau}}\left[\log\left(\frac{x}{K}\right) + \left(r - \frac{\sigma^2}{2}\right)\tau\right]. \end{aligned} \tag{9}$$

Risk-Neutral Representation of the BSM PDE Solution

- Let
$$\Theta(t) := \frac{\mu - r}{\sigma}$$
- We have
$$W^{\mathbb{Q}}(t) = W(t) + \int_0^t \frac{\mu - r}{\sigma} du = W(t) + \frac{\mu - r}{\sigma} t$$
is a Brownian motion under $\mathbb{Q}$ satisfying $\frac{d\mathbb{Q}}{d\mathbb{P}} = Z(T)$.
- By the Girsanov theorem, we have that
$$\begin{aligned} dS(t) &= \mu S(t)dt + \sigma S(t)dW(t) \\ &= \mu S(t)dt + \sigma S(t)\left(dW^{\mathbb{Q}}(t) - \frac{\mu - r}{\sigma}dt\right) \\ &= rS(t)dt + \sigma S(t)dW^{\mathbb{Q}}(t). \end{aligned}$$
- $\Theta(t)$: the Sharpe ratio or market price of risk.

Risk-Neutral Representation of the BSM PDE Solution

- The probability measure $\mathbb{Q}$ is called the risk-neutral (martingale) measure.
- Under $\mathbb{Q}$, we have
$$dS(u) = rS(u)du + \sigma S(u)dW^{\mathbb{Q}}(u).$$
This is a special case of the general SDE: $\beta(u, x) = rx$ and $\gamma(u, x) = \sigma x$.
- Let
$$v(t, S(t)) = \mathbb{E}^{\mathbb{Q}}[e^{-r(T-t)}(S(T) - K)^+ | \mathcal{F}(t)].$$
- By the Feynman-Kac theorem, $v(t, x)$ solves the BSM PDE:
$$v_t(t, x) + rxv_x(t, x) + \frac{1}{2}\sigma^2 x^2 v_{xx}(t, x) = rv(t, x), \qquad (10)$$
with terminal condition $v(T, x) = (x - K)^+$.

A rigorous application of Feynman-Kac theorem: uniqueness

- Because
$$(S(T)-K)^+ \leq S(T),$$
the no-arbitrage price of the call option must be dominated by the value of the underlying asset, i.e.
$$u(t,S(t)) \leq S(t).$$
- Therefore, the price function u satisfies the polynomial growth condition:
$$\max_{0\leq t\leq T} |u(t,s)| \leq s \leq M(1+|s|^{2\mu}).$$
- By the Feynman-Kac theorem, we have the Black-Scholes solution, i.e.
$$u(t,S(t)) = \mathbb{E}^Q\left[e^{-r(T-t)}(S(T)-K)^+|\mathcal{F}(t)\right].$$
- As an exercise, prove the case for put options.

Risk-neutral Valuation

- Under $\mathbb{Q}$, $e^{-rt}S(t)$ (the discounted underlying asset price), $e^{-rt}v(t,S(t))$ (the discounted option price) and $e^{-rt}X(t)$ (the discounted replicating portfolio value) are all martingales
- Risk-Neutral Representation of the BSM PDE Solution:
$$v(t,S(t)) = \mathbb{E}^Q[e^{-r(T-t)}(S(T)-K)^+|\mathcal{F}(t)]$$
- This expresses the option price as the risk-neutral expectation of the discounted payoff.
- Using the explicit solution of $S(T)$ to derive the Black-Scholes-Merton formula.

Existence of Risk-Neutral Implies No Arbitrage

- General definition of **Risk Neutral Probability Measure**: an equivalent probability measure under which the discounted security prices of the market are martingales.
- **Theorem**: If we can find a risk-neutral probability measure (for the whole market), the market is free of arbitrage.
- A straightforward proof by contradiction: If there exists an arbitrage, then beginning with $X(0) = 0$, we can construct a portfolio such that
  - $X(T)(\omega) \geq 0$ for all $\omega \in \Omega$.
  - $\mathbb{P}(X(T) > 0) > 0$.
- Equivalence implies $\mathbb{Q}(X(T) > 0) > 0$.
- Therefore, $\mathbb{E}^{\mathbb{Q}} X(0) = 0$ but $\mathbb{E}^{\mathbb{Q}} \left[e^{-rT} X(T) \right] > 0$. Contradiction!

Overview

- Beyond Black-Scholes-Merton: stochastic volatility

Overview

- Empirical evidence for the Black-Scholes model: historical and implied
- Possible explanations
- Derivative valuation under stochastic volatility models

Empirical Evidence

Is the Black-Scholes model an accurate description of the financial world?

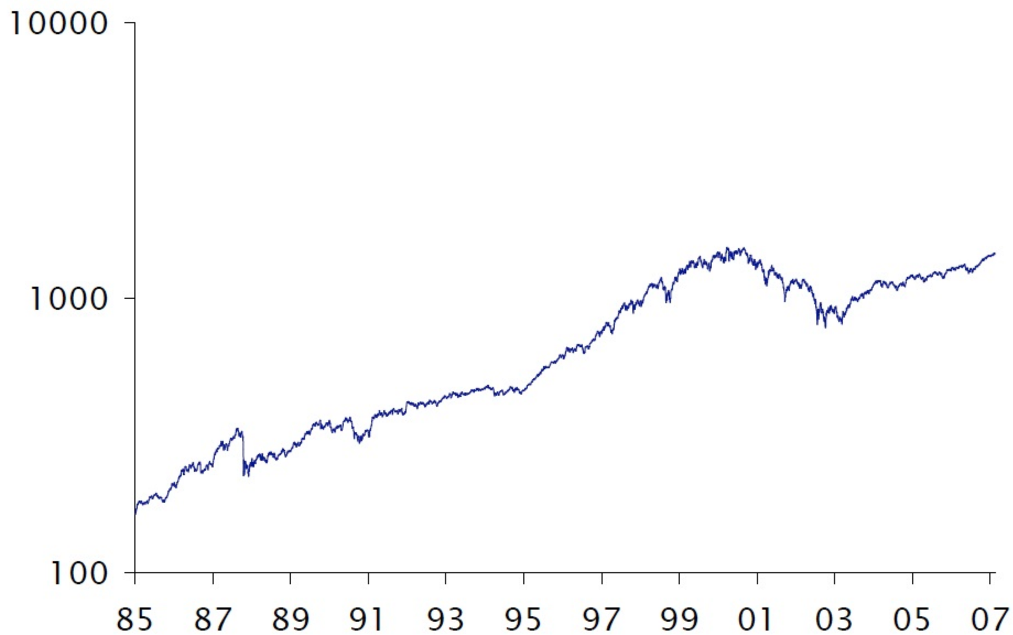
Historical (underlying asset)

- distribution of returns (check lognormal assumption)
- time series properties of returns (check independence and constant volatility assumptions)

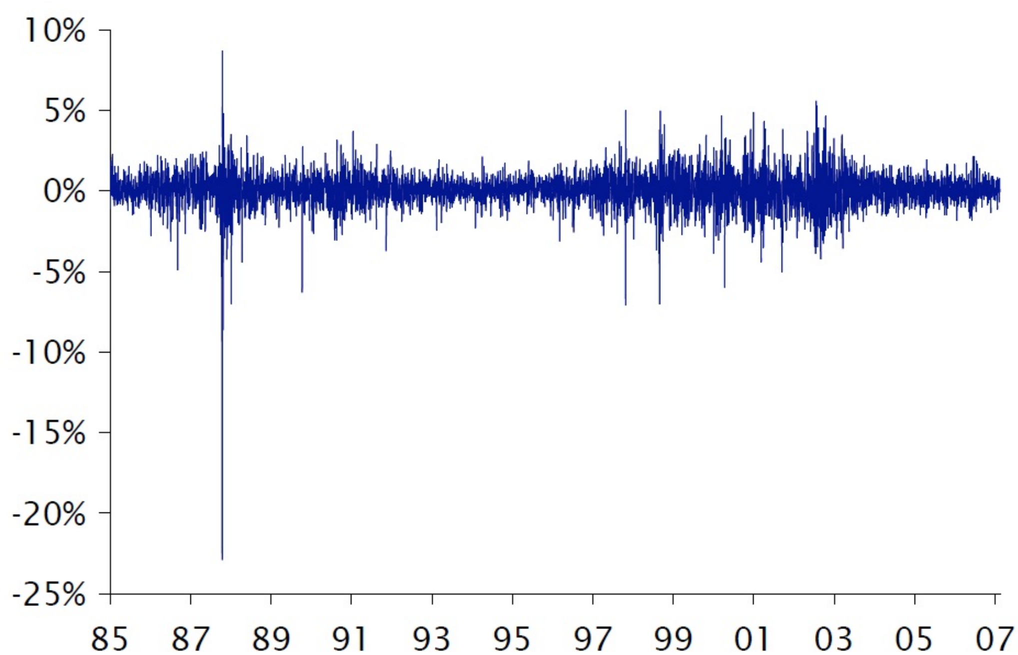
Implied (option)

- implied volatility (check dependence on strike and maturity; check change over time)

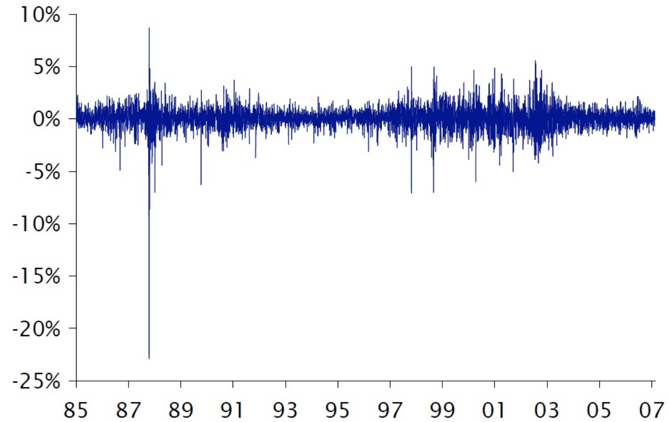
S&P500 Prices:1985-2007



Daily S&P500 Ln-Returns: 1985-2007

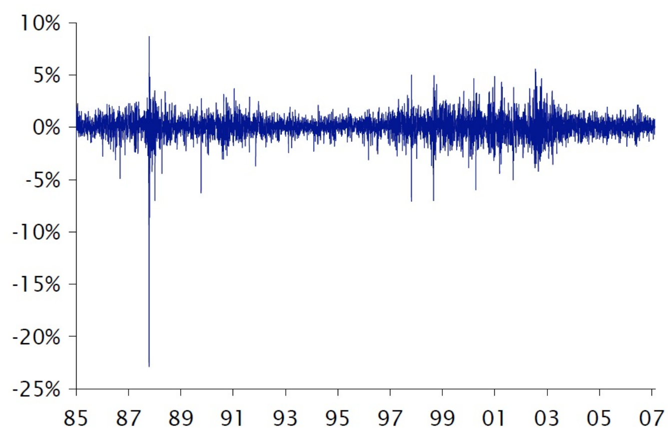


Non-Normality of Ln-Returns



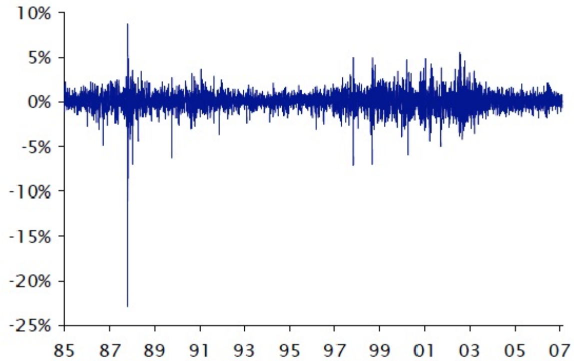
- Daily mean of Ln-returns: 0.04%
- Daily standard deviation of Ln-returns: 1.1%
- Daily Ln-return on 10/19/87: -22.9%
- Under normality, $\mathbb{P}(X \leq -22.9\%) = N(-20.9) \approx 10^{-96}$ (every 10^{93} years)

Non-Normality of Ln-Returns

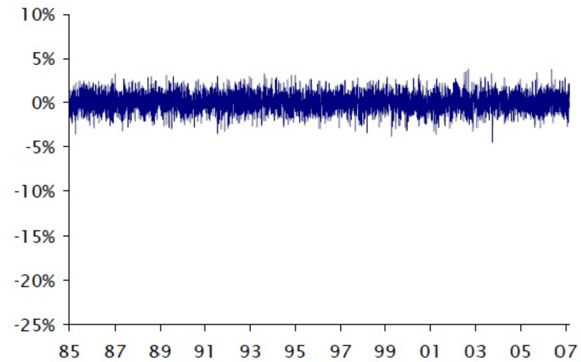


- 10/13/1989, $\mathbb{P}(X \leq -6.31\%) = N(-5.8) \approx 10^{-9}$ (every 1,000,000 years)
- Extreme movements are much more common than log-normal assumption suggests!

Non-Normality of Ln-Returns: Data versus Simulation



Data



Simulated returns

Non-Normality of Ln-Returns

$$\text{Kurtosis}(X) = \frac{E[(X-\bar{X})^4]}{\sigma^4} - 3.$$

Kurtosis is one measure of "fat tails", or the probability of extreme events.

$$\text{Kurtosis}(\text{normal random variable}) = 0.$$

$$\text{Sample Kurtosis}(\text{S\&P500 Ln-returns}) = 45.$$

Is this statistically significant?

Autocorrelation of Ln-Returns

Given a time series of Ln-returns $R_1, R_2, ..., R_n$, define the autocorrelation with a lag of k by

$$C(k) = \frac{E[(R_i - \bar{R})(R_{i-k} - \bar{R})]}{\sigma_R^2}$$

Time	Price	Ln-Return Lag 0	Ln-Return Lag 1	Ln-Return Lag 2
0	S_0			
Δt	$S_{\Delta t}$	$R_1 = \ln\left(\frac{S_{\Delta t}}{S_0}\right)$		
$2\Delta t$	$S_{2\Delta t}$	$R_2 = \ln\left(\frac{S_{2\Delta t}}{S_{\Delta t}}\right)$	R_1	
$3\Delta t$	$S_{3\Delta t}$	$R_3 = \ln\left(\frac{S_{3\Delta t}}{S_{2\Delta t}}\right)$	R_2	R_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n\Delta t$	$S_{n\Delta t}$	$R_n = \ln\left(\frac{S_{n\Delta t}}{S_{(n-1)\Delta t}}\right)$	R_{n-1}	R_{n-2}

Autocorrelation of Ln-Returns

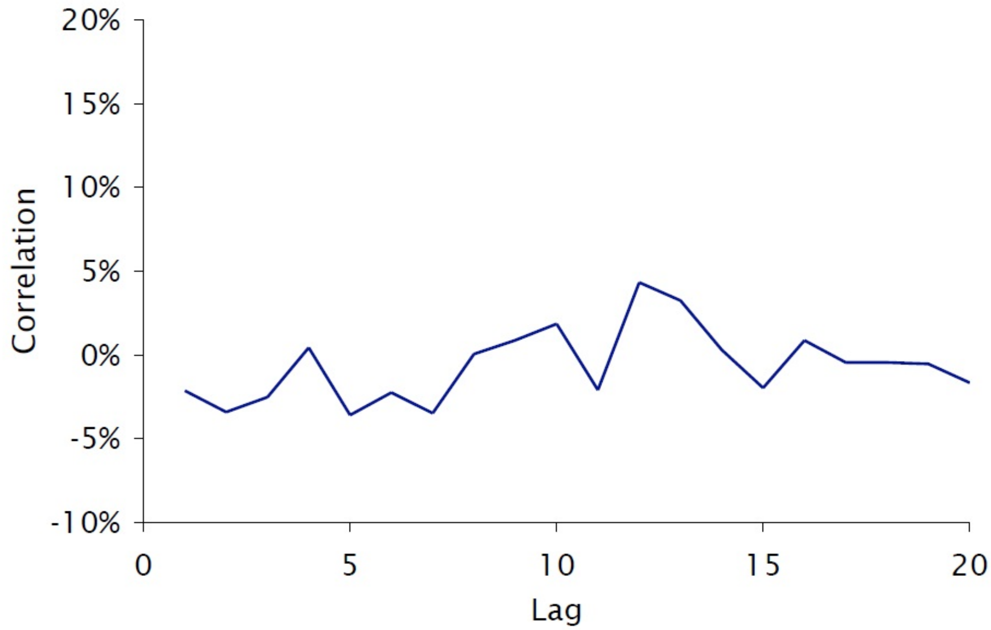
Under Black-Scholes assumptions,

$$R_i = \ln\left(\frac{S_{i\Delta t}}{S_{(i-1)\Delta t}}\right) = (\mu - \sigma^2/2)\Delta t + \sigma\sqrt{\Delta t}Z_i,$$

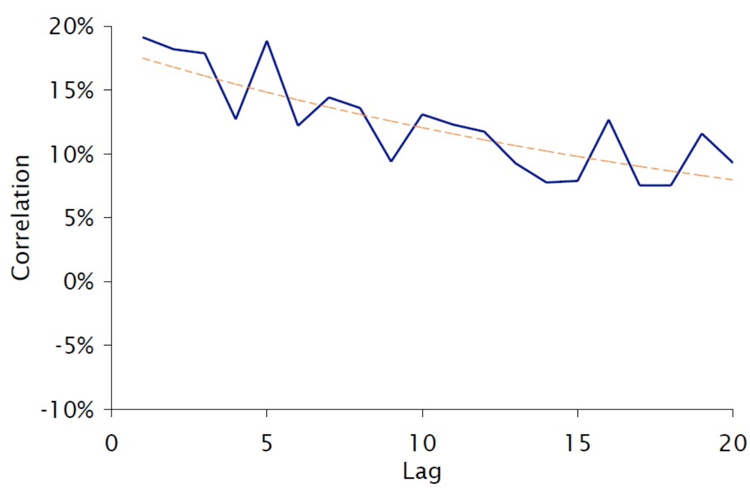
where $Z_1, Z_2, ...$are independent $N(0,1)$ random variables. Thus, $R_1, R_2, ...$should be independent, and

$$C(k) = 0 \quad \text{for } k \geq 1.$$

Autocorrelation of Ln-Returns



Autocorrelation of S&P500 Squared Ln-Returns



Autocorrelation of squared Ln-returns is hightly significant, even at a 20-day lag.

Big move today $\Rightarrow$ Big move tomorrow

$\Rightarrow$ volatility clustering!

Implied volatility

IMPLIED VOLATILITY: for option pricing

- The constant volatility parameter plugged into the Black-Scholes-Merton model for option pricing

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t, \tag{11}$$

- The **implied volatility** σ^* is the volatility that equates model and market option prices, i.e.

$$C(\sigma^*) = C_{Market}. \tag{12}$$

- Different options imply different implied volatility!
- Different maturities and strikes $\Longrightarrow$ implied volatility surface

Realized volatility

REALIZED VOLATILITY: for risk management and forecasting etc.

- Statistical definition: capture the real fluctuation of the asset return!
- Independent of any model
- **Historical Observation**: $\{S(t_i)\}$,
- **Realized variance** for the period of $[0, T]$ is defined as:

$$RV_{0,T} := \frac{1}{(n-1)\Delta t} \sum_{i=0}^{n-1} \left(\log \frac{S_{t_{i+1}}}{S_{t_i}} \right)^2 .$$

- **Realized volatility**: $\sqrt{RV_{0,T}}$

FTSE100 Index Option Prices, Spot$= 6045$

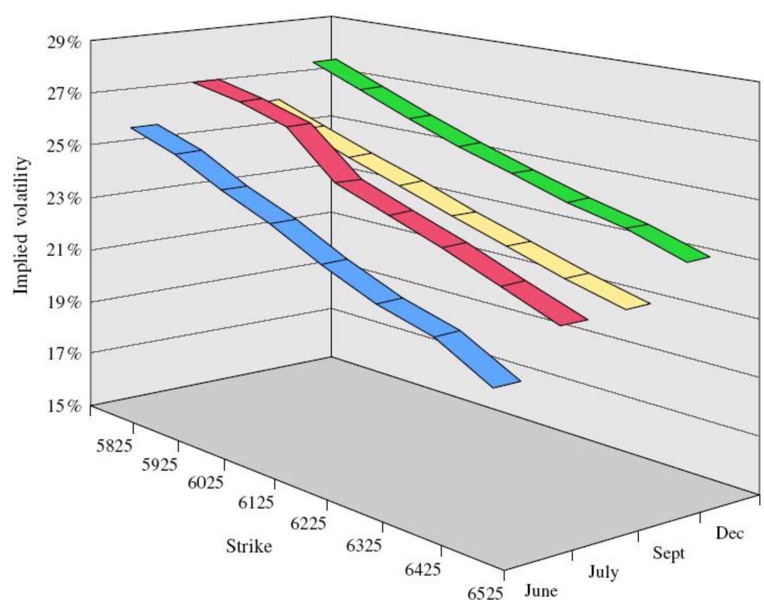
Source: Financial Times

■ **EURO STYLE FTSE 100 INDEX OPTION** (LIFFE) £10 per full index point — 19 May

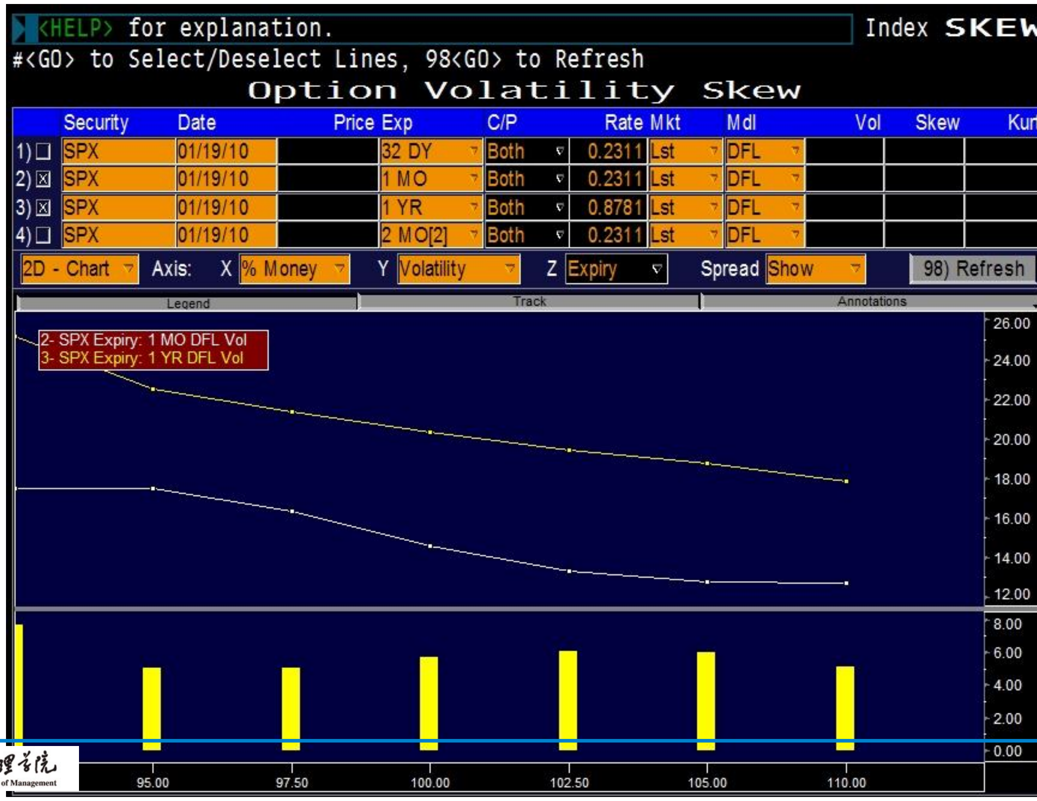
	5825		5925		6025		6125		6225		6325		6425		6525	
	C	P	C	P	C	P	C	P	C	P	C	P	C	P	C	P
May	**216½**	¼	**116½**	¼	**16½**	¼	¼	83½	¼	183½	¼	283½	¼	383½	¼	483½
Jun	**310½**	76½	**241½**	107	**179**	144	**127**	119½	**84**	248½	**52**	316	**30½**	393½	**15**	477½
Jul	**410**	144½	**347**	181	**288**	221	**224**	256	**175½**	306	**134**	363½	**98**	426½	**69**	496½
Sep	**506½**	216	**441½**	249	**380**	286	**323½**	327	**271**	373	**224**	424	**181½**	479½	**145½**	541½
Dec†	**663½**	301½	**597**	331	**533½**	364½	**474**	401½	**418½**	442½	**366½**	487½	**320**	537	**273½**	587½

Calls 15,531; Puts 32,579. * Underlying Index value. Premiums shown are based on settlement prices. † Long dated expiry months.

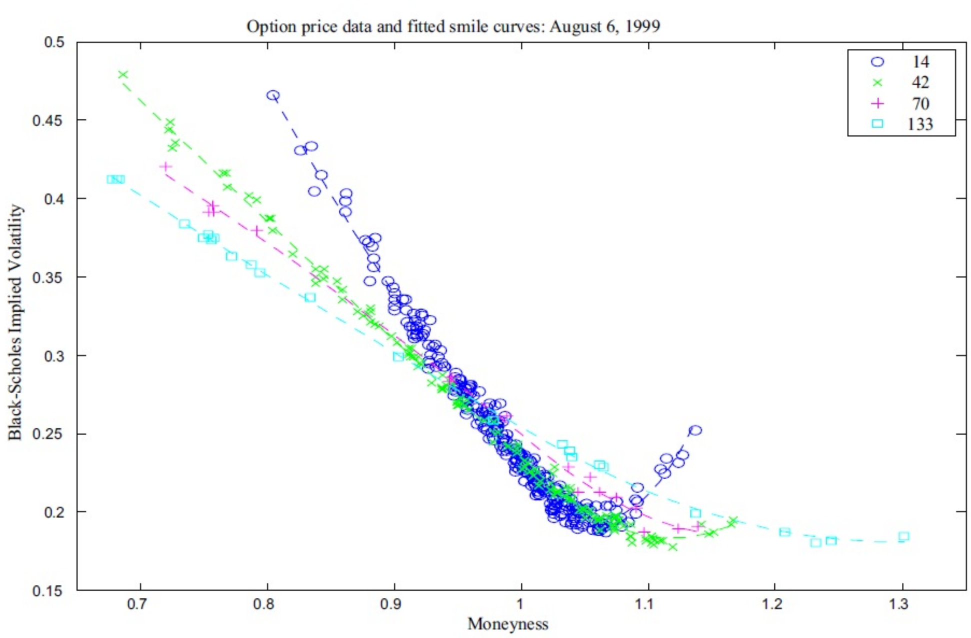
Implied Volatility Surface for FTSE100 Index Option



S&P500 Option Implied Volatility Smile from Bloomberg



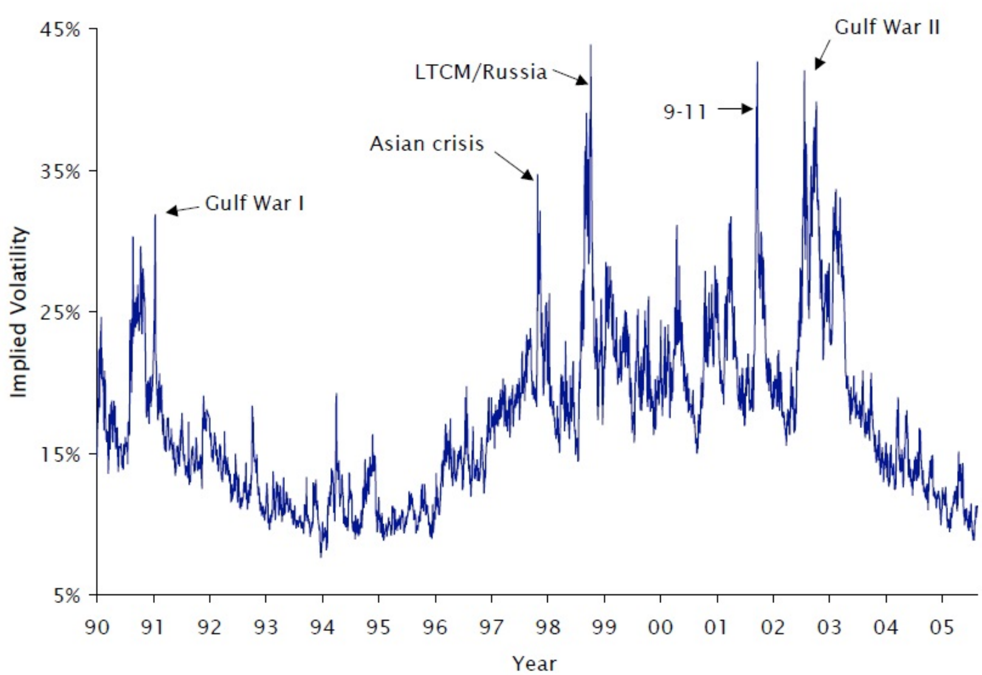
Implied Volatility Smile



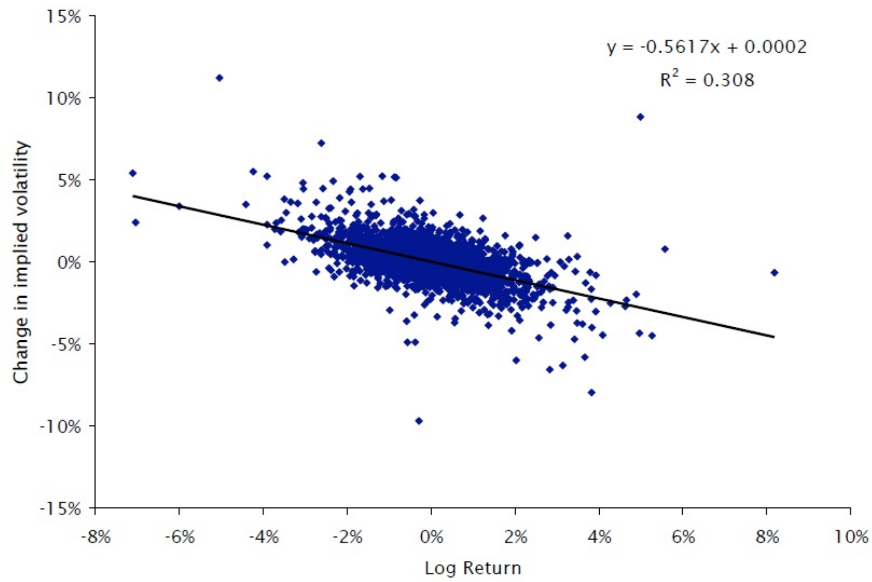
Implied Volatility Smile

- Option prices are often expressed in units of Black-Scholes implied volatility.
- Under the Blask-Scholes model, implied volatility should be constant as a function of strike price and maturity.
- Non-constant implied volatilities are direct evidence that the market does not price options with the Black-Sholes model.

S&P500 ATM Implied Volatility: 1990-2005



Change in Implied Volatility vs. Ln-Returns



The change in Implied volatility decreases when Ln-returns increases.

Empirical Evidence

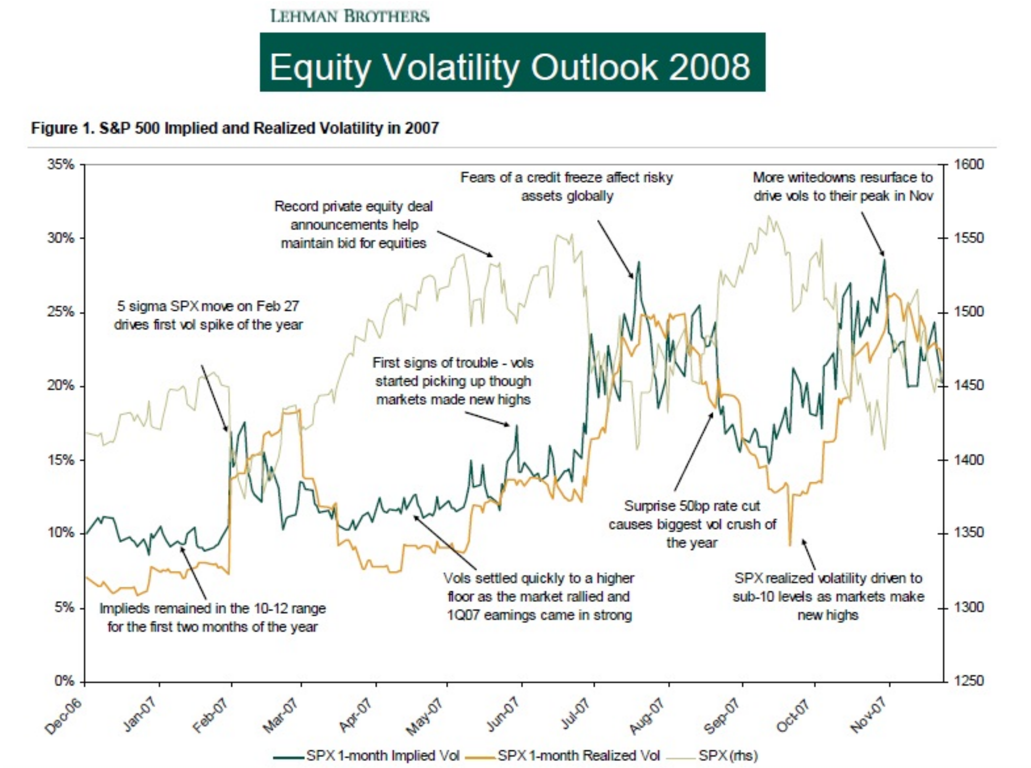
- Ln-returns exhibit fatter tails that the normal distribution suggests
- Autocorrelation of squared Ln-returns implies dependence (volatility clustering)
- Downward sloping implied volatility curve (after 1987)
- Implied volatility changes over time
- Implied volatility changes are correlated with Ln-returns

What models are consistent with this?

$$(1) \text{ and } (3) \Rightarrow \text{jumps}$$

$$(1), (2), (3), (4), \text{ and } (5) \Rightarrow \text{stochastic volatility}$$

Empirical Evidence



Stochastic Volatility Models

So based on the empirical evidences, one of the directions is to generalize the Black-Scholes model by adding stochastic volatility.

Why not model volatility in the same way a stock price is modeled (i.e., log-normal distribution)?

Properties of a stochastic volatility model:

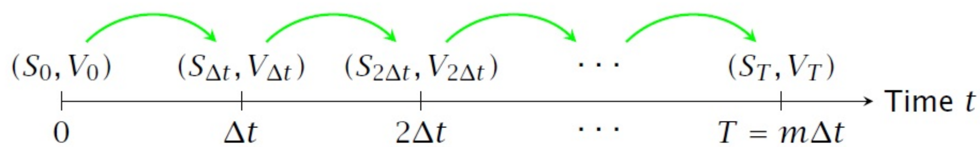
- Mean-reversion (long-run volatility parameter $\sqrt{\theta}$)
- Speed of mean-reversion (κ)
- Volatility of variance (σ_v)
- Correlation of variance and stock processes (ρ)

Formal Stochastic Volatility Model

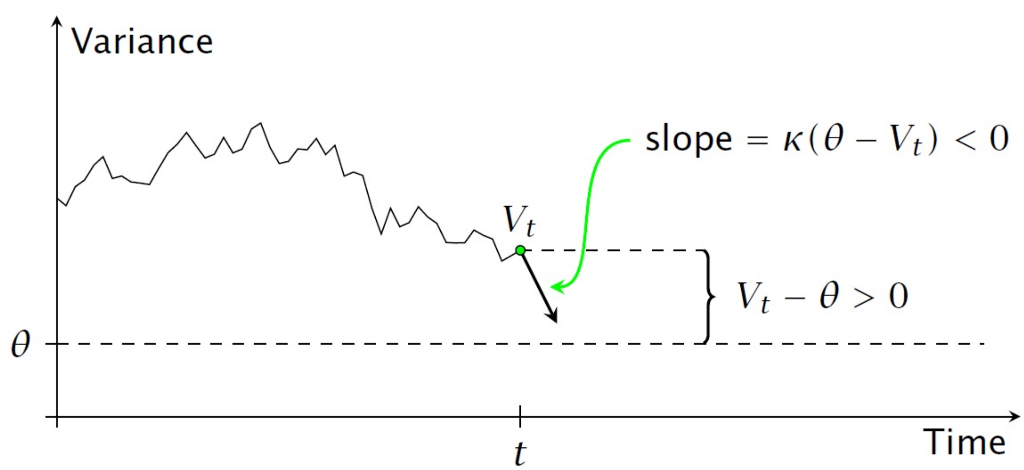
Heston (1993) Model

$$
\begin{aligned}
\frac{dS_t}{S_t} &= \mu dt + \sqrt{V_t} dW_t^1 \\
dV_t &= \kappa(\theta - V_t)dt + \sigma_v \sqrt{V_t} dW_t^2
\end{aligned}
$$

$S_t =$ stock price at time t $\quad V_t =$ variance at time t

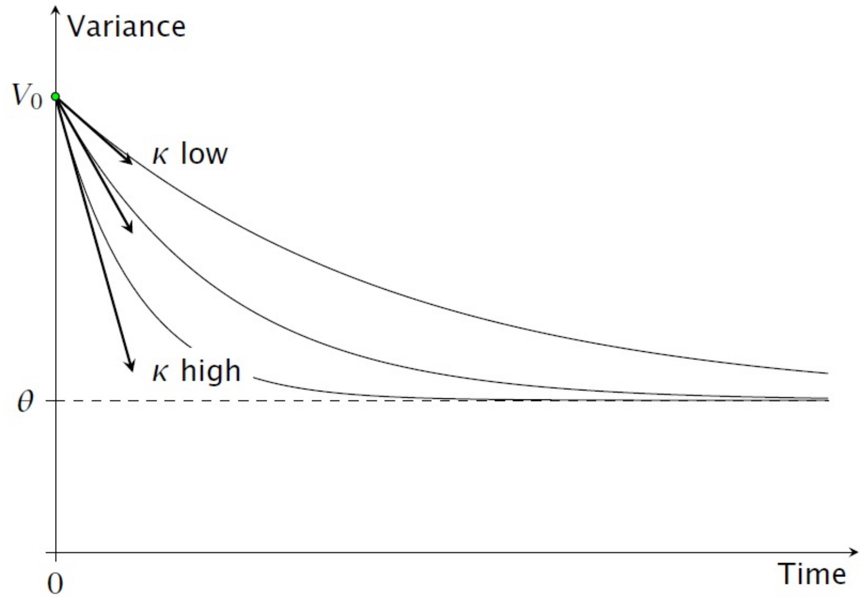


Formal Stochastic Volatility Model



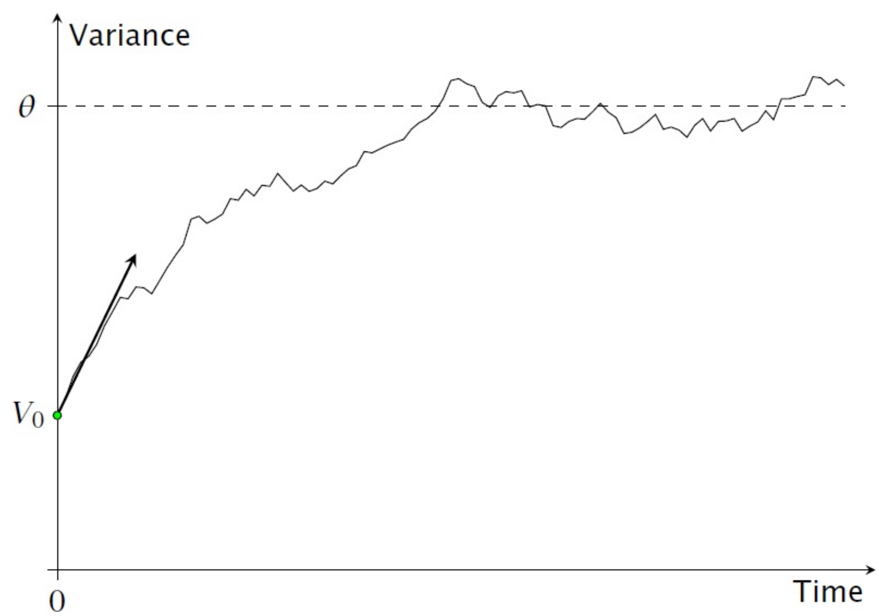
- Long-run variance θ (long-run volatility $\sqrt{\theta}$)
- Speed of mean reversion κ
- Volatility of variance σ_v
- $\sqrt{V_t}$ term guarantees positive variance (in the limit as $\Delta t \to 0$)

Mean Reversion Rate



$$\sigma_V = 0 \quad V_0 > \theta$$

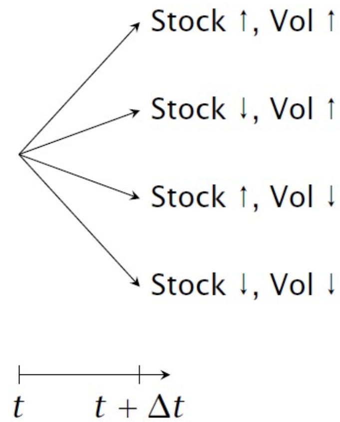
Mean Reversion Rate



$$\sigma_V = 0 \quad V_0 < \theta$$

Market Incompleteness

The stochastic volatility model is an incomplete market.



- Options cannot be replicated by dynamic trading of a stock and bond
- Can hedge volatility risk by trading options
- “Less” incomplete than jump models

How to price an option under a stochastic volatility model?

Review from Black-Sholes-Merton: An Understanding

We assume that real world dynamics of the underlying asset is

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma dW^P(t).$$

For a call option with maturity T and strike K, its price $v(t, S(t))$ satisfies

$$\begin{aligned}
& dv(t, S(t)) \\
= \; & \frac{\partial v}{\partial t} dt + \frac{\partial v}{\partial x} dS(t) + \frac{1}{2}\frac{\partial^2 v}{\partial x^2} d[S, S](t) \\
= \; & \frac{\partial v}{\partial t} dt + \frac{\partial v}{\partial x}[\mu S(t) dt + \sigma S(t) dW^P(t)] + \frac{1}{2}\frac{\partial^2 v}{\partial x^2}\sigma^2 S(t)^2 dt.
\end{aligned}$$

Review from Black-Sholes-Merton: An Understanding

By the Black-Scholes-Merton equation, the above equation deduce to

$$dv(t,S(t)) = \left(rv(t,S(t)) - rS(t)\frac{\partial v}{\partial x}\right)dt + \frac{\partial v}{\partial x}[\mu S(t)dt + \sigma S(t)dW^P(t)].$$

Thus,

$$dv(t,S(t)) - rv(t,S(t))dt = \frac{\partial v}{\partial x}\sigma S(t)\left(\frac{\mu - r}{\sigma}dt + dW^P(t)\right), \quad (13)$$

which is equivalent to

$$\frac{dv(t,S(t))}{v(t,S(t))} - rdt = \frac{\frac{\partial v}{\partial x}\sigma S(t)}{v(t,S(t))}\left(\frac{\mu - r}{\sigma}dt + dW^P(t)\right). \quad (14)$$

The term $\frac{dv(t,S(t))}{v(t,S(t))} - rdt$ can be understood as an **excess return**.

Review from Black-Sholes-Merton: An Understanding

Integrating both sides of (13) and taking conditional expectation E_t,

$$E_t v(t+\Delta, S(t+\Delta)) - v(t,S(t)) - \int_t^{t+\Delta} rE_t v(u,S(u))du$$
$$= \int_t^{t+\Delta} E_t\left(\frac{\partial v}{\partial x}S(u)\right)(\mu - r)du.$$

Thus,

$$\frac{1}{\Delta}[E_t v(t+\Delta, S(t+\Delta)) - v(t,S(t))] - \frac{1}{\Delta}\int_t^{t+\Delta} rE_t v(u,S(u))du$$
$$= \frac{1}{\Delta}\int_t^{t+\Delta} E_t\left(\frac{\partial v}{\partial x}S(u)\right)(\mu - r)du.$$

Let $\Delta \to 0$,

$$\lim_{\Delta\to 0}\frac{1}{\Delta}\frac{E_t v(t+\Delta, S(t+\Delta)) - v(t,S(t))}{v(t,S(t))} - r = \frac{\frac{\partial v}{\partial x}\sigma S(t)}{v(t,S(t))}\frac{\mu - r}{\sigma} \quad (15)$$

Market Price of Risk: the Black-Sholes-Merton Case

The Sharpe Ratio or the **market price of risk** of the underlying asset is

$$\lambda = \frac{\mu - r}{\sigma}.$$

Here, if we view

$$\frac{1}{\Delta}\frac{E_t v(t+\Delta, S(t+\Delta)) - v(t, S(t))}{v(t, S(t))}$$

as an expected return per time, (15) can be interpreted as a "CAPM" type result. Here $\frac{\frac{\partial v}{\partial x} S(t)}{v(t,S(t))}$ plays a role as the "beta".

The LHS of (15) is an instantenous excess return. We can somehow regard $\frac{\frac{\partial v}{\partial x}\sigma S(t)}{v(t,S(t))}$ as a percentage of "Brownian risk corresponding to $W^P(t)$" and $\frac{\mu-r}{\sigma}$ is the excess premium per unit of $dW^P(t)$.

A Mathematical Characterization of Market Price of Risk

We have

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma dW^P(t) = r dt + \sigma\left(\frac{\mu - r}{\sigma} dt + dW^P(t)\right).$$

No-arbitrage argument yields the risk-neutral probability measure, under which the dynamics of the underlying asset is written as

$$\frac{dS(t)}{S(t)} = r dt + \sigma dW^Q(t).$$

By the Girsanov theorem, $W^Q(t)$ can be constructed through

$$W^Q(t) = \lambda t + W^P(t) = \frac{\mu - r}{\sigma} t + W^P(t).$$

So, the market price of risk is exactly the "drift" in the Girsanov change of measure.

Stochastic Volatility Model

By analogy, we work on the stochastic volatility case in which the market with the underlying asset and a money market account is incomplete.

First, we assume the model under the physical probability measure as

$$\begin{aligned}\frac{dS(t)}{S(t)} &= \mu dt + \sigma(t) dW_1^P(t), \\ d\sigma(t) &= a(\sigma(t))dt + b(\sigma(t))\left[\rho dW_1^P(t) + \sqrt{1-\rho^2} dW_2^P(t)\right],\end{aligned}$$

where the $(W_1^P(t), W_2^P(t))$ is a two-dimensional standard Brownian motion.

Construction of Replicating Portfolio

There are underlying asset $S(t)$, a kind of asset $V_1(t)$ solely depending on volatility (e.g., a variance swap or a delta-hedged portfolio) and a money market account for us to replicate an option (with maturity T and strike K) with value $V(t)$.

To replicate an option, we use $\Delta(t)$ shares of the underlying asset with price $S(t)$, $\Delta_1(t)$ shares of an arbitrary asset with value $V_1(t)$. And put the rest in money market account. The change of the a self-financing replicating portfolio value satisfies

$$\begin{aligned}d\Pi(t) =& \Delta(t) dS(t) + \Delta_1(t) dV_1(t) \\ &+ r(\Pi(t) - \Delta(t)S(t) - \Delta_1(t)V_1(t))dt.\end{aligned} \tag{16}$$

Construction of Replicating Portfolio

We can assume that $\Pi(t) = v(t, S(t), \sigma(t))$ for some smooth function $v(t, x, y)$ and $V_1(t) = v_1(t, \sigma(t))$ for some smooth function $v(t, y)$.

$$
\begin{aligned}
& dv(t, S(t), \sigma(t)) = \frac{\partial v}{\partial t} dt + \frac{\partial v}{\partial x} dS(t) + \frac{\partial v}{\partial y} d\sigma(t) \\
+ \quad & \frac{1}{2}\frac{\partial^2 v}{\partial x^2} d[S, S](t) + \frac{1}{2}\frac{\partial^2 v}{\partial y^2} d[\sigma, \sigma](t) + \frac{\partial^2 v}{\partial x \partial y} d[S, \sigma](t) \\
= \quad & \left(\frac{\partial v}{\partial t} + \frac{\partial v}{\partial x}\mu S(t) + \frac{\partial v}{\partial y} a(\sigma(t)) + \frac{1}{2}\frac{\partial^2 v}{\partial x^2}\sigma^2(t) S^2(t) \right. \\
& \left. + \frac{1}{2}\frac{\partial^2 v}{\partial y^2}\left(b^2(\sigma(t))\right) + \frac{\partial^2 v}{\partial x \partial y} S(t)\sigma(t) b(\sigma(t))\rho \right) dt \\
+ \quad & \left(\frac{\partial v}{\partial x}\sigma(t) S(t) + \frac{\partial v}{\partial y} b(\sigma(t))\rho \right) dW_1^P(t) \\
& + \frac{\partial v}{\partial y} b(\sigma(t))\sqrt{1-\rho^2} dW_2^P(t).
\end{aligned}
$$

Construction of Replicating Portfolio

On the other hand, from (16), we also use Ito formula

$$
\begin{aligned}
& d\Pi(t) \\
= \quad & \Delta(t)[\mu S(t)dt + \sigma(t)S(t)dW_1^P(t)] + \Delta_1(t)dv_1(t, \sigma(t)) \\
& + r(v(t, S(t), \sigma(t)) - \Delta(t)S(t) - \Delta_1(t)v_1(t, \sigma(t)))dt \\
= \quad & \left(\Delta(t)\mu S(t) + \Delta_1(t)\left(\frac{\partial v_1}{\partial t} + \frac{\partial v_1}{\partial y} a(\sigma(t)) + \frac{1}{2}\frac{\partial^2 v_1}{\partial y^2} b^2(\sigma(t)) \right) \right. \\
& \left. + r(v(t, S(t), \sigma(t)) - \Delta(t)S(t) - \Delta_1(t)v_1(t, \sigma(t))) \right) dt \\
& + \left(\Delta(t)\sigma(t)S(t) + \Delta_1(t)\frac{\partial v_1}{\partial y} b(\sigma(t))\rho \right) dW_1^P(t) \\
& + \Delta_1(t)\frac{\partial v_1}{\partial y} b(\sigma(t))\sqrt{1-\rho^2} dW_2^P(t).
\end{aligned}
$$

Then, replication requires to equate the above two equations.

Replicating Strategy

Thus, we should find the following two equations for the replicating strategy $(\Delta(t), \Delta_1(t))$ as

$$\begin{aligned}\frac{\partial v}{\partial x}\sigma(t)S(t)+\frac{\partial v}{\partial y}b(\sigma(t))\rho &= \Delta(t)\sigma(t)S(t)+\Delta_1(t)\frac{\partial v_1}{\partial y}b(\sigma(t))\rho,\\ \frac{\partial v}{\partial y}b(\sigma(t))\sqrt{1-\rho^2} &= \Delta_1(t)\frac{\partial v_1}{\partial y}b(\sigma(t))\sqrt{1-\rho^2}.\end{aligned}$$

Solving this equation system, we obtain the following replicating strategy

$$\Delta_1(t)=\frac{\partial v}{\partial y}(t,S(t),\sigma(t))/\frac{\partial v_1}{\partial y}(t,\sigma(t))$$

and

$$\Delta(t)=\frac{\partial v}{\partial x}(t,S(t),\sigma(t)).$$

PDE for Option Pricing

Equate the above two equations of $d\Pi(t)$ and $dv(t,S(t),\sigma(t))$, we can also get a PDE:

$$\frac{\frac{\partial v}{\partial t}+\frac{1}{2}x^2y^2\frac{\partial^2 v}{\partial x^2}+\frac{1}{2}b^2(y)\frac{\partial^2 v}{\partial y^2}+\rho b(y)xy\frac{\partial^2 v}{\partial x\partial y}+rx\frac{\partial v}{\partial x}+a(y)\frac{\partial v}{\partial y}-rv}{\frac{\partial v}{\partial y}}$$
$$=\frac{\frac{\partial v_1}{\partial t}+\frac{1}{2}b^2(y)\frac{\partial^2 v_1}{\partial y^2}+a(y)\frac{\partial v_1}{\partial y}-rv_1}{\frac{\partial v_1}{\partial y}}.$$

Note that the RHS is only a function on the independent variable t and y. And if you have $\frac{\partial v}{\partial x}=0$, the left-hand side reduced to the right-hand side. We assume such a function to be

$$f(t,y)=\frac{\frac{\partial v_1}{\partial t}+\frac{1}{2}b^2(y)\frac{\partial^2 v_1}{\partial y^2}+a(y)\frac{\partial v_1}{\partial y}-rv_1}{\frac{\partial v_1}{\partial y}}.$$

PDE for Option Pricing

So, we have

$$\frac{\partial v}{\partial t}+\frac{1}{2}x^2y^2\frac{\partial^2 v}{\partial x^2}+\frac{1}{2}b^2(y)\frac{\partial^2 v}{\partial y^2}+\rho b(y)xy\frac{\partial^2 v}{\partial x\partial y} + rx\frac{\partial v}{\partial x}+a(y)\frac{\partial v}{\partial y}-f(t,y)\frac{\partial v}{\partial y}-rv=0 \tag{17}$$

and

$$\frac{\partial v_1}{\partial t}+\frac{1}{2}b^2(y)\frac{\partial^2 v_1}{\partial y^2}+a(y)\frac{\partial v_1}{\partial y}-f(t,y)\frac{\partial v_1}{\partial y}-rv_1=0. \tag{18}$$

- These PDEs are obtained from the replication procedure.
- Note that for option valuation, $f(t,y)$ has to be pre-specified as part of the real world model.

Feymann-Kac Representation for Option Pricing

For pricing an option with maturity T and payoff function $P(x)$, we impose the terminal condition $v(T,x,y)=P(x)$. We have

$$v(t,S(t),\sigma(t))=e^{-r(T-t)}E_t^Q P(S(T)).$$

Here, Q is the risk neutral measure under which the dynamics of $(S(t),\sigma(t))$ follows that

$$\begin{aligned}\frac{dS(t)}{S(t)} &= rdt+\sigma(t)dW_1^Q(t),\\ d\sigma(t) &= [a(\sigma(t))-f(t,\sigma(t))]dt\\ &\quad +b(\sigma(t))\left[\rho dW_1^Q(t)+\sqrt{1-\rho^2}dW_2^Q(t)\right],\end{aligned}$$

where $(W_1^Q(t),W_2^Q(t))$ is a two-dimensional standard Brownian motion under the martingale pricing measure Q. This can be shown in a similar way to the one-dimensional case discussed in Lecture 9. How to obtain such a Q from P?

$V_1(t) = v_1(t, \sigma(t))$ plays a role as the "Delta-hedged" option. Now, let us look at the excess return of such an asset. Use Ito Formula on $dv_1(t, \sigma(t))$, and apply

$$\frac{\partial v_1}{\partial t} + \frac{1}{2} b^2 (y) \frac{\partial^2 v_1}{\partial y^2} + a (y) \frac{\partial v_1}{\partial y} - f(t, y) \frac{\partial v_1}{\partial y} - r v_1 = 0.$$

We have

$$dv_1(t, \sigma(t)) - rv_1(t, \sigma(t))dt = b(\sigma(t)) \frac{\partial v_1}{\partial y} \left[\frac{f(t, \sigma(t))}{b(\sigma(t))} dt + dW_v^P(t) \right],$$

where

$$W_v^P(t) = \rho W_1^P(t) + \sqrt{1 - \rho^2} W_2^P(t)$$

represent a Brownian motion driving the volatility process.

Analogy to (15), we have

$$\lim_{\Delta \to 0} \frac{1}{\Delta} \frac{E_t v_1(t + \Delta, \sigma(t + \Delta)) - v_1(t, \sigma(t))}{v_1(t, \sigma(t))} - r = \frac{b(\sigma(t)) \frac{\partial v_1}{\partial y}}{v_1(t, \sigma(t))} \frac{f(t, \sigma(t))}{b(\sigma(t))}.$$

- This is an analog to the CAPM as many people claimed.
- The LHS is an instantaneous excess return.
- We can somehow regard $\frac{b(\sigma(t)) \frac{\partial v_1}{\partial y}}{v_1(t,\sigma(t))}$ as a percentage of "Brownian risk corresponding to $W_v^P(t)$" and $\frac{f(t,\sigma(t))}{b(\sigma(t))}$ as the excess risk premium per unit of $dW_v^P(t)$.

MPR and MPVR

Now, we look at the excess return of the option. Based on Ito formula and the pricing equation (17), we can get

$$\begin{aligned} & dv(t,S(t),\sigma(t)) - rv(t,S(t),\sigma(t))dt \\ = \; & \sigma(t)\,S(t)\,\frac{\partial v}{\partial x}\left[\frac{\mu - r}{\sigma(t)}dt + dW_1^P(t)\right] \\ & + b(\sigma(t))\,\frac{\partial v_1}{\partial y}\left[\frac{f(t,\sigma(t))}{b(\sigma(t))}dt + dW_v^P(t)\right]. \end{aligned}$$

Thus,

$$\lim_{\Delta\to 0}\frac{1}{\Delta}\frac{E_t v(t+\Delta,S(t+\Delta),\sigma(t+\Delta)) - v(t,S(t),\sigma(t))}{v(t,S(t),\sigma(t))} - r =$$

$$\frac{S(t)\,\frac{\partial v}{\partial x}}{v(t,S(t),\sigma(t))}(\mu - r) + \frac{b(\sigma(t))\,\frac{\partial v_1}{\partial y}}{v(t,S(t),\sigma(t))}\frac{f(t,\sigma(t))}{b(\sigma(t))}.$$

$\Rightarrow$ economic interpretation for option return.

MPR and MPVR

Now, we formally call

$$\lambda_1(t) = \frac{\mu - r}{\sigma(t)}, \quad \lambda_2(t) = \frac{f(t,\sigma(t))}{b(\sigma(t))}$$

as the **market price of risk**.

- We can call $\lambda_1(t)$ the market price of return risk (MPR) and call $\lambda_2(t)$ the market price of volatility risk (MPVR).
- Again, note that these two items render the drifts in the Girsanov change of measure (from the physical measure P to the risk-neutral measure Q).

MPR and MPVR

Similar to previous discussion in one-dimensional case, we just need to find two drifts such that

$$\begin{aligned}\frac{\mu - r}{\sigma(t)}dt + dW_1^P(t) &= \gamma_1(t)dt + dW_1^P(t) \\ \frac{f(t,\sigma(t))}{b(\sigma(t))}dt + dW_v^P(t) &= \rho\left[\gamma_1(t)dt + dW_1^P(t)\right] \\ &+ \sqrt{1-\rho^2}\left[\gamma_2(t)dt + dW_2^P(t)\right],\end{aligned}$$

i.e.,

$$\begin{aligned}\gamma_1(t) &= \frac{\mu - r}{\sigma(t)}, \\ \rho\gamma_1(t) + \sqrt{1-\rho^2}\gamma_2(t) &= \frac{f(t,\sigma(t))}{b(\sigma(t))}.\end{aligned}$$

The Girsanov Theorem: Multi-dimensional Case

Theorem. Let $W(t) = (W_1(t), ..., W_d(t), 0 \le t \le T$ be a d-dimensional Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\{\mathcal{F}(t);\ 0 \le t \le T\}$ be a filtration for this Brownian motion. Let $\Theta = (\Theta_1(t), ..., \Theta_d(t))$ is a d-dimensional adapted process. Define

$$Z(t) = \exp\left(-\int_0^t \Theta(u)\dot{d}W(u) - \frac{1}{2}\int_0^t \|\Theta(u)\|^2 du\right),$$

and

$$\widetilde{W}(t) = W(t) + \int_0^t \Theta(u)du.$$

Assume that

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

Then under the probability measure $\widetilde{\mathbb{P}}(A) = \int_A Z(\omega)d\widetilde{\mathbb{P}}(\omega),\ \forall A \in \mathcal{F}$, the process $\widetilde{W}(t),\ 0 \le t \le T$ is a Brownian motion.

MPR and MPVR under Girsanov Theorem

Thus, we can construct the probability measure Q through

$$\left.\frac{dQ}{dP}\right|_{\mathcal{F}_t} = \exp\left(\int_0^t \gamma_1(s)dW_1^P(s) + \int_0^t \gamma_2(s)dW_2^P(s) \right. \\ \left. -\frac{1}{2}\int_0^t \gamma_1(s)^2 ds - \frac{1}{2}\int_0^t \gamma_2(s)^2 ds\right).$$

So, under Q,

$$\begin{aligned} W_1^Q(t) &= \int_0^t \gamma_1(s)ds + W_1^P(t), \\ W_2^Q(t) &= \int_0^t \gamma_2(s)ds + W_2^P(t), \end{aligned}$$

is a standard two-dimensional Brownian motion.

MPR and MPVR under Girsanov Theorem

Now, we can see that under Q, the original model can be expressed as

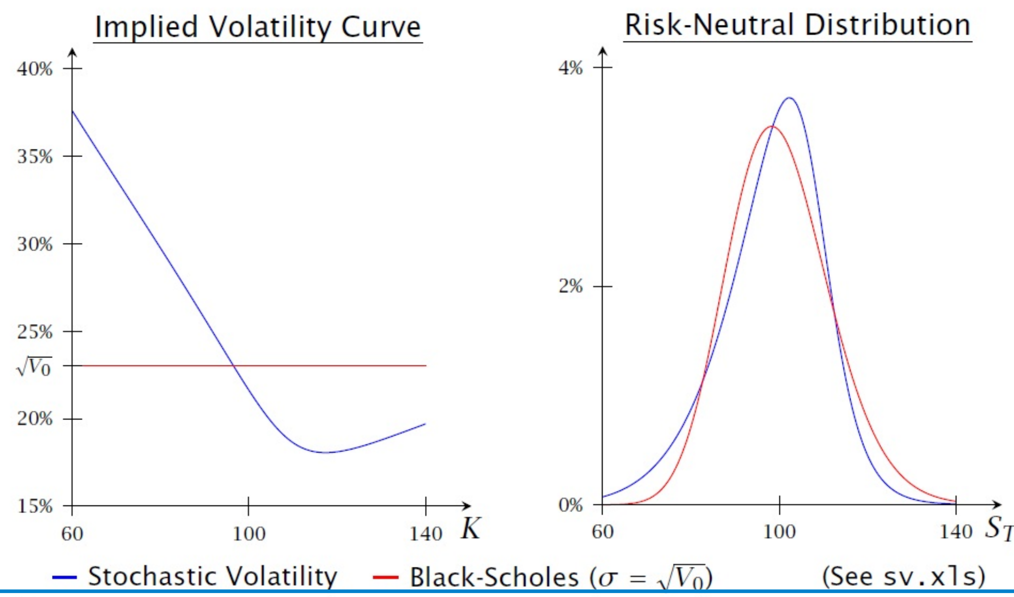
$$\begin{aligned} \frac{dS(t)}{S(t)} &= rdt + \sigma(t)dW_1^Q(t), \\ d\sigma(t) &= [a(\sigma(t)) - f(t,\sigma(t))]dt \\ &\quad + b(\sigma(t))\left[\rho dW_1^Q(t) + \sqrt{1-\rho^2}dW_2^Q(t)\right], \end{aligned}$$

Further tasks:

- find a closed-form formula for option pricing, or,
- numerically evaluate the option price
- then, calibrate the model (find proper value of parameters) by fitting the formula to option price data

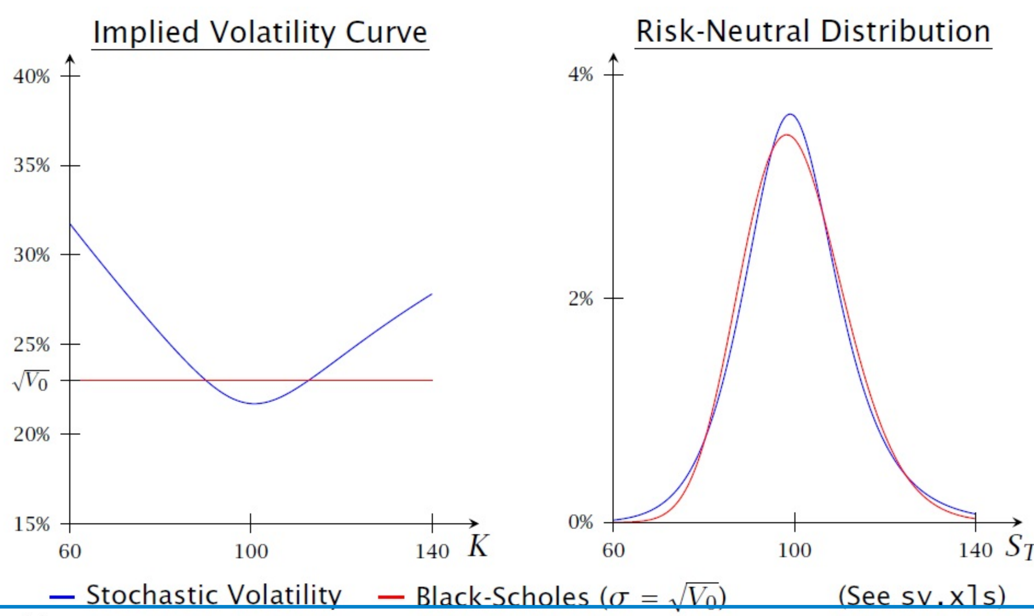
Possible calibration results for the Heston SV model

$$\begin{aligned} r &= 3\%,\ S_0 = 100,\ \sqrt{V_0} = \sqrt{\theta} = 23\%, \\ \kappa &= 4,\ \sigma_V = 0.8,\ \rho = -60\%, T = 0.25 \end{aligned}$$



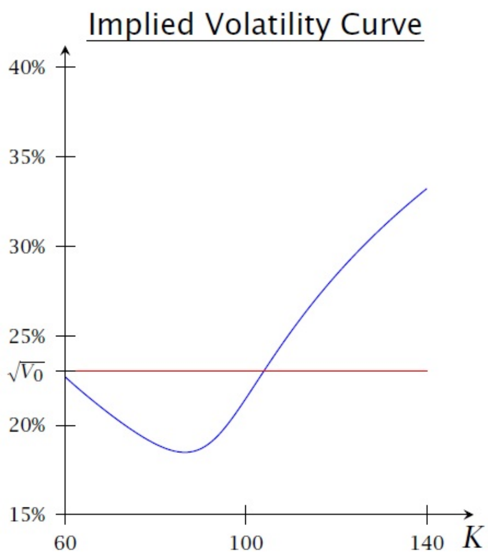
Possible calibration results for the Heston SV model

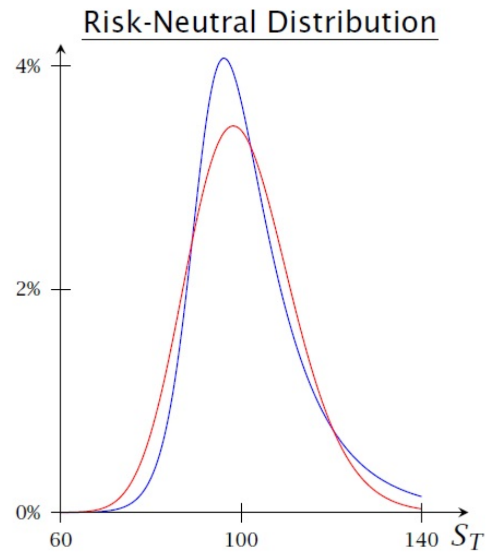
$$\begin{aligned} r &= 3\%,\ S_0 = 100,\ \sqrt{V_0} = \sqrt{\theta} = 23\%, \\ \kappa &= 4,\ \sigma_V = 0.8,\ \rho = 0\%, T = 0.25 \end{aligned}$$



Possible calibration results for the Heston SV model

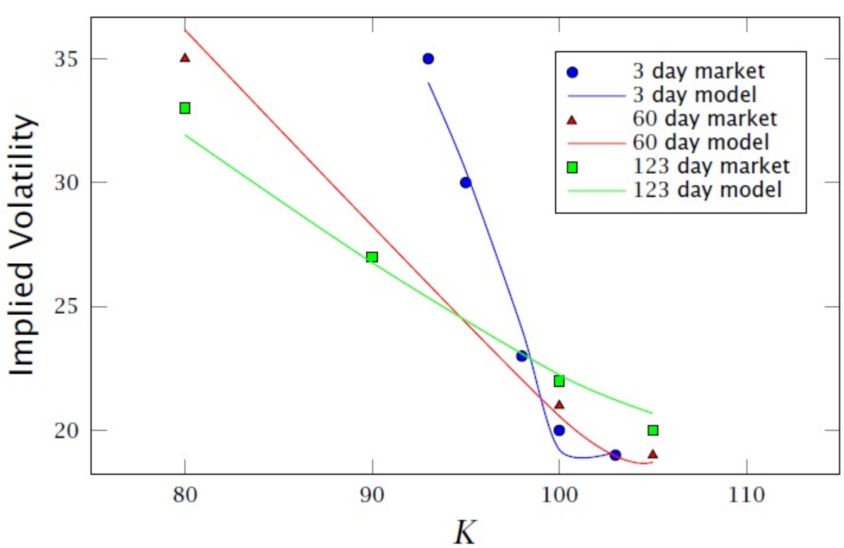
$$
\begin{aligned}
r &= 3\%,\ S_0 = 100,\ \sqrt{V_0} = \sqrt{\theta} = 23\%, \\
\kappa &= 4,\ \sigma_V = 0.8,\ \rho = 60\%, T = 0.25
\end{aligned}
$$





Stochastic Volatility — Black-Scholes ($\sigma = \sqrt{V_0}$) (See sv.xls)

Results of Fitting

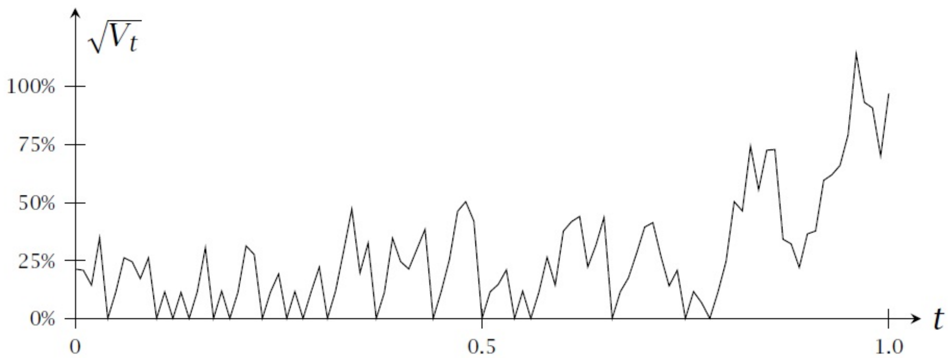


$S_0 = 100$	$\sqrt{V_0} = 21.38\%$	$\sqrt{\theta} = 26.15\%$	$\kappa = 19.66$
$r = 5\%$	$\sigma_V = 4.25$	$\rho = -44.55\%$	

Results of Fitting

$$S_0 = 100 \qquad \sqrt{V_0} = 21.38\% \qquad \sqrt{\theta} = 26.15\% \qquad \kappa = 19.66$$
$$r = 5\% \qquad \sigma_V = 4.25 \qquad \rho = -44.55\%$$

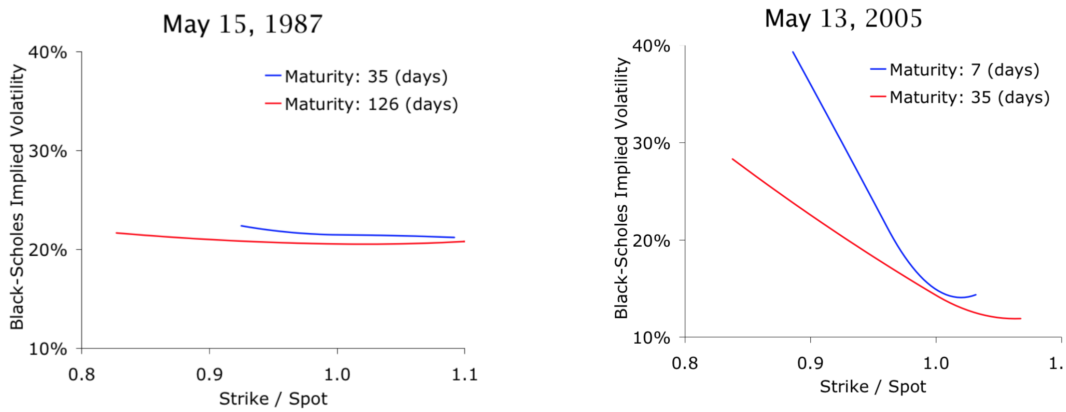
Resulting Sample Path:



Overview

- Beyond Black-Scholes-Merton: jumps

Implied Volatility Curves: Pre- and Post 1987



- Prior to 1987 crash: implied volatilities were nearly constant (consistent with Black-Scholes)
- Post 1987: implied volatility curves or smiles or smirks (no longer consistent with Black-Scholes)

Merton Jump-Diffusion Model

Empirical Evidence

1. Log-returns exhibit fat tails
2. Volatility clustering
3. Downward sloping implied volatility curve
4. Implied volatility changes over time
5. Implied volatility changes are correlated with log-returns

The Merton jump-diffusion model attempts to address (1) and (3) by allowing for jumps in asset prices.

Implied vol curve $\Leftrightarrow$ RN dist of ST $\Leftrightarrow$ Fat tails/jumps

Jump-Diffusion Models

We'd like to generalize the Black-Scholes (diffusion) model by adding jumps A jump is a large price change in the underlying asset

in a time period so short that trades cannot be executed as the price goes up or down

When do jumps occur?

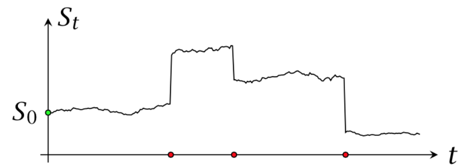
- Random or known?
- How frequently do jumps occur?

What are the jump sizes when they occur? When do jumps occur?

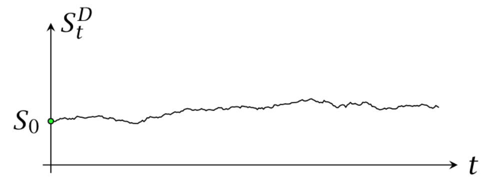
- What mean and variance?
- What distribution?

Jump-Diffusion Models

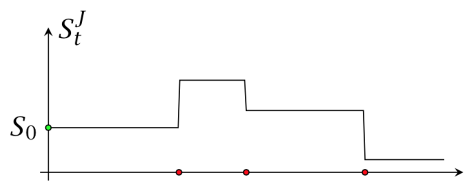
Jump-diffusion



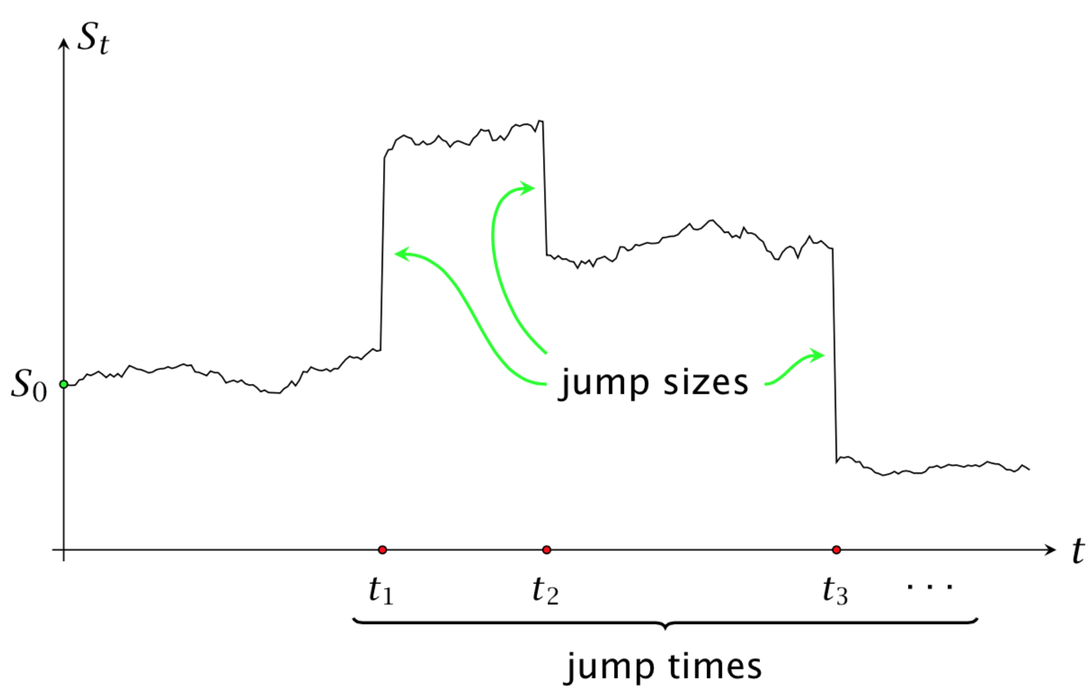
Pure diffusion
(Black-Scholes)



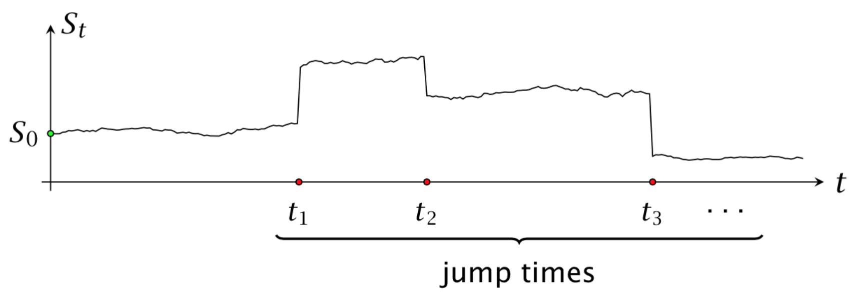
Pure jump



How to Include Jumps?



Jump Times



Jump times $t_1, t_2, t_3, \ldots$ are distributed according to a Poisson process:

$$\lambda = \text{arrival rate (jumps/year)}$$

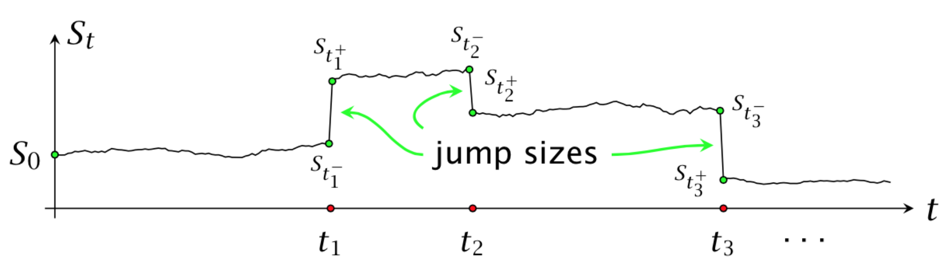
$$N(t) = \text{number of jumps in}[0, t].$$

We have

$$P(N(t) = i) = \frac{e^{-\lambda t}(\lambda t)^i}{i!}$$

for $i \geq 0$

Jump Times



If a jump occurs at time t, $S(t)$ changes from $S(t-)$ to $S(t)$:

$$S(t) = XS(t-)$$

Jump-diffuion Model

We start from the physical probability measure $\mathbb{P}$. We employ a compound Poisson process (CPP hereafter) $Q(t) = \sum_{n=1}^{N(t)} Y_n$ with $Y_n = X_n - 1$, where X_n has i.i.d. distributions with mean $\mu_S + 1$ ($\mathbb{E}^{\mathbb{P}} Y_n = \mu_S$). We note that the compensated compound Poisson (CCPP hereafter) process $Q(t) - \lambda\mu_S t$ is a martingale (exercise). The Merton jump-diffusion model is specified as follows

$$\frac{dS(t)}{S(t-)} = \mu dt + \sigma dW(t) + d[Q(t) - \lambda\mu_S t].$$

The reason why we employ the CCPP $Q(t) - \lambda\mu_S t$ in the model is that we hope to take the advantage of its martingale property. Intuitively, we have

$$\mathbb{E}_t^{\mathbb{P}}\left(\frac{dS(t)}{S(t-)}\right) = \mu dt + \sigma\mathbb{E}_t^{\mathbb{P}} dW(t) + \mathbb{E}_t^{\mathbb{P}} d[Q(t) - \lambda\mu_S t] = \mu dt,$$

i.e., the local mean rate of return is μ. This is in analogy to the Black-Scholes case.

Explicit Solution

This equation can be equivalently written as

$$\begin{aligned} \frac{dS(t)}{S(t-)} &= (\mu - \lambda\mu_S)dt + \sigma dW(t) + dQ(t) && (19) \\ &= (\mu - \lambda\mu_S)dt + \sigma dW(t) + d\left(\sum_{n=1}^{N(t)} (X_n - 1)\right). \end{aligned}$$

The solution of this equation is given by

$$S(t) = S(0)\exp\left(\sigma W(t) + \left(\mu - \lambda\mu_S - \frac{1}{2}\sigma^2\right)t\right)\prod_{n=1}^{N(t)} X_n. \qquad (20)$$

A formal proof of this claim requires systematical knowledge in stochastic calculus with jumps, see, e.g., Chapter 11 in Shreve's book. We provide an intuitive argument to convince ourselves.

Obtain the Risk-Neutral

- Suppose we hope to price a call option with strike K and maturity T. How jump-diffusions are applied in option pricing?
- Because of the nature of the jump component, it is not possible to use a finite number of assets to replicate option payoff. (Please have a try.)
- So, we seek for a solution in weaker sense. We know the following principle. If there exists a risk-neutral measure, under which all discounted assets price are martingales, the market is free of arbitrage.
- Thus, once we have such a risk-neutral measure, denoted by $\mathbb{Q}$, we don't need to worry about the existence of arbitrage, as assumed by any efficient markets.

Obtain the Risk-Neutral

- Assume that the $\mathbb{Q}-$dynamics of $\{S(t)\}$ has the following pattern

$$\begin{aligned}\frac{dS(t)}{S(t-)} &= \mu^{\mathbb{Q}}dt + \sigma dW^{\mathbb{Q}}(t) + d[Q(t) - \lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}t] \qquad (21)\\ &= \left[\mu^{\mathbb{Q}} - \lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}\right]dt + \sigma dW^{\mathbb{Q}}(t) + dQ(t).\end{aligned}$$

- The drift changes from $\mu - \lambda\mu_S$ to $\mu^{\mathbb{Q}} - \lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}$.
- A necessary condition for $\mathbb{Q}$ being the desired risk-neutral measure is to make sure the discounted price $e^{-rt}S(t)$ becomes a $\mathbb{Q}-$martingale.
- When $\mu^{\mathbb{Q}} = r$ and $Q(t) = \sum_{n=1}^{N(t)} Y_n$ is a CPP with intensity $\lambda^{\mathbb{Q}}$ and $\mathbb{E}^{\mathbb{Q}}Y_n = \mu_S^{\mathbb{Q}}$, it becomes true.
- proof of martingale condition by calculations

Change of measure

- The change from measure from $\mathbb{P}$ to $\mathbb{Q}$ consists of two parts: the Brownian motion part and the jump part, respectively.
- Comparing the $\mathbb{P}$ and $\mathbb{Q}-$dynamics of $S(t)$, we have

$$(\mu - \lambda\mu_S)\,dt + \sigma dW(t) + dQ(t) \equiv \left(r - \lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}\right)dt + \sigma dW^{\mathbb{Q}}(t) + dQ(t).$$

  For the Brownian motion part, simular to Black-Scholes, we consider the Girsanov type change of measure such that

$$W^{\mathbb{Q}}(t) = W(t) + \theta t,$$

  becomes a Brownian motion, where

$$\theta = \frac{1}{\sigma}\left(\mu - r + \lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}} - \lambda\mu_S\right). \tag{22}$$

- According to the Girsanov theorem, the Brownian part of the Radon-Nykodim derivative can be given by

$$Z_1(t) = \exp\left(-\theta W(t) - \frac{1}{2}\theta^2 t\right).$$

Change of measure

- For the jump part, by applying the following Radon-Nykodim derivative:

$$Z_2(t) = \exp((\lambda - \lambda^{\mathbb{Q}})t)\prod_{n=1}^{N(t)}\frac{\lambda^{\mathbb{Q}} f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)},$$

  the intensity of the CPP changes from λ to $\lambda^{\mathbb{Q}}$ and the density the jump size changed from $f_{\mathbb{P}}$ to $f_{\mathbb{Q}}$.
- Prove this claim as an exercise. First, prove that $Z_2(t)$ is a martingale using the definition directly. Then, prove the change of measure effect by calculating moment generating functions.
- According to our current context, we just need to make sure $f_{\mathbb{Q}}$ satisfies

$$\mathbb{E}^{\mathbb{Q}} Y_n = \int_R y f_{\mathbb{Q}}(y)dy = \mu_S^{\mathbb{Q}}.$$

Change of measure

- Putting the two parts together, the risk-neutral measure can be constructed from the following

$$\left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathcal{F}(T)} = Z_1(T)Z_2(T)$$

- $\lambda^{\mathbb{Q}}$ and $\mu_S^{\mathbb{Q}}$ show up in a product in the market price of risk (22); so are λ and μ_S. So, to make sure jump generates contribution in the market price of risk, it is enough to allow change either in intensity or jump size.
- In econometrics, it is conventional to make a trade-off by changing the intensity while keeping the jumps size distribution. In such a case, the Radon-Nykodim derivative for jump part becomes

$$Z_2(t) = \exp((\lambda - \lambda^{\mathbb{Q}})t)\left(\frac{\lambda^{\mathbb{Q}}}{\lambda}\right)^{N(t)}.$$

A closed-form formula for option pricing

- Merton (1976) assumes that X_n has a lognormal distribution. More precisely, under the risk-neutral probability measure $\mathbb{Q}$, we assume $Z_n = \log X_n$ have a normal distribution with mean $\mu_Z^{\mathbb{Q}}$ and variance $\left(\sigma_Z^{\mathbb{Q}}\right)^2$. Suppose, we hope to price an option with
- Since $\mathbb{Q}$ is assumed to be the risk-neutral probability measure, the discounted option price process $\{V(t)\}$ has to be a $\mathbb{Q}$-martingale also. Thus, we must have

$$V(0) = \mathbb{E}^{\mathbb{Q}}[e^{-rT}V(T)] = \mathbb{E}^{\mathbb{Q}}[e^{-rT}(S(T) - K)^+].$$

- It follows from conditioning that

$$V(0) = \sum_{n=0}^{\infty} e^{-rT}\mathbb{E}^{\mathbb{Q}}[(S(T) - K)^+ | N(T) = n]\mathbb{Q}(N(T) = n).$$

A closed-form formula for option pricing

- Conditioning on $N(T) = n$,

$$S_0(T) = s_0 \exp\left(\left(r - \lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}} - \frac{1}{2}\sigma^2\right)T + \sigma W_1(T) + \sum_{n=1}^{n} Z_n\right)$$

has a lognormal distribution.
- Standard calculation yields that

$$V(0) = \sum_{n=0}^{\infty} e^{-\lambda^{\mathbb{Q}}T}\frac{(\lambda^{\mathbb{Q}}T)^n}{n!}\left[s_0 e^{-\lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}T + n\left(\mu_Z^{\mathbb{Q}} + \frac{(\sigma_Z^{\mathbb{Q}})^2}{2}\right)} N(d_{1,n}) - e^{-rT}KN(d_{2,n})\right],$$

with

$$d_{1,n}(s) \;:=\; \frac{1}{\sqrt{\sigma^2 T + n(\sigma_Z^{\mathbb{Q}})^2}}\left[\log\frac{s_0}{s} + \left(r - \lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}} + \frac{\sigma^2}{2}\right)T + n(\mu_Z^{\mathbb{Q}} + (\sigma_Z^{\mathbb{Q}})^2)\right],$$

$$d_{2,n}(s) \;:=\; \frac{1}{\sqrt{\sigma^2 T + n(\sigma_Z^{\mathbb{Q}})^2}}\left[\log\frac{s_0}{s} + \left(r - \lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}} - \frac{\sigma^2}{2}\right)T + n\mu_Z^{\mathbb{Q}}\right].$$

Tips for computing

- Formula is a weighted average of Black-Scholes prices
- Infinite sum since the number of jumps in $[0, T]$ is unlimited
- In practice, the probability of many jumps in $[0, T]$ is quite small, so typically only a small number of terms are needed for an accurate result
- Keep adding terms until the terms are negligible
- Implementation of $p_j(\lambda')$
  - Note that $\frac{(\lambda^{\mathbb{Q}}T)^{n+1}}{(n+1)!} = \frac{\lambda^{\mathbb{Q}}T}{n+1}\frac{(\lambda^{\mathbb{Q}}T)^n}{n!}$
  - Compute such component by iteration; never compute $(\lambda^{\mathbb{Q}}T)^n$ and then $n!$ and then divide

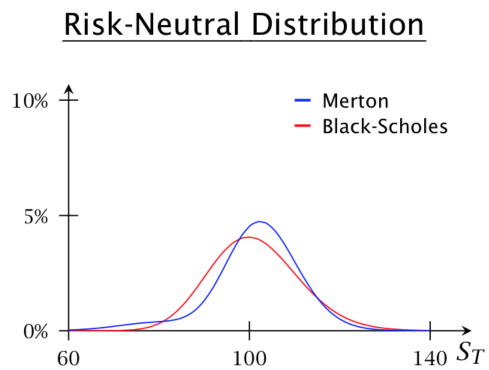
Risk-Neutral Distribution

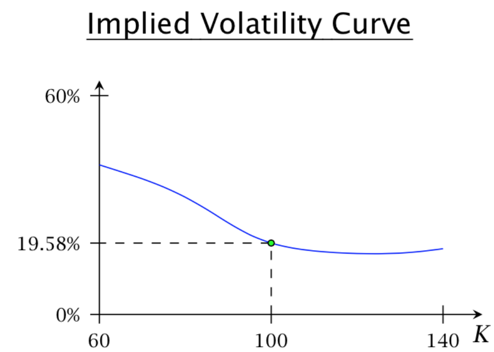
Let $\mathbb{E}X_n = \sigma_S^2$.

$$S_0 = 100 \quad r = 5\% \quad T = 0.25$$

$$\sigma = 15\% \quad \lambda = 40\% \quad \mu_S = -20\% \quad \sigma_S = 15\%$$

Merton price of ATM call = 4.532
Black-Scholes implied volatility = 19.58%



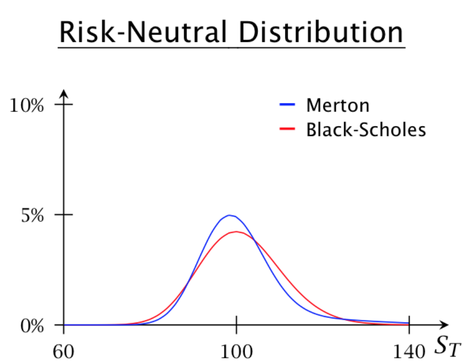


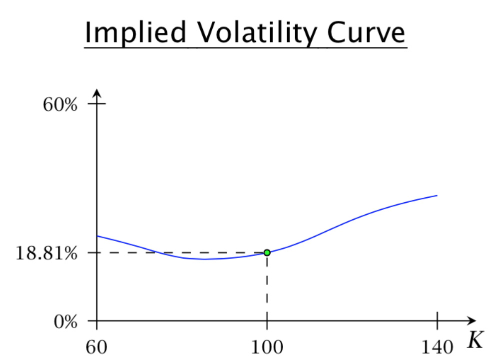
Risk-Neutral Distribution

$$S_0 = 100 \quad r = 5\% \quad T = 0.25$$

$$\sigma = 15\% \quad \lambda = 40\% \quad \mu_S = 20\% \quad \sigma_S = 15\%$$

Merton price of ATM call = 4.381
Black-Scholes implied volatility = 18.81%





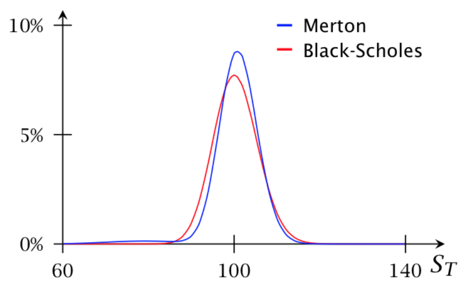
Risk-Neutral Distribution

$$S_0 = 100 \;\; r = 5\% \;\; T = \frac{1}{12}$$

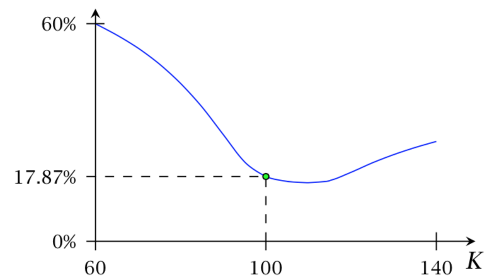
$$\sigma = 15\% \;\; \lambda = 40\% \;\; \mu_S = -20\% \;\; \sigma_S = 15\%$$

Merton price of ATM call = 2.268
Black-Scholes implied volatility = 17.87%

Risk-Neutral Distribution



Implied Volatility Curve



Combining stochastic volatility and jumps

As an extension of stochastic volatility and jumps, we combine them to have stochastic volatility with jump models (SVJ), e.g.,

$$\begin{aligned}\frac{dS(t)}{S(t)} &= \mu dt + \sqrt{V(t)}dW_1^P(t) + d[Q(t) - \lambda\mu_S t],\\ dV(t) &= a(V(t))dt + b(V(t))\left[\rho dW_1^P(t) + \sqrt{1-\rho^2}dW_2^P(t)\right].\end{aligned}$$

And, we can even consider models with jumps in volatility (SVJJ), e.g.,

$$\begin{aligned}\frac{dS(t)}{S(t)} &= \mu dt + \sqrt{V(t)}dW_1^P(t) + d[Q(t) - \lambda\mu_S t],\\ dV(t) &= a(V(t))dt + b(V(t))\left[\rho dW_1^P(t) + \sqrt{1-\rho^2}dW_2^P(t)\right] + dJ_V(t).\end{aligned}$$

The way to study these models involves the previous methods for studying stochastic volatility and jumps.

Overview

- Affine jump-diffusions and Fourier pricing methods

Affine jump-diffusions

- a large class of continuous-time models
- many famous diffusion and jump-diffusion models belong to this class
- Affine processes admit several nice properties, e.g., analytical tractability in derivatives pricing
- initiated formally in Duffie, Pan and Singleton (2000, Econometrica)

Definition: affine jump-diffusions

Consider a process $X(t)$ governed by the following SDE

$$dX(t) = \mu\left(X(t)\right)dt + \sigma\left(X(t)\right)dW(t) + dJ(t), \qquad (23)$$

for some functions $\mu(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma(\cdot) : \mathbb{R}^d \to \mathbb{R}^{d\times d}$, where $\{W(t)\}$ is a $d-$dimensional standard Brownian motion; $\{J(t)\}$ is a jump process with jump intensity $\lambda\left(X(t)\right)$ for some function $\lambda(\cdot)$. Process (23) is affine if and only if the following specifications hold

$$\begin{aligned} \mu(x) &= K_0 + K_1 x, \text{ with } K_0 \in \mathbb{R}^d, K_1 \in \mathbb{R}^{d\times d}, \\ \left(\sigma(x)\sigma^\top(x)\right)_{ij} &= (H_0)_{ij} + (H_1)_{ij} x, \text{ with } (H_0)_{ij} \in \mathbb{R} \text{ and } (H_1)_{ij} \in \mathbb{R}^d, \\ \lambda(x) &= \lambda_0 + \lambda_1^\top x, \text{ with } \lambda_0 \in \mathbb{R}, \lambda_1 \in \mathbb{R}^d. \end{aligned}$$

After some suitable transformations (e.g., taking log of asset price), many popular continuous-time models fall into this class.

An alternative expression of affine jump-diffusions

Indeed, a $d-$dimensional affine jump-diffusion model can be equivalently specified as follows (exercise, please refer to Proposition A.1 and Corollary A.1 in Duffie and Kan (1996))

$$dX(t) = \mathcal{K}(\Theta - X(t))dt + \Sigma\sqrt{V(t)}dW(t) + dJ(t),$$

where $\{W(t)\}$ is a $d-$dimensional standard Brownian motion; $\mathcal{K}$ and Σ are $d \times d$ matrices; V is a diagonal matrix with the ith diagonal element given by

$$V_{ii}(t) = \alpha_i + \beta_i^\top X(t).$$

In other words, we have

$$V(t) = \begin{pmatrix} \alpha_1 + \beta_1^\top x & & & \\ & \alpha_2 + \beta_2^\top x & & \\ & & \ddots & \\ & & & \alpha_d + \beta_d^\top x \end{pmatrix}.$$

An exercise

According to the definition, show that the Heston stochastic volatility model is affine:

$$
\begin{aligned}
dS(t) &= \mu S(t)dt + \sqrt{V(t)}S(t)dW^S(t), \\
dV(t) &= \kappa\left(\theta - V(t)\right)dt + \xi\sqrt{V(t)}dW^V(t),
\end{aligned}
$$

with ρ representing the correlation between the standard Brownian motions $\{W^S(t)\}$ and $\{W^V(t)\}$. According to the definition, show that the Parallel double Heston model

$$
\begin{aligned}
dS(t) &= rS(t)dt + \sqrt{V_1(t)}S(t)dW_1(t) + \sqrt{V_2(t)}S(t)dW_2(t), \\
dV_1(t) &= \left(a_1 - b_1 V_1(t)\right)dt + \sigma_1\sqrt{V_1(t)}dW_3(t), \\
dV_2(t) &= \left(a_1 - b_1 V_2(t)\right)dt + \sigma_2\sqrt{V_2(t)}dW_4(t),
\end{aligned}
$$

where $W_1(t)$ and $W_3(t)$, $W_2(t)$ and $W_4(t)$ are constantly correlated.

Analytical transforms: Fourier transforms

- Most affine processes do not admit closed-form transition densities, but their Laplace transforms are explicitly available.
- The expectation $\mathbb{E}\left(e^{-sX(t)}\right)$ is in closed-form for any complex number s.
- Suppose X is a one-dimensional continuous random variable and its law is implicit.
- Characteristic function (CF hereafter) is defined by

$$
f(t) := \mathbb{E}\left(e^{itX}\right) = \int_{-\infty}^{+\infty} e^{itx}p(x)dx,
$$

where $p(x)$ is law of X.

- Indeed, the function $f(\cdot)$ is the Fourier transform of $p(\cdot)$.
- By applying inverse Fourier transform, we will get $p(x)$.
- Fortunately, the characteristic function $f(t)$ usually admits explicit form.

Analytical transform: Laplace transforms

- A related notion is Laplace transform

$$L(s) := \mathbb{E}\left(e^{-sX}\right) = \int_{-\infty}^{+\infty} e^{-sx} p(x) dx,$$

  for any complex number s. If we ignore some technical details, we can also deduce the law of X by inverse Laplace transform.
- Similarly, if X is a random vector, Laplace transform can be defined as

$$L(s) = \mathbb{E}\left(e^{-s^\top X}\right) = \int_{-\infty}^{+\infty} e^{-s^\top x} p(x) dx, \tag{24}$$

  where s is a column vector.

Analytical transform: Laplace transforms

Motivated by some applications, we can replace expectation (24) by conditional expectation, which is called "conditional Laplace transform". Duffie and Kan (1996) and Duffie, Pan and Singleton (2000) indicate that the conditional Laplace transform of affine processes admits exponential affine form, i.e.,

$$\mathbb{E}\left(e^{-s^\top X(T)} | \mathcal{F}(t)\right) = \exp\left(A(T-t) + B^\top (T-t) X(t)\right),$$

i.e.,

$$\mathbb{E}\left(e^{-s^\top X(T)} | X(t) = x\right) = \exp\left(A(T-t) + B^\top (T-t) x\right),$$

a group of ODEs, subject to initial conditions

$$A(0) = 0 \text{ and } B(0) = -s. \tag{25}$$

ODEs

Martingale properties of $\mathbb{E}\left(e^{-s^\top X(T)}|\mathcal{F}(t)\right)$ lead to:

- for affine-diffusions (without jump)

$$\begin{aligned} A'(T-t) &= K_0^\top B^\top(T-t) + \frac{1}{2}B(T-t)^\top H_0 B(T-t), \\ B'(T-t) &= K_1^\top B^\top(T-t) + \frac{1}{2}B(T-t)^\top H_1 B(T-t). \end{aligned}$$

- For affine jump-diffusion model, thanks to the jump component, we will obtain additional terms using stochastic calculus for jumps, a topic beyond this course. The corresponding ODEs can be given as follow

$$\begin{aligned} A'(T-t) &= K_0^T B'(T-t) + \frac{1}{2}B(T-t)^\top H_0 B(T-t) + \lambda_0(\theta(B(T-t)) - 1), \\ B'(T-t) &= K_1^\top B^\top(T-t) + \frac{1}{2}B(T-t)^\top H_1 B(T-t) + \lambda_1(\theta(B(T-t)) - 1), \end{aligned}$$

with $\theta(c) = E\left[\exp\left(c^\top Z\right)\right]$, where Z is the vector of jump sizes.

Classification of affine-diffusion models

Consider an $N-$dimensional affine diffusion process governed by the following SDE

$$dX(t) = \mathcal{K}\left(\Theta - X(t)\right)dt + \Sigma\sqrt{V(t)}dW(t), \qquad (26)$$

for some $\mathcal{K}, \Sigma \in \mathbb{R}^{N\times N}$ and $\Theta \in \mathbb{R}^N$; $\{W(t)\}$ is an $N-$dimensional standard Brownian motion; $V(t)$ is a diagonal matrix with entry $V_{ii}(t) = \alpha_i + \beta_i^\top X(t)$.

- To ensure the square root of $V(t)$ is well-defined, we need affine process $\{X(t)\}$ to be "admissible", i.e., $V_{ii}(t) > 0$ for all i. Duffie and Kan (1996) gives restrictions for admissibility in Condition A.
- In univariate case, this condition is usually called "Feller condition".

Canonical forms

- We call (26) an $A_m(N)$ process if the rank of $\beta_i, 1 \le i \le N$ is m. Note that for $N-$dimensional affine processes, it has $N+1$ subclasses, $A_0(N), \cdots, A_N(N)$. Here, m characterize the number of volatility factors. See, e.g., Dai and Singleton (2000, JoF).
- Another canonical form for $A_m(N)$ processes is developed by Collin-Dufresne. This type of canonical form is based on unspanned stochastic volatility (USV) and Taylor expansions.

Fourier pricing methods: an example from the Heston model

We start some investigations on the Heston model which is governed by the following SDE under the risk-neutral measure Q

$$\frac{dS(t)}{S(t)} = rdt + \sqrt{V(t)}dW_1(t),\ S(0) = s_0, \quad (27)$$

$$dV(t) = \kappa(\theta - V(t))dt + \sigma\sqrt{V(t)}\left[\rho dW_1(t) + \sqrt{1-\rho^2}dW_2(t)\right],\ V(0) = v_0. \quad (28)$$

where $\{W(t)\} = \{(W_1(t), W_2(t))\}$ is a two dimensional standard Brownian motion; $\mu, \kappa, \rho, \sigma$ and θ are all scalar parameters. To show that the Heston model is affine, we take logarithm of $S(t)$. Then, model (27)-(28) becomes

$$dY(t) = (r - \frac{1}{2}V(t))dt + \sqrt{V(t)}dW_1(t),\ S(0) = s_0, \quad (29)$$

$$dV(t) = \kappa(\theta - V(t))dt + \sigma\sqrt{V(t)}\left[\rho dW_1(t) + \sqrt{1-\rho^2}dW_2(t)\right],\ V(0) = v_0, \quad (30)$$

with $Y(t) = \log S(t)$.

Affine transform of the Heston model

We have

$$
\begin{aligned}
& \psi\left(u, t, T, y, v\right) \\
= \; & E^{Q}\left[\exp\left(u_y Y(T) + u_v V(T)\right) | Y(t) = y, V(t) = v\right] \\
= \; & \exp\left(A(T-t) + B_1(T-t)y + B_2\left(T-t\right)v\right),
\end{aligned}
$$

where the functions $A(T-t)$, $B_1(T-t)$ and $B_2\left(T-t\right)$ satisfy

$$
\begin{aligned}
A'(T-t) &= rB_1(T-t) + \kappa\theta B_2\left(T-t\right), && (31)\\
B_1'(T-t) &= 0 && (32)\\
B_2'(T-t) &= -\kappa B_2(T-t) - \frac{1}{2}[B_1(T-t) + B_1^2(T-t) && (33)\\
&\quad +2\sigma\rho B_1(T-t)B_2(T-t) + \sigma^2 B_2^2(T-t)]
\end{aligned}
$$

with the following initial conditions

$$
A(0) = 0,\ B_1(0) = u_y,\ B_2(0) = u_v.
$$

Closed-form solution

It is obvious that the solution to (32) is

$$
B_1(T-t) = u_y. \tag{34}
$$

Plugging (34) into (33), we obtain

$$
B_2'(T-t) = \frac{1}{2}\sigma^2 B_2^2(T-t) + (\sigma\rho u_y - \kappa)B_2(T-t) + \frac{1}{2}u_y^2 - \frac{1}{2}u_y,
$$

which is a Riccati equation. Solving the equation, we have

$$
B_2(T-t) = \frac{(\kappa - \sigma\rho u_y + d) - (\kappa - \sigma\rho u_y - d)ge^{d(T-t)}}{\sigma^2(1 - ge^{d(T-t)})}, \tag{35}
$$

where

$$
d = \sqrt{(\sigma\rho w_R - \kappa)^2 - \sigma^2(w_R^2 - w_R)},
$$

Closed-form solution

and

$$g = \frac{(\kappa - \sigma\rho u_y) + \sqrt{(\sigma\rho u_y - \kappa)^2 - \sigma^2(u_y^2 - u_y) - u_v\sigma^2}}{(\kappa - \sigma\rho u_y) - \sqrt{(\sigma\rho u_y - \kappa)^2 - \sigma^2(u_y^2 - u_y) - u_v\sigma^2}}.$$

Finally, plugging (34) and (35) into (31), we obtain

$$A(T-t) = r u_y (T-t) + \frac{\kappa\theta}{\sigma^2}\left[(\kappa - \sigma\rho u_y + d)(T-t) - 2\log\left(\frac{1 - g e^{d(T-t)}}{1-g}\right)\right]. \tag{36}$$

Fourier pricing method

Now, our aim is to recover option price from the above Laplace transform. It is well-known that, under risk neutral measure, the discounted option price is a martingale. Thus, the time-t price $C(Y(t), V(t), T-t, K)$ of a call option with maturity T and strike K with payoff $(S(t) - K)^+$ can be represented as follows:

$$C(Y(t), V(t), T-t, K) = e^{-r(T-t)} E^Q\left[(S(T) - K)^+ \middle| \mathcal{F}(t)\right],$$

where $\{\mathcal{F}(t)\}$ denote the filtration. Now, introduce a vector $u := (u_y, u_v)^\top = (1, 0)^\top$. Then, we have

$$\begin{aligned} & C(Y(t), V(t), T-t, K) \qquad (37) \\ = \;& e^{-r(T-t)} E^Q\left[\left(e^{u_y Y(T) + u_v V(T)} - K\right) 1\{u_y Y(T) + u_v V(T) \ge \log K\} \middle| \mathcal{F}(t)\right] \\ = \;& e^{-r(T-t)} E^Q\left[e^{u_y Y(T) + u_v V(T)} 1\{u_y Y(T) + u_v V(T) \ge \log K\} \middle| \mathcal{F}(t)\right] \\ & - K e^{-r(T-t)} E^Q\left[1\{u_y Y(T) + u_v V(T) \ge \log K\} \middle| \mathcal{F}(t)\right]. \end{aligned}$$

Fourier pricing method

Let $G_{a,b}(x; Y(t), V(t), T-t)$ denote the price of a security that pays $e^{a_1 Y(t)+a_2 V(t)}$ at time T in the event that $b_1 Y(t) + b_2 V(t) \leq x$. Mathematically, define $G_{a,b}(x; Y(t), V(t), T-t)$ as follow

$$\begin{aligned} & G_{a,b}(x; Y(t), V(t), T-t) : \\ = \quad & e^{-r(T-t)} E^Q \left[e^{a_1 Y(T)+a_2 V(T)} 1\{b_1 Y(T) + b_2 V(T) \leq x\} \middle| \mathcal{F}(t) \right]. \end{aligned}$$

Then, option price (37) admits the following representation:

$$\begin{aligned} & C(Y(t), V(t), T-t, K) \\ = \quad & G_{u,-u}(-\log K; T-t, Y(t), V(t)) - K G_{0,-u}(-\log K; T-t, Y(t), V(t)). \end{aligned}$$

Fourier pricing method: closed-form transform

The Fourier-Stieltjes transform $\mathcal{G}_{a,b}(\cdot; Y(t), V(t), T-t)$ of $G_{a,b}(\cdot; Y(t), V(t), T-t)$, if well-defined, is given by

$$\mathcal{G}_{a,b}(v; Y(t), V(t), T-t) = \int_{\mathbb{R}} e^{ivx} dG_{a,b}(x; Y(t), V(t), T-t).$$

We have

$$\begin{aligned} & \mathcal{G}_{a,b}(v; Y(t), V(t), T-t) \\ = \quad & E^Q \left[e^{(a_1+ivb_1)Y(T)+(a_2+ivb_2)V(T)} \middle| \mathcal{F}(t) \right] = \psi(a+ivb, Y(t), V(t), t, T), \end{aligned}$$

which is the Laplace transform for $X(t) = (Y(t), V(t))^\top$.

Fourier pricing method: transform inversion

Therefore, applying inverse Fourier transform, we recover $G_{a,b}(x; Y(t), V(t), T-t)$ and then price vanilla options analytically. Gil-Pelaez (1951) proposes a useful formula to obtain cumulative distribution function from characteristic function. A direct application leads to

$$G_{a,b}(x; Y(t), V(t), T-t) = \frac{\psi(a, Y(t), V(t), t, T)}{2}$$
$$+ \frac{1}{2\pi} \int_0^{+\infty} \frac{e^{ivx}\psi(a - ivb, Y(t), V(t), t, T) - e^{-ivx}\psi(a + ivb, Y(t), V(t), t, T)}{iv} dv.$$

Note that $e^{ivx}\psi(a - ivb, \cdot)$ is the complex conjugate of $e^{-ivx}\psi(a + ivb, \cdot)$. Thus, the integral on the RHS simplifies to

$$\begin{aligned} & \int_0^{+\infty} \frac{e^{ivx}\psi(a - ivb, Y(t), V(t), t, T) - e^{-ivx}\psi(a + ivb, Y(t), V(t), t, T)}{iv} dv. \\ = & \; -2 \int_0^{+\infty} \frac{Im\left[e^{-ivx}\psi(a + ivb, Y(t), V(t), t, T)\right]}{iv}. \end{aligned}$$

Overview

- Applications in optimal portfolio choice

A very brief literature review

- The pioneering work can be traced back to Markowitz (1952) in a mean-variance setting, assuming quadratic utility function.
- Then, Samuelson and Merton's work in 1969 and 1971 used continuous-time dynamic programming techniques, applying HJB equations to solve the problem.
- Then, it comes Karatzas, Lehoczky, and Shreve (1987), Cox and Huang $(1989, 1991)$ which apply martingale tools to find explicit solutions.
- The most recent milestones include Ocone and Karatzas (1991) and Detemple Garcia, and Rindisbacher (2003).

The model

- The financial market consists of a risk-free money market account B_t appreciating a spot interest rate r_t. The dynamics of the price B_t is characterized by

$$dB_t = r_t B_t dt.$$

- There are also d stocks $S_t \in \mathbb{R}_+^d$, which pay dividends. The evolution of the stock vector process is

$$dS_t + D_t dt = I_t^S \left(\mu_t dt + \sigma_t dW_t \right),$$

where D_t is $d-$dimensional vector representing the dividend rate; μ_t is a $d-$dimensional vector representing the total expected return; σ_t is a $d \times d$ matrix representing the volatility; $I_t^S = diag\left(S_t\right).$

Utility

- Let c_t denote the consumption process, which is progressively measurable, integrable and satifying $c_t \geq 0$. Denote by C the collection of such consumption processes c_t. We consider the following expected total utility of the consumptions

$$U(c) = E\left[\int_0^T u(c_t, t)\, dt\right].$$

- Here $u(\cdot, t)$ is utility function, which is continuous, twice continuously differentiable and concave. We assume the following *Inada* conditions:

$$u'(0, t) = \infty \text{ and } u'(\infty, t) = 0.$$

The portfolio planning

- Let π_t be the dollar amount invested at each risky asset. Denote by P^π the collection of π_t satisfying this condition.
- Denote by X_t the wealth process. Thus, it has dynamics

$$dX_t = -c_t dt + \pi_t' \left(I_t^S\right)^{-1} (dS_t + D_t dt) + \left(X_t - \pi_t' 1_{d\times 1}\right) r_t dt,\ X_0 = x. \tag{38}$$

- We can understand the second term as the numbers of shares in the risk assets (i.e., $\pi_t'\left(I_t^S\right)^{-1}$) multiplied by the dollar earned (i.e., $dS_t + D_t dt$) by holding them in an infinitesimal time period. Collecting terms, we obtain that

$$\begin{aligned} dX_t &= (X_t r_t - c_t)\, dt + \pi_t' \left[(\mu_t - r_t 1_{d\times 1})\, dt + \sigma_t dW_t\right] \\ &= (X_t r_t - c_t)\, dt + \pi_t' \sigma_t \left[\theta_t dt + dW_t\right], \\ X_0 &= x. \end{aligned}$$

The optimal portfolio choice

- We define $(c, \pi) \in C \otimes P^{\pi}$ to be admissible by $(c, \pi) \in A(x)$ iff $X_t \geq 0$, $\forall t \in [0, T]$. This is the no-bankruptcy condition. We say that an admissible pair $(c, \pi) \in A(x)$ is optimal $(c, \pi) \in A^*(x)$ iff no other $(c', \pi') \in A(x)$ satisfies that $U(c') > U(c)$.

- We seek for the solution to the following optimization problem:

$$\max_{(\pi_t, c_t)} E\left[\int_0^T u(c_t, t)dt\right].$$

A static consumption-portfolio problem

- First, we define:

$$\eta_t = \exp\left(-\int_0^t \theta_v' \cdot dW_v - \frac{1}{2}\int_0^t \theta_v' \theta_v' dv\right) < \infty, \forall t \in [0, T]\,,\; P\!-\!a.s.$$

  Suppose Novikov condition is satisfied. Thus η_t is a martingale and $E[\eta_t] = 1, \forall t \in [0, T]$. Girsanov theorem applies here. On $\mathcal{F}_T$, we definte Q by

$$\left.\frac{dQ}{dP}\right|_{\mathcal{F}_T} = \eta_T$$

- Define $\xi_t = b_t \eta_t$, where $b_t = \exp\left(-\int_0^t r_v dv\right)$. Here, ξ_t is called the state price density (SPD hereafter). We usually call Q the risk-neutral probability measure.

A static consumption-portfolio problem

- (Static problem) Suppose $c \in C$. Define

$$B(x) = \left\{ c \in C | E\left(\int_0^T \xi_t c_t dt \right) \leq x \right\}.$$

$B(x)$ is called a budget set. $c \in B(x)$ is called optimal for static problem $c \in B^*(x)$ iff there does not exist $c' \in B(x)$ such that $U(c') > U(c)$.

Theorem

$(c, \pi) \in A^*(x) \Rightarrow c \in B^*(x)$ and if $c \in B^*(x)$ then there exists an admissible π such that $(c, \pi) \in A^*(x)$.

Solving the static problem

The static problem is

$$\max_c E\left[\int_0^T u(c_t, t)\, dt \right]$$

subject to

$$E\left(\int_0^T \xi_v c_v dv \right) \leq x.$$

The principle of Lagrange multiplier in constraint optimization suggest that $c \in B^*(x)$ iff $(c, y) \in \arg\max L(c, y)$, where

$$L(c, y) = E\left[\int_0^T u(c_t, t)\, dt \right] + y \left[x - E\left(\int_0^T \xi_t c_t dt \right) \right]. \quad (39)$$

A heuristic application of the first-order conditions w.r.t. c and y yields that

$$u_c'(c_t, t) = y \xi_t,\ y > 0,\ \text{and } E\left(\int_0^T \xi_v c_v dv \right) = x. \quad (40)$$

Solving the static problem

Thus, we state the following theorem:

Theorem

Suppose that (c, y) satisfies (40), then c is optimal for the static problem. And if c is optimal for the static problem, there exists a $y > 0$ such that (c, y) satisfies (40).

Lemma

There exists a unique $y > 0$, such that $x = \chi(y)$.

Road map to the solution

Putting everything together, indeed, we have the following route or solving the optimal consumption problem

$$\left.\begin{array}{c} y^* = \chi^{-1}(x) \Longrightarrow c_t^* = I(y^*\xi_t, t) \\ \Longrightarrow X_t^* = E_t\left(\int_t^T \xi_{t,v} c_v^* dv\right) \\ \text{Martingale representation} \Longrightarrow \phi_t \end{array}\right\} \Longrightarrow \pi_t^* = X_t^*\left(\sigma_t'\right)^{-1}\theta_t + \xi_t^{-1}\left(\sigma_t'\right)^{-1}\phi_t$$

The solution

Theorem

We have the following solution to the static problem:

$$c_t^* = I\left(y^*\xi_t, t\right)$$

where $y^* = \chi^{-1}(x)$ *. We have*

$$\pi_t^* = X_t^* \left(\sigma_t'\right)^{-1} \theta_t + \xi_t^{-1} \left(\sigma_t'\right)^{-1} \phi_t,$$

where

$$E_t\left[\int_0^T \xi_t I\left(y^*\xi_t, t\right) dt\right] - E\left[\int_0^T \xi_t I\left(y^*\xi_t, t\right) dt\right] = \int_0^t \phi_s \cdot dW_s.$$

Thus, we have

$$X_t^* = E_t\left[\int_t^T \xi_{t,v} c_v^* dv\right] = E_t\left[\int_t^T \xi_{t,v} I\left(y^*\xi_v, v\right) dv\right].$$

Explicit representation for optimal portfolio choice

Theorem

We have the structure of optimal portfolio as

$$\pi_t^* = \left(\sigma_t'\right)^{-1} \theta_t E_t\left[\int_t^T \xi_{t,v} \frac{c_v^*}{R_v^*} dv\right] - \left(\sigma_t'\right)^{-1} E_t\left[\int_t^T \xi_{t,v} c_v^* \left(1 - \frac{1}{R_v^*}\right) H_{t,v} dv\right],$$

where $\xi_{t,v} = \frac{\xi_v}{\xi_t}$, $c_v^* = I\left(y^*\xi_v, v\right)$ *, and* $R_v^* = R\left(c_v^*\right)$ *with* $R(c) := -\frac{u_c''(c,t)}{u_c'(c,t)} c$*. Here,* y^* *solves the equation*

$$x = E\left[\int_0^T \xi_v I\left(y\xi_v, v\right) dv\right].$$

Moreover,

$$H_{t,v}' = \int_t^v D_t r_s ds + \int_t^v \left(dW_s' + \theta_s' ds\right) D_t \theta_s.$$

Here, $H_{t,v}$ *is a* $d \times 1$ *vector; the Malliavin derivative* $D_t r_s$ *is a* $1 \times d$ *vector; the Malliavin derivative* $D_t \theta_s$ *is a* $d \times d$ *matrix.*

More cases

1. The bequest case:
$$\max_{(\pi_t, X_T)} E\left[U(X_T)\right],$$
subject to
$$dX_t = r_t X_t dt + X_t \pi_t^\top \left[(\mu_t - r_t \mathbf{1}_m)dt + \sigma_t dW_t\right], \; X_0 = x,$$
$$X_t \geq 0, \text{ for all } t \in [0, T].$$

2. The consumption-bequest case:
$$\max_{(\pi_t, c_t, X_T)} E\left[\int_0^T u(c_t, t)dt + U(X_T, T)\right],$$
subject to
$$dX_t = r_t X_t dt + X_t \pi_t^\top \left[(\mu_t - r_t \mathbf{1}_m)dt + \sigma_t dW_t\right] - c_t dt, \; X_0 = x,$$
$$X_t \geq 0, \text{ for all } t \in [0, T].$$

Excellent exercises

Following the frameworks, to calculate closed-form optimal portfolio choice under some simple models, e.g., to solve the Merton problem under the geometric Brownian motion model. A closed-form solution was proposed by Merton. So, one should be able to obtain the same formula by calculating the expectations involved in our framework. Once you are able to complete this exercise, you fully master a lot of contents.

Motivation:

- Martingale representation theorem
  - Wiener spaces: martingales with finite variance are weighted sums of BM increments
  - $$M_t = M_0 + \int_0^t \phi_s dW_s$$
    for some progressively measurable process ϕ
  - ϕ represents volatility coefficient of martingale
- Malliavin calculus identifies integrand ϕ (Clark-Ocone formula)
- Thus, the optimal portfolio choice is solved!
- See supplementary notes.

Supplementary Notes on "Applications of Stochastic Analysis"

Chenxu Li *
Department of Business Statistics
and Econometrics
Guanghua School of Management
Peking University

Tuesday 7$^{\text{th}}$ February, 2017

Abstract

In this set of notes, we provide some technical details supplementary to our lecture on "Applications of Stochastic Analysis." We mainly discuss about two streams of applications: derivatives pricing and optimal portfolio choice. As this is a very preliminary version, please keep them for your own use and don't circulate them. We will try to write them into a text book. I appreciate any reports of errors and typos. With this set of notes, you don't need worry about the problem of viewing clustered formula derivations the board from a distant point in our big classroom. However, please note that it is not an exhaustive list of our in-class discussion. A combination of course slides, these supplementary notes, homework problem sets and your notes taken from our in-class discussions would be quite helpful to your study.

*Email: cxli@gsm.pku.edu.cn

Contents

*I thank Chenxu (my PhD student) for his kind and careful preparation of the notes of this section.

Part I

Applications in Derivatives Pricing

This part is devoted to applications in derivatives pricing.

1 On Black-Scholes-Merton (1973)

1.1 Discounted stock price under risk neutral probability

We will prove that, under risk neutral probability $\mathbb{Q}$, the discounted stock price $e^{-rt}S(t)$ is a martingale. Applying Itô product formula, we obtain that

$$\begin{aligned}
de^{-rt}S(t) &= e^{-rt}dS(t) + S(t)de^{-rt} + de^{-rt}dS(t) \\
&= e^{-rt}\left[rS(t)dt + \sigma S(t)dW^{\mathbb{Q}}(t)\right] - re^{-rt}S(t)dt \\
&= e^{-rt}\sigma S(t)dW^{\mathbb{Q}}(t)
\end{aligned}$$

Thus, we have a $\mathbb{Q}-$martingale

$$e^{-rt}S(t) = S(0) + \int_0^t e^{-rt}\sigma S(t)dW^{\mathbb{Q}}(t).$$

1.2 A rigorous application of Feynman-Kac theorem to Black-Scholes-Merton PDE

Because

$$(S(T) - K)^+ \leq S(T),$$

the no-arbitrage price of the call option must be dominated by the value of the underlying asset, i.e.

$$u(t, S(t)) \leq S(t).$$

Therefore, the price function u satisfies the polynomial growth condition:

$$\max_{0\leq t\leq T} |u(t,s)| \leq s \leq M(1+|s|^{2\mu}).$$

By Theorem 5.7.6 in Karatzas and Shreve (1988) [15], we have the Black-Scholes solution, i.e.

$$u(t,S(t)) = \mathbb{E}^Q\left[e^{-r(T-t)}(S(T)-K)^+|\mathcal{F}(t)\right].$$

As an exercise, prove the case for put options.

1.3 Derivation of the Black-Scholes-Merton Formula

Under the risk-neutral measure $\mathbb{Q}$, the Black-Scholes-Merton Model follows

$$dS(t) = rS(t)dt + \sigma S(t)dW^{\mathbb{Q}}(t).$$

For lightening notations, we simply write $W^{\mathbb{Q}}(t)$ as $W(t)$ in what follows. The time t price of a call option with strike K and maturity T admits the following representation

$$c(t,S(t)) = \mathbb{E}^{\mathbb{Q}}\left[e^{-r(T-t)}(S(T)-K)^+|\mathcal{F}(t)\right] = \mathbb{E}^{\mathbb{Q}}\left[e^{-r(T-t)}(S(T)-K)^+|S(t)\right]$$

where the second equality follows the Markov property of $\{S(t)\}$. In other words, we need to calculate

$$c(t,s) = \mathbb{E}^{\mathbb{Q}}\left[e^{-r(T-t)}(S(T)-K)^+|S(t)=s\right].$$

We note that

$$(S(T)-K)^+ = (S(T)-K)1_{\{S(T)\geq K\}} = S(T)1_{\{S(T)\geq K\}} - K1_{\{S(T)\geq K\}}.$$

We will calculate $\mathbb{E}^{\mathbb{Q}}\left[S(T)1_{\{S(T)\geq K\}}|\mathcal{F}(t)\right]$ and $\mathbb{E}^{\mathbb{Q}}\left[K1_{\{S(T)\geq K\}}|\mathcal{F}(t)\right]$.

Because

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

we have

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

where Y is the standard normal random variable

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

and $\tau = T - t$.

For $\mathbb{E}^{\mathbb{Q}}\left[K1_{\{S(T)\geq K\}}|S(t) = s\right]$, we have

$$\begin{aligned} & \mathbb{E}^{\mathbb{Q}}\left[K1_{\{S(T)\geq K\}}|S(t) = s\right] \\ =\ & K\mathbb{E}^{\mathbb{Q}}\left[1_{\{S(T)\geq K\}}|S(t) = s\right] \\ =\ & K\mathbb{Q}\left(S(T) \geq K|S(t) = s\right) \\ =\ & K\mathbb{Q}\left(s\exp\left(\sigma\sqrt{\tau}Y + \left(r - \frac{1}{2}\sigma^2\right)\tau\right) \geq K|S(t) = s\right) \\ =\ & K\mathbb{Q}\left(s\exp\left(\sigma\sqrt{\tau}Y + \left(r - \frac{1}{2}\sigma^2\right)\tau\right) \geq K\right), \end{aligned}$$

where we eliminate the conditioning in the last equality because of the independence between Y and $S(t)$. Because

$$s\exp\left(\sigma\sqrt{\tau}Y + \left(r - \frac{1}{2}\sigma^2\right)\tau\right) \geq K \Longleftrightarrow Y \geq \frac{1}{\sigma\sqrt{\tau}}\left[\log\left(\frac{K}{s}\right) - \left(r - \frac{1}{2}\sigma^2\right)\tau\right],$$

we have

$$
\begin{aligned}
& \mathbb{Q}\left(s \exp\left(\sigma\sqrt{\tau}Y + \left(r - \frac{1}{2}\sigma^2\right)\tau\right) \geq K\right) \\
= \; & \mathbb{Q}\left(Y \geq \frac{1}{\sigma\sqrt{\tau}}\left[\log\left(\frac{K}{s}\right) - \left(r - \frac{1}{2}\sigma^2\right)\tau\right]\right) \\
= \; & 1 - N\left(\frac{1}{\sigma\sqrt{\tau}}\left[\log\left(\frac{K}{s}\right) - \left(r - \frac{1}{2}\sigma^2\right)\tau\right]\right) \\
= \; & N\left(-\frac{1}{\sigma\sqrt{\tau}}\left[\log\left(\frac{K}{s}\right) - \left(r - \frac{1}{2}\sigma^2\right)\tau\right]\right) \\
= \; & N\left(\frac{1}{\sigma\sqrt{\tau}}\left[\log\left(\frac{s}{K}\right) + \left(r - \frac{1}{2}\sigma^2\right)\tau\right]\right).
\end{aligned}
$$

Thus, it is easy to find

$$
\mathbb{E}^{\mathbb{Q}}\left[K1_{\{S(T)\geq K\}}|S(t) = s\right] = KN\left(d_-\left(\tau, s\right)\right),
$$

where

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

For $\mathbb{E}^{\mathbb{Q}}\left[S(T)1_{\{S(T)\geq K\}}|S(t)=s\right]$, we note that

$$
\begin{aligned}
& \mathbb{E}^{\mathbb{Q}}\left[S(T)1_{\{S(T)\geq K\}}|S(t)=s\right] \\
= & \mathbb{E}^{\mathbb{Q}}\left[s\exp\left(\sigma\sqrt{\tau}Y+\left(r-\frac{1}{2}\sigma^2\right)\tau\right)1_{\{s\exp\left(\sigma\sqrt{\tau}Y+\left(r-\frac{1}{2}\sigma^2\right)\tau\right)\geq K\}}|S(t)=s\right] \\
= & \int_{\frac{1}{\sigma\sqrt{\tau}}\left[\log\left(\frac{K}{s}\right)-\left(r-\frac{1}{2}\sigma^2\right)\tau\right]}^{\infty} s\exp\left(\sigma\sqrt{\tau}y+\left(r-\frac{1}{2}\sigma^2\right)\tau\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}dy \\
= & s\exp(r\tau)\int_{\frac{1}{\sigma\sqrt{\tau}}\left[\log\left(\frac{K}{s}\right)-\left(r-\frac{1}{2}\sigma^2\right)\tau\right]}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y^2+\sigma\sqrt{\tau}y-\frac{1}{2}\sigma^2\tau\right)dy \\
= & s\exp(r\tau)\int_{\frac{1}{\sigma\sqrt{\tau}}\left[\log\left(\frac{K}{s}\right)-\left(r-\frac{1}{2}\sigma^2\right)\tau\right]}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(y-\sigma\sqrt{\tau}\right)^2\right)dy \\
= & s\exp(r\tau)\int_{\frac{1}{\sigma\sqrt{\tau}}\left[\log\left(\frac{K}{s}\right)-\left(r-\frac{1}{2}\sigma^2\right)\tau\right]-\sigma\sqrt{\tau}}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right)dz \\
= & s\exp(r\tau)\int_{\frac{1}{\sigma\sqrt{\tau}}\left[\log\left(\frac{K}{s}\right)-\left(r+\frac{1}{2}\sigma^2\right)\tau\right]}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right)dz \\
= & s\exp(r\tau)\left[1-N\left(\frac{1}{\sigma\sqrt{\tau}}\left[\log\left(\frac{K}{s}\right)-\left(r+\frac{1}{2}\sigma^2\right)\tau\right]\right)\right] \\
= & s\exp(r\tau)N\left(-\frac{1}{\sigma\sqrt{\tau}}\left[\log\left(\frac{K}{s}\right)-\left(r+\frac{1}{2}\sigma^2\right)\tau\right]\right) \\
= & s\exp(r\tau)N\left(d_+(\tau,s)\right),
\end{aligned}
$$

where

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

Therefore, we finally have

$$
\begin{aligned}
c(t,s) &= \mathbb{E}^{\mathbb{Q}}\left[e^{-r(T-t)}(S(T)-K)^+|S(t)=s\right] \\
&= e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\left[e^{-r(T-t)}S(T)1_{\{S(T)\geq K\}}|S(t)=s\right]-e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\left[K1_{\{S(T)\geq K\}}|S(t)=s\right] \\
&= sN\left(d_+\left(\tau,s\right)\right)-e^{-r\tau}KN\left(d_-\left(\tau,s\right)\right).
\end{aligned}
$$

1.4 Market price of risk: a view from Black-Scholes

It is interesting for us to understand the market price of risk. First, assume the real world dynamics of the underlying asset is

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma dW^P(t).$$

We know that

$$\lambda = \frac{\mu - r}{\sigma}$$

is usually called the market price of risk. For a call option with maturity T and strike K, its price $v(t, S(t))$ satisfies

$$\begin{aligned} dv(t, S(t)) &= \frac{\partial v}{\partial t} dt + \frac{\partial v}{\partial x} dS(t) + \frac{1}{2}\frac{\partial^2 v}{\partial x^2} d[S, S](t) \\ &= \frac{\partial v}{\partial t} dt + \frac{\partial v}{\partial x}[\mu S(t)dt + \sigma S(t) dW^P(t)] + \frac{1}{2}\frac{\partial^2 v}{\partial x^2}\sigma^2 S^2(t) dt. \end{aligned}$$

By the Black-Scholes equation, we know that the above equation is carried out to

$$dv(t, S(t)) = \left(rv(t, S(t)) - rS(t)\frac{\partial v}{\partial x} \right) dt + \frac{\partial v}{\partial x}[\mu S(t)dt + \sigma S(t) dW^P(t)].$$

Thus, we have

$$dv(t, S(t)) - rv(t, S(t))dt = \frac{\partial v}{\partial x}\sigma S(t) \left(\frac{\mu - r}{\sigma} dt + dW^P(t) \right), \tag{1.1}$$

which is equivalent to

$$\frac{dv(t, S(t))}{v(t, S(t))} - rdt = \frac{\frac{\partial v}{\partial x}\sigma S(t)}{v(t, S(t))} \left(\frac{\mu - r}{\sigma} dt + dW^P(t) \right). \tag{1.2}$$

This equation has very clear economic meaning. The term $\frac{dv(t,S(t))}{v(t,S(t))} - rdt$ can be understood as an excess return. Integrating both sides of (1.2), we obtain that

$$\int_t^{t+\Delta} \frac{dv(u, S(u))}{v(u, S(u))} - \int_t^{t+\Delta} rdu = \int_t^{t+\Delta} \frac{\frac{\partial v}{\partial x}\sigma S(u)}{v(u, S(u))} \left(\frac{\mu - r}{\sigma} du + dW^P(u) \right).$$

Taking conditional expectation E_t, we obtain that

$$E_t \int_t^{t+\Delta} \frac{dv(u,S(u))}{v(u,S(u))} - \int_t^{t+\Delta} r du = \int_t^{t+\Delta} E_t \frac{\frac{\partial v}{\partial x}\sigma S(u)}{v(u,S(u))} \frac{\mu - r}{\sigma} du.$$

In some reference book, people like to write

$$E_t \frac{dv(t,S(t))}{v(t,S(t))} - r dt = \frac{\frac{\partial v}{\partial x}\sigma S(t)}{v(t,S(t))} \frac{\mu - r}{\sigma} dt \equiv \frac{\frac{\partial v}{\partial x} S(t)}{v(t,S(t))} (\mu - r) dt.$$

An equivalent way of writing this relation is

$$\frac{1}{v(t,S(t))} \frac{E_t dv(t,S(t))}{dt} - r = \frac{\frac{\partial v}{\partial x} S(t)}{v(t,S(t))} (\mu - r). \tag{1.3}$$

I don't like this expression, as $E_t \frac{dv(t,S(t))}{v(t,S(t))}$ doesn't make mathematical sense. Maybe it is a physical way for understanding these things. Instead, I would integrate (1.1) to obtain that

$$v(t+\Delta, S(t+\Delta)) - v(t,S(t)) - \int_t^{t+\Delta} r v(u,S(u)) du = \int_t^{t+\Delta} \frac{\partial v}{\partial x} \sigma S(u) \left(\frac{\mu - r}{\sigma} dt + dW^P(u) \right).$$

Taking E_t, we obtain that

$$E_t v(t+\Delta, S(t+\Delta)) - v(t,S(t)) - \int_t^{t+\Delta} r E_t v(u,S(u)) du = \int_t^{t+\Delta} E_t \left(\frac{\partial v}{\partial x} S(u) \right) (\mu - r) du.$$

Now, we have

$$\frac{1}{\Delta} \left[E_t v(t+\Delta, S(t+\Delta)) - v(t,S(t)) \right] - \frac{1}{\Delta} \int_t^{t+\Delta} r E_t v(u,S(u)) du = \frac{1}{\Delta} \int_t^{t+\Delta} E_t \left(\frac{\partial v}{\partial x} S(u) \right) (\mu - r) du.$$

Let $\Delta \to 0$, we have

$$\lim_{\Delta \to 0} \frac{1}{\Delta} \left[E_t v(t+\Delta, S(t+\Delta)) - v(t,S(t)) \right] - r v(t,S(t)) = \frac{\partial v}{\partial x} S(t) (\mu - r).$$

Now, we have

$$\lim_{\Delta\to 0}\frac{1}{\Delta}\frac{E_t v(t+\Delta, S(t+\Delta)) - v(t, S(t))}{v(t, S(t))} - r = \frac{\frac{\partial v}{\partial x}S(t)}{v(t, S(t))}(\mu - r), \tag{1.4}$$

which is equivalent to (1.3). Indeed, this gives a rigorous formulation of (1.3). Now, if we view

$$\frac{1}{\Delta}\frac{E_t v(t+\Delta, S(t+\Delta)) - v(t, S(t))}{v(t, S(t))}$$

as an expected return per time (in an infinitesimal sense), (1.4) can be interpreted as a "CAPM" type result. Here, $\frac{\frac{\partial v}{\partial x}S(t)}{v(t,S(t))}$ plays a role as the "beta". But, I am now not toally sure how to fully interpret everything. Can you? To what maximal extent? The LHS is no doubt an instantenous excess return. I can somehow regard $\frac{\frac{\partial v}{\partial x}\sigma S(t)}{v(t,S(t))}$ as a percentage of "Brownian risk corresponding to $W^P(t)$" and $\frac{\mu-r}{\sigma}$ is the excess premium per unit of $dW^P(t)$. In particular, we know that, *because call (put) options are positively (negatively) exposed to stock prices and stock expected excess returns are generally positive, the expected excess return of calls is positive and the expected excess return of puts is negative.*

Now, we mention another characterization of the market price of risk. This is a mathematical characterization. We have

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma dW^P(t) = r dt + \sigma\left(\frac{\mu - r}{\sigma}dt + dW^P(t)\right).$$

On the other hand, no-arbitrage argument (or, say replication argument) yields the risk-neutral probability measure, under which the dynamics of the underlying asset is written as

$$\frac{dS(t)}{S(t)} = r dt + \sigma dW^Q(t).$$

By the Girsanov theorem, $W^Q(t)$ can be constructed (in the sense of distributional equivalence) through

$$W^Q(t) = \lambda t + W^P(t) = \frac{\mu - r}{\sigma}t + W^P(t).$$

So, the market price of risk is exactly the "drift" in the Girsanov change of measure.

2 Beyond Black-Scholes: stochastic volatility

2.1 Derivatives pricing under stochastic volatility models

By analogy to the Black-Scholes case, we work on the stochastic volatility case. In this case, the market with the underlying asset and a money market account is incomplete in the sense that a derivative security may not be replicated using these two instruments only. So, we need to derive everything carefully. I would suggest to correct a number of things from the book "the volatility surface".

First, we assume the model under the physical probability measure as

$$\begin{aligned}
\frac{dS(t)}{S(t)} &= \mu dt + \sqrt{V(t)} dW_1^P(t), \\
dV(t) &= a(V(t))dt + b(V(t)) \left[\rho dW_1^P(t) + \sqrt{1-\rho^2} dW_2^P(t) \right],
\end{aligned}$$

where the $(W_1^P(t), W_2^P(t))$ is a two-dimensional standard Brownian motion. Instead of constructing a risk-free portfolio, we perform replication as it is more rigorous. We use $\Delta(t)$ shares of the underlying asset with price $S(t)$, $\Delta_1(t)$ shares of an arbitrary asset with value $V_1(t)$ solely depending on volatility (Note that such an asset can be, e.g., a variance swap or a delta-hedged portfolio.) as well as a money market account to replicate an option (with maturity T and strike K) with value $V_2(t)$, Self-financing condition yields that the change of the replicating portfolio value satisfies

$$d\Pi(t) = \Delta(t)dS(t) + \Delta_1(t)dV_1(t) + r(\Pi(t) - \Delta(t)S(t) - \Delta_1(t)V_1(t))dt. \tag{2.1}$$

Now, we can assume that $\Pi(t) = v(t, S(t), V(t))$ for some smooth function $v(t,x,y)$ and $V_1(t) = v_1(t, V(t))$ for some smooth function $v(t, y)$.

Using the Ito formula, we that

$$
\begin{aligned}
& dv(t,S(t),V(t)) \\
= \; & \frac{\partial v}{\partial t}dt + \frac{\partial v}{\partial x}dS\left(t\right) + \frac{\partial v}{\partial y}dV(t) + \frac{1}{2}\frac{\partial^2 v}{\partial x^2}d\left[S,S\right](t) + \frac{1}{2}\frac{\partial^2 v}{\partial y^2}d\left[V,V\right](t) + \frac{\partial^2 v}{\partial x \partial y}d\left[S,V\right](t) \\
= \; & \frac{\partial v}{\partial t}dt + \frac{\partial v}{\partial x}\left(\mu S\left(t\right)dt + \sqrt{V(t)}S\left(t\right)dW_1^P\left(t\right)\right) \\
& + \frac{\partial v}{\partial y}\left(a\left(V(t)\right)dt + b\left(V(t)\right)\rho dW_1^P\left(t\right) + b\left(V(t)\right)\sqrt{1-\rho^2}dW_2^P\left(t\right)\right) \\
& + \frac{1}{2}\frac{\partial^2 v}{\partial x^2}V(t)S^2\left(t\right)dt + \frac{1}{2}\frac{\partial^2 v}{\partial y^2}\left(b^2\left(V(t)\right)\rho^2 + b^2\left(V(t)\right)\left(1-\rho^2\right)\right)dt \\
& + \frac{\partial^2 v}{\partial x \partial y}\sqrt{V(t)}S(t)b\left(V(t)\right)\rho dt \\
= \; & \left(\begin{array}{c} \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x}\mu S\left(t\right) + \frac{\partial v}{\partial y}a\left(V(t)\right) \\ +\frac{1}{2}\frac{\partial^2 v}{\partial x^2}V(t)S^2\left(t\right) + \frac{1}{2}\frac{\partial^2 v}{\partial y^2}b^2\left(V(t)\right) + \frac{\partial^2 v}{\partial x \partial y}S(t)\sqrt{V(t)}b\left(V(t)\right)\rho \end{array}\right)dt \\
& + \left(\frac{\partial v}{\partial x}\sqrt{V(t)}S\left(t\right) + \frac{\partial v}{\partial y}b\left(V(t)\right)\rho\right)dW_1^P\left(t\right) + \frac{\partial v}{\partial y}b\left(V(t)\right)\sqrt{1-\rho^2}dW_2^P\left(t\right).
\end{aligned}
$$

And, on the other hand, from (2.1), we obtain that

$$
\begin{aligned}
d\Pi(t) \; = \; & \Delta(t)[\mu S(t)dt + \sqrt{V(t)}S(t)dW_1^P(t)] \\
& + \Delta_1(t)dv_1(t,V(t)) + r(v(t,S(t),V(t)) - \Delta(t)S(t) - \Delta_1(t)v_1(t,V(t)))dt \\
= \; & \Delta(t)\mu S\left(t\right)dt + \Delta(t)\sqrt{V(t)}S(t)dW_1^P(t) + \Delta_1(t)\left(\frac{\partial v_1}{\partial t}dt + \frac{\partial v_1}{\partial y}dV(t) + \frac{1}{2}\frac{\partial^2 v_1}{\partial y^2}d\left[V,V\right](t)\right) \\
& + r(v(t,S(t),V(t)) - \Delta(t)S(t) - \Delta_1(t)v_1(t,V(t)))dt \\
= \; & \Delta(t)\mu S\left(t\right)dt + \Delta(t)\sqrt{V(t)}S(t)dW_1^P(t) \\
& + \Delta_1(t)\left(\begin{array}{c} \frac{\partial v_1}{\partial t}dt + \frac{\partial v_1}{\partial y}\left(a\left(V(t)\right)dt + b\left(V(t)\right)\rho dW_1^P\left(t\right) + b\left(V(t)\right)\sqrt{1-\rho^2}dW_2^P\left(t\right)\right) \\ +\frac{1}{2}\frac{\partial^2 v_1}{\partial y^2}b^2\left(V(t)\right)dt \end{array}\right) \\
& + r(v(t,S(t),V(t)) - \Delta(t)S(t) - \Delta_1(t)v_1(t,V(t)))dt \\
= \; & \left(\begin{array}{c} \Delta(t)\mu S\left(t\right) + \Delta_1(t)\left(\frac{\partial v_1}{\partial t} + \frac{\partial v_1}{\partial y}a\left(V(t)\right) + \frac{1}{2}\frac{\partial^2 v_1}{\partial y^2}b^2\left(V(t)\right)\right) \\ +r(v(t,S(t),V(t)) - \Delta(t)S(t) - \Delta_1(t)v_1(t,V(t))) \end{array}\right)dt \\
& + \left(\Delta(t)\sqrt{V(t)}S(t) + \Delta_1(t)\frac{\partial v_1}{\partial y}b\left(V(t)\right)\rho\right)dW_1^P\left(t\right) + \Delta_1(t)\frac{\partial v_1}{\partial y}b\left(V(t)\right)\sqrt{1-\rho^2}dW_2^P\left(t\right).
\end{aligned}
$$

Then, replicating require to equate the above two equations. Thus, we should find the following two equations for the replicating strategy $(\Delta(t), \Delta_1(t))$ as

$$\begin{aligned}
\frac{\partial v}{\partial x}\sqrt{V(t)}S(t)+\frac{\partial v}{\partial y}b(V(t))\rho &= \Delta(t)\sqrt{V(t)}S(t)+\Delta_1(t)\frac{\partial v_1}{\partial y}b(V(t))\rho, \\
\frac{\partial v}{\partial y}b(V(t))\sqrt{1-\rho^2} &= \Delta_1(t)\frac{\partial v_1}{\partial y}b(V(t))\sqrt{1-\rho^2}.
\end{aligned}$$

Solving this equation system, we obtain the following replicating strategy

$$\Delta_1(t)=\frac{\partial v}{\partial y}(t,S(t),V(t))/\frac{\partial v_1}{\partial y}(t,V(t))$$

and

$$\Delta(t)=\frac{\partial v}{\partial x}(t,S(t),V(t)).$$

Also, an equation similar to the one on the top of page 6 in Gatheral's book:

$$\begin{aligned}
&\frac{\frac{\partial v}{\partial t}+\frac{1}{2}x^2y\frac{\partial^2 v}{\partial x^2}+\frac{1}{2}b^2(y)\frac{\partial^2 v}{\partial y^2}+\rho b(y)x\sqrt{y}\frac{\partial^2 v}{\partial x\partial y}+rx\frac{\partial v}{\partial x}+a(y)\frac{\partial v}{\partial y}-rv}{\frac{\partial v}{\partial y}} \\
=\;&\frac{\frac{\partial v_1}{\partial t}+\frac{1}{2}b^2(y)\frac{\partial^2 v_1}{\partial y^2}+a(y)\frac{\partial v_1}{\partial y}-rv_1}{\frac{\partial v_1}{\partial y}}.
\end{aligned}$$

Thus, we note that this equation holds true for any arbitrary variance sensitive asset with $V_1(t) = v_1(t, V(t))$. So, it must be equal to a function on the independent variable t and y. We assume such a function to be $f(t, y)$, which is related to the definition of the market price of volatility risk (to be formally assumed momentarily). Please note that the above equation is a beauty. If you have $\frac{\partial v}{\partial x} = 0$, the left-hand side reduced to the right-hand side. So, we have

$$\frac{\partial v}{\partial t}+\frac{1}{2}x^2y\frac{\partial^2 v}{\partial x^2}+\frac{1}{2}b^2(y)\frac{\partial^2 v}{\partial y^2}+\rho b(y)x\sqrt{y}\frac{\partial^2 v}{\partial x\partial y}+rx\frac{\partial v}{\partial x}+a(y)\frac{\partial v}{\partial y}-f(t,y)\frac{\partial v}{\partial y}-rv=0$$

and

$$\frac{\partial v_1}{\partial t}+\frac{1}{2}b^2(y)\frac{\partial^2 v_1}{\partial y^2}+a(y)\frac{\partial v_1}{\partial y}-f(t,y)\frac{\partial v_1}{\partial y}-rv_1=0.$$

For option valuation, $f(t,y)$ has to be pre-specified as part of the real world model. In a lot of reference this issue was not clearly discussed. Indeed, to use stochastic volatility, a full model should include two parts: the real world dynamics and an assumption of a functional form of the market price of volatility risk. The risk-neutral probably measure is totally auxiliary. And, it is implied from the Feymann-Kac's theorem. For example, from the pricing equation

$$\frac{\partial v}{\partial t}+\frac{1}{2}x^2y\frac{\partial^2 v}{\partial x^2}+\frac{1}{2}b^2\left(y\right)\frac{\partial^2 v}{\partial y^2}+\rho b\left(y\right)x\sqrt{y}\frac{\partial^2 v}{\partial x\partial y}+rx\frac{\partial v}{\partial x}+a\left(y\right)\frac{\partial v}{\partial y}-f(t,y)\frac{\partial v}{\partial y}-rv=0,$$

subject to

$$v(T,x,y)=P(x),$$

we have the price of an option with maturity T and payoff function $P(x)$ as

$$v(t,S(t),V(t))=e^{-r(T-t)}E_t^Q P(S(T)).$$

Here, Q is the risk neutral measure under which the dynamics of $(S(t),V(t))$ follows that

$$\begin{aligned}\frac{dS(t)}{S(t)} &= rdt+\sqrt{V(t)}dW_1^Q(t),\\ dV(t) &= [a(V(t))-f(t,V(t))]dt+b(V(t))\left[\rho dW_1^Q(t)+\sqrt{1-\rho^2}dW_2^Q(t)\right],\end{aligned}$$

where $(W_1^Q(t),W_2^Q(t))$ is a two-dimensional standard Brownian motion under the martingale pricing measure Q.

2.2 Understanding the market price of risk

Now, I am revealing something very interesting but seldom clarified in the literature of stochastic volatility modeling. The question is how do we model $f(t,V(t))$ in a parametric form so that the model has both concrete mathematical meaning and the mathematical tractability. We start from an investigation of the economic reasons. Now, $V_1(t)=v_1(t,V(t))$ plays a role as the "Delta-hedged" option as discussed in page 7 of Gatheral's book. Now, let us look at the excess return of such an asset.

Indeed, we have

$$\begin{aligned}
& dv_1(t, V(t)) \\
= \;& \frac{\partial v_1}{\partial t}dt + \frac{\partial v_1}{\partial y}\left(a\left(V(t)\right)dt + b\left(V(t)\right)\rho dW_1^P(t) + b\left(V(t)\right)\sqrt{1-\rho^2}dW_2^P(t)\right) \\
& + \frac{1}{2}\frac{\partial^2 v_1}{\partial y^2}b^2\left(V(t)\right)dt \\
= \;& \left(\frac{\partial v_1}{\partial t} + a\left(V(t)\right)\frac{\partial v_1}{\partial y} + \frac{1}{2}\frac{\partial^2 v_1}{\partial y^2}b^2\left(V(t)\right)\right)dt \\
& + \frac{\partial v_1}{\partial y}\left(b\left(V(t)\right)\rho dW_1^P(t) + b\left(V(t)\right)\sqrt{1-\rho^2}dW_2^P(t)\right) \\
= \;& \left(f(t, V(t))\frac{\partial v_1}{\partial y} + rv_1(t, V(t))\right)dt + \frac{\partial v_1}{\partial y}\left(b\left(V(t)\right)\rho dW_1^P(t) + b\left(V(t)\right)\sqrt{1-\rho^2}dW_2^P(t)\right),
\end{aligned}$$

where we have applied

$$\frac{\partial v_1}{\partial t} + \frac{1}{2}b^2(y)\frac{\partial^2 v_1}{\partial y^2} + a(y)\frac{\partial v_1}{\partial y} - f(t, y)\frac{\partial v_1}{\partial y} - rv_1 = 0.$$

So, we have

$$dv_1(t, V(t)) - rv_1(t, V(t))dt = b\left(V(t)\right)\frac{\partial v_1}{\partial y}\left[\frac{f(t, V(t))}{b\left(V(t)\right)}dt + dW_v^P(t)\right],$$

where

$$W_v^P(t) = \rho W_1^P(t) + \sqrt{1-\rho^2}W_2^P(t)$$

represent a Brownian motion driving the volatility process. Analogy to (1.4), we have

$$\lim_{\Delta\to 0}\frac{1}{\Delta}\frac{E_t v_1(t+\Delta, V(t+\Delta)) - v_1(t, V(t))}{v_1(t, V(t))} - r = \frac{b\left(V(t)\right)\frac{\partial v_1}{\partial y}}{v_1(t, V(t))}\frac{f(t, V(t))}{b\left(V(t)\right)}.$$

This is an analog to the CAPM as many people claimed. But, I am now not toally sure how to fully interpret everything. Can you? To what maximal extent? The LHS is no doubt an instantenous excess return. I can somehow regard $\frac{b(V(t))\frac{\partial v_1}{\partial y}}{v_1(t,V(t))}$ as a percentage of "Brownian risk corresponding to $W_v^P(t)$" and $\frac{f(t,V(t))}{b(V(t))}$ is the excess premium per unit of $dW_v^P(t)$.

Now, we look at the excess return of the option. We deduce that

$$
\begin{aligned}
& dv(t,S(t),V(t)) \\
= & \left(\frac{\partial v}{\partial x}\mu S\left(t\right)-rS\left(t\right)\frac{\partial v}{\partial x}+f(t,V(t))\frac{\partial v}{\partial y}+rv\right)dt \\
& +\left(\frac{\partial v}{\partial x}\sqrt{V(t)}S\left(t\right)+\frac{\partial v}{\partial y}b\left(V(t)\right)\rho\right)dW_1^P\left(t\right)+\frac{\partial v}{\partial y}b\left(V(t)\right)\sqrt{1-\rho^2}dW_2^P\left(t\right).
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
& dv(t,S(t),V(t))-rv(t,S(t),V(t))dt \\
= & \frac{\partial v}{\partial x}S\left(t\right)\left[(\mu-r)dt+\sqrt{V(t)}dW_1^P\left(t\right)\right]+f(t,V(t))\frac{\partial v}{\partial y}dt+\frac{\partial v}{\partial y}b\left(V(t)\right)dW_v^P\left(t\right) \\
= & \sqrt{V(t)}S\left(t\right)\frac{\partial v}{\partial x}\left[\frac{\mu-r}{\sqrt{V(t)}}dt+dW_1^P(t)\right]+b\left(V(t)\right)\frac{\partial v_1}{\partial y}\left[\frac{f(t,V(t))}{b\left(V(t)\right)}dt+dW_v^P(t)\right].
\end{aligned}
$$

This is just an application of the Ito formula and the pricing equation we just obtained. We can see that

$$
\begin{aligned}
& \lim_{\Delta\to 0}\frac{1}{\Delta}\frac{E_t v(t+\Delta,S(t+\Delta),V(t+\Delta))-v(t,S(t),V(t))}{v(t,S(t),V(t))}-r \\
= & \frac{\sqrt{V(t)}S\left(t\right)\frac{\partial v}{\partial x}}{v(t,S(t),V(t))}\frac{\mu-r}{\sqrt{V(t)}}+\frac{b\left(V(t)\right)\frac{\partial v_1}{\partial y}}{v(t,S(t),V(t))}\frac{f(t,V(t))}{b\left(V(t)\right)}.
\end{aligned}
$$

Similar to our previous discussion, we can give the above equation a very interesting economics interpretation from excess returns.

Now, we formally call

$$
\lambda_1(t)=\frac{\mu-r}{\sqrt{V(t)}},\quad \lambda_2(t)=\frac{f(t,V(t))}{b\left(V(t)\right)}
$$

as the market price of risk. In more detail, I would call $\lambda_1(t)$ the market price of return risk and call $\lambda_2(t)$ the market price of volatility risk. Let us give a mathematical interpretation of the market price of risk. Indeed, if we hope to make everything rigorous, there are two things to consider. First of all, this two items render the drifts in the Girsanov change of measure. Look at the previous discussions on the risk-neutral measure Q. We can construct it using the Girsanov theorem. Now, we just need to find two drifts such that

$$\begin{aligned}
\frac{\mu - r}{\sqrt{V(t)}}dt + dW_1^P(t) &= \gamma_1(t)dt + dW_1^P(t) \\
\frac{f(t, V(t))}{b(V(t))}dt + dW_v^P(t) &= \rho\left[\gamma_1(t)dt + dW_1^P(t)\right] + \sqrt{1-\rho^2}\left[\gamma_2(t)dt + dW_2^P(t)\right],
\end{aligned}$$

i.e.,

$$\begin{aligned}
\gamma_1(t) &= \frac{\mu - r}{\sqrt{V(t)}} \\
\rho\gamma_1(t) + \sqrt{1-\rho^2}\gamma_2(t) &= \frac{f(t, V(t))}{b(V(t))}.
\end{aligned}$$

Thus, we can construct a new probability measure through

$$\frac{d\widehat{Q}}{dP}|_{\mathcal{F}_t} = \exp\left(\int_0^t \gamma_1(s)dW_1^P(s) + \int_0^t \gamma_2(s)dW_2^P(s) - \frac{1}{2}\int_0^t \gamma_1(s)^2 ds - \frac{1}{2}\int_0^t \gamma_2(s)^2 ds\right).$$

Under $\widehat{Q}$,

$$\begin{aligned}
W_1^{\widehat{Q}}(t) &= \int_0^t \gamma_1(s)ds + W_1^P(t), \\
W_2^{\widehat{Q}}(t) &= \int_0^t \gamma_2(s)ds + W_2^P(t),
\end{aligned}$$

is a standard Brownian motion. Now, we can see that

$$\begin{aligned}
\frac{dS(t)}{S(t)} &= rdt + \sqrt{V(t)}dW_1^{\widehat{Q}}(t), \\
dV(t) &= [aV(t)) - f(t, V(t))]dt + b(V(t))\left[\rho dW_1^{\widehat{Q}}(t) + \sqrt{1-\rho^2}dW_2^{\widehat{Q}}(t)\right].
\end{aligned}$$

So, we can see that $\widehat{Q}$ is a construction of the risk-neutral implied from the Feymann-Kac formula. Please note that we also have to make sure the exponential local-martingale generated by the market price of risk is a true martingale. So, technically, some sort of Novikov conditions should be verified. This is sometimes very challenging. So, I can see a lot of finance papers tried to by pass this kind of discussions.

Or, we say, in the research of finance, this kind theoretical mathematical issue is not quite necessary. However, this mathematical issue also partially serves as a rule for us to design a concrete stochastic volatility model. Morever, to obtain some good analytical formulas or approximations, we usually need to make sure the risk neutral dynamics of the volatility process is relatively analytical tractable. For example, we usually require some good form of the drift $a(V(t)) - f(t, V(t))$. For example, we consider a mean-reverting CEV diffusion stochastic volatility model with

$$\begin{aligned} a(x) &= \kappa(\theta - x), \\ b(x) &= \alpha x^{\beta}. \end{aligned}$$

Among many we have very simple choices of $f(t, V(t))$ as a constant or a linear function in $V(t)$. For example, if $f(t, V(t)) = \xi V(t)$, we will have the market price of volatility risk as

$$\lambda_2(t) = \frac{\xi V(t)}{\alpha V(t)^{\beta}} = \frac{\xi}{\alpha} V(t)^{1-\beta}.$$

Note that in the SPX options market, people like to assue $0 \leq \beta \leq 1$. So, such a model looks OK.

An interesting question is the estimation of the market price of volatility risk. I think the key is

$$\begin{aligned} &\lim_{\Delta \to 0} \frac{1}{\Delta} \frac{E_t v(t+\Delta, S(t+\Delta), V(t+\Delta)) - v(t, S(t), V(t))}{v(t, S(t), V(t))} - r \\ = &\frac{\sqrt{V(t)} S(t) \frac{\partial v}{\partial x}}{v(t, S(t), V(t))} \frac{\mu - r}{\sqrt{V(t)}} + \frac{\frac{\partial v_1}{\partial y}}{v(t, S(t), V(t))} f(t, V(t)). \end{aligned}$$

The question is how? What is a best method? Also, my puzzle is what is a economic meaning full specification of the market price of volatility risk? Does it increase as the volatility increase or decrease? Also, we have many interesting questions for stochastic volatility.

3 Beyond Black-Scholes: jumps

We start from the physical probability measure $\mathbb{P}$. We employ a compound Poisson process (CPP hereafter) $Q(t) = \sum_{n=1}^{N(t)} Y_n$ with $Y_n = X_n - 1$, where X_n has i.i.d. distributions with mean $\mu_S + 1$

($\mathbb{E}^{\mathbb{P}} Y_n = \mu_S$). We note that the compensated compound Poisson (CCPP hereafter) process $Q(t) - \lambda\mu_S t$ is a martingale (exercise). The Merton jump-diffusion model is specified as follows

$$\frac{dS(t)}{S(t-)} = \mu dt + \sigma dW(t) + d[Q(t) - \lambda\mu_S t].$$

The reason why we employ the CCPP $Q(t) - \lambda\mu_S t$ in the model is that we hope to take the advantage of its martingale property. Intuitively, we have

$$\mathbb{E}_t^{\mathbb{P}}\left(\frac{dS(t)}{S(t-)}\right) = \mu dt + \sigma \mathbb{E}_t^{\mathbb{P}} dW(t) + \mathbb{E}_t^{\mathbb{P}} d[Q(t) - \lambda\mu_S t] = \mu dt,$$

i.e., the local mean rate of return is μ. This is in analogy to the Black-Scholes case.

3.1 Explicit solution

This equation can be equivalently written as

$$\begin{aligned} \frac{dS(t)}{S(t-)} &= (\mu - \lambda\mu_S)dt + \sigma dW(t) + dQ(t) && (3.1) \\ &= (\mu - \lambda\mu_S)dt + \sigma dW(t) + d\left(\sum_{n=1}^{N(t)} (X_n - 1)\right). \end{aligned}$$

The solution of this equation is given by

$$S(t) = S(0)\exp\left(\sigma W(t) + \left(\mu - \lambda\mu_S - \frac{1}{2}\sigma^2\right) t\right) \prod_{n=1}^{N(t)} X_n. \tag{3.2}$$

A formal proof of this claim requires systematical knowledge in stochastic calculus with jumps, see, e.g., Chapter 11 in [21]. Here, I provide an intuitive argument to convince ourselves. Indeed, when jumps do not occur, the dynamics of asset price evolves according to

$$\frac{dS(t)}{S(t)} = (\mu - \lambda\mu_S)dt + \sigma dW(t); \tag{3.3}$$

when jump occurs, we have

$$\frac{dS(t)}{S(t-)} = d\left(\sum_{n=1}^{N(t)} (X_n - 1)\right),$$

i.e.,

$$\frac{S(t) - S(t-)}{S(t-)} = X_k - 1,$$

if the kth jump occurs at time t. This is equivalent to have $S(t) = S(t-)X_k$.

Assume jumps occur at $0 < \tau_1 < \tau_2 < \cdots < \tau_{N(t)} \leq t$. Thus, by solving (3.3), we have

$$S(\tau_1-) = S(0)\exp\left(\sigma W(\tau_1) + \left(\mu - \lambda\mu_S - \frac{1}{2}\sigma^2\right)\tau_1\right).$$

After the first jump, we have

$$S(\tau_1) = S(\tau_1-)X_1 = S(0)\exp\left(\sigma W(\tau_1) + \left(\mu - \lambda\mu_S - \frac{1}{2}\sigma^2\right)\tau_1\right)X_1.$$

Regard $S(\tau_1)$ as the new starting point and use the time-translation property of Brownian motion, we have

$$S(\tau_2-) = S(\tau_1)\exp\left(\sigma\left[W(\tau_2) - W(\tau_1)\right] + \left(\mu - \lambda\mu_S - \frac{1}{2}\sigma^2\right)(\tau_2 - \tau_1)\right).$$

After the second jump, we have

$$S(\tau_2) = S(\tau_2-)X_1 = S(\tau_1)\exp\left(\sigma\left[W(\tau_2) - W(\tau_1)\right] + \left(\mu - \lambda\mu_S - \frac{1}{2}\sigma^2\right)(\tau_2 - \tau_1)\right)X_2.$$

In general, we will have

$$S(\tau_{j+1}) = S(\tau_{j+1}-)X_j = S(\tau_j)\exp\left(\sigma\left[W(\tau_{j+1}) - W(\tau_j)\right] + \left(\mu - \lambda\mu_S - \frac{1}{2}\sigma^2\right)(\tau_{j+1} - \tau_j)\right)X_{j+1}$$

for all $j = 1, 2, 3, ..., N(t) - 1$. And, eventually, we have

$$S(t) = S(\tau_{N(t)})\exp\left(\sigma\left[W(t) - W(\tau_{N(t)})\right] + \left(\mu - \lambda\mu_S - \frac{1}{2}\sigma^2\right)(t - \tau_{N(t)})\right).$$

According to all the above equations, we obtain (3.2).

3.2 Risk-neutral

Suppose we hope to price a call option with strike K and maturity T. How jump-diffusions are applied in option pricing? Because of the nature of the jump component, it is not possible to use a finite number of assets to replicate option payoff. (Please have a try. And, some heuristic replication arguments can be found in Chapter 11 in [21] and Chapter 5 in [11].) So, we seek for a solution in weaker sense. We know the following principle. If there exists a risk-neutral measure, under which all discounted assets price are martingales, the market is free of arbitrage. Thus, once we have such a risk-neutral measure, denoted by $\mathbb{Q}$, we don't need to worry about the existence of arbitrage, as assumed by any efficient markets. Assume that the $\mathbb{Q}-$dynamics of $\{S(t)\}$ has the following pattern

$$\begin{aligned} \frac{dS(t)}{S(t-)} &= \mu^{\mathbb{Q}}dt + \sigma dW^{\mathbb{Q}}(t) + d[Q(t) - \lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}t] \\ &= \left[\mu^{\mathbb{Q}} - \lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}\right] dt + \sigma dW^{\mathbb{Q}}(t) + dQ(t). \end{aligned} \tag{3.4}$$

In other words, the drift changes from $\mu - \lambda\mu_S$ to $\mu^{\mathbb{Q}} - \lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}$. A necessary condition for $\mathbb{Q}$ being the desired risk-neutral measure is to make sure the discounted price $e^{-rt}S(t)$ becomes a $\mathbb{Q}-$martingale. When $\mu^{\mathbb{Q}} = r$ and $Q(t) = \sum_{n=1}^{N(t)} Y_n$ is a CPP with intensity $\lambda^{\mathbb{Q}}$ and $\mathbb{E}^{\mathbb{Q}}Y_n = \mu_S^{\mathbb{Q}}$, it becomes true.

To see this, we begin by solving (3.4) as

$$S(t) = S(0)\exp\left(\sigma W^{\mathbb{Q}}(t) + \left(\mu^{\mathbb{Q}} - \lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}} - \frac{1}{2}\sigma^2\right)t\right)\prod_{n=1}^{N(t)} X_n.$$

Thus, it is obvious to have

$$e^{-rt}S(t) = S(0)\exp\left(\sigma W^{\mathbb{Q}}(t) + \left(-\lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}} - \frac{1}{2}\sigma^2\right)t\right)\prod_{n=1}^{N(t)} X_n$$

For any $t > s > 0$, we will show

$$\mathbb{E}_s^{\mathbb{Q}}\left[e^{-rt}S(t)\right] = e^{-rs}S(s).$$

We note that

$$
\begin{aligned}
\mathbb{E}_s^{\mathbb{Q}} &= \mathbb{E}_s^{\mathbb{Q}}\left(S(0)\exp\left(\sigma W^{\mathbb{Q}}(t)+\left(-\lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}-\frac{1}{2}\sigma^2\right)t\right)\prod_{n=1}^{N(t)}X_n\right)\\
&= S(0)\mathbb{E}_s^{\mathbb{Q}}\left(\exp\left(\sigma W^{\mathbb{Q}}(t)-\frac{1}{2}\sigma^2 t\right)\right)\mathbb{E}_s^{\mathbb{Q}}\left(\exp\left(-\lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}t\right)\prod_{n=1}^{N(t)}X_n\right).
\end{aligned}
$$

It is obvious that

$$
\mathbb{E}_s^{\mathbb{Q}}\left(\exp\left(\sigma W^{\mathbb{Q}}(t)-\frac{1}{2}\sigma^2 t\right)\right)=\exp\left(\sigma W^{\mathbb{Q}}(s)-\frac{1}{2}\sigma^2 s\right).
$$

And, we have

$$
\begin{aligned}
\mathbb{E}_s^{\mathbb{Q}}\left(\exp\left(-\lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}t\right)\prod_{n=1}^{N(t)}X_n\right) &= \exp\left(-\lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}t\right)\mathbb{E}_s^{\mathbb{Q}}\left(\prod_{n=N(s)+1}^{N(t)}X_n\prod_{n=1}^{N(s)}X_n\right)\\
&= \exp\left(-\lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}t\right)\prod_{n=1}^{N(s)}X_n\mathbb{E}_s^{\mathbb{Q}}\left(\prod_{n=N(s)+1}^{N(t)}X_n\right),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbb{E}_s^{\mathbb{Q}}\left(\prod_{n=N(s)+1}^{N(t)}X_n\right) &= \sum_{k=0}^{\infty}\mathbb{E}_s^{\mathbb{Q}}\left(\prod_{n=N(s)+1}^{N(t)}X_n|N(t)-N(s)=k\right)\mathbb{Q}(N(t)-N(s)=k)\\
&= \sum_{k=0}^{\infty}\left(\mu_S^{\mathbb{Q}}+1\right)^k\frac{(t-s)^k\left(\lambda^{\mathbb{Q}}\right)^k\exp(-(t-s)\lambda^{\mathbb{Q}})}{k!}\\
&= \sum_{k=0}^{\infty}\frac{(t-s)^k\left(\lambda^{\mathbb{Q}}\right)^k\left(\mu_S^{\mathbb{Q}}+1\right)^k}{k!}\exp(-(t-s)\lambda^{\mathbb{Q}})\\
&= \exp\left((t-s)\lambda^{\mathbb{Q}}\left(\mu_S^{\mathbb{Q}}+1\right)\right)\exp(-(t-s)\lambda^{\mathbb{Q}})\\
&= \exp\left((t-s)\lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}\right).
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
\mathbb{E}_s^{\mathbb{Q}}\left(\exp\left(-\lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}t\right)\prod_{n=1}^{N(t)}X_n\right) &= \exp\left(-\lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}t\right)\prod_{n=1}^{N(s)}X_n\exp\left((t-s)\lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}\right) \\
&= \prod_{n=1}^{N(s)}X_n\exp\left(-\lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}s\right).
\end{aligned}
$$

Up to now, we obtain the risk-neutral dynamics

$$
\begin{aligned}
\frac{dS(t)}{S(t-)} &= rdt + \sigma dW^{\mathbb{Q}}(t) + d[Q(t) - \lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}t] \\
&= \left(r - \lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}\right)dt + \sigma dW^{\mathbb{Q}}(t) + dQ(t). \qquad (3.5)
\end{aligned}
$$

3.3 Change of measure

The change from measure from $\mathbb{P}$ to $\mathbb{Q}$ consists of two parts: the Brownian motion part and the jump part, respectively. Comparing the $\mathbb{P}$ and $\mathbb{Q}-$dynamics of $S(t)$, i.e., (3.1) and (3.5), we have

$$
(\mu - \lambda\mu_S)\,dt + \sigma dW(t) + dQ(t) \equiv \left(r - \lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}\right)dt + \sigma dW^{\mathbb{Q}}(t) + dQ(t).
$$

For the Brownian motion part, simular to Black-Scholes, we consider the Girsanov type change of measure such that

$$
W^{\mathbb{Q}}(t) = W(t) + \theta t,
$$

becomes a Brownian motion, where

$$
\theta = \frac{1}{\sigma}\left(\mu - r + \lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}} - \lambda\mu_S\right). \qquad (3.6)
$$

According to the Girsanov theorem, the Brownian part of the Radon-Nykodim derivative can be given by

$$
Z_1(t) = \exp\left(-\theta W(t) - \frac{1}{2}\theta^2 t\right).
$$

For the jump part, by applying the following Radon-Nykodim derivative:

$$Z_2(t) = \exp((\lambda - \lambda^{\mathbb{Q}})t) \prod_{n=1}^{N(t)} \frac{\lambda^{\mathbb{Q}} f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}, \tag{3.7}$$

the intensity of the CPP changes from λ to $\lambda^{\mathbb{Q}}$ and the density the jump size changed from $f_{\mathbb{P}}$ to $f_{\mathbb{Q}}$. According to our current context, we just need to make sure $f_{\mathbb{Q}}$ satisfies

$$\mathbb{E}^{\mathbb{Q}} Y_n = \int_R y f_{\mathbb{Q}}(y) dy = \mu_S^{\mathbb{Q}}.$$

Putting the two parts together, the risk-neutral measure can be constructed from the following

$$\left. \frac{d\mathbb{Q}}{d\mathbb{P}} \right|_{\mathcal{F}(T)} = Z_1(T) Z_2(T)$$

Indeed, it seems that the risk-neutral martingale property of discounted asset price can be obtained even without changing the jump intensity and jump size distribution. And we also note that $\lambda^{\mathbb{Q}}$ and $\mu_S^{\mathbb{Q}}$ show up in a product in the market price of risk (3.6); so are λ and μ_S. So, to make sure jump generates contribution in the market price of risk, it is enough to allow change either in intensity or jump size. In econometrics, it is conventional to make a trade-off by changing the intensity while keeping the jumps size distribution. In such a case, the Radon-Nykodim derivative for jump part becomes

$$Z_2(t) = \exp((\lambda - \lambda^{\mathbb{Q}})t) \left(\frac{\lambda^{\mathbb{Q}}}{\lambda} \right)^{N(t)}.$$

We will return to this point when we consider estimation of jump-diffusion models.

In the end of this section, we prove the change-of-measure property associated with (3.7). First, we prove that $Z_2(t)$ is a martingale using the definition directly. Then, prove the change of measure effect by calculating moment generating functions.

Indeed, to show $Z_2(t)$ is a martingale under $\mathbb{P}$, we just need to prove that

$$\mathbb{E}^{\mathbb{P}}\left(Z_2(t)|\mathcal{F}(s)\right) = Z_2(s).$$

By substitution, we get

$$\begin{aligned}
& \mathbb{E}^{\mathbb{P}}\left(\exp((\lambda-\lambda^{\mathbb{Q}})t)\prod_{n=1}^{N(t)}\frac{\lambda^{\mathbb{Q}}f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}|\mathcal{F}(s)\right) \\
= \;& \exp((\lambda-\lambda^{\mathbb{Q}})t)\prod_{n=1}^{N(s)}\frac{\lambda^{\mathbb{Q}}f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}\mathbb{E}^{\mathbb{P}}\left(\prod_{n=N(s)+1}^{N(t)}\frac{\lambda^{\mathbb{Q}}f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}\right).
\end{aligned}$$

According to the stationarity-increments property of Poisson processes, we have

$$\mathbb{E}^{\mathbb{P}}\left(\prod_{n=N(s)+1}^{N(t)}\frac{\lambda^{\mathbb{Q}}f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}\right) = \mathbb{E}^{\mathbb{P}}\left(\prod_{n=1}^{N(t-s)}\frac{\lambda^{\mathbb{Q}}f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}\right).$$

To obviate the random term $N(t-s)$, we take conditional expectation on it

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}\left(\prod_{n=1}^{N(t-s)}\frac{\lambda^{\mathbb{Q}}f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}\right) &= \mathbb{E}^{\mathbb{P}}\left(\mathbb{E}^{\mathbb{P}}\left(\prod_{n=1}^{N(t-s)}\frac{\lambda^{\mathbb{Q}}f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}|N(t-s)\right)\right) \\
&= \sum_{k=0}^{\infty}\mathbb{E}^{\mathbb{P}}\left(\prod_{n=1}^{N(t-s)}\frac{\lambda^{\mathbb{Q}}f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}|N(t-s)=k\right)\mathbb{P}(N(t-s)=k) \\
&= \left(\frac{\lambda^{\mathbb{Q}}}{\lambda}\right)^k\sum_{k=0}^{\infty}\mathbb{E}^{\mathbb{P}}\left(\prod_{n=1}^{k}\frac{f_{\mathbb{Q}}(Y_n)}{f_{\mathbb{P}}(Y_n)}|N(t-s)=k\right)\mathbb{P}(N(t-s)=k).
\end{aligned}$$

By independence, we obtain that

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}\left(\prod_{n=1}^{k}\frac{f_{\mathbb{Q}}(Y_n)}{f_{\mathbb{P}}(Y_n)}|N(t-s)=k\right) &= \mathbb{E}^{\mathbb{P}}\left(\prod_{n=1}^{k}\frac{f_{\mathbb{Q}}(Y_n)}{f_{\mathbb{P}}(Y_n)}\right) \\
&= \prod_{n=1}^{k}\mathbb{E}^{\mathbb{P}}\left(\frac{f_{\mathbb{Q}}(Y_n)}{f_{\mathbb{P}}(Y_n)}\right) \\
&= \prod_{n=1}^{k}\int_R\frac{f_{\mathbb{Q}}(y)}{f_{\mathbb{P}}(y)}f_{\mathbb{P}}(y)dy = 1.
\end{aligned}$$

Thus, plugging the above results into (5.4), we obtain that

$$
\begin{aligned}
& \mathbb{E}^{\mathbb{P}}\left(\exp((\lambda-\lambda^{\mathbb{Q}})t)\prod_{n=1}^{N(t)}\frac{\lambda^{\mathbb{Q}}f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}|\mathcal{F}(s)\right) \\
= \;& \exp((\lambda-\lambda^{\mathbb{Q}})t)\prod_{n=1}^{N(s)}\frac{\lambda^{\mathbb{Q}}f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}\sum_{k=0}^{\infty}\left(\frac{\lambda^{\mathbb{Q}}}{\lambda}\right)^k\frac{(t-s)^k(\lambda)^k}{k!}\exp(-(t-s)\lambda) \\
= \;& \exp((\lambda-\lambda^{\mathbb{Q}})t)\prod_{n=1}^{N(s)}\frac{\lambda^{\mathbb{Q}}f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}\sum_{k=0}^{\infty}\left(\lambda^{\mathbb{Q}}\right)^k\frac{(t-s)^k}{k!}\exp(-(t-s)\lambda) \\
= \;& \exp\left((\lambda-\lambda^{\mathbb{Q}})s\right)\prod_{n=1}^{N(s)}\frac{\lambda^{\mathbb{Q}}f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)} = Z_2(s).
\end{aligned}
$$

where we have applied the following fact from calculus

$$
\sum_{k=0}^{\infty}\left(\lambda^{\mathbb{Q}}\right)^k\frac{(t-s)^k}{k!} = \exp\left(\lambda^{\mathbb{Q}}(t-s)\right),
$$

Thus, $Z_2(t)$ is a martingale under $\mathbb{P}$. The martingale property of $\{\exp((\lambda-\lambda^{\mathbb{Q}})t)\left(\frac{\lambda^{\mathbb{Q}}}{\lambda}\right)^{N(t)}\}$ is a special case.

Then, we prove that by change of measure from $\mathbb{P}$ to $\mathbb{Q}$ by $Z_2(t)$, the intensity of compound Poisson process $\{Q(t)\}$ changes from λ to $\lambda^{\mathbb{Q}}$. First, we calculate the moment generating function if $Q(t)$ is a CPP with intensity $\lambda^{\mathbb{Q}}$ under $\mathbb{Q}$. We deduce that

$$
\begin{aligned}
\mathbb{E}^{\mathbb{Q}}\left(e^{\theta Q(t)}\right) &= \mathbb{E}^{\mathbb{Q}}\left(\prod_{n=1}^{N(t)}e^{\theta Y_n}\right) \\
&= \mathbb{E}^{\mathbb{Q}}\left(\mathbb{E}^{\mathbb{Q}}\left(\prod_{n=1}^{N(t)}e^{\theta Y_n}|N(t)\right)\right) \\
&= \sum_{k=0}^{\infty}\mathbb{E}^{\mathbb{Q}}\left(\prod_{n=1}^{N(t)}e^{\theta Y_n}|N(t)=k\right)\mathbb{Q}(N(t)=k). \qquad (3.8)
\end{aligned}
$$

By independence, we have

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}\left(\prod_{n=1}^{N(t)} e^{\theta Y_n}|N(t)=k\right) &= \mathbb{E}^{\mathbb{Q}}\left(\prod_{n=1}^{k} e^{\theta Y_n}|N(t)=k\right) \\
&= \prod_{n=1}^{k}\mathbb{E}^{\mathbb{Q}}\left(e^{\theta Y_n}\right) = \prod_{n=1}^{k}\int_R e^{\theta y} f_{\mathbb{Q}}(y)dy \\
&= \left(\int_R e^{\theta y} f_{\mathbb{Q}}(y)dy\right)^k.
\end{aligned}$$

Then, plugging the above reseults to (3.8) , we get

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}\left(e^{\theta Q(t)}\right) &= \sum_{k=0}^{\infty}\left(\int_R e^{\theta y} f_{\mathbb{Q}}(y)dy\right)^k \frac{(t\lambda^{\mathbb{Q}})^k}{k!}\exp(-t\lambda^{\mathbb{Q}}) \\
&= \exp\left(\lambda^{\mathbb{Q}} t\int_R \exp\left(\theta y\right) f_{\mathbb{Q}}(y)dy\right)\exp(-t\lambda^{\mathbb{Q}}) \\
&= \left(\lambda^{\mathbb{Q}} t\left(\int_R \exp\left(\theta y\right) f_{\mathbb{Q}}(y)dy - 1\right)\right).
\end{aligned}$$

On the other hand, assuming $Q(t)$ is a CPP with intensity λ under probability measure $\mathbb{P}$, we calculate the moment-generating function of $Q(t)$ under probability measure $\mathbb{Q}$ via change of measure from $\mathbb{P}$ to $\mathbb{Q}$ by $Z_2(t)$. We will show that

$$\mathbb{E}^{\mathbb{Q}}\left(e^{\theta Q(t)}\right) = \left(\lambda^{\mathbb{Q}} t\left(\int_R \exp\left(\theta y\right) f_{\mathbb{Q}}(y)dy - 1\right)\right).$$

Therefore, since moment generating function characterize distribution, we claim that after measure transformation, the intensity of $Q(t)$ becomes $\lambda^{\mathbb{Q}}$ under probability measure $\mathbb{Q}$. The proofs are as follows:

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}\left(e^{\theta Q(t)}\right) &= \mathbb{E}^{\mathbb{P}}\left(e^{\theta Q(t)} Z_2(t)\right) \qquad (3.9) \\
&= \mathbb{E}^{\mathbb{P}}\left(\exp\left(\theta\sum_{n=1}^{N(t)} Y_n\right)\exp((\lambda-\lambda^{\mathbb{Q}})t)\prod_{n=1}^{N(t)}\frac{\lambda^{\mathbb{Q}} f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}\right).
\end{aligned}$$

Take out what is known, we have

$$
\begin{aligned}
& \mathbb{E}^{\mathbb{P}}\left(\exp\left(\theta\sum_{n=1}^{N(t)} Y_n\right)\exp((\lambda-\lambda^{\mathbb{Q}})t)\prod_{n=1}^{N(t)}\frac{\lambda^{\mathbb{Q}} f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}\right) \\
= \quad & \exp((\lambda-\lambda^{\mathbb{Q}})t)\mathbb{E}^{\mathbb{P}}\left(\exp\left(\theta\sum_{n=1}^{N(t)} Y_n\right)\prod_{n=1}^{N(t)}\frac{\lambda^{\mathbb{Q}} f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}\right) \\
= \quad & \exp((\lambda-\lambda^{\mathbb{Q}})t)\mathbb{E}^{\mathbb{P}}\left(\prod_{n=1}^{N(t)}\left(\exp\left(\theta Y_n\right)\frac{\lambda^{\mathbb{Q}} f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}\right)\right).
\end{aligned}
$$

Then take conditional expectation on it, we can get

$$
\begin{aligned}
& \exp((\lambda-\lambda^{\mathbb{Q}})t)\mathbb{E}^{\mathbb{P}}\left(\prod_{n=1}^{N(t)}\left(\exp\left(\theta Y_n\right)\frac{\lambda^{\mathbb{Q}} f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}\right)\right) \\
= \quad & \exp((\lambda-\lambda^{\mathbb{Q}})t)\mathbb{E}^{\mathbb{P}}\left(\mathbb{E}^{\mathbb{P}}\left(\prod_{n=1}^{N(t)}\left(\exp\left(\theta Y_n\right)\frac{\lambda^{\mathbb{Q}} f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}\right)|N(t)\right)\right) \\
= \quad & \exp((\lambda-\lambda^{\mathbb{Q}})t)\sum_{k=0}^{\infty}\mathbb{E}^{\mathbb{P}}\left(\prod_{n=1}^{N(t)}\left(\exp\left(\theta Y_n\right)\frac{\lambda^{\mathbb{Q}} f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}\right)|N(t)=k\right)\mathbb{P}(N(t)=k).
\end{aligned}
$$

By independence, we have

$$
\begin{aligned}
\mathbb{E}^{\mathbb{P}}\left(\prod_{n=1}^{N(t)}\left(\exp\left(\theta Y_n\right)\frac{\lambda^{\mathbb{Q}} f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}\right)|N(t)=k\right) \quad &= \quad \prod_{n=1}^{k}\mathbb{E}^{\mathbb{P}}\left(\exp\left(\theta Y_n\right)\frac{\lambda^{\mathbb{Q}} f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}|N(t)=k\right) \\
&= \quad \prod_{n=1}^{k}\mathbb{E}^{\mathbb{P}}\left(\exp\left(\theta Y_n\right)\frac{\lambda^{\mathbb{Q}} f_{\mathbb{Q}}(Y_n)}{\lambda f_{\mathbb{P}}(Y_n)}\right) \\
&= \quad \left(\frac{\lambda^{\mathbb{Q}}}{\lambda}\right)^k\prod_{n=1}^{k}\int_R \exp\left(\theta y\right)\frac{f_{\mathbb{Q}}(y)}{f_{\mathbb{P}}(y)}f_{\mathbb{P}}(y)dy \\
&= \quad \left(\frac{\lambda^{\mathbb{Q}}}{\lambda}\right)^k\left(\int_R \exp\left(\theta y\right)f_{\mathbb{Q}}(y)dy\right)^k.
\end{aligned}
$$

Then, by plugging the above results to (3.9), we obtain that

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\left(e^{\theta Q(t)}\right) &= \exp((\lambda-\lambda^{\mathbb{Q}})t)\sum_{k=0}^{\infty}\left(\frac{\lambda^{\mathbb{Q}}}{\lambda}\right)^k\left(\int_R \exp(\theta y)\, f_{\mathbb{Q}}(y)dy\right)^k\frac{(t\lambda)^k}{k!}\exp(-t\lambda)\\ &= \exp\left(\left(\lambda-\lambda^{\mathbb{Q}}\right)t\right)\exp\left(\left(\lambda^{\mathbb{Q}}\int_R \exp(\theta y)\, f_{\mathbb{Q}}(y)dy-\lambda\right))t\right)\\ &= \exp\left(\lambda^{\mathbb{Q}}t\left(\int_R \exp(\theta y)\, f_{\mathbb{Q}}(y)dy-1\right)\right).\end{aligned}$$

That is exactly what we want. We have proved that by changing measure from $\mathbb{P}$ to $\mathbb{Q}$, the intensity of CPP $Q(t)$ changes from λ to $\lambda^{\mathbb{Q}}$.

3.4 A closed-form formula for option pricing

Before closing this section, we discuss how a closed-form formula for option pricing can be derived. Merton [19] assumes that X_n has a lognormal distribution. More precisely, under the risk-neutral probability measure $\mathbb{Q}$, we assume $Z_n = \log X_n$ have a normal distribution with mean $\mu_Z^{\mathbb{Q}}$ and variance $\left(\sigma_Z^{\mathbb{Q}}\right)^2$. Since $\mathbb{Q}$ is assumed to be the risk-neutral probability measure, the discounted option price process $\{V(t)\}$ has to be a $\mathbb{Q}$-martingale also. Thus, we must have

$$V(0) = \mathbb{E}^{\mathbb{Q}}[e^{-rT}V(T)] = \mathbb{E}^{\mathbb{Q}}[e^{-rT}(S(T)-K)^+].$$

It follows from conditioning that

$$V(0) = \sum_{n=0}^{\infty} e^{-rT}\mathbb{E}^{\mathbb{Q}}[(S(T)-K)^+|N(T)=n]\mathbb{Q}(N(T)=n) = \sum_{n=0}^{\infty} T_{m,n}(K)e^{-\lambda^{\mathbb{Q}}T}\frac{(\lambda^{\mathbb{Q}}T)^n}{n!},$$

Conditioning on $N(T)=n$,

$$S_0(T) = s_0\exp\left(\left(r-\lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}-\frac{1}{2}\sigma^2\right)T+\sigma W_1(T)+\sum_{n=1}^{n} Z_n\right)$$

has a lognormal distribution. It follows from standard calculation that

$$V(0)=\sum_{n=0}^{\infty}e^{-\lambda^{\mathbb{Q}}T}\frac{(\lambda^{\mathbb{Q}}T)^n}{n!}\left[s_0e^{-\lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}T+n\left(\mu_Z^{\mathbb{Q}}+\frac{(\sigma_Z^{\mathbb{Q}})^2}{2}\right)}N\left(d_{1,n}\right)-e^{-rT}KN\left(d_{2,n}\right)\right],$$

where N represents the probability distribution function (c.d.f. hereafter) of a standard normal distribution throughout the paper; $d_{1,n}$ and $d_{2,n}$ are defined as

$$\begin{aligned}
d_{1,n}(s) &:= \frac{1}{\sqrt{\sigma^2T+n(\sigma_Z^{\mathbb{Q}})^2}}\left[\log\frac{s_0}{s}+\left(r-\lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}+\frac{\sigma^2}{2}\right)T+n(\mu_Z^{\mathbb{Q}}+(\sigma_Z^{\mathbb{Q}})^2)\right],\\
d_{2,n}(s) &:= \frac{1}{\sqrt{\sigma^2T+n(\sigma_Z^{\mathbb{Q}})^2}}\left[\log\frac{s_0}{s}+\left(r-\lambda^{\mathbb{Q}}\mu_S^{\mathbb{Q}}-\frac{\sigma^2}{2}\right)T+n\mu_Z^{\mathbb{Q}}\right].
\end{aligned}$$

3.5 Combining stochastic volatility and jumps

As an extension of stochastic volatility and jumps, we combine them to have stochastic volatility with jump models (SVJ), e.g.,

$$\begin{aligned}
\frac{dS(t)}{S(t)} &= \mu dt+\sqrt{V(t)}dW_1^P(t)+d[Q(t)-\lambda\mu_S t],\\
dV(t) &= a(V(t))dt+b(V(t))\left[\rho dW_1^P(t)+\sqrt{1-\rho^2}dW_2^P(t)\right].
\end{aligned}$$

And, we can even consider models with jumps in volatility (SVJJ), e.g.,

$$\begin{aligned}
\frac{dS(t)}{S(t)} &= \mu dt+\sqrt{V(t)}dW_1^P(t)+d[Q(t)-\lambda\mu_S t],\\
dV(t) &= a(V(t))dt+b(V(t))\left[\rho dW_1^P(t)+\sqrt{1-\rho^2}dW_2^P(t)\right]+dJ_V(t).
\end{aligned}$$

The way to study these models involves the previous methods for studying stochastic volatility and jumps.

4 Affine jump-diffusion models†

In these notes, I will give an introduction to affine processes, a large class of continuous-time models. Many famous diffusion and jump-diffusion models belong to this class. Affine processes admit several nice properties, e.g., analytical tractability in derivatives pricing. As generalizations or related alternatives, some affine based processes are proposed, such as quadratic models (see, e.g., Ahn, Dittmar, and Gallant [1], Gourieroux and Sufana [13], and Leippold and Wu [16]), Wishart processes, linear generating models (see, e.g., Carr, Gabaix, and Wu [3], Filipovic, Larsson, and Trolle [9]).

4.1 Definition

Here, without lose of generality, we define affine jump-diffusion processes directly. According to Duffie, Pan, and Singleton [8], the standard definition for $d-$dimensional jump-diffusion models is given as follows. Consider a process $X(t)$ governed by the following SDE

$$dX(t) = \mu\left(X(t)\right)dt + \sigma\left(X(t)\right)dW(t) + dJ(t), \tag{4.1}$$

for some functions $\mu(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma(\cdot) : \mathbb{R}^d \to \mathbb{R}^{d\times d}$, where $\{W(t)\}$ is a $d-$dimensional standard Brownian motion; $\{J(t)\}$ is a jump process with jump intensity $\lambda\left(X(t)\right)$ for some function $\lambda(\cdot)$. Process (4.1) is affine if and only if the following specifications hold

$$\begin{aligned}
\mu(x) &= K_0 + K_1 x, \text{ with } K_0 \in \mathbb{R}^d, K_1 \in \mathbb{R}^{d\times d},\\
\left(\sigma(x)\sigma^\top(x)\right)_{ij} &= (H_0)_{ij} + (H_1)_{ij}x, \text{ with } (H_0)_{ij} \in \mathbb{R} \text{ and } (H_1)_{ij} \in \mathbb{R}^d,\\
\lambda(x) &= \lambda_0 + \lambda_1^\top x, \text{ with } \lambda_0 \in \mathbb{R}, \lambda_1 \in \mathbb{R}^d.
\end{aligned}$$

After some suitable transformations (e.g., taking log of asset price), many popular continuous-time models fall into this class.

Indeed, a $d-$dimensional affine jump-diffusion model can be equivalently specified as follows (exer-

†I thank Chenxu (my PhD student) for his kind and careful preparation of the notes of this section.

cise, please refer to Proposition A.1 and Corollary A.1 in Duffie and Kan [7])

$$dX(t) = \mathcal{K}(\Theta - X(t))dt + \Sigma\sqrt{V(t)}dW(t) + dJ(t),$$

where $\{W(t)\}$ is a $d-$dimensional standard Brownian motion; $\mathcal{K}$ and Σ are $d \times d$ matrices; V is a diagonal matrix with the ith diagonal element given by

$$V_{ii}(t) = \alpha_i + \beta_i^\top X(t).$$

In other words, we have

$$V(t) = \begin{pmatrix} \alpha_1 + \beta_1^\top x & & & \\ & \alpha_2 + \beta_2^\top x & & \\ & & \ddots & \\ & & & \alpha_d + \beta_d^\top x \end{pmatrix}.$$

Thus,

$$\sqrt{V(t)} = \begin{pmatrix} \sqrt{\alpha_1 + \beta_1^\top x} & & & \\ & \sqrt{\alpha_2 + \beta_2^\top x} & & \\ & & \ddots & \\ & & & \sqrt{\alpha_d + \beta_d^\top x} \end{pmatrix}$$

Comparing with the standard notation, we have

- the drift $\mu(x) = K_0 + K_1 x$ with $K_0 = \mathcal{K}\Theta$ and $K_1 = -\mathcal{K}$,
- the diffusion $\left(\sigma(x)\sigma(x)^\top\right)_{ij} = (H_0)_{ij} + (H_1)_{ij}x = \left(\Sigma\sqrt{V(t)}\sqrt{V(t)}\Sigma^\top\right)_{ij} = \left(\Sigma V(t)\Sigma^\top\right)_{ij}$

Exercise 1. According to the definition, show that the Heston stochastic volatility model is affine:

$$\begin{aligned} dS(t) &= \mu S(t)dt + \sqrt{V(t)}S(t)dW^S(t), \\ dV(t) &= \kappa\left(\theta - V(t)\right)dt + \xi\sqrt{V(t)}dW^V(t), \end{aligned}$$

with ρ representing the correlation between the standard Brownian motions $\{W^S(t)\}$ and $\{W^V(t)\}$.

Exercise 2. According to the definition, show that the Parallel double Heston model (see Christoffersen, Heston, and Jacobs [5])

$$\begin{aligned} dS(t) &= rS(t)dt + \sqrt{V_1(t)}S(t)dW_1(t) + \sqrt{V_2(t)}S(t)dW_2(t), \\ dV_1(t) &= (a_1 - b_1 V_1(t))\,dt + \sigma_1\sqrt{V_1(t)}dW_3(t), \\ dV_2(t) &= (a_1 - b_1 V_2(t))\,dt + \sigma_2\sqrt{V_2(t)}dW_4(t), \end{aligned}$$

where $W_1(t)$ and $W_3(t)$, $W_2(t)$ and $W_4(t)$ are constantly correlated.

4.2 Analytical transform

Most affine processes do not admit closed-form transition densities, but their Laplace transforms are explicitly available. In other words, the expectation $\mathbb{E}\left(e^{-sX(t)}\right)$ is in closed-form for any complex number s. Suppose X is a one-dimensional continuous random variable and its law is implicit. How can we characterize the distribution of X? Characteristic function (CF hereafter) is defined by

$$f(t) := \mathbb{E}\left(e^{itX}\right) = \int_{-\infty}^{+\infty} e^{itx}p(x)dx,$$

where $p(x)$ is law of X. Indeed, the function $f(\cdot)$ is the Fourier transform of $p(\cdot)$. By applying inverse Fourier transform, we will get $p(x)$. Fortunately, the characteristic function $f(t)$ usually admits explicit form.

A related notion is Laplace transform

$$L(s) := \mathbb{E}\left(e^{-sX}\right) = \int_{-\infty}^{+\infty} e^{-sx}p(x)dx,$$

for any complex number s. If we ignore some technical details, we can also deduce the law of X by inverse Laplace transform.

Similarly, if X is a random vector, Laplace transform can be defined as

$$L(s) = \mathbb{E}\left(e^{-s^\top X}\right) = \int_{-\infty}^{+\infty} e^{-s^\top x} p(x)dx, \tag{4.2}$$

where s is a column vector. Motivated by some applications, we can replace expectation (4.2) by conditional expectation, which is called "conditional Laplace transform". Duffie and Kan [7] and Duffie, Pan, and Singleton [8] indicate that the conditional Laplace transform of affine processes admits exponential affine form, i.e.,

$$\mathbb{E}\left(e^{-s^\top X(T)} | \mathcal{F}(t)\right) = \exp\left(A(T-t) + B^\top (T-t) X(t)\right),$$

i.e.,

$$\mathbb{E}\left(e^{-s^\top X(T)} | X(t) = x\right) = \exp\left(A(T-t) + B^\top (T-t) x\right),$$

where the functions $A(\cdot), B(\cdot)$ satisfy a group of ODEs, subject to initial conditions

$$A(0) = 0 \text{ and } B(0) = -s. \tag{4.3}$$

Exercise 3. Derive the ODEs for $A(\cdot)$ and $B(\cdot)$. Hint: $\mathbb{E}\left(e^{-s^\top X(T)} | \mathcal{F}(t)\right)$ is a martingale.

We consider affine diffusion model first. Denote by

$$\psi(-s, X(t), t, T) := E\left(e^{-s^\top X(T)} \middle| \mathcal{F}(t)\right). \tag{4.4}$$

Then, Itô's formula implies

$$d\psi(-s, X(t), t, T) = \frac{\partial \psi}{\partial t} dt + \nabla_x^\top \psi dX(t) + \frac{1}{2} dX(t)^\top \nabla_x^2 \psi dX(t), \tag{4.5}$$

where $\nabla_x^\top \psi$ and $\nabla_x^2 \psi$ are gradient and Hessian matrix of ψ with respect to $X(t)$, respectively. According to the exponential affine form of $\psi(-s, X(t), t, T)$

$$\psi(-s, X(t), t, T) = \exp\left(A(T-t) + B^\top (T-t) X(t)\right),$$

we deduce

$$\begin{aligned}
\frac{\partial \psi}{\partial t} &= \psi \cdot \left(-A'(T-t) - B'(T-t)^\top x\right), \\
\nabla_x \psi &= \psi \cdot B(T-t) = \psi \cdot (B_1(T-t), B_2(T-t), \cdots, B_d(T-t))^\top, \\
\nabla_x^2 \psi &= \psi \cdot B(T-t)B(T-t)^\top = \psi \cdot (B_i(T-t)B_j(T-t))_{ij}.
\end{aligned}$$

Also, recall the affine specifications, i.e.,

$$\mu(x) = K_0 + K_1 x, \quad \left(\sigma(x)\sigma^\top(x)\right)_{ij} = (H_0)_{ij} + (H_1)_{ij} x.$$

Then, we have

$$\begin{aligned}
\nabla_x^\top \psi dX(t) &= \psi B^\top (T-t)\left(K_0 + K_1 X(t)\right) dt + \psi B^\top (T-t)\left(H_0 + H_1 X(t)\right) dW(t), \\
dX(t)^\top \nabla_x^2 \psi dX(t) &= B(T-t)^\top \left(H_0 + H_1 X(t)\right) B(T-t) dt.
\end{aligned}$$

Plugging the above two equations into (4.5), we obtain

$$\begin{aligned}
& d\psi(-s, t, T, X(t)) \\
= \; & \psi \cdot \Big[\Big(-A'(T-t) - B'(T-t)^\top X(t) + B^\top (T-t)\left(K_0 + K_1 X(t)\right) \\
& + \frac{1}{2} B(T-t)^\top \left(H_0 + H_1 X(t)\right) B(T-t)\Big) dt + B^\top (T-t)\left(H_0 + H_1 X(t)\right) dW(t)\Big].
\end{aligned}$$

Since $\psi(-s, X(t), t, T)$ is a martingale, drift term on the RHS of above SDE should be identically zero, i.e.,

$$\begin{aligned}
0 \equiv \; & -A'(T-t) + B^\top (T-t) K_0 + \frac{1}{2} B(T-t)^\top H_0 B(T-t) \\
& -B'(T-t)^\top X(t) + B^\top (T-t) K_1 X(t) + \frac{1}{2} B(T-t)^\top H_1 X(t) B(T-t).
\end{aligned}$$

Here, there is a little trick to simplify $B(T-t)^\top H_1 X(t) B(T-t)$. As it is a scalar, we can rewrite is as a summation, i.e.,

$$B(T-t)^\top H_1 X(t) B(T-t) = \sum_{i,j=1}^{d} B_i(T-t)^\top \left(H_1 X(t)\right)_{ij} B_j(T-t),$$

where $(H_1X(t))_{ij}$ refers to ijth entry of $H_1X(t)$, which, by definition, is $(H_1)_{ij}X(t)$. Then, we further have

$$B(T-t)^\top H_1 X(t) B(T-t) = \sum_{i,j=1}^{d} B_i(T-t)^\top (H_1)_{ij} X(t) B_j(T-t).$$

Since $B_j(T-t)$ is a scalar function, we can interchange $B_j(T-t)$ and $X(t)$. It follows that

$$\begin{aligned} B(T-t)^\top H_1 X(t) B(T-t) &= \sum_{i,j=1}^{d} B_i(T-t)^\top (H_1)_{ij} B_j(T-t) X(t) \\ &= B(T-t)^\top H_1 B(T-t) X(t). \end{aligned}$$

Then, the martingale condition simplifies to

$$\begin{aligned} 0 \equiv\ & -A'(T-t) + B^\top(T-t)K_0 + \frac{1}{2}B(T-t)^\top H_0 B(T-t) \\ & + \left(-B'(T-t)^\top + B^\top(T-t)K_1 + \frac{1}{2}B(T-t)^\top H_1 B(T-t)\right) X(t). \end{aligned}$$

Note that, in the above equation, the first line on the RHS is free of $X(t)$, and the second line linearly depends on $X(t)$. To make sure RHS is identically zero, it is necessary and sufficient that

$$\begin{aligned} A'(T-t) &= K_0^\top B^\top(T-t) + \frac{1}{2}B(T-t)^\top H_0 B(T-t), \\ B'(T-t) &= K_1^\top B^\top(T-t) + \frac{1}{2}B(T-t)^\top H_1 B(T-t). \end{aligned}$$

Combining with initial conditions (4.3), we obtain the final result. For affine jump-diffusion model, thanks to the jump component, we will obtain additional terms using stochastic calculus for jumps, a topic beyond this course. The corresponding ODEs can be given as follow

$$
\begin{aligned}
A'(T-t) &= K_0^T B'(T-t) + \frac{1}{2}B(T-t)^\top H_0 B(T-t) + \lambda_0(\theta(B(T-t)) - 1), \\
B'(T-t) &= K_1^\top B^\top(T-t) + \frac{1}{2}B(T-t)^\top H_1 B(T-t) + \lambda_1(\theta(B(T-t)) - 1),
\end{aligned}
$$

with $\theta\left(c\right) = E\left[\exp\left(c^\top Z\right)\right]$, where Z is the vector of jump sizes.

4.3 Classification of affine models

Consider an $N-$dimensional affine diffusion process governed by the following SDE

$$dX(t) = \mathcal{K}\left(\Theta - X(t)\right)dt + \Sigma\sqrt{V(t)}dW(t), \tag{4.6}$$

for some $\mathcal{K}, \Sigma \in \mathbb{R}^{N\times N}$ and $\Theta \in \mathbb{R}^N$; $\{W(t)\}$ is an $N-$dimensional standard Brownian motion; $V(t)$ is a diagonal matrix with entry $V_{ii}(t) = \alpha_i + \beta_i^\top X(t)$. To ensure the square root of $V(t)$ is well-defined, we need affine process $\{X(t)\}$ to be "admissible", i.e., $V_{ii}(t) > 0$ for all i. Duffie and Kan [7] gives restrictions for admissibility in Condition A. In univariate case, this condition is usually called "Feller condition". In the following discussion, we always assume affine process (4.6) is admissible. We call (4.6) an $A_m(N)$ process if the rank of $\beta_i, 1 \le i \le N$ is m. Note that for $N-$dimensional affine processes, it has $N+1$ subclasses, $A_0(N), \cdots, A_N(N)$. Here, m characterize the number of volatility factors. See Cheridito, Filipović, and Kimmel [4] for detailed discussion. Singleton [22] and Dai and Singleton [6] propose the canonical representation of $A_m(N)$. Another canonical form for $A_m(N)$ processes is developed by Collin-Dufresne. This type of canonical form is based on unspanned stochastic volatility (USV) and Taylor expansions.

4.4 Fourier pricing methods: an example from the Heston model

We start some investigations on the Heston model which is governed by the following SDE under the risk-neutral measure Q

$$
\begin{aligned}
\frac{dS(t)}{S(t)} &= rdt + \sqrt{V(t)}dW_1(t),\ S(0) = s_0, && (4.7a)\\
dV(t) &= \kappa(\theta - V(t))dt + \sigma\sqrt{V(t)}\left[\rho dW_1(t) + \sqrt{1-\rho^2}dW_2(t)\right],\ V(0) = v_0. && (4.7b)
\end{aligned}
$$

where $\{W(t)\} = \{(W_1(t), W_2(t))\}$ is a two dimensional standard Brownian motion; $\mu, \kappa, \rho, \sigma$ and θ are all scalar parameters. To show that the Heston model is affine, we take logarithm of $S(t)$. Then, model (4.7a)-(4.7b) becomes

$$
\begin{aligned}
dY(t) &= (r - \frac{1}{2}V(t))dt + \sqrt{V(t)}dW_1(t),\ S(0) = s_0, && (4.8a)\\
dV(t) &= \kappa(\theta - V(t))dt + \sigma\sqrt{V(t)}\left[\rho dW_1(t) + \sqrt{1-\rho^2}dW_2(t)\right],\ V(0) = v_0, && (4.8b)
\end{aligned}
$$

with $Y(t) = \log S(t)$.

In previous exercises, we have proved that the Heston model is a special affine process, whose Laplace transform (moment generating function) admits exponential affine form, i.e.,

$$
\begin{aligned}
& \psi\left(u, t, T, y, v\right)\\
=\ & E^Q\left[\exp\left(u_y Y(T) + u_v V(T)\right)|Y(t) = y, V(t) = v\right]\\
=\ & \exp\left(A(T-t) + B_1(T-t)y + B_2\left(T-t\right)v\right),
\end{aligned}
$$

where the functions $A(T-t)$, $B_1(T-t)$ and $B_2\left(T-t\right)$ satisfy

$$
\begin{aligned}
A'(T-t) &= rB_1(T-t) + \kappa\theta B_2\left(T-t\right), && (4.9a)\\
B_1'(T-t) &= 0 && (4.9b)\\
B_2'(T-t) &= -\kappa B_2(T-t) - \frac{1}{2}[B_1(T-t) + B_1^2(T-t) && (4.9c)\\
&\quad +2\sigma\rho B_1(T-t)B_2(T-t) + \sigma^2 B_2^2(T-t)]
\end{aligned}
$$

with the following initial conditions

$$A(0) = 0,\ B_1(0) = u_y,\ B_2(0) = u_v.$$

It is obvious that the solution to (4.9b) is

$$B_1(T-t) = u_y. \tag{4.10}$$

Plugging (4.10) into (4.9c), we obtain

$$B_2'(T-t) = \frac{1}{2}\sigma^2 B_2^2(T-t) + (\sigma\rho u_y - \kappa)B_2(T-t) + \frac{1}{2}u_y^2 - \frac{1}{2}u_y,$$

which is a Riccati equation. Solving the equation, we have

$$B_2(T-t) = \frac{(\kappa - \sigma\rho u_y + d) - (\kappa - \sigma\rho u_y - d)ge^{d(T-t)}}{\sigma^2(1 - ge^{d(T-t)})}, \tag{4.11}$$

where

$$d = \sqrt{(\sigma\rho w_R - \kappa)^2 - \sigma^2(w_R^2 - w_R)},$$

and

$$g = \frac{(\kappa - \sigma\rho u_y) + \sqrt{(\sigma\rho u_y - \kappa)^2 - \sigma^2(u_y^2 - u_y)} - u_v\sigma^2}{(\kappa - \sigma\rho u_y) - \sqrt{(\sigma\rho u_y - \kappa)^2 - \sigma^2(u_y^2 - u_y)} - u_v\sigma^2}.$$

Finally, plugging (4.10) and (4.11) into (4.9a), we obtain

$$A(T-t) = ru_y(T-t) + \frac{\kappa\theta}{\sigma^2}\left[(\kappa - \sigma\rho u_y + d)(T-t) - 2\log\left(\frac{1 - ge^{d(T-t)}}{1-g}\right)\right]. \tag{4.12}$$

Now, our aim is to recover option price from the above Laplace transform. It is well-known that, under risk neutral measure, the discounted option price is a martingale. Thus, the time-t price $C(Y(t), V(t), T-t, K)$ of a call option with maturity T and strike K with payoff $(S(t) - K)^+$ can be represented as follows:

$$C(Y(t),V(t),T-t,K)=e^{-r(T-t)}E^Q\left[(S(T)-K)^+\middle|\mathcal{F}(t)\right]=e^{-r(T-t)}E^Q\left[\left(e^{Y(T)}-K\right)^+\middle|\mathcal{F}(t)\right],$$

where $\{\mathcal{F}(t)\}$ denote the filtration. Now, introduce a vector $u:=(u_y,u_v)^\top=(1,0)^\top$. Then, we have

$$\begin{aligned}
& C(Y(t),V(t),T-t,K) \qquad (4.13)\\
= \;& e^{-r(T-t)}E^Q\left[\left(e^{u_yY(T)+u_vV(T)}-K\right)1\left\{u_yY(T)+u_vV(T)\geq\log K\right\}\middle|\mathcal{F}(t)\right]\\
= \;& e^{-r(T-t)}E^Q\left[e^{u_yY(T)+u_vV(T)}1\left\{u_yY(T)+u_vV(T)\geq\log K\right\}\middle|\mathcal{F}(t)\right]\\
& -Ke^{-r(T-t)}E^Q\left[1\left\{u_yY(T)+u_vV(T)\geq\log K\right\}\middle|\mathcal{F}(t)\right].
\end{aligned}$$

Let $G_{a,b}(x;Y(t),V(t),T-t)$ denote the price of a security that pays $e^{a_1Y(t)+a_2V(t)}$ at time T in the event that $b_1Y(t)+b_2V(t)\leq x$. Mathematically, define $G_{a,b}(x;Y(t),V(t),T-t)$ as follow

$$G_{a,b}(x;Y(t),V(t),T-t):=e^{-r(T-t)}E^Q\left[e^{a_1Y(T)+a_2V(T)}1\left\{b_1Y(T)+b_2V(T)\leq x\right\}\middle|\mathcal{F}(t)\right]. \qquad (4.14)$$

Then, option price (4.13) admits the following representation:

$$C(Y(t),V(t),T-t,K)=G_{u,-u}(-\log K;T-t,Y(t),V(t))-KG_{0,-u}(-\log K;T-t,Y(t),V(t)).$$

The Fourier-Stieltjes transform $\mathcal{G}_{a,b}(\cdot;Y(t),V(t),T-t)$ of $G_{a,b}(\cdot;Y(t),V(t),T-t)$, if well-defined, is given by

$$\mathcal{G}_{a,b}(v;Y(t),V(t),T-t)=\int_{\mathbb{R}}e^{ivx}dG_{a,b}(x;Y(t),V(t),T-t).$$

Note that

$$dG_{a,b}(x;Y(t),V(t),T-t)=e^{-r(T-t)}E^Q\left[e^{a_1Y(T)+a_2V(T)}\delta\left(b_1Y(T)+b_2V(T)-x\right)\middle|\mathcal{F}(t)\right]dx,$$

where $\delta(\cdot)$ is the Dirac Delta function. See Kanwal [14] for its definition. A popular definition is given by generalized integral: we call $\delta(\cdot)$ the Dirac Delta function if and only if

$$\int_{\mathbb{R}} f(x)\,\delta(x)\,dx = f(0),$$

for any sufficiently smooth function $f(\cdot)$. If $\mathcal{G}_{a,b}(v; Y(t), V(t), T-t)$ is well-defined, Fubini theorem implies that we can interchange the integral and the conditional expectation and obtain

$$\mathcal{G}_{a,b}(v; Y(t), V(t), T-t) = E^Q\left[\int_{\mathbb{R}} e^{ivx} e^{a_1 Y(T)+a_2 V(T)} \delta\left(b_1 Y(T) + b_2 V(T) - x\right) dx \,\middle|\, \mathcal{F}(t)\right]$$

By the definition of Dirac Delta function, we further deduce

$$\mathcal{G}_{a,b}(v; Y(t), V(t), T-t) = E^Q\left[e^{(a_1+ivb_1)Y(T)+(a_2+ivb_2)V(T)} \,\middle|\, \mathcal{F}(t)\right] = \psi\left(a+ivb, Y(t), V(t), t, T\right), \tag{4.15}$$

which is the Laplace transform (4.4) for $X(t) = (Y(t), V(t))^\top$. In the Section 4.5, we provide a more elementary approach to prove the relation (4.15).

Therefore, applying inverse Fourier transform, we recover $G_{a,b}(x; Y(t), V(t), T-t)$ and then price vanilla options analytically. Gil-Pelaez [12] proposes a useful formula to obtain cumulative distribution function from characteristic function. A direct application leads to

$$\begin{aligned} & G_{a,b}(x; Y(t), V(t), T-t) \\ = \; & \frac{\psi(a, Y(t), V(t), t, T)}{2} + \frac{1}{2\pi}\int_0^{+\infty} \frac{e^{ivx}\psi(a-ivb, Y(t), V(t), t, T) - e^{-ivx}\psi(a+ivb, Y(t), V(t), t, T)}{iv} dv. \end{aligned}$$

Note that $e^{ivx}\psi(a-ivb, \cdot)$ is the complex conjugate of $e^{-ivx}\psi(a+ivb, \cdot)$. Thus, the integral on the RHS simplifies to

$$\begin{aligned} & \int_0^{+\infty} \frac{e^{ivx}\psi(a-ivb, Y(t), V(t), t, T) - e^{-ivx}\psi(a+ivb, Y(t), V(t), t, T)}{iv} dv. \\ = \; & -2\int_0^{+\infty} \frac{\operatorname{Im}\left[e^{-ivx}\psi(a+ivb, Y(t), V(t), t, T)\right]}{iv}. \end{aligned}$$

These are heuristic derivation of inverse transform since we omit some technical details. Duffie, Pan, and Singleton [8] gives rigorous expression of inverse transform as stated in the following proposition.

PROPOSITION 1. *Suppose for fix* $T \in [0,+\infty)$, a *and* $b \in \mathbb{R}^2$, *that* $\mathcal{G}_{a,b}(v;Y(t),V(t),T-t)$ *is well-defined for any* $v \in \mathbb{R}$, *and that*

$$\int_{\mathbb{R}} \left|\psi\left(a+ivb,y,v',t,T\right)\right| dv < \infty.$$

Then, $G_{a,b}(\cdot;Y(t),V(t),T-t)$ *is well-defined by (4.14) and given by*

$$\begin{aligned} & G_{a,b}(x;Y(t),V(t),T-t) \\ = \quad & \frac{\psi(a,Y(t),V(t),t,T)}{2} - \frac{1}{\pi}\int_0^{+\infty} \frac{\operatorname{Im}\left[\psi(a+ivb,Y(t),V(t),t,T)\,e^{-ivx}\right]}{v} dv, \end{aligned}$$

where $\operatorname{Im}(c)$ *denotes the imaginary part of* $c \in \mathbb{C}$.

4.5 An alternative proof of (4.15)

The Fourier-Stieltjes transform $\mathcal{G}_{a,b}(\cdot;Y(t),V(t),T-t)$ of $G_{a,b}(\cdot;Y(t),V(t),T-t)$, if well-defined, is given by

$$\mathcal{G}_{a,b}(v;Y(t),V(t),T-t) = \int_{\mathbb{R}} e^{ivx} dG_{a,b}(x;Y(t),V(t),T-t),$$

where $G_{a,b}(x;Y(t),V(t),T-t)$ is defined in (4.14). In what follows, we consider three case: $b_2 > 0$, $b_2 < 0$ and $b_2 = 0$.

Case 1. In the first case, note that

$$G_{a,b}(x;Y(t),V(t),T-t) = e^{-r(T-t)} \int_{\mathbb{R}} \int_{-\infty}^{(x-b_1y)/b_2} e^{a_1y+a_2v'} p\left(y,v' \middle| Y(t),V(t)\right) dv'dy,$$

where $p\left(y,v' \middle| Y(t),V(t)\right)$ denotes the transition density of $(Y(T),V(T))^{\top}$. Differentiating x on both sides of $G_{a,b}(x;Y(t),V(t),T-t)$ implies

$$\frac{dG_{a,b}(x;Y(t),V(t),T-t)}{dx} = \frac{e^{-r(T-t)}}{b_2} \int_{\mathbb{R}} e^{a_1y+a_2(x-b_1y)/b_2} p\left(y, \frac{x-b_1y}{b_2} \middle| Y(t),V(t)\right) dy.$$

Plugging it into $\mathcal{G}_{a,b}\left(v; Y(t), V(t), T-t\right)$, we have

$$\mathcal{G}_{a,b}\left(v; Y(t), V(t), T-t\right) = \frac{e^{-r(T-t)}}{b_2} \int_{\mathbb{R}^2} e^{ivx} e^{a_1 y + a_2 (x - b_1 y)/b_2} p\left(y, \frac{x-b_1 y}{b_2}\right| \left. Y(t), V(t)\right) dydx.$$

Introduce $v' = (x - b_1 y)/b_2$. Replacing x by $b_2 v' + b_1 y$ on the RHS of the above equation, we have

$$\begin{aligned} & \mathcal{G}_{a,b}\left(v; Y(t), V(t), T-t\right) \\ = \; & e^{-r(T-t)} \int_{\mathbb{R}^2} e^{iv(b_2 v' + b_1 y)} e^{a_1 y + a_2 v'} p\left(y, v' \middle| Y(t), V(t)\right) dydv' \\ = \; & e^{-r(T-t)} \int_{\mathbb{R}^2} e^{(a_1 + ivb_1) y + (a_2 + ivb_2) v'} p\left(y, v' \middle| Y(t), V(t)\right) dydv' \\ = \; & E^Q\left[e^{(a_1 + ivb_1) Y(T) + (a_2 + ivb_2) V(T)} \middle| \mathcal{F}(t) \right]. \end{aligned}$$

The final equation is exactly conditional Laplace transform $\psi\left(a + ivb, Y(t), V(t), t, T\right)$.

Case 2. For the second case, the integral form of $G_{a,b}\left(x; Y(t), V(t), T-t\right)$ is given by

$$G_{a,b}\left(x; Y(t), V(t), T-t\right) = e^{-r(T-t)} \int_{\mathbb{R}} \int_{(x-b_1 y)/b_2}^{+\infty} e^{a_1 y + a_2 v'} p\left(y, v' \middle| Y(t), V(t)\right) dv' dy.$$

Differentiating x on both sides of $G_{a,b}\left(x; Y(t), V(t), T-t\right)$ implies

$$\frac{dG_{a,b}\left(x; Y(t), V(t), T-t\right)}{dx} = -\frac{e^{-r(T-t)}}{b_2} \int_{\mathbb{R}} e^{a_1 y + a_2 (x - b_1 y)/b_2} p\left(y, \frac{x-b_1 y}{b_2}\right| \left. Y(t), V(t)\right) dy.$$

Plugging into $\mathcal{G}_{a,b}\left(v; Y(t), V(t), T-t\right)$, we have

$$\mathcal{G}_{a,b}\left(v; Y(t), V(t), T-t\right) = -\frac{e^{-r(T-t)}}{b_2} \int_{\mathbb{R}^2} e^{ivx} e^{a_1 y + a_2 (x - b_1 y)/b_2} p\left(y, \frac{x-b_1 y}{b_2}\right| \left. Y(t), V(t)\right) dydx.$$

Similarly, introduce $v' = (x - b_1 y)/b_2$. Replacing x by $b_2 v' + b_1 y$ on the RHS of above equation, we have

$$
\begin{aligned}
& \mathcal{G}_{a,b}\left(v; Y(t), V(t), T-t\right) \\
= \;& e^{-r(T-t)} \int_{\mathbb{R}^2} e^{iv(b_2 v' + b_1 y)} e^{a_1 y + a_2 v'} p\left(y, v' \middle| Y(t), V(t)\right) dy dv' \\
= \;& e^{-r(T-t)} \int_{\mathbb{R}^2} e^{(a_1 + ivb_1) y + (a_2 + ivb_2) v'} p\left(y, v' \middle| Y(t), V(t)\right) dy dv' \\
= \;& E^Q\left[e^{(a_1 + ivb_1) Y(T) + (a_2 + ivb_2) V(T)} \middle| \mathcal{F}(t) \right].
\end{aligned}
$$

This result is consistent with what we obtained in previous case.

Case 3. If $b_2 = 0$, $G_{a,b}\left(x; Y(t), V(t), T-t\right)$ degenerates to

$$
G_{a,b}\left(x; Y(t), V(t), T-t\right) = e^{-r(T-t)} \int_{\mathbb{R}^2} e^{a_1 y + a_2 v'} 1\left\{b_1 y \leq x\right\} dv' dy.
$$

We separete this case into three sub-cases: $b_1 > 0$, $b_1 < 0$ and $b_1 = 0$. When $b_1 > 0$, we can rewrite the above integral as

$$
G_{a,b}\left(x; Y(t), V(t), T-t\right) = e^{-r(T-t)} \int_{\mathbb{R}} \int_{-\infty}^{x/b_1} e^{a_1 y + a_2 v'} p\left(y, v' \middle| Y(t), V(t)\right) dy dv'.
$$

Differentiating x on both sides of $G_{a,b}\left(x; Y(t), V(t), T-t\right)$ implies

$$
\frac{dG_{a,b}\left(x; Y(t), V(t), T-t\right)}{dx} = \frac{e^{-r(T-t)}}{b_1} \int_{\mathbb{R}} e^{a_1 x / b_1 + a_2 v'} p\left(\frac{x}{b_1}, v' \middle| Y(t), V(t)\right) dv'.
$$

Plugging into $\mathcal{G}_{a,b}\left(v; Y(t), V(t), T-t\right)$, we have

$$
\mathcal{G}_{a,b}\left(v; Y(t), V(t), T-t\right) = \frac{e^{-r(T-t)}}{b_1} \int_{\mathbb{R}^2} e^{ivx} e^{a_1 x / b_1 + a_2 v'} p\left(\frac{x}{b_1}, v' \middle| Y(t), V(t)\right) dv' dx.
$$

Introduce $y = x/b_1$. Replacing x by $b_1 y$ on the RHS of the above equation, we have

$$\begin{aligned}
\mathcal{G}_{a,b}\left(v; Y(t), V(t), T-t\right) &= e^{-r(T-t)} \int_{\mathbb{R}^2} e^{ivb_1 y} e^{a_1 y + a_2 v'} p\left(y, v' \middle| Y(t), V(t)\right) dv' dy \\
&= e^{-r(T-t)} \int_{\mathbb{R}^2} e^{(a_1 + ivb_1) y + a_2 v'} p\left(y, v' \middle| Y(t), V(t)\right) dv' dy \\
&= E^Q \left[e^{(a_1 + ivb_1) Y(T) + a_2 V(T)} \middle| \mathcal{F}(t) \right].
\end{aligned}$$

The final equation is $\psi\left(a + ivb, Y(t), V(t), t, T\right)$ with $b_2 = 0$. When $b_1 < 0$, $G_{a,b}\left(x; Y(t), V(t), T-t\right)$ can be represented by

$$G_{a,b}\left(x; Y(t), V(t), T-t\right) = e^{-r(T-t)} \int_{\mathbb{R}} \int_{x/b_1}^{+\infty} e^{a_1 y + a_2 v'} p\left(y, v' \middle| Y(t), V(t)\right) dy dv'.$$

Differentiating x on both sides of $G_{a,b}\left(x; Y(t), V(t), T-t\right)$ implies

$$\frac{dG_{a,b}\left(x; Y(t), V(t), T-t\right)}{dx} = -\frac{e^{-r(T-t)}}{b_1} \int_{\mathbb{R}} e^{a_1 x/b_1 + a_2 v'} p\left(\frac{x}{b_1}, v' \middle| Y(t), V(t)\right) dv'.$$

Plugging into $\mathcal{G}_{a,b}\left(v; Y(t), V(t), T-t\right)$, we have

$$\mathcal{G}_{a,b}\left(v; Y(t), V(t), T-t\right) = -\frac{e^{-r(T-t)}}{b_1} \int_{\mathbb{R}^2} e^{ivx} e^{a_1 x/b_1 + a_2 v'} p\left(\frac{x}{b_1}, v' \middle| Y(t), V(t)\right) dv' dx.$$

Introduce $y = x/b_1$. Replacing x by $b_1 y$ on the RHS of the above equation, we have

$$\begin{aligned}
\mathcal{G}_{a,b}\left(v; Y(t), V(t), T-t\right) &= e^{-r(T-t)} \int_{\mathbb{R}^2} e^{ivb_1 y} e^{a_1 y + a_2 v'} p\left(y, v' \middle| Y(t), V(t)\right) dv' dy \\
&= e^{-r(T-t)} \int_{\mathbb{R}^2} e^{(a_1 + ivb_1) y + a_2 v'} p\left(y, v' \middle| Y(t), V(t)\right) dv' dy \\
&= E^Q \left[e^{(a_1 + ivb_1) Y(T) + a_2 V(T)} \middle| \mathcal{F}(t) \right].
\end{aligned}$$

This is consistent with the case $b_1 > 0$. When $b_1 = b_2 = 0$, note that

$$1\left\{b_1 y + b_2 v' \leq x\right\} = 1\left\{x \geq 0\right\},$$

which is free of y and v'. Thus, indicator function $1\{x \geq 0\}$ can be taken out of the integral. Then, $G_{a,b}(x; Y(t), V(t), T-t)$ can be written as

$$dG_{a,b}(x; Y(t), V(t), T-t) = e^{-r(T-t)} d(1\{x \geq 0\}) \int_{\mathbb{R}^2} e^{a_1 y + a_2 v'} p\left(y, v' \middle| Y(t), V(t)\right) dydv'.$$

Plugging into $\mathcal{G}_{a,b}(v; Y(t), V(t), T-t)$, we have

$$\mathcal{G}_{a,b}(v; Y(t), V(t), T-t) = e^{-r(T-t)} \int_{\mathbb{R}^3} e^{ivx} e^{a_1 y + a_2 v'} p\left(y, v' \middle| Y(t), V(t)\right) d(1\{x \geq 0\}) dydv'.$$

Let us handle the integral according to x first, i.e., we need to calculate

$$\int_{\mathbb{R}} e^{ivx} d(1\{x \geq 0\}).$$

Note that $1\{x \geq 0\}$ is the CDF of a random variable Z, which aggregates its value at 0 with probability 1. Thus, this integral can be regarded as $E\left[e^{ivZ}\right] = 1$. Based on this fact, Fourier-Stieltjes transform $\mathcal{G}_{a,b}(v; Y(t), V(t), T-t)$ simplifies to

$$\begin{aligned}
\mathcal{G}_{a,b}(v; Y(t), V(t), T-t) &= e^{-r(T-t)} \int_{\mathbb{R}^2} e^{a_1 y + a_2 v'} p\left(y, v' \middle| Y(t), V(t)\right) dydv' \\
&= E^Q\left[e^{a_1 Y(T) + a_2 V(T)} \middle| \mathcal{F}(t) \right],
\end{aligned}$$

which is $\psi(a + ivb, Y(t), V(t), t, T)$ with $b_1 = b_2 = 0$.

In summary, we conclude that

$$\mathcal{G}_{a,b}(v; Y(t), V(t), T-t) = \psi(a + ivb, Y(t), V(t), t, T),$$

for any $a, b \in \mathbb{R}^2$.

Part II

Applications in Optimal Portfolio Choice

This part is devoted to applications in optimal portfolio choice.

5 Optimal portfolio choice: a martingale approach

5.1 A very brief literature review

The pioneering work can be traced back to Markowitz (1952) in a mean-variance setting, assuming quadratic utility function. Then, Samuelson and Merton's work in 1969 and 1971 used continuous-time dynamic programming techniques, applying HJB equations to solve the problem. Then, it comes Karatzas, Lehoczky, and Shreve (1987), Cox and Huang (1989, 1991) which apply martingale tools to find explicit solutions. The most recent milestones include Ocone and Karatzas (1991) and Detemple Garcia, and Rindisbacher (2003).

5.2 The model

Uncertainty structure is a tuple (Ω, F, P), a filtration $\mathcal{F}$, and a d-dimentional Brownian motion W. The financial market consists of a risk-free money market account B_t appreciating a spot interest rate r_t, which is progressively measurable and integrable:

$$\int_0^t r_s ds < \infty,\ \mathbb{P} - a.s., \forall t \in [0, T] .$$

The dynamics of the price B_t is characterized by

$$dB_t = r_t B_t dt.$$

There are also d stocks $S_t \in \mathbb{R}_+^d$, which pay dividends. The evolution of the stock vector process is

$$dS_t + D_t dt = I_t^S \left(\mu_t dt + \sigma_t dW_t\right),$$

where D_t is $d-$dimensional vector representing the dividend rate; μ_t is a $d-$dimensional vector representing the total expected return; σ_t is a $d \times d$ matrix representing the volatility; $I_t^S = diag\left(S_t\right)$. In one-dimensional case, this becomes

$$dR_t = S_t^{-1}(dS_t + D_t dt).$$

The dynamics can also be written in terms of returns as

$$dR_t = \left[I_t^S\right]^{-1}(dS_t + D_t dt) = \mu_t dt + \sigma_t dW_t,$$

where $\left[I_t^S\right]^{-1} S_t$ is return on capital gain; $\left[I_t^S\right]^{-1} D_t dt$ is the return on dividends. The processes are all progressively measurable and integrable P-a.s. in the following sense

$$\int_0^T \|\mu_t\| \, dt < \infty, \ \int_0^T \|D_t\| \, dt < \infty, \ \int_0^T \|\sigma_t\|^2 \, dt < \infty, \ \int_0^T \sigma_t \sigma_t' dt < \infty.$$

We assume that σ_t^{-1} exists for any t.

We define the market price of risk (Sharpe ratio) process as

$$\theta_t = \sigma_t^{-1}\left(\mu_t - r_t \times 1_{d \times 1}\right).$$

We assume that θ_t satisfies

$$\int_0^T \|\theta_t\|^2 \, dt < \infty.$$

as well as the Novikov condition:

$$E\left[\exp\left(\frac{1}{2}\int_0^T \|\theta_t\|^2 \, dt\right)\right] < \infty.$$

The existence of σ_t^{-1} implies that

$$dW_t = \sigma_t^{-1}\left[\left[I_t^S\right]^{-1}(dS_t + D_t dt) - \mu_t dt\right].$$

It means that we can always trade in the stocks and money market account in order to replicate the "uncertainty" W_t. The market is complete in this sense.

Let c_t denote the consumption process, which is progressively measurable, integrable and satifying $c_t \geq 0$. Denote by C the collection of such consumption processes c_t. We consider the following expected total utility of the consumptions

$$U(c) = E\left[\int_0^T u(c_t, t)\, dt\right].$$

Here $u(\cdot, t)$ is utility function, which is continuous, twice continuously differentiable and concave. We assume the following *Inada* conditions:

$$u'(0, t) = \infty \text{ and } u'(\infty, t) = 0.$$

Here are some examples:

1. The Log utility:
$$u(c, t) = \log c.$$

2. The Constant Relative Risk Aversion (CRRA) utility:
$$u(c, t) = \frac{c^{1-R}}{1-R},$$

3. The Hyperbolic Absolute Risk Aversion (HARA) utility:
$$u(c, t) = \frac{(c+A)^{1-R}}{1-R},$$

Now we define portfolio policies. Let π_t be the dollar amount invested at each risky asset. Assume that π_t is progressively measurable and satisfying

$$\int_0^T \left(\left\|\pi_t'(\mu_t - r_t \times 1_{d\times 1})\right\| + \left\|\pi_t'\sigma_t\sigma_t'\pi_t\right\|\right) dt < \infty, P - a.s..$$

Denote by P^π the collection of π_t satisfying this condition. We assume jointly $(c, \pi) \in C \otimes P^\pi$. Denote by X_t the wealth process. Thus, it has dynamics

$$dX_t = -c_t dt + \pi_t' \left(I_t^S\right)^{-1} (dS_t + D_t dt) + \left(X_t - \pi_t' 1_{d\times 1}\right) r_t dt,\ X_0 = x. \tag{5.1}$$

We can understand the second term as the numbers of shares in the risk assets (i.e., $\pi_t' \left(I_t^S\right)^{-1}$) multiplied by the dollar earned (i.e., $dS_t + D_t dt$) by holding them in an infinitesimal time period. Collecting terms, we obtain that

$$\begin{aligned} dX_t &= (X_t r_t - c_t)\, dt + \pi_t' \left[(\mu_t - r_t 1_{d\times 1})\, dt + \sigma_t dW_t\right] \\ &= (X_t r_t - c_t)\, dt + \pi_t' \sigma_t \left[\theta_t dt + dW_t\right], \\ X_0 &= x. \end{aligned}$$

We define $(c,\pi) \in C \otimes P^\pi$ to be admissible by $(c,\pi) \in A(x)$ iff $X_t \geq 0,\ \forall t \in [0,T]$. This is the no-bankruptcy condition. We say that an admissible pair $(c,\pi) \in A(x)$ is optimal $(c,\pi) \in A^*(x)$ iff no other $(c',\pi') \in A(x)$ satisfies that $U(c') > U(c)$. We seek for the solution to the following optimization problem:

$$\max_{(\pi_t, c_t)} E\left[\int_0^T u(c_t, t)dt\right].$$

5.3 A static consumption-portfolio problem

First, we define:

$$\eta_t = \exp\left(-\int_0^t \theta_v' \cdot dW_v - \frac{1}{2}\int_0^t \theta_v' \theta_v' dv\right) < \infty, \forall t \in [0,T],\ P - a.s.$$

Suppose Novikov condition is satisfied. Thus η_t is a martingale and $E[\eta_t] = 1, \forall t \in [0,T]$. Girsanov theorem applies here. On $\mathcal{F}_T$, we definte Q by

$$\left.\frac{dQ}{dP}\right|_{\mathcal{F}_T} = \eta_T$$

And, let

$$W_t^Q = W_t + \int_0^t \theta_s ds, \text{ for } t \leq T.$$

Thus, W^Q is a Q Brownian motion process. Under this measure, the dynamics of the risky asset becomes

$$dS_t + D_t dt = I_t^S \left(r_t dt + \sigma_t dW_t^Q \right).$$

Define $\xi_t = b_t \eta_t$, where $b_t = \exp\left(-\int_0^t r_v dv\right)$. Here, ξ_t is called the state price density (SPD hereafter). We usually call Q the risk-neutral probability measure.

DEFINITION 1. (Static problem) Suppose $c \in C$. Define

$$B(x) = \left\{ c \in C | E\left(\int_0^T \xi_t c_t dt\right) \leq x \right\}.$$

$B(x)$ is called a budget set. $c \in B(x)$ is called optimal for static problem $c \in B^*(x)$ iff there does not exist $c' \in B(x)$ such that $U(c') > U(c)$.

REMARK 1. By the Girsanov theorem and the properties of conditional expectations, we have

$$\begin{aligned}
E\left(\int_0^T \xi_t c_t dt\right) &= \int_0^T E^P(\xi_t c_t)\, dt \\
&= \int_0^T E^P(b_t \eta_t c_t)\, dt \\
&= \int_0^T E^P\left(b_t c_t E_t^P \eta_T\right) dt \\
&= \int_0^T E^P\left[E_t^P(b_t c_t \eta_T)\right] dt \\
&= \int_0^T E^P(b_t c_t \eta_T)\, dt \\
&= \int_0^T E^Q(b_t c_t)\, dt = E^Q\left(\int_0^T b_t c_t dt\right).
\end{aligned}$$

Thus, the inequality condition in $B(x)$ becomes

$$E^Q\left(\int_0^T b_t c_t dt\right) \leq x, \tag{5.2}$$

i.e., the risk-netrual expectation of total discounted consumption is no more than the initial capital x. It is natural to conjecture as follows: in a risk-neutral world, in which all assets appreciate a mean-rate of return r_t, the expected total future consumptions should be less than or equal to the initial capital.

THEOREM 1. *$(c, \pi) \in A^* (x) \Rightarrow c \in B^* (x)$ and if $c \in B^* (x)$ then there exists an admissible π such that $(c, \pi) \in A^* (x)$.*

Proof. We first prove that $(c, \pi) \in A (x) \Rightarrow c \in B (x)$ and if $c \in B (x)$ then there exists an admissible π such that $(c, \pi) \in A (x)$.

Indeed, based on the wealth equation and the definition of the SPD

$$dX_t = (X_t r_t - c_t)\, dt + \pi_t' \sigma_t \left[\theta_t dt + dW_t\right],$$

and

$$d\xi_t = -\xi_t \left[r_t dt + \theta_t' dW_t\right]$$

By using the Ito product formula, we obtain that

$$d\left(\xi_t X_t\right) = \left(-\xi_t c_t\right) dt + \xi_t \left(\pi_t' \sigma_t - X_t \theta_t'\right) dW_t,$$

which is equivalent to

$$\xi_t X_t - x = -\int_0^t \xi_v c_v dv + \int_0^t \xi_v \left(\pi_v' \sigma_v - X_v \theta_v'\right) dW_v.$$

Thus, we have the following universally true result

$$\xi_t X_t + \int_0^t \xi_v c_v dv = x + \int_0^t \xi_v \left(\pi_v' \sigma_v - X_v \theta_v'\right) dW_v. \tag{5.3}$$

Note that $\int_0^t \xi_v \left(\pi_v' \cdot \sigma_v - X_v \theta_v'\right) \cdot dW_v$ is a local martingale.

First, assume $(c, \pi) \in A (x)$. We will show that $c \in B (x)$. By definition we have $X_t \geq 0$, $\forall t \in [0, T]$.

So $\xi_t X_t \geq 0$, $\int_0^t \xi_v c_v dv \geq 0$, $\forall t \in [0,T]$. Thus, we have

$$x + \int_0^t \xi_v \left(\pi_v' \sigma_v - X_v \theta_v'\right) dW_v \geq 0.$$

Because $\int_0^t \xi_v \left(\pi_v' \sigma_v - X_v \theta_v'\right) dW_v$ is a local martingale and it is bounded from below, thus it's a supermartingale, see, e.g., Exercise 1.9 in Karatzas' tutorial. So, we have

$$E\left(\xi_t X_t + \int_0^t \xi_v c_v dv\right) \leq x, \forall t \in [0,T].$$

Especially, for $t = T$, we have

$$E\left(\xi_T X_T + \int_0^T \xi_v c_v dv\right) \leq x.$$

Thus, we have

$$E\left(\int_0^T \xi_v c_v dv\right) \leq x,$$

i.e., $c \in B(x)$.

Next let's assume that $c \in B(x)$, we will show that there exists π such that $(c,\pi) \in A(x)$. That is to show that, by trading according to the wealth equation (5.1), we always have $X_t \geq 0$. Denote by

$$M_t = E_t\left(\int_0^T \xi_v c_v dv\right).$$

It's a martingale. By martingale representation theorem, there exists a unique predicable, integrable in the sense of P^2 process $\phi \in R^d$, such that

$$M_t = M_0 + \int_0^t \phi_v' dW_v.$$

Let π_t be

$$\pi_t = X_t \left(\sigma_t'\right)^{-1} \theta_t + \xi_t^{-1} \left(\sigma_t'\right)^{-1} \phi_t.$$

And we can see that this ϕ depend on c. Thus, by plugging this into (5.3), we have

$$\begin{aligned}
\xi_t X_t + \int_0^t \xi_v c_v dv &= x + \int_0^t \xi_v \left(\left[X_t \left(\sigma_t'\right)^{-1} \theta_t + \xi_t^{-1} \left(\sigma_t'\right)^{-1} \phi_t \right]' \sigma_v - X_v \theta_v' \right) dW_v \\
&= x + \int_0^t \xi_v \left(\left[X_t \theta_t' \sigma_t^{-1} + \xi_t^{-1} \phi_t' \sigma_t^{-1} \right] \sigma_v - X_v \theta_v' \right) dW_v \\
&= x + \int_0^t \xi_v \left(X_t \theta_t' + \xi_t^{-1} \phi_t' - X_v \theta_v' \right) dW_v \\
&= x + \int_0^t \phi_v' dW_v \\
&= x + M_t - M_0. \qquad (5.4)
\end{aligned}$$

Because $c \in B(x)$, i.e., $x \geq M_0$, we deduce that

$$\xi_t X_t + \int_0^t \xi_v c_v dv \geq M_t = E_t \left[\int_0^T \xi_v c_v dv \right].$$

So, we have

$$\xi_t X_t \geq E_t \left[\int_t^T \xi_v c_v dv \right] \geq 0.$$

Next, we go one more step to show that $(c, \pi) \in A^*(x) \Rightarrow c \in B^*(x)$ and if $c \in B^*(x)$ then there exists an admissible π such that $(c, \pi) \in A^*(x)$. We first prove the sufficiency. Suppose $(c, \pi) \in A^*(x)$. Then, for any $(c', \pi') \in A(x)$, we have $U(c') \leq U(c)$. We claim that $c \in B^*(x)$. Indeed, otherwise, there exists $c'' \in B(x)$ such that $U(c'') > U(c)$. Using what we proved before, there exists a corresponding π'' such that $(c'', \pi'') \in A(x)$, which results in a contradiction.

Next, slightly override the notation, we prove the necessity. Indeed, suppose $c \in B^*(x)$, for any $c' \in B(x)$, we have $U(c') \leq U(c)$. We have proved previously that there exists π such that $(c, \pi) \in A(x)$. We will show that $(c, \pi) \in A^*(x)$, i.e., for any $(c', \pi') \in A(x)$, we have $U(c') \leq U(c)$. Otherwise, there exists $(c'', \pi'') \in A(x)$, we have $U(c'') > U(c)$. This results in a contradiction since we have proved that $c'' \in B(x)$. $\square$

REMARK 2. Based on (5.4), we have

$$\begin{aligned}\xi_t X_t + \int_0^t \xi_v c_v dv &= x + M_t - M_0 \\ &= x + E_t\left(\int_0^T \xi_v c_v dv\right) - E\left(\int_0^T \xi_v c_v dv\right) \\ &\geq E_t\left(\int_0^T \xi_v c_v dv\right). \qquad (5.5)\end{aligned}$$

Thus, we have

$$\xi_t X_t \geq E_t\left(\int_t^T \xi_v c_v dv\right).$$

This can be equivalently written as

$$E_t\left(\int_t^T \xi_{t,v} c_v dv\right) \leq X_t \qquad (5.6)$$

where

$$\xi_{t,v} = \frac{\xi_v}{\xi_t}.$$

Using the Girsanov change of measure (the conditional case, see Appendix), we have that (5.6) is equivalent to

$$E_t^Q\left(\int_t^T b_{t,v} c_v dv\right) \leq X_t,$$

where $b_{t,v}$ is the discounting factor from time t to v, i.e.,

$$b_{t,v} = \exp\left(-\int_t^v r_s ds\right).$$

This is a time-t analog of the time-zero budget condition (5.2).

5.4 The solution to the static problem

The static problem is

$$\max_c E\left[\int_0^T u(c_t, t)\, dt\right]$$

subject to

$$E\left(\int_0^T \xi_v c_v dv\right) \leq x.$$

The principle of Lagrange multiplier in constraint optimization suggest that $c \in B^*(x)$ iff $(c, y) \in \arg\max L(c, y)$, where

$$L(c, y) = E\left[\int_0^T u(c_t, t)\, dt\right] + y\left[x - E\left(\int_0^T \xi_t c_t dt\right)\right]. \tag{5.7}$$

A heuristic application of the first-order conditions w.r.t. c and y yields that

$$u_c'(c_t, t) = y\xi_t,\ y > 0,\ \text{and } E\left(\int_0^T \xi_v c_v dv\right) = x. \tag{5.8}$$

Thus, we state the following theorem.

THEOREM 2. *Suppose that (c, y) satisfies (5.8), then c is optimal for the static problem. And if c is optimal for the static problem, there exists a $y > 0$ such that (c, y) satisfies (5.8).*

REMARK 3. When the condition (5.8) holds, the last inequality in (5.5) becomes equality, i.e.,

$$\xi_t X_t + \int_0^t \xi_v c_v dv = E_t\left(\int_0^T \xi_v c_v dv\right).$$

This leads to

$$E_t\left(\int_t^T \xi_v c_v dv\right) = \xi_t X_t.$$

This is equivalent to

$$E_t\left(\int_t^T \xi_{t,v} c_v dv\right) = E_t^Q\left(\int_t^T b_{t,v} c_v dv\right) = X_t.$$

That is the price of the optimal portfolio is the present value of future discounted consumption.

Proof. (Necessity) We employ the perturbation theory. Let

$$\begin{aligned} c_t + \varepsilon_1 \Delta c_t,\ \Delta c_t &= c_t' - c_t, \varepsilon_1 > 0, \\ y + \varepsilon_2 \Delta y,\ \Delta y &= y' - y, \varepsilon_2 > 0. \end{aligned}$$

Let $y+\varepsilon_2\Delta y$ be non-negative constant and $c_t+\varepsilon_1\Delta c_t$ be still in $B(x)$. To use tools from classical calculus, we introduce a function

$$f(\varepsilon_1,\varepsilon_2):=L(c_t+\varepsilon_1\Delta c_t,y+\varepsilon_2\Delta y)$$

Then (c,y) is optimal for $L(c,y)$ implies that $(\varepsilon_1,\varepsilon_2)=(0,0)$ is optimal for $f(\varepsilon_1,\varepsilon_2)$. Thus, we have the following first order conditions

$$\frac{\partial L}{\partial\varepsilon_1}|_{\varepsilon=(0,0)}=\frac{\partial L}{\partial\varepsilon_2}|_{\varepsilon=(0,0)}=0.$$

Because

$$\begin{aligned}\frac{\partial L}{\partial\varepsilon_1}|_{\varepsilon=(0,0)} &= E\left[\int_0^T u_c'(c_t,t)\,\Delta c_t dt\right]-yE\left[\int_0^T \xi_t\Delta c_t dt\right]\\ &= E\left[\int_0^T \left(u_c'(c_t,t)-y\xi_t\right)\Delta c_t dt\right]=0.\end{aligned}$$

has to hold for any Δc_t, the only possibility is

$$u_c'(c_t,t)-y\xi_t=0.$$

And, we have

$$\frac{\partial L}{\partial\varepsilon_2}|_{\varepsilon=(0,0)}=\Delta y\left(x-E\left[\int_0^T \xi_t c_t dt\right]\right)=0.$$

This has to hold for any Δy; thus, we have

$$x-E\left[\int_0^T \xi_t c_t dt\right]=0.$$

(Sufficiency) Suppose that (c,y) satisfies

$$\begin{aligned}u_c'(c_t,t) &= y\xi_t, y>0,\\ x &= E\left[\int_0^T \xi_t c_t dt\right].\end{aligned}$$

And suppose that there exists a $c' \in B(x)$, $c' \geq 0$, progressively measurable and integrable, such that $U(c') > U(c)$. From the concavity of u, we know that

$$u(c_t, t) \geq u(c'_t, t) + u'_c(c_t, t)(c_t - c'_t), \forall c_t, c'_t, t \in [0, T], \omega \in \Omega.$$

Integrate with respect to t from 0 to T on both sides and taking expectations, we get

$$E\left[\int_0^T u(c_t, t)\, dt\right] \geq E\left[\int_0^T u(c'_t, t)\, dt\right] + E\left[\int_0^T u'_c(c_t, t)(c_t - c'_t)\, dt\right].$$

This is equivalent to

$$\begin{aligned} U(c) &\geq U(c') + E\left[\int_0^T y\xi_t (c_t - c'_t)\, dt\right] \\ &= U(c') + y\left(x - E\left[\int_0^T \xi_t c'_t dt\right]\right) \geq U(c'), \end{aligned}$$

which result in a contradiction. □

Suppose that u'_c is continuous, decreasing by concavity, satisfying Inada conditions

$$u'_c(0, t) = \infty, \text{ and } u'_c(\infty, t) = 0.$$

Then, there exists unique inverse $I(y, t)$, such that $u'_c(I(y, t), t) = y$. We know from the definition that

$$I(0, t) = \infty, \text{ and } I(\infty, t) = 0.$$

And $I(\cdot, t)$ decreases strictly. So, from $u'_c(c_t, t) = y\xi_t$, we know that

$$c_t = I(y\xi_t, t), \forall t \in [0, T], \omega \in \Omega.$$

So, we can solve for y by

$$x = E\left[\int_0^T \xi_t I(y\xi_t, t)\, dt\right] = \chi(y).$$

The existence of the y can be obtained from the lemma.

LEMMA 1. *There exists a unique $y > 0$, such that $x = \chi(y)$.*

Proof. Since $\chi(\cdot)$ is continuous, decreasing, and $\chi(0) = \infty$, $\chi(\infty) = 0$, for $x > 0$, we can find $y^* = \chi^{-1}(x)$. $\square$

Putting everything together, indeed, we have the following route or solving the optimal consumption problem

$$\left.\begin{array}{r} y^* = \chi^{-1}(x) \Longrightarrow c_t^* = I(y^*\xi_t, t) \\ \Longrightarrow X_t^* = E_t\left(\int_t^T \xi_{t,v} c_v^* dv\right) \\ \text{Martingale representation} \Longrightarrow \phi_t \end{array}\right\} \Longrightarrow \pi_t^* = X_t^* \left(\sigma_t'\right)^{-1} \theta_t + \xi_t^{-1} \left(\sigma_t'\right)^{-1} \phi_t$$

we have proved the following theorem.

THEOREM 3. *We have the following solution to the static problem:*

$$c_t^* = I(y^*\xi_t, t)$$

where

$$y^* = \chi^{-1}(x).$$

We have

$$\pi_t^* = X_t^* \left(\sigma_t'\right)^{-1} \theta_t + \xi_t^{-1} \left(\sigma_t'\right)^{-1} \phi_t,$$

where

$$E_t\left[\int_0^T \xi_t I(y^*\xi_t, t)\, dt\right] - E\left[\int_0^T \xi_t I(y^*\xi_t, t)\, dt\right] = \int_0^t \phi_s \cdot dW_s$$

with

$$\xi_t = b_t \cdot \exp\left(-\int_0^t \theta_v' \cdot dW_v - \frac{1}{2}\int_0^t \theta_v' \cdot \theta_v dv\right)$$

and

$$b_t = \exp\left(-\int_0^t r_v dv\right).$$

Thus, we have

$$X_t^* = E_t\left[\int_t^T \xi_{t,v} c_v^* dv\right] = E_t\left[\int_t^T \xi_{t,v} I\left(y^* \xi_v, v\right) dv\right]$$

and, the value function is given by

$$J_t = E_t\left[\int_t^T u\left(I\left(y^* \xi_v, v\right), v\right) dv\right].$$

The first example is log utility:

$$u\left(c_t, t\right) = a_t \log c_t, a_t > 0.$$

Then:

$$u_c'\left(c_t, t\right) = \frac{a_t}{c_t}.$$

and

$$I\left(y, t\right) = \left(\frac{y}{a_t}\right)^{-1}.$$

So, we have

$$c_t^* = \left(\frac{y^* \xi_t}{a_t}\right)^{-1}$$

and

$$x = E\left[\int_0^T \xi_t \left(\frac{y^* \xi_t}{a_t}\right)^{-1} dt\right].$$

So we get

$$\left(y^*\right)^{-1} = \frac{x}{E\left[\int_0^T a_t dt\right]}.$$

Then, we have

$$X_t^* = E_t\left[\int_t^T \xi_{t,v} \left(\frac{y^* \xi_t \xi_{t,v}}{a_t a_{t,v}}\right)^{-1} dv\right] = \left(\frac{y^* \xi_t}{a_t}\right)^{-1} E_t\left[\int_t^T a_{t,v} dv\right] = c_t^* E_t\left[\int_t^T a_{t,v} dv\right].$$

So, we have

$$c_t^* = X_t^* \cdot m_t,$$

where

$$m_t = \frac{1}{E_t\left[\int_t^T a_{t,v}dv\right]}$$

This is called feedback equation. Derive π_t^* as an exercise. Assume a_t is not stochastic, then we have:

$$J_t = -E_t\left[\int_t^T a_v\left(\log\left(\frac{y}{a_v}\right) + \log\xi_v\right)dv\right].$$

Simplify this expression as an exercise. The second example is CRRA utility function:

$$u\left(c_t, t\right) = a_t\frac{c_t^{1-R}}{1-R}, R > 0.$$

Derive the solution to the static problem as an exercise.

Suppose that $u_c'\left(0,t\right) < \infty$, where the Inada condition does not hold. We have to impose $c_t \geq 0$ for the problem. We can show that the solution is:

$$c_t^* = \max\left(I\left(y^*\xi_t, t\right), 0\right).$$

All the rest remains the same. Do this as an exercise. See Cox and Huang's paper as a hint.

5.5 The explicit representation of the optimal portfolio

With some necessary preparations on Malliavin calculus (see Section 5.7), we establish the following theorem for optimal portfolio consumption choice.

THEOREM 4. *We have the structure of optimal portfolio as*

$$\pi_t^* = \left(\sigma_t'\right)^{-1}\theta_t E_t\left[\int_t^T \xi_{t,v}\frac{c_v^*}{R_v^*}dv\right] - \left(\sigma_t'\right)^{-1}E_t\left[\int_t^T \xi_{t,v}c_v^*\left(1-\frac{1}{R_v^*}\right)H_{t,v}dv\right],$$

where

$$\xi_{t,v} = \frac{\xi_v}{\xi_t} \text{ with } \xi_t = \exp\left(-\int_0^t \theta_s' dW_s - \frac{1}{2}\int_0^t \theta_s'\theta_s ds - \int_0^t r_s ds\right),$$

and

$$c_v^* = I\left(y^* \xi_v, v\right),$$

and

$$R_v^* = R\left(c_v^*\right),$$

with

$$R\left(c\right) := -\frac{u_c''\left(c,t\right)}{u_c'\left(c,t\right)} c.$$

Here, y^* solves the equation

$$x = E\left[\int_0^T \xi_v I\left(y \xi_v, v\right) dv\right].$$

Moreover,

$$H_{t,v}' = \int_t^v D_t r_s ds + \int_t^v \left(dW_s' + \theta_s' ds\right) D_t \theta_s.$$

Here, $H_{t,v}$ is a $d \times 1$ vector; $D_t r_s$ is a $1 \times d$ vector; $D_t \theta_s$ is a $d \times d$ matrix.

Under a general model, we expressed the optimal portfolio weight in an explicit expectation form. Essentially, we give an explicit expectation expression of the solution to a very general HJB equation for the optimal portfolio problem. This is analogy to the Feymann-Kac representation for solutions to linear parabolic PDEs. This is the merit of this result. To convince yourself the power of the martingale approach, I suggest to follow these expectation expression to further obtain closed-form formulas under some certain models, e.g., the Merton model (multidimensional Black-Scholes). Then, check the agreement on the solution obtained from the HJB approach. For some other “analytically tractable” models, we can carry out similar calculations to explicitly compute the expectations. For example, to consider mean-reversion of returns and affine stochastic volatility and interest rate, etc.

If we define relative risk aversion as

$$R\left(c\right) = -\frac{u_c''\left(c,t\right)}{u_c'\left(c,t\right)} c = R^A\left(c\right) c,$$

where $R^A(c)$ is absolute risk aversion. And define the absolute risk tolerance as

$$\Gamma(c) = \frac{1}{R^A(c)} = \frac{c}{R(c)}.$$

The term

$$E_t\left[\int_t^T \xi_{t,v}\frac{c_v^*}{R_v^*}dv\right] = E_t\left[\int_t^T \xi_{t,v}\Gamma(c_v^*)\,dv\right]$$

is the cost at t, the present value of future risk tolerance. The term

$$E_t\left[\int_t^T \xi_{t,v}\left(c_v^* - \Gamma(c_v^*)\right)H_{t,v}dv\right]$$

is the cost at time t of a contingent claim that pays $\left(c_v^* - \Gamma(c_v^*)\right)H_{t,v}$ at time v continuously.

We already know that

$$\xi_t X_t^* = E_t\left[\int_t^T \xi_v c_v^* dv\right],$$

and

$$\pi_t^* = X_t^*\left[\frac{(\sigma_t')^{-1}\theta_t E_t\left[\int_t^T \xi_v \frac{c_v^*}{R_v^*}dv\right]}{E_t\left[\int_t^T \xi_v c_v^* dv\right]} - \frac{(\sigma_t')^{-1}\cdot E_t\left[\int_t^T \xi_{t,v}c_v^*\left(1-\frac{1}{R_v^*}\right)H_{t,v}dv\right]}{E_t\left[\int_t^T \xi_v c_v^* dv\right]}\right].$$

The portfolio weight becomes π_t^*/X_t^*. Denote by

$$\rho_{t,v} = \frac{\xi_v c_v^*}{E_t\left[\int_t^T \xi_v c_v^* dv\right]}.$$

By plugging in this expression, we can clean up the formula further, i.e.,

$$\pi_t^* = X_t^*\left[(\sigma_t')^{-1}\theta_t E_t\left(\int_t^T \frac{\rho_{t,v}}{R_v^*}dv\right) - \xi_t^{-1}(\sigma_t')^{-1}E_t\left[\int_t^T \rho_{t,v}\left(1-\frac{1}{R_v^*}\right)H_{t,v}dv\right]\right].$$

We can also have the portfolio decomposition formula:

$$\pi_t^* = \pi_t^m + \pi_t^h,$$

where

$$\pi_t^m = \left(\sigma_t'\right)^{-1} \theta_t E_t \left[\int_t^T \xi_{t,v} \frac{c_v^*}{R_v^*} dv\right].$$

This is exactly the Markowitz mean variance portfolio demand called diversification demand component. And, we have

$$\pi_t^h = \pi_t^{hR} + \pi_t^{h\theta},$$

where

$$\begin{aligned}
\pi_t^{hR} &= -\left(\sigma_t'\right)^{-1} E_t \left[\int_t^T \xi_{t,v} c_v^* \left(1 - \frac{1}{R_v^*}\right) H_{t,v}^R dv\right], \\
\pi_t^{h\theta} &= -\left(\sigma_t'\right)^{-1} E_t \left[\int_t^T \xi_{t,v} c_v^* \left(1 - \frac{1}{R_v^*}\right) H_{t,v}^\theta dv\right].
\end{aligned}$$

The term π_t^{hR} correspond to the interest rate hedge and the term $\pi_t^{h\theta}$ denote the market price of risk hedge. And:

$$\begin{aligned}
H_{t,v}^R &= \int_t^v D_t r_s ds, \\
H_{t,v}^\theta &= \int_t^v \left(dW_s' + \theta_s' ds\right) D_t \theta_s,
\end{aligned}$$

Using the Bayes formula (change of measure for condtional expectations, we obtain the following expression of the optimal portfolio weight under the measure Q, i.e.,

$$\pi_t^* = \left(\sigma_t'\right)^{-1} \theta_t E_t^Q \left[\int_t^T b_{t,v} \frac{c_v^*}{R_v^*} dv\right] - \left(\sigma_t'\right)^{-1} E_t^Q \left[\int_t^T b_{t,v} c_v^* \left(1 - \frac{1}{R_v^*}\right) H_{t,v} dv\right],$$

with

$$H_{t,v}' = \int_t^v D_t r_s ds + \int_t^v dW_s'^Q D_t \theta_s.$$

This formula looks much simpler and thus easier to implement numerically.

Proof. According to previous discussions, we have the optimal portfolio weight as

$$\pi_t^* = X_t^* \left(\sigma_t'\right)^{-1} \theta_t + \xi_t^{-1} \left(\sigma_t'\right)^{-1} \phi_t^*.$$

Here ϕ_t satisfies

$$M_t^* - M_0^* = \int_0^t \phi_v^{*\prime} dW_v$$

where

$$M_t^* = E_t \left[\int_0^T \xi_v I\left(y^* \xi_v, v\right) dv \right].$$

Using the Clark-Ocone formula, we have

$$\phi_v^{*\prime} = E_v \left[D_v \left(\int_0^T \xi_s I\left(y^* \xi_s, s\right) ds \right) \right].$$

Using the chain rule of Malliavin calculus, this is equal to

$$\phi_v^{*\prime} = E_v \left[\int_v^T \left(I\left(y^* \xi_s, s\right) + y^* \xi_s I'\left(y^* \xi_s, s\right) \right) D_v \xi_s ds \right].$$

Because

$$I\left(y^* \xi_s, s\right) = c_s^*,$$

and

$$y^* \xi_s = u_c'\left(c_s^*, s\right),$$

and

$$I'\left(y^* \xi_s, s\right) = \frac{1}{u_c''\left(I\left(y^* \xi_s, s\right), s\right)} = \frac{1}{u_c''\left(c_s^*, s\right)},$$

we have

$$I\left(y^* \xi_s, s\right) + y^* \xi_s I'\left(y^* \xi_s, s\right) = c_s^* \left(1 - \frac{1}{R_s^*}\right).$$

And, we have

$$D_v \xi_s = -\xi_s \left(\theta_v' + \int_v^s dW_u' D_v \theta_u + \int_v^s \theta_u' D_v \theta_u du + \int_v^s D_v r_u du \right) = -\xi_s \left(\theta_v' + H_{v,s}' \right).$$

Thus, we have

$$
\begin{aligned}
\phi_v^{*\prime} &= -E_v\left[\int_v^T c_s^*\left(1-\frac{1}{R_s^*}\right)\xi_s\left(\theta_v'+H_{v,s}'\right)ds\right] \\
&= -\theta_v' E_v\left[\int_v^T c_s^*\left(1-\frac{1}{R_s^*}\right)\xi_s ds\right] - E_v\left[\int_v^T c_s^*\left(1-\frac{1}{R_s^*}\right)\xi_s H_{v,s}' ds\right].
\end{aligned}
$$

And, we have

$$
X_t^* = E_t\left[\int_t^T \xi_{t,v} c_v^* dv\right]
$$

Collecting terms, we get the final answer. □

COROLLARY 1. If we have deterministic opportunity set: (r, θ) are only functions of time, then $H \equiv 0$ and $\pi_t^h \equiv 0$. There is no hedge at all, the optimal portfolio is Markowitz one.

COROLLARY 2. If we have stochastic opportunity set (r, θ), but if the utility function is log utility, then there is still no hedge.

REMARK 4. log utility is called myopic utility function, because $R(c) \equiv 1$ and thus $\pi_t^h \equiv 0$.

5.6 Excellent exercises

Consider the following very helpful exercises.

1. Develop the above framework for the bequest case. Suppose the investor wants to maximize his expected utility over the terminal wealth by allocating his wealth in the market following a dynamic portfolio policy. Then the problem can be formulated as

$$
\max_{(\pi_t, X_T)} E\left[U(X_T)\right],
$$

subject to

$$
\begin{aligned}
& dX_t = r_t X_t dt + X_t \pi_t^\top\left[(\mu_t - r_t \mathbf{1}_m)dt + \sigma_t dW_t\right],\ X_0 = x, \\
& X_t \geq 0,\ \text{for all } t \in [0, T].
\end{aligned}
$$

2. Develop the above framework for the consumption-bequest case. Suppose the investor wants to maximize his expected utility over both intermediate consumption and terminal wealth by allocating his wealth in the market following a dynamic portfolio policy. Then the problem can be formulated as

$$\max_{(\pi_t, c_t, X_T)} E\left[\int_0^T u(c_t, t)dt + U(X_T, T)\right],$$

subject to

$$\begin{aligned} & dX_t = r_t X_t dt + X_t \pi_t^\top \left[(\mu_t - r_t \mathbf{1}_m)dt + \sigma_t dW_t\right] - c_t dt, \ X_0 = x, \\ & X_t \geq 0, \ \text{for all } t \in [0, T]. \end{aligned}$$

3. Following the frameworks, to calculate closed-form optimal portfolio choice under some simple models, e.g., to solve the Merton problem under the geometric Brownian motion model. A closed-form solution was proposed by Merton. So, one should be able to obtain the same formula by calculating the expectations involved in our framework. Once you are able to complete this exercise, you fully master a lot of contents.

5.7 An introduction to Malliavin Calculus: a useful tool

Malliavin calculus is a calculus of variations for stochastic processes.

- Development of theory:
  - Malliavin, Stroock, Bismut,...
  - Originally motiated by "under what conditions the distribution of a random variable is smooth?", by Paul Malliavin, in 1970s.
  - Existence and smoothness of densities
- A new calculus that extended Newton's into the realm of the random
- Applies to Brownian functionals: random variables and stochastic processes that depend on trajectories of Brownian motion

- Malliavin derivative measures impact of small change in trajectory of Brownian motion on value of Brownian functional. How the distribution changes as the path of the process change? This is a very natural question to consider, with many potential applications, e.g., asymptotic expansions.

- Main text on both theory and applications can be found in, e.g., [2], [10], [17], [20], and [18].

- It has a fearsome reputation for being heavily technical. "It is still thought of as more complicated than it actually is, ..., Some of the early presentations were quite scary and technical, so people thought the costs would outweigh the benefits" — Professor Bernt Oksendal, University of Oslo.

- Applications in financial economics: sensitivity simulations, asymptotic expansions, and portfolio choice problems, etc.

- By inspring new works, the financial legacy of Malliavin calculus may turn out to be greater than some once ever thought!

Outline

- Definition

- Riemann, Wiener and Ito integrals

- Clark-Ocone formula

- Chain rule

- Stochastic differential equations

5.7.1 Definition

Smooth Brownian functionals

- Space of (smooth) functions: $C_p^\infty(R^{nd})$

  - $f(\cdot) : R^{nd} \to R$

  - Infinitely differentiable

- Polynomial growth

- Wiener space generated by d-dimensional Brownian motion $W = (W_1, \cdots, W_d)'$

  - A space of trajectories: an infinite dimensional space
  - Each state of nature corresponds to a trajectory of BM
  - Set of states is space of trajectories

- Let $(t_1, ..., t_n)$ be a partition of $[0, T]$

  - Sample BM at points of this partition: $(W_{t_1}, \cdots, W_{t_n})$
  - Construct random variable $F(W) \equiv f(W_{t_1}, \cdots, W_{t_n})$, where $f \in C_p^\infty(R^{nd})$
  - f has n arguments, each of which has d components.
  - F is smooth Brownian functional
  - By doing so, we use finite dimensional property to characterize and approximate infinite dimensional property.

Examples: assume W is one-dimensional

- Qaudratic function: $W_T^2, \sum_{j=1}^n W_{t_j}^2$

- Any polynomial: $\sum_{k=1}^K a_k W_T^k$, $\sum_{j=1}^n (\sum_{k=1}^K a_k W_{t_j}^k)$

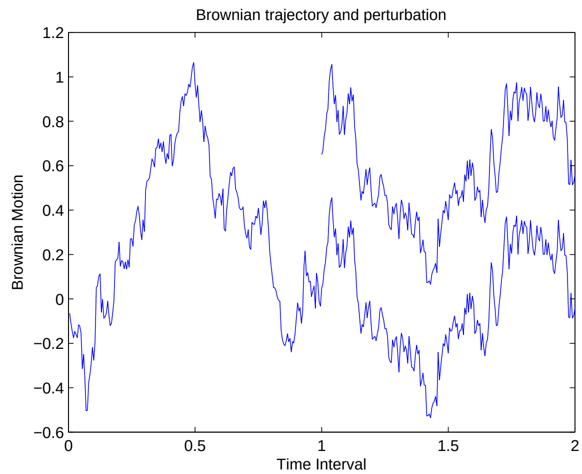
- Stock price in Black-Scholes model: (limit of sequence of SBF (Smooth Brownian Functional))

  - Geometric Brownian motion:

$$S_T = S_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right) T + \sigma W_T\right)$$

  - Write $S_T = f(W_T)$ with $f(x) = S_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right) T + \sigma x\right)$, the limit of polynomials
  - S_T is (limit of) smooth Brownian functional (sampled at one point)

Figure 5.1: Perturbation of a Brownian Trajectory



Let us conduct the following experiment:

- Analog: for $x \in R^d$, perturb x to $x + \Delta x$
- Start from a one-dimensional case, i.e., consider a one-dimensional Brownian motion.
- Perturbate trajectory of BM from some time t onward
- Shift W by ε starting at t, where $t_k \leq t < t_{k+1}$ for some $k = 1, ..., n$
  - $W_s \to W_s + \varepsilon 1_{[t,\infty[}(s)$, i.e.,

$$\begin{cases} s < t, W_s \to W_s \text{ (no effect before t)} \\ s \geq t, W_s \to W_s + \varepsilon \text{(Perturbation at t by } \varepsilon) \end{cases}$$

    and $W_{t_k} \to W_{t_k} + \varepsilon 1_{[t,\infty[}(t_k)$
  - where $1_{[t,\infty[}$ is indicator of $[t, \infty)$ (i.e., $1_{[t,\infty[}(s) = 1$ for $s \in [t, \infty); 0$ otherwise)

Malliavin derivative of smooth Brownian functional (assume $d = 1$)

- MD at t of F is change in F due to a change in path of W starting at t.

- MD of F at t is defined by[‡]

$$
\begin{aligned}
D_t F(W) &= \lim_{\varepsilon \to 0} \frac{F(W + \varepsilon 1_{[t,\infty[}) - F(W)}{\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{f(W_{t_1} + \varepsilon 1_{[t,\infty[}(t_1), ..., W_{t_n} + \varepsilon 1_{[t,\infty[}(t_n)) - f(W_{t_1}, ..., W_{t_n})}{\varepsilon} \\
&\equiv \left. \frac{\partial f(W_{t_1} + \varepsilon 1_{[t,\infty[}(t_1), ..., W_{t_n} + \varepsilon 1_{[t,\infty[}(t_n))}{\partial \varepsilon} \right|_{\varepsilon=0}.
\end{aligned}
$$

- By using the chain rule of calculus, we obtain

$$
D_t F(W) = \sum_{j=1}^{n} \partial_j f(W_{t_1}, ..., W_{t_k}, ..., W_{t_n}) 1_{[t,\infty[}(t_j) \tag{5.9}
$$

where $\partial_j f$ is derivative of f with respect to j^{th} argument of f.

- MD of F is

$$
DF(W) = \{D_t F(W) : t \in [0, T]\}
$$

  - Stochastic process
  - Not adapted: depends on future values of BM

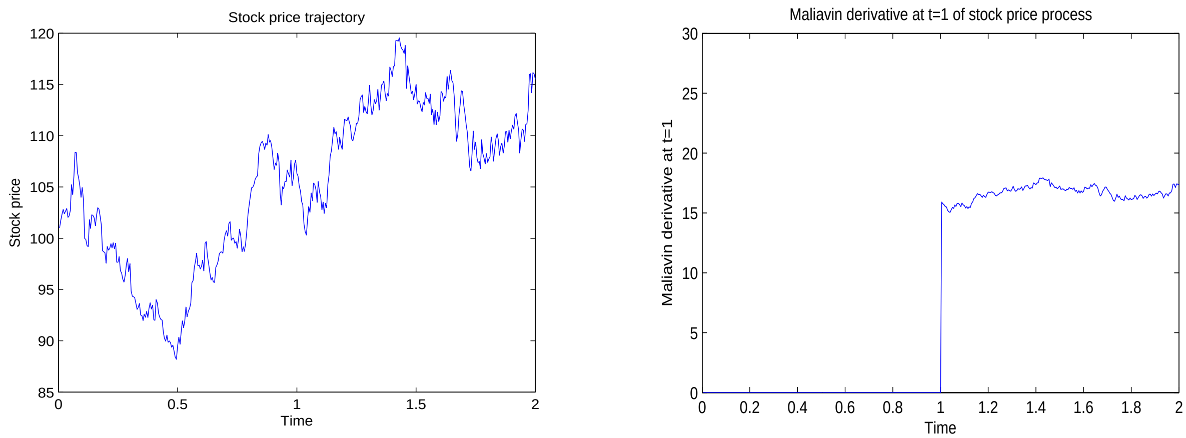
Example 1. Black-Scholes model

- Recall $S_T = f(W_T)$ with $f(x) = S_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right) T + \sigma x\right)$
- Direct application of the formula (5.9) gives (a case with $d = 1$ and $n = 1$)

$$
\begin{aligned}
D_t S_T &= \partial f(W_T) \mathbf{1}_{[t,\infty[}(T) \\
&= \sigma S_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right) T + \sigma W_T\right) \mathbf{1}_{[t,\infty[}(T) = \sigma S_T \mathbf{1}_{[t,\infty[}(T).
\end{aligned}
$$

- $D_t S_T = 0$, if $t > T$.

[‡] Indeed, we can consider perturbation with respect to other general functions in time. The resulting definition is consistent with our current proposal and can be naturally regarded as a generalization, see, e.g., Section 5.7.6.



(a) Before taking Malliavin derivative

(b) After taking Malliavin derivative

- Malliavin derivative is derivative with respect to W_T:
  - Perturbation of path of W from t onward affects S_T only through W_T
  - Analog to derivaives in R^d
- Malliavin derivative for S_v :
  - Double indexed stochastic process $D_t S_v = \sigma S_v 1_{[t,\infty[}(v)$
  - $D_t S_v$ not $\mathcal{F}_t$-measurable, but $\mathcal{F}_v$-measurable

Multidimensional case: $d > 1$

- Malliavin derivative of F at t is now $1 \times d$- dimensional vector

$$D_t F = (D_{1t} F, ..., D_{dt} F)$$

- The ith coordinate $D_{it}F$ measures impact of perturbation in W_i by ε starting at t, i.e.,

$$\begin{aligned}
& D_{it}F(W) \\
= \;& \lim_{\varepsilon \to 0} \frac{F(W + (0, \cdots, 0, \varepsilon 1_{[t,\infty[}, 0, \cdots, 0)) - F(W)}{\varepsilon} \\
= \;& \lim_{\varepsilon \to 0} \frac{f(W_{t_1} + (0, \cdots, 0, \varepsilon 1_{[t,\infty[}, 0, \cdots, 0), ..., W_{t_n} + (0, \cdots, 0, \varepsilon 1_{[t,\infty[}, 0, \cdots, 0)) - f(W_{t_1}, ..., W_{t_n}))}{\varepsilon} \\
\equiv \;& \left. \frac{\partial f(W_{t_1} + (0, \cdots, 0, \varepsilon 1_{[t,\infty[}, 0, \cdots, 0), ..., W_{t_n} + (0, \cdots, 0, \varepsilon 1_{[t,\infty[}, 0, \cdots, 0))}{\partial \varepsilon} \right|_{\varepsilon = 0}
\end{aligned}$$

- (Exercise) We can write one-dimensional definition for this derivative

$$D_{it}F = \sum_{j=1}^{n} \frac{\partial f}{\partial x_{ij}}(W_{t_1}, ..., W_{t_k}, ..., W_{t_n}) \mathbf{1}_{[t,\infty[}(t_j)$$

  which can be simplified to

$$D_{it}F = \sum_{j=k}^{n} \frac{\partial f}{\partial x_{ij}}(W_{t_1}, ..., W_{t_k}, ..., W_{t_n}) \mathbf{1}_{[t,\infty[}(t_j)$$

  for $t_k \leq t < t_{k+1}$.

  - where $\partial f / \partial x_{ij}$ is derivative with respect to i^{th} component of j^{th} argument of f (i.e. derivative with respect to W_{it_j})

- MD of F is $DF(W) = \{D_t F(W) : t \in [0, T]\}$; d-dimensional (row) stoch. proc.

Domain of Malliavin derivative operator

- Defined for smooth Brownian functionals in straightforward way
- Works for limits of smooth Brownian functionals
  - Extends to functionals of path of BM over continuous interval $[0, T]$
  - Path-dependent functional can be approximated by sequence of SBF
  - MD of path-dependent functional is limit of MD of SBF in approximating sequence

– Space of random variables for which Malliavin derivatives defined is $D^{1,2}$

– Refer to the following diagram:

$$\begin{array}{ccc} F & \longleftarrow & \text{SBF} \\ \| & & \downarrow \\ DF & \longleftarrow & D\text{SBF} \end{array}$$

– This is completion of set of smooth Brownian functionals in norm

$$\|F\|_{1,2} = \left(E(F^2) + \mathbf{E}\left[\int_0^T \|D_t F\|^2 \, dt \right] \right)^{\frac{1}{2}}.$$

where $\|D_t F\|^2 = \sum_i (D_{it} F)^2$.

5.7.2 Derivatives of Wiener, Riemann & Ito integrals (some useful properties)

Wiener Integral

- $F(W) = \int_0^T h(s) dW_s$, where $h(t)$ is fct of time and W is one-dim. Can we guess what is $D_t F(W)$?
- Malliavin derivative:

$$\begin{aligned} D_t F(W) &= \lim_{\varepsilon \to 0} \frac{F(W + \varepsilon 1_{[t,\infty[}) - F(W)}{\varepsilon} \\ &= \lim_{\varepsilon \to 0} \frac{\int_0^T h(s) d(W_s + \varepsilon 1_{[t,\infty[}(s)) + \int_0^T h(s) dW_s}{\varepsilon} \\ &= \lim_{\varepsilon \to 0} \frac{\int_0^T h(s) d(\varepsilon 1_{[t,\infty[}(s))}{\varepsilon} \\ &= \lim_{\varepsilon \to 0} \frac{\int_0^T h(s) \delta(s-t) ds \varepsilon}{\varepsilon} = h(t) 1_{\{t \le T\}} \end{aligned}$$

- Here, $1_{[t,\infty[}(s)$ is a Heaviside function; $\delta(s-t)$ is the Dirac Delta function centered at t.
- MD of F at t is volatility $h(t)$ of stochastic integral at t (the perturbation time) for $0 \le t \le T$. This is a very natural result!

Random Riemann integral with integrand depending on path of BM

- $F(W) \equiv \int_0^T h_s(W)ds$ where h progressively measurable

- Malliavin derivative:

$$\begin{aligned}
D_t F &= \lim_{\varepsilon \to 0} \frac{F(W + \varepsilon 1_{[t,\infty[}) - F(W)}{\varepsilon} \\
&= \frac{1}{\varepsilon}\left[\int_0^T h_s(W + \varepsilon 1_{[0,+\infty)})ds - \int_0^T h_s(W)ds\right] \\
&= \int_0^T \frac{1}{\varepsilon}(h_s(W + \varepsilon 1_{[t,+\infty)}) - h_s(W))ds \\
&= \int_0^T D_t h_s ds = \int_t^T D_t h_s ds,
\end{aligned}$$

  because $D_t h_s = 0$ for $s < t$ (no effect prior to the perturbation time).

Ito integral

- $F(W) = \int_0^T h_s(W)dW(s)$ where h progressively measurable

- Malliavin derivative:

$$D_t F = \lim_{\varepsilon \to 0} \frac{F(W + \varepsilon \mathbf{1}_{[t,\infty[}) - F(W)}{\varepsilon} = \int_t^T D_t h_s dW_s + h_t$$

Malliavin derivatives of Riemann, Ito integrals depending on multi-dimensional BM defined in same way, but component by component

Example 2. $F = \int_0^T r_s ds$ where r is the interest rate $\to D_t F = \int_t^T D_t r_s ds$.

Example 3. $F = \int_0^T \theta_v dW_v$, $D_t F = \int_t^T D_t \theta_v dW_v + \theta_t$

5.7.3 Clark-Ocone formula

Motivation

- Martingale representation theorem
  - Wiener spaces: martingales with finite variance are weighted sums of BM increments

–
$$M_t = M_0 + \int_0^t \phi_s dW_s$$

for some progressively measurable process ϕ

– ϕ represents volatility coefficient of martingale

- Malliavin calculus identifies integrand ϕ (Clark-Ocone formula)

Clark-Ocone formula

- Any $\mathcal{F}_T$ measurable random variable $F \in D^{1,2}$ can be decomposed as

$$F = E[F] + \int_0^T E_t[D_t F] dW_t \tag{5.10}$$

- $D_t F$ is $1 \times d$ dimensional and is $\mathcal{F}_T$ measurable, not $\mathcal{F}_t$ measurable. But, $E_t[D_t F]$ is always $\mathcal{F}_t$ measurable. So, the integral $\int_0^T E_t[D_t F] dW_t$ is an Ito stochastic integral, which is a martingale.

- How to obtain the martingale representation? Indeed, let $F = M_T$ in (5.10), we obtain that

$$M_T = M_0 + \int_0^T E_t[D_t M_T] dW_t.$$

- Implication: consider martingale closed by $F \in D^{1,2}$ (i.e. $M_t = E_t[F]$)

  – Take conditional expectations!

  – We use (5.10) and the martingale property of stochastic integral to obtain

$$M_t = E[F] + E_t\left(\int_0^T E_s[D_s F] dW_s\right) = E[F] + \int_0^t E_s[D_s F] dW_s.$$

- An informal derivation

  – Assume $F \in D^{1,2}$

- We consider the case with $d = 1$ first. Martingale representation theorem gives $F = E[F] + \int_0^T \phi_s dW_s$.
- Take MD on each side & simplify $D_t F = \phi_t + \int_t^T D_t \phi_s dW_s$
- Take conditional expectations on each side, we have $E_t[D_t F] = \phi_t$.
- Now, for multidimensional case, we just do the similar thing!
- Martingale representation theorem gives $F = E[F] + \int_0^T \phi_s dW_s$, where ϕ_s is $1 \times d$ dimensional
- (Exercise) By definition, we try to find

$$D_{it} \int_0^T \phi_s dW_s = ?$$

Guess

$$D_{it} \int_0^T \phi_s dW_s = \phi_{it} + ?$$

Then, we have

$$\begin{aligned} D_t F &= (D_{1t} F, D_{2t} F, \cdots, D_{dt} F) \\ &= \phi_t + ??? \end{aligned}$$

Taking conditional expectation, we obtain $E_t[D_t F] = \phi_t$.

Remark (commutativity)

- Results can be used to show that MD and conditional expectation commute
- For martingale $M_v = E_v[F]$, we have

$$M_v = E_v[F] = E[F] + \int_0^v E_s[D_s F] dW_s.$$

- Thus, the Malliavin derivatives of M_v is

$$D_t M_v = D_t E_v[F] = E_v[D_t F].$$

5.7.4 Chain rule of Malliavin calculus

In applications often need MD of function of path-dependent random variable

- Chain rule also applies in Malliavin calculus

Let $G = g(F)$ where

- $F = (F_1, ..., F_n)$ is vector of random variables in $D^{1,2}$
- g is a differentiable function of F with bounded derivatives
- Malliavin derivatie of $G = g(F)$ is

$$D_t G = D_t g(F) = \sum_{i=1}^{n} \frac{\partial g}{\partial x_i}(F) D_t F_i$$

  where $\frac{\partial g}{\partial x_i}(F)$ is derivative relative to the ith argument of ϕ.

Example 4. $F = \int_0^T r(Y_{1s}, ..., Y_{ns}) ds \rightarrow D_t F = \int_0^T \sum_{i=1}^{n} \frac{\partial r}{\partial y_i} D_t Y_{is} ds$

Exercise: Let $\xi_t = b_t \eta_t$, where

$$b_t = e^{-\int_0^t r_s ds} \text{ and } \eta_t = \exp\left(-\int_0^t \theta_v dW_v - \frac{1}{2}\int_0^t \theta_v^2 dv\right).$$

Find $D_s b_t, D_s \eta_t$, and $D\xi_t$.

5.7.5 Stochastic differential equations

For portfolio models need to handle diffusions (SDEs)

- What is Malliavin derivative of solution of SDE?
- Rules described above can be used for that purpose

Suppose state variable Y_t follows diffusion process

- $dY_t = \mu^Y(Y_t, t)dt + \sigma^Y(Y_t, t)dW_t$ where Y_0 given; W one dimensional
  - Assumptions:
    - Coefficients satisfy Lipschitz and Growth conditions
    - Smoothness: coefficients are continuously differentiable w.r.t. y
  - Integral form

$$Y_s = Y_0 + \int_0^s \mu^Y(Y_v, v)dv + \int_0^s \sigma^Y(Y_v, v)dW_v$$

Malliavin derivative:

- Apply rules of Malliavin calculus and take MD on each side
- For $s \geq t$

$$\begin{aligned} D_t Y_s &= D_t \left(Y_0 + \int_0^s \mu^Y(Y_v, v)dv + \int_0^s \sigma^Y(Y_v, v)dW_v \right) \\ &= D_t Y_0 + \int_t^s \partial\mu^Y D_t Y_v dv + \int_t^s \partial\sigma^Y D_t Y_v dW_v + \sigma^Y(Y_t, t) \\ &= \int_t^s \partial\mu^Y D_t Y_v dv + \int_t^s \partial\sigma^Y D_t Y_v dW_v + \sigma^Y(Y_t, t) \end{aligned}$$

  where second equality follows from $D_t Y_0 = 0$

- Conclusion: MD follows linear SDE

$$d(D_t Y_s) = \left[\partial\mu^Y(Y_s, s)ds + \partial\sigma^Y(Y_s, s)dW_s\right](D_t Y_s)$$

  subject to initial condition

$$\lim_{s \to t} D_t Y_s = \sigma^Y(Y_t), \text{ for } s \geq t.$$

  Here, t is a parameter; s is the time variable.

Solution to this linear SDE:

$$\begin{aligned} D_t Y_s &= D_t Y_t \exp\left(\int_t^s (\partial\mu^Y(Y_v, v) - \frac{1}{2}(\partial\sigma^Y(Y_v, v))^2)dv + \int_t^s \partial\sigma^Y(Y_v, v)dW_v\right) \\ &= \sigma^Y(Y_t, t)\exp\left(\int_t^s (\partial\mu^Y(Y_v, v) - \frac{1}{2}(\partial\sigma^Y(Y_v, v))^2)dv + \int_t^s \partial\sigma^Y(Y_v, v)dW_v\right) \end{aligned}$$

- Explicit formula: MD as a function of trajectory of W, Y
- Analog: solution to the Black-Scholes model
- Computation:
  - Method 1: Monte Carlo simulation of (W, Y)+ explicit solution
  - Method 2: Monte Carlo simulation of linear SDE (6)
  - Other methods

Example 5. Vascek

$$\begin{aligned} dr_t &= k(\bar{r} - r_t)dt + \sigma dW_t \\ d(D_t r_v) &= -kD_t r_v dv; D_t r_t = \sigma \\ D_t r_v &= \sigma \exp(-k(v - t)) \end{aligned}$$

Multidimentional case:

- Y_t is a k dimensional vector diffusion process
- SDE:

$$dY_t = \mu^Y(Y_t, t)dt + \sigma^Y(Y_{t,}t)dW_t$$

  where Y_0 given

  - W d-dimensional vector; $\mu^Y(Y_t, t)$ k-dimensional vector

- $\sigma^Y(Y_t,t)$ $k \times d$-dimensional matrix

- Assumptions:
  - Coefficients satisfy Lipschitz and Growth conditions
  - Smoothness: coefficients are continuously differentiable

- Definte Malliavin calculus for a random vector as follows:

$$D_tY_s = \begin{pmatrix} D_tY_{1v} \\ D_tY_{2v} \\ \cdot \\ D_tY_{kv} \end{pmatrix} = \begin{pmatrix} (D_{1t}Y_{1v}, D_{2t}Y_{1v}, \cdots, D_{dt}Y_{1v}) \\ (D_{1t}Y_{2v}, D_{2t}Y_{2v}, \cdots, D_{dt}Y_{2v}) \\ \vdots \\ (D_{1t}Y_{kv}, D_{2t}Y_{kv}, \cdots, D_{dt}Y_{kv}) \end{pmatrix}$$

- Similar to the one-dimensional case, we have

$$\begin{aligned} dD_tY_s &= D_t\mu^Y(Y_s,s)ds + D_t(\sigma^Y(Y_s,s)dW_s) \\ &= \partial_y\mu^Y(Y_s,s)D_tY_sds + D_t\left(\sum_j \sigma^Y_{\cdot,j}(Y_s,s)dW_{js}\right) \\ &= \partial_y\mu^Y(Y_s,s)D_tY_sds + \left(\sum_j \partial_y\sigma^Y_{\cdot,j}(Y_s,s)D_tY_sdW_{js}\right), \end{aligned} \tag{5.11}$$

with initial condition

$$\lim_{s\to t} D_tY_s = \sigma^Y(Y_t,t), \text{ for } s \geq t.$$

- Here, we have

$$\partial_y\mu^Y(y,s) = \begin{pmatrix} \partial_{y_1}\mu^Y_1(y,s) & \partial_{y_2}\mu^Y_1(y,s) & \cdots & \partial_{y_k}\mu^Y_1(y,s) \\ \partial_{y_1}\mu^Y_2(y,s) & \partial_{y_2}\mu^Y_2(y,s) & \cdots & \partial_{y_k}\mu^Y_2(y,s) \\ \vdots & \vdots & \vdots & \vdots \\ \partial_{y_1}\mu^Y_k(y,s) & \partial_{y_2}\mu^Y_k(y,s) & \cdots & \partial_{y_k}\mu^Y_k(y,s) \end{pmatrix}_{k\times k}$$

- By writing $\sigma^Y(Y_s, s)dW_s$ in a column vector form $\sum_j \sigma^Y_{\cdot,j}(Y_s, s)dW_{js}$ explicitly, the Malliavin differentiation follows in a similar as it is for the drift.

  - System of linear SDEs - coupled
  - Computation: Monte Carlo simulation

- An Excellent Exercise: since the essence of Malliavin calculus can be seen from the one-dimensional case ($d = 1$) for scalar variable ($k = 1$), it is natural to write the element of the SDE system and take Malliavin derivaives. Follow this route to check (5.11) is indeed correct!

- **Example 6.** $\begin{cases} dr_t = k(\overline{r_t} - r_t)dt + \sigma dW_t \\ d\overline{r_t} = \alpha(\overline{r_t})dt + \gamma(\overline{\eta})dW_t \end{cases}$

5.7.6 A formal definition

In this section, we give a formal definition of Malliavin derivative. This can be regarded as a consistent and more general version of all defined before. Let $(\Omega, P, \mathcal{F}, \{\mathcal{F}(t)\})$ denote the d-dimensional filtered Wiener space, where $\Omega = \mathcal{C}_0([0, T], \mathbb{R}^d)$. The coordinate process $\{w(t)\}$ is a d-dimensional Brownian motion under the Wiener measure $\mathbb{P}$. Let H be the Cameron-Martin subspace of Ω, i.e.

$$H = \left\{ h = \left(\int_0^{\cdot} \dot{h}^1(s)ds, ..., \int_0^{\cdot} \dot{h}^d(s)ds \right); \quad \dot{h} \in L^2[0, T] \right\}.$$

The inner product of the Hilbert space H is defined as $\langle h_1, h_2 \rangle_H = \sum_{k=1}^d \int_0^T \dot{h}_1^k(s)\dot{h}_2^k(s)ds$, for any $h_1, h_2 \in H$. Thus, the norm is equipped with $||h||_H = \left(\sum_{k=1}^d \int_0^T |\dot{h}^k(t)|^2 dt \right)^{\frac{1}{2}}$, for $h \in H$. Let $F : \Omega \to R$ be an $\mathcal{F}(T)-$measurable random variable, which is also called Wiener functional. For $F \in L^p(\Omega)$, where $p > 1$, we define the directional derivative of F along $h \in H$ as

$$D_h F(w) := \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} F(w + \epsilon h) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[F(w + \epsilon h) - F(w) \right]. \tag{5.12}$$

Thus, $D.F(w)$ is defined as a linear functional on Hilbert space H. By the Riesz representation theory, there exists an element $D_sF(w) := (D^1F(w), D^2F(w), ..., D^dF(w)) \in H$, such that

$$D_hF(w) = \langle \dot{h}, DF(w) \rangle_H = \sum_{k=1}^{d} \int_0^T \dot{h}^k(s) D_s^k F(w) ds.$$

Let $L^p(\Omega : H)$ denote the collection of measurable maps f from Ω to H such that $||f||_H \in L^p(\Omega)$. If $DF \in L^p(\Omega : H)$, DF is defined as the Malliavin derivative of F. The analog of Malliavin derivative operator D in finite-dimensional space $\mathbb{R}^n$ is the gradient operator. Given a $\mathcal{C}^1$-function $f : \mathbb{R}^n \to \mathbb{R}$, the derivative along direction n can be computed as $\frac{\partial f}{\partial n} = \nabla f \cdot n$. This resemblance convinces us that the Malliavin calculus generalizes the ordinary finite-dimensional calculus to infinite-dimensional settings. The Malliavin derivative DF can be intuitively regarded as an H-valued random variable or a d-dimensional stochastic process. Consequently, one is able to define higher order Malliavin derivatives. Let $\mathcal{P}$ denote the collection of polynomials on the Wiener space Ω. Let us define the s-times Malliavin norm $|| \cdot ||_{D_p^s}$ as

$$\|F\|_{D_p^s} = \left[\mathbb{E}\|F\|^p + \sum_{j=1}^{s} \mathbb{E}\|D^{(j)}F\|^p_{H^{\otimes j}} \right]^{\frac{1}{p}}. \tag{5.13}$$

By completing $\mathcal{P}$ (in $L^p(\Omega)$) according to norm $|| \cdot ||_{D_p^s}$, one constructs a Banach space denoted by D_p^s, which collects all s-times Malliavin differentiable variables.

References

[1] Ahn, D.-H., R. F. Dittmar, and A. R. Gallant (2002): "Quadratic Term Structure Models: Theory and Evidence," *Review of Financial Studies*, 16, 459–485.

[2] Bismut, J.-M. (1984): *Large deviations and the Malliavin calculus.* Birkhäuser, Boston.

[3] Carr, P., X. Gabaix, and L. Wu (2011): "Linearity-generating processes, unspanned stochastic volatility, and interest-rate option pricing," *Unspanned Stochastic Volatility, and Interest-Rate Option Pricing (March 18, 2011).*

[4] Cheridito, P., D. Filipović, and R. L. Kimmel (2007): "Market Price of Risk Specifications for Affine Models: Theory and Evidence," *Journal of Financial Economics*, 83, 123–170.

[5] Christoffersen, P., S. Heston, and K. Jacobs (2009): "The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work So Well," *Management Science*, 55, 1914–1932.

[6] Dai, Q., and K. J. Singleton (2000): "Specification Analysis of Affine Term Structure Models," *The Journal of Finance*, 55, 1943–1978.

[7] Duffie, D., and R. Kan (1996): "A Yield-Factor Model of Interest Rates," *Mathematical Finance*, 6, 379–406.

[8] Duffie, D., J. Pan, and K. J. Singleton (2000): "Transform analysis and asset pricing for affine jump-diffusions," *Econometrica*, 68, 1343–1376.

[9] Filipovic, D., M. Larsson, and A. B. Trolle (2014): "Linear-rational term structure models," *Swiss Finance Institute Research Paper*, (14-15).

[10] Friz, P. K. (2005): "An Introduction to Malliavin Calculus," *lecture notes.*

[11] Gatheral, J. (2006): *The Volatility Surface, a Practitioner's Guide*. John Wiley and Sons, Hoboken, NJ.

[12] Gil-Pelaez, J. (1951): "Note on the inversion theorem," *Biometrika*, 38(3-4), 481–482.

[13] Gourieroux, C., and R. Sufana (2003): "Wishart quadratic term structure models," *Les Cahiers du CREF of HEC Montreal Working Paper*, (03-10).

[14] Kanwal, R. P. (2004): *Generalized Functions: Theory and Applications*. Birkhäuser Boston, 3 edn.

[15] Karatzas, I., and S. E. Shreve (1991): *Brownian Motion and Stochastic Calculus*, vol. 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edn.

[16] Leippold, M., and L. Wu (2002): "Asset Pricing Under The Quadratic Class," *Journal of Financial and Quantitative Analysis*, 37, 271–295.

[17] Malliavin, P. (1997): *Stochastic Analysis.* Grundlehren der Mathematischen Wissenschaften. 313. Berlin: Springer. xi, 342 p.

[18] Malliavin, P., and A. Thalmaier (2006): *Stochastic Calculus of Variations in Mathematical Finance.* Springer Finance. Berlin: Springer.

[19] Merton, R. C. (1976): "Option pricing when underlying stock returns are discontinuous," *Journal of Financial Economics*, 3, 125–144.

[20] Nualart, D. (2006): *The Malliavin Calculus and Related Topics. 2nd ed.* Probability and Its Applications. Berlin: Springer.

[21] Shreve, S. E. (2004): *Stochastic Calculus for Finance. II*, Springer Finance. Springer-Verlag, New York.

[22] Singleton, K. J. (2006): *Empirical Dynamic Asset Pricing: Model Specification and Econometric Assessment.* Princeton University Press.

Homework Assignment – 1
Stochastic Analysis and Applications
Spring 2017
Due: in three weeks after the distribution, in class

§1 Enhancement of understanding

Please pick up and complete any five questions or issues discussed in class.

§2 Assigned questions

1. Prove that, if a martingale with continuous sample paths has bounded first variation, it must be a constant. (Hint: consider its quadratic variation and apply the Doob-Meyer decomposition.)
2. 1.5.7 Problem from p. 36 of Karatzas and Shreve [1]
3. 1.5.20 Exercise from p. 37 of Karatzas and Shreve [1]
4. 2.8.12 – 2.8.14 Problems from p. 100 of Karatzas and Shreve [1] (Hint: use the Optional Sampling Theorem)
5. 3.3.18 Problem from p. 158 of Karatzas and Shreve [1] (Hint: verify the definition)
6. 3.3.14 Problem from p. 156 of Karatzas and Shreve [1]
7. Exercise 11.3 from p. 526 of Shreve [2]
8. Exercise 4.13 from p. 197 of Shreve [2]
9. Exercise 4.15 from p. 199 of Shreve [2]
10. Exercise 4.16 from p. 200 of Shreve [2]

References

[1] Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*, vol. 113 of *Graduate Texts in Mathematics*. 2nd ed. Springer-Verlag, New York.

[2] Shreve, S. E. (2004). *Stochastic Calculus for Finance. II.* Springer Finance, Springer-Verlag, New York.

Homework Assignment – 2
Stochastic Analysis and Applications
Spring 2017
Due: in three weeks after the distribution, in class

§1 Enhancement of understanding

Please pick up and complete any ten questions or issues discussed in class. In particular, we prefer questions regarding applications of stochastic analysis in financial engineering.

§2 Assigned questions

1. Given (8.3) and (8.4) on p. 96, use the Girsanov theorem to solve 3.5.9 Exercise on p. 197 of Karatzas and Shreve [1]
2. 5.1.2 Problem from p. 283 of Karatzas and Shreve [1]
3. 5.6.15 Problem from p. 360 of Karatzas and Shreve [1]
4. Exercise 6.8 from p. 291 of Shreve [2]
5. Exercise 6.9 from p. 291 of Shreve [2]

References

[1] Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*, vol. 113 of *Graduate Texts in Mathematics*. 2nd ed. Springer-Verlag, New York.

[2] Shreve, S. E. (2004). *Stochastic Calculus for Finance. II.* Springer Finance, Springer-Verlag, New York.