

Notes on Financial Mathematics

Yunfei Luo



RIVIAL, or not trivial: that is the question:
Whether 'tis brighter in the mind to neglect
The spaces and fields of irritating randomness,
Or to watch Lebesgue dive into an ocean of discontinuities,
And by measuring handle them? To indicate, to integrate;
No more; and by an integral we mitigate
the crisis and the million latent madnesses
That mathematicians adhere to, 'tis a salvation
Timelessly to be honored. To indicate, to integrate;
To integrate: proclivity to formulate: ay, there's the trap;
For in that integral of indicators what formula may come
Since we have evaded the Riemann sum,
Must pardon us rest: there's the reason
that triggers agony of so many graduates
For who would stomach the burns and freezes of notations
The definition's twists, the operator's ambiguity
The detours of recurring conditions, the law's gravity
The convolution of transforms and the chaos
That possible lack of the prerequisite causes
When they themselves might their courses finish
With hollow skulls? who would backpacks shoulder
To whine and whirl in a heavy snow
But that the fear of something beneath answers,
The hidden riddles by whose bewildered
No challenger solves, erodes the bravery
And makes us rather tackle those problems we have
Than hasten to others that we are unaware of?
Thus sophistication does make fools of us all.
And thus the vivid scenery of curiosity
Is blurred o'er with gloomy mist of hesitation
And serendipity of great ideas and concepts
With this concern their compasses become pointless
And lose the purpose of action. Soft you now.
The gifted Kolmogorov! Dude, in thy books
Be all my hints suggested.

Contents

1	Measure Theory	1
2	Probability	10
3	Brownian Motion	16
4	Martingale	27
5	Stochastic Calculus	33
6	Stochastic Models and Computational Methods	39
7	Fundamentals of Finance	68
8	High-Frequency Trading (Draft)	90
A	Analysis	105
B	Univariate Distributions	122
C	Multivariate Distributions	127
D	The Formula of PI	128

SECTION 1

Measure Theory

Foreword

It all begins with the coin tossing.

Theorem 1.1. *The collection of the Bernoulli infinite process is uncountably infinite.*

Proof. Let $f : \mathbb{N} \rightarrow (\epsilon_i \in \{H, T\})_{i \in \mathbb{N}}$, $f_n = \{\epsilon_{n1}, \epsilon_{n2}, \dots, \epsilon_{nk}, \dots\}$ such that ϵ_{nk} represents the k^{th} outcome of the n^{th} sequence. And now we define a new sequence g_n as follows.

$$g_n = \begin{cases} H, & \epsilon_{nn} = T \\ T, & \epsilon_{nn} = H \end{cases} \quad (1.1)$$

Ω_∞ is uncountable since g_n cannot be found in f_n . \square

Remark 1.2. *The process of tossing the coin infinitely is equivalent to the power set $2^{\mathbb{N}}$ which is uncountable.*

Theorem 1.3. \mathbb{N}^k is countable for $k \in \mathbb{N}_{++}$.

Proof. Suppose we choose the first k primes and denote them as p_1, p_2, \dots, p_k . Let the function $f : \mathbb{N}^k \rightarrow \mathbb{N}$, $f(p_1, p_2, \dots, p_k) = p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$. Given the Fundamental Theorem of Algebra, f is a bijection since every positive natural number has a unique factorization of primes, and thus \mathbb{N}^k is countable. Since there are infinite primes, \mathbb{N}^∞ is countable \square

Remark 1.4. *The rationals are countable since $\mathbb{Q} \sim \mathbb{N}^2 \sim \mathbb{N}$.*

Theorem 1.5. *The union of any countable collection of countable sets is countable*

Proof. Let the set $S = \bigcup_{n \in \mathbb{N}} S_n$ where S_n is countable for any $n \in \mathbb{N}$. S_0 is countable. By induction, suppose $\bigcup_{n=0}^k S_n$ is countable for some $k \geq 1$. Let $f : \bigcup_{n=0}^k S_n \rightarrow \bigcup_{n=0}^{k+1} S_n$ be the Cartesian product of elements of $\bigcup_{n=0}^k S_n$ and the elements of S_{k+1} and thus $\bigcup_{n=0}^{k+1} S_n$ is also countable since there exists a bijection. \square

Definition 1.6. *Let Ω be any nonempty set. A nonempty collection of subsets, $\mathcal{F} \subseteq 2^\Omega$, is said to be a σ -field (σ -algebra) if*

1. $\emptyset \in \mathcal{F}$.
2. $A \in \mathcal{F} \iff A^c \in \mathcal{F}$.
3. $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}$ for any sequence $(A_i \in \mathcal{F})_{i \in \mathbb{N}}$.

Remark 1.7. *If the sequence is finite in the last axiom, then \mathcal{F} is an algebra, denoted as \mathcal{A} .*

Theorem 1.8. *(Induced σ -field) Let Ω be any nonempty set and $\mathcal{F} \subseteq 2^\Omega$ be any σ -field. For every nonempty subset $A \subseteq \Omega$, the family $A \cap \mathcal{F} = \{A \cap B : B \in \mathcal{F}\} \subseteq 2^A$ is a σ -field on the set A .*

Proof. $\emptyset \in A \cap \mathcal{F}$. Given any set $B \in \mathcal{F}$ such that $A \cap B \in A \cap \mathcal{F}$, $B^c \in \mathcal{F}$ by the second axiom of σ -field and thus $A \cap B^c \in A \cap \mathcal{F}$. Given any sequence $(B_i \in \mathcal{F})_{i \in \mathbb{N}}$ such that $A \cap B_i \in A \cap \mathcal{F}$ for every i , $\bigcup_{i \in \mathbb{N}} B_i \in \mathcal{F}$ by the third axiom of σ -field and thus $A \cap \bigcup_{i \in \mathbb{N}} B_i \in A \cap \mathcal{F}$. \square

Theorem 1.9. *If \mathcal{R} is a ring, then the family $\mathcal{A} = \mathcal{R} \cup \{A^c : A \in \mathcal{R}\}$ is an algebra.*

Proof. $\emptyset \in \mathcal{R} \subseteq \mathcal{A}$. Let the set $B \in \mathcal{A}$. $B \in \mathcal{R}$ or $B \in \{A^c : A \in \mathcal{R}\}$. Then, $B^c \in \{A^c : A \in \mathcal{R}\}$ or $B^c \in \mathcal{R}$ which implies $B^c \in \mathcal{A}$. \square

Theorem 1.10. *A ring is an algebra if $\Omega \in \mathcal{R}$.*

Proof. Denote a ring as \mathcal{R} . $\emptyset \in \mathcal{R}$ by definition. Since the sets $A, B \in \mathcal{R}$ implies $A \setminus B \in \mathcal{R}$, let $A = \Omega \in \mathcal{R}$ and $B \in \mathcal{R}$ be any subset and thus $B^c = \Omega \setminus B \in \mathcal{R}$. Since the sets $A, B \in \mathcal{R}$ implies $A \cup B \in \mathcal{R}$, for any sequence $(A_i \in \mathcal{R})_{i \in \mathbb{N}}$, $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{R}$ by induction. \square

Theorem 1.11. *An algebra is a ring.*

Proof. Denote an algebra as \mathcal{A} . $\emptyset \in \mathcal{A}$ given the first axiom. If the set $A \in \mathcal{A}$ and the set $B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}$ given the third axiom and $A \setminus B = (A^c \cup B)^c \in \mathcal{A}$ given the second and the third axioms. \square

Theorem 1.12. *A nonempty collection of subsets, $\mathcal{A} \in 2^\Omega$, is an algebra if and only if it is closed under finite union and complement.*

Proof. If \mathcal{A} is an algebra, then it is closed under complement given the second axiom and closed under finite union given the third axiom. Given a nonempty collection of subsets \mathcal{A} that is closed under finite union and complement, $\emptyset = \Omega^c \in \mathcal{A}$ and thus all axioms are satisfied. \square

Theorem 1.13. *If $\Omega \neq \emptyset$ and \mathcal{F} is any σ -field on Ω , then $\Omega = \emptyset^c \in \mathcal{F}$ and $\bigcap_{i \in \mathbb{N}} A_i = (\bigcup_{i \in \mathbb{N}} A_i^c)^c \in \mathcal{F}$.*

Theorem 1.14. *The intersection of σ -field is a σ -field.*

Proof. Denote n σ -field as $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n$ for $n \in \mathbb{N}_{++}$. $\emptyset \in \bigcap_{i=1}^n \mathcal{F}_i$ since the empty set belongs to any space. Suppose the set $A \in \bigcap_{i=1}^n \mathcal{F}_i$. Then, $A \in \mathcal{F}_1, A \in \mathcal{F}_2, \dots, A \in \mathcal{F}_n$. Then, $A^c \in \mathcal{F}_1, A^c \in \mathcal{F}_2, \dots, A^c \in \mathcal{F}_n$ by the property of σ -field and thus $A^c \in \bigcap_{i=1}^n \mathcal{F}_i$. Given any sequence $(A_i \in \mathcal{F})_{i \in \mathbb{N}}$, $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}_1, \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}_2, \dots, \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}_n$ by the property of σ -field. Then, $\bigcup_{i \in \mathbb{N}} A_i \in \bigcap_{k=1}^n \mathcal{F}_k$. \square

Remark 1.15. *The union of σ -field might not be a σ -field. Let $\Omega = \{0, 1\}$, $\mathcal{F}_1 = \{\emptyset, \{0\}\}$, and $\mathcal{F}_2 = \{\emptyset, \{1\}\}$. Both \mathcal{F}_1 and \mathcal{F}_2 are σ -field while $\mathcal{F}_1 \cup \mathcal{F}_2 = \{\emptyset, \{0\}, \{1\}\}$ is not.*

Definition 1.16. *Let Ω be any nonempty set and \mathcal{F} be any σ -field of subsets of Ω . $P : \mathcal{F} \rightarrow [0, 1]$ is a probability measure if*

1. (Full Measure) $P(\Omega) = 1$.
2. (Zero Measure) $P(\emptyset) = 0$.
3. (Countable Additivity) For any sequence $(A_i \in \mathcal{F})_{i \in \mathbb{N}}$ such that $A_i \cap A_j = \emptyset$ for any $i \neq j$, $P(\bigcup_{i \in \mathbb{N}} A_i) = \sum_{i \in \mathbb{N}} P(A_i)$.

Remark 1.17. *(Ω, \mathcal{F}, P) is a probability space if all axioms fulfilled. Remove the first axiom and define $P : \mathcal{F} \rightarrow [0, \infty)$. P is a measure on \mathcal{F} . Let $P(\Omega) < \infty$. P is finite. For a sequence $(A_i \in \mathcal{F})_{i \in \mathbb{N}}$ such that $P(A_i) < \infty$ for any i and $\bigcup_{i \in \mathbb{N}} A_i = \Omega$, P is σ -finite. Finite implies σ -finite.*

Theorem 1.18. (Induced Measure) Let $(\mathbb{X}, \mathcal{S}, \mu)$ be any measure and let A be nonempty element of \mathcal{S} . The set function $A \cap \mathcal{S} \ni B \mapsto \mu(B)$ is a measure.

Proof. Let $P(\emptyset) = 0$. Let $B_1, B_2, \dots, B_k, \dots$ be disjoint sets of $A \cap \mathcal{S}$. Since the countable additivity holds for the space \mathcal{S} and $A \in \mathcal{S}$, $P(\bigcup_{i \in \mathbb{N}} B_i) = \sum_{i \in \mathbb{N}} P(B_i)$ and thus countable additivity holds for the measure μ . \square

Theorem 1.19. Let (Ω, \mathcal{F}, P) be any probability space and let $(A_i \in \mathcal{F})_{i \in \mathbb{N}}$ be any sequence of random events in that space.

1. $P(\bigcup_{i \in \mathbb{N}} A_i) \leq \sum_{i \in \mathbb{N}} P(A_i)$.
2. (Inner Continuity) If $A_i \subseteq A_{i+1}$ for any $i \in \mathbb{N}$ and $A = \bigcup_{i \in \mathbb{N}} A_i$, then $P(A) = \lim_i P(A_i)$.
3. (Outer Continuity) If $A_i \supseteq A_{i+1}$ for any $i \in \mathbb{N}$ and $A = \bigcap_{i \in \mathbb{N}} A_i$, then $P(A) = \lim_i P(A_i)$.

Remark 1.20. $\bigcup_{i \in \mathbb{N}} A_i = A_0 \cup (\bigcup_{i \in \mathbb{N}} A_{i+1} \setminus A_i)$.

Theorem 1.21. For every fixed $u > t$, $P_u(A) = P_t(A)$ for any $A \in \mathcal{F}_t$.

Proof. Let A_1, A_2, \dots, A_n be partitions of A at time t such that $P(A) = \sum_{i=1}^n P(A_i)$. Let $A_{k1}, A_{k2}, \dots, A_{km}$ be the refining chain of A_k for any $k \in \mathbb{N}_{++}$ at time $t+1$.

$$P_{t+1}(A) = \sum_{i=1}^n \sum_{j=1}^m P(A_{ij}) = \sum_{i=1}^n P(A_i) = P_t(A), \quad (1.2)$$

and thus $P_{t+1}(A) = P_t(A)$ for any $A \in \mathcal{P}_t$. By induction, for every fixed $u > t$, $P_u(A) = P_t(A)$ for any $A \in \mathcal{F}_t$. \square

Theorem 1.22. For any $\omega \in \Omega_\infty$, the singleton $\{\omega\} \in \mathcal{F}_\infty$ and $P_\infty(\{\omega\}) = 0$.

Proof. Let $\{\omega_1\}, \{\omega_2\}, \dots, \{\omega_k\}, \dots$ be the partitions of Ω_∞ . We have known that $\{\Omega_\infty\} \in \mathcal{F}_\infty$ by the definition of σ -field. Then,

$$\begin{aligned} \{\Omega_\infty\} = \bigcup_{n \in \mathbb{N}} \bigcup_{k=1}^n \{\omega_k\} \in \mathcal{F}_\infty &\implies \left(\bigcup_{n \in \mathbb{N}} \bigcup_{k=1}^n \{\omega_k\} \right)^c \in \mathcal{F}_\infty \implies \bigcap_{n \in \mathbb{N}} \left(\bigcup_{k=1}^n \{\omega_k\} \right)^c \in \mathcal{F}_\infty \\ &\implies \lim_{k \rightarrow \infty} \bigcap_{k=1}^n \{\omega_k\} \in \mathcal{F}_\infty \implies \{\omega_k\} \in \mathcal{F}_\infty. \end{aligned} \quad (1.3)$$

This is equivalent to $\{\omega\} \in \mathcal{F}$ since and the indexing is arbitrary. Also,

$$P_\infty(\{\omega\}) = P_\infty \left[\left(\bigcup_{n \in \mathbb{N}} \bigcup_{k=1}^n \{\omega_k\} \right)^c \right] = 1 - P_\infty \left(\bigcup_{n \in \mathbb{N}} \bigcup_{k=1}^n \{\omega_k\} \right) = 1 - 1 = 0. \quad (1.4)$$

\square

Definition 1.23. Let $(\mathbb{X}, \mathcal{S}, \mu)$ be any measure space such that $\{x\} \in \mathcal{S}$ for any $x \in \mathbb{X}$. μ is nonatomic if $\mu(\{x\}) = 0$ for any $x \in \mathbb{X}$. μ is purely atomic if there exists a countable set $\mathcal{X} \subset \mathbb{X}$ such that

$$\mu(A) = \sum_{x \in \mathcal{X} \cap A} \mu(\{x\}) \quad \text{for any } A \in \mathcal{S}.$$

Definition 1.24. A_i occurs infinitely often is equivalent to the event

$$\limsup A_i = \bigcap_{i \in \mathbb{N}} \bigcup_{j \geq i} A_j.$$

Theorem 1.25. (First Borel-Cantelli Lemma) Let $(\mathbb{X}, \mathcal{S}, \mu)$ be any measure space and let $(A_i \in \mathcal{S})_{i \in \mathbb{N}}$ be any sequence of sets from \mathcal{S} . Then,

$$\sum_{i \in \mathbb{N}} \mu(A_i) < \infty \implies \mu(\limsup A_i) = 0. \quad (1.5)$$

Theorem 1.26. (Second Borel-Cantelli Lemma) Let (Ω, \mathcal{F}, P) be any probability measure space and let $(A_i \in \mathcal{F})_{i \in \mathbb{N}}$ be any sequence of independent events from \mathcal{S} . Then,

$$\sum_{i \in \mathbb{N}} P(A_i) = \infty \implies P(\limsup A_i) = 1. \quad (1.6)$$

Remark 1.27. The second Borel-Cantelli Lemma is the converse of the first Borel-Cantelli Lemma. The independence can be weakened to the pairwise independence.

Theorem 1.28. Let the set function $\mu_* : \mathcal{R} \rightarrow \mathbb{R}$, $\mu_*([a, b]) = F(b) - F(a)$ defined on an increasing function $F : \mathbb{R} \rightarrow \mathbb{R}$. F is right-continuous if F is countably additive.

Proof. Suppose F is not a right-continuous function on $[a, b]$. Then, there exists c such that F is not continuous at c .

$$\mu_*([a, b]) = F(b) - F(a). \quad (1.7)$$

$$\mu_*([a, c]) + \mu_*([c, b]) = F(b) - F(c+) + F(c-) - F(a). \quad (1.8)$$

Since $F(x)$ is not continuous at c , $F(c-) \neq F(c+)$ and thus $\mu_*([a, b]) \neq \mu_*([a, c]) + \mu_*([c, b])$ which contradicts the countable additivity. Thus, $F(x)$ must be right-continuous. \square

Theorem 1.29. $\{x\} \in \mathcal{B}(\mathbb{R})$ for any $x \in \mathbb{R}$ and $\Lambda(\{x\}) = 0$.

Proof. Let $x \in \mathbb{R}$. The family $\{(x - \frac{1}{n}, x + \frac{1}{n}) : n \in \mathbb{N}_{++}\}$ is a Borel σ -field by definition. By the properties of σ -field,

$$\begin{aligned} \left(x - \frac{1}{n}, x + \frac{1}{n}\right)^c \in \mathcal{B}(\mathbb{R}) &\implies \bigcup_{n \in \mathbb{N}} \left(x - \frac{1}{n}, x + \frac{1}{n}\right)^c \in \mathcal{B}(\mathbb{R}) \\ &\implies \left(\bigcap_{n \in \mathbb{N}} \left(x - \frac{1}{n}, x + \frac{1}{n}\right)\right)^c \in \mathcal{B}(\mathbb{R}) \\ &\implies \bigcap_{n \in \mathbb{N}} \left(x - \frac{1}{n}, x + \frac{1}{n}\right) \in \mathcal{B}(\mathbb{R}) \implies \{x\} \in \mathcal{B}(\mathbb{R}) \end{aligned} \quad (1.9)$$

By definition, $\Lambda(\{x\}) \geq 0$. Suppose $\Lambda(\{x\}) = \epsilon > 0$, we can find ϵ' such that $0 < \epsilon' < \epsilon$ and

$$\Lambda([x - \epsilon_0, x]) = \epsilon' < \epsilon = \Lambda(\{x\}). \quad (1.10)$$

However, $\Lambda([x - \epsilon_0, x]) \geq \Lambda(\{x\})$ by definition. This is a contradiction and thus $\Lambda(\{x\}) = 0$. \square

Corollary 1.30. For $-\infty < a < b < \infty$, $\Lambda([a, b]) = \Lambda([a, b]) = \Lambda([a, b]) = \Lambda([a, b])$.

Proof.

$$\Lambda([a, b]) = \Lambda([a, b]) + \Lambda(\{a\}) = \Lambda([a, b]). \quad (1.11)$$

$$\Lambda([a, b]) = \Lambda([a, b]) + \Lambda(\{b\}) = \Lambda([a, b]). \quad (1.12)$$

$$\Lambda([a, b]) = \Lambda([a, b]) + \Lambda(\{a\}) + \Lambda(\{b\}) = \Lambda([a, b]). \quad (1.13)$$

□

Theorem 1.31. For any countable set $C \subset \mathbb{R}$, $C \in \mathcal{B}(\mathbb{R})$ and $\Lambda(C) = 0$.

Proof. Given C is countable, $C = \bigcup_{i \in \mathbb{N}} \{c_i\} \in \mathcal{B}(\mathbb{R})$ since the countable union of Borel sets is a Borel set. Given the countable additivity of measures,

$$\Lambda(C) = \Lambda\left(\sum_{i \in \mathbb{N}} \{c_i\}\right) = \sum_{i \in \mathbb{N}} \Lambda(\{c_i\}) = 0. \quad (1.14)$$

□

Corollary 1.32. Let $\mathcal{D} = \{(a, b) \in \mathbb{R}^2 : a, b \in \mathbb{Q}\}$. Since \mathbb{Q} and \mathcal{D} are countable, $\mathbb{Q} \in \mathcal{B}(\mathbb{R})$ and $\mathcal{D} \in \mathcal{B}(\mathbb{R})$ and $\Lambda(\mathbb{Q}) = \Lambda(\mathcal{D}) = 0$.

Definition 1.33. Let $(\mathbb{X}, \mathcal{S})$ be any measurable space and let μ and ν be any two measures on \mathcal{S} such that $\mu(\mathbb{X}) > 0$ and $\nu(\mathbb{X}) > 0$. μ and ν are singular ($\mu \perp \nu$) if there exists a set $A \in \mathcal{S}$ such that $\mu(A) = \nu(A^c) = 0$.

Theorem 1.34. If μ is any purely atomic measure on \mathbb{R} , then $\mu \perp \Lambda$.

Proof. Let $A = \mathbb{R} \setminus \mathbb{Q}$. $\Lambda(A^c) = \Lambda(\mathbb{Q}) = 0$. Since μ is a purely atomic measure on \mathbb{R} , there exists a countable set $\mathcal{X} \subset \mathbb{R}$ such that

$$\mu(\mathbb{R} \setminus \mathbb{Q}) = \sum_{x \in \mathcal{X} \cap (\mathbb{R} \setminus \mathbb{Q})} \mu(\{x\}). \quad (1.15)$$

Given the collection of x is a countable set and the singleton $\{x\}$ has measure 0, $\mu(\mathbb{R} \setminus \mathbb{Q}) = 0$. $\mu \perp \Lambda$ since we have found a set $A = \mathbb{R} \setminus \mathbb{Q}$ such that $\mu(A) = \Lambda(A^c) = 0$. □

Theorem 1.35. The Cantor set $\mathcal{C} \in \mathcal{B}([0, 1])$ and $\Lambda(\mathcal{C}) = 0$.

Proof. The Cantor set can be defined as

$$\mathcal{C} = \bigcap_{n=1}^{\infty} \bigcup_{k=0}^{2^n-1} \left(\left[\frac{3k}{3^n}, \frac{3k+1}{3^n} \right] \cup \left[\frac{3k+2}{3^n}, \frac{3k+3}{3^n} \right] \right). \quad (1.16)$$

Since \mathcal{C} is closed under countable intersections, \mathcal{C} is a Borel set.

$$\Lambda(\mathcal{C}^c) = \frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \cdots = \sum_{n=0}^{\infty} \frac{2^n}{3^{n+1}} = \frac{1}{3} \cdot \sum_{n=0}^{\infty} \left(\frac{2}{3}\right)^n = \frac{1}{3} \cdot \frac{1}{1 - \frac{2}{3}} = 1. \quad (1.17)$$

Thus, $\Lambda(\mathcal{C}) = \Lambda([0, 1]) - \Lambda(\mathcal{C}^c) = 1 - 1 = 0$. □

Theorem 1.36. $[0, 1]$ and \mathbb{R} are uncountably infinite.

Proof. Any real number on $[0, 1]$ can be written the decimal format $0.d_1d_2\cdots d_k\cdots$. Define an injection

$$f : \mathbb{N} \rightarrow [0, 1], \quad f(n) = 0.d_{n1}d_{n2}\cdots d_{nk}\cdots \quad (1.18)$$

such that the d_{nk} represents the k -th decimal of the n -th number. And now we define a number g_n as follows.

$$g_n = \begin{cases} 7 & d_{nn} \neq 7 \\ 5 & d_{nn} = 7 \end{cases} \quad (1.19)$$

Since g_n cannot be found in $f(n)$, $[0, 1]$ is uncountable. $[0, 1]$ and \mathbb{R} have the same cardinality and thus \mathbb{R} is uncountable. \square

Theorem 1.37. Let Ω be any nonempty set, $(\mathbb{X}, \mathcal{S})$ be any measurable space, and $X : \Omega \rightarrow \mathbb{X}$ be any function. $X^{-1}(\mathcal{S}) = \{X^{-1}(B) : B \in \mathcal{S}\} \subseteq 2^\Omega$ is a σ -field.

Proof. $\emptyset \in X^{-1}(\mathcal{S})$ since the empty set belongs to any space. If $X^{-1}(A) \in \mathcal{F}$ for some $A \in \mathcal{S}$, then $X^{-1}(A^c) = \{X^{-1}(B) : B \in A^c\} = \{X^{-1}(B) : B \in A\}^c = (X^{-1}(A))^c \in \mathcal{F}$. $\cup_{i \in \mathbb{N}} X^{-1}(B_i) = \cup_{i \in \mathbb{N}} \{X^{-1}(B_i) : B_i \in \mathcal{S}\} = \{X^{-1}(\cup_{i \in \mathbb{N}} B_i) : \cup_{i \in \mathbb{N}} B_i \in \mathcal{S}\} = X^{-1}(\cup_{i \in \mathbb{N}} B_i)$. \square

Definition 1.38. Given any two measurable spaces, (Ω, \mathcal{F}) and $(\mathbb{X}, \mathcal{S})$ and a function $X : \Omega \rightarrow \mathbb{X}$, X is \mathcal{S}/\mathcal{F} -measurable if $X^{-1}(\mathcal{S}) \subseteq \mathcal{F}$, abbreviated as $X \prec \mathcal{S}/\mathcal{F}$.

Theorem 1.39. Let (Ω, \mathcal{F}) and $(\mathbb{X}, \mathcal{S})$ be two measurable spaces and suppose \mathcal{S} is generated by some family $\mathcal{C} \subseteq 2^\mathbb{X}$, in that \mathcal{S} is the smallest σ -field over \mathbb{X} that includes \mathcal{C} . Then, the function $X : \Omega \rightarrow \mathbb{X}$ is \mathcal{S}/\mathcal{F} -measurable if and only if $X^{-1}(A) \in \mathcal{F}$ for any $A \in \mathcal{C}$.

Theorem 1.40. Let (Ω, \mathcal{F}) , $(\mathbb{X}, \mathcal{S})$, and $(\mathbb{Y}, \mathcal{T})$ be any three measurable spaces and let $f : \Omega \rightarrow \mathbb{X} \times \mathbb{Y}$ be any function. f is $(\mathcal{S} \otimes \mathcal{T})/\mathcal{F}$ -measurable if and only if $f^{-1}(A \times B) \in \mathcal{F}$ for any $A \in \mathcal{S}$ and for any $B \in \mathcal{T}$.

Proof. If f is $(\mathcal{S} \otimes \mathcal{T})/\mathcal{F}$ -measurable, $f^{-1}(\mathcal{S} \times \mathcal{T}) \in \mathcal{F}$. Then,

$$\begin{aligned} f^{-1}(\mathcal{S} \times \mathcal{T}) \in \mathcal{F} &\implies (X, Y)^{-1}(\mathcal{S} \times \mathcal{T}) \in \mathcal{F} \implies X^{-1}(\mathcal{S}) \cap Y^{-1}(\mathcal{T}) \in \mathcal{F} \\ &\implies X^{-1}(A) \cap Y^{-1}(B) \subseteq X^{-1}(\mathcal{S}) \cap Y^{-1}(\mathcal{T}) \in \mathcal{F} \\ &\implies (X, Y)^{-1}(A \times B) \in \mathcal{F} \implies f^{-1}(A \times B) \in \mathcal{F}. \end{aligned} \quad (1.20)$$

If $f^{-1}(A \times B) \in \mathcal{F}$ is true for every $A \in \mathcal{S}$ and $B \in \mathcal{T}$. Let $A = \mathcal{S}$ and $B = \mathcal{T}$. $f^{-1}(\mathcal{S} \times \mathcal{T}) \in \mathcal{F}$ and thus f is $(\mathcal{S} \otimes \mathcal{T})/\mathcal{F}$ -measurable. \square

Theorem 1.41. Given a measurable space (Ω, \mathcal{F}) , the function $X : \Omega \rightarrow \mathbb{R}$ is Borel-measurable if and only if $X^{-1}(I) \in \mathcal{F}$ for all intervals $I \subset \mathbb{R}$ of a particular type.

Proof. If $X \in \mathcal{B}(\mathbb{R})$, then we can write X as countable unions or intersections of intervals on \mathbb{R} . \mathbb{R} is equivalent to $\bigcup_{I \subset \mathbb{R}} I$.

$$X^{-1}(\mathbb{R}) = \Omega \in \mathcal{F} \iff X^{-1}\left(\bigcup_{I \subset \mathbb{R}} I\right) \in \mathcal{F} \iff \bigcup_{I \subset \mathbb{R}} X^{-1}(I) \in \mathcal{F} \iff X^{-1}(I) \in \mathcal{F}. \quad (1.21)$$

The other direction is trivial. Since the unions of those intervals are \mathbb{R} and X^{-1} for any interval is a σ -field, $X^{-1}(\mathbb{R}) \in \mathcal{F}$ and thus $X \in \mathcal{B}(\mathbb{R})$. \square

Theorem 1.42. *The composition of measurable functions is measurable.*

Proof. Let X be \mathcal{S}/\mathcal{F} -measurable and Y be \mathcal{T}/\mathcal{S} -measurable. $X^{-1}(\mathcal{S}) \subseteq \mathcal{F}$ and $Y^{-1}(\mathcal{T}) \subseteq \mathcal{S}$. Then, $(X^{-1} \circ Y^{-1})(\mathcal{T}) \subseteq X^{-1}(\mathcal{S}) \subseteq \mathcal{F}$ and thus $Y \circ X$ is \mathcal{T}/\mathcal{F} -measurable. \square

Theorem 1.43. *Let (Ω, \mathcal{F}) , $(\mathbb{X}, \mathcal{S})$, and $(\mathbb{Y}, \mathcal{T})$ be any three measurable spaces and let $f : \Omega \times \mathbb{X} \rightarrow \mathbb{Y}$ be any $\mathcal{T}/(\mathcal{F} \otimes \mathcal{S})$ -measurable function. For any fixed $x^* \in \mathbb{X}$, the mapping $\omega \mapsto f(\omega, x^*)$ is a \mathcal{T}/\mathcal{F} -measurable function from Ω to \mathbb{Y} .*

Proof. Let $g : \Omega \rightarrow \Omega \times \mathbb{X}$. Since x^* is fixed and (Ω, \mathcal{F}) is measurable, $g^{-1}((\omega, x^*)) = \omega \in \mathcal{F}$ for any $(\omega, x^*) \in \Omega \times \mathbb{X}$ and thus g is $\mathcal{F} \otimes \mathcal{S}/\mathcal{F}$ -measurable. Given that the composition of measurable functions is measurable, $f \circ g : \Omega \rightarrow \mathbb{Y}$ is \mathcal{T}/\mathcal{F} -measurable. \square

Theorem 1.44. *Let $(\mathbb{X}, \mathcal{S})$ and $(\mathbb{Y}, \mathcal{T})$ be any two topological spaces. Any function $f : \mathbb{X} \rightarrow \mathbb{Y}$ that is continuous for the topologies \mathcal{S} and \mathcal{T} must be measurable for the Borel σ -field that these two topologies generate.*

Definition 1.45. *Let (Ω, \mathcal{F}, P) be any probability measure and X be any real-valued random variable on that space. The probability density function of X is*

$$\mathcal{L}_X(B) = P \circ X^{-1}(B). \quad (1.22)$$

Definition 1.46. *Let X be any random variable and let \mathcal{L}_X be its probability density function. The cumulative distribution function of X is*

$$F_X(x) = \mathcal{L}_X([-\infty, x]). \quad (1.23)$$

Theorem 1.47. *F_X is càdlàg and increasing.*

Proof. Suppose there exists a decreasing sequence $\{x_n\}_{n \in \mathbb{N}}$ such that $x_n \rightarrow x$ as $n \rightarrow \infty$ and we will show $\lim_{n \rightarrow \infty} F_X(x_n) = F_X(x)$. Define $A_n = \{\omega : X(\omega) \leq x_n, n \in \mathbb{N}\}$ and $A = \{\omega : X(\omega) \leq x\}$. $A_1 \supseteq A_2 \supseteq \dots \supseteq A$ and

$$\lim_{n \rightarrow \infty} F_X(x_n) = \lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right) = P(A) = F_X(x). \quad (1.24)$$

Given any x_i and x_j such that $x_i < x_j$, $A_i \subseteq A_j \implies P(A_i) \leq P(A_j) \implies F_X(x_i) \leq F_X(x_j)$. \square

Theorem 1.48. *$\lim_{x \rightarrow -\infty} F_X(x) = 0$ and $\lim_{x \rightarrow \infty} F_X(x) = 1$.*

Proof. Let $\{x_n : n \in \mathbb{N}\}$ be a decreasing sequence such that $x_n \rightarrow -\infty$ as $n \rightarrow \infty$. $A_1 \supseteq A_2 \supseteq \dots \supseteq A$ and $\bigcap_{n=1}^{\infty} A_n = \emptyset$.

$$\lim_{x \rightarrow -\infty} F_X(x) = \lim_{n \rightarrow \infty} F_X(x_n) = \lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right) = P(\emptyset) = 0. \quad (1.25)$$

Let $\{x_n : n \in \mathbb{N}\}$ be an increasing sequence such that $x_n \rightarrow \infty$ as $n \rightarrow \infty$. $A_1 \subseteq A_2 \subseteq \dots \subseteq A$ and $\bigcup_{n=1}^{\infty} A_n = \Omega$.

$$\lim_{x \rightarrow \infty} F_X(x) = \lim_{n \rightarrow \infty} F_X(x_n) = \lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = P(\Omega) = 1. \quad (1.26)$$

\square

Theorem 1.49. F_X is continuous if and only if the probability density function \mathcal{L}_X is nonatomic.

Proof. Suppose F_X is continuous. Then, F_X is continuous at a for any $a \in \mathbb{R}$. Let $\{a_n : n \in \mathbb{N}\}$ be an increasing sequence such that $a_n \rightarrow a$ as $n \rightarrow \infty$. Define $A_n =] - \infty, a_n[$ for $n \in \mathbb{N}$.

$$F_X(a) = \mathcal{L}_X(] - \infty, a]) = \mathcal{L}_X(] - \infty, a]) + \mathcal{L}_X(\{a\}). \quad (1.27)$$

$$\lim_{x \rightarrow a^-} F_X(x) = \lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathcal{L}_X(] - \infty, a]). \quad (1.28)$$

Since F_X is continuous, $\lim_{x \rightarrow a^-} F_X(x) = F_X(a) \implies \mathcal{L}_X(\{a\}) = 0$ for any $a \in \mathbb{R}$ which implies \mathcal{L}_X is nonatomic. Suppose \mathcal{L}_X is nonatomic, then $\mathcal{L}_X(\{a\}) = 0$ for any $a \in \mathbb{R}$ and consequently $\mathcal{L}_X(] - \infty, a]) = \mathcal{L}_X(] - \infty, a[)$ which implies the left continuity. Since F_X is right-continuous, F_X is continuous. \square

Remark 1.50. It is possible that the continuous probability density function \mathcal{L}_X and the Lebesgue measure Λ are singular. One example could be Cantor distribution. Let \mathcal{C} be the Cantor set. Then, $\Lambda \perp \mathcal{L}_X$ since there exists a set $\mathcal{C} \in [0, 1]$ such that $\Lambda(\mathcal{C}) = \mathcal{L}_X(\mathcal{C}^c) = 0$.

Theorem 1.51. Let $f : \mathbb{R} \rightarrow \mathbb{R}_+$ be a Borel-measurable function.

1. If $\Lambda(\{x \in \mathbb{R} : f(x) > 0\}) = 0$, then $\int_{\mathbb{R}} f(x) \Lambda(dx) = 0$.
2. If $\Lambda(\{x \in \mathbb{R} : f(x) = \infty\}) > 0$, then $\int_{\mathbb{R}} f(x) \Lambda(dx) = \infty$.
3. If $\Lambda(\{x \in \mathbb{R} : f(x) > 0\}) > 0$, then $\int_{\mathbb{R}} f(x) \Lambda(dx) > 0$.

Theorem 1.52. Let $f(x)$ be the indicator function of rational numbers on some interval I . $f(x)$ is Borel-measurable. $f(x)$ is not Riemann-integrable but Lebesgue-integrable.

Proof. Let the set $A = \{x \in I : f(x) \leq a, a \in \mathbb{R}\}$.

$$A = \begin{cases} \emptyset & a < 0 \\ (\mathbb{R} \setminus \mathbb{Q}) \cap I & 0 \leq a < 1 \\ \mathbb{R} \cap I & a \geq 1 \end{cases} \quad (1.29)$$

A is a Borel set for any $a \in \mathbb{R}$ and thus $f(x)$ is Borel-measurable. This function is nowhere continuous since both rationals and irrationals are dense in any interval. No matter how small the subdivisions of the interval are, the supremum is 1 and the infimum is 0 and thus this function is not well-defined as a Riemann integral. By the definition of Lebesgue integral,

$$\int_I f(x) \Lambda(dx) = \int_{\mathbb{R}} 1_{\mathbb{Q} \cap I} \Lambda(dx) = \Lambda(\mathbb{Q} \cap I) = 0. \quad (1.30)$$

\square

Theorem 1.53. If $f \in \mathcal{L}^1(\mathbb{X}, \mathcal{S}, \mu)$, $|\int_{\mathbb{X}} f(x) \mu(dx)| \leq \int_{\mathbb{X}} |f(x)| \mu(dx)$.

Proof. $f(x) = f^+(x) - f^-(x)$ where $f^+(x) = \max\{f(x), 0\}$ and $f^-(x) = \max\{-f(x), 0\}$.

$$\left| \int_{\mathbb{X}} f(x) \mu(dx) \right| = \left| \int_{\mathbb{X}} (f^+(x) - f^-(x)) \mu(dx) \right| \leq \left| \int_{\mathbb{X}} (f^+(x) + f^-(x)) \mu(dx) \right| = \int_{\mathbb{X}} |f(x)| \mu(dx). \quad (1.31)$$

\square

Definition 1.54. Let $(\mathbb{X}, \mathcal{S}, \mu)$ be a measure space. Let the set $E \in \mathcal{S}$ with $\mu(E) = 0$. The sequence $(f_i(x))_{i \in \mathbb{N}}$ converges a.e. if it is defined and converges for any $x \in \mathbb{X} \setminus E$. We use a.s. instead of a.e. if μ is a probability measure.

Remark 1.55. The limit of the integral is equivalent to the integral of the limit if f_i converges a.e.

Definition 1.56. (Absolute Continuity and Equivalence) Let μ and ν be any two measures on the measurable space $(\mathbb{X}, \mathcal{S})$. ν is absolutely continuous with respect to μ ($\nu \preceq \mu$) if every μ -null set from \mathcal{S} is also a ν -null set ($\mu(A) = 0$ for some $A \in \mathcal{S}$ implies $\nu(A) = 0$). μ and ν are equivalent ($\mu \approx \nu$) if $\nu \preceq \mu$ and $\mu \preceq \nu$.

Theorem 1.57. Let $(\mathbb{X}, \mathcal{S})$ be any measurable space and let μ and ν be any two measures on $(\mathbb{X}, \mathcal{S})$ such that $\nu \preceq \mu$. Then, there exists a Radon-Nikodym derivative $R \in \mathcal{L}^1(\mathbb{X}, \mathcal{S}, \mu)$ such that $\nu(A) = \int_A R d\mu$, abbreviated as $\nu = R \odot \mu$ or $R = d\nu/d\mu$, for any $A \in \mathcal{S}$.

Theorem 1.58. (Computation of Expected Values) Let (Ω, \mathcal{F}, P) be a probability space, X be any finite real-valued random variable on Ω , and $f : \mathbb{R} \rightarrow \mathbb{R}$ be a Borel function. Recall the probability density function is defined as $\mathcal{L}_X = P \circ X^{-1}$ and there exists a density function $\varphi_X = d\mathcal{L}_X/d\Lambda \in \mathcal{L}^1(\Lambda)$. Then,

$$\mathbb{E}[f(X)] = \int_{\Omega} f \circ X dP = \int_{\mathbb{R}} f d\mathcal{L}_X = \int_{\mathbb{R}} f \varphi_X d\Lambda. \quad (1.32)$$

Definition 1.59. Let $I \subseteq \mathbb{R}$ be any left-closed interval and let $F : I \rightarrow \mathbb{R}$ be any increasing function. The span of F is defined as the set $\{y \in \mathbb{R} : F(x) \leq y \leq F(x') \text{ for some } x, x' \in I\}$. The inverse of F is defined as the function $F^{-1}(y) = \inf\{x \in I : F(x) \geq y\}$.

Remark 1.60. If F is càdlàg, then $x \geq F^{-1}(y)$ and $F(x) \geq y$ are equivalent. If F is continuous, then $F(x) = \inf\{y : F^{-1}(y) \geq x\}$ for any $x \in I$.

SECTION 2

Probability

Foreword

An event that almost always happens does not always happen.

Theorem 2.1. (*Computation of Densities*) Let x be a random variable that admits density φ and $f : \mathbb{R} \rightarrow \mathbb{R}$ be a Borel-measurable function. Let $y = g(x)$.

$$\frac{d\mathcal{L}_y}{d\Lambda}(y) = \varphi(g^{-1}(y)) \cdot |\partial g^{-1}(y)|. \quad (2.1)$$

Theorem 2.2. Let $x \sim \mathcal{N}(a, \sigma^2)$ and $y = e^x$. The log-normal density is

$$f_y(y) = \frac{d \ln y}{dy} \cdot f(\ln y) = \frac{1}{y\sqrt{2\pi}\sigma} e^{-\frac{(\ln y - a)^2}{2\sigma^2}}. \quad (2.2)$$

Theorem 2.3. Let $x \sim \mathcal{N}(a, \sigma^2)$ and $\theta \in \mathbb{R}$. Then,

$$\mathbb{E} \left[e^{-\theta \frac{x-a}{\sigma^2} - \frac{|\theta|^2}{2\sigma^2}} \right] = \mathbb{E} \left[e^{\theta \frac{x-a}{\sigma^2} - \frac{|\theta|^2}{2\sigma^2}} \right] = 1. \quad (2.3)$$

Definition 2.4. Given a probability space (Ω, \mathcal{F}, P) and any two random events $A, B \in \mathcal{F}$ with $P(B) > 0$, the conditional probability of A given B is

$$P(A|B) = \frac{P(AB)}{P(B)}. \quad (2.4)$$

Theorem 2.5. (*Product Probability Formula*) For any $A_1, A_2, \dots, A_n \in \mathcal{F}$ with every product probability greater than 0,

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \cdots P(A_n|A_1 A_2 \cdots A_{n-1}). \quad (2.5)$$

Theorem 2.6. (*Total Probability Formula*) If $\mathcal{P} = \{H_1, H_2, \dots, H_n\}$ is some measurable partition of Ω , for all $A \in \mathcal{F}$,

$$P(A) = \sum_{i=1}^n P(AH_i). \quad (2.6)$$

Corollary 2.7. (*Bayes' Formula*) If $A \in \mathcal{F}$ and $P(A) > 0$, for all $i \in \mathbb{N}_{|n|}$,

$$P(H_i|A) = \frac{P(A|H_i)}{\sum_{k=1}^n P(A|H_k)P(H_k)} \cdot P(H_i). \quad (2.7)$$

Definition 2.8. If the joint law of x and y is absolutely continuous with respect to Λ^2 , the joint density is

$$\varphi_{x,y} = \frac{d\mathcal{L}_{x,y}}{d\Lambda^2}. \quad (2.8)$$

The existence of the joint density implies the existence of marginal densities.

$$\varphi_x(x) = \int_{\mathbb{R}} \varphi_{x,y}(x, y) dy. \quad \varphi_y(y) = \int_{\mathbb{R}} \varphi_{x,y}(x, y) dx. \quad (2.9)$$

Definition 2.9. Given n real-valued variables x_1, x_2, \dots, x_n defined on (Ω, \mathcal{F}, P) , they are (jointly) independent if $\mathcal{L}_{x_1, x_2, \dots, x_n} = \mathcal{L}_{x_1} \times \mathcal{L}_{x_2} \times \dots \times \mathcal{L}_{x_n}$. They are pairwise independent if $\mathcal{L}_{x_i, x_j} = \mathcal{L}_{x_i} \times \mathcal{L}_{x_j}$ for $i, j \in \mathbb{N}_n$ with $i \neq j$.

Remark 2.10. Pairwise independence does not imply joint independence. Joint independence does not imply pairwise independence.

Theorem 2.11. Let x and y be any two independent variables from the space $\mathcal{L}^1(\Omega, \mathcal{F}, P)$. Then, $xy \in \mathcal{L}^1(P)$ and $\mathbb{E}[xy] = \mathbb{E}[x]\mathbb{E}[y]$.

Proof.

$$\begin{aligned} \mathbb{E}[xy] &= \int_{\mathbb{R}} \int_{\mathbb{R}} xy \varphi_{x,y}(x, y) dx dy = \int_{\mathbb{R}} \int_{\mathbb{R}} xy \varphi_x(x) \varphi_y(y) dx dy \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} (x \varphi_x(x)) (y \varphi_y(y)) dx dy = \left(\int_{\mathbb{R}} x \varphi_x(x) dx \right) \left(\int_{\mathbb{R}} y \varphi_y(y) dy \right) = \mathbb{E}[x] \mathbb{E}[y]. \end{aligned} \quad (2.10)$$

Given the calculation above, we have shown that xy is P -integrable and thus $xy \in \mathcal{L}^1(P)$. \square

Remark 2.12. Independence implies uncorrelation. Uncorrelation generally does not implies independence except for the multivariate Gaussian distribution.

Theorem 2.13. Let x and y be any two independent random variables with densities $\varphi(x)$ and $\psi(y)$. The density of $x + y$ must be given by the convolution of the two individual densities. Then,

$$(\varphi * \psi)(z) = \int_{\mathbb{R}} \varphi(z - y) \psi(y) dy. \quad (2.11)$$

Theorem 2.14. The mean minimizes the mean square errors. The median minimizes the mean absolute errors. The mode minimizes logistic prediction errors.

Theorem 2.15. Suppose that the random variables x_1, x_2, \dots, x_n have finite second moments and let $s_n = x_1 + x_2 + \dots + x_n$. s_n has a finite second moment and

$$\text{Var}(s_n) = \sum_{i=1}^n \text{Var}(x_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{Cov}(x_i, x_j). \quad (2.12)$$

Theorem 2.16. The conditional expectation of conditional expectation equals the conditional expectation. Let $\mathcal{G} \subseteq \mathcal{G}' \subseteq \mathcal{F}$ and X be any random variable with $\mathbb{E}[X]$ defined.

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}']|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}]. \quad (2.13)$$

Proof.

$$\begin{aligned} \mathbb{E}[\mathbb{E}[X|\mathcal{G}']|\mathcal{G}] &= \mathbb{E}\left[\int_X x P(X = x|\mathcal{G}') dx |\mathcal{G}\right] = \int_{\mathcal{G}'} \left(\int_X x P(X = x|\mathcal{G}') dx\right) P(\mathcal{G}' = g'|\mathcal{G}) dg' \\ &= \int_{\mathcal{G}'} \int_X x P(X = x|\mathcal{G}') P(\mathcal{G}' = g'|\mathcal{G}) dx dg' = \int_X x P(X = x|\mathcal{G}) dx = \mathbb{E}[X|\mathcal{G}]. \end{aligned} \quad (2.14)$$

\square

Theorem 2.17. Let x and y be two random variables with joint density $\varphi_{x,y}$ and suppose that $\mathbb{E}[|f(x, y)|] < \infty$ for some Borel function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. Then,

$$\mathbb{E}[f(x, y)|y] = \varphi(y) = \frac{\int_{\mathbb{R}} f(x, y) \varphi_{x,y}(x, y) dx}{\int_{\mathbb{R}} \varphi_{x,y}(x, y) dx}. \quad (2.15)$$

Proof.

$$\begin{aligned} \mathbb{E}[\varphi(y)g(y)] &= \int_{\mathbb{R}} \frac{\int_{\mathbb{R}} f(x, y) \varphi_{x,y}(x, y) dx}{\int_{\mathbb{R}} \varphi_{x,y}(x, y) dx} g(y) \varphi_y(y) dy = \int_{\mathbb{R}} \frac{\int_{\mathbb{R}} f(x, y) \varphi_{x,y}(x, y) dx}{\varphi_y(y)} \varphi_y(y) g(y) dy \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, y) g(y) \varphi_{x,y}(x, y) dx dy = \mathbb{E}[f(x, y)g(y)]. \end{aligned} \quad (2.16)$$

Let $g(y) = 1$. $\mathbb{E}[f(x, y)] = \mathbb{E}[\varphi(y)] \implies \mathbb{E}[f(x, y)|y] = \mathbb{E}[\varphi(y)|y] = \varphi(y)$. \square

Theorem 2.18. (Bayes' Formula for Conditional Expectations) Let (Ω, \mathcal{F}, P) be any probability space. Let a σ -field $\mathcal{G} \subseteq \mathcal{F}$. Let another probability measure Q on \mathcal{F} such that $Q \sim P$ and $R = dQ/dP$. Since $R \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$, the conditional expectation $\mathbb{E}[R|\mathcal{G}]$ is defined. Let X be any random variable such that $\mathbb{E}[|X|R] < \infty$ and $X \in \mathcal{L}(\Omega, \mathcal{F}, P)$. Then, the conditional expectation $\mathbb{E}^Q[X|\mathcal{G}]$ is defined and

$$\mathbb{E}^Q[X|\mathcal{G}] = \frac{\mathbb{E}^P[XR|\mathcal{G}]}{\mathbb{E}^P[R|\mathcal{G}]}. \quad (2.17)$$

Definition 2.19. The sequence (x_n) converges a.s in probability to a random variable X if

$$P\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} x_n(\omega) = x(\omega)\right\}\right) = 1. \quad (2.18)$$

Definition 2.20. The sequence $(x_n) \in \mathcal{L}^0(\Omega, \mathcal{F}, P)$ converges in probability if there is a random variable $x_* \in \mathcal{L}^0(\Omega, \mathcal{F}, P)$ such that, for any $\epsilon \in \mathbb{R}_{++}$,

$$\lim_{n \rightarrow \infty} P(|x_n - x_*| > \epsilon) = 0. \quad (2.19)$$

Theorem 2.21. Convergence a.s. in probability implies convergence in probability.

Theorem 2.22. A sequence converges in probability if and only if every subsequence contains a sub-subsequence that converges a.s in probability.

Definition 2.23. For $x, y \in \mathcal{L}^0(\Omega, \mathcal{F}, P)$, the Ky Fan distance is

$$d_p(x, y) = \inf \{ \epsilon \in \mathbb{R}_+ : P(|x - y| > \epsilon) \leq \epsilon \}. \quad (2.20)$$

Theorem 2.24. (Cauchy Criteria for Convergence) The sequence of random variables (x_n) converges in probability if and only if

$$\lim_{n \rightarrow \infty} \sup_{m \geq n} d_p(x_n, x_m) = 0. \quad (2.21)$$

The sequence converges a.s. if and only if, for any $\epsilon \in \mathbb{R}_{++}$,

$$\lim_{n \rightarrow \infty} P\left(\sup_{m \geq n} |x_n - x_m| \geq \epsilon\right) = 0. \quad (2.22)$$

Definition 2.25. (\mathcal{L}^p Convergence) Let $p \in [1, \infty[$, the sequence of random variables $(x_n) \in \mathcal{L}^p$ converges in \mathcal{L}^p -norm if there exists a random variable $x_* \in \mathcal{L}^p$ such that

$$\lim_{n \rightarrow \infty} \mathbb{E} [|x_n - x_*|^p] = 0. \quad (2.23)$$

Definition 2.26. Let ξ be a real-valued random variable defined on (Ω, \mathcal{F}, P) . For any $c \in \mathbb{R}_{++}$, ξ is integrable if

$$\mathbb{E}[|\xi|] < \infty \iff \mathbb{E}[|\xi|1_{\{|\xi| \geq c\}}] < \infty \iff \lim_{c \rightarrow \infty} \mathbb{E}[|\xi|1_{\{|\xi| \geq c\}}] = 0. \quad (2.24)$$

ξ is uniformly integrable if

$$\lim_{c \rightarrow \infty} \sup_{\xi \in \Xi} \mathbb{E}[|\xi|1_{\{|\xi| \geq c\}}] = 0. \quad (2.25)$$

Theorem 2.27. If the family Ξ is uniformly integrable, there exists a constant $c \in \mathbb{R}_{++}$ such that $\mathbb{E}[|\xi|] \leq c$ for any $\xi \in \Xi$.

Proof. Since the family Ξ is uniformly integrable, there exists k such that $\sup_{\xi \in \Xi} \mathbb{E}[|\xi|1_{\{|\xi| \geq k\}}] \leq 1$. For any $\xi \in \Xi$, there is a constant $c \in \mathbb{R}_{++}$ such that

$$\sup_{\xi \in \Xi} \mathbb{E}[|\xi|] = \sup_{\xi \in \Xi} (\mathbb{E}[|\xi|1_{\{|\xi| < k\}}] + \mathbb{E}[|\xi|1_{\{|\xi| \geq k\}}]) \leq k + 1 = c. \quad (2.26)$$

□

Theorem 2.28. Let Ξ be a family of integrable random variables. Ξ is uniformly integrable if and only if there exists a positive increasing convex function $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that $\lim_{x \rightarrow \infty} \frac{f(x)}{x} = \infty$ and $\sup_{\xi \in \Xi} \mathbb{E}[f(|\xi|)] < \infty$.

Theorem 2.29. Let ξ be any integrable random variable on (Ω, \mathcal{F}, P) . Then the collection of all random variables on (Ω, \mathcal{F}, P) that can be expressed as conditional expectations of the form $\mathbb{E}[\xi|\mathcal{G}]$, for any sub- σ -field $\mathcal{G} \subseteq \mathcal{F}$, is a uniformly integrable family.

Theorem 2.30. Let \mathcal{A} be any family of random variables and suppose that there is an integrable random such that $|\xi| \leq |\eta|$ P -a.s. for all $\xi \in \mathcal{A}$. Then, the family \mathcal{A} is uniformly integrable.

Proof. Since η is an integrable random variable, $\lim_{c \rightarrow \infty} \sup_{\eta \in H} \mathbb{E}[|\eta|1_{\{|\eta| \geq c\}}] = 0$. Given that $|\xi| \leq |\eta|$ P -a.s. for all $\xi \in \mathcal{A}$, $\mathbb{E}[|\xi|1_{\{|\xi| \geq c\}}] \leq \mathbb{E}[|\eta|1_{\{|\eta| \geq c\}}]$ for any $c \in \mathbb{R}_{++}$. Therefore,

$$0 \leq \lim_{c \rightarrow \infty} \sup_{\xi \in \Xi} \mathbb{E}[|\xi|1_{\{|\xi| \geq c\}}] \leq \lim_{c \rightarrow \infty} \sup_{\eta \in H} \mathbb{E}[|\eta|1_{\{|\eta| \geq c\}}] = 0 \implies \lim_{c \rightarrow \infty} \sup_{\xi \in \Xi} \mathbb{E}[|\xi|1_{\{|\xi| \geq c\}}] = 0. \quad (2.27)$$

Since ξ is arbitrarily chosen from \mathcal{A} , we have proved that the family \mathcal{A} is uniformly integrable. □

Definition 2.31. Let ξ be a sequence of independent random variables. The finite tail σ -field is $\mathcal{T}_k = \sigma(\xi_k, \xi_{k+1}, \dots)$ for $k \in \mathbb{N}$. The infinite tail σ -field is $\mathcal{T}_\infty = \bigcap \mathcal{T}_k$.

Theorem 2.32. (Kolmogorov's 0-1 Law) Let ξ be a sequence of independent random variables defined on (Ω, \mathcal{F}, P) . Let $\mathcal{T}_\infty \subseteq \mathcal{F}$ be the tail σ -field and $A \in \mathcal{T}_\infty$ be any tail event. Then, $P(A) = 0$ or $P(A) = 1$.

Theorem 2.33. Given a coin toss space $(\Omega_\infty, \mathcal{F}_\infty, P_\infty)$ and a randomly chosen sequence $\omega = (\epsilon_1 \epsilon_2 \dots \epsilon_k) \in \Omega_\infty$, the probability for ω to contain infinitely many runs of any possible finite size is 1.

Proof. Let the event A_n be the occurrence of ω in any sequence with the length n . Then,

$$P(A_n) = \begin{cases} 0 & n < k \\ \frac{(n-k+2)!}{2^n} & n \geq k \end{cases} \quad (2.28)$$

Given the second Borel-Cantelli Lemma,

$$\sum_{n \in \mathbb{N}} P(A_n) = \sum_{n \geq k} \frac{(n-k+2)!}{2^n} = \infty \implies P(A_n \text{ i.o.}) = 1. \quad (2.29)$$

□

Theorem 2.34. *Let (x_n) be a sequence of random variables on (Ω, \mathcal{F}, P) such that $\mathbb{E}[|x_n|] < \infty$. Then, $\lim x_n = 0$ P -a.s.*

Proof. Given that $\sum_{n \in \mathbb{N}} \mathbb{E}[|x_n|] < \infty$ and the Chebyshev's inequality, for any fixed $\epsilon > 0$,

$$\sum_{n \in \mathbb{N}} P(|x_n| > \epsilon) < \sum_{n \in \mathbb{N}} \frac{\mathbb{E}[|x_n|]}{\epsilon} < \infty. \quad (2.30)$$

By the first Borel-Cantelli lemma,

$$\sum_{n \in \mathbb{N}} P(|x_n| > \epsilon) < \infty \implies P(\{|x_n| > \epsilon\} \text{ i.o.}) = 0 \implies P(x_n \rightarrow 0) = 1. \quad (2.31)$$

Therefore, $\lim_{n \rightarrow \infty} x_n = 0$ P -a.s. □

Theorem 2.35. *(Levy's Equivalence Theorem and Three-Series Theorem) Let (x_n) be a sequence of independent real-valued random variables. Then, the following statements are equivalent.*

1. The series $\sum x_n$ converges a.s.
2. The series $\sum x_n$ converges in probability.
3. The series $\sum x_n$ converges in distribution.
4. The series $\sum P(|x_n| \geq 1)$, $\sum \mathbb{E}[x_n 1_{\{|x_n| \leq 1\}}]$, and $\sum \text{Var}(x_n 1_{\{|x_n| \leq 1\}})$ converge.

Remark 2.36. *The Levy's Equivalence Theorem is trivial if all x_n are positive.*

Theorem 2.37. *Let the mapping $X_t : \Omega_\infty \rightarrow \{-1, 1\}$. Define the event A as*

$$A = \left\{ \omega \in \Omega_\infty : \sum \frac{X_t(\omega)}{t} \text{ converges to a finite limit} \right\}. \quad (2.32)$$

Then, $P(A) = 1$.

Proof. Let A be any tail event. Given the Kolmogorov's 0 – 1 Law, since the event A is a tail σ -field, either $P(A) = 0$ or $P(A) = 1$. Given the second Borel-Cantelli Lemma, since $P(A_n) = \frac{1}{2}$ for $n \in \mathbb{N}$, $\sum_{n \in \mathbb{N}} P(A_n) = \infty \implies P(A_n \text{ i.o.}) = 1$. Thus, $P(A)_\infty \neq 0 \implies P(A)_\infty = 1$. □

Theorem 2.38. (Central Limit Theorem) Let (x_n) be any sequence of independent and identically distributed random variables such that $\mathbb{E}[|x_n|^2] < \infty$. Let $\mu = \mathbb{E}[x_n]$ and $\sigma^2 = \text{Var}(x_n)$. Then, the sequence

$$\frac{x_1 + x_2 + \cdots + x_n - n\mu}{\sigma\sqrt{n}} \quad (2.33)$$

converges to the standard normal distribution $\mathcal{N}(0, 1)$.

Theorem 2.39. $\mathbb{E}[\hat{\sigma}_n^2] = \sigma^2$.

Proof.

$$\begin{aligned} \mathbb{E}[\hat{\sigma}_n^2] &= \mathbb{E}\left[\frac{\sum_{i=1}^n (x_i - \bar{x}_n)^2}{n-1}\right] = \mathbb{E}\left[\frac{\sum_{i=1}^n (x_i - a)^2 - n(\bar{x}_n - a)^2}{n-1}\right] \\ &= \frac{n}{n-1} \mathbb{E}\left[\frac{\sum_{i=1}^n (x_i - a)^2}{n}\right] - \frac{n}{n-1} \mathbb{E}[(\bar{x}_n - a)^2] = \frac{n}{n-1} \cdot \sigma^2 - \frac{n}{n-1} \cdot \frac{\sigma^2}{n} = \sigma^2. \end{aligned} \quad (2.34)$$

□

Theorem 2.40. (Layer Cake Formula for the Covariance) If x and y do not change signs and are integrable, then

$$\text{Cov}(x, y) = \int_{\mathbb{R}_+ \times \mathbb{R}_+} \text{Cov}(1_{[u, \infty[}(x), 1_{[v, \infty[}(y)) \, dudv. \quad (2.35)$$

Theorem 2.41. Let x be any random variable and let u and v be any real numbers. Then,

$$\text{Cov}(1_{[u, +\infty[}(x), 1_{[v, +\infty[}(x)) \geq 0. \quad (2.36)$$

Proof. Without loss of generality, assume $u < v$. The product of $1_{[u, +\infty[}(x)$ and $1_{[v, +\infty[}(x)$ are $1_{[v, +\infty[}(x)$. $0 \leq \mathbb{E}[1_{[u, +\infty[}(x)] \leq 1$ and $0 \leq \mathbb{E}[1_{[v, +\infty[}(x)] \leq 1$. Given the Layer Cake Formula for the Covariance,

$$\begin{aligned} \text{Cov}(1_{[u, +\infty[}(x), 1_{[v, +\infty[}(x)) &= \mathbb{E}[1_{[u, +\infty[}(x)1_{[v, +\infty[}(x)] - \mathbb{E}[1_{[u, +\infty[}(x)]\mathbb{E}[1_{[v, +\infty[}(x)] \\ &= \mathbb{E}[1_{[v, +\infty[}(x)](1 - \mathbb{E}[1_{[u, +\infty[}(x)]) \geq 0. \end{aligned} \quad (2.37)$$

□

Theorem 2.42. Let x be any random variable. Let $f : \mathbb{R} \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$ be two increasing and càdlàg functions that do not change sign. Then, $\text{Cov}(f(x), g(x))$ is well-defined and $\text{Cov}(f(x), g(x)) \geq 0$.

Proof. The inverse functions of f and g are defined as $f^{-1}(x) = \inf\{x \in \mathbb{R} : f(x) \geq y\}$ and $g^{-1}(x) = \inf\{x \in \mathbb{R} : g(x) \geq y\}$. Since f and g are increasing and càdlàg, $1_{[u, \infty[}(f(x)) = 1_{[f^{-1}(u), +\infty[}(x)$ and $1_{[v, \infty[}(g(x)) = 1_{[g^{-1}(v), +\infty[}(x)$. Given the Layer Cake Formula for the Covariance,

$$\begin{aligned} \text{Cov}(f(x), g(x)) &= \int_{\mathbb{R}_+ \times \mathbb{R}_+} \text{Cov}(1_{[u, +\infty[}(f(x)), 1_{[v, +\infty[}(g(x))) \, dudv \\ &= \int_{\mathbb{R}_+ \times \mathbb{R}_+} \text{Cov}(1_{[f^{-1}(u), +\infty[}(x), 1_{[g^{-1}(v), +\infty[}(x)) \, dudv \geq 0 \end{aligned} \quad (2.38)$$

□

Theorem 2.43. If x is distributed symmetrically and $f : \mathbb{R} \rightarrow \mathbb{R}_+$ is a positive, increasing and continuous function, then

$$\text{Var}\left(\frac{f(x) + f(-x)}{2}\right) \leq \frac{1}{2} \text{Var}(f(x)). \quad (2.39)$$

SECTION 3

Brownian Motion

Foreword

Brownian motion is Gaussian, Markov, a diffusion, and a martingale.

Definition 3.1. A finite real stochastic process W defined on some probability space (Ω, \mathcal{F}, P) , is a linear Brownian motion if

1. $W_0 = k$ P -a.s.
2. $W_t - W_s \in \mathcal{N}(0, t - s)$ for $0 \leq s < t < \infty$.
3. For any $0 = t_0 < t_1 < \dots < t_n < \infty$, the increments

$$W_{t_1} - W_{t_0}, W_{t_2} - W_{t_1}, \dots, W_{t_n} - W_{t_{n-1}}$$

are independent.

4. W is continuous a.s.

Remark 3.2. W is a standard Brownian motion if $k = 0$. Brownian motion is non-differentiable a.s.

Theorem 3.3. (Density Function of Brownian Motion) Let W be a Brownian motion and let $0 \leq t_1 < t_2 < \dots < t_n < \infty$ be arbitrarily chosen. $W_{t_i} = W_{t_i} - W_0 \sim \mathcal{N}(0, t_i)$. Thus, the family of the finite dimensional distributions of the Brownian motion is equivalent to the one of multivariate normal distribution. Let

$$\Sigma = \begin{pmatrix} t_1 & t_1 & t_1 & \dots & t_1 \\ t_1 & t_2 & t_2 & \dots & t_2 \\ t_1 & t_2 & t_3 & \dots & t_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & t_3 & \dots & t_n \end{pmatrix} \quad \text{and} \quad W = \begin{pmatrix} w_{t_1} \\ w_{t_2} \\ w_{t_3} \\ \vdots \\ w_{t_n} \end{pmatrix}.$$

The density function is

$$\begin{aligned} P(W) &= \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} e^{\frac{1}{2} W^T (\Sigma^{-1}) W} \\ &= \frac{1}{\sqrt{(2\pi)^n t_1(t_2 - t_1) \dots (t_n - t_{n-1})}} \cdot \exp \left(-\frac{w_1^2}{2t_1} - \frac{(w_2 - w_1)^2}{2(t_2 - t_1)} \dots - \frac{(w_n - w_{n-1})^2}{2(t_n - t_{n-1})} \right). \end{aligned} \quad (3.1)$$

Definition 3.4. A Gaussian process is any \mathbb{R} -valued stochastic process with the finite dimensional multivariate Gaussian distribution. X is centered if $\mathbb{E}[X_t] = 0$ for all $t \in \mathbb{R}_+$.

Definition 3.5. Let $X = (X_t)_{t \in \mathbb{R}_+}$ be any real stochastic process such that $\mathbb{E}[X_t^2] < \infty$ and $\mathbb{E}[X_t] = 0$ for any $t \in \mathbb{R}_+$. The covariance function of X is defined as

$$\mathbb{R}_+ \times \mathbb{R}_+ \ni (s, t) \rightsquigarrow C(s, t) = \mathbb{E}[X_s X_t].$$

Theorem 3.6. $C(s, t) = \text{Cov}(X_s, X_t)$.

Remark 3.7. *There are two properties of the covariance function. For any $s, t \in \mathbb{R}_+$,*

1. *(Symmetry) $C(s, t) = C(t, s)$.*
2. *(Positive Definiteness) $C(s, t) > 0$.*

Theorem 3.8. *W is a Brownian motion if and only if it is a centered Gaussian process with $\mathbb{E}[W_s W_t] = s \wedge t$ for any $s, t \in \mathbb{R}_+$.*

Proof. Let W be a Brownian motion.

1. W is Gaussian since $W_t \sim \mathcal{N}(0, t)$ for any $t \in \mathbb{R}_+$.

2. For any $t \in \mathbb{R}_+$,

$$\mathbb{E}[W_t] = \mathbb{E}[W_t - W_0] + \mathbb{E}[W_0] = 0. \quad (3.2)$$

3. Let $s < t$ without loss of generality.

$$\begin{aligned} \mathbb{E}[W_s W_t] &= \mathbb{E}[W_s(W_t - W_s) + W_s^2] = \mathbb{E}[(W_s - W_0)(W_t - W_s) + W_s^2] \\ &= \mathbb{E}[W_s - W_0]\mathbb{E}[W_t - W_s] + \mathbb{E}[W_s^2] = \mathbb{E}[W_s^2] \\ &= \text{Var}[W_s] + \mathbb{E}[W_s]^2 = \text{Var}[W_s - W_0] + \mathbb{E}[W_s - W_0]^2 = s = s \wedge t. \end{aligned} \quad (3.3)$$

Let X be a centered Gaussian process with $\mathbb{E}[X_s X_t] = s \wedge t$ for any $s, t \in \mathbb{R}_+$.

1. $X_0 = 0$
2. $X_t - X_s \in N(0, t - s)$ for $0 \leq s < t < \infty$

$$\mathbb{E}[X_t - X_s] = \mathbb{E}[X_t] - \mathbb{E}[X_s] = 0. \quad (3.4)$$

$$\begin{aligned} \text{Var}(X_t - X_s) &= \mathbb{E}[(X_t - X_s)^2] - \mathbb{E}[X_t - X_s]^2 = \mathbb{E}[(X_t - X_s)^2] \\ &= \mathbb{E}[X_t^2] + \mathbb{E}[X_s^2] - 2\mathbb{E}[X_s X_t] \\ &= \mathbb{E}[X_t]^2 + \text{Var}(X_t) + \mathbb{E}[X_s]^2 + \text{Var}(X_s) - 2\mathbb{E}[X_s X_t] \\ &= t + s - 2(s \wedge t) = t + s - 2s = t - s. \end{aligned} \quad (3.5)$$

3. For any $0 = t_0 < t_1 < \dots < t_n < \infty$, the random variables

$$X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$$

are independent. Let $0 \leq u < v < w < \infty$.

$$\begin{aligned} \mathbb{E}[(X_w - X_v)(X_v - X_u)] &= \mathbb{E}[X_w X_v - X_w X_u - X_v^2 + X_v X_u] \\ &= \mathbb{E}[X_w X_v] - \mathbb{E}[X_w X_u] - \mathbb{E}[X_v^2] + \mathbb{E}[X_v X_u] \\ &= w - v - w + v = 0. \end{aligned} \quad (3.6)$$

$$\mathbb{E}[(X_w - X_v)]\mathbb{E}[(X_v - X_u)] = (\mathbb{E}[X_w] - \mathbb{E}[X_v])(\mathbb{E}[X_v] - \mathbb{E}[X_u]) = 0. \quad (3.7)$$

Since X_t has a multivariate Gaussian distribution, uncorrelation implies independence.

□

Theorem 3.9. Let $(\xi_i)_{i \in \mathbb{N}}$ be any infinite sequence of i.i.d. $\mathcal{N}(0, 1)$ random variables and let (φ_i) be any orthonormal basis in the Hilbert space. For $t \in [0, T]$,

$$W_t = \sum_{i \in \mathbb{N}} \xi_i \left(\int_0^t \varphi_i(s) ds \right) \quad (3.8)$$

is a Brownian motion.

Proof.

$$\begin{aligned} \mathbb{E}[W_t] &= \mathbb{E} \left[\sum_{i \in \mathbb{N}} \xi_i \left(\int_0^t \varphi_i(x) dx \right) \right] = \sum_{i \in \mathbb{N}} \mathbb{E} \left[\xi_i \left(\int_0^t \varphi_i(x) dx \right) \right] \\ &= \sum_{i \in \mathbb{N}} \mathbb{E}[\xi_i] \cdot \mathbb{E} \left[\int_0^t \varphi_i(x) dx \right] = 0. \end{aligned} \quad (3.9)$$

$$\begin{aligned} \mathbb{E}[W_s W_t] &= \mathbb{E} \left[\sum_{i, j \in \mathbb{N}} \xi_i \xi_j \int_0^s \int_0^t \varphi_i(x) \varphi_j(y) dx dy \right] \\ &= \mathbb{E} \left[\sum_{i \in \mathbb{N}} \xi_i^2 \int_0^s \int_0^t \varphi_i(x) \varphi_i(y) dx dy \right] = \sum_{i \in \mathbb{N}} \mathbb{E} \left[\xi_i^2 \int_0^s \int_0^t \varphi_i(x) \varphi_i(y) dx dy \right] \\ &= \sum_{i \in \mathbb{N}} \mathbb{E}[\xi_i^2] \mathbb{E} \left[\int_0^s \int_0^t \varphi_i(x) \varphi_i(y) dx dy \right] = \sum_{i \in \mathbb{N}} \int_0^s \int_0^t \varphi_i(x) \varphi_i(y) dx dy \\ &= \int_{\mathbb{R}} 1_{[0, s]}(u) 1_{[0, t]}(u) du = s \wedge t. \end{aligned} \quad (3.10)$$

Since W_t is a Gaussian process, W_t is a Brownian motion. □

Theorem 3.10. (*Transformations of Brownian Motion*) Let W be a Brownian motion.

- (Shifting Invariance) For any fixed $s \in \mathbb{R}_+$, $W_{s+t} - W_s$ is a Brownian motion that is independent of $(W_t)_{t \in [0, s]}$.
- (Symmetry and Scaling Invariance) For any $a \in \mathbb{R} \setminus \{0\}$, aW_{t/a^2} is a Brownian motion.
- (Time Reversal Property) $(W_{a-t} - W_t)_{t \in [0, a]}$ is a Brownian motion.
- (Time Inversion Property) Let $B_t = tW_{1/t}$ for any $t \in \mathbb{R}_{++}$ and $B_0 = 0$. B is a Brownian motion.

Proof. a. Denote $W'_t = W_{s+t} - W_s$.

- $W'_0 = W_{s+0} - W_s = 0$.
- $W'_v - W'_u = (W_{s+v} - W_s) - (W_{s+u} - W_s) = W_{s+v} - W_{s+u} \in N(0, v - u)$ for $0 \leq u < v < \infty$.

3. For any $0 = t_0 < t_1 < \dots < t_n < \infty$, the random variables

$$\begin{aligned} (W_{s+t_1} - W_s) - (W_{s+t_0} - W_s) &= W_{s+t_1} - W_{s+t_0}, \\ (W_{s+t_2} - W_s) - (W_{s+t_1} - W_s) &= W_{s+t_2} - W_{s+t_1}, \\ &\dots \\ (W_{s+t_n} - W_s) - (W_{s+t_{n-1}} - W_s) &= W_{s+t_n} - W_{s+t_{n-1}} \end{aligned}$$

are independent.

4. This process is independent of $(W_t)_{t \in [0, s]}$.

$$\mathbb{E}[(W_{s+t} - W_s)(W_t)] = \mathbb{E}[(W_{s+t} - W_s)(W_t - W_0)] = \mathbb{E}[(W_{t+s} - W_t)]\mathbb{E}[(W_s - W_0)]. \quad (3.11)$$

b. Denote $W'_t = aW_{t/a^2}$ for any $a \in \mathbb{R} \setminus \{0\}$.

1. $W'_0 = aW_{0/a^2} = 0$.

2. $W'_t - W'_s = a(W_{t/a^2} - W_{s/a^2}) \in N(0, t - s)$ for $0 \leq s < t < \infty$. $W'_t - W'_s$ is Gaussian.

$$\mathbb{E}[a(W_{t/a^2} - W_{s/a^2})] = a\mathbb{E}[W_{t/a^2} - W_{s/a^2}] = 0. \quad (3.12)$$

$$\text{Var}(a(W_{t/a^2} - W_{s/a^2})) = a^2 \text{Var}(W_{t/a^2} - W_{s/a^2}) = a^2 \cdot \frac{t - s}{a^2} = t - s. \quad (3.13)$$

3. For any $0 = t_0 < t_1 < \dots < t_n < \infty$, the random variables

$$\begin{aligned} W'_{t_1} - W'_{t_0} &= a(W_{t_1/a^2} - W_{t_0/a^2}), \quad W'_{t_2} - W'_{t_1} = a(W_{t_2/a^2} - W_{t_1/a^2}), \\ &\dots, \quad W'_{t_n} - W'_{t_{n-1}} = a(W_{t_n/a^2} - W_{t_{n-1}/a^2}) \end{aligned}$$

are independent.

$$\begin{aligned} \mathbb{E}[(aX)(aY)] &= \mathbb{E}[a^2XY] = a^2\mathbb{E}[XY] \\ &= a^2\mathbb{E}[X]\mathbb{E}[Y] = (a\mathbb{E}[X])(a\mathbb{E}[Y]) = \mathbb{E}[aX]\mathbb{E}[aY]. \end{aligned} \quad (3.14)$$

Since W'_t has a multivariate Gaussian distribution, uncorrelation implies independence.

c. Denote $W'_t = W_{1-t} - W_1$ for $t \in [0, 1]$.

1. $W'_0 = W_{1-0} - W_1 = 0$.

2. $W'_v - W'_u = (W_{1-v} - W_1) - (W_{1-u} - W_1) = -(W_{1-u} - W_{1-v}) \in \mathcal{N}(0, v - u)$ for every choice of $0 \leq u < v \leq 1$ given the symmetry of the Gaussian distribution.

3. For any $0 = t_0 < t_1 < \dots < t_n = 1$, $0 = 1 - t_n < 1 - t_{n-1} < \dots < 1 - t_0 = 1$, the random variables

$$\begin{aligned} W'_{t_1} - W'_{t_0} &= -(W_{1-t_1} - W_{1-t_0}), \quad W'_{t_2} - W'_{t_1} = -(W_{1-t_2} - W_{1-t_1}), \\ &\dots, \quad W'_{t_n} - W'_{t_{n-1}} = -(W_{1-t_n} - W_{1-t_{n-1}}) \end{aligned}$$

are independent.

d.

1. $B_0 = 0$ by definition.

2. If $s = 0$, $B_t - B_s = tW_{1/t} \sim \mathcal{N}(0, t)$ for $0 < t < \infty$. $B_t - B_s$ is Gaussian.

$$\mathbb{E}[B_t - B_s] = \mathbb{E}[tW_{1/t}] = t\mathbb{E}[W_{1/t}] = 0. \quad (3.15)$$

$$\text{Var}(B_t - B_s) = \text{Var}(tW_{1/t}) = t^2 \text{Var}(W_{1/t}) = t^2 \cdot \frac{1}{t} = t. \quad (3.16)$$

3. If $s \neq 0$, $B_t - B_s = tW_{1/t} - sW_{1/s} = (t-s)W_{1/t} + s(W_{1/s} - W_{1/t}) = (t-s)W_{1/t} - s(W_{1/s} - W_{1/t}) \sim \mathcal{N}(0, t-s)$ for $0 < s < t < \infty$.

$$\begin{aligned} \mathbb{E}[B_t - B_s] &= \mathbb{E}[(t-s)(W_{1/t} - W_0) + s(W_{1/s} - W_{1/t})] \\ &= (t-s)\mathbb{E}[W_{1/t} - W_0] + s\mathbb{E}[W_{1/s} - W_{1/t}] = 0. \end{aligned} \quad (3.17)$$

$$\begin{aligned} \text{Var}(B_t - B_s) &= \text{Var}((t-s)(W_{1/t} - W_0) + s(W_{1/s} - W_{1/t})) \\ &= (t-s)^2 \text{Var}(W_{1/t} - W_0) + s^2 \text{Var}(W_{1/s} - W_{1/t}) \\ &= (t-s)^2 \cdot \frac{1}{t} + s^2 \cdot \left(\frac{1}{s} - \frac{1}{t}\right) = t-s. \end{aligned} \quad (3.18)$$

4. For any $0 = t_0 < t_1 < \dots < t_n < \infty$, the random variables

$$B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}$$

are independent. Suppose $x = 0$ and let $0 < y < z < \infty$. Then, $0 < \frac{1}{z} < \frac{1}{y} < \infty$ and

$$\begin{aligned} \mathbb{E}[(B_z - B_y)(B_y - B_x)] &= \mathbb{E}[zyW_{1/z}W_{1/y} - y^2W_{1/y}^2] \\ &= zy\mathbb{E}[W_{1/z}W_{1/y}] - y^2\mathbb{E}[W_{1/y}^2] = \frac{zy}{z} - \frac{y^2}{y} = 0. \end{aligned} \quad (3.19)$$

$$\mathbb{E}[(B_z - B_y)]\mathbb{E}[(B_y - B_x)] = (\mathbb{E}[B_z] - \mathbb{E}[B_y])\mathbb{E}[B_y] = 0. \quad (3.20)$$

Suppose $x \neq 0$ and let $0 < x < y < z < \infty$. Then, $0 < \frac{1}{z} < \frac{1}{y} < \frac{1}{x} < \infty$ and

$$\begin{aligned} \mathbb{E}[(B_z - B_y)(B_y - B_x)] &= \mathbb{E}[zyW_{1/z}W_{1/y} - y^2W_{1/y}^2 - zxW_{1/z}W_{1/x} + xyW_{1/x}W_{1/y}] \\ &= zy\mathbb{E}[W_{1/z}W_{1/y}] - y^2\mathbb{E}[W_{1/y}^2] - zx\mathbb{E}[W_{1/z}W_{1/x}] + xy\mathbb{E}[W_{1/x}W_{1/y}] \\ &= \frac{zy}{z} - \frac{y^2}{y} - \frac{zx}{z} + \frac{xy}{y} = 0. \end{aligned} \quad (3.21)$$

$$\mathbb{E}[B_z - B_y]\mathbb{E}[B_y - B_x] = (\mathbb{E}[B_z] - \mathbb{E}[B_y])(\mathbb{E}[B_y] - \mathbb{E}[B_x]) = 0. \quad (3.22)$$

Since B_t has a multivariate Gaussian distribution, uncorrelation implies independence.

□

Theorem 3.11. (*Kolmogorov's Continuity Criterion*) Let X be a stochastic process on some probability space (Ω, \mathcal{F}, P) . If there exist constants α, β , and c such that, for all $t, h \in \mathbb{R}_{++}$

$$\mathbb{E}[|X_{t+h} - X_t|^\alpha] \leq ch^{1+\beta}, \quad (3.23)$$

then there exists a P -a.s. continuous process \tilde{X} (a modification of X) on that space.

Corollary 3.12. Every Brownian motion admits an a.s. continuous modification.

Proof. Let W be a Brownian motion such that $W_{t+h} - W_t \sim \mathcal{N}(0, h)$. Let $\alpha = 4$, $\beta = 1$, and $c = 3$.

$$\mathbb{E}[|W_{t+h} - W_t|^4] = \mathbb{E}[(W_{t+h} - W_t)^4] = 3h^2 \leq 3h^2. \quad (3.24)$$

□

Definition 3.13. A filtration $\{\mathcal{F}_t\}$ on the measurable space (Ω, \mathcal{F}) in any family of sub- σ -fields.

Definition 3.14. Let (Ω, \mathcal{F}, P) be a probability space and $\{\mathcal{F}_t\}$ be a filtration. A stopping time is any function $\tau : \Omega \rightarrow \mathbb{R}_+$ such that $\{\tau \leq t\} \in \mathcal{F}_t$ for any $t \in \mathbb{R}_+$.

Theorem 3.15. If τ is a stopping time, then the random variable ω is $\mathcal{F}_{\tau-}$ -measurable and thus \mathcal{F}_τ -measurable.

Proof. By the definition of stopping time, the event $A = \{\tau > t\} \in \mathcal{F}_t$. By the definition of \mathcal{F}_{t-} , $A \cap \{\tau > t\} = \{\tau > t\} \in \mathcal{F}_{t-} \subseteq \mathcal{F}_t$. □

Theorem 3.16. Let τ and τ' be any two stopping times such that $\tau' \leq \tau$. Then, $\mathcal{F}_{\tau'} \subseteq \mathcal{F}_\tau$ and $\mathcal{F}_{\tau'-} \subseteq \mathcal{F}_{\tau-}$.

Proof. Given that $A \in \mathcal{F}_{\tau'}$ and $\tau' \leq \tau$, for all $t \in \mathbb{R}_+$,

$$A \cap \{\tau \leq t\} = A \cap \{\tau' \leq t\} \cap \{\tau \leq t\} \subseteq A \cap \{\tau' \leq t\} \in \mathcal{F}_\tau \implies A \in \mathcal{F}_\tau. \quad (3.25)$$

Since $A \in \mathcal{F}_{\tau'} \implies A \in \mathcal{F}_\tau$, $\mathcal{F}_{\tau'} \subseteq \mathcal{F}_\tau$. Given that $A \in \mathcal{F}_{\tau'-}$ and $\tau' \leq \tau$, for all $t \in \mathbb{R}_+$,

$$A \cap \{\tau' > t\} = A \cap \{\tau > t\} \cap \{\tau' > t\} \subseteq A \cap \{\tau > t\} \in \mathcal{F}_{\tau-} \implies A \in \mathcal{F}_{\tau-}. \quad (3.26)$$

Since $A \in \mathcal{F}_{\tau'-} \implies A \in \mathcal{F}_{\tau-}$, $\mathcal{F}_{\tau'-} \subseteq \mathcal{F}_{\tau-}$. □

Theorem 3.17. If τ and τ' are stopping times, then $\tau' \wedge \tau$ and $\tau' \vee \tau$ are stopping times.

Proof. Since τ' and τ are stopping times, $\{\tau' \leq t\} \in \mathcal{F}_t$ and $\{\tau \leq t\} \in \mathcal{F}_t$.

$$\{\tau' \wedge \tau \leq t\} = \{\tau' \leq t\} \cup \{\tau \leq t\} \in \mathcal{F}_t. \quad (3.27)$$

$$\{\tau' \vee \tau \leq t\} = \{\tau' \leq t\} \cap \{\tau \leq t\} \in \mathcal{F}_t. \quad (3.28)$$

□

Theorem 3.18. If τ' and τ are stopping times and $A \in \mathcal{F}_{\tau'}$, then

$$A \cap \{\tau' \leq \tau\} \in \mathcal{F}_{\tau'} \cap \mathcal{F}_\tau \subseteq \mathcal{F}_\tau, \quad A \cap \{\tau' < \tau\} \in \mathcal{F}_{\tau'} \cap \mathcal{F}_\tau \subseteq \mathcal{F}_\tau, \quad A \cap \{\tau' = \tau\} \in \mathcal{F}_{\tau'} \cap \mathcal{F}_\tau \subseteq \mathcal{F}_\tau.$$

Proof.

$$\begin{aligned} A \cap \{\tau' \leq \tau\} \cap \{\tau \leq t\} &= (A \cap \{\tau' \leq t\}) \cap \{\tau' \wedge t \leq \tau \wedge t\} \cap \{\tau \leq t\} \\ &\in \mathcal{F}_{\tau'} \cap \{\tau' \wedge t \leq \tau \wedge t\} \cap \mathcal{F}_\tau \subseteq \mathcal{F}_{\tau'} \cap \mathcal{F}_\tau \subseteq \mathcal{F}_\tau. \end{aligned} \quad (3.29)$$

$$\begin{aligned} A \cap \{\tau' < \tau\} \cap \{\tau \leq t\} &= (A \cap \{\tau' \leq t\}) \cap \{\tau' \wedge t < \tau \wedge t\} \cap \{\tau \leq t\} \\ &\in \mathcal{F}_{\tau'} \cap \{\tau' \wedge t < \tau \wedge t\} \cap \mathcal{F}_\tau \subseteq \mathcal{F}_{\tau'} \cap \mathcal{F}_\tau \subseteq \mathcal{F}_\tau. \end{aligned} \quad (3.30)$$

$$\begin{aligned} A \cap \{\tau' = \tau\} \cap \{\tau \leq t\} &= (A \cap \{\tau' \leq t\}) \cap \{\tau' \wedge t = \tau \wedge t\} \cap \{\tau \leq t\} \\ &\in \mathcal{F}_{\tau'} \cap \{\tau' \wedge t = \tau \wedge t\} \cap \mathcal{F}_\tau \subseteq \mathcal{F}_{\tau'} \cap \mathcal{F}_\tau \subseteq \mathcal{F}_\tau. \end{aligned} \quad (3.31)$$

□

Theorem 3.19. *Let τ be a stopping time. $\{\tau < t\} \in \mathcal{F}_{t-} \subseteq \mathcal{F}_t$ for any $t \in \mathbb{R}_+$ and therefore $\{\tau = t\} \in \mathcal{F}_t$. If \mathcal{F} is right-continuous, then $\{\tau < t\} \in \mathcal{F}_t$ implies $\{\tau \leq t\} \in \mathcal{F}_t$ for any $t \in \mathbb{R}_+$.*

Proof.

$$\{\tau < t\} = \bigcup_{n \in \mathbb{N}_{++}} \{\tau \leq t - \frac{1}{n}\} = \mathcal{F}_0 \vee \bigcap_{n \in \mathbb{N}_{++}} \{\tau \geq t - \frac{1}{n}\} = \mathcal{F}_0 \vee \{\tau > t\} \in \mathcal{F}_{t-} \subseteq \mathcal{F}_t. \quad (3.32)$$

$$\{\tau = t\} = \{\tau \leq t\} \setminus \{\tau < t\} \in \mathcal{F}_t. \quad (3.33)$$

$$\{\tau < t + \frac{1}{n}\} \in \mathcal{F}_t \implies \bigcap_{n \in \mathbb{N}_{++}} \{\tau < t + \frac{1}{n}\} \in \mathcal{F}_t \implies \{\tau \leq t\} \in \mathcal{F}_t. \quad (3.34)$$

□

Theorem 3.20. *Let $\phi \in \mathcal{C}_K^\infty(\mathbb{R}; \mathbb{R})$ and $x \in \mathbb{R}$. The infinitesimal generator of the Brownian motion is*

$$\begin{aligned} \mathcal{G}\phi(x) &= \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} \mathbb{E}[\phi(W_{t+\epsilon}) - \phi(W_t) | W_t = x] = \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} \mathbb{E}[\phi(W_{t+\epsilon}) - \phi(W_t)] \\ &= \lim_{\epsilon \rightarrow 0^+} \mathbb{E}\left[\frac{\phi(W_{t+\epsilon}) - \phi(W_t)}{\epsilon}\right] = \mathbb{E}[\partial\phi(W_t) | W_t = x] = \mathbb{E}[\partial\phi(x)] = \int_{\mathbb{R}} \partial\phi(x) dx = \frac{1}{2} \partial^2 \phi(x). \end{aligned} \quad (3.35)$$

Definition 3.21. *Let $\mathcal{P} = \{t_0, t_1, \dots, t_n\}$ be any partition of the interval $[0, t]$. The variation of a function $f : [0, t] \rightarrow \mathbb{R}$ is*

$$\mathcal{V}^{\mathcal{P}}(f) = \sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)|. \quad (3.36)$$

The total variation is the supremum of all variations of a function $f : I \rightarrow \mathbb{R}$ such that

$$\mathcal{T}^{\mathcal{P}}(f) = \sup_{\mathcal{P}} \sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)|.$$

Remark 3.22. The variation is increasing with respect to \mathcal{P} . f has a finite variation on $[0, t]$ if $\mathcal{V}^{\mathcal{P}}(f)_t < \infty$. f has a bounded variation if $\lim_{t \rightarrow \infty} \mathcal{V}^{\mathcal{P}}(f)_t < \infty$.

Definition 3.23. Let $\mathcal{P} = \{t_0, t_1, \dots, t_n\}$ be any partition of the interval $[0, t]$. The quadratic variation of a function $f : [0, t] \rightarrow \mathbb{R}$ is

$$\mathcal{Q}^{\mathcal{P}}(f, f) = \sum_{i=0}^{n-1} (f(t_{i+1}) - f(t_i))^2. \quad (3.37)$$

Remark 3.24. The total quadratic variation is not defined here since it is not increasing. f has a finite quadratic variation if $\lim_{n \rightarrow \infty} \mathcal{Q}^{\mathcal{P}}(f, f) < \infty$ for any $t \in \mathbb{R}_{++}$.

Lemma 3.25. Since the Brownian process from $u = 0$ to $u = s$ is independent of the one from $u = s$ to $u = t$,

$$\begin{aligned} \mathbb{E} \left[\int_0^s W_{\omega}(u) du \int_s^t W_{\omega}(u) du \right] &= \mathbb{E} \left[\int_0^s W_{\omega}(u) du \right] \mathbb{E} \left[\int_s^t W_{\omega}(u) du \right] \\ &= \int_0^s \mathbb{E} [W_{\omega}(u)] du \int_s^t \mathbb{E} [W_{\omega}(u)] du = 0. \end{aligned} \quad (3.38)$$

Theorem 3.26. Let W be a Brownian motion and, for any $t \in \mathbb{R}_+$,

$$X_t = \int_0^t W(s) ds.$$

Then, $X_t \sim \mathcal{N}(0, \frac{t^3}{3})$.

Proof.

$$\mathbb{E}[X_t] = \mathbb{E} \left[\int_0^t W_{\omega}(s) ds \right] = \int_0^t \mathbb{E} [W_{\omega}(s)] ds = \int_0^t 0 ds = 0. \quad (3.39)$$

Let $0 \leq s < t < \infty$ without loss of generality.

$$\begin{aligned} C(s, t) &= \mathbb{E}[X_s X_t] = \mathbb{E} \left[\int_0^s W_{\omega}(u) du \int_0^t W_{\omega}(u) du \right] = \mathbb{E} \left[\left(\int_0^s W_{\omega}(u) du \right)^2 \right] \\ &= \mathbb{E} \left[\int_0^s \int_0^s W_{\omega}(u) W_{\omega}(v) dudv \right] = \int_0^s \int_0^s \mathbb{E} [W_{\omega}(u) W_{\omega}(v)] dudv \\ &= \int_0^s \int_0^s u \wedge v dudv = \int_0^s \int_0^v u dudv + \int_0^s \int_0^u v dvdu = \frac{s^3}{3}. \end{aligned} \quad (3.40)$$

If $s \geq t$, $C(s, t) = \frac{t^2}{2}$ by the similar argument and thus $C(s, t) = \frac{s^3}{3} \wedge \frac{t^3}{3}$. \square

Remark 3.27. X_t is not a martingale.

Lemma 3.28. Let $\mathcal{P} = \{t_0, t_1, \dots, t_n\}$ be a partition of $[0, t]$. Then,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\left(\sum_{i=0}^{n-1} ((W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i)) \right)^2 \right] = 0. \quad (3.41)$$

Proof. Given the Brownian motion W , $W_{t_{i+1}} - W_{t_i} \sim N(0, t_{i+1} - t_i)$. As $t_{i+1} - t_i \rightarrow 0$,

$$\begin{aligned}
 & \mathbb{E} \left[\left(\sum_{i \in \mathbb{N}} ((W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i)) \right)^2 \right] \\
 &= \mathbb{E} \left[\sum_{i \in \mathbb{N}} (W_{t_{i+1}} - W_{t_i})^2 - \sum_{i \in \mathbb{N}} (t_{i+1} - t_i) \right]^2 + \text{Var} \left(\sum_{i \in \mathbb{N}} (W_{t_{i+1}} - W_{t_i})^2 - \sum_{i \in \mathbb{N}} (t_{i+1} - t_i) \right) \\
 &= \sum_{i \in \mathbb{N}} \text{Var}((W_{t_{i+1}} - W_{t_i})^2) = \sum_{i \in \mathbb{N}} \left(\mathbb{E}[(W_{t_{i+1}} - W_{t_i})^4] - \mathbb{E}[(W_{t_{i+1}} - W_{t_i})^2]^2 \right) \\
 &= \sum_{i \in \mathbb{N}} \left(\mathbb{E}[(W_{t_{i+1}} - W_{t_i})^4] - \left(\text{Var}(W_{t_{i+1}} - W_{t_i}) + \mathbb{E}[W_{t_{i+1}} - W_{t_i}]^2 \right)^2 \right) \\
 &= \sum_{i \in \mathbb{N}} (3(t_{i+1} - t_i)^2 - (t_{i+1} - t_i)^2) = 2 \sum_{i \in \mathbb{N}} (t_{i+1} - t_i)^2. \quad (3.42)
 \end{aligned}$$

□

Remark 3.29. It is trivial that $\sum_{i \in \mathbb{N}} (t_{i+1} - t_i) = t$. Since $W_{t_1} - W_{t_0}, W_{t_2} - W_{t_1}, \dots, W_{t_n} - W_{t_{n-1}}$ are independent, $(W_{t_1} - W_{t_0})^2, (W_{t_2} - W_{t_1})^2, \dots, (W_{t_n} - W_{t_{n-1}})^2$ are independent. Recall that $W_{t_{i+1}} - W_{t_i} \sim N(0, t_{i+1} - t_i)$.

$$\begin{aligned}
 \mathbb{E} \left[\sum_{i \in \mathbb{N}} (W_{t_{i+1}} - W_{t_i})^2 \right] &= \sum_{i \in \mathbb{N}} \mathbb{E}[(W_{t_{i+1}} - W_{t_i})^2] \\
 &= \sum_{i \in \mathbb{N}} \mathbb{E}[(W_{t_{i+1}} - W_{t_i})]^2 + \text{Var}(W_{t_{i+1}} - W_{t_i}) = \sum_{i \in \mathbb{N}} (t_{i+1} - t_i) = t. \quad (3.43)
 \end{aligned}$$

Additionally, $\mathbb{E}[X^4] = 3\sigma^4$ for $X \sim N(0, \sigma^2)$.

Theorem 3.30. The Brownian motion W has finite quadratic variation and $[W, W] = t$.

Proof. Since the expectation of the ordinary least squares of the difference between the quadratic variance of the Brownian motion and the corresponding time interval can be arbitrarily small as the time interval is sufficiently small. Thus,

$$[W, W] = \lim_{n \rightarrow \infty} \mathcal{Q}^n(X_\omega, X_\omega)_t = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} (W_{t_{i+1}} - W_{t_i})^2 = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} (t_{i+1} - t_i) = t. \quad (3.44)$$

□

Theorem 3.31. Let X and Y be any two stochastic processes such that X is continuous a.s. and Y has finite variation on every finite interval. $[X, Y] = 0$.

Proof. Since Y has finite variance, for sufficiently small $\Delta t = t_{i+1} - t_i$, there exists a constant c such that

$$\lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} |Y_{t_{i+1}} - Y_{t_i}| \leq c. \quad (3.45)$$

Since X is continuous almost surely, for sufficiently small $\Delta t = t_{i+1} - t_i$, there exists $\frac{\epsilon}{c} > 0$ such that

$$|X_{t_{i+1}} - X_{t_i}| < \frac{\epsilon}{c}. \quad (3.46)$$

Thus,

$$\begin{aligned} [X, Y] &= \lim_{n \rightarrow \infty} \mathcal{Q}^{\mathcal{P}_n}(X_\omega, Y_\omega)_t = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} (X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i}) \\ &\leq \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} |X_{t_{i+1}} - X_{t_i}| \cdot |Y_{t_{i+1}} - Y_{t_i}| < \frac{\epsilon}{c} \cdot c = \epsilon. \end{aligned} \quad (3.47)$$

□

Theorem 3.32. (Law of the Iterated Logarithm) Let W be a Brownian motion.

$$P \left[\limsup_{t \rightarrow 0^+} \frac{W_t}{\sqrt{2t \log(\log(\frac{1}{t}))}} = 1 \right] = 1. \quad (3.48)$$

Corollary 3.33. Let W be a Brownian motion. $-W$ and $\pm W_{1/t}$ are also Brownian motions.

$$P \left[\liminf_{t \rightarrow 0^+} \frac{W_t}{\sqrt{2t \log(\log(\frac{1}{t}))}} = -1 \right] = P \left[\limsup_{t \rightarrow 0^+} \frac{-W_t}{\sqrt{2t \log(\log(\frac{1}{t}))}} = 1 \right] = 1. \quad (3.49)$$

$$P \left[\limsup_{t \rightarrow \infty} \frac{W_t}{\sqrt{2t \log(\log(t))}} = 1 \right] = P \left[\limsup_{1/t \rightarrow 0^+} \frac{W_{1/t}}{\sqrt{2t \log(\log(\frac{1}{t}))}} = 1 \right] = 1. \quad (3.50)$$

$$P \left[\liminf_{t \rightarrow \infty} \frac{W_t}{\sqrt{2t \log(\log(t))}} = -1 \right] = P \left[\limsup_{1/t \rightarrow 0^+} \frac{-W_{1/t}}{\sqrt{2t \log(\log(\frac{1}{t}))}} = -1 \right] = 1. \quad (3.51)$$

Definition 3.34. The running maximum of the Brownian motion W is

$$\mathcal{S}_t(W) = \sup_{s \in [0, t]} W_s. \quad (3.52)$$

Theorem 3.35. (Reflection Principle) Let $b \in \mathbb{R}_{++}$ and $a \in]-\infty, b]$. Then,

$$P(\mathcal{S}_t \geq b, W_t < a) = P(W_t > 2b - a). \quad (3.53)$$

Corollary 3.36. For every fixed $t \in \mathbb{R}_{++}$ and all $b \in \mathbb{R}_{++}$,

$$P(\mathcal{S}_t \geq b) = P(|W_t| \geq b). \quad (3.54)$$

Proof.

$$P(\mathcal{S}_t \geq b) = P(W_t \geq b) + P(\mathcal{S}_t \geq b, W_t < b) = P(W_t \geq b) + P(W_t \geq b) = P(|W_t| \geq b). \quad (3.55)$$

□

Theorem 3.37. *For the domain $\mathcal{D} = \{(a, b) \in \mathbb{R}^2 : b > 0, a \leq b\}$, the joint density of the Brownian motion and its running maximum at time t is*

$$\varphi(a, b) = \sqrt{\frac{2}{\pi t^3}} (2b - a) e^{-\frac{(a-2b)^2}{2t}}. \quad (3.56)$$

Proof.

$$\begin{aligned} \varphi(a, b) &= \frac{\partial^2}{\partial a \partial b} (P(\mathcal{S}_t \geq b, W_t < a)) = \frac{\partial^2}{\partial a \partial b} \left(\frac{1}{\sqrt{2\pi t}} \int_{-\infty}^a e^{-\frac{(x-2b)^2}{2t}} dx \right) \\ &= \frac{\partial}{\partial b} \left(\frac{1}{\sqrt{2\pi t}} e^{-\frac{(a-2b)^2}{2t}} \right) = \sqrt{\frac{2}{\pi t^3}} (a - 2b) e^{-\frac{(a-2b)^2}{2t}}. \end{aligned} \quad (3.57)$$

□

Theorem 3.38. *For fixed $b \in \mathbb{R}_{++}$, the distribution density of the stopping τ_b in \mathbb{R}_{++} is*

$$\varphi_b(t) = \frac{b}{\sqrt{2\pi t^3}} e^{-\frac{b^2}{2t}}. \quad (3.58)$$

Proof.

$$\varphi_b(t) = \frac{\partial}{\partial t} P(t_b \leq t) = \frac{\partial}{\partial t} \left(\sqrt{\frac{2}{\pi t}} \int_b^\infty e^{-\frac{x^2}{2t}} dt \right) = \frac{b}{\sqrt{2\pi t^3}} e^{-\frac{b^2}{2t}}. \quad (3.59)$$

□

SECTION 4

Martingale

Definition 4.1. (Martingale) Let $0 \leq s < t$. A real-valued process X is a martingale if

1. X is adapted to \mathcal{F} .
2. $\mathbb{E}[|X_t|] < \infty$.
 - a. (Submartingale) $\mathbb{E}[X_t^+] < \infty$.
 - b. (Supermartingale) $\mathbb{E}[X_t^-] < \infty$.
3. $X_s = \mathbb{E}[X_t | \mathcal{F}_s]$.
 - a. (Submartingale) $X_s \leq \mathbb{E}[X_t | \mathcal{F}_s]$.
 - b. (Supermartingale) $X_s \geq \mathbb{E}[X_t | \mathcal{F}_s]$.

Theorem 4.2. If X is a supermartingale such that $\mathbb{E}[X_T] = \mathbb{E}[X_0]$ for some $T \in \mathbb{R}_{++}$, then $(X_t)_{t \in [0, T]}$ is a martingale.

Proof. Since X is a supermartingale, $X_0 \geq \mathbb{E}[X_t | \mathcal{F}_0]$ and $X_T \geq \mathbb{E}[X_T | \mathcal{F}_t]$. Given that $\mathcal{F}_0 \subseteq \mathcal{F}_t$,

$$X_0 \geq \mathbb{E}[X_t | \mathcal{F}_0] \geq \mathbb{E}[\mathbb{E}[X_T | \mathcal{F}_t] | \mathcal{F}_0] = \mathbb{E}[X_T | \mathcal{F}_0]. \quad (4.1)$$

Since $\mathbb{E}[X_T] = \mathbb{E}[X_0]$,

$$\mathbb{E}[X_T | \mathcal{F}_0] = \mathbb{E}[X_0 | \mathcal{F}_0] = X_0. \quad (4.2)$$

Thus, for every $t \in [0, T]$,

$$\mathbb{E}[X_t | \mathcal{F}_0] = \mathbb{E}[\mathbb{E}[X_T | \mathcal{F}_t] | \mathcal{F}_0] \implies \mathbb{E}[\mathbb{E}[X_t | \mathcal{F}_0]] = \mathbb{E}[\mathbb{E}[\mathbb{E}[X_T | \mathcal{F}_t] | \mathcal{F}_0]] \implies X_t = \mathbb{E}[X_T | \mathcal{F}_t]. \quad (4.3)$$

□

Lemma 4.3. Let W be a Brownian motion. Then, for $0 \leq s < t$,

$$\mathbb{E}[e^{\sigma(W_t - W_s)}] = e^{\frac{1}{2}\sigma^2(t-s)}. \quad (4.4)$$

Proof. $W_t - W_s \sim N(0, t-s) \implies \sigma(W_t - W_s) \sim N(0, \sigma^2(t-s))$.

$$\begin{aligned} \mathbb{E}[e^{\sigma(W_t - W_s)}] &= \int_{-\infty}^{\infty} e^x \frac{1}{\sqrt{2\pi}\sigma\sqrt{t-s}} e^{-\frac{x^2}{2\sigma^2(t-s)}} dx \\ &= e^{\frac{1}{2}\sigma^2(t-s)} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma\sqrt{t-s}} e^{-\frac{(x-2\sigma^2(t-s))^2}{2\sigma^2(t-s)}} dx = e^{\frac{1}{2}\sigma^2(t-s)}. \end{aligned} \quad (4.5)$$

□

Theorem 4.4. Let W be a Brownian motion. For any $\sigma \in \mathbb{R}_+$ and $0 \leq s < t$,

- a. σW is a martingale.
- b. $\sigma W + \sigma(t-s)$ is a submartingale and $\sigma W - \sigma(t-s)$ is a supermartingale.
- c. $\sigma^2 W^2$ and $e^{\sigma W}$ are submartingales.

d. $\sigma^2(W^2 - (t - s)^2)$ is a martingale.

e. $e^{\sigma W - \frac{\sigma^2(t-s)}{2}}$ is a martingale.

Proof. a. σW is adapted to \mathcal{F}^W .

$$\mathbb{E}[|\sigma W|] = \sigma \mathbb{E}[|W|] = \sigma \sqrt{\frac{2t}{\pi}} < \infty. \quad (4.6)$$

$$\mathbb{E}[\sigma W_t | \mathcal{F}_s^W] = \mathbb{E}[\sigma(W_t - W_s) | \mathcal{F}_s^W] + \mathbb{E}[\sigma W_s | \mathcal{F}_s^W] = \mathbb{E}[\sigma W_s | \mathcal{F}_s^W] = \sigma W_s. \quad (4.7)$$

b. $\sigma W \pm \sigma(t - s)$ is adapted to \mathcal{F}^W .

$$\mathbb{E}[|\sigma W \pm \sigma(t - s)|] \leq \sigma \mathbb{E}[|W|] + \sigma(t - s) = \sigma \sqrt{\frac{2t}{\pi}} + \sigma(t - s) < \infty. \quad (4.8)$$

$$\mathbb{E}[W_t + \sigma(t - s) | \mathcal{F}_s^W] = \mathbb{E}[W_t | \mathcal{F}_s^W] + \sigma(t - s) = \sigma W_s + \sigma(t - s) \geq \sigma W_s. \quad (4.9)$$

$$\mathbb{E}[W_t - \sigma(t - s) | \mathcal{F}_s^W] = \mathbb{E}[W_t | \mathcal{F}_s^W] - \sigma(t - s) = \sigma W_s - \sigma(t - s) \leq \sigma W_s. \quad (4.10)$$

c. $\sigma^2 W^2$ and $e^{\sigma W}$ is adapted to \mathcal{F}^W .

$$\mathbb{E}[|\sigma^2 W_t^2|] = \sigma^2 \mathbb{E}[W_t^2] = \sigma^2(\mathbb{E}[W_t]^2 + \text{Var}(W_t)) = \sigma^2 \text{Var}(W_t) = \sigma^2 t < \infty. \quad (4.11)$$

$$\mathbb{E}[|e^{\sigma W_t}|] = \mathbb{E}[(e^\sigma)^{W_t}] = e^{\frac{1}{2}\sigma^2 t} < \infty. \quad (4.12)$$

$$\begin{aligned} \mathbb{E}[\sigma^2 W_t^2 | \mathcal{F}_s^W] &= \sigma^2 \mathbb{E}[(W_t - W_s + W_s)^2 | \mathcal{F}_s^W] \\ &= \sigma^2(\mathbb{E}[(W_t - W_s)^2 | \mathcal{F}_s^W] + 2\mathbb{E}[(W_t - W_s)W_s | \mathcal{F}_s^W] + \mathbb{E}[W_s^2 | \mathcal{F}_s^W]) \\ &= \sigma^2(t - s) + \sigma^2 W_s^2 \geq \sigma^2 W_s^2. \end{aligned} \quad (4.13)$$

$$\begin{aligned} \mathbb{E}[e^{\sigma W_t} | \mathcal{F}_s^W] &= \mathbb{E}[e^{\sigma(W_t - W_s) + \sigma W_s} | \mathcal{F}_s^W] = e^{\sigma W_s} \mathbb{E}[e^{\sigma(W_t - W_s)} | \mathcal{F}_s^W] \\ &= e^{\sigma W_s} \mathbb{E}[e^{\sigma(W_t - W_s)}] = e^{\sigma W_s + \frac{1}{2}\sigma^2(t-s)} \geq e^{\sigma W_s}. \end{aligned} \quad (4.14)$$

d. $\sigma^2 W^2 - \sigma^2(t - s)$ is adapted to \mathcal{F}^W .

$$\mathbb{E}[|\sigma^2 W_t^2 - \sigma^2(t - s)|] \leq \mathbb{E}[\sigma^2 W_t^2] + \sigma^2(t - s) = \sigma^2 t + \sigma^2(t - s) = \sigma^2(2t - s) < \infty. \quad (4.15)$$

$$\mathbb{E}[\sigma^2 W_t^2 - \sigma^2(t - s) | \mathcal{F}_s^W] = \mathbb{E}[\sigma^2 W_t^2 | \mathcal{F}_s^W] - \sigma^2(t - s) = \sigma^2 W_s^2. \quad (4.16)$$

e. $e^{\sigma W - \frac{1}{2}\sigma^2(t-s)}$ is adapted to \mathcal{F}^W .

$$\mathbb{E}[|e^{\sigma W_t - \frac{1}{2}\sigma^2(t-s)}|] = e^{-\frac{1}{2}\sigma^2(t-s)} \mathbb{E}[e^{\sigma W_t}] = e^{-\frac{1}{2}\sigma^2(t-s) + \frac{1}{2}\sigma^2 t} = e^{\frac{1}{2}\sigma^2 s} < \infty \quad (4.17)$$

$$\mathbb{E}[e^{\sigma W_t - \frac{1}{2}\sigma^2(t-s)} | \mathcal{F}_s^W] = e^{-\frac{1}{2}\sigma^2(t-s)} \mathbb{E}[e^{\sigma W_t} | \mathcal{F}_s^W] = e^{-\frac{1}{2}\sigma^2(t-s)} e^{\sigma W_s + \frac{1}{2}\sigma^2(t-s)} = e^{\sigma W_s}. \quad (4.18)$$

□

Theorem 4.5. *The exponential distribution lacks memory.*

Proof. For any $t, s \in \mathbb{R}_+$ with $s \leq t$,

$$P(\xi > t | \xi > s) = \frac{P(\{\xi > t\} \cap \{\xi > s\})}{P(\xi > s)} = \frac{P(\xi > t)}{P(\xi > s)} = \frac{e^{-ct}}{e^{-cs}} = e^{-c(t-s)} = P(\xi > t-s). \quad (4.19)$$

□

Lemma 4.6.

$$\begin{aligned} \int_{\mathbb{R}} (z-y)^{m-1} y^{n-1} dy &= \int_{\mathbb{R}} \sum_{k=0}^{m-1} \frac{\Gamma(m)}{\Gamma(k+1) \cdot \Gamma(m-k)} z^{m-1-k} y^k y^{n-1} dy \\ &= \sum_{k=0}^{m-1} \frac{\Gamma(m)}{(k+n) \cdot \Gamma(k+1) \cdot \Gamma(m-k)} z^{m+n-1} = \frac{\Gamma(m) \cdot \Gamma(n)}{\Gamma(m+n)} z^{m+n-1}. \end{aligned} \quad (4.20)$$

Theorem 4.7. *Let $\xi \sim \Gamma(c, m)$ and $\eta \sim \Gamma(c, n)$. Then, $\xi + \eta \sim \Gamma(c, m+n)$.*

Proof. The density of $\xi + \eta$ is given by the convolution of the two individual densities.

$$\begin{aligned} (\xi * \eta)(z) &= \int_{\mathbb{R}} \xi(z-y) \eta(y) dy = \int_{\mathbb{R}} \frac{c^m}{\Gamma(m)} (z-y)^{m-1} e^{-c(z-y)} \cdot \frac{c^n}{\Gamma(n)} y^{n-1} e^{-cy} dy \\ &= \frac{c^{m+n} e^{-cz}}{\Gamma(m) \cdot \Gamma(n)} \int_{\mathbb{R}} (z-y)^{m-1} y^{n-1} dy = \frac{c^{m+n} e^{-cz}}{\Gamma(m+n)} z^{m+n-1}. \end{aligned} \quad (4.21)$$

□

Theorem 4.8. *Let (ξ_i) be any sequence of i.i.d. random variables such that $P(\xi > t) = e^{-ct}$ for $t \in \mathbb{R}_+$ and $\tau_n = \sum_{i=1}^n \xi_i$. For any fixed $t \in \mathbb{R}_{++}$, $P(\tau_n \leq t) = 0$ for $n \in \mathbb{N}_{++}$.*

Proof.

$$\begin{aligned} \sum_{n \in \mathbb{N}} P(\tau_n < t) &= \sum_{n=1}^{\infty} \int_0^t \frac{c^n x^{n-1}}{\Gamma(n)} e^{-cx} dx = \int_0^t \left(\sum_{n=1}^{\infty} \frac{(cx)^{n-1}}{\Gamma(n)} \right) \cdot ce^{-cx} dx \\ &= \int_0^t e^{cx} \cdot ce^{-cx} dx = ct. \end{aligned} \quad (4.22)$$

Given the first Borel-Cantelli Lemma,

$$\sum_{n \in \mathbb{N}} P(\tau_n < t) = ct < \infty \implies P(\{\tau_n < t\} \text{ i.o.}) = 0. \quad (4.23)$$

□

Theorem 4.9. *If N is a Poisson process of parameter $c \in \mathbb{R}_{++}$, then for any fixed $t \in \mathbb{R}_{++}$, the random variable N_t is a Poisson process of parameter ct .*

Proof.

$$\begin{aligned} P(N_t = k) &= P(\tau_k \leq t, \tau_{k+1} > t) = P(\tau_k \leq t, \xi_{k+1} > t - \tau_k) \\ &= \int_0^t \frac{c^k x^{k-1}}{(k-1)!} e^{-cx} \cdot e^{-c(t-x)} dx = e^{-ct} \frac{(ct)^k}{k!}. \end{aligned} \quad (4.24)$$

□

Theorem 4.10. *Let W be a Brownian motion.*

$$\lim_{t \rightarrow \infty} e^{W_t - \frac{t}{2}} = 0. \quad (4.25)$$

Proof. Given the squeeze theorem,

$$\begin{aligned} \liminf_{t \rightarrow \infty} e^{-\sqrt{2t \log(\log(t))} - \frac{t}{2}} &\leq \lim_{t \rightarrow \infty} e^{W_t - \frac{t}{2}} \leq \limsup_{t \rightarrow \infty} e^{\sqrt{2t \log(\log(t))} - \frac{t}{2}} \\ &\implies e^{-\infty} \leq \lim_{t \rightarrow \infty} e^{W_t - \frac{t}{2}} \leq e^{-\infty} \implies \lim_{t \rightarrow \infty} e^{W_t - \frac{t}{2}} = 0. \end{aligned} \quad (4.26)$$

□

Theorem 4.11. *(Doob's Maximal Inequality) If X is a right-continuous submartingale and $S_t = \sup_{s \in [0, t]} X_s$ is its running maximum, for any $\lambda \in \mathbb{R}_{++}$ and $t \in \mathbb{R}_+$,*

$$\lambda^p P(S_t \geq \lambda^p) \leq \mathbb{E}[X_t^+]. \quad (4.27)$$

Theorem 4.12. *(Doob's Convergence Theorem) Let X be any right-continuous submartingale with $\sup_{t \in \mathbb{R}_+} \mathbb{E}[X_t^+] < \infty$ (or supermartingale with $\sup_{t \in \mathbb{R}_+} \mathbb{E}[X_t^-] < \infty$). Then, there exists a measurable random variable $X_\infty < \infty$ such that $\lim_{t \rightarrow \infty} X_t = X_\infty$.*

Theorem 4.13. *Let W be a Brownian motion. Then,*

$$\sup_{t \in \mathbb{R}_+} \mathbb{E}[W_t^+] = \infty. \quad (4.28)$$

Proof. Given the Doob's maximal theorem,

$$\begin{aligned} \mathbb{E}[W_t^+] &\geq \lambda^p P(S_t \geq \lambda^p) = 2\lambda^p P(|W_t| \geq \lambda^p) \\ &= \frac{2\lambda^p}{\sqrt{2\pi t}} \int_{-\infty}^{\lambda^p} e^{-\frac{(x-2\lambda^p)^2}{2t}} dx = \lambda^p \left(1 - \frac{2}{\sqrt{\pi}} \int_0^{\frac{\lambda^p}{\sqrt{2t}}} e^{-x^2} dx \right). \end{aligned} \quad (4.29)$$

Given that

$$\lim_{t \rightarrow \infty} \int_0^{\frac{\lambda^p}{\sqrt{2t}}} e^{-x^2} dx = 0 \quad (4.30)$$

and λ can be arbitrarily large, $\mathbb{E}[W_t^+]$ is unbounded above and thus $\sup_{t \in \mathbb{R}_+} \mathbb{E}[W_t^+] = \infty$. □

Theorem 4.14. *Let X be a martingale. The following conditions are equivalent.*

1. *The family of random variables X_t is uniformly integrable.*
2. *The limit exists.*
3. *There exists an integrable random variable X_∞ such that $X_t = \mathbb{E}[X_\infty | \mathcal{F}_t]$ for any $t \in \mathbb{R}_+$.*

Theorem 4.15. *(Optional Stopping Theorem) Let τ and τ' be any two stopping times such that $P(\tau' \leq \tau) = 1$.*

1. *If X be a right-continuous and integrable submartingale, the random variables X_τ and $X_{\tau'}$ are both integrable and $X_{\tau'} \leq \mathbb{E}[X_\tau | \mathcal{F}_s]$.*

2. If X be a right-continuous and integrable supermartingale, the random variables X_τ and $X_{\tau'}$ are both integrable and $X_{\tau'} \geq \mathbb{E}[X_\tau | \mathcal{F}_s]$.
3. If X be a right-continuous and uniformly integrable martingale, the random variables X_τ and $X_{\tau'}$ are both uniformly integrable and $X_{\tau'} = \mathbb{E}[X_\tau | \mathcal{F}_s]$.
 - Since X is uniformly integrable, the collection of all random variables that can be expressed as conditional expectation $\mathbb{E}[X | \mathcal{F}]$ for $\mathcal{F} \subseteq \mathcal{F}$ is uniformly integrable. Given any two \mathcal{F} -stopping times τ' and τ with $P(\tau' \leq \tau) = 1$, since X is a uniformly integrable and right-continuous martingale,

$$X_{\tau'} = \mathbb{E}[X_\tau | \mathcal{F}_{\tau'}] = \mathbb{E}[X_\infty | \mathcal{F}_{\tau'}]. \quad (4.31)$$

Definition 4.16. Let M be an adapted process. M is a local martingale if there exists a sequence of stopping times (τ_i) such that

1. $P(\tau_i < \tau_{i+1}) = 1$.
2. $P(\lim_{i \rightarrow \infty} \tau_i = \infty) = 1$.
3. $M_{t \wedge \tau_i}$ is a martingale.

Theorem 4.17. a. Every martingale is a local martingale.

- b. The sum or the difference of two local martingales is a local martingale.
- c. A local martingale stopped at any stopping time is still a local martingale.
- d. A postive local martingale is a supermartingale.

Proof. a. For any right-continuous martingale M , M is also a local martingale since there exists a sequence of stopping times $(\tau_i = i)_{i \in \mathbb{N}}$ such that

1. $\tau_i = i \leq i + 1 = \tau_{i+1}$.
2. $\lim_{i \rightarrow \infty} \tau_i = \infty$.
3. $1_{\{\tau_i > 0\}} M^{\tau_i}$ is a martingale.

b. Let M and N be two local martingales with the localizing sequences τ and τ' . The sum or difference of two local martingales is a local martingale since there exist a sequence of stopping times $\tau \wedge \tau'$ such that

1. $\tau_i \leq \tau_{i+1}, \tau'_i \leq \tau'_{i+1} \implies \tau_i \wedge \tau'_i \leq \tau_{i+1} \wedge \tau'_{i+1}$.
2. $\lim_i \tau_i = \infty, \lim_i \tau'_i = \infty \implies \lim_i (\tau_i \wedge \tau'_i) = \infty$.
3. $1_{\{\tau_i + \tau'_i > 0\}} M^{\tau_i + \tau'_i}$ is a martingale.

c. Let M be a local martingale and M' be M stopped at the stopping time τ' . M' is a local martingale since there exists a sequence $\tau_i \wedge \tau'$ such that,

1. $\tau_i \wedge \tau' \leq \tau_{i+1} \wedge \tau'$.
2. $\lim_{i \rightarrow \infty} \tau_i \wedge \tau' = \infty$.

3. $1_{\tau_i \wedge \tau' > 0} M^{\tau_i \wedge \tau'}$ is a martingale.

d. Let M be a local martingale with the sequence τ .

$$M_s = \lim_{n \rightarrow \infty} X_s^{\tau_n} = \lim_{n \rightarrow \infty} \mathbb{E}[X_t^{\tau_n} | \mathcal{F}_s] \geq \mathbb{E}[\lim_{n \rightarrow \infty} X_t^{\tau_n} | \mathcal{F}_s] = \mathbb{E}[X_t | \mathcal{F}_s]. \quad (4.32)$$

Thus, M is a supermartingale. \square

Lemma 4.18. *Let (Ω, \mathcal{F}, P) be a probability space and Q be another probability measure such that $Q \precsim P$. Let \mathcal{F}_t be a filtration inside \mathcal{F} and $R_t = dQ/dP$.*

$$\mathbb{E}^P[R_t | \mathcal{F}_s] = R_s. \quad (4.33)$$

Theorem 4.19. *A process M is an (\mathcal{F}, Q) -martingale if and only if the process MR is an (\mathcal{F}, Q) -martingale.*

Proof.

$$E^P[M_t R_t | \mathcal{F}_s] = M_s R_s \iff E^Q[M_t | \mathcal{F}_s] = \frac{E^P[M_t R_t | \mathcal{F}_s]}{E^P[R_t | \mathcal{F}_s]} = M_s. \quad (4.34)$$

\square

Theorem 4.20. *A process M is a local (\mathcal{F}, Q) -martingale if and only if the process MR is a local (\mathcal{F}, Q) -martingale.*

Definition 4.21. *A real-valued continuous process X is a semimartingale if it can be decomposed into a continuous local martingale denoted as M and a continuous adapted process denoted as A that starts from 0 and has paths of finite variation on finite intervals.*

Theorem 4.22. *Any two continuous semimartingales have finite quadratic covariation.*

SECTION 5

Stochastic Calculus

Definition 5.1. Let \mathcal{C} be the class of functions $f(t, \omega) : [0, \infty] \times \Omega \rightarrow \mathbb{R}$ such that

1. $f(t, \omega)$ is $\mathcal{B} \otimes \mathcal{F}_t$ -measurable.
2. There exists an increasing family of σ -field \mathcal{H}_t such that B_t is a martingale of \mathcal{H}_t and $f(t, \omega)$ is \mathcal{H}_t -adapted.
3. $\mathbb{E} \left[\int f(t, \omega)^2 dt \right] < \infty$.

Definition 5.2. (Itô Integral) Let $f(t, \omega) \in \mathcal{C}$. The Itô integral of $f(t, \omega)$ is

$$\int f(t, \omega) dW_t(\omega) = \lim_{n \rightarrow \infty} \int \varphi_n(t, \omega) dW_t(\omega) \in \mathcal{L}_{loc}^2 \quad (5.1)$$

where $\{\varphi_n\}$ is a sequence of elementary functions such that

$$\mathbb{E} \left[\lim_{n \rightarrow \infty} \int (f(t, \omega) - \varphi_n(t, \omega))^2 dt \right] = 0. \quad (5.2)$$

Remark 5.3. The integral above can be written as $f \cdot W$ for short.

Lemma 5.4. (Itô Isometry) For all $f \in \mathcal{C}$,

$$\mathbb{E} \left[(f \cdot W)^2 \right] = \mathbb{E} \left[\int f^2 \right]. \quad (5.3)$$

Corollary 5.5. For all $f, g \in \mathcal{C}$,

$$\mathbb{E} [(f \cdot W) (g \cdot W)] = \mathbb{E} \left[\int fg \right]. \quad (5.4)$$

Theorem 5.6. Given any continuous semimartingale $X = M + A$ and any locally bounded and predictable process Y ,

$$Y \cdot X = Y \cdot M + Y \cdot A. \quad (5.5)$$

Theorem 5.7. (Stochastic integrals of semimartingales) Let $X, Y \in \mathcal{C}$ and M be a continuous semimartingale.

a. If $a, b \in \mathbb{R}$, then

$$(aX + bY) \cdot M = aX \cdot M + bY \cdot M. \quad (5.6)$$

b. If τ is a stopping time, then

$$(X 1_{[0, \tau]}) \cdot M_t = X \cdot M^\tau = (X \cdot M)^\tau. \quad (5.7)$$

c. If M is a bounded and continuous martingale, so as $X \cdot M$. If M is L^2 -bounded, then

$$\mathbb{E}[(X \cdot M)^2] = \mathbb{E}[X^2 \cdot \langle M, M \rangle]. \quad (5.8)$$

d. $X \cdot M$ is a continuous local martingale and

$$\langle X \cdot M, X \cdot M \rangle = X^2 \cdot \langle M, M \rangle. \quad (5.9)$$

If N is a continuous local martingale, then

$$\langle X \cdot M, N \rangle = X \cdot \langle M, N \rangle. \quad (5.10)$$

Theorem 5.8. (Integration by Parts) Let X and Y be any two continuous semimartingales. Then,

$$XY = X_0Y_0 + X \cdot Y + Y \cdot X + \langle X, Y \rangle. \quad (5.11)$$

Proof.

$$\begin{aligned} X_t Y_t - X_0 Y_0 &= \sum_{i \in \mathbb{N}} (X_{t_{i+1}} Y_{t_{i+1}} - X_{t_i} Y_{t_i}) = \sum_{i \in \mathbb{N}} (X_{t_{i+1}} Y_{t_{i+1}} - X_{t_{i+1}} Y_{t_i} + X_{t_{i+1}} Y_{t_i} - X_{t_i} Y_{t_i}) \\ &= \sum_{i \in \mathbb{N}} (X_{t_{i+1}} (Y_{t_{i+1}} - Y_{t_i}) + (X_{t_{i+1}} - X_{t_i}) Y_{t_i}) \\ &= \sum_{i \in \mathbb{N}} (X_{t_i} (Y_{t_{i+1}} - Y_{t_i}) + (X_{t_{i+1}} - X_{t_i}) Y_{t_i} + (X_{t_{i+1}} - X_{t_i}) (Y_{t_{i+1}} - Y_{t_i})) \\ &= X \cdot Y + Y \cdot X + \langle X, Y \rangle. \end{aligned} \quad (5.12)$$

□

Corollary 5.9. Let X , Y , and Z be any three continuous semimartingales. Then,

$$X_t Y_t Z_t = X_0 Y_0 Z_0 + (XY) \cdot Z + Y \cdot \langle X, Z \rangle + X \cdot \langle Y, Z \rangle + (YZ) \cdot X + (XZ) \cdot Y + Z \cdot \langle X, Y \rangle. \quad (5.13)$$

Lemma 5.10. If M is a continuous local martingale and

$$\int X^2 d\langle W, W \rangle < \infty, \quad (5.14)$$

then $X \cdot M$ is a continuous local martingale.

Theorem 5.11. Let M be any continuous local martingale and A be any adapted and cadlag process of finite variation that starts from 0. Then, the process $AM - M \cdot A$ is a local martingale.

Proof.

$$\begin{aligned} A_t M_t - A_0 M_0 - M \cdot A &= \sum_{i \in \mathbb{N}} (A_{t_{i+1}} M_{t_{i+1}} - A_{t_i} M_{t_i} - M_{t_i} (A_{t_{i+1}} - A_{t_i})) \\ &= \sum_{i \in \mathbb{N}} \liminf_{s_{i+1} \rightarrow t_{i+1}^-} A_{s_{i+1}} (M_{t_{i+1}} - M_{t_i}) = A_- \cdot M \end{aligned} \quad (5.15)$$

□

Definition 5.12. (Itô Process) Let $B(t, \omega) = (B_1(t, \omega), \dots, B_m(t, \omega))$ be an m -dimensional Brownian motion. For $i \in [1, n]$ and $j \in [1, m]$, the processes u_i and v_{ij} satisfy the conditions

$$P \left[\int_0^t |u(s, \omega)| ds < \infty \right] = P \left[\int_0^t v(s, \omega)^2 ds < \infty \right] = 1. \quad (5.16)$$

The process X_t is an Itô process if

$$dX(t) = u dt + v dB(t) \quad (5.17)$$

where

$$X(t) = \begin{pmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_n(t) \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \quad v = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1m} \\ v_{21} & v_{22} & \cdots & v_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nm} \end{pmatrix}, \quad dW(t) = \begin{pmatrix} dW_1(t) \\ dW_2(t) \\ \vdots \\ dW_m(t) \end{pmatrix}.$$

Theorem 5.13. (Itô Formula) Let $X(t) = (X_1(t), \dots, X_n(t))$ be n -dimensional continuous semimartingale and $f(t, x)$ be a C^2 map from $[0, \infty) \times \mathbb{R}^n$ into \mathbb{R}^m . Then, the process $Y(t) = f(t, X(t))$ is an Itô process whose component numbers k and Y_k are given by

$$dY_k = \frac{\partial f_k}{\partial t}(t, X) dt + \sum_i \frac{\partial f_k}{\partial X_i}(t, X) dX_i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f_k}{\partial X_i \partial X_j}(t, X) dX_i dX_j \quad (5.18)$$

where $dW_i dW_j = \delta_{ij} dt$ and $dW_i dt = dt dW_i = 0$.

Theorem 5.14. Every continuous and strictly positive local martingale M can be expressed in the form $M = e^{L - \frac{1}{2}\langle L, L \rangle}$ for some continuous local martingale L .

Proof. Let $L = \frac{1}{M} \cdot M$. Given the Itô's formula,

$$\log(M) = \log(M_0) + \frac{1}{M} \cdot M + \frac{1}{2M^2} \cdot \langle M, M \rangle \implies M = e^{\frac{1}{M} \cdot M + \frac{1}{2M^2} \cdot \langle M, M \rangle} = e^{L - \frac{1}{2}\langle L, L \rangle}. \quad (5.19)$$

□

Theorem 5.15. Let B and W be any two independent Brownian motions. Their joint quadratic variation is null.

Proof.

$$\begin{aligned} \mathbb{E} \left[(\mathcal{Q}^{\mathcal{P}_n}(W, B)_t)^2 \right] &= \mathbb{E} \left[\sum ((W_{t_{i+1}} - W_{t_i}) (B_{t_{i+1}} - B_{t_i}))^2 \right] \\ &= \mathbb{E} \left[\sum (W_{t_{i+1}} - W_{t_i}) (B_{t_{i+1}} - B_{t_i}) \right]^2 + \text{Var} \left(\sum (W_{t_{i+1}} - W_{t_i}) (B_{t_{i+1}} - B_{t_i}) \right) \\ &= \text{Var} \left(\sum (W_{t_{i+1}} - W_{t_i}) (B_{t_{i+1}} - B_{t_i}) \right) = \sum \text{Var} ((W_{t_{i+1}} - W_{t_i}) (B_{t_{i+1}} - B_{t_i})) \\ &= \sum \left(\mathbb{E} \left[((W_{t_{i+1}} - W_{t_i}) (B_{t_{i+1}} - B_{t_i}))^2 \right] - \mathbb{E} [(W_{t_{i+1}} - W_{t_i}) (B_{t_{i+1}} - B_{t_i})]^2 \right) \\ &= \sum \mathbb{E} [(W_{t_{i+1}} - W_{t_i})^2] \cdot \mathbb{E} [(B_{t_{i+1}} - B_{t_i})^2] = \sum (t_{i+1} - t_i)^2 \rightarrow 0. \end{aligned}$$

□

Theorem 5.16. Let W be a Brownian motion. Then,

$$W^2 = W_0 W_0 + 2W \cdot W + \langle W, W \rangle \implies W \cdot W = \frac{W^2 - t}{2}. \quad (5.20)$$

Theorem 5.17. Let $X = e^{\sigma W + (b - \sigma^2/2)t}$ for some fixed parameters σ and b . Then,

$$X = 1 + \sigma X \cdot W + bX \cdot t. \quad (5.21)$$

Theorem 5.18. (Girsanov's Theorem) Let W_t be a Brownian motion on the probability space (Ω, \mathcal{F}, P) and N be a continuous local martingale adapted to the filtration \mathcal{F} . Define the Doléans-Dade exponential as

$$\mathcal{E}(N) = e^{N - \frac{1}{2}\langle N, N \rangle}. \quad (5.22)$$

If $\mathcal{E}(N)$ is martingale, then another probability measure Q can be defined such that

$$\frac{dQ}{dP} = \mathcal{E}(N). \quad (5.23)$$

Then, $Q \approx P$. If M is a P -local martingale, then $M - \langle M, N \rangle$ is a Q -local martingale.

Proof.

$$(M - \langle M, N \rangle)\mathcal{E}(N) = M_0\mathcal{E}(N_0) + (M - \langle M, N \rangle) \cdot \mathcal{E}(N) + \mathcal{E}(N) \cdot M + \langle M, \mathcal{E}(N) \rangle. \quad (5.24)$$

Thus, $(M - \langle M, N \rangle)\mathcal{E}(N)$ is a local martingale since it can be written as the sum of the stochastic integrals of local martingales. $(M - \langle M, N \rangle)$ is a continuous local (\mathcal{F}, Q) -martingale since $(M - \langle M, N \rangle)\mathcal{E}(N)$ is a continuous local (\mathcal{F}, P) -martingale. \square

Remark 5.19. The condition $\mathbb{E}[\mathcal{E}(N)] = 1$ is equivalent to the condition that $\mathcal{E}(N)$ is martingale. Since N is a continuous local martingale, $\mathcal{E}(N)$ is a positive local martingale and thus a supermartingale. Since N is a supermartingale and $\mathbb{E}[\mathcal{E}(N)_T] = \mathbb{E}[\mathcal{E}(N)_0] = 1$, $\mathcal{E}(N)$ is a martingale on $[0, T]$.

Theorem 5.20. (Tanaka's Formula) For all $t \in \mathbb{R}_+$, let

$$L_t^a = \lim_{\epsilon \rightarrow 0^+} \int_0^t \frac{1}{2\epsilon} 1_{]a-\epsilon, a+\epsilon[}(W_s) ds. \quad (5.25)$$

Then,

$$|W_t - a| - |a| = \int_0^t \text{sgn}(W_s - a) dW_s + L_t^a. \quad (5.26)$$

Theorem 5.21. (Extension of Itô Formula) Let X be any continuous semimartingale and let $f : \mathbb{R} \rightarrow \mathbb{R}$ be any finite convex function. There is a continuous increasing and adapted process A^f on f and X such that

$$f(X) = f(X_0) + \partial_- f(X) \cdot X + \frac{1}{2} A^f. \quad (5.27)$$

Theorem 5.22. If W is a Brownian motion, then,

$$L^0(|W|) = 2L^0(W). \quad (5.28)$$

Proof.

$$L^0(|W|) = \lim_{\epsilon \rightarrow 0} \int \frac{1}{\epsilon} 1_{[0, \epsilon[}(|W|) d\langle |W|, |W| \rangle = 2 \lim_{\epsilon \rightarrow 0} \int \frac{1}{2\epsilon} 1_{]-\epsilon, \epsilon[}(W) d\langle W, W \rangle = 2L^0(W). \quad (5.29)$$

\square

Theorem 5.23. *Let A be a continuous process of finite variation and Z be a continuous semimartingale. The stochastic differential equation*

$$X = X_0 + X \cdot A + Z \quad (5.30)$$

has a unique solution

$$X = e^A(X_0 + e^{-A} \cdot Z). \quad (5.31)$$

Theorem 5.24. *Let Y and Z be two continuous semimartingales. Let $\mathcal{E}(Y) = e^{Y - \frac{1}{2}\langle Y, Y \rangle}$ be the Doléan-Dade exponent. The stochastic differential equation*

$$X = X_0 + X \cdot Y + Z \quad (5.32)$$

has a unique solution

$$X = \mathcal{E}(Y) \left(X_0 + \mathcal{E}(Y)^{-1} \cdot Z - \mathcal{E}(Y)^{-1} \cdot \langle Y, Z \rangle \right). \quad (5.33)$$

Theorem 5.25. *Given some fixed $T \in \mathbb{R}_{++}$ and any two deterministic functions $f, g : [0, T[\rightarrow \mathbb{R}$, the stochastic differential function has a unique strong solution on $[0, T[$ that*

$$X = e^{f \cdot t} \left(X_0 + e^{-f \cdot t} \cdot W + (ge^{-f \cdot t}) \cdot t \right). \quad (5.34)$$

Corollary 5.26. *Let $t \in [0, T]$. For some fixed scalars $x, \xi \in \mathbb{R}$, the equation*

$$X = x + W + \frac{\xi - X}{T - t} \cdot t \quad (5.35)$$

has a unique strong solution

$$X = (T - t) \left(\frac{1}{T - t} \cdot W \right) + \frac{T - t}{T} x + \frac{t}{T} \xi. \quad (5.36)$$

Proof. Let $X_0 = x$, $f(t) = -\frac{1}{T-t}$, and $g(t) = \frac{\xi}{T-t}$. Then,

$$f \cdot t = \int_0^t -\frac{1}{T-s} ds = \ln\left(\frac{T-t}{T}\right). \quad (5.37)$$

Thus,

$$\begin{aligned} X &= \frac{T-t}{T} \left(x + \frac{T}{T-t} \cdot W + \frac{\xi T}{(T-t)^2} \cdot t \right) \\ &= \frac{T-t}{T} x + (T-t) \left(\frac{1}{T-t} \cdot W \right) + \xi(T-t) \left(\frac{1}{(T-t)^2} \cdot t \right) \\ &= (T-t) \left(\frac{1}{T-t} \cdot W \right) + \frac{T-t}{T} x + \frac{t}{T} \xi. \end{aligned} \quad (5.38)$$

□

Definition 5.27. *Let $t \in [0, T]$. The Brownian bridge is the pinned Brownian motion with $X_0 = X_T = 0$ and can be expressed as*

$$X_t = (T-t) \int_0^t \frac{1}{T-s} dW_s. \quad (5.39)$$

Theorem 5.28. $\int_0^t \frac{1}{T-u} dW_u$ is a Brownian motion.

Proof. Let $Y_t = \int_0^t \frac{1}{T-u} dW_u$. Y_t is Gaussian.

$$\mathbb{E} \left[\int_0^t \frac{1}{T-u} dW_u \right] = 0. \quad (5.40)$$

$$\mathbb{E} \left[\int_0^s \frac{1}{T-u} dW_u \int_0^t \frac{1}{T-u} dW_u \right] = \int_0^{s \wedge t} \mathbb{E} \left[\frac{1}{(T-u)^2} \right] du = \frac{s}{T(T-s)}. \quad (5.41)$$

□

Theorem 5.29. Let $T \in \mathbb{R}_{++}$ and $t \in [0, T]$. Assume the Gaussian random variable $\xi \in \mathcal{N}(0, T)$ is independent of the Brownian motion. Then,

$$B_t = (T-t) \int_0^t \frac{1}{T-s} dW_s + \frac{t}{T} \xi \quad (5.42)$$

is Brownian motion on $[0, T]$.

Proof.

$$\mathbb{E}[B_t] = \mathbb{E} \left[(T-t) \int_0^t \frac{1}{T-u} dW_u + \frac{t}{T} \xi \right] = (T-t) \mathbb{E} \left[\int_0^t \frac{1}{T-u} dW_u \right] + \frac{t}{T} \mathbb{E}[\xi] = 0. \quad (5.43)$$

For $0 \leq s < t < \infty$,

$$\begin{aligned} \mathbb{E}[B_s B_t] &= \mathbb{E} \left[\left((T-s) \int_0^s \frac{1}{T-u} dW_u + \frac{s}{T} \xi \right) \left((T-t) \int_0^t \frac{1}{T-u} dW_u + \frac{t}{T} \xi \right) \right] \\ &= (T-s)(T-t) \mathbb{E} \left[\int_0^s \frac{1}{T-u} dW_u \int_0^t \frac{1}{T-u} dW_u \right] + \frac{st}{T^2} \mathbb{E}[\xi^2] \\ &= (T-s)(T-t) \int_0^{s \wedge t} \mathbb{E} \left[\frac{1}{(T-u)^2} \right] du + \frac{st}{T} = (T-s)(T-t) \cdot \frac{s}{(T-s)T} + \frac{st}{T} = s. \end{aligned} \quad (5.44)$$

□

Remark 5.30. The Brownian motion B_t can also be expressed as

$$B_t = W_t + \int_0^t \frac{B_T - B_s}{T-s} ds. \quad (5.45)$$

Proof. Apply Itô's formula,

$$\begin{aligned} d \left[e^{-\int_0^t r_s ds} p \right] &= e^{-\int_0^t r_s ds} \left[-r_t p dt + \frac{\partial p}{\partial t} dt + \frac{\partial p}{\partial r_t} dr_t + \frac{1}{2} \frac{\partial^2 p}{\partial r_t^2} d\langle r_t, r_t \rangle \right] \\ &= e^{-\int_0^t r_s ds} \left[\frac{\partial p}{\partial t} + \mu(t, r_t) \frac{\partial p}{\partial r} + \frac{1}{2} \sigma^2(t, r_t) \frac{\partial^2 p}{\partial r_t^2} - r_t p \right] dt + e^{-\int_0^t r_s ds} \frac{\partial p}{\partial t} \sigma(t, r_t) dW_t^*. \end{aligned} \quad (6.4)$$

Since the discounted price is a martingale, its drift should be zero. \square

Remark 6.3. For a zero-coupon bond, the boundary condition is

$$p(T, r) = B(T, T) = 1. \quad (6.5)$$

Definition 6.4. The Ornstein-Uhlenbeck (OU) process, also known as mean-reverting Brownian motion, is driven by a Brownian motion W with parameters $\mu \in \mathbb{R}$ and $\kappa, \sigma \in \mathbb{R}_{++}$.

$$r = r_0 + \kappa(\mu - r) \cdot t + \sigma W \iff dr_t = \kappa(\mu - r_t) dt + \sigma dW_t. \quad (6.6)$$

It has a solution

$$r_t = e^{-\kappa t} (r_0 + e^{\kappa t} \cdot (\sigma W + \kappa \mu t)). \quad (6.7)$$

Remark 6.5. It is also called Vasicek process.

Definition 6.6. The Hull-White (extended one-factor Vasicek) process is

$$dr_t = \kappa(\mu_t - r_t) dt + \sigma dW_t^*. \quad (6.8)$$

Theorem 6.7. Let $\tau = T - t$. Suppose that the Hull-White process admits an exponential affine solution for the price of a pure discount bond

$$p(\tau, r_t) = e^{A(\tau) + B(\tau) \cdot r_t}. \quad (6.9)$$

The PDE is

$$\left(-A'(\tau) + \frac{1}{2} \sigma^2 B^2(\tau) + \kappa \mu(\tau) B(\tau) \right) - r (\kappa B(\tau) + B'(\tau) + 1) = 0. \quad (6.10)$$

$$B'(\tau) = -\kappa B(\tau) - 1. \quad (6.11)$$

$$A'(\tau) = \frac{1}{2} \sigma^2 B^2(\tau) + \kappa \mu(\tau) B(\tau). \quad (6.12)$$

Given the boundary condition $p(T, r) = 1$, $A(0) = B(0) = 0$. Thus,

$$B(\tau) = \frac{1}{\kappa} (e^{-\kappa \tau} - 1). \quad (6.13)$$

$$A(\tau) = A(0) + \int_0^\tau A'(s) ds = \int_0^\tau \left(\kappa \mu(\tau - s) B(s) + \frac{1}{2} \sigma^2 B(s)^2 \right) ds. \quad (6.14)$$

$$\begin{aligned}
 r(t, T) &= -\frac{A(\tau)}{\tau} - \frac{B(\tau)}{\tau} r_t \\
 &= L_0 - S_0 \left(\frac{1 - e^{-at}}{at} \right) + K_0 t \left(\frac{1 - e^{-at}}{at} \right)^2 - R_0 \left(\frac{1 - e^{-bt}}{bt} \right) + M_0 t \left(\frac{1 - e^{-bt}}{bt} \right)^2. \quad (6.15)
 \end{aligned}$$

L_0 is the long-term rate. S_0 is the long-short spread. K_0 and M_0 are curvature parameters related to volatility. This family is consistent with the generalized model where both the short rate and its long term mean follow OU processes.

Remark 6.8. If $\mu(t) = \mu$,

$$A(\tau) = -\frac{\sigma^2}{4\kappa} B(\tau)^2 - r_\infty (\tau + B(\tau)) \quad (6.16)$$

where the long term rate

$$r_\infty = \mu + \frac{\sigma}{\kappa} \left(\lambda - \frac{\sigma}{2\kappa} \right). \quad (6.17)$$

Since the term structure of spot rates is $p(\tau, r_t) = e^{-\tau r(\tau, r_t)}$,

$$r(\tau, r_t) = -\frac{A(\tau)}{\tau} - \frac{B(\tau)}{\tau} r_t = r_\infty - (r_\infty - r_t) \left(\frac{1 - e^{-\kappa\tau}}{\tau} \right) + \frac{\sigma^2 \tau}{4\kappa} \left(\frac{1 - e^{-\kappa\tau}}{\kappa\tau} \right)^2. \quad (6.18)$$

Definition 6.9. The continuous compounded interest rate given the (extended) Nelson-Siegel model is

$$r_t = \beta_0 + \beta_1 \left(\frac{1 - e^{-\alpha t}}{\alpha t} \right) + \beta_2 \left(\frac{1 - (1 + \alpha t)e^{-\alpha t}}{\alpha t} \right) + \beta_3 \left(\frac{1 - (1 + bt)e^{-bt}}{bt} \right). \quad (6.19)$$

1. α is a scale. β_0 is a level or long-term rate. β_1 is a slope or long-short spread. β_2 is a curvature.
2. For the extended model, b and β_3 allow for more flexibility in fitting the short term end of the term structure. It is consistent with a model in which the short rate follows an OU process with deterministic long term mean.

Theorem 6.10. The relative bond dynamics for the Hull-White process is

$$\begin{aligned}
 \frac{dp(\tau, r_t)}{p(\tau, r_t)} &= e^{-A(\tau) - B(\tau)r_t} d \left(e^{A(\tau) + B(\tau)r_t} \right) = d(A(\tau) + B(\tau)r_t) \\
 &= -A'(\tau) dt - B'(\tau)r_t dt + B(\tau) dr_t + \frac{1}{2} \sigma^2 B(\tau)^2 d\langle r_t, r_t \rangle \\
 &= \left(-A'(\tau) - B'(\tau)r_t + \kappa(\mu - r_t)B(\tau) + \frac{1}{2} \sigma^2 B(\tau)^2 \right) dt + \sigma B(\tau) dW_t^* \\
 &= \left(-A'(\tau) + \frac{1}{2} \sigma^2 B(\tau)^2 + \kappa\mu B(\tau) - r_t (B'(\tau) + \kappa B(\tau)) \right) dt + \sigma B(\tau) dW_t^* \\
 &= r_t dt + \sigma B(\tau) dW_t^*. \quad (6.20)
 \end{aligned}$$

Remark 6.11. The drift r_t means that all traded assets must grow at the risk-free rate of return under the risk neutral probability. Since the exponential growth rate is normally distributed, the bond price is log-normally distributed.

Theorem 6.12. *The solution to the Hull-White process is*

$$r_t = e^{-\kappa t} r_0 + \int_0^t e^{-\kappa(t-s)} \mu(s) ds + \int_0^t \sigma e^{-\kappa(t-s)} dW_s. \quad (6.21)$$

Proof. By Itô's lemma,

$$\begin{aligned} d(e^{\kappa t} r_t) &= \kappa e^{\kappa t} r_t dt + e^{\kappa t} dr_t = \kappa e^{\kappa t} r_t dt + e^{\kappa t} (\kappa (\mu(t) - r_t) dt + \sigma dW_t) \\ &= \kappa e^{\kappa t} \mu(t) dt + e^{\kappa t} \sigma dW_t. \end{aligned} \quad (6.22)$$

□

Theorem 6.13. *The short rate is normally distributed with the conditional mean*

$$\mathbb{E}[r_t | r_0] = e^{-\kappa t} r_0 + \int_0^t e^{-\kappa(t-s)} \mu(s) ds \quad (6.23)$$

and the conditional variance

$$\begin{aligned} \text{Var}(r_t | r_0) &= \text{Var} \left(\int_0^t \sigma e^{\kappa(s-t)} dW_s \right) = \mathbb{E} \left[\left(\int_0^t \sigma e^{\kappa(s-t)} dW_s \right)^2 \right] \\ &= \int_0^t \sigma^2 e^{2\kappa(s-t)} ds = \frac{\sigma^2}{2\kappa} (1 - e^{-2\kappa t}). \end{aligned} \quad (6.24)$$

Remark 6.14.

$$\int_0^t e^{-\kappa(t-s)} \mu(s) ds = \alpha(t) - \alpha(0) e^{\kappa t} \quad \text{where} \quad \alpha(t) = f^M(0, t) + \frac{\sigma^2}{2\kappa^2} (1 - e^{-2\kappa t})^2. \quad (6.25)$$

Theorem 6.15. *Let $f(0, t)$ be the forward rate. The time-varying long-term mean*

$$\mu(t) = f(0, t) + \frac{1}{\kappa} \frac{\partial f(0, t)}{\partial t} + \frac{\sigma^2}{2\kappa^2} (1 - e^{-2\kappa t}). \quad (6.26)$$

Proof.

$$\begin{aligned} \frac{\partial A(t)}{\partial t} &= \frac{\partial}{\partial t} \int_0^t \left(\kappa \mu(t-s) B(s) + \frac{1}{2} \sigma^2 B^2(s) \right) ds \\ &= \frac{\partial}{\partial t} \int_0^t \kappa \mu(t-s) B(s) ds + \frac{\partial}{\partial t} \int_0^t \frac{1}{2} \sigma^2 B^2(s) ds \\ &= \kappa \mu(0) B(t) + \boxed{\int_0^t \kappa \mu'(t-s) B(s) ds} + \frac{\sigma^2}{2} B(t)^2 \\ &= \frac{\sigma^2}{2} B(t)^2 - \int_0^t \kappa e^{-\kappa s} \mu(t-s) ds. \end{aligned} \quad (6.27)$$

$$\frac{\partial B(t)}{\partial t} = -e^{-\kappa t}. \quad (6.28)$$

$$f(0, t) = -\frac{\partial \ln p(t, r_0)}{\partial t} = -\frac{\partial A(t)}{\partial t} - \frac{\partial B(t)}{\partial t} \cdot r_0 = e^{-\kappa t} \cdot r_0 - \frac{\sigma^2}{2} B(t)^2 + \int_0^t \kappa e^{-\kappa s} \mu(t-s) ds. \quad (6.29)$$

$$\begin{aligned}
 \frac{\partial f(0, t)}{\partial t} &= -\kappa e^{-\kappa t} \cdot r_0 - \sigma^2 B(t) B'(t) + \kappa e^{-\kappa t} \mu(0) + \boxed{\int_0^t \kappa e^{-\kappa s} \mu'(t-s) ds} \\
 &= -\kappa e^{-\kappa t} \cdot r_0 - \frac{\sigma^2}{\kappa} (e^{-\kappa t} - 1)(-e^{-\kappa t}) + \kappa \mu(t) - \kappa \int_0^t \kappa e^{-\kappa s} \mu(t-s) ds \\
 &= -\kappa e^{-\kappa t} \cdot r_0 - \frac{\sigma^2}{\kappa} (e^{-\kappa t} - 1)(-e^{-\kappa t}) + \kappa \mu(t) - \kappa \left(f(0, t) - e^{-\kappa t} \cdot r_0 + \frac{\sigma^2}{2} B(t)^2 \right) \\
 &= -\frac{\sigma^2}{\kappa} (e^{-\kappa t} - 1)(-e^{-\kappa t}) + \kappa \mu(t) - \kappa f(0, t) - \frac{\sigma^2}{2\kappa} (e^{-\kappa t} - 1)^2 \\
 &= \kappa \mu(t) - \kappa f(0, t) + \frac{\sigma^2}{2\kappa} (e^{-2\kappa t} - 1) \implies \mu(t) = f(0, t) + \frac{1}{\kappa} \frac{\partial f(0, t)}{\partial t} + \frac{\sigma^2}{2\kappa^2} (1 - e^{-2\kappa t}).
 \end{aligned} \tag{6.30}$$

□

Remark 6.16. Given integration by parts,

$$\begin{aligned}
 \int_0^t \kappa \mu'(t-s) B(s) ds &= -\kappa \mu(t-s) B(s) \Big|_0^t - \int_0^t -\kappa \mu(t-s) B'(s) ds \\
 &= -\kappa \mu(0) B(t) - \int_0^t \kappa e^{-\kappa s} \mu(t-s) ds.
 \end{aligned} \tag{6.31}$$

$$\begin{aligned}
 \int_0^t \kappa e^{-\kappa s} \mu'(t-s) ds &= -\kappa e^{-\kappa s} \mu(t-s) \Big|_0^t - \int_0^t \kappa^2 e^{-\kappa s} \mu(t-s) ds \\
 &= -\kappa e^{-\kappa t} \mu(0) + \kappa \mu(t) - \kappa \int_0^t \kappa e^{-\kappa s} \mu(t-s) ds.
 \end{aligned} \tag{6.32}$$

Definition 6.17. The Cox-Ingersoll-Ross (CIR) process is driven by a Brownian motion W with parameters $\mu \in \mathbb{R}$ and $\kappa, \sigma \in \mathbb{R}_{++}$.

$$r = r_0 + \kappa(\mu - r) \cdot t + \sigma \sqrt{|r|} \cdot W \iff dr_t = \mu(\theta - r_t) dt + \sigma \sqrt{r_t} dW_t. \tag{6.33}$$

Theorem 6.18. Let $\tau = T - t$. Suppose that the CIR process admits an exponential affine solution for the price of a pure discount bond.

$$p(\tau, r_t) = e^{A(\tau) + B(\tau)r_t}. \tag{6.34}$$

The PDE is

$$(-A'(\tau) + \kappa \mu B(\tau)) - r \left(\kappa B(\tau) + B'(\tau) - \frac{1}{2} \sigma^2 B(\tau)^2 + 1 \right) = 0. \tag{6.35}$$

$$B'(\tau) = \frac{1}{2} \sigma^2 B(\tau)^2 - \kappa B(\tau) - 1. \tag{6.36}$$

$$A'(\tau) = \kappa \mu B(\tau). \tag{6.37}$$

Given the boundary condition $p(T, r) = 1$, $A(0) = B(0) = 0$. Thus,

$$B(\tau) = \frac{-2(e^{\tau\gamma} - 1)}{(\kappa + \gamma)(e^{\tau\gamma} - 1) + 2\gamma}. \tag{6.38}$$

$$A(\tau) = A(0) + \int_0^\tau A'(s) ds = \int_0^\tau \kappa \mu B(s) ds = \frac{2\mu\kappa}{\sigma^2} \ln \left(\frac{2\gamma e^{\tau(\kappa+\gamma)/2}}{(\kappa + \sigma)(e^{\tau\gamma} - 1) + 2\gamma} \right). \quad (6.39)$$

$$\gamma = \sqrt{\kappa^2 + 2\sigma^2}. \quad (6.40)$$

Theorem 6.19. *The relative bond dynamics for the CIR process is*

$$\frac{dp(\tau, r_t)}{p(\tau, r_t)} = r_t dt + \sigma \sqrt{|r_t|} B(\tau) dW_t^*. \quad (6.41)$$

The drift r_t means that all traded assets must grow at the risk-free rate of return under the risk neutral probability. The discounted bond price is not log-normally distributed since it has a stochastic volatility.

Remark 6.20. *OU and CIR processes are most commonly used in interest rate and volatility modeling assuming asset prices are normally distributed. μ represents the speed of mean reversion and θ represents the level of mean reversion.*

Theorem 6.21. *Suppose the one-factor short rate model follows*

$$dr_t = (K_0(t) + K_1(t)r_t) dt + (H_0(t) + H_1(t)r_t)^\alpha dW_t^* \quad (6.42)$$

for some deterministic functions (K_0, K_1, H_0, H_1) and $\alpha > 0$. The model has an exponential affine structure if and only if $r_t = r$ and $\boxed{\alpha = 1/2 \text{ or } H_1(t) = 0}$.

Model	K_0	K_1	H_0	H_1	α	Affine
Cox, Ingersoll & Ross	•	•		•	0.5	Yes
Dothan				•	1	No
Merton (Ho & Lee)	•		•		1	Yes
Vasicek	•	•	•		1	Yes
Brennan & Schwartz	•	•		•	1	No
Constantinides & Ingersoll				•	1.5	No

Definition 6.22. *The exponential OU process is driven by a Brownian motion W with parameters $\theta \in \mathbb{R}$ and $\kappa, \sigma \in \mathbb{R}_{++}$.*

$$X = X_0 + \sigma X \cdot W + \kappa(\theta - \log(X))X \cdot t. \quad (6.43)$$

Corollary 6.23. *If $Y = \log(X)$ and X is an exponential OU process, then*

$$\log(X) = \log(X_0) + \sigma W + k \left(\alpha - \frac{1}{2k} \sigma^2 - Y \right) \cdot t. \quad (6.44)$$

Notice that Y is an OU process.

Proof. By Itô's Formula,

$$\begin{aligned} \log(X) &= \log(X_0) + \frac{d\log(X)}{dX} \cdot X + \frac{1}{2} \frac{d^2 \log(X)}{dX^2} \cdot \langle X, X \rangle \\ &= \log(X_0) + \frac{1}{X} k(\alpha - \log(X))X \cdot t + \frac{1}{X} \sigma X \cdot W + \frac{1}{2} \left(-\frac{1}{X^2} \right) \sigma^2 X^2 \cdot t \\ &= \log(X_0) + \sigma W + k \left(\alpha - \frac{1}{2k} \sigma^2 - Y \right) \cdot t. \end{aligned} \quad (6.45)$$

□

Definition 6.24. The Dixit and Pindyck process driven by a Brownian motion W with parameters $\kappa, \sigma, \theta \in \mathbb{R}_{++}$.

$$X = X_0 + \sigma X \cdot W + \kappa(\theta - X)X \cdot t. \quad (6.46)$$

Corollary 6.25. If $Y = \log(X)$ and X is a Dixit and Pindyck process, then

$$Y = \log(X_0) + \sigma W - \kappa X \cdot t + \left(\kappa\theta - \frac{\sigma^2}{2}\right) \cdot t. \quad (6.47)$$

Remark 6.26. The exponential OU process and the Dixit and Pindyck process can be used to price commodities.

Definition 6.27. The constant elasticity of variance (CEV) process is driven by a Brownian motion W with parameters $r, \sigma, \beta \in \mathbb{R}$.

$$S = S_0 + rS \cdot t + \sigma S^\beta \cdot W \iff dS_t = rS_t dt + \sigma S_t^\beta dW \quad (6.48)$$

where β can be used to account for some degree of skew in the volatility surface. The following PDE can be deduced from the SDE.

$$\frac{\partial c}{\partial t} + \frac{1}{2}\sigma^2 s^{2\beta} \frac{\partial^2 c}{\partial s^2} + rs \frac{\partial c}{\partial s} - rc = 0. \quad (6.49)$$

Proof. Let $c(s, t)$ be the payoff of a derivative instrument at time t . By Itô's lemma,

$$\begin{aligned} dc(s, t) &= \frac{\partial c}{\partial s} ds + \frac{1}{2} \frac{\partial^2 c}{\partial s^2} (ds)^2 + \frac{\partial c}{\partial t} dt \\ &= \frac{\partial c}{\partial s} rs dt + \frac{\partial c}{\partial s} \sigma s^\beta dW + \frac{1}{2} \frac{\partial^2 c}{\partial s^2} \sigma^2 s^{2\beta} dt + \frac{\partial c}{\partial t} dt \\ &= \left(\frac{\partial c}{\partial t} + rs \frac{\partial c}{\partial s} + \frac{1}{2} \sigma^2 s^{2\beta} \frac{\partial^2 c}{\partial s^2} \right) dt + \frac{\partial c}{\partial s} \sigma s^\beta dW. \end{aligned} \quad (6.50)$$

We construct a delta-hedged portfolio to eliminate the stochastic term.

$$\Pi(s, t) = c(s, t) - \frac{\partial c}{\partial s} s. \quad (6.51)$$

This is a risk-free instrument at time t . By the standard arbitrage argument, its evolution equation must be

$$d\Pi = r\Pi dt \iff \frac{\partial \Pi}{\partial t} = r\Pi. \quad (6.52)$$

Thus,

$$\left(\frac{\partial c}{\partial t} + rs \frac{\partial c}{\partial s} + \frac{1}{2} \sigma^2 s^{2\beta} \frac{\partial^2 c}{\partial s^2} \right) - \frac{\partial c}{\partial s} rs = r \left(c - \frac{\partial c}{\partial s} s \right) \quad (6.53)$$

and the CEV PDE is acquired after the rearrangement of terms. \square

Theorem 6.28. (*Explicit Euler Scheme*) We discretize partial derivatives in the CEV PDE by backward difference and Talyor expansion.

$$\begin{aligned} \frac{c(s_i, t_j) - c(s_i, t_{j-1})}{h_t} + \frac{1}{2} \sigma^2 s_i^{2\beta} \frac{c(s_{i+1}, t_j) - 2c(s_i, t_j) + c(s_{i-1}, t_j)}{h_s^2} \\ + rs_i \frac{c(s_{i+1}, t_j) - c(s_{i-1}, t_j)}{2h_s} - rc(s_i, t_j) = 0 \end{aligned} \quad (6.54)$$

with the time mesh $h_t = \frac{t}{n}$ and the price mesh $h_s = \frac{s_m}{m}$. We can rearrange to isolate $c(s_i, t_{j-1})$.

$$c(s_i, t_{j-1}) = l_i c(s_{i-1}, t_j) + a_i c(s_i, t_j) + u_i c(s_{i+1}, t_j) \quad (6.55)$$

where

$$l_i = \frac{\sigma^2 s_i^2 h_t}{2h_s^2} - \frac{rs_i h_t}{2h_s} = \frac{1}{2}(\sigma^2 i^{2\beta} h_s^{2\beta-2} - ri)h_t, \quad (6.56)$$

$$a_i = 1 - \sigma^2 s_i^2 \frac{h_t}{h_s^2} - rh_t = 1 - \left(\sigma^2 i^{2\beta} h_s^{2\beta-2} + r \right) h_t, \quad (6.57)$$

$$u_i = \frac{\sigma^2 s_i^2 h_t}{2h_s^2} + \frac{rs_i h_t}{2h_s} = \frac{1}{2}(\sigma^2 i^{2\beta} h_s^{2\beta-2} + ri)h_t. \quad (6.58)$$

The boundary conditions for European options are

Bound	Call	Put	Call-Spread	Put-Spread
Right	$(S_i - K)_+$	$(K - S_i)_+$	$(S_i - K_2)_+ - (S_i - K_1)_+$	$(K_1 - S_i)_+ - (K_2 - S_i)_+$
Bottom	0	$Ke^{-r\Delta t} - S_m$	0	$(K_1 - K_2)e^{-r\Delta t}$
Top	$S_m - Ke^{-r\Delta t}$	0	$(K_1 - K_2)e^{-r\Delta t}$	0

where $\Delta t = t_n - t_j$. For American options, during each iteration,

$$c_j = \max \left\{ f(t_j), c_{j+1}e^{-rh_t} \right\}. \quad (6.59)$$

Therefore,

$$c(s_1, t_{j-1}) = l_1 c(s_0, t_j) + a_1 c(s_1, t_j) + u_1 c(s_2, t_j). \quad (6.60)$$

$$c(s_{m-1}, t_{j-1}) = l_{m-1} c(s_{m-2}, t_j) + a_{m-1} c(s_{m-1}, t_j) + u_{m-1} c(s_m, t_j). \quad (6.61)$$

In matrix form,

$$\begin{pmatrix} c(s_1, t_{j-1}) \\ c(s_2, t_{j-1}) \\ \vdots \\ c(s_{m-2}, t_{j-1}) \\ c(s_{m-1}, t_{j-1}) \end{pmatrix} = \begin{pmatrix} a_1 & u_1 & & & \\ l_2 & a_2 & u_2 & & \\ & \ddots & \ddots & \ddots & \\ & & l_{m-2} & a_{m-2} & u_{m-2} \\ & & & l_{m-1} & a_{m-1} \end{pmatrix} \begin{pmatrix} c(s_1, t_j) \\ c(s_2, t_j) \\ \vdots \\ c(s_{m-2}, t_j) \\ c(s_{m-1}, t_j) \end{pmatrix} + \begin{pmatrix} b_j \\ 0 \\ \vdots \\ 0 \\ b'_j \end{pmatrix} \quad (6.62)$$

where $b_j = l_1 c(s_0, t_j)$ and $b'_j = u_{m-1} c(s_m, t_j)$ or $\vec{C}_{j-1} = A\vec{C}_j + \vec{b}_j$ in short. By induction¹,

$$\vec{C}_0 = A^n \vec{C}_n + \sum_{k=0}^{n-1} A^k \vec{b}_{k+1}. \quad (6.63)$$

Since they are all known a priori, this should provide us the solution as long as there are no surprises from A .

¹Do NOT use the induction form for the actual programming since the matrix might be close to singular or badly scaled. Beside, the optimal time complexity for solving tridiagonal systems of equations is $O(n)$.

Remark 6.29. (Courant–Friedrichs–Lewy Condition) For the explicit Euler scheme to be stable and numerically solvable,

$$h_t < \frac{h_s^2}{\sigma^2 s_M^2} = O(h_s^2). \quad (6.64)$$

Theorem 6.30. (Implicit Euler Scheme) We discretize partial derivatives in the CEV PDE by forward difference and Talyor expansion.

$$\begin{aligned} \frac{c(s_i, t_{j+1}) - c(s_i, t_j)}{h_t} + \frac{1}{2} \sigma^2 s_i^2 \frac{c(s_{i+1}, t_j) - 2c(s_i, t_j) + c(s_{i-1}, t_j)}{h_s^2} \\ + r s_i \frac{c(s_{i+1}, t_j) - c(s_{i-1}, t_j)}{2h_s} - r c(s_i, t_j) = 0. \end{aligned} \quad (6.65)$$

We can rearrange to isolate $c(s_i, t_{j+1})$.

$$c(s_i, t_{j+1}) = -l_i c(s_{i-1}, t_j) + (2 - a_i) c(s_i, t_j) - u_i c(s_{i+1}, t_j) \quad (6.66)$$

with a_i , l_i , and u_i denoted previously. With the boundary conditions,

$$c(s_1, t_{j+1}) = -l_1 c(s_0, t_j) + (2 - a_1) c(s_1, t_j) - u_1 c(s_2, t_j). \quad (6.67)$$

$$c(s_{m-1}, t_{j+1}) = -l_{m-1} c(s_{m-2}, t_j) + (2 - a_{m-1}) c(s_{m-1}, t_j) - u_{m-1} c(s_m, t_j). \quad (6.68)$$

In matrix form,

$$\begin{pmatrix} c(s_1, t_{j+1}) \\ c(s_2, t_{j+1}) \\ \vdots \\ c(s_{m-2}, t_{j+1}) \\ c(s_{m-1}, t_{j+1}) \end{pmatrix} = \begin{pmatrix} 2 - a_1 & -u_1 & & & \\ -l_2 & 2 - a_2 & -u_2 & & \\ & \ddots & \ddots & \ddots & \\ & & -l_{m-2} & 2 - a_{m-2} & -u_{m-2} \\ & & & -l_{m-1} & 2 - a_{m-1} \end{pmatrix} \begin{pmatrix} c(s_1, t_j) \\ c(s_2, t_j) \\ \vdots \\ c(s_{m-2}, t_j) \\ c(s_{m-1}, t_j) \end{pmatrix} - \begin{pmatrix} b_j \\ 0 \\ \vdots \\ 0 \\ b'_j \end{pmatrix} \quad (6.69)$$

with b_j and b'_j denoted previously or $\vec{C}_{j+1} = B\vec{C}_j - \vec{b}_j$ in short. If B is invertible,

$$\vec{C}_j = B^{-1} (\vec{C}_{j+1} + \vec{b}_j). \quad (6.70)$$

By induction¹,

$$\vec{C}_0 = B^{-n} \vec{C}_n + \sum_{k=0}^{n-1} B^{-(k+1)} \vec{b}_k. \quad (6.71)$$

Since they are all known a priori, this should provide us the solution as long as there are no surprises from B .

Theorem 6.31. (Crank–Nicolson Scheme) We shift back the Euler implicit scheme by one unit and combine it with the explicit one.

$$\begin{aligned} c(s_i, t_j) - c(s_i, t_{j-1}) = -l_i c(s_{i-1}, t_{j-1}) + (2 - a_i) c(s_i, t_{j-1}) - u_i c(s_{i+1}, t_{j-1}) \\ - l_i c(s_{i-1}, t_j) - a_i c(s_i, t_j) - u_i c(s_{i+1}, t_j) \end{aligned} \quad (6.72)$$

which can be rearranged as

$$\begin{aligned} -l_i c(s_{i-1}, t_{j-1}) + (3 - a_i) c(s_i, t_{j-1}) - u_i c(s_{i+1}, t_{j-1}) \\ = l_i c(s_{i-1}, t_j) + (a_i + 1) c(s_i, t_j) + u_i c(s_{i+1}, t_j). \end{aligned} \quad (6.73)$$

In matrix form,

$$\begin{aligned} & \begin{pmatrix} 3 - a_1 & -u_1 & & & \\ -l_2 & 3 - a_2 & -u_2 & & \\ & \ddots & \ddots & \ddots & \\ & & -l_{m-2} & 3 - a_{m-2} & -u_{m-2} \\ & & & -l_{m-1} & 3 - a_{m-1} \end{pmatrix} \begin{pmatrix} c(s_1, t_{j-1}) \\ c(s_2, t_{j-1}) \\ \vdots \\ c(s_{m-2}, t_{j-1}) \\ c(s_{m-1}, t_{j-1}) \end{pmatrix} - \begin{pmatrix} b_{j-1} \\ 0 \\ \vdots \\ 0 \\ b'_{j-1} \end{pmatrix} \\ &= \begin{pmatrix} a_1 + 1 & u_1 & & & \\ l_2 & a_2 + 1 & u_2 & & \\ & \ddots & \ddots & \ddots & \\ & & l_{m-2} & a_{m-2} + 1 & u_{m-2} \\ & & & l_{m-1} & a_{m-1} + 1 \end{pmatrix} \begin{pmatrix} c(s_1, t_j) \\ c(s_2, t_j) \\ \vdots \\ c(s_{m-2}, t_j) \\ c(s_{m-1}, t_j) \end{pmatrix} + \begin{pmatrix} b_j \\ 0 \\ \vdots \\ 0 \\ b'_j \end{pmatrix} \end{aligned} \quad (6.74)$$

or $A_1 \vec{C}_{j-1} - \vec{b}_{j-1} = A_2 \vec{C}_j + \vec{b}_j$ in short. If A_1 is invertible,

$$\vec{C}_{j-1} = A_1^{-1} (A_2 \vec{C}_j + \vec{b}_j + \vec{b}_{j-1}). \quad (6.75)$$

Since they are all known a priori, this should provide us the solution as long as there are no surprises from A_1 or A_2 .

Remark 6.32. The implicit Euler and Crank-Nicolson schemes are not subject to CFL conditions. The Crank-Nicolson scheme converges faster than the implicit Euler one.

Theorem 6.33. Let $\beta = 0$ in the CEV process. The process is reduced to the Bachelier process. Asset prices are normally distributed.

$$dS_t = rS_t dt + \sigma dW. \quad (6.76)$$

The solution is

$$S_t = S_0 e^{rt} + \sigma \int_0^t e^{r(t-s)} dW_s \sim \mathcal{N} \left(S_0 e^{rt}, \frac{\sigma^2}{2r} (e^{2rt} - 1) \right). \quad (6.77)$$

Theorem 6.34. Given the Bachelier process, the European call option price is

$$c = \mathcal{N} \left(\frac{S_0 e^{rt} - K}{\sqrt{\nu}} \right) (S_0 - K e^{-rt}) + \mathcal{N}' \left(\frac{S_0 e^{rt} - K}{\sqrt{\nu}} \right) e^{-rt} \sqrt{\nu}, \quad (6.78)$$

$$\nu = \frac{\sigma^2}{2r} (1 - e^{-2rt}). \quad (6.79)$$

where $\mathcal{N}(\cdot)$ and $\mathcal{N}'(\cdot)$ are the cumulative distribution function and the probability density function of standard normal distribution.

Theorem 6.35. *Let $\beta = 1$ in the CEV process and further denote δ as the dividend rate. The process is reduced to the Black-Scholes process. Asset prices are log-normally distributed.*

$$dS_t = (r - \delta)S_t dt + \sigma S_t dW_t. \quad (6.80)$$

Theorem 6.36. *(European Option Pricing under BS Model) The European call option price is*

$$c_t = \mathcal{N}(d_1)S_0e^{-\delta t} - \mathcal{N}(d_2)Ke^{-rt} \quad (6.81)$$

and the European put option price is

$$p_t = \mathcal{N}(-d_2)Ke^{-rt} - \mathcal{N}(-d_1)S_0e^{-\delta t} \quad (6.82)$$

where

$$d_1 = \frac{\ln\left(\frac{S}{K}\right) + \left(r - \delta + \frac{\sigma^2}{2}\right)t}{\sigma\sqrt{t}}, \quad d_2 = d_1 - \sigma\sqrt{t}. \quad (6.83)$$

Proof. For the call option,

$$\begin{aligned} c_t &= \mathbb{E}_t [b_{t,T}(S_T - K)_+] = S_t b_{t,T} e^{(r-\delta)\tau} \mathbb{E}_t [\eta_{t,T} 1_{S_t N_t \geq K}] - K b_{t,T} \mathbb{E}_t [1_{S_t N_t \geq K}] \\ &= S_t e^{-\delta\tau} \mathbb{E}_t^\sigma [1_{S_t N_t \geq K}] - K b_{t,T} \mathbb{E}_t [1_{S_t N_t \geq K}] = S_t e^{-\delta\tau} \mathbb{Q}^\sigma(\epsilon) - K e^{-r\tau} \mathbb{Q}(\epsilon). \end{aligned} \quad (6.84)$$

Since $dW^\sigma = dW - \sigma dt$ is a \mathbb{Q}^σ -martingale,

$$\begin{aligned} \epsilon : N_{t,T} \geq \frac{K}{S_t} &\iff W \geq \frac{1}{\sigma\sqrt{\tau}} \left[\ln\left(\frac{K}{S_t}\right) - \left(r - \delta - \frac{1}{2}\sigma^2\right)\tau \right] \\ &= -\frac{1}{\sigma\sqrt{\tau}} \left[\ln\left(\frac{S_t}{K}\right) + \left(r - \delta + \frac{1}{2}\sigma^2\right)\tau \right] + \sigma\sqrt{\tau} \\ &= -d + \sigma\sqrt{\tau} \iff W^\sigma \geq -d. \end{aligned} \quad (6.85)$$

Thus,

$$\mathbb{Q}^\sigma(\epsilon) = 1 - N(-d) = N(d), \quad \mathbb{Q}(\epsilon) = N(d - \sigma\sqrt{\tau}). \quad (6.86)$$

Similarly, for the put option,

$$\begin{aligned} p_t &= \mathbb{E}_t [b_{t,T}(K - S_T)_+] = K b_{t,T} \mathbb{E}_t [1_{S_t N_t \leq K}] - S_t e^{-\delta\tau} \mathbb{E}_t [\eta_{t,T} 1_{S_t N_t \leq K}] \\ &= K b_{t,T} \mathbb{E}_t [1_{S_t N_t \leq K}] - S_t e^{-\delta\tau} \mathbb{E}_t^\sigma [1_{S_t N_t \leq K}] = K b_{t,T} \mathbb{Q}(\bar{\epsilon}) - S_t e^{-\delta\tau} \mathbb{Q}^\sigma(\bar{\epsilon}) \end{aligned} \quad (6.87)$$

where

$$\mathbb{Q}^\sigma(\bar{\epsilon}) = N(-d), \quad \mathbb{Q}(\bar{\epsilon}) = N(-d + \sigma\sqrt{\tau}). \quad (6.88)$$

□

Remark 6.37. *We denote $\mathcal{A}(\tau)$ as the price of call or put at τ and*

$$\alpha^\pm = r - \delta \pm \frac{\sigma^2}{2}. \quad (6.89)$$

Theorem 6.38. (Exotic Option Pricing under BS Model) The reset strike call option price is

$$RSC(S_t, t; T_1, T_2) = \mathbb{E}_t \left[b_{t,T_2} (S_{T_2} - K)_+ 1_{S_{T_1} \geq K} + b_{t,T_2} (S_{T_2} - S_{T_1})_+ 1_{S_{T_1} < K} \right]. \quad (6.90)$$

The first part is evaluated as

$$\begin{aligned} RSC_1 &= \mathbb{E}_t \left[b_{t,T_2} (S_{T_2} - K) 1_{S_{T_1} \geq K, S_{T_2} \geq K} \right] \\ &= S_t \mathbb{E}_t \left[b_{t,T_2} e^{(r-\delta)\tau_2} \eta_{t_1, T_2} 1_{\epsilon_1 \wedge \epsilon_2} \right] - K e^{-r\tau_2} \mathbb{E}_t [1_{\epsilon_1 \wedge \epsilon_2}] \\ &= S_t e^{-\delta\tau_2} \mathbb{E}_t^\sigma [1_{\epsilon_1 \wedge \epsilon_2}] - K e^{-r\tau_2} \mathbb{E}_t [1_{\epsilon_1 \wedge \epsilon_2}] \\ &= S_t e^{-\delta\tau_2} \mathbb{Q}^\sigma [\epsilon_1 \wedge \epsilon_2] - K e^{-r\tau_2} \mathbb{Q} [\epsilon_1 \wedge \epsilon_2] \\ &= S_t e^{-\delta\tau_2} N_2(d_1, d_2, \rho) - K e^{-r\tau_2} N_2(d_1 - \sigma\sqrt{\tau_1}, d_2 - \sigma\sqrt{\tau_2}, \rho) \end{aligned} \quad (6.91)$$

where

$$\rho = \text{Corr}(\omega_1 \sqrt{\tau_1}, \omega_2 \sqrt{\tau_2}) = \frac{\text{Cov}(\omega_1 \sqrt{\tau_1}, \omega_2 \sqrt{\tau_2})}{\sqrt{\tau_1 \tau_2}} = \sqrt{\frac{\tau_1}{\tau_2}}. \quad (6.92)$$

The second part is evaluated as

$$\begin{aligned} RSC_2 &= \mathbb{E}_t \left[b_{t,T_2} S_{T_2} \mathcal{A}(\tau_{12}) 1_{S_{T_1} < K} \right] \\ &= S_t e^{-\delta\tau_2} \mathbb{E}_t^\sigma [1_{S_{T_1} < K}] \mathcal{A}(\tau_{12}) = S_t e^{-\delta\tau_2} \mathbb{Q}^\sigma (S_{T_1} < K) \mathcal{A}(\tau_{12}) \\ &= S_t e^{-\delta\tau_2} N(-d_1) \mathcal{A}(\tau_{12}). \end{aligned} \quad (6.93)$$

The reset strike put option price is

$$RSP(S_t, t; T_1, T_2) = \mathbb{E}[b_{t,T_2} (K - S_{T_2})^+ 1_{S_{T_1} \geq K} + b_{t,T_2} (S_{T_1} - S_{T_2})^+ 1_{S_{T_1} < K}]. \quad (6.94)$$

The first part is evaluated as

$$\begin{aligned} RSP_1 &= \mathbb{E}_t [b_{t,T_2} (K - S_{T_2}) 1_{S_{T_1} \geq K, K \geq S_{T_2}}] \\ &= K e^{-r\tau_2} \mathbb{E}_t [1_{\epsilon_1 \wedge \bar{\epsilon}_2}] - S_t \mathbb{E}_t [b_{t,T_2} e^{(r-\delta)\tau_2} \eta_{t_1, T_2} 1_{\epsilon_1 \wedge \bar{\epsilon}_2}] \\ &= K e^{-r\tau_2} \mathbb{E}_t [1_{\epsilon_1 \wedge \bar{\epsilon}_2}] - S_t e^{-\delta\tau_2} \mathbb{E}_t^\sigma [1_{\epsilon_1 \wedge \bar{\epsilon}_2}] \\ &= K e^{-r\tau_2} \mathbb{Q} [\epsilon_1 \wedge \bar{\epsilon}_2] - S_t e^{-\delta\tau_2} \mathbb{Q}^\sigma [\epsilon_1 \wedge \bar{\epsilon}_2] \\ &= K e^{-r\tau_2} N_2(d_1 - \sigma\sqrt{\tau_1}, -d_2 + \sigma\sqrt{\tau_2}, \rho) - S_t E_t [b_{t,T_2} e^{(r-\delta)\tau_2} \eta_{t_1, T_2} N_2(d_1, -d_2, \rho)]. \end{aligned} \quad (6.95)$$

The second part is evaluated as

$$RSP_2 = E_t [b_{t,T_2} S_{T_2} \mathcal{A}(\tau_{12}) 1_{S_{T_1} < K}] = S_t e^{-\delta\tau_2} \mathbb{E}_t [1_{S_{T_1} < K}] \mathcal{A}(\tau_{12}) = S_t e^{-\delta\tau_2} N(-d_1) \mathcal{A}(\tau_{12}) \quad (6.96)$$

The call on call option price is

$$\begin{aligned} CC(S_t, t; T_1, T_2) &= \mathbb{E}_t \left[b_{t,T_2} (S_{T_2} - K_2) 1_{S_{T_1} \geq S^*, S_{T_2} \geq K_2} - b_{t,T_1} K_1 1_{S_{T_1} \geq S^*} \right] \\ &= S_t e^{-\delta\tau_2} N_2(d_1^*, d_2, \rho) - K_2 e^{-r\tau_2} N_2(d_1^* - \sigma\sqrt{\tau_1}, d_2 - \sigma\sqrt{\tau_2}, \rho) \\ &\quad - K_1 e^{-\delta\tau_1} N(d_1^* - \sigma\sqrt{\tau_1}). \end{aligned} \quad (6.97)$$

The call on put option price is

$$\begin{aligned} CP(S_t, t; T_1, T_2) &= \mathbb{E} \left[b_{t, T_2}(K_2 - S_{T_2}) 1_{S_{T_1} \geq S^*, K_2 \geq S_{T_2}} - b_{t, T} K_1 1_{S_{T_1} \geq S^*} \right] \\ &= K_2 e^{-r\tau_2} N_2(d_1^* - \sigma\sqrt{\tau_1}, -d_2 + \sigma\sqrt{\tau_2}, \rho) - S_t e^{-\delta\tau_2} N_2(d_1^*, -d_2, \rho) \\ &\quad - K_1 e^{-\delta\tau_1} N(d_1^* - \sigma\sqrt{\tau_1}). \end{aligned} \quad (6.98)$$

The put on call option price is

$$\begin{aligned} PC(S_t, t; T_1, T_2) &= \mathbb{E} \left[b_{t, T} K_1 1_{S^* \geq S_{T_1}} - b_{t, T_2}(S_{T_2} - K_2) 1_{S^* \geq S_{T_1}, S_{T_2} \geq K_2} \right] \\ &= K_1 e^{-\delta\tau_1} N(-d_1^* + \sigma\sqrt{\tau_1}) \\ &\quad - \left(S_t e^{-\delta\tau_2} N_2(-d_1^*, d_2, \rho) - K_2 e^{-r\tau_2} N_2(-d_1^* + \sigma\sqrt{\tau_1}, d_2 - \sigma\sqrt{\tau_2}, \rho) \right). \end{aligned} \quad (6.99)$$

The put on put option price is

$$\begin{aligned} PP(S_t, t; T_1, T_2) &= \mathbb{E} \left[b_{t, T} K_1 1_{S^* \geq S_{T_1}} - b_{t, T_2}(K_2 - S_{T_2}) 1_{S^* \geq S_{T_1}, K_2 \geq S_{T_2}} \right] \\ &= K_1 e^{-\delta\tau_1} N(\sigma\sqrt{\tau_1} - d_1^*) \\ &\quad - \left(K_2 e^{-r\tau_2} N_2(-d_1^* + \sigma\sqrt{\tau_1}, -d_2 + \sigma\sqrt{\tau_2}, \rho) - S_t e^{-\delta\tau_2} N_2(-d_1^*, -d_2, \rho) \right). \end{aligned} \quad (6.100)$$

The chooser option price is

$$\begin{aligned} V_t &= \mathbb{E}_t \left[b_{t, T_2}(S_{T_2} - K)_+ 1_{S_{T_1} \geq S^*, S_{T_2} \geq K} + b_{t, T_2}(K - S_{T_2})_+ 1_{S_{T_1} < S^*, S_{T_2} < K} \right] \\ &= S_t e^{-\delta\tau_2} N_2(d_1^*, d_2, \rho) - K e^{-r\tau_2} N_2(d_1^* - \sigma\sqrt{\tau_1}, d_2 - \sigma\sqrt{\tau_2}, \rho) \\ &\quad + K e^{-r\tau_2} N_2(-d_1^* + \sigma\sqrt{\tau_1}, -d_2 + \sigma\sqrt{\tau_2}, \rho) - S_t e^{-\delta\tau_2} N_2(-d_1^*, -d_2, \rho). \end{aligned} \quad (6.101)$$

(Barrier Option) Define R as the rebate rate. We set

$$A = \phi S_t e^{-\delta\tau} N(\phi x_1) - \phi K e^{-r\tau} N(\phi x_1 - \phi\sigma\sqrt{\tau}), \quad (6.102)$$

$$B = \phi S_t e^{-\delta\tau} N(\phi x_2) - \phi K e^{-r\tau} N(\phi x_2 - \phi\sigma\sqrt{\tau}), \quad (6.103)$$

$$C = \phi S_t \left(\frac{H}{S_t} \right)^{2(\mu+1)} e^{-\delta\tau} N(\eta y_1) - \phi K e^{-r\tau} \left(\frac{H}{S_t} \right)^{2\mu} N(\eta y_1 - \eta\sigma\sqrt{\tau}), \quad (6.104)$$

$$D = \phi S_t \left(\frac{H}{S_t} \right)^{2(\mu+1)} e^{-\delta\tau} N(\eta y_2) - \phi K e^{-r\tau} \left(\frac{H}{S_t} \right)^{2\mu} N(\eta y_2 - \eta\sigma\sqrt{\tau}), \quad (6.105)$$

$$E = R e^{-r\tau} \left[N(\eta x_2 - \eta\sigma\sqrt{\tau}) - \left(\frac{H}{S_t} \right)^{2\mu} N(\eta y_2 - \eta\sigma\sqrt{\tau}) \right], \quad (6.106)$$

$$F = R \left(\frac{H}{S_t} \right)^{\mu+\lambda} N(\eta z) + \left(\frac{H}{S_t} \right)^{\mu-\lambda} N(\eta z - 2\eta\lambda\sigma\sqrt{\tau}) \quad (6.107)$$

where

$$x_1 = \frac{\ln\left(\frac{S_t}{K}\right)}{\sigma\sqrt{\tau}} + (1 + \mu)\sigma\sqrt{\tau}, \quad x_2 = \frac{\ln\left(\frac{S_t}{H}\right)}{\sigma\sqrt{\tau}} + (1 + \mu)\sigma\sqrt{\tau}, \quad (6.108)$$

$$y_1 = \frac{\ln\left(\frac{H^2}{S_t K}\right)}{\sigma\sqrt{\tau}} + (1 + \mu)\sigma\sqrt{\tau}, \quad y_2 = \frac{\ln\left(\frac{H}{S_t}\right)}{\sigma\sqrt{\tau}} + (1 + \mu)\sigma\sqrt{\tau}, \quad (6.109)$$

$$z = \frac{\ln\left(\frac{H}{S_t}\right)}{\sigma\sqrt{\tau}} + \lambda\sigma\sqrt{\tau}, \quad \mu = \alpha^-, \quad \lambda = \sqrt{\mu^2 + \frac{2r}{\sigma^2}}, \quad \eta, \phi \in \{-1, 1\}. \quad (6.110)$$

(Down-Out Call) Let $\eta = 1$ and $\phi = 1$.

$$c^{do} = \begin{cases} A - C + F, & K > L \\ B - D + F, & K \leq L \end{cases} \quad (6.111)$$

(Down-Out Put) Let $\eta = 1$ and $\phi = -1$.

$$p^{do} = \begin{cases} A - B + C - D + F, & K > L \\ F, & K \leq L \end{cases} \quad (6.112)$$

(Up-Out Call) Let $\eta = -1$ and $\phi = 1$.

$$p^{do} = \begin{cases} F, & K > L \\ A - B + C - D + F, & K \leq L \end{cases} \quad (6.113)$$

(Up-Out Put) Let $\eta = -1$ and $\phi = -1$.

$$p^{uo} = \begin{cases} B - D + F, & K > L \\ A - C + F, & K \leq L \end{cases} \quad (6.114)$$

(Down-In Call) Let $\eta = 1$ and $\phi = 1$.

$$c^{di} = \begin{cases} C + E, & K > L \\ A - B + D + E, & K \leq L \end{cases} \quad (6.115)$$

(Down-In Put) Let $\eta = 1$ and $\phi = -1$.

$$p^{di} = \begin{cases} B - C + D + E, & K > L \\ A + E, & K \leq L \end{cases} \quad (6.116)$$

(Up-In Call) Let $\eta = -1$ and $\phi = 1$.

$$c^{ui} = \begin{cases} A + E, & K > L \\ B - C + D + E, & K \leq L \end{cases} \quad (6.117)$$

(Up-In Put) Let $\eta = -1$ and $\phi = -1$.

$$p^{ui} = \begin{cases} A - B + D + E, & K > L \\ C + E, & K \leq L \end{cases} \quad (6.118)$$

(Lookback Options) Let $M_{t,T} = \sup_{t \leq \tau \leq T} S_\tau$ and $m_{t,T} = \inf_{t \leq \tau \leq T} S_\tau$.

$$d_1 = \frac{\ln(\frac{S_t}{K}) + \alpha^+ \tau}{\sigma \sqrt{\tau}}. \quad d_2 = d_1 - \sigma \sqrt{\tau}. \quad (6.119)$$

$$e_1 = \frac{\ln(\frac{S_t}{M_{0,t}}) + \alpha^+ \tau}{\sigma \sqrt{\tau}}. \quad e_2 = e_1 - \sigma \sqrt{\tau}. \quad (6.120)$$

$$f_1 = \frac{\ln(\frac{S_t}{m_{0,t}}) + \alpha^+ \tau}{\sigma \sqrt{\tau}}. \quad f_2 = f_1 - \sigma \sqrt{\tau}. \quad (6.121)$$

(Fixed Lookback Call) If $K > M_{0,t}$, then

$$\begin{aligned} C &= S_0 e^{-\delta \tau} N(d_1) - K e^{-r \tau} N(d_2) \\ &+ S_0 e^{-r \tau} \frac{\sigma^2}{2(r-\delta)} \left[- \left(\frac{S_0}{K} \right)^{-\frac{2(r-\delta)}{\sigma^2}} N \left(d_1 - \frac{2(r-\delta)}{\sigma} \right) + e^{(r-\delta)\tau} N(d_1) \right]. \end{aligned} \quad (6.122)$$

Otherwise,

$$\begin{aligned} C &= e^{-r \tau} (M_{0,t} - K) + S_0 e^{-\delta \tau} N(e_1) - M_{0,t} e^{-r \tau} N(e_2) \\ &+ S_0 e^{-r \tau} \frac{\sigma^2}{2(r-\delta)} \left[- \left(\frac{x}{M_{0,t}} \right)^{-\frac{2(r-\delta)}{\sigma^2}} N \left(e_1 - \frac{2(r-\delta)}{\sigma} \sqrt{\tau} \right) + e^{(r-\delta)\tau} N(e_1) \right]. \end{aligned} \quad (6.123)$$

(Fixed Lookback Put) If $K < m_{0,t}$, then

$$\begin{aligned} P &= K e^{-r \tau} N(-d_2) - S_0 e^{-\delta \tau} N(-d_1) \\ &+ S_0 e^{-r \tau} \frac{\sigma^2}{2(r-\delta)} \left[\left(\frac{x}{K} \right)^{-\frac{2(r-\delta)}{\sigma^2}} N \left(-d_1 + \frac{2b}{\sigma} \sqrt{\tau} \right) - e^{(r-\delta)\tau} N(-d_1) \right]. \end{aligned} \quad (6.124)$$

Otherwise,

$$\begin{aligned} P &= e^{-r \tau} (K - m_{0,t}) - S_0 e^{-\delta \tau} N(-f_1) + m_{0,t} e^{-r \tau} N(-f_2) \\ &+ S_0 e^{-r \tau} \frac{\sigma^2}{2(r-\delta)} \left[\left(\frac{S_0}{m_{0,t}} \right)^{-\frac{2(r-\delta)}{\sigma^2}} N(-f_1 + \frac{2b}{\sigma} \sqrt{\tau}) - e^{(r-\delta)\tau} N(-f_1) \right]. \end{aligned} \quad (6.125)$$

(Floating Lookback Call)

$$\begin{aligned} C &= S_0 e^{-\delta \tau} N(f_1) - m_{0,t} e^{-r \tau} N(f_2) \\ &+ S_0 e^{-r \tau} \frac{\sigma^2}{2b} \left[\left(\frac{S_0}{m_{0,t}} \right)^{-\frac{2(r-\delta)}{\sigma^2}} N(-f_1 + \frac{2(r-\delta)}{\sigma} \sqrt{\tau}) - e^{(r-\delta)\tau} N(-f_1) \right]. \end{aligned} \quad (6.126)$$

(Floating Lookback Put)

$$\begin{aligned} P &= M_{0,t} e^{-r \tau} N(-e_2) - S_0 e^{-\delta \tau} N(-e_1) \\ &+ S_0 e^{-r \tau} \frac{\sigma^2}{2(r-\delta)} \left[- \left(\frac{S_0}{M_{0,t}} \right)^{-\frac{2b}{\sigma^2}} N \left(e_1 - \frac{2(r-\delta)}{\sigma} \sqrt{\tau} \right) + e^{(r-\delta)\tau} N(b_1) \right]. \end{aligned} \quad (6.127)$$

The shout option price is

$$sh_t = \mathbb{E}_t \left[b_{t,T}(S_T - K)_+ 1_{S_T \geq S_{\tau_{sh}}} b_{t,T} + (S_{\tau_{sh}} - K) 1_{S_T < S_{\tau_{sh}}} \right] \quad (6.128)$$

The first part is evaluated as

$$\begin{aligned} \mathbb{E}_t[b_{t,T}(S_T - K) 1_{S_T \geq K, S_T \geq S_{sh}}] &= S_t \mathbb{E}_t \left[b_{t,T} e^{(r-\delta)\tau} \eta_{t,T} 1_{\epsilon_1 \wedge \epsilon_2} \right] - K e^{-r\tau} \mathbb{E}_t[1_{\epsilon_1 \wedge \epsilon_2}] \\ &= S_t e^{-\delta\tau} N(d(S_t, S_{sh}, \tau)) - K e^{-r\tau} N(d(S_t, S_{sh}, \tau) - \sigma\sqrt{\tau}). \end{aligned} \quad (6.129)$$

The second part is evaluated as

$$\mathbb{E}[b_{t,T}(S_{sh} - K) 1_{S_{sh} > K, S_{sh} > S_T}] = e^{-r(T-t)} (S_{sh} - K) N(-d(S_t, S_{sh}, \tau) + \sigma\sqrt{\tau}). \quad (6.130)$$

The geometric Asian option price is

$$\begin{aligned} APAC_t &= \mathbb{E}_t[b_{t,T}(G_T - K)_+] \\ &= S e^{(b^G - r)\tau} N(d(S_t, K, \tau)) - K e^{-r\tau} N(d(S_t, K, \tau) - \sigma^G \sqrt{\tau}). \end{aligned} \quad (6.131)$$

By definition,

$$\begin{aligned} G_T &= \exp \left(\frac{1}{T} \int_0^T (\ln S_0 + \alpha^- u + \sigma W_u) du \right) \\ &= S_0 \exp \left(\frac{\alpha^- T}{2} + \frac{\sigma}{T} \int_0^T W_u du \right) \end{aligned} \quad (6.132)$$

where

$$\int_0^T W_u du \sim \mathcal{N} \left(0, \frac{T^3}{3} \right) \sim \frac{T}{\sqrt{3}} W_T. \quad (6.133)$$

Thus,

$$\sigma^G = \frac{\sigma}{\sqrt{3}}. \quad (6.134)$$

$$b^G - \frac{1}{2}(\sigma^G)^2 = \frac{1}{2} \left(b - \frac{1}{2}\sigma^2 \right) \implies b^G = \frac{1}{2} \left(b - \frac{\sigma^2}{6} \right). \quad (6.135)$$

Definition 6.39. The *Greeks* are the quantities, denoted as Greek letters, representing the sensitivity of the price of derivatives to a change in underlying parameters on which the value of an instrument or portfolio of financial instruments is dependent.

Greek	European Call Option (BS)	European Put Option (BS)
$\Delta = \frac{\partial V}{\partial S}$	$N(d_1)$	$N(d_1) - 1$
$\Gamma = \frac{\partial \Delta}{\partial S} = \frac{\partial^2 V}{\partial S^2}$		$\frac{N'(d_1)}{S\sigma\sqrt{t}}$
$\nu = \frac{\partial V}{\partial \sigma}$		$SN'(d_1)\sqrt{t}$
$\Theta = \frac{\partial V}{\partial t}$	$-\frac{SN'(d_1)\sigma}{2\sqrt{t}} - rKe^{-rt}N(d_2)$	$-\frac{SN'(d_1)\sigma}{2\sqrt{t}} + rKe^{-rt}N(-d_2)$
$\rho = \frac{\partial V}{\partial r}$	$Kte^{-rt}N(d_2)$	$-Kte^{-rt}N(-d_2)$

Remark 6.40. We can rewrite the CEV PDE in Greeks.

$$\Theta + \frac{1}{2}\sigma^2 s^{2\beta}\Gamma + rs\Delta = rc. \quad (6.136)$$

There is a trade-off between gamma and theta. If gamma is large (which is good), then theta decay is also large (which is not good).

Definition 6.41. The Heston process is driven by two Brownian motions W_1 and W_2 with parameters $r, \kappa, \theta, \rho, \sigma$.

$$dS_t = rS_t dt + \sqrt{\nu_t}S_t dW_t^1, \quad (6.137)$$

$$d\nu_t = \kappa(\theta - \nu_t)dt + \sigma\sqrt{\nu_t}dW_t^2, \quad (6.138)$$

$$\text{Cov}(dW_t^1, dW_t^2) = \rho dt. \quad (6.139)$$

The price process is log-normal and the volatility process follows CIR. The initial volatility σ_0 must be chosen.

Definition 6.42. The characteristic function is the Fourier Transform of the probability density function. The probability density function can be recovered from the Fourier Inverse.

$$\psi(u) = \int e^{iux} \varphi(x) dx. \quad \Longleftrightarrow \quad \varphi(x) = \frac{1}{2\pi} \int e^{-iux} \psi(u) du. \quad (6.140)$$

Theorem 6.43. (Fourier Transform of Call Price with a Damping Factor) Let $\alpha > 0$ be a damping factor and $C_T(k)$ be the call price function of the log-strike k . Define $\tilde{C}_T(k)$ as

$$\tilde{C}_T(k) = e^{\alpha k} C_T(k). \quad (6.141)$$

Let $\Psi(\nu)$ be the Fourier Transform of $\tilde{C}_T(k)$ and $\psi(\nu)$ be the characteristic function.

$$\Psi(\nu) = \int e^{\alpha k} e^{i\nu k} C_T(k) dk \quad \Longleftrightarrow \quad C_T(k) = \frac{e^{-\alpha k}}{2\pi} \int e^{-i\nu k} \Psi(\nu) d\nu \quad (6.142)$$

where

$$\Psi(\nu) = \frac{e^{-\int_0^T r_u du}}{(\alpha + i\nu)(\alpha + i\nu + 1)} \psi(\nu - (\alpha + 1)i). \quad (6.143)$$

Proof. Let $\varphi(s)$ be the probability density function of log-prices. Given the Fourier Transform,

$$\begin{aligned} \Psi(\nu) &= \int_{-\infty}^{\infty} e^{i\nu k} \tilde{C}_T(k) dk \\ &= e^{-\int_0^T r_u du} \int_{-\infty}^{\infty} e^{(i\nu + \alpha)k} \int_k^{\infty} (e^s - e^k) \varphi(s) ds dk \\ &= e^{-\int_0^T r_u du} \int_{-\infty}^{\infty} \varphi(s) \int_{-\infty}^s e^{(i\nu + \alpha)k} (e^s - e^k) dk ds \\ &= e^{-\int_0^T r_u du} \int_{-\infty}^{\infty} \varphi(s) \cdot \frac{e^{(\alpha + i\nu + 1)s}}{(\alpha + i\nu)(\alpha + i\nu + 1)} ds \\ &= \frac{e^{-\int_0^T r_u du}}{(\alpha + i\nu)(\alpha + i\nu + 1)} \int_{-\infty}^{\infty} e^{i(\nu - (\alpha + 1)i)s} \varphi(s) ds \\ &= \frac{e^{-\int_0^T r_u du}}{(\alpha + i\nu)(\alpha + i\nu + 1)} \psi(\nu - (\alpha + 1)i). \end{aligned} \quad (6.144)$$

□

Theorem 6.44. (Option pricing via FFT) Choose $N = 2^n$, $\Delta\nu$ and α and let $\Delta k = \frac{2\pi}{N\delta\nu}$. The vector of x_j is obtained by

$$x_j = \frac{2 - \delta_{j-1}}{2} \Delta\nu \frac{\exp(-i(s_0 - \frac{\Delta k N}{2})\nu_j - \int_0^T r_u du)}{(\alpha + i\nu_j)(\alpha + i\nu_j + 1)} \psi(v_j - (\alpha + 1)i). \quad (6.145)$$

where $\nu_j = (j - 1)\Delta\nu$. Call FFT to obtain y_j .

$$C_T(k_j) = \frac{e^{-\alpha(s_0 - \Delta k(\frac{N}{2} - (j-1)))}}{\pi} \text{Re}(y_j). \quad (6.146)$$

Definition 6.45. The characteristic function of the Heston model is

$$\psi(u) = \frac{\exp\left(iu(s_0 + (r - q)t) + \frac{\kappa\theta t(\kappa - i\rho\sigma u)}{\sigma^2} - \frac{(u^2 + iu)\nu_0}{\lambda \coth(\frac{\lambda t}{2}) + \kappa - i\rho\sigma u}\right)}{\left(\cosh\left(\frac{\lambda t}{2}\right) + \frac{\kappa - i\rho\sigma u}{\lambda} \cdot \sinh\left(\frac{\lambda t}{2}\right)\right)^{\frac{2\kappa\theta}{\sigma^2}}} \quad (6.147)$$

where

$$\lambda = \sqrt{\sigma^2(u^2 + iu) + (\kappa - i\rho\sigma u)^2}. \quad (6.148)$$

Definition 6.46. The Stochastic-Alpha-Beta-Rho (SABR) process is driven by the two Brownian motions W_1 and W_2 .

$$dS_t = rS_t dt + \sigma_t S_t^\beta dW_t^1, \quad (6.149)$$

$$d\sigma_t = \alpha\sigma_0 dW_t^2, \quad (6.150)$$

$$\text{Cov}(dW_t^1, dW_t^2) = \rho dt. \quad (6.151)$$

The price process follows CEV and the volatility process is log-normal. The initial volatility σ_0 must be chosen. This model is most commonly used in modeling rates and foreign exchange markets.

Remark 6.47. Under certain sets of parameters, we can recover BS-model from both process. This is of crucial importance in practice because it gives us a set of natural tests of our code. This extra parameter in the Heston model enable us to fit the slope of the volatility surface in expiry in addition to strike. As a result, when calibrating a SABR model, we generally calibrate a set of calibration parameters per expiry. Conversely, when calibrating a Heston model, we are able to calibrate all expiries in a single calibration.

Theorem 6.48. (SABR calibartion)

$$F_t = S_t e^{rt}. \quad (6.152)$$

$$\tilde{\sigma}_t = \frac{\alpha \ln\left(\frac{F_t}{K}\right)}{x(z)} \cdot \frac{A(F_t, K)}{B(F_t, K)}. \quad (6.153)$$

$$A(F_t, K) = 1 + \left[\frac{\sigma_0^2(1 - \beta)^2}{24(F_t K)^{1-\beta}} + \frac{\alpha\beta\rho\sigma_0}{4(F_t K)^{\frac{1-\beta}{2}}} + \alpha^2 \frac{2 - 3\rho^2}{24} \right] t + \dots \quad (6.154)$$

$$B(F_t, K) = 1 + \frac{1}{24} \left[(1 - \beta) \ln\left(\frac{F_t}{K}\right) \right]^2 + \frac{1}{1920} \left[(1 - \beta) \ln\left(\frac{F_t}{K}\right) \right]^4 + \dots \quad (6.155)$$

$$x(z) = \ln \left(\frac{\sqrt{1 - 2\rho z + z^2} + z - \rho}{1 - \rho} \right). \quad (6.156)$$

$$z = \frac{\alpha}{\sigma_0} (F_t K)^{\frac{1-\beta}{2}} \ln \left(\frac{F_t}{K} \right). \quad (6.157)$$

Definition 6.49. The Variance Gamma process is driven by a Brownian motion W .

$$X(t; \sigma, \nu, \theta) = \theta \gamma(t; 1, \nu) + \sigma W(\gamma(t; 1, \nu)), \quad (6.158)$$

$$\ln(S_t) = \ln(S_0) + (r - q + \omega)t + X(t; \sigma, \nu, \theta). \quad (6.159)$$

This model uses a jump process which can be seen as a Brownian motion process with a random clock. Time changes follow a Gamma distribution with mean 1 and variance ν . Allows us to control skewness and kurtosis of distribution. ω is selected to ensure that the process is a martingale.

Remark 6.50. Jump processes are of particular use when trying to model extreme volatility skews of very short dated options. In these cases, implied volatility for out-of-the-money options will need to be raised to an unreasonable level in order to fit the market price. Variance Gamma is comparable to SABR in that when are calibrating the model we generally pick a different set of parameters per option expiry.

Definition 6.51. (Parameter calibration) Let \hat{c} and c be the model prices and market prices. The best set of parameters can be found by minimizing the squared pricing errors.

$$\vec{p}_{\min} = \operatorname{argmin}_{\vec{p}} \sum_{\tau, K} \omega \cdot |\hat{c}(\tau, K, \vec{p}) - c(\tau, K)|^n \quad (6.160)$$

Remark 6.52. The implied volatility $\tilde{\sigma}$ can be acquired by calibration.

Theorem 6.53. (Breeden-Litzenberger) Let c be the payoff of a European call option. Then, the probability density function of strikes can be extracted from the market option prices.

$$\varphi(K) \approx e^{\int_0^T r_u du} \frac{c(K-h) - 2c(K) + c(K+h)}{h^2}. \quad (6.161)$$

This method only works for one-dimensional density. It also assumes that prices are available and exact for every strike.

Definition 6.54. Given a probability space and a filtration that is right-continuous and complete, given a sequence of stopping times $(T^k)_{k \geq 0}$, the counting process is

$$N_t = \sum 1_{T^k \leq t} \quad (6.162)$$

such that $N_t < \infty$ and N_t has the same information as T^k .

Definition 6.55. The process λ_t is the intensity of the counting process N if

$$\mathbb{E}_t [N_{t+\Delta} - N_t] \approx \lambda_t \Delta \quad (6.163)$$

where λ_t , mean conditional arrival rate of jumps of N , gives a measure of how many jumps of N over a small time period after time t .

Theorem 6.56. Any counting process is a submartingale because it is an increasing process given that $N_t \leq \mathbb{E}_t[N_t]$ almost surely for $t < T$.

Theorem 6.57. Given the Doob-Meyer Decomposition, the counting process N_t , a submartingale, can be decomposed into a local martingale M_t and an increasing predictable process A_t , also known as the compensator. It has continuous paths if and only if the jump times T^k are totally inaccessible. Since any increasing and continuous function is differentiable almost everywhere. A can be written as

$$A_t = \int_0^t \lambda_s ds \quad (6.164)$$

for some non-negative predictable process λ_t .

Theorem 6.58. The process $M_t = N_t - A_t$ is a martingale if and only if $\mathbb{E} \left[\int_0^t \lambda_s ds \right] < \infty$ for all $t \geq 0$.

Corollary 6.59. If $\mathbb{E} \left[\int_0^t \lambda_s ds \right] < \infty$, then

$$\begin{aligned} 0 &= \mathbb{E}_t[M_{t+\Delta}] = \mathbb{E}_t[N_{t+\Delta} - N_t] - \mathbb{E}_t[A_{t+\Delta} - A_t] \\ &\iff \mathbb{E}_t[N_{t+\Delta} - N_t] = \mathbb{E}_t \left[\int_t^{t+\Delta} \lambda_s ds \right] \approx \lambda_t \Delta. \end{aligned} \quad (6.165)$$

Thus, the process λ_t is the intensity of N .

Theorem 6.60. The process

$$Z_t(u) = \exp(\phi(iu)A_t - iuN_t) \quad (6.166)$$

is a martingale if $\phi(u) = 1 - e^{-u}$. This process is also known as the characteristic martingale of N .

Remark 6.61. It follows that

$$1 = \mathbb{E}_t \left[\frac{Z_T}{Z_t} \right] = \mathbb{E}_t [\exp(\phi(iu)(A_T - A_t) - iu(N_T - N_t))]. \quad (6.167)$$

Then,

$$\mathbb{E}_t \left[e^{-iu(N_T - N_t)} \right] = \exp((e^{-iu} - 1)(A_T - A_t)). \quad (6.168)$$

The right-hand side is exactly the characteristic function of a Poisson distribution with the rate $(A_T - A_t)$ given the condition of the jump increment $(N_T - N_t)$ which means

$$\begin{aligned} P_t[N_T - N_t = n] &= \frac{1}{n!} (A_T - A_t)^n e^{-(A_T - A_t)} \\ &= \frac{1}{n!} \left(\int_t^T \lambda_s ds \right)^n e^{-\int_t^T \lambda_s ds}. \end{aligned} \quad (6.169)$$

Theorem 6.62. (Watanabe's Theorem) A counting process is a Poisson process if and only if it has a deterministic compensator.

Definition 6.63. (Types of Poisson Processes) The compensator A is deterministic (stochastic) if λ is deterministic (stochastic). N is an inhomogeneous Poisson process if the intensity $\lambda_t = f(t)$ is time-varying. N is a homogeneous (standard) Poisson process if $\lambda_t := \lambda$ is constant (one).

Theorem 6.64. *The homogeneous Poisson process has three properties.*

1. For any $0 \leq t < s$, the increment $N_s - N_t$ is independent of \mathcal{F}_t .
2. For any $0 \leq t < s$ and $0 \leq u < v$ with $s-t = v-u$, two increments have the same distribution.
3. For any $0 < t$, N_t is Poisson distributed with λt

Remark 6.65. *It resembles three properties of Brownian motion.*

Definition 6.66. *Suppose A is random, then the process N is doubly-stochastic Poisson process. With $G_t = \sigma(A_t) \subset F_t$ the filtration generated by A ,*

$$\mathbb{E}_t \left[e^{-iu(N_T - N_t)} \right] = \mathbb{E}_t \left[e^{-iu(A_T - A_t)} \right]. \quad (6.170)$$

Remark 6.67. *(Affine Doubly-Stochastic Process) Take $\lambda_t = \Lambda(X_{t-})$ for an affine function $\Lambda(x) = \Lambda_0 + \Lambda_1 x$. Suppose that X solves the affine SDE*

$$dX_t = \mu(X_t)dt + \sigma(X_t) dW_t \quad (6.171)$$

with $\mu(x) = K_0 + K_1 x$ and $\sigma^2(x) = H_0 + H_1 x$ are affine. Also,

$$\mathbb{E}_t \left[\exp \left(-\phi(iu) \int_t^T \Lambda(X_s) ds \right) \right] = \exp(a(t) + b(t)X_t) \quad (6.172)$$

where the functions a and b satisfy Ricatti equations. As a result,

$$\mathbb{E}_t \left[e^{-iu(N_T - N_t)} \right] = \mathbb{E}_t \left[e^{-iu(A_T - A_t)} \right] = \exp(a(t) + b(t)X_t). \quad (6.173)$$

Definition 6.68. *A point process is defined as*

$$L_t = \sum_{n=1}^{N_t} L^n, \quad (6.174)$$

where N_t is a counting process and (L^n) is a sequence of random variables.

Remark 6.69. *If N_t is a Poisson process, then L_t is a compound Poisson process.*

Definition 6.70. *Suppose $\lambda_0 = c$. The generalized Hawkes process is*

$$d\lambda_t = \kappa(c - \lambda_t) dt + \sigma \sqrt{\lambda_t} dW_t + \delta dK_t. \quad (6.175)$$

Let $\sigma = 0$ to eliminate the stochastic component. The Hawkes process is

$$d\lambda_t = \kappa(c - \lambda_t) dt + \delta dK_t, \quad (6.176)$$

whose solution is

$$\lambda_t = c + \delta \int_0^t e^{-\kappa(t-s)} dK_s = c + \delta \sum_{n=1}^{N_t} e^{-\kappa(t-T^n)} K^n \quad (6.177)$$

1. The intensity jumps up after a jump (self-exciting)

2. If no further jump arrives, then the intensity reverts back to its mean c at an exponential rate κ (mean-reverting)

Let $\kappa = 0$. The birth process is

$$d\lambda_t = \delta dK_t \quad (6.178)$$

whose solution is

$$\lambda_t = c + \delta \sum_{n=1}^{N_t} K^n. \quad (6.179)$$

1. A jump of N_t causes a jump in the point process L_t and also a jump in the intensity of future jump arrivals.
2. This process resembles the process of population births.

A self-exciting process is a birth process with non-decreasing λ_t . Let $\delta = 0$. Then, it is reduced to the homogeneous Poisson process.

Remark 6.71. A drawback of the birth process is that $\lambda_t \rightarrow \infty$ as $t \rightarrow \infty$. In many realistic applications of point processes, the intensity of jump arrival will tend to be self-exciting but then revert back to an average value as if no jumps occur, which can be simulated by the (generalized) .

Definition 6.72. The jump-diffusion process is

$$dX_t = \mu(X_t) dt + \sigma(X_t) dB_t + dL_t, \quad L_t = \sum_{n=1}^{N_t} \gamma(X_{T_n-}). \quad (6.180)$$

Theorem 6.73. Suppose a jump-diffusion X is affine SDE such that $\mu(x) = k_0 + k_1x$ and $\sigma^2(x) = h_0 + h_1x$ and the jump is i.i.d. with density p . The jump counting process N has intensity $\lambda_t = \Lambda(X_{t-})$ for the affine function $\Lambda(x) = l_0 + l_1x$. The characteristic function is

$$\psi_{t,T}(u) = e^{\alpha(t) + \beta(t)X_t} \quad (6.181)$$

where α and β solve the Ricatti equations

$$\alpha'(t) = -k_0\beta(t) - \frac{1}{2}h_0\beta^2(t) - l_0 \left(\mathbb{E} \left[e^{\beta(t)L^1} \right] - 1 \right), \quad \alpha(T) = 0. \quad (6.182)$$

$$\beta'(t) = -k_1\beta(t) - \frac{1}{2}h_1\beta^2(t) - l_1 \left(\mathbb{E} \left[e^{\beta(t)L^1} \right] - 1 \right), \quad \beta(T) = iu. \quad (6.183)$$

which are integro-differential equations and are difficult to solve unless $l_1 = 0$.

Theorem 6.74. (Feynman-Kac with Jumps) Suppose that the functions $\mu(x)$ and $\Sigma(x)$ are Lipschitz continuous in x and $\mathbb{E} \left[\int_0^T \Lambda(X_s) ds \right] < \infty$. Suppose there exists a function $c : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ that solves the partial differential equation

$$\begin{aligned} 0 = & g(t, x) + \frac{\partial c(t, x)}{\partial t} + \mu(x) \frac{\partial c(t, x)}{\partial x} + \frac{\sigma^2(x)}{2} \frac{\partial^2 c(t, x)}{\partial x^2} - \rho(t, x) c(t, x) \\ & + \lambda(x) \mathbb{E} [c(t, x + L^1) - c(t, x)] \end{aligned} \quad (6.184)$$

for $0 \leq t < T$ with boundary condition $c(T, x) = h(x)$. In addition, assume that the function c satisfies

$$\mathbb{E} \left[\int_0^T e^{-2 \int_0^t \rho(s, X_s) ds} \left(\frac{\partial c(t, X_t)}{\partial x} \sigma(X_t) \right)^2 dt \right] < \infty \quad (6.185)$$

and

$$\sup_{t \in [0, T]} \sup_{x \in \mathbb{R}} \mathbb{E} [|c(t, x + L^1) - c(t, x)|] < \infty. \quad (6.186)$$

Then, almost surely,

$$c_t = \mathbb{E}_t \left[e^{-\int_t^T \rho(s, X_s) ds} h(X_T) + \int_t^T e^{-\int_t^s \rho(u, X_u) du} g(s, X_s) ds \right] \quad (6.187)$$

Theorem 6.75. (Merton Model) The stock price process S solves the SDE:

$$dS_t = \mu S_{t-} dt + \sigma S_{t-} dB_t + S_{t-} dJ_t \quad (6.188)$$

for a point process of the type

$$J_t = \sum_{j=1}^{N_t} (Y^j - 1) \quad (6.189)$$

with positive i.i.d. random variables (Y^j) . The counting process $N_t = \sum_{k \geq 1} 1_{T^k \leq t}$ has constant intensity λ and jump times $(T^k)_{k \geq 1}$.

$$S_t = S_0 e^{(\mu - \frac{\sigma^2}{2})t + \sigma B_t} \sum_{j=1}^{N_t} Y^j \quad (6.190)$$

In particular, if the jump sizes Y^j are log-normally distributed such that $Y^j \sim \mathcal{LN}(a, b^2)$, then S_t is log-normally distributed given $N_t = n$ such that

$$S_t |_{N_t=n} \sim \mathcal{LN} \left(\ln S_0 + \left(\mu - \frac{\sigma^2}{2} \right) t + an, \sigma^2 t + b^2 n \right). \quad (6.191)$$

Remark 6.76. (Options in the Merton model) Jumps are an additional source of risk for options. Option prices will reflect the fact that the underlying stock price may jump.

Theorem 6.77. If $\lambda_t > 0$ for all $t > 0$, then the variables

$$S^k = A_{T^k} = \int_0^{T^k} \lambda_s ds \quad (6.192)$$

are the arrival times of a standard Poisson process in time-scaled filtration defined by the stopping-time σ -fields. This means that T^k is the hitting time of A to the random level $S^k = E^1 + \dots + E^k$ where the E^n are i.i.d. $\exp(1)$:

$$T^k = A_{S^k}^{-1} = \inf \left\{ t : \int_0^t \lambda_s ds \geq S^k \right\} \quad (6.193)$$

Therefore, we can simulate the jump times T^k sequentially using the following reformulation of Meyer's Theorem: Conditional on $T^{k-1} = t$, the k -th jump is given by

$$T^k = \inf \left\{ T > t : \int_t^T \lambda_s ds \geq E^k \right\}, \quad (6.194)$$

where E^k is a sequence of i.i.d. exponential random variables.

Theorem 6.78. (Euler Discretization with Jumps) Suppose that X is a jump-diffusion process that solves

$$dX_t = \mu(X_t) dt + \sigma(X_t) dB_t + dJ_t \quad (6.195)$$

for a standard Brownian motion B and a jump process $J_t = \sum_{n=1}^{N_t} \Gamma(X_{T_{n-}}, Z^n)$, where Z^n is a sequence of i.i.d. random variables. The counting process N has intensity $\lambda_t = \Lambda(X_t) > 0$ for all $t > 0$. Let (E^k) be a sequence of i.i.d. standard exponential random variables.

1. Initialize $t = 0$, $\hat{X}_t = X_0$, $\hat{A}_t = 0$, $S = E^1$, $n = 0$, $\hat{N}_t = 0$.
2. Compute $A^{temp} = \hat{A}_t + \Lambda(\hat{X}_t)\Delta$
3. IF: $A^{temp} \geq S$, a jump has occurred between times t and $t + \Delta$.

$$(a) \text{ Set } \hat{T}^k = t + \frac{S - \hat{A}_t}{A^{temp} - \hat{A}_t} \Delta.$$

(b) Compute $\hat{X}_{\hat{T}^k-} = \hat{X}_t + \mu(\hat{X}_t) (\hat{T}^k - t) + \sigma(\hat{X}_t) (B_{\hat{T}^k} - B_t)$. (Euler step until right before the jump)

(c) Compute $\hat{X}_{\hat{T}^k} = \hat{X}_{\hat{T}^k-} + \Gamma(X_{\hat{T}^k-}, Z^k)$ (Incorporate Jump of X)

(d) Update $t \leftarrow \hat{T}^k$, $\hat{A}_t \leftarrow S$, $n \leftarrow n + 1$, $S \leftarrow S + E^n$, and $\hat{N}_t \leftarrow n$.

ELSE: No jump has occurred between times t and $t + \Delta$.

(a) Compute the standard Euler step

(b) Update $t \leftarrow t + \Delta$, $\hat{A}_t \leftarrow A^{temp}$, and $N_{t+\Delta} \leftarrow N_t$.

4. If $t < T$, go to Step 2.

Definition 6.79. (Merton Model) Under the risk neutral measure \mathbb{Q} , the value of the firm's assets is currently $V_0 > 0$ and follows a geometric Brownian motion

$$dV_t = rV_t dt + \sigma V_t dW_t. \quad (6.196)$$

At time T , the firm's managers will be faced with a payment obligation of \bar{D} . If $V_T < \bar{D}$, managers will declare bankruptcy of the firm. By the APR, debt holders receive V_T and the shareholders receive 0. If $V_T \geq \bar{D}$, by the APR, the bondholders receive \bar{D} and the shareholders receive the residual value $S = V_T - \bar{D}$.

Remark 6.80. Both equity and debt can be viewed as derivative securities on the value V of the firm's assets, whose payoffs are as follows.

$$S(T) = \max\{V_T - \bar{D}, 0\}. \quad (6.197)$$

$$\bar{B}_M(T, T) = \min\{V_T, \bar{D}\} = \bar{D} - \max\{\bar{D} - V_T, 0\}. \quad (6.198)$$

Definition 6.81. The survival probability of the firm is

$$P(0, T) = \mathbb{Q}(V_T \geq \bar{D}) = \mathcal{N}(d_2). \quad (6.199)$$

The price of the defaultable debt is

$$\bar{B}_M(0, T) = \bar{D}e^{-rT}\mathcal{N}(d_2) + V_0\mathcal{N}(-d_1) \quad (6.200)$$

where

$$d_{1,2} = \frac{\ln V_0/\bar{D} + (r \pm \sigma^2/2)T}{\sigma\sqrt{T}}. \quad (6.201)$$

Definition 6.82. The continuous yield to maturity $y(T)$ for the defaultable debt is defined as the solution to

$$\bar{B}_M(0, T) = \bar{D}e^{-y(T)T} \iff y(T) = \frac{\ln(\bar{D}/\bar{B}_M(0, T))}{T}. \quad (6.202)$$

Definition 6.83. The credit spread is defined as the risk premium with respect to risk-free rate.

$$s(T) = y(T) - r. \quad (6.203)$$

The credit spread is given by

$$s(T) = -\frac{1}{T} \ln \left(\mathcal{N}(d_2) + \frac{V_0}{\bar{D}e^{-rT}} \mathcal{N}(-d_1) \right) \quad (6.204)$$

Definition 6.84. The debt ratio is defined as

$$L = \frac{\bar{D}e^{-rT}}{V_0} \quad (6.205)$$

which represents a measure of leverage.

Definition 6.85. Black and Cox model uses the same setting in Merton model and includes a safety covenant which monitors poor performance and triggers a premature default which can be modeled as a stopping time. The standard for poor performance is set in terms of a time-dependent deterministic barrier $\bar{v}_t = Ke^{-r(T-t)}$. As soon as the value of the assets crosses this threshold before maturity, the debt holders take over the firm, liquidate the assets and recover the amount \bar{v}_T . Otherwise, default occurs at maturity or not depends on whether $V_T < \bar{D}$ or not.

Remark 6.86. The following inequalities hold.

- $V_0 > \bar{v}_0$. It is imposed to avoid default at inception date.
- $Ke^{-r(T-\tau)} \leq \bar{D}$. It ensures that the payoff to the bond holders at the default time τ never exceeds the face value of debt.

Definition 6.87. (Decomposition of default) If the default time $\tau = \inf\{t \in [0, T] : V_t \leq v_t\}$ exists, we have $V_\tau = v_\tau$. Let $\tau = \bar{\tau} \wedge \hat{\tau}$ where $\bar{\tau}$ is the early default time and $\hat{\tau}$ is the default time at T .

Theorem 6.88. Let $M_t = \sup W_t$ be the running maximum. For every $x \geq 0$,

$$P(M_t \geq x) = 2P(W_t \geq x) = 2\mathcal{N}(-x/\sqrt{t}). \quad (6.206)$$

Theorem 6.89. Let $X_t = \alpha t + W_t$ and $M_t^X = \sup X_t$. Let $y \geq 0$ and $y \geq x$. Then,

$$P(X_t \leq x, M_t^X \geq y) = e^{2\alpha y} \mathcal{N}\left(\frac{x - 2y - \alpha t}{\sqrt{t}}\right). \quad (6.207)$$

Proof. Let \mathbb{Q}^* be the probability measure on (Ω, \mathcal{F}_t) given by $d\mathbb{Q}^* = e^{-\alpha W_t - \frac{\alpha^2}{2}t} d\mathbb{Q}$ so that $W_t^* = X_t - \alpha t$ follows a standard Brownian motion under \mathbb{Q}^* and $d\mathbb{Q}^* = e^{\alpha W_t^* - \frac{\alpha^2}{2}t} d\mathbb{Q}$. Then, the joint probability follows

$$\begin{aligned} P(X_t \leq x, M_t^X \geq y) &= \mathbb{E}^{\mathbb{Q}} \left[1_{X_t \leq x, M_t^X \geq y} \right] = \mathbb{E}^{\mathbb{Q}^*} \left[e^{\alpha W_t^* - \frac{\alpha^2}{2}t} 1_{W_t^* \leq x, M_t^{W^*} \geq y} \right] \\ &= \mathbb{E}^{\mathbb{Q}^*} \left[e^{\alpha \bar{W}_t - \frac{\alpha^2}{2}t} 1_{\bar{W}_t \leq x, M_t^{\bar{W}} \geq y} \right] \end{aligned} \quad (6.208)$$

where $\bar{W}_t = W_t^* + (2y - W_t^*)1_{\tau_y < t}$.

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}^*} \left[e^{\alpha \bar{W}_t - \frac{\alpha^2}{2} t} 1_{\bar{W}_t \leq x, M_t^{\bar{W}} \geq y} \right] &= \mathbb{E}^{\mathbb{Q}^*} \left[e^{\alpha(2y - W_t^*) - \frac{\alpha^2}{2} t} 1_{2y - W_t^* \leq x} \right] = e^{2\alpha y} \mathbb{E}^{\mathbb{Q}^*} \left[e^{-\alpha W_t^* - \frac{\alpha^2}{2} t} 1_{2y - W_t^* \leq x} \right] \\ &= e^{2\alpha y} \mathbb{E}^{\bar{\mathbb{Q}}} [1_{W_t^* \geq 2y - x}] = e^{2\alpha y} \bar{\mathbb{Q}}(W_t^* \geq 2y - x) \\ &= e^{2\alpha y} \bar{\mathbb{Q}}(\bar{W}_t - \alpha t \geq 2y - x) = e^{2\alpha y} \bar{\mathbb{Q}}(\bar{W}_t \geq 2y - x + \alpha t) \end{aligned} \quad (6.209)$$

where $\bar{\mathbb{Q}}$ is an equivalent probability measure $d\bar{\mathbb{Q}} = e^{-\alpha W_t - \frac{\alpha^2}{2} t} d\mathbb{Q}^*$. \square

Corollary 6.90. *Let $y \geq 0$ and $y \geq x$. Then,*

$$P(X_t \leq x, M_t^X \leq y) = \mathcal{N}\left(\frac{x - \alpha t}{\sqrt{t}}\right) - e^{2\alpha y} \mathcal{N}\left(\frac{x - 2y - \alpha t}{\sqrt{t}}\right). \quad (6.210)$$

Proof.

$$\begin{aligned} P(X_t \leq x, M_t^X \leq y) &= P(X_t \leq x) - P(X_t \leq x, M_t^X \geq y) \\ &= P\left(\frac{W_t}{\sqrt{t}} \leq \frac{x - \alpha t}{\sqrt{t}}\right) - e^{2\alpha y} \mathcal{N}\left(\frac{x - 2y - \alpha t}{\sqrt{t}}\right) \\ &= \mathcal{N}\left(\frac{x - \alpha t}{\sqrt{t}}\right) - e^{2\alpha y} \mathcal{N}\left(\frac{x - 2y - \alpha t}{\sqrt{t}}\right). \end{aligned} \quad (6.211)$$

\square

Corollary 6.91. *Let $y \geq 0$. Then,*

$$P(M_t^X \leq y) = \mathcal{N}\left(\frac{y - \alpha t}{\sqrt{t}}\right) - e^{2\alpha y} \mathcal{N}\left(\frac{-y - \alpha t}{\sqrt{t}}\right). \quad (6.212)$$

Proof.

$$\begin{aligned} P(M_t^X \geq y) &= P(X_t > y, M_t^X \geq y) + P(X_t \leq y, M_t^X \geq y) = P(X_t > y) + P(X_t \leq y, M_t^X \geq y) \\ &= P\left(\frac{W_t}{\sqrt{t}} > \frac{y - \alpha t}{\sqrt{t}}\right) + e^{2\alpha y} \mathcal{N}\left(\frac{-y - \alpha t}{\sqrt{t}}\right) \\ &= 1 - P\left(\frac{W_t}{\sqrt{t}} \leq \frac{y - \alpha t}{\sqrt{t}}\right) + e^{2\alpha y} \mathcal{N}\left(\frac{-y - \alpha t}{\sqrt{t}}\right) \\ &= 1 - \mathcal{N}\left(\frac{y - \alpha t}{\sqrt{t}}\right) + e^{2\alpha y} \mathcal{N}\left(\frac{-y - \alpha t}{\sqrt{t}}\right). \end{aligned} \quad (6.213)$$

\square

Corollary 6.92. *Let $y \geq 0$ and $y \geq x$. Then,*

$$\begin{aligned} P(X_t > x, M_t^X \leq y) &= P(M_t^X \leq y) - P(X_t \leq x, M_t^X \leq y) \\ &= \mathcal{N}\left(\frac{y - \alpha t}{\sqrt{t}}\right) - e^{2\alpha y} \mathcal{N}\left(\frac{-y - \alpha t}{\sqrt{t}}\right) \\ &\quad - \mathcal{N}\left(\frac{x - \alpha t}{\sqrt{t}}\right) + e^{2\alpha y} \mathcal{N}\left(\frac{x - 2y - \alpha t}{\sqrt{t}}\right). \end{aligned} \quad (6.214)$$

For the following theorems,

$$h_{5,1} = \frac{\ln \frac{V_0}{\bar{D}} + (r \pm \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}. \quad (6.215)$$

$$h_{6,2} = \frac{\ln \frac{V_0}{\bar{D}} + 2\ln \frac{\bar{v}_0}{V_0} + (r \pm \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}. \quad (6.216)$$

$$h_3 = \frac{\ln \frac{V_0}{\bar{v}_0} + T\sigma^2/2}{\sigma\sqrt{T}}. \quad (6.217)$$

$$h_{(4,7),8} = \frac{\ln \frac{\bar{v}_0}{V_0} \pm T\sigma^2/2}{\sigma\sqrt{T}}. \quad (6.218)$$

Theorem 6.93. *The survival probability of the firm is defined as*

$$P(0, T) = \mathcal{Q}(\tau > T) = \mathcal{Q}(\bar{\tau} = \infty, V_T \geq \bar{D}) = \mathcal{N}(h_1) - \frac{V_0}{\bar{v}_0} \mathcal{N}(h_2). \quad (6.219)$$

Proof. For every $s \in [0, T)$,

$$\begin{aligned} \{\bar{\tau} = \infty\} &= \{V_s \geq \bar{v}_s\} = \left\{ V_0 \exp \left(\left(r - \frac{\sigma^2}{2} \right) s + \sigma W_s \right) \geq K e^{-r(T-s)} \right\} \\ &= \left\{ \left(r - \frac{\sigma^2}{2} \right) s + \sigma W_s \geq \ln (K e^{-rT} / V_0) + rs \right\} \\ &= \left\{ 1/2\sigma s - W_s \leq -\frac{\ln(\bar{v}_0/V_0)}{\sigma} \right\} = \{\bar{\alpha}s - W_s \leq \bar{y}\} \\ &= \{\sup(\bar{\alpha}s - W_s) \leq \bar{y}\} = \{\sup X_s \leq \bar{y}\} = \{M_T^X \leq \bar{y}\}. \end{aligned} \quad (6.220)$$

$$\{\hat{\tau} = \infty\} = \{V_T \geq \bar{D}\} = \left\{ V_0 e^{(r - \frac{\sigma^2}{2})T + \sigma W_T} \geq \bar{D} \right\} = \left\{ (r - \frac{\sigma^2}{2})T + \sigma W_T \geq \ln \frac{\bar{D}}{V_0} \right\} \quad (6.221)$$

$$= \left\{ \frac{1}{2}\sigma T - W_T \leq -\frac{\ln \frac{\bar{D}}{V_0} - rT}{\sigma} \right\} = \{\bar{\alpha}T - W_T \leq \bar{x}\} = \{X_T \leq \bar{x}\}. \quad (6.222)$$

$$\begin{aligned} P(0, T) &= \mathcal{Q}(\tau > T) = \mathcal{Q}(\bar{\tau} = \infty, V_T \geq \bar{D}) = \mathcal{Q}(M_T^X \leq \bar{y}, X_T \leq \bar{x}) \\ &= \mathcal{N}\left(\frac{\bar{x} - \bar{\alpha}T}{\sqrt{t}}\right) - e^{2\bar{\alpha}\bar{y}} \mathcal{N}\left(\frac{\bar{x} - 2\bar{y} - \bar{\alpha}T}{\sqrt{t}}\right). \end{aligned} \quad (6.223)$$

We also need verify that $0 \leq \bar{x} \leq \bar{y}$.

$$\bar{y} = -\frac{1}{\sigma} \ln(\bar{v}_0/V_0) \geq 0. \quad (6.224)$$

$$\bar{y} - \bar{x} = -\frac{1}{\sigma} \ln(\bar{v}_0/V_0) + \frac{\ln(\bar{D}/V_0) - rT}{\sigma} = -\frac{1}{\sigma} (\ln K - \ln \bar{D}) \geq 0. \quad (6.225)$$

□

Theorem 6.94. *The price of the defaultable bond is*

$$\begin{aligned} \bar{B}(0, T) = DB(0, T) \left(\mathcal{N}(h_1) - \frac{V_0}{\bar{v}_0} \mathcal{N}(h_2) \right) + KB(0, T) \left(\mathcal{N}(h_7) + \frac{V_0}{\bar{v}_0} \mathcal{N}(h_8) \right) \\ + V_0 \left(\mathcal{N}(h_3) - \frac{\bar{v}_0}{V_0} \mathcal{N}(h_4) - \mathcal{N}(h_5) + \frac{\bar{v}_0}{V_0} \mathcal{N}(h_6) \right) \end{aligned} \quad (6.226)$$

Proof. Let J_1 be the price for no default, J_2 be the price for early default and J_3 be the price for default at T . $J = J_1 + J_2 + J_3$.

$$J_1 = \mathbb{E} [\bar{D} e^{-rT} 1_{\tau > T}] = \bar{D} e^{-rT} \mathbb{Q}(\tau > T) = \bar{D} e^{-rT} P(0, T). \quad (6.227)$$

$$\begin{aligned} J_2 &= \mathbb{E} [V_\tau e^{-r\tau} 1_{\tau < T}] = \mathbb{E} [\bar{v}_\tau e^{-r\bar{\tau}} 1_{\bar{\tau} < T}] = \mathbb{E} [K e^{-r(T-\bar{\tau})} e^{-r\bar{\tau}} 1_{\bar{\tau} < T}] \\ &= K e^{-rT} \mathbb{E} [1_{\bar{\tau} < T}] = K e^{-rT} \mathbb{Q}(\bar{\tau} < T) = K e^{-rT} (1 - \mathbb{Q}(\bar{\tau} = \infty)) \\ &= K e^{-rT} (1 - \mathbb{Q}(M_T^X \leq \bar{y})) = K e^{-rT} \left(1 - \mathcal{N} \left(\frac{\bar{y} - \bar{\alpha}T}{\sqrt{T}} \right) + e^{2\alpha\bar{y}} \mathcal{N} \left(\frac{-\bar{y} - \bar{\alpha}T}{\sqrt{T}} \right) \right). \end{aligned} \quad (6.228)$$

$$\begin{aligned} J_3 &= \mathbb{E} [V_T e^{-rT} 1_{\tau=T}] = \mathbb{E} [V_T e^{-rT} 1_{\bar{\tau}=\infty, V_T < \bar{D}}] \\ &= \mathbb{E} [V_0 e^{(r-\frac{1}{2}\sigma^2)T + \sigma W_T} e^{-rT} 1_{\bar{\tau}=\infty, V_T < \bar{D}}] = V_0 \mathbb{E} [e^{-\frac{1}{2}\sigma^2 T + \sigma W_T} 1_{\bar{\tau}=\infty, V_T < \bar{D}}] \\ &= V_0 \bar{\mathbb{Q}}(\bar{\tau} = \infty, V_T < \bar{D}) = V_0 \bar{\mathbb{Q}}(M_T^{\bar{X}} \leq \bar{y}, \bar{X}_T > \bar{x}) \\ &= V_0 \left(\mathcal{N} \left(\frac{\bar{y} - \alpha T}{\sqrt{T}} \right) - e^{2\alpha\bar{y}} \mathcal{N} \left(\frac{-\bar{y} - \alpha T}{\sqrt{T}} \right) - \mathcal{N} \left(\frac{\bar{x} - \alpha T}{\sqrt{T}} \right) + e^{2\alpha\bar{y}} \mathcal{N} \left(\frac{\bar{x} - 2\bar{y} - \alpha T}{\sqrt{T}} \right) \right) \end{aligned} \quad (6.229)$$

□

Definition 6.95. (*Zhou's model*) Let U_i be a sequence of i.i.d. random variables with finite expected value $\phi = \mathbb{Q}(U_i)$. Suppose U_i takes values on $(-1, \infty)$, we assume that N_t is a Poisson process with intensity $\lambda > 0$ under the probability measure \mathbb{Q} . The σ -fields generated by the process are independent under \mathbb{Q} . The dynamics of V under \mathbb{Q} is

$$dV_t = V_{t-} ((r - \phi\lambda)dt + \sigma_v dW_t + dJ_t) \quad (6.230)$$

where $J_t = \sum_{i=1}^{N_t} U_i$. The solution is

$$V_t = V_0 \exp \left(\sigma_V W_t + \left(r - \frac{1}{2} \sigma_V^2 - \lambda \phi \right) t \right) \sum_{i=1}^{N_t} (1 + U_i). \quad (6.231)$$

Theorem 6.96. Assume that $U_n + 1$ has the log-normal distribution under \mathcal{Q} such that $\ln(1 + U_j) \sim \mathcal{N}(\mu, \sigma^2)$. This implies that

$$\phi = \mathbb{E} [U_i] = e^{\mu + \frac{1}{2}\sigma^2} - 1. \quad (6.232)$$

The survival probability of the firm, within $[0, T]$, is given by

$$P(0, T) = \mathcal{Q}(V_T \geq \bar{D}) = 1 - \sum_{n=0}^{\infty} e^{-\lambda T} \frac{(\lambda T)^n}{n!} \mathcal{N}(-d_{2,n}). \quad (6.233)$$

The price of the defaultable bond is

$$\bar{B}(0, T) = e^{-rT} \left(\bar{D}P(0, T) + V_0 \sum_{n=0}^{\infty} e^{\mu_n(T) + \frac{1}{2}\sigma_n^2(T)} \cdot \frac{e^{-\lambda T} (\lambda T)^n}{n!} \mathcal{N}(-d_{1,n}) \right) \quad (6.234)$$

where

$$d_{2,n} = \frac{\ln(V_0/\bar{D}) + \mu_n(T)}{\sigma_n(T)}, \quad d_{1,n} = d_{2,n} + \sigma_n(T) \quad (6.235)$$

with

$$\mu_n(T) = T(r - \frac{1}{2}\sigma_V^2 - \lambda\phi) + n\mu, \quad \sigma_n^2(T) = T\sigma_V^2 + n\sigma^2. \quad (6.236)$$

Definition 6.97. The KMV model derives the implied default probability can be acquired from the stock prices.

1. Estimate the market value and volatility of the firm's assets
2. Calculate the distance to default, an index measure of default risk
3. Scale the distance of default to actual probabilities of default using an empirical default database.

Remark 6.98. Suppose we want to estimate V_0 and σ_V by S_0 and σ_S in the context of Merton's model.

$$S_0 = C^{BS}(V_0, \sigma_V, r, T, \bar{D}). \quad (6.237)$$

$$\sigma_S = \frac{1}{S_0} \mathcal{N} \left(\frac{\ln(V_0/\bar{D}) + (r + \sigma_V^2/2)T}{\sigma_V \sqrt{T}} \right) \sigma_V V_0. \quad (6.238)$$

Definition 6.99. The distance to default DD is defined by

$$DD = \frac{V_0 - \tilde{D}}{V_0 \sigma_V} \quad (6.239)$$

where the default point \tilde{D} is defined as the short-term debt plus half of the long-term debt. In KMV model,

$$DD = \frac{\ln V_0 - \ln \bar{D} + (\mu_V - \frac{1}{2}\sigma_V^2)}{\sigma_V}. \quad (6.240)$$

Definition 6.100. Expected default frequencies (EDF), default probabilities in KMV terminology, is calculated as

$$EDF = \frac{\text{number of firms with } DD = x \text{ that defaulted within one year}}{\text{number of total firms with } DD = x}. \quad (6.241)$$

SECTION 7

Fundamentals of Finance

Definition 7.1. *Finance* is the science of wealth management by individuals, companies and governments that analyzes changes in values of assets over time and relates the changes to the prices of financial instruments that promise future cash flows if certain conditions apply.

Definition 7.2. A *derivative* is a financial instrument whose value depends on or is derived from the value of the underlying asset. Derivatives are traded either on exchanges such as the Chicago Board Options Exchange or in over-the-counter (OTC) market.

Remark 7.3. Derivatives are used to hedge, speculate, or arbitrage. They change the nature of a liability or an investment without incurring the cost of selling or buying other assets.

Definition 7.4. A *forward* is an agreement to buy or sell an asset at maturity to a predetermined price (or interest). Forwards are traded in OTC.

Definition 7.5. A *futures* has the same payoff as a forward. In contrast, your broker will require that you deposit money into a margin account. As time evolves, the broker deposits money into the account when the futures price rises, and the investor deposits money into the account when the future price falls. Futures are traded on exchanges.

The asset	The contract size	Delivery arrangements
Delivery months	Price quotes	Price limits and position limits

Remark 7.6. Entering a forward or futures contract costs nothing except for a small transactions fee. Compared with forwards, futures minimize the risk that you will not get paid off in the end. However, futures require more liquidity.

Definition 7.7. An *arbitrage* is the possibility to make profits without taking risks. It is defined as a portfolio with value process (X_t) with maturity T such that $X_0 = 0$ and

1. $P(X_t = 0) = 1$ for $0 < t < T$;
2. $P(X_T < 0) = 0$;
3. $P(X_T > 0) > 0$.

Notation	Interpretation	Notation	Interpretation
c	European call price	C	American call price
p	European put price	P	American put price
π	European derivative price	Π	American derivative price
S_0	Stock price today	S_t	Stock price at time t
K	Strike	T	Maturity
μ	Return rate of stock	d	Dividend rate of stock
σ	Volatility of stock price	r	Risk-free interest rate

Definition 7.8. A call or put option gives you the right but no obligation to buy or sell an asset later. The future point of time to exercise the option is known as maturity. The exercised price is known as strike.

Definition 7.9. European-style options can only be exercised at maturity. American-style options can be exercised at any time during their life. They are vanilla options. Bermudan-style options can be exercised on specified dates on or before maturity.

Definition 7.10. The straddle option is a combination of call and put at the same strike. The straggle option is a combination of call and put at different strikes.

Definition 7.11. Asian options are options whose payoff is determined by the average underlying price. Lookback options are options whose payoff is determined by the maximum or minimum underlying price. Barrier options are options that the underlying price must pass (or not pass) a certain level before it can be exercised. There are four types of barrier options.

1. Up-and-out: starts below the barrier and has to move up for the option to become null.
2. Down-and-out: starts above the barrier and has to move down for the option to become null.
3. Up-and-in: starts below the barrier and has to move up for the option to become activated.
4. Down-and-in: starts above the barrier and has to move down for the option to become activated.

Type	Payoff	Type	Payoff
call	$(S_T - K)_+$	put	$(K - S_T)_+$
AA Asian-fixed call	$(A_T - K)_+$	AA Asian-fixed put	$(K - A_T)_+$
AA Asian-float call	$(S_T - kA_T)_+$	AA Asian-float put	$(kA_T - S_T)_+$
lookback-fixed call	$(S_{\max} - K)_+$	lookback-fixed put	$(K - S_{\min})_+$
lookback-float call	$S_T - S_{\min}$	lookback-float put	$S_{\max} - S_T$

Definition 7.12. The party that has agreed to buy or sell has the long or short position.

Definition 7.13. The intrinsic value indicates the payoff of the option if the option were exercised immediately.

Definition 7.14. Moneyness refers to the relationship between the stock price and the strike

1. At-the-money: $S \approx K$ with intrinsic values close to 0.
2. In-the-money: $S > K$ for calls and $S < K$ for puts with positive intrinsic values.
3. Out-of-the-money: $S < K$ for calls and $S > K$ for puts with negative intrinsic values.

Definition 7.15. Time value is difference between the price of the option and the intrinsic value.

Definition 7.16. Option class refers to all options of the same type (put or call) of one underlying asset. Option series refers to all options of one given option class with the same maturity.

Theorem 7.17. (Bounds for calls) $(S_0 - Ke^{-rT})_+ \leq c \leq C \leq S_0$.

Proof. By no-arbitrage argument, assume that there are two investment strategies. You can either buy an European call option and lend Ke^{-rT} to others or only buy a share of the stock. The payoff for the first strategy is $(S_T - K)_+ + K$ and the payoff for the second strategy is S_T . Since the strategy 1 pays off at least as much as strategy 2, the initial value of strategy 1 needs to be at least as high as the initial value of strategy 2. Also, the price for the call option is non-negative. Thus,

$$(S_T - K)_+ + K \geq S_T \implies C + Ke^{-rT} \geq c + Ke^{-rT} \geq S_0 \implies C \geq c \geq (S_0 - Ke^{-rT})_+. \quad (7.1)$$

The call option gives you the right to buy a stock, and thus its price cannot be higher than the price of the stock. \square

Theorem 7.18. (*Bounds for European puts*) $(Ke^{-rT} - S_0)_+ \leq p \leq Ke^{-rT}$.

Proof. By no-arbitrage argument, assume that there are two investment strategies. You can either sell an European put option and lend Ke^{-rT} to a friend or only buy a share of the stock. The payoff for strategy one is $-(K - S_T)_+ + K$ and the payoff for strategy two is S_T . Since the strategy 2 pays off at least as much as strategy 1, the initial value of strategy 2 needs to be at least as high as the initial value of strategy. Also, the price for the put option is non-negative. Thus,

$$-(K - S_T)_+ + K \leq S_T \implies -p + Ke^{-rT} \leq S_0 \implies p \geq (Ke^{-rT} - S_0)_+. \quad (7.2)$$

The European put option gives you the right to sell a stock at strike K at maturity T , and thus its price cannot be higher than the strike discounted to its present value. \square

Corollary 7.19. (*Bounds for American puts*) $(K - S_0)_+ \leq P \leq K$. The proof is similar except that there is no need to discount since the American option can be exercised immediately.

Variable	c	p	C	P
S_0	+	-	+	-
K	-	+	-	+
T	?	?	+	+
σ	+	+	+	+
r	+	-	+	-
d	-	+	-	+
μ	0	0	0	0

Theorem 7.20. (*put-call parity for European options*) $c - (S_0 - Ke^{-rT}) = p$. You can buy a European call and sell a forward on the same stock for forward price K to replicate the European put.

Notation	Interpretation	Notation	Interpretation
S_0	stock price at $t = 0$	u	upward factor for the stock
S_1	stock price at $t = 1$	d	downward factor for the stock
B_0	bond price at $t = 0$	r	interest rate for the bond
B_1	bond price at $t = 1$		

Definition 7.21. In the *binomial model*, The stock price can only goes up or down. Consider a market position that consists of x shares of the stock and y shares of the bond. Let $\omega \in \{H, T\}$. The value of this position is $V_0 = xS_0 + yB_0$ and $V_1(\omega) = xS_1(\omega) + yB_1(\omega)$.

1. $V_1(H) = 1$ and $V_1(T) = 0$.

$$x_H S_1(H) + y_H B_1(H) = 1. \quad x_H S_1(T) + y_H B_1(T) = 0. \quad (7.3)$$

The solution is

$$x_H = -\frac{B_1(T)}{B_1(H)S_1(T) - B_1(T)S_1(H)} = \frac{1}{(u-d)S_0}. \quad (7.4)$$

$$y_H = \frac{S_1(T)}{B_1(H)S_1(T) - B_1(T)S_1(H)} = -\frac{d}{(1+r)(u-d)B_0}. \quad (7.5)$$

2. $V_1(H) = 0$ and $V_1(T) = 1$.

$$x_T S_1(H) + y_T B_1(H) = 0. \quad x_T S_1(T) + y_T B_1(T) = 1. \quad (7.6)$$

The solution is

$$x_T = -\frac{B_1(H)}{B_1(T)S_1(H) - B_1(H)S_1(T)} = -\frac{1}{(u-d)S_0}. \quad (7.7)$$

$$y_T = \frac{S_1(H)}{B_1(T)S_1(H) - B_1(H)S_1(T)} = \frac{u}{(1+r)(u-d)B_0}. \quad (7.8)$$

$$\varphi_0^H = x_H S_0 + y_H B_0 = \frac{1+r-d}{(1+r)(u-d)}. \quad \varphi_0^T = x_T S_0 + y_T B_0 = \frac{u-(1+r)}{(1+r)(u-d)}. \quad (7.9)$$

$$\varphi_0^H + \varphi_0^T = \frac{1}{1+r}. \quad (7.10)$$

Remark 7.22. The solution reveals that φ_0^H and φ_0^T only depend on u , d , and r . φ_0^H and φ_0^T are *Arrow-Debreu securities*.

Definition 7.23. Let $\tilde{p} = (1+r)\varphi_0^H$ and $\tilde{q} = (1+r)\varphi_0^T$ and observe that $\tilde{p} + \tilde{q} = 1$. \tilde{p} and \tilde{q} are *risk-neutral probabilities*.

Remark 7.24.

$$\frac{S_1(H)\tilde{p} + S_1(T)\tilde{q}}{1+r} = S_0. \quad \frac{B_1(H)\tilde{p} + B_1(T)\tilde{q}}{1+r} = B_0. \quad (7.11)$$

Remark 7.25. There are several assumptions to the binomial model.

1. There exists no arbitrage.
2. Any amount of share can be traded.
3. Borrowing and lending interest rates are the same.
4. The stock price at maturity can only take one of two values.

Definition 7.26. (*Replicating Strategy under the one-period binomial pricing model*) The replicating portfolio holds a loan of W and Δ shares of the stock. Assume that $d < r < u$. We set

$$S_T^u \Delta - W e^{rT} = (S_T^u - K)_+ = S_T^u - K. \quad (7.12)$$

$$S_T^d \Delta - W e^{rT} = (S_T^d - K)_+ = 0. \quad (7.13)$$

We obtain

$$\Delta = \frac{S_T^u - K}{S_T^u - S_T^d} \quad (7.14)$$

$$W = S_T^d e^{-rT} \Delta \quad (7.15)$$

$$c = S_0 \Delta - W = S_0 \left(1 - e^{(d-r)T}\right) \frac{e^{uT} - \frac{K}{S_0}}{e^{uT} - e^{dT}}. \quad (7.16)$$

This portfolio will always pay off as much as the call option, no matter what the stock price S_T is. The cost c is also known as the delta hedge price of a European call option. It is also the risk-neutral price of a European call option.

Remark 7.27. (*Completeness*) This multiperiod binomial model is complete because every derivative on the stock can be replicated by a portfolio consisting of the savings and the stock.

Remark 7.28. The binomial model becomes the Black-Scholes model when the time interval is infinitely small.

Theorem 7.29. Define $n \in \{0, 1, \dots, N\}$ as stopping times or random spots at which an American derivative is exercised. There exists an optimal exercise time τ^* such that the only possible optimal price for an American call option is

$$\Pi = \sup_n \mathbb{E} [e^{-rn\Delta t} g(S_n)] = \mathbb{E} [e^{-r\tau^* \Delta t} g(S_{\tau^*})]. \quad (7.17)$$

Proof. Because an American option gives you the right to exercise immediately, the American option has to be priced not lower than a European option that only pays off at any time τ . Due to risk aversion, the price cannot be higher than the supremum. The supremum can be achieved or there will be no demand for the option. Thus, there exists an optimal stopping time. \square

Theorem 7.30. (*Optimal stopping time*) Let Π_n be the price of an American derivative where $n \in \{0, 1, \dots, N\}$. Define $\Pi_{N+1} = 0$ and $\Pi_N = g(S_N)$. It is optimal to exercise the derivative if the current payoff is not lower than the discounted value of the derivative from the next period. Thus, we need to find n such that

$$g(S_n) \geq \tilde{\mathbb{E}}_n [e^{r\Delta t} C_{n+1}]. \quad (7.18)$$

Note that risk-averse investors will exercise as soon as they believe that it is worth to exercise and avoid further risks. Thus,

$$\tau^* = \inf_n \left\{ n \in \{0, 1, \dots, N\} : g(S_n) \geq \tilde{\mathbb{E}}_n [e^{r\Delta t} C_{n+1}] \right\} \quad (7.19)$$

is the optimal exercise time for the American call option.

Definition 7.31. A super-replicating portfolio is a portfolio that is always valued at least as much as the value of the underlying option. Thus

$$V_n \geq C_n \geq \Pi_n. \quad (7.20)$$

Remark 7.32. A super-replicating portfolio provides a hedge for the American call option. Even when the option is not exercised optimally, the super-replicating portfolio provides enough liquidity to the option seller to pay off the option. A portfolio that holds a long position in a super-replicating portfolio and a short position in the corresponding American option has non-negative cash flows.

Remark 7.33. The price of the American derivative is equal to the initial value of the super-replicating portfolio with the smallest initial value. This means that the price of the derivative is equal to the maximum profit a risk-neutral investor expects from the option, and equal to the smallest cost of entering a hedging portfolio.

Theorem 7.34. American calls on non-dividend paying stock should never be exercised early.

Proof. If the investor exercises immediately, she will pay the strike K and receive a stock that does not pay any dividends. If she holds the stock, she may risk losing money when the stock price goes down but will enjoy the upside of a rising stock price.

If the investor does not exercise immediately, she will not have to pay anything today and will still get the upside of a rising stock price when exercising the option at a later point of time. However, the investor is not exposed to the downside of a falling stock price. Her payoff is bounded below by zero.

Because of the time value of money, the investor will prefer to pay the strike at a later point of time in order to get the same upside but no downside. \square

Remark 7.35. The above discussion does not hold for a put. The holder of an American put may decide to exercise the option early if the stock price is sufficiently small. In this case, she exercises the put right away and sells the stock at the higher strike price, receives the strike price, and invest this money in the financial market. This may be more profitable than holding the option and exercising at a later point of time.

Theorem 7.36. (Replication for Derivatives in the Multiperiod Binomial Model) Consider an N -period binomial model. Assume that $d < r < u$ and let \tilde{p} be the risk-neutral probability. The risk-neutral price is defined through backward induction for European derivatives or snell envelope for American derivatives.

$$\pi_N(\omega_N) = g(S_N). \quad (7.21)$$

$$\pi_n = e^{-r\Delta t} \tilde{\mathbb{E}}_n [\pi_{n+1}]. \quad \Pi_n = \max \left\{ g(S_n), e^{-r\Delta t} \tilde{\mathbb{E}}_n [\Pi_{n+1}] \right\}. \quad (7.22)$$

Let $v_0 = \pi_0$ and

$$\Delta_n(\omega_n) = \frac{\pi_{n+1}(\omega_n u) - \pi_{n+1}(\omega_n d)}{S_{n+1}(\omega_n u) - S_{n+1}(\omega_n d)}. \quad (7.23)$$

Define

$$v_{n+1}(\omega_{n+1}) = \Delta_n(\omega_n) S_{n+1}(\omega_{n+1}) + e^{r\Delta t} (v_n(\omega_n) - S_n(\omega_n) \Delta_n(\omega_n)). \quad (7.24)$$

Then, $v_n(\omega_n) = \pi_n(\omega_n)$ for any $n \leq N$ and $V_n(\omega_n) = \Pi_n(\omega_n)$ for any $n \leq \tau^*$.

Proof. Fix N and proceed by forward induction. The claim for $n = 0$ holds by definition. Suppose now the claim holds for $n < N$ for European derivatives and $n < \tau^*$ for American derivatives. We will show that it also holds for $n + 1$. Suppose that $\omega_{n+1} = u$ and $S_{n+1} = S_{n+1}(u)$. The induction assumption implies that $v_n = \pi_n$.

$$\begin{aligned} v_{n+1}(u) &= \frac{\pi_{n+1}(u) - \pi_{n+1}(d)}{S_{n+1}(u) - S_{n+1}(d)} S_{n+1}(u) + e^{r\Delta t} \left(v_n - \frac{\pi_{n+1}(u) - \pi_{n+1}(d)}{S_{n+1}(u) - S_{n+1}(d)} S_n \right) \\ &= (\pi_{n+1}(u) - \pi_{n+1}(d)) \frac{S_{n+1} - e^{r\Delta t} S_n}{S_{n+1}(u) - S_{n+1}(d)} + e^{r\Delta t} v_n \\ &= (\pi_{n+1}(u) - \pi_{n+1}(d))(1 - \tilde{p}) + e^{r\Delta t} \pi_n. \end{aligned} \quad (7.25)$$

1. (European derivatives)

$$\begin{aligned} v_{n+1}(u) &= (\pi_{n+1}(u) - \pi_{n+1}(d))(1 - \tilde{p}) + e^{r\Delta t} e^{-r\Delta t} (\tilde{p}\pi_{n+1}(u) + (1 - \tilde{p})\pi_{n+1}(d)) \\ &= \pi_{n+1}(u). \end{aligned} \quad (7.26)$$

Similarly, $v_{n+1}(d) = \pi_{n+1}(d)$. Thus, $v_{n+1}(\omega) = \pi_{n+1}(\omega)$ for $n < N$ and the portfolio is replicated.

2. (American derivatives)

$$\begin{aligned} V_{n+1}(u) &\geq (\Pi_{n+1}(u) - \Pi_{n+1}(d))(1 - \tilde{p}) + e^{r\Delta t} e^{-r\Delta t} (\tilde{p}\Pi_{n+1}(u) + (1 - \tilde{p})\Pi_{n+1}(d)) \\ &= \Pi_{n+1}(u). \end{aligned} \quad (7.27)$$

Similarly, $V_{n+1}(d) \geq \Pi_{n+1}(d)$. Thus, $V_{n+1}(\omega) \geq \Pi_{n+1}(\omega)$. We conclude $V_{n+1}(\omega) = \Pi_{n+1}(\omega)$ for $n < \tau^*$ since it is not optimal to exercise the option before time τ^* and the portfolio is replicated.

□

Remark 7.37. Replace r with $r - \delta$ in case of paying dividends.

Definition 7.38. A contingent claim X is a financial security which has value X_n in period n and scenario $\omega_1 \cdots \omega_n$.

Definition 7.39. An equivalent martingale measure (EMM) is a probability measure \mathbb{Q} on (Ω, \mathcal{F}) satisfying:

1. \mathbb{Q} and \mathbb{P} are equivalent. For all $A \in \mathcal{F}$,

$$\mathbb{P}[A] > 0 \iff \mathbb{Q}[A] > 0. \quad (7.28)$$

2. Any discounted portfolio process is a martingale relative to \mathbb{Q} .

$$e^{-r_0, n} = \mathbb{E}_N^{\mathbb{Q}}[e^{-r_0, n+1} V_{n+1}]. \quad (7.29)$$

Remark 7.40. The risk-neutral probability is an EMM. Potentially, there may exist more than one EMM.

Theorem 7.41. (*First Fundamental Theorem of Asset Pricing*) A market is arbitrage-free if and only if there exists an EMM.

Proof. By contradiction, suppose V is an arbitrage under an EMM. Then, by definition,

$$V_0 = 0. \quad \mathbb{P}[V_N < 0] = 0. \quad \mathbb{P}[V_N > 0] > 0. \quad (7.30)$$

Given that \mathbb{Q} and \mathbb{P} assign positive probabilities to the same events.

$$\mathbb{Q}[V_N < 0] = 0. \quad \mathbb{Q}[V_N > 0] > 0. \quad (7.31)$$

Also, $e^{-r_{0,N}} > 0$ implies

$$\mathbb{Q}[e^{-r_{0,N}} V_N < 0] = 0. \quad \mathbb{Q}[e^{-r_{0,N}} V_N > 0] > 0. \quad (7.32)$$

The martingale property under \mathbb{Q} yields $V_0 = \mathbb{E}^{\mathbb{Q}}[e^{-r_{0,N}} V_N] > 0$ which contradicts the assumption that V is an arbitrage. The other direction is trivial since we have developed the risk-neutral probability measure previously for an arbitrage-free market. \square

Definition 7.42. Suppose in period n , we have Δ_n shares of the stock and ρ_n zero-coupon bonds with maturity $n + 1$. The portfolio value at the beginning of period n is

$$V_n = \Delta_n S_n + \rho_n B_{n,n+1} = \Delta_n S_n + \rho_n e^{-r_n}. \quad (7.33)$$

The portfolio is self-financing if no money is deposited into, or withdrawn from the portfolio.

$$\Delta_{n-1} S_n + \rho_{n-1} = \Delta_n S_n + \rho_n B_{n,n+1}. \quad (7.34)$$

The portfolio is adapted if Δ_n and ρ_n depend only on the scenario until period n . Since we hold Δ_n and ρ_n constant over $[n, n + 1]$, we say that Δ and ρ are predictable over $[n, n + 1]$.

Definition 7.43. A portfolio replicates a contingent claim X when

1. The portfolio is both self-financing and adapted.
2. The portfolio value is equal to the contingent claim in every period with probability one.

The market is complete if every contingent claim can be replicated.

Theorem 7.44. (*Second Fundamental Theorem of Asset Pricing*) Under the assumption of no arbitrage, a market is complete if and only if there exists a unique EMM.

Proof. By no-arbitrage argument, there exist at least one EMM by the first FTAP. By contradiction, suppose there are two EMMs, \mathbb{Q} and $\tilde{\mathbb{Q}}$. Fix a specific scenario $\omega = (\omega_1, \dots, \omega_n)$ in Ω and consider the claim X_N with payoff X_N . Since the market is complete, X_N can be replicated by a self-financing and predictable portfolio V . In other words, $\mathbb{P}[X_N = V_N] = 1$ for a portfolio generated by adapted Δ and ρ . The martingale property for the discounted V under \mathbb{Q} or $\tilde{\mathbb{Q}}$ implies $V_0 = \mathbb{E}^{\mathbb{Q}}[e^{-r_{0,N}} V_N]$.

$$V_0 = \mathbb{E}^{\mathbb{Q}}[e^{-r_{0,N}} V_N] = \mathbb{E}^{\mathbb{Q}}[e^{-r_{0,N}} X_N] = \mathbb{E}^{\mathbb{Q}}[1_\omega] = \mathbb{Q}[\omega]. \quad (7.35)$$

Thus, $\mathbb{Q} = \tilde{\mathbb{Q}}$. For the other direction. Suppose there exists a unique EMM \mathbb{P} . Take a contingent claim X . The martingale property implies that $X_n = \mathbb{E}[e^{-r_n} X_{n+1}]$. For such a price process, we know that the portfolio with $V_0 = X_0$ and

$$\Delta_n = \frac{X_{n+1}(u) - X_{n+1}(d)}{S_{n+1}(u) - S_{n+1}(d)}, \quad \rho_n = \frac{X_n - \Delta_n S_n}{B_n}, \quad (7.36)$$

for $n = 0, 1, \dots, N - 1$ is by construction an adapted and self-financing replicating portfolio for the contingent claim. Thus, the contingent claim X is redundant. This holds for any contingent claim so that the market is complete. \square

Definition 7.45. (Preference relation) Let Δ be a portfolio and use \succeq to rank portfolios.

1. (Completeness) Either $\Delta_1 \succeq \Delta_2$ or $\Delta_2 \succeq \Delta_1$.
2. (Transitivity) If $\Delta_1 \succeq \Delta_2$ and $\Delta_2 \succeq \Delta_3$, then $\Delta_1 \succeq \Delta_3$.
3. (Monotonicity) If $W^{\Delta_1} \geq W^{\Delta_2}$ for all ω , then $\Delta_1 \succeq \Delta_2$.
4. (Convexity / Diversification) Let $\Delta_3 = \alpha\Delta_1 + (1 - \alpha)\Delta_2$ for $\alpha \in (0, 1)$. If $\Delta_1 \succeq \Delta$ and $\Delta_2 \succeq \Delta$, then $\Delta_3 \succeq \Delta$.

Definition 7.46. Let $U(x)$ be a twice continuously differentiable utility function defined on the interval $I \subseteq \mathbb{R}$.

1. $U(x)$ is increasing since the investor prefers more wealth.
2. $U(x)$ is concave due to the law of diminishing return and implies that risk aversion.

The Arrow-Pratt coefficient of absolute risk aversion is

$$A(x) = -\frac{\partial^2 U(x)}{\partial U(x)}. \quad (7.37)$$

The Arrow-Pratt coefficient of relative risk aversion is

$$R(x) = A(x)x = -\frac{\partial^2 U(x)}{\partial U(x)}x. \quad (7.38)$$

Type	$U(\omega)$	$R(\omega)$	$A(\omega)$	Comments
Exponential	$-e^{a\omega}$	a	$a\omega$	$a > 0$ and $\omega \in \mathbb{R}$
Logarithmic	$\ln(\omega)$	$\frac{1}{\omega}$	1	$\omega > 0$
Power	$\frac{\omega^{1-b}}{1-b}$	$\frac{b}{\omega}$	b	$b > 0, b \neq 1$ and $\omega > 0$
Quadratic	$\omega - c\omega^2$	$\frac{1}{2c - \omega}$	$\frac{\omega}{2c - \omega}$	$c > 0$ and $\omega < \frac{1}{2c}$

Definition 7.47. The certainty equivalent is the guaranteed amount of cash that would yield the same exact expected utility as a given risky asset with absolute certainty.

$$CE = U^{-1}(\mathbb{E}[U(x)]). \quad (7.39)$$

The risk premium is the difference between the certainty equivalent and the expected return.

Definition 7.48. A twice continuously differentiable utility function $U(x)$ defined on some interval $I \subseteq \mathbb{R}$ has constant absolute risk aversion (CARA) if

$$A(x) = -\frac{\partial^2 U(x)}{\partial U(x)} = k > 0. \quad (7.40)$$

$$\begin{aligned} A(x) = -\frac{\partial^2 U(x)}{\partial U(x)} = k > 0 &\implies \partial^2 U(x) + k\partial U(x) = 0 \implies \partial(e^{kx}\partial U(x)) = 0 \\ &\implies \partial U(x) = c_0 \cdot e^{-kx} \implies U(x) = c_1 - \frac{c_0}{k}e^{-kx}. \end{aligned} \quad (7.41)$$

Definition 7.49. A twice continuously differentiable utility function $U(x)$ defined on some interval $I \subseteq \mathbb{R}$ has hyperbolic absolute risk aversion (HARA) if

$$A(x) = -\frac{\partial^2 U(x)}{\partial U(x)} = \frac{1-\gamma}{x}. \quad (7.42)$$

$$\begin{aligned} A(x) = -\frac{\partial^2 U(x)}{\partial U(x)} = \frac{1-\gamma}{x} &\implies \partial^2 U(x) \cdot x + (1-\gamma)\partial U(x) = 0 \\ &\implies \partial(x^{1-\gamma}\partial U(x)) = 0 \implies \partial U(x) = c_0 \cdot x^{\gamma-1} \\ &\implies U(x) = \begin{cases} c_0 \log(x) + c_1 & \gamma = 0 \\ \frac{c_0}{\gamma} x^\gamma + c_1 & \gamma \in]0, 1[\end{cases} \end{aligned} \quad (7.43)$$

Definition 7.50. A twice continuously differentiable utility function $U(x)$ defined on some interval $I \subseteq \mathbb{R}$ has constant relative risk aversion (CRRA) if

$$R(x) = -\frac{\partial^2 U(x)}{\partial U(x)} x = k > 0. \quad (7.44)$$

$$\begin{aligned} R(x) = -\frac{\partial^2 U(x)}{\partial U(x)} \cdot x = k > 0 &\implies \partial^2 U(x) \cdot x + k\partial U(x) = 0 \\ &\implies \partial(x^k \partial U(x)) = 0 \implies \partial U(x) = c_0 \cdot x^{-k} \\ &\implies U(x) = \begin{cases} c_0 \log(x) + c_1 & k = 1 \\ \frac{c_0}{1-k} x^{1-k} + c_1 & k \in]0, 1[\cup]1, \infty[\end{cases} \end{aligned} \quad (7.45)$$

Theorem 7.51. (State Price Densities) The process $(\Psi_t)_{t \in \mathbb{N}_T}$ is \mathcal{F} -adapted and strictly positive. Let $S_T = 0$ since no security is traded in the final period. For any traded security S ,

$$\Psi_t S_t = \mathbb{E}[\Psi_{t+1}(S_{t+1} + D_{t+1})|F_t]. \quad (7.46)$$

Corollary 7.52.

$$S_t = \sum_{s=t+1}^T \mathbb{E}\left[\frac{\Psi_s}{\Psi_t} D_s | F_t\right]. \quad (7.47)$$

Proof.

$$\begin{aligned} S_t &= \mathbb{E}\left[\frac{\Psi_{t+1}}{\Psi_t}(S_{t+1} + D_{t+1})|F_t\right] = \mathbb{E}\left[\frac{\Psi_{t+1}}{\Psi_t} S_{t+1} | F_t\right] + \mathbb{E}\left[\frac{\Psi_{t+1}}{\Psi_t} D_{t+1} | F_t\right] \\ &= \mathbb{E}\left[\frac{\Psi_{t+2}}{\Psi_{t+1}} \frac{\Psi_{t+1}}{\Psi_t}(S_{t+2} + D_{t+2})|F_t\right] + \mathbb{E}\left[\frac{\Psi_{t+1}}{\Psi_t} D_{t+1} | F_t\right] \\ &= \mathbb{E}\left[\frac{\Psi_{t+2}}{\Psi_t} S_{t+2} | F_t\right] + \mathbb{E}\left[\frac{\Psi_{t+2}}{\Psi_t} D_{t+2} | F_t\right] + \mathbb{E}\left[\frac{\Psi_{t+1}}{\Psi_t} D_{t+1} | F_t\right] \\ &= \dots = \mathbb{E}\left[\frac{\Psi_T}{\Psi_t} S_T | F_t\right] + \sum_{s=t+1}^T \mathbb{E}\left[\frac{\Psi_s}{\Psi_t} D_s | F_t\right] = \sum_{s=t+1}^T \mathbb{E}\left[\frac{\Psi_s}{\Psi_t} D_s | F_t\right]. \end{aligned} \quad (7.48)$$

□

Definition 7.53. (*Stochastic Discount Factors*) The process $(\xi_{t+1})_{t \in \mathbb{N}_{|T-1}}$ is agent-specific.

$$\xi_{t+1} = \frac{\Psi_{t+1}}{\Psi_t}. \quad (7.49)$$

Corollary 7.54.

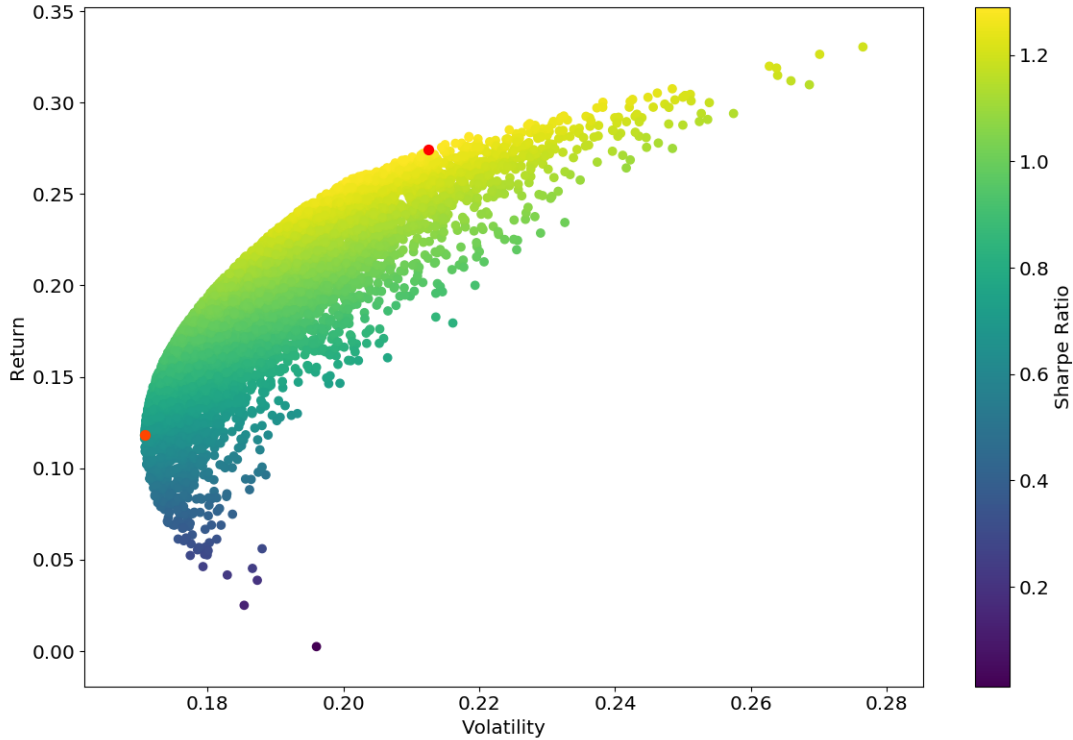
$$S_t = \sum_{s=t+1}^T \mathbb{E} \left[\left(\prod_{u=t+1}^s \xi_u \right) D_s | F_t \right]. \quad (7.50)$$

Theorem 7.55. The investor allocates the initial wealth W_0 among a risky security with the initial price S_0 and a zero-coupon bond with constant return rate r . The optimal share position satisfies

$$\Delta^* \approx \frac{1}{A(W_0)} \cdot \frac{\mathbb{E}[(S_1 - S_0 e^r)]}{\mathbb{E}[(S_1 - S_0 e^r)^2]}. \quad (7.51)$$

There are trade-offs between return, risk, and risk aversion.

1. Higher expected excess return leads to more risky assets.
2. Higher return variance leads to less risky assets.
3. Higher risk aversion leads to less risky assets.



The Efficient Frontier Plot

Definition 7.56. Suppose there are n risky securities. Let $i, j \in \mathbb{N}$ with $1 \leq i, j \leq n$. For the i^{th} security, its expected return is \bar{r}_i and its volatility is σ_i , and its correlation with the j^{th} security is ρ_{ij} . For a portfolio $\Delta = (\Delta_1, \Delta_2, \dots, \Delta_N)$ of fractions of wealth, its expected return is

$$\bar{r}(\Delta) = \sum_{i=1}^n \Delta_i \bar{r}_i. \quad (7.52)$$

and its return variance is

$$\sigma^2(\Delta) = \sum_{i=1}^n \sum_{j=1}^n \Delta_i \Delta_j \sigma_i \sigma_j \rho_{ij}. \quad (7.53)$$

Definition 7.57. The efficient frontier² refers to the map $r \mapsto \sigma(\Delta(r))$ where $\Delta(r)$ is the optimal portfolio to minimize the volatility given r .

Theorem 7.58. The Markowitz portfolio allocation problem without the risk-free asset is to minimize variance portfolio given a target expected return \bar{r} . With matrix notations,

$$\min_{\Delta} \Delta^T \Sigma \Delta \quad \text{s.t.} \quad \Delta^T \vec{r} = \bar{r} \quad \text{and} \quad \Delta^T \vec{1} = 1. \quad (7.54)$$

Σ covariance matrix \vec{r} vector of expected returns
 Δ vector of portfolio weights $\vec{1}$ vector of ones

Let λ_1 and λ_2 be two Lagrange multipliers and define

$$\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \quad \text{and} \quad (g\Delta - c)^T = \Delta^T g^T - c^T = \Delta^T (\vec{r} \quad \vec{1}) - (\bar{r} \quad 1). \quad (7.55)$$

Then, the problem can be transformed to unconstrained one

$$\min_{\Delta, \lambda} \left(\Delta^T \Sigma \Delta - (g\Delta - c)^T \lambda \right). \quad (7.56)$$

The optimal solution (Δ^*, λ^*) satisfies the Karush-Kuhn-Tucker (KKT) conditions

$$2\Sigma\Delta^* - g^T\lambda = 0 \quad \text{and} \quad g\Delta^* = c. \quad (7.57)$$

Thus,

$$\Delta^* = \Sigma^{-1}g^T (g\Sigma^{-1}g^T)^{-1}c. \quad (7.58)$$

Theorem 7.59. (Two-Fund Theorem) The maximum sharpe ratio (MSR) Δ_{MSR} gives the highest expected return per unit volatility.

$$\max_{\Delta} \frac{\Delta^T \vec{r}}{\sqrt{\Delta^T \Sigma \Delta}} \quad \text{s.t.} \quad \Delta^T \vec{1} = 1 \quad \implies \quad \Delta_{MSR} = \frac{\Sigma^{-1}\vec{r}}{\vec{1}^T \Sigma^{-1}\vec{r}}. \quad (7.59)$$

The global minimum variance (GMV) Δ_{GMV} gives the lowest possible variance. Any rational investor would invest in portfolio that lies on the efficient frontier and above this point.

$$\min_{\Delta} \Delta^T \Sigma \Delta \quad \text{s.t.} \quad \Delta^T \vec{1} = 1 \quad \implies \quad \Delta_{GMV} = \frac{\Sigma^{-1}\vec{1}}{\vec{1}^T \Sigma^{-1}\vec{1}}. \quad (7.60)$$

²Given the stock returns of Apple, Cisco, IBM, and Amazon, we simulate 8000 portfolios. The red dot represents the (simulated) MSR and the orange dot represents the (simulated) GMV.

Given that $\Delta_{MSR}^T \vec{1} = 1$, $\Delta_{GMV}^T \vec{1} = 1$ and $(\Delta^*)^T \vec{1} = 1$, there exists α^* such that

$$\Delta^* = \alpha^* \Delta_{MSR} + (1 - \alpha^*) \Delta_{GMV}. \quad (7.61)$$

To achieve the target return,

$$\bar{r} = \alpha^* \bar{r}_{MSR} + (1 - \alpha^*) \bar{r}_{GMV} \implies \alpha^* = \frac{\bar{r} - \bar{r}_{GMV}}{\bar{r}_{MSR} - \bar{r}_{GMV}}. \quad (7.62)$$

Thus, instead of choosing Δ , the investor can choose $(\alpha^*, 1 - \alpha^*)$ to the core portfolios Δ_{MSR} and Δ_{GMV} to achieve the target return \bar{r} .

Theorem 7.60. Let $\vec{r}_e = \vec{r} - r_f \vec{1}$ be the vector of excess returns and $\bar{r}_e = \bar{r} - \bar{r}_f$ be the expected excess return. The Markowitz portfolio allocation problem with the risk-free asset is to minimize variance portfolio of risky assets given an expected excess return \bar{r}_e . Since the return of a risk-free asset r_f is constant, its variance and covariances with other risky assets are 0. In other words, investing in risk-free assets changes the return but does not add variance. With matrix notations,

$$\min_{\Delta} \Delta^T \Sigma \Delta \quad \text{s.t.} \quad \Delta^T \vec{r}_e = \bar{r}_e. \quad (7.63)$$

Let λ be the Lagrange multiplier. The problem can be transformed to the unconstrained one

$$\min_{\Delta, \lambda} \Delta^T \Sigma \Delta - \lambda (\Delta^T \vec{r}_e - \bar{r}_e). \quad (7.64)$$

The optimal solution (Δ^*, λ^*) satisfies

$$2\Sigma\Delta - \lambda \vec{r}_e = 0. \quad (7.65)$$

$$\vec{r}_e^T \Delta = \bar{r}_e. \quad (7.66)$$

Thus,

$$\Delta^* = \frac{\bar{r}_e \Sigma^{-1} \vec{r}_e}{\vec{r}_e^T \Sigma^{-1} \vec{r}_e}. \quad (7.67)$$

Theorem 7.61. (One-Fund Theorem) With the risk-free asset, the maximum sharpe ratio in the excess return space $\Delta_{MSR,e}$ gives the highest expected excess return per unit volatility.

$$\max_{\Delta} \frac{\Delta^T \vec{r}_e}{\sqrt{\Delta^T \Sigma \Delta}} \quad \text{s.t.} \quad \Delta^T \vec{1} = 1 \implies \Delta_{MSR,e} = \frac{\Sigma^{-1} \vec{r}_e}{\vec{1}^T \Sigma^{-1} \vec{r}_e}. \quad (7.68)$$

The risky asset position is

$$\Delta^* = \alpha^* \Delta_{MSR,e} \quad (7.69)$$

where

$$\alpha^* = \bar{r}_e \cdot \frac{\vec{1}^T \Sigma^{-1} \vec{r}_e}{\vec{r}_e^T \Sigma^{-1} \vec{r}_e}. \quad (7.70)$$

and the risk-free asset position is $1 - \alpha^*$.

Definition 7.62. The capital market line is

$$r = r_f + \frac{\bar{r}^* - r_f}{\sigma^*} \sigma \quad (7.71)$$

where $\bar{r}^* = \bar{r}_{MSR,e}$ and $\sigma^* = \sigma_{MSR,e}$ are the return and the volatility of the market portfolio.

Theorem 7.63. (CAPM pricing formula) *The market portfolio solves*

$$\max_{\Delta} \frac{\Delta^T (\bar{r} - r_f \mathbf{1})}{\sqrt{\Delta^T \Sigma \Delta}} \quad \text{s.t.} \quad \Delta^T \mathbf{1} = 1. \quad (7.72)$$

The optimal solution Δ^ satisfies*

$$\bar{r} - r_f \mathbf{1} = \frac{(\Delta^*)^T (\bar{r} - r_f \mathbf{1})}{(\Delta^*)^T \Sigma \Delta^*} \cdot \Sigma \Delta^* = \frac{\bar{r}^* - r_f}{(\sigma^*)^2} \Sigma \Delta^*. \quad (7.73)$$

Let r_m be the return of the m^{th} risky asset and r^ be the return of the market portfolio. By definition,*

$$\text{Var}(r^*) = (\sigma^*)^2 \quad \text{and} \quad \text{Cov}(r_m, r^*) = \sum_{k=1}^M \Delta_k^* \text{Cov}(r_m, r_k) = (\Sigma \Delta^*)_m \quad (7.74)$$

Thus,

$$\bar{r}_m - r_f = \frac{\bar{r}^* - r_f}{(\sigma^*)^2} (\Sigma \Delta^*)_m = \frac{\text{Cov}(r_m, r^*)}{\text{Var}(r^*)} (\bar{r}^* - r_f) = \beta_m (\bar{r}^* - r_f). \quad (7.75)$$

where beta measures the level of dependence between a risky asset and the market portfolio.

Theorem 7.64. *Any efficient portfolio is composed by a fraction α invested in the market portfolio and a fraction $1 - \alpha$ invested in the risk-free asset. Denote by r the return of this portfolio.*

$$r = \alpha r^* + (1 - \alpha) r_f. \quad (7.76)$$

The beta of an efficient portfolio is

$$\beta = \frac{\text{Cov}(r, r^*)}{\text{Var}(r^*)} = \frac{\alpha \text{Cov}(r^*, r^*)}{\text{Var}(r^*)} = \alpha. \quad (7.77)$$

Definition 7.65. (Regression Analysis) *Given CAPM, the m -th risky asset has expected return*

$$\bar{r}_m = r_f + \beta_m (\bar{r}^* - r_f) + \epsilon_m \quad (7.78)$$

with $\mathbb{E}[\epsilon_m] = 0$ and $\text{Cov}(\epsilon_m, r^) = 0$. Let $\text{Var}(\epsilon_m) = \sigma_{\epsilon_m}^2$.*

$$\sigma_m^2 = \text{Var}(r_m) = \text{Var}(r_f + \beta_m (\bar{r}^* - r_f) + \epsilon_m) = \beta_m^2 (\sigma^*)^2 + \sigma_{\epsilon_m}^2. \quad (7.79)$$

The first component, known as systematic risk, captures all variance arising from the fluctuations of the market return. The second component, known as idiosyncratic risk, captures all variance that is orthogonal to the market arising from the fluctuations of the specific security and it can be diversified away. The CAPM price is

$$S_0^{(m)} = \mathbb{E} \left[\frac{S_1^{(m)}}{1 + \bar{r}_m} \right]. \quad (7.80)$$

Remark 7.66. *The risk-neutral price compensates for risks by adjusting the probability distribution of outcomes. Unfavorable outcomes become more likely as the security becomes more risky. The CAPM price compensates for risks by adjusting the interest rates. Though being conceptually different, both approaches lead to the same price in complete and arbitrage-free markets.*

Definition 7.67. The Sharpe ratio of the portfolio Δ that measures the excess return per unit of volatility is defined as

$$s(\Delta) = \frac{\bar{r}(\Delta) - r_f}{\sigma(\Delta)} \leq s(\Delta_{MSR,e}). \quad (7.81)$$

Definition 7.68. A fixed income security is a financial claim which provides a return in the form of fixed periodic payments and an eventual return of principal at maturity.

Definition 7.69. The spot rate or the short rate, denoted as r_t is the rate paid on a short-term loan that measures in the next period. The interest rate, denoted as $r_{t,T}$, is the rate paid on a loan initiated at time t and matures at time T . The map $T \rightarrow r_{t,T}$ is the term structure of interest rates.

Definition 7.70. A caplet is a call option on a future interest rate. A cap is a combination of multiple caplets for different maturities. A floorlet is a put option on a future interest rate. A floor is a combination of multiple floorlets for different maturities. Suppose there are N periods and let K be the strike.

$$\text{Cap}_{0,N} = \sum_{n=1}^N \text{Caplet}_{0,n} = \sum_{n=1}^N \mathbb{E}[e^{-r_{0,n}}(e^{r_{n-1}} - K)_+] \quad (7.82)$$

$$\text{Floor}_{0,N} = \sum_{n=1}^N \text{Floorlet}_{0,n} = \sum_{n=1}^N \mathbb{E}[e^{-r_{0,n}}(K - e^{r_{n-1}})_+] \quad (7.83)$$

Definition 7.71. A swap is a contract in which one side agrees to pay a fixed interest rate and the other side pays a variable interest rate at predetermined times. Suppose there are N periods. The fixed interest rate, denoted as s_N , is the swap rate.

Theorem 7.72. (Determining Swap Rates) The risk-neutral pricing formula indicates

$$0 = \sum_{n=1}^N \mathbb{E}[e^{-r_{0,n}}(e^{s_N} - e^{r_{n-1}})] = e^{s_N} \sum_{n=1}^N \mathbb{E}[e^{-r_{0,n}}] - \sum_{n=1}^N \mathbb{E}[e^{-r_{0,n-1}}].$$

$$\iff s_N = \ln \left(\frac{\sum_{n=1}^N B_{0,n-1}}{\sum_{n=1}^N B_{0,n}} \right). \quad (7.84)$$

Remark 7.73. Let K be the strike and s_N be the swap rate. Suppose $K = e^{s_N}$. Then, the value of the swap is given by

$$\text{Swap}_{0,N} = \text{Cap}_{0,N} - \text{Floor}_{0,N}. \quad (7.85)$$

Definition 7.74. A zero-coupon bond (ZCB) is a security that pays at maturity. It can be understood as a bond. A coupon bond is a security that pays at maturity and at certain times before maturity.

Theorem 7.75. (The Law of One Price) Two traded securities or portfolio of traded securities which promise the same sequence of future cash flows should have the same price at any point of time.

Definition 7.76. (Pricing Bonds under Risk-Neutral Probability) Suppose there is a zero-coupon bond that pays off one dollar at maturity T . Let $B_{t,T}$ be the price of the zero-coupon bond at time t . The risk-neutral pricing formula states

$$B_{t,T} = \mathbb{E}[e^{-r_{t,T}}] = e^{-(T-t)r_{t,T}}. \quad (7.86)$$

No arbitrage implies that $0 < B_{t,T} \leq 1$ and $B_{T,T} = 1$. It also implies that the coupon bond is equivalent to the portfolio of zero-coupon bonds. Thus, the (actual or dirty) price of a coupon bond with coupon rate c is

$$P_t = c \sum_{i=t+1}^T B_{t,i} + B_{t,T}. \quad (7.87)$$

Definition 7.77. The fixed interest rate y_t for your cash flows to grow to the bond price is the yield.

$$P_t = c \sum_{i=t+1}^T e^{-(i-t)y_t} + e^{-(T-t)y_t}. \quad (7.88)$$

The mapping $t \rightarrow y_t$ is the yield curve or term structure of yields.

Remark 7.78. The yield curve tells us how we expect a dollar invested in the US government to grow through time. Long-term investments normally promise a higher yield. Investors require high compensation to put aside liquidity for a long time. An increasing yield curve is normal. Sometimes, the yield curve is decreasing. This may occur during recessions when low interest rates on long-term investments are used to acquire cash, to jump start the economy. Inverted.

Definition 7.79. The accrued interest at time $t \in (T_{i-1}, T_i]$ is

$$AI_t = c_i \frac{t - T_{i-1}}{T_i - T_{i-1}}. \quad (7.89)$$

The quoted or clean price is

$$P'_t = P_t - AI_t. \quad (7.90)$$

Definition 7.80. Let δ be the time interval. The par yield is the coupon rate c_t which produces a par value given the term structure of interest rates such that

$$1 = \delta c_t \sum_{i=t+1}^T B_{t,i} + B_{t,T} \iff c_t = \frac{1 - B_{t,T}}{\delta \cdot \sum_{i=t+1}^T B_{t,i}}. \quad (7.91)$$

Definition 7.81. The (Macaulay) duration of a coupon bond is

$$D_t = \frac{c \sum_{i=t+1}^T i \cdot B_{t,i} + T \cdot B_{t,T}}{P_t}. \quad (7.92)$$

The convexity of a coupon bond is

$$C_t = \frac{c \sum_{i=t+1}^T i(i+1)B_{t,i} + T(T+1)B_{t,T}}{P_t}. \quad (7.93)$$

Definition 7.82. A callable or puttable bond is the bond that includes a call or put provision to allow the issuer to redeem the bond for a predetermined price. It can be seen as a portfolio of longing a bond and shorting a call or put option. A convertible bond allows its holder to convert the bond into a predetermined number of shares of the issuer's stocks. It can be seen as a portfolio of a bond and an American exchange option.

Theorem 7.83. (Bootstrap) Let P be the vector of observed prices and B be the vector of unknown discount factors. Let C be the matrix of cash flows. Then,

$$P = CB \implies B = C^{-1}P \quad (7.94)$$

Note that C is invertible only if it is a square matrix. In practice, there could be way more unknown factors than market prices and that would result in imperfect fits.

Definition 7.84. A cubic spline approximation of the discount function is defined by

$$\begin{aligned} \hat{B}_t &= \sum_{i=0}^{k-1} 1_{\{\tau_i \leq t\}} \left(\alpha_i + \beta_i (t - \tau_i) + \gamma_i (t - \tau_i)^2 + \delta_i (t - \tau_i)^3 \right) \\ &= \alpha_0 + \beta_0 t + \gamma_0 t^2 + \sum_{i=0}^{k-1} \delta_i (t - \tau_i)^3. \end{aligned} \quad (7.95)$$

Theorem 7.85. (Pricing Forwards under Risk-Neutral Probability) Suppose S is the price process of a fixed income product which satisfies the discounted martingale property. Denote the forward price at period n as $\text{For}_{n,N}$. Since entering a forward at period n contract costs nothing, buyers and sellers will determine a rational forward price for the asset to be zero. The risk-neutral pricing formula indicates

$$0 = \mathbb{E}[e^{-r_{n,N}}(S_N - \text{For}_{n,N})] \iff \text{For}_{n,N} = \frac{\mathbb{E}[e^{-r_{n,N}} S_N]}{\mathbb{E}[e^{-r_{n,N}}]} = \frac{S_n}{B_{n,N}}. \quad (7.96)$$

Remark 7.86. The forward price is the current asset price scaled up with the expected growth rate of a risk-free investment.

Definition 7.87. Let $S_n = 1$. Denote the forward rate from period n to period N as $f_{n,N}$ such that $\text{For}_{n,N} = e^{f_{n,N}}$. Then,

$$f_{n,N} = \ln \left(\frac{S_n}{B_{n,N}} \right) = -\ln(B_{n,N}). \quad (7.97)$$

Theorem 7.88. (Pricing Futures under Risk-Neutral Probability) Suppose S is the price process of a fixed income product which satisfies the discounted martingale property. Denote the futures price at period n as $\text{Fut}_{n,N}$. Since entering a futures at period n contract costs nothing, buyers and sellers will determine a rational futures price for the asset. Futures are settled daily through the margin account, so gains and losses are realized during the lifetime of the futures. $\text{Fut}_{N,N} = S_N$ by no arbitrage. The risk-neutral pricing formula indicates

$$0 = \mathbb{E}_n \left[\sum_{k=n}^{N-1} e^{-r_{n,k+1}} (\text{Fut}_{k+1,N} - \text{Fut}_{k,N}) \right] \iff \text{Fut}_{n,N} = \mathbb{E}[S_N]. \quad (7.98)$$

Proof. Clearly, $\text{Fut}_{N,N} = \mathbb{E}[S_N] = S_N$. Also, for $k = n, \dots, N-1$,

$$\begin{aligned} \mathbb{E}_n[e^{-r_{n,k+1}} \text{Fut}_{k+1,N}] &= \mathbb{E}_n[e^{-r_{n,k+1}} \mathbb{E}_{k+1}[S_N]] = \mathbb{E}_n[\mathbb{E}_{k+1}[e^{-r_{n,k+1}} S_N]] \\ &= \mathbb{E}_n[\mathbb{E}_k[\mathbb{E}_{k+1}[e^{-r_{n,k+1}} S_N]]] = \mathbb{E}_n[\mathbb{E}_k[e^{-r_{n,k+1}} S_N]] \\ &= \mathbb{E}_n[e^{-r_{n,k+1}} \mathbb{E}_k[S_N]] = \mathbb{E}_n[e^{-r_{n,k+1}} \text{Fut}_{k,N}]. \end{aligned} \quad (7.99)$$

□

Corollary 7.89. *To compare the forward price with the futures one, note that*

$$\begin{aligned} \text{For}_{n,N} &= \frac{\mathbb{E}[e^{-r_{n,N}} S_N]}{\mathbb{E}[e^{-r_{n,N}}]} = \frac{\text{Cov}_n(e^{-r_{n,N}}, S_N) + \mathbb{E}[e^{-r_{n,N}}] \mathbb{E}[S_N]}{\mathbb{E}[e^{-r_{n,N}}]} \\ &= \frac{\text{Cov}_n(e^{-r_{n,N}}, S_N)}{\mathbb{E}[e^{-r_{n,N}}]} + \text{Fut}_{n,N}. \end{aligned} \quad (7.100)$$

Thus, $\text{For}_{n,N} - \text{Fut}_{n,N}$ and $\text{Cov}_n(e^{-r_{n,N}}, S_N)$ have the same sign.

Remark 7.90. *We derived the futures price under the assumptions that margin account balances do not accumulate interests and a futures investor has infinite liquidity to cover any margin calls while those are not the cases in reality. Most brokers pay interests on margin account balances and most investors have to borrow money to cover margin calls, and thus the futures rate will generally be higher than the forward rate.*

Definition 7.91. (Principal Component Analysis) *Let X be the matrix of the empirical data which are t observations of k market variables. Note that the market variables could be standardized or normalized or differenced in order to be stationary.*

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{t1} & x_{t2} & \dots & x_{tk} \end{pmatrix}$$

Let V be the sample correlation matrix.

$$V = \frac{X^T X}{t} = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1k} \\ \rho_{21} & 1 & \dots & \rho_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k1} & \rho_{k2} & \dots & 1 \end{pmatrix} \quad (7.101)$$

Let W be the eigenvector matrix and Λ be the eigenvalue matrix. Note that W is an orthogonal matrix and Λ is a diagonal matrix. The diagonalization is

$$\Lambda = W^T V W \quad (7.102)$$

where

$$\Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_k \end{pmatrix} \quad \text{and} \quad W = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1k} \\ w_{21} & w_{22} & \dots & w_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ w_{k1} & w_{k2} & \dots & w_{kk} \end{pmatrix} = [\vec{w}_1 \quad \vec{w}_2 \quad \dots \quad \vec{w}_k].$$

Every eigenvector w_i is a principal component (PC) or an uncorrelated risk factor. The first PC is referred as level. The second PC is referred as twist. The third PC is referred as curvature. The sum of k eigenvalues $\sum_{i=1}^k \lambda_i$ is equal to the total variance of the original k variables. The factor score P_i is the amount of a particular PC which was observed to be present in the change of the market variables on a given day.

$$P_i = X \cdot w_i. \quad (7.103)$$

Definition 7.92. A credit derivative is a derivative security that is primarily used to transfer, hedge or manage credit risk and whose payoff is materially affected by credit risk and is conditioned on the occurrence of a credit event.

- A reference credit or reference entity is an entity that issues debts and whose default triggers the credit event.
- A reference obligation or reference credit asset is an underlying asset, issued by the reference credit, used in credit derivatives.
- A credit event triggers the default payment by the protection seller to the protection buyer.
- A default payment is the payment which has to be made if a credit event has happened.
- A recovery rate is the value of the defaulted asset expressed in percentage of its nominal value.

Definition 7.93. An asset swap package is a combination of a defaultable bond with an interest rate swap contract that swaps the coupon of the bond into a payoff stream of Libor plus a spread (chosen so that the value of package is equal to the par value of the defaultable bond).

Definition 7.94. A default digital swap (DDS) is a credit derivative whose default payment is a predetermined fixed amount.

Definition 7.95. A total return swap (TRS) is a credit derivative where the total return receiver receives all the cash flows associated with a given reference asset without actual buying or owning.

Definition 7.96. A credit default swap (CDS) is a contractual agreement to transfer the default risk of the reference credit from the protection buyer to the protection seller.

Definition 7.97. A basket default swap (BDS) is a credit derivative written on a portfolio issued by more than one reference entities, usually between three and ten.

Definition 7.98. A portfolio default swap (PDS) transfers portions of the credit risk associated with a portfolio from a protection buyer to a protection seller.

Definition 7.99. A collateralized debt obligation (CDO) is a security with different tranches of seniority and with interest and principal payments that are backed by the cash flows of an underlying portfolio of debt instruments.

- Equity tranches (the most subordinated tranches) receive interest and principal payments only if all other investors received their promised payments.
- Mezzanines and senior tranches receive interest and principal payments by seniority.

Definition 7.100. (Steps of Issuing CDO) Suppose we have a commercial bank with a loan portfolio. The bank transfers the credit risk associated with the portfolio and keeps the loans on its balanced sheet. A special purpose vehicle (SPV) is a business entity or trust formed specifically to issue that CDO.

1. The risk transfer is made via a PDS where the SPV is the counterparty.
2. The bank makes periodic premium payments to SPV and the SPV promises to stand ready to step in to cover any default-related losses.

3. The SPV issues notes to various classes of investors where each class corresponds to claims with a given level of seniority.
4. The SPV invests the proceeds of the notes in high-grade instruments.
5. The SPV uses these assets both as collateral for its obligation toward the sponsoring bank and the investors.

Definition 7.101. Define τ as the default time of the reference asset. The survival probability over $(0, T]$ is denoted by $P(0, T) = \mathbb{Q}(\tau > T)$. The default probability is the complement. $D(0, T) = 1 - P(0, T)$. Defaultable means that $\mathbb{Q}(\tau \leq T)$ and $\mathbb{Q}(\tau > T)$ are strictly positive for any $T > 0$. Note that $\mathbb{Q}(\tau = 0) = 0$. Similarly, the conditional survival probability at time $t > 0$ is

$$P(t, T) = \mathbb{Q}(\tau > T | \mathcal{F}_t \wedge \{\tau > t\}). \quad (7.104)$$

$$P(t, T_1, T_2) = \frac{P(t, T_2)}{P(t, T_1)}. \quad (7.105)$$

Definition 7.102. The implied hazard rate of default, in discrete time, over $[T, T + \Delta t]$, as seen from any time $t \leq T$ is

$$H(t, T, T + \Delta t) = \frac{1}{\Delta t} \frac{D(t, T, T + \Delta t)}{P(t, T, T + \Delta t)} \quad (7.106)$$

Definition 7.103. The discount factor between time $t \geq 0$ and T is given by

$$\beta(t, T) = e^{-\int_t^T r_s ds}. \quad (7.107)$$

Definition 7.104. Let the default-free zero-coupon bond (ZCB) prices at time t with maturity T be $B(t, T)$. Let the defaultable zero-coupon bond prices at time t with maturity T be $\bar{B}(t, T)$. To ensure no arbitrage, $0 \leq \bar{B}(t, T) \leq B(0, T)$ for all $0 \leq t < T$. The bond prices are normally decreasing, non-negative functions of maturity. For any $0 \leq t < T_1 < T_2$,

$$B(t, T_1) \geq B(t, T_2) > 0. \quad (7.108)$$

$$\bar{B}(t, T_1) \geq \bar{B}(t, T_2) > 0. \quad (7.109)$$

We also write

$$B(t, T) = \mathbb{E}_t[\beta(t, T)]. \quad (7.110)$$

$$\bar{B}(t, T) = B(t, T)P(t, T) \quad (7.111)$$

$P(t, T)$ is called implied survival probability. We may assume $B(t, t) = \bar{B}(t, t) = 1$. $P(t, T)$ is non-negative and decreasing.

Definition 7.105. The non-defaultable simply compounded forward rate over the period $(T_1, T_2]$ is

$$F(t, T_1, T_2) = \frac{B(t, T_1)/B(t, T_2) - 1}{T_2 - T_1}. \quad (7.112)$$

Let $T_2 - T_1 \rightarrow 0$.

$$f(t, T) = \frac{\partial}{\partial T} \ln B(t, T). \quad (7.113)$$

Remark 7.106. *An asset with positive recovery can be viewed as an asset with zero recovery and an additional positive payoff at default.*

Definition 7.107. *Let $e(t, T_{n-1}, T_n)$ be the price of one dollar at time $t \geq 0$ that is paid if default happens in $(T_{n-1}, T_n]$.*

$$e(t, T_{n-1}, T_n) = \mathbb{E} [\beta(t, T_n) 1_{T_{n-1} < \tau \leq T_n} | \mathcal{F}_t \wedge \{\tau > t\}]. \quad (7.114)$$

The price $e(t, T_{n-1}, T_n)$ is given by

$$e(t, T_{n-1}, T_n) = B(t, T_n)(P(t, T_{n-1}) - P(t, T_n)) = \Delta \bar{B}(t, T_n) H(t, T_{n-1}, T_n). \quad (7.115)$$

Within the whole time interval $(0, T_M]$, we have

$$e(0, 0, T_M) = \sum_{n=1}^M e(0, T_{n-1}, T_n) \quad (7.116)$$

Definition 7.108. *(Term structures of forward rates and hazard rates) Let $\delta_{i-1} = T_i - T_{i-1}$.*

$$B(0, T_n) = \prod_{i=1}^n \frac{1}{1 + \delta_{i-1} F(0, T_{i-1}, T_i)}. \quad (7.117)$$

$$P(0, T_n) = \prod_{i=1}^n \frac{1}{1 + \delta_{i-1} H(0, T_{i-1}, T_i)}. \quad (7.118)$$

Let $\delta_i \rightarrow 0$.

$$B(0, T) = \exp \left(- \int_0^T f(0, s) ds \right). \quad (7.119)$$

$$\bar{B}(0, T) = \exp \left(- \int_0^T f(0, s) + h(0, s) ds \right). \quad (7.120)$$

$$e(0, T) = h(0, T) \bar{B}(0, T) \quad (7.121)$$

Definition 7.109. *The price of a defaultable fixed-coupon bond is*

$$\bar{C}(0) = \sum_{n=1}^M \bar{c}_n \bar{B}(0, T_n) + \bar{B}(0, T_M) + \pi \sum_{n=1}^M e(0, T_{n-1}, T_n). \quad (7.122)$$

The expected recovery rate π is assumed to be time and default independent percentage of the reference entity's par value. Assume that the payment is made at T_n if a default occurs in $(T_{n-1}, T_n]$.

Definition 7.110. *A CDS consists of two payment legs: the fixed leg (fee payment) and the floating leg (default insurance).*

- *Fixed leg: payment of $\bar{s}\delta$ at T_n if no default until T_n .*

$$\sum_{n=0}^M \delta \bar{B}(0, T_n)$$

- *Floating leg: payment of $(1 - \pi)$ at T_n if default in $(T_{n-1}, T_n]$.*

$$(1 - \pi) \sum_{n=1}^M e(0, T_{n-1}, T_n) = (1 - \pi) \sum_{n=1}^M \delta H(0, T_{n-1}, T_n) \bar{B}(0, T_n).$$

Thus, the market CDS spread is

$$\bar{s} = (1 - \pi) \frac{\sum_{n=1}^M e(0, T_{n-1}, T_n)}{\sum_{n=0}^M \delta \bar{B}(0, T_n)} = (1 - \pi) \frac{\sum_{n=1}^M H(0, T_{n-1}, T_n) \bar{B}(0, T_n)}{\sum_{n=0}^M \bar{B}(0, T_n)} \quad (7.123)$$

Definition 7.111. (*Absolute Priority Rule*) Upon bankruptcy or liquidation, senior claims (debt claims) are repaid before junior ones (equity).

Definition 7.112. Let the loss L be a random variable. The value at risk (VaR) at confidence level $\alpha \in (0, 1)$ is defined as

$$\text{VaR}_\alpha = \inf\{l \in \mathbb{R} | P(L > l) \leq 1 - \alpha\} = \inf\{l \in \mathbb{R} | F(l) \leq \alpha\}. \quad (7.124)$$

It is the smallest level l so that the probability L exceeds l is no larger than $1 - \alpha$.

Definition 7.113. Let the loss L be a random variable with $\mathbb{E}[|L|] < \infty$. The expected shortfall (ES) or cumulative value at risk (CVaR) at confidence level $\alpha \in (0, 1)$ is defined as

$$\text{ES}_\alpha = \frac{1}{1 - \alpha} \int_\alpha^1 \text{VaR}_u \, du. \quad (7.125)$$

Obviously, $\text{ES}_\alpha \geq \text{VaR}_\alpha$.

Definition 7.114. Let $\rho : \mathbb{R} \rightarrow (0, 1)$ be a risk measure. ρ should have the following desired properties.

1. (*Monotonicity*) If $L_1 \leq L_2$, $\rho(L_1) \leq \rho(L_2)$. Lower losses should be less risky.
2. (*Cash-additivity*) For $l \in \mathbb{R}$, $\rho(L - l) = \rho(L) - l$. Adding cash to position reduces risky by the cash amount.
3. (*Sub-additivity / Diversification*) $\rho(L_1 + L_2) \leq \rho(L_1) + \rho(L_2)$. Diversification reduces risk.
4. (*Positive Homogeneity*) For $\lambda > 0$, $\rho(\lambda L) = \lambda \rho(L)$. Adding same losses will not diversify.
5. (*Convexity*) For $\lambda \in [0, 1]$, $\rho(\lambda L_1 + (1 - \lambda)L_2) \leq \lambda \rho(L_1) + (1 - \lambda)\rho(L_2)$. Diversification for weighted portfolios reduces risk.

ρ is coherent if 1 - 4 are satisfied. ρ is convex if 1, 2 and 5 are satisfied.

Remark 7.115. ES is coherent while VaR is not. ES is convex while VaR is not.

SECTION 8

High-Frequency Trading (Draft)

Definition 8.1. *Algorithmic trading is computer-assisted trading*

1. *largely automated, in contrast to pit trading*
2. *large order execution algorithms, market making strategies*

High-Frequency Trading (Characteristics from SEC report on HFT)

1. Use of extraordinary high speed and sophisticated programs from generating, routing and executing orders.
2. Use of co-location services and individual data feeds offered by exchanges and others to

8.1 Trading Direction

NYSE Trade data does not indicate the direction of market orders. Only traded prices and quotes are recorded. The question is how can we infer market order directions from traded prices.

1. Quotes appear earlier than trades. Lee and Ready (1991) argues quotes lead trades by 5 seconds. However, the latency in quote issuance and trade execution has decreased over time. More recent research has recommended the use of 2 seconds or 0 seconds lags instead.
2. Traded prices can be inside quoted spread.

Two steps of Lee and Ready algorithm

Step One: Quote Test.

If the price of a trade is lower than the midpoint of the matched bid and ask, then the trade is classified as seller-initiated. If the price is higher, the trade is considered to be buyer-initiated. When it is equal, use step two.

Step Two: Tick and Reverse Tick Test

1. Uptick: if price is higher than previous trade
2. Downtick: if price is lower than previous trade
3. Zero-uptick: the price is the same as the previous trade, and the last price change was an uptick
4. Zero-downtick: the price is the same as the previous trade, and the last price change was a downtick
5. Tick test: a trade is classified as a buy if it occurs on an uptick or a zero-uptick; otherwise it is classified as a sell
6. Reverse tick test: If the current trade is followed by a trade with higher (lower) price, the reverse tick test classifies the current trade as a sell (buy).
7. If the price reverses (continues), tick test and reverse tick test agree (disagree).

8.2 Lec 4 Grossman-Miller Market Making model

Market maker

1. responsible for providing liquidity to market participants by quoting prices to buy and sell the assets
2. provides immediacy of trades
3. provides liquidity by quoting buy and sell prices
4. quotes buy prices lower than sell prices to earn the spread

Trade off for market maker

1. provides liquidity to earn spread
 2. inventory subject to price movement risk
-
1. trades with market maker if liquidity is not available from other liquidity traders
 2. may willing to lose some values (does not accept the best prices) to get immediate executions
-
1. How much is liquidity trader willing to pay to get immediate execution
 2. How much does market maker charge to bare inventory risk
 3. How does the price depend on size of trade / volatility of asset price / risk aversion of market maker

8.3 Lec 5 Trading on an informational advantage (Kyle Model)

Different agents have different information. Agents trade when they have or believes to have better information about what the price is going to do than what is reflected in current prices.

8.4 Optimal Execution with Continuous Trading, Lec 7

Background

1. Investors who regularly come to the market with large orders are institutional traders
2. Pension funds, hedge funds, mutual funds, etc
3. These investors usually delegate their trades to an broker
4. Broker will slice the parent order into small parts and try to execute each child order
5. Broker needs to balance between price impact (trade quicker) and price risk (take longer time to complete all trades)
6. The liquidation horizon T is important. Short T leads to faster trading hence more price impact. Long T means that broker can trade slowly but is subject to fundamental price change

Execution costs and slippage

1. Benchmark price: mid-price (price without price impact)
2. Execution costs are measured as the difference between the benchmark price and the actual price (measured as the average price per share)
3. The execution cost is also called cost slippage

Optimal liquidation using lit and dark pool

1. Dark pools are trading venues which do not display bid and ask quotes.
2. We focus on one kind of dark pool known as a crossing network - systems that allow participants to enter unpriced orders to buy and sell securities, these orders are crossed at a specified time at a price derived from another market.
3. Typically, the price is crossed at the midprice in a corresponding lit trading venue
4. When a trader places an order in a dark pool, she may have to wait for some time until a matching order arrives so that her order is executed at the mid price at that time
5. The size of execution depends on the size of arrival orders, so the execution may be partial
6. Dark pool avoids temporary price impact, but the trader is exposed to fundamental price change.

8.5 Liquidation Model, Lec 7 Notes

Liquidation Model

1. Q^ν is the agent's inventory when the agent trades at the speed of ν . Then, Q^ν follows

$$dQ^\nu = -\nu_t dt \quad Q_0 = R. \quad (8.1)$$

2. The midprice follows

$$dS_t^\nu = -g(\nu_t) dt + \sigma dW_t. \quad S_0^\nu = S. \quad (8.2)$$

3. The execution price satisfies

$$\hat{S}_t^\nu = S_t^\nu - \left(\frac{\Delta}{2} + f(\nu_t) \right) \quad (8.3)$$

where $f(\nu)$ is the temporary price impact. Δ is the spread.

4. Agent's cash position follows

$$dX_t = \hat{S}_t^\nu \nu_t dt. \quad X_0 = X. \quad (8.4)$$

The expected revenue from the sale is $\mathbb{E}[\int_0^T \hat{S}_t^\nu \nu_t dt]$. The execution cost is

$$\mathbb{E}[C] = RS_0 - \mathbb{E} \left[\int_0^T \hat{S}_t^\nu \nu_t dt \right] \quad (8.5)$$

Liquidation with only temporary impact

1. We assume linear temporary price impact $f(\nu) = k\nu$.
2. We assume that the bid-ask spread $\Delta = 0$ or equivalently, that S_t represents the best bid price.
3. The trader needs to liquidate a ; R shares by time T .

The trader aims to maximize the expected revenue of liquidation

$$\sup_{\nu} \mathbb{E} \left[\int_0^T (S_u - k\nu_u) \nu_u du \right] \quad (8.6)$$

where we look for optimal selling strategy, $\nu_t \geq 0$, $t \in [0, T]$. Define the value function

$$H(t, s, q) = H_t = \sup_{\nu} \mathbb{E}_t \left[\int_t^T (s_u - k\nu_u) \nu_u du \right] \quad (8.7)$$

By DPP, $H_t + \int_0^t (s_u - k\nu_u) \nu_u du$ is a supermartingale for all ν and is a martingale for some ν^* . Recall that

$$dQ = -\nu_t dt \quad dS_t = \sigma dW_t. \quad (8.8)$$

Then, the trader's HJB equation is

$$\partial_t H + \frac{1}{2} \sigma^2 \partial_{ss} H + \sup_{\nu} \{ (s - k\nu) \nu - \nu \partial_q H \} = 0. \quad (8.9)$$

With terminal conditions

$$H(t, s, q) = -\infty. (*) \quad H(T, s, 0) = 0. \quad (8.10)$$

(*) If there are still shares not liquidated by T , then there is infinite penalty.

From the FOC in 8.9, we have

$$\nu_* = \frac{1}{2k} (s - \partial_q H) \quad (8.11)$$

plug in 8.11 back into 8.9. 8.9 is transformed to

$$\partial_t H + \frac{1}{2} \sigma^2 \partial_{ss} H + \frac{1}{4k} (s - \partial_q H)^2 = 0. \quad (8.12)$$

Guess H of the following form,

$$H(t, s, q) = qs + h(t, q). \quad (8.13)$$

Plug the previous form of H back to 8.12. We obtain

$$\partial_t h + \frac{1}{4k} (\partial_q h)^2 = 0. \quad (8.14)$$

Further guess that

$$h(t, q) = q^2 h_2(t) \quad (8.15)$$

for another function $h_2(t)$.

Then h_2 satisfies

$$\partial_t h_2 + \frac{1}{k} h_2^2 = 0. \quad (8.16)$$

with boundary condition $h_2 \rightarrow -\infty$ as $t \rightarrow T$. 8.16 is equivalent to

$$\frac{d}{dt} \left(-\frac{1}{h_2} \right) = \frac{\partial_t h_2}{h_2^2} = -\frac{1}{k} \quad (8.17)$$

Integrate both sides from t to T and we obtain

$$h_2(t) = \left(-\frac{1}{h_2(T)} + \frac{T-t}{k} \right)^{-1} = \left(\frac{T-t}{k} \right)^{-1} \quad (8.18)$$

Plug in the form of H and h back to 8.11

$$\nu^* = \nu^*(t, S_t, Q_t) = \frac{1}{2k} (S_t - \partial_q H(t, S_t, Q_t)) = -\frac{1}{k} h_2(t) Q_t. \quad (8.19)$$

Then,

$$dQ_t = \frac{1}{k} h_2(t) Q_t \implies \frac{dQ_t}{Q_t} = \frac{1}{k} h_2(t) \quad (8.20)$$

$$\implies \int_0^t \frac{dQ_u}{Q_u} = \int_0^t \frac{1}{k} h_2(u) du \quad (8.21)$$

$$\implies Q_t = \frac{T-t}{T} R \quad (8.22)$$

$\nu_t = \frac{R}{T}$. The shares are liquidated at a constant rate.

8.5.1 Liquidation with permanent and temporary price impact

$$dS_t = -g(\nu_t) dt + \sigma dW_t \quad (8.23)$$

$$dQ_t = -\nu_t dt \quad (8.24)$$

Trader aims to maximize

$$\mathbb{E} \left[X_T + Q_T (S_T - \partial Q_T) - \varphi \int_t^T (Q_u)^2 du \right] \quad (8.25)$$

$$H(t, x, s, q) = \sup_{\nu} \dots \quad (8.26)$$

DPP implies $H(t, x_t, s_t, q_t) - \varphi \int_0^t Q_u^2 du$ is a supermartingale for all ν and is a martingale for some ν^* . Write down the HJB,

$$\partial_t H + \frac{1}{2} \sigma^2 \partial_{ss} H - \varphi q^2 + \sup_{\nu} \{ \nu (s - f(\nu)) \partial_x H - g(\nu) \partial_s H - \nu \partial_q H \} = 0. \quad (8.27)$$

with terminal condition

$$H(t, x, s, q) = x + sq - \partial q^2 \quad (8.28)$$

We consider linear permanent and temporary price impact $f(\nu) = k\nu$ and $g(\nu) = b\nu$. Then, the FOC in 8.27 yields

$$\nu^* = \frac{1}{2k} \frac{s\partial_x H - b\partial_s H - \partial_q H}{\partial_x H} \quad (8.29)$$

Since the terminal condition is $H(T, x, s, q) = x + sq - \partial q^2$, we guess

$$H(t, x, s, q) = x + sq + h(t, s, q) \quad (8.30)$$

$$h(T, s, q) = -\partial q^2 \quad (8.31)$$

Plug 8.30 back to 8.27, we obtain

$$0 = \partial_t h + \frac{1}{2} \partial_{ss} h - \varphi q^2 \frac{1}{4k} [b(q + \partial_s h) + \partial_q h]^2 \quad (8.32)$$

Since this equation depends on derivatives of S and $h(T, s, q) = -\partial q^2$ does not depend on s . Then, we guess h does not depend on s . The previous equation is reduced to

$$0 = \partial_t h(t, q) - \varphi q^2 + \frac{1}{4k} [bq + \partial_q h(t, q)]^2 \quad (8.33)$$

$$\nu^* = -\frac{1}{2k} (\partial_q h + bq) \quad (8.34)$$

As in the previous section, we introduce h_2 such that

$$h(t, q) = h_2(t)q^2. \quad h_2(T) = -2. \quad (8.35)$$

Plug this form of h into 8.33, and introduce X as $h_2(t) = -\frac{b}{2} + X(t)$. Then, 8.33 is transformed to

$$\begin{aligned} \frac{\partial_t X}{k\varphi - X^2} = \frac{1}{k} &\implies \frac{\partial_t X}{(\sqrt{k\varphi} - X)(\sqrt{k\varphi} + X)} = \frac{1}{k} \\ &\implies \frac{\partial_t X}{\sqrt{k\varphi} - X} + \frac{\partial_t X}{\sqrt{k\varphi} + X} = \sqrt{\frac{\varphi}{k}} \end{aligned} \quad (8.36)$$

Integrating both sides on $[t, T]$, we obtain

$$X(t) = \sqrt{k\varphi} \frac{1 + \xi e^{2\gamma(T-t)}}{1 - \xi e^{2\gamma(T-t)}} \quad (8.37)$$

where

$$\gamma = \sqrt{\frac{\varphi}{k}} \quad \text{and} \quad \xi = \frac{\gamma - \frac{b}{2} + \sqrt{k\varphi}}{\gamma - \frac{b}{2} - \sqrt{k\varphi}}. \quad (8.38)$$

Plug X , h back to ν^* , we obtain

$$\nu_t^* = \gamma \frac{\xi e^{\gamma(T-t)} + e^{\gamma(T-t)}}{\xi e^{\gamma(T-t)} - e^{\gamma(T-t)}} Q^{\nu^*} \quad (8.39)$$

We can also obtain (see page 147 in Cartea, Jaimungal, Penalva)

$$Q_t = \frac{\xi e^{\gamma(T-t)} - e^{-\gamma(T-t)}}{\xi e^{\gamma T} - e^{-\gamma T}} R \quad (8.40)$$

when the terminal penalty goes to infinity

The optimal liquidation strategies in these two sections are both deterministic. They are not affected by the market mid-price.

8.5.2 Optimal liquidation in lit and dark markets

1. Lit market is subject to temporary price impact

$$\hat{S}_t = S_t - k\nu_t \quad (8.41)$$

2. The agent can post $y_t \leq q_t$ units of shares in the dark market where $q_t \leq R$ is the remaining shares to be liquidated. The agent can continuously adjust y_t . The agent receives S_t per share for each unit executed in the dark pool, which may not be the whole amount y_t .
3. We assume order arrives at Poisson time in the dark pool. Let N_t be a Poisson process with intensity λ and $\{\epsilon_j : j = 1, 2, \dots\}$ be i.i.d. rvs representing the size of arriving orders in the dark pool. $\{\epsilon_j\}$ is also independent of N . Then, the total arriving orders follow a compound Poisson process $\nu_t = \sum_{n=1}^{N_t} \sigma_n$.
4. The inventory of the agent follows

$$dQ_t = -\nu_t dt - \min(y_t, \epsilon_{N_t-}) dN_t \quad (8.42)$$

5. Cash Poisson

$$dX_t = (s_t - k\nu_t)\nu_t dt + s_t \min(y_t, \epsilon_{N_t-}) dN_t \quad (8.43)$$

The objective function is the same

$$H(t, x, s, q) = \sup_{\nu, y} \mathbb{E}^{x, s, q} \left[x_T + Q_t^{\nu, y} (S_T - \alpha Q_T) - \phi \int_t^T (Q_u)^2 du \right] \quad (8.44)$$

Note that when N jumps by 1. $X \rightarrow x + s \min(y, \epsilon)$. $q \rightarrow q - \min(y, \epsilon)$. The HJB equation is

$$\begin{aligned} \partial_t H + \frac{1}{2} \sigma^2 \partial_{ss} H - \varphi q^2 + \sup_{\nu} \{ (s - k\nu) \nu \partial_x H - \nu \partial_q H \} \\ + \sup_{\nu} \{ \lambda \mathbb{E} [H(t, x + s \min(y, \epsilon), s, s, q - \min(y, \epsilon)) - H(t, x, s, q)] \} = 0. \end{aligned} \quad (8.45)$$

with terminal condition

$$H(t, x, s, q) = x + q(s - \alpha q). \quad (8.46)$$

We still guess

$$H(t, x, s, q) = x + qs + h(t, q). \quad (8.47)$$

Then HJB is reduced to

$$\partial_t h - \varphi q^2 + \sup_{\nu} \{-k\nu^2 - \nu \partial_q h\} + \lambda \sup \mathbb{E} [h(t, q - \min(y, \epsilon)) - h(t, q)] = 0. \quad (8.48)$$

The first order condition in ν yields

$$\nu^* = -\frac{1}{2k} \partial_q h. \quad (8.49)$$

In general, the guessing HJB does not admit a closed form solution. We can consider special case, $\epsilon_j \geq k, j = 1, 2, 3, \dots$, then $\min(\epsilon, y) = y$ l. i.e. the arrival of orders in the dark pool is large enough so that y is fully executed when an order arrives in the dark pool.

The HJB1 is transformed to

$$\partial_t h - \varphi q^2 + \sup_{\nu} \{-k\nu^2 - \nu \partial_q h\} + \lambda \sup \{h(t, q - y) - h(t, q)\} = 0 \quad (8.50)$$

with terminal condition

$$h(T, q) = -\alpha q^2 \quad (8.51)$$

HJB2 has an explicit solution

$$\nu^* = -\frac{1}{k} h_2(t) Q_t^{\nu, y}, \quad y^* = Q_t^{\nu, y} \quad (8.52)$$

for some function $h_2(t)$. This means the optimal amount to send to the dark pool is always the remaining orders. This is because the execution is at mid-price in the dark pool, and the agent can adjust y continuously.

8.6 Targeting Volume, Lec 8

We consider an example of μ which models the rate of total market sell orders submitted by other participants. Consider μ satisfies a mean reverting process

$$d\mu_t = -\kappa \mu_t dt + \eta_{1+N_t-} dN_t \quad (8.53)$$

where N is a homogeneous Poisson process with intensity λ and η_1, η_2, \dots are non-negative i.i.d. random variables. μ satisfies

$$\mu_t = e^{-\kappa t} \mu_0 + \sum_{m=1}^{N_t} e^{-\kappa(t-\tau_m)} \eta_m \quad (8.54)$$

where η_m is the m -th arrival time of the Poisson process

$$\begin{aligned}
 \mathbb{E}[\mu_t] &= e^{-kt} \mu_0 + \mathbb{E} \left[\int_0^t e^{-k(t-u)} \eta_{1+} N_{u-} dN_u \right] \\
 &= e^{-kt} \mu_0 + \int_0^t e^{-k(t-u)} \mathbb{E}[\eta_1] \lambda du \\
 &= e^{-kt} \mu_0 + \frac{\lambda}{k} \mathbb{E}[\eta_1] (1 - e^{-kt}) \rightarrow \frac{\lambda}{k} \mathbb{E}[\eta_1]
 \end{aligned} \tag{8.55}$$

We can also calculate

$$\mathbb{E}[\mu_u | \mathcal{F}_t] = e^{-k(u-t)} \mu_t + \frac{\lambda}{k} \mathbb{E}[\eta_1] (e^{-kt} - e^{-ku}) \tag{8.56}$$

When liquidating an order, a trader needs to consider his volume relative to the total volume sent by all other market participants. The trader wants to camouflage his orders among orders submitted by other participants. One way to do this is to choose a rate of trading which targets a pre-determined fraction of the total volume traded.

8.7 Targeting percentage of market's speed of trading

- * When liquidating, the trader targets a specified percentage of arrival rate of all market sell MOS.
- * When the trader sells at rate ν , then his inventory follows
 - (a) $dQ_t = -\nu_t dt$. $Q_0 = R$.
- * Let μ_t denote the speed at which all other market participants are selling shares using MOS. (This rate can be estimated by summing all shares executed over a small time window, and dividing by the time window.)
- * The trader targets $\rho\mu_t$ at every time.
- * Given a liquidation rate ν , the trader's performance criteria is

$$H = \sup_{\nu} \mathbb{E} \left[X_T + Q_T (S_T - \partial Q_T) - \varphi \int_t^T (\nu_u - \rho\mu_u)^2 du \right] \tag{8.57}$$

- * Mid-price: dt term is the permanent price impact

$$dS_t = -b\nu_t dt + \sigma dW_t \tag{8.58}$$

- * Execution price: temporary price impact

$$\hat{S}_t = S_t - k\nu_t \tag{8.59}$$

- * Cash position

$$dX_t = \hat{S}_t \nu_t dt \tag{8.60}$$

From the dynamic programming principle, H satisfies the following HJB equation.

$$0 = (\partial_t + \frac{1}{2}\sigma^2\partial_{ss} + L)H + \sup_{\nu}\{(s - k\nu)\nu\partial_x H - \nu\partial_q H - b\nu\partial_s H - \varphi(\nu - \rho\mu)^2\} \quad (8.61)$$

with the terminal condition

$$H(T) = x + q(s - \partial q) \quad (8.62)$$

FOC yields

$$\nu^* = \frac{s\partial_x H - \partial_q H - b\partial_s H + 2\varphi\rho\mu}{2(k + \varphi)} \quad (8.63)$$

To solve 8.61, we guess the following form of H .

$$H = x + qs + h(t, \mu, q). \quad (8.64)$$

Plug in and we get the following equation of H

$$0 = (\partial_t + L)h + \frac{1}{4(k + \varphi)}(\partial_q h + bq + 2\varphi\rho\mu)^2 - \varphi\rho^2\mu^2. \quad (8.65)$$

with the terminal condition

$$h = -\partial q^2 \quad (8.66)$$

To solve 8.65, observe that the right-hand side is a quadratic function of q . Then, we guess

$$h(t, \mu, q) = h_0(t, \mu) + qh_1(t, \mu) + q^2h_2(t, \mu). \quad (8.67)$$

We get

$$\nu^* = \frac{1}{k + \varphi} \left\{ [\varphi\rho\mu - \frac{1}{2}h_1(t, \mu)] - [\frac{1}{2}b + h_2(t, \mu)]q \right\} \quad (8.68)$$

We get a quadratic function in q on the right-hand side. Since the equation holds for all q , we get the following three equations for h_0, h_1, h_2 .

$$0 = (\partial_t + L)h_2 + \frac{(h_2 + b/2)^2}{k + \varphi} \quad (8.69)$$

$$0 = (\partial_t + L)h_1 + \frac{h_1 - 2\varphi\rho\mu}{k + \varphi}(h_2 + b/2) \quad (8.70)$$

$$0 = (\partial_t L)h_0 + \frac{1}{4(k + \varphi)}(h_1 - 2\varphi\rho\mu)^2 - \varphi\rho^2\mu^2 \quad (8.71)$$

with terminal conditions

$$h_2(T, \mu) = \partial \quad h_1(T, \mu) = 0 \quad (8.72)$$

We first solve 8.69, then 8.70 and 8.71. The solution of 8.69 is

$$h_2(t, \mu) = - \left(\frac{T-t}{k+\varphi} + \frac{1}{2-b/2} \right)^{-1} - \frac{b}{2} \quad (8.73)$$

Note that the terminal condition and the second term do not depend on μ . Then, h_2 does not depend on μ . 8.73 implies that

$$\frac{b}{2} + h_2(t, \mu) = - \frac{k+\varphi}{T-t+\xi} \quad (8.74)$$

where

$$\xi = \frac{k+\varphi}{2-b/2} \quad (8.75)$$

is a constant. Go back to 8.68, the optimal liquidation rate is

$$\nu^* = \frac{1}{k+\varphi} [\varphi \rho \mu - \frac{1}{2} h_1(t, \mu)] + \frac{q}{T-t+\xi} \quad (8.76)$$

Replace μ by μ_t and q by Q_t . The optimal liquidation rate is

$$\nu^* = \frac{1}{k+\varphi} [\varphi \rho \mu_t - \frac{1}{2} h_1(t, \mu_t)] + \frac{Q_t}{T-t+\xi} \quad (8.77)$$

Remark: Let us focus on the second term. When $\partial \rightarrow \infty$, then the trader cannot afford the terminal liquidation. Then the trader must liquidate all inventory by time T . In this case, $\xi \rightarrow 0$. The second term is a constant liquidation strategy, i.e., liquidate at the rate R/T from 0 to T . We have seen this strategy in the first example of the last chapter. This strategy is called TWAP (time weighted average price).

$$\frac{\int_0^T S_t \frac{R}{T} dt}{T} = \frac{R}{T} \frac{\int_0^T S_t dt}{T} \quad (8.78)$$

To understand the first term in 8.77, we need to solve h_1 . Let us recall the Feynman-Kac representation. Consider the following equation.

$$\partial_t v + Lv - cv + h(\mu, t) = 0 \quad (8.79)$$

with the terminal condition

$$v(T, \mu) = g(\mu) \quad (8.80)$$

Then, Feynman-Kac representation says

$$v(\mu, t) = \mathbb{E} \left[e^{-c(T-t)} g(\mu_T) + \int_t^T e^{-c(u-t)} h(\mu_u, u) du \right] \quad (8.81)$$

Indeed, from 8.81, we expect that

$$e^{-ct} v(\mu_t, t) + \int_0^t e^{-cu} h(\mu_u, u) du \quad (8.82)$$

is a martingale. Apply Ito's formula to calculate its drift, we get

$$e^{-ct} (-cv + \partial_t v + L^\mu v + h(\mu, t)) = 0. \quad (8.83)$$

Come back to 8.70, consider

$$-c = \frac{h_2 + b/2}{k + \varphi} \quad (8.84)$$

$$h = -\frac{2\varphi\rho\mu}{k + \varphi}(h_2 + b/2) = \frac{2\varphi\rho\mu}{T - t + \xi} \quad (8.85)$$

Then, we have from the Feynman-Kac that the solution of 8.70, with the terminal condition $h_1(T, \mu) = 0$, is

$$\begin{aligned} h_1(\mu, t) &= \mathbb{E} \left[\int_t^T e^{\int_t^s \frac{1}{T-s+\xi} ds} \frac{2\varphi\rho\mu_u}{T - u + \xi} du \right] \\ &= 2\varphi\rho \frac{\int_t^T \mathbb{E}[\mu_u] du}{T - t + \xi} \end{aligned} \quad (8.86)$$

When $\alpha \rightarrow \infty$, then r is the average arrival rate of the MOS submitted by other participants during the remaining time.

The optimal speed is

$$\begin{aligned} \nu_t^* &= \frac{\varphi}{k + \varphi} \rho \mu_t \\ &+ \frac{1}{T - t + \xi} \left(Q_t - \frac{\varphi\rho}{k + \varphi} \int_t^T \mathbb{E}[\mu_u | \mathcal{F}_t] du \right) \end{aligned} \quad (8.87)$$

8.7.1 Including Impact of Other Traders

In the previous model, we assume that the liquidating agent's trades move the mid-price, but other traders' trades do not. This assumption is not consistent. In this section, we assume all market participants' trades have permanent price impact. The mid-price follows

$$dS_t = b(\mu_t^+ - (\nu_t + \mu_t^-)) dt + \sigma dW_t \quad (8.88)$$

where μ_t denotes the rate of trading for buy and sell MOs sent by other traders and $\nu_t > 0$ is the agent's liquidation rate. The execution price of the liquidation trader is

8.7.2 stat

$$RV_t = \left(1 - \frac{\bar{n}}{n}\right)^{-1} \left(RV_t^{ave} - \frac{\bar{n}}{n} RV_{0,t}\right) \quad (8.89)$$

$$RV_t^{ave} = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^{n_i} (r_{i,j}^\circ)^2, \quad RV_{0,t} = \sum_{j=1}^n (r_j^\circ)^2 \quad (8.90)$$

$$r_t^\circ = r_t + e_t. \quad (8.91)$$

$$\sum_{t=1}^n (r_t^\circ)^2 \approx \sum_{t=1}^n r_t^2 + n \operatorname{Var}(e). \quad (8.92)$$

$$\frac{\bar{n}}{n} RV_{0,t} \approx \frac{\bar{n}}{n} \sum_{t=1}^n r_t^2 + \bar{n} \operatorname{Var}(e). \quad (8.93)$$

8.8 Pair Trading lec 9

Many traders do not only look at one asset, but consider interactions between different assets. This works best with groups of assets that share common stocks. For example, assets from the same industry. Consider two assets INTC and SMHC. We model the dynamics of these two assets as a 2-dim mean-reverting process

$$dS_{t,1} = k_{11}(\theta_{11} - S_{t,1}) dt + k_{12}(\theta_{12} - S_{t,2}) dt + \sigma_{11} dW_1 + \sigma_{12} dW_2. \quad (8.94)$$

$$dS_{t,2} = k_{21}(\theta_{11} - S_{t,1}) dt + k_{22}(\theta_{22} - S_{t,2}) dt + \sigma_{11} dW_1 + \sigma_{12} dW_2. \quad (8.95)$$

k represents the speed of mean-reverting. Now we transform the system by

$$\tilde{S}_{t,1} = \alpha_{11} S_{t,1} + \alpha_{12} S_{t,2} \quad (8.96)$$

$$\tilde{S}_{t,2} = \alpha_{21} S_{t,1} + \alpha_{22} S_{t,2} \quad (8.97)$$

and thus

$$d\tilde{S}_{t,1} = k_{11}(\theta_{11} - \tilde{S}_{t,1}) dt + \sigma_{11} dW_1 + \sigma_{12} dW_2. \quad (8.98)$$

$$d\tilde{S}_{t,2} = k_{21}(\theta_{11} - \tilde{S}_{t,1}) dt + \sigma_{11} dW_1 + \sigma_{12} dW_2. \quad (8.99)$$

$$k = \begin{pmatrix} \tilde{k}_1 & 0 \\ 0 & \tilde{k}_2 \end{pmatrix} \quad (8.100)$$

\tilde{k}_1, \tilde{k}_2 are two eigenvalues of k with $\tilde{k}_1 > \tilde{k}_2$. $\tilde{S}_{t,1}$ has the largest mean-reverting speed, called the co-integration factor.

To estimate the stochastic model after transformation, we define a vector autoregressive process (VAR)

$$\Delta S_t = A + B \Delta S_{t-1} + \epsilon_t \quad (8.101)$$

A is a vector of constants and B is a matrix of constants and ϵ is a vector of white noises. In discrete versions,

$$S_t - S_{t-1} = k(\theta - S_{t-1})\Delta t + \sigma(W_t - W_{t-1}) \quad (8.102)$$

$$S_{t-1} - S_{t-2} = k(\theta - S_{t-2})\Delta t + \sigma(W_{t-1} - W_{t-2}) \quad (8.103)$$

$$(S_t - S_{t-1}) = (1 - k\Delta t)(S_{t-1} - S_{t-2}) + \sigma(W_t - 2W_{t-1} + W_{t-2}). \quad (8.104)$$

$$k = \frac{I - B}{\Delta t} \quad (8.105)$$

Pair trading bets on the empirical fact of mean-reverting behavior. It is a class of statistical arbitrage strategy. This is not a true arbitrage. It may lead to loss. Consider the co-integration factor, which is a portfolio of two assets. The strategy is to go long the portfolio factor when it is cheap and close the position when the portfolio's value increases, or go short when the portfolio is expensive and close the position when portfolio's value decreases.

8.9 Ad Hoc Bands

A simple strategy to profit the co-integration factor's mean-reverting is to place bands which are 1-sd above and below the mean-reverting level.

1. Buy (sell) one unit of the portfolio if the lower (upper) band is hit. Close the position when the value of the portfolio is within a small interval, say 1/10-sd of the mean-reverting level.
2. At the end of the trading day, the strategy may end in a long or short position which never reverts back to the inner band by the end of the trading day. The wider the trigger bands, the more likely it is to end with inventory.

8.10 Optimal Stopping Problem

Given a process

$$dX_t = \mu(t, X_t) dt + \sigma(t, X_t) dW_t \quad (8.106)$$

with the time t and the state X such that X starts from X_t at time t , the agent aims to maximize $\mathbb{E}[G(X_\tau)]$, where G is the payoff function when the process is stopped. Define the agent's value function

$$H(t, X) = \sup_{\tau \in J_{[t, T]}} \mathbb{E}_{t, X}[G(X_\tau)] \quad (8.107)$$

where $J_{[t, T]}$ is the set of F -stopping time with value in $[t, T]$. Replace x by X_t , then the process $H(t, X_t)$ is the so called snell envelop. By the dynamic programming principle, $H(t, X_t)$ is a supermartingale for all t and is a martingale until the optimal stopping time t^* .

Apply Ito's formula to $H(t, X_t)$ and get

$$\partial_t H + \mu \partial_x H + \frac{1}{2} \sigma^2 \partial_{xx} H \leq 0. \quad (8.108)$$

Until the optimal stopping time,

$$\max \left\{ \partial_t H + \mu \partial_x H + \frac{1}{2} \sigma^2 \partial_{xx} H, G - H \right\} = 0 \quad (8.109)$$

It can be transformed to a free boundary problem

$$\partial_t H + \mu \partial_x H + \frac{1}{2} \sigma^2 \partial_{xx} H = 0. \quad (8.110)$$

$$H(t, X) = G(t, X). \quad (8.111)$$

Optimal Band Selection

We have seen that the P and L of a pair trading strategy depends on the choice of the bands. We will choose them optimally here. Consider the co-integration factor S which follows mean-reverting. We need to determine buy time η and sell time τ optimally.

Let ρ be the trader's discounting rate (the larger ρ is, the more impatient the trader is). Let c be the transaction cost for each trade. Given that the process S start from S_t at time t , the optimal exiting problem is

$$H(t, S) = \sup_{\tau} \mathbb{E}_{t,S}[e^{-\rho(\tau-t)}(S_t - c)]. \quad (8.112)$$

$H(t, S)$ is the best value if the trader sells the position optimally in the future. Once we obtain H , we can consider P_2 , Given the process S starts from S_t at time t , the optimal entering problem is

$$G(t, S) = \sup_{\eta} \mathbb{E}_{t,S}[e^{-\rho(\eta-t)}(H(\eta, S_{\eta}) - S_t - c)]. \quad (8.113)$$

Consider an infinite horizon problem, H and G do not depend on t . Solve 8.112. The HJB equation associated is

$$\max\{(L - P)H(\epsilon); \epsilon - c - H(\epsilon)\} = 0 \quad (8.114)$$

SECTION A

Analysis

Theorem A.1. (Algebraic Properties of the Field) A field \mathbb{F} is a set with two binary operations denoted as $+$ and \cdot as well as called addition and multiplication respectively that satisfies the following properties. Let $a, b, c \in \mathbb{F}$.

1. (Commutative Property of Addition) $a + b = b + a$.
2. (Associative Property of Addition) $(a + b) + c = a + (b + c)$.
3. (Existence of a Zero Element) There exists 0 in \mathbb{F} such that $0 + a = a + 0 = a$.
4. (Existence of Negative Elements) There exists $-a \in \mathbb{F}$ such that $a + (-a) = (-a) + a = 0$.
5. (Commutative Property of Multiplication) $a \cdot b = b \cdot a$.
6. (Associative Property of Multiplication) $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.
7. (Existence of a One Element) There exists 1 in \mathbb{F} such that $1 \cdot a = a \cdot 1 = a$.
8. (Existence of Reciprocals) If $a \neq 0$, there exists $a^{-1} \in \mathbb{F}$ such that $a \cdot a^{-1} = a^{-1} \cdot a = 1$.
9. (Distributive Property) $a \cdot (b + c) = a \cdot b + a \cdot c$.

Remark A.2. Let $a, b \in \mathbb{F}$. The subtraction is defined as $a - b = a + (-b)$. If $b \neq 0$, the division is defined as $\frac{a}{b} = a \cdot b^{-1}$. The sets \mathbb{Q} , \mathbb{R} , and \mathbb{C} are all fields.

Theorem A.3. Let $a, b \in \mathbb{F}$.

1. If $a + b = a$, then $b = 0$.
2. If $a \neq 0$ and $a \cdot b = a$, then $b = 1$.
3. $a \cdot 0 = 0$.
4. If $a \neq 0$ and $a \cdot b = 1$, then $b = a^{-1}$.
5. If $a \cdot b = 0$, then either $a = 0$ or $b = 0$.

Definition A.4. (The Order Properties of Real Numbers) Denote \mathbb{R}_{++} as the set of positive real numbers. Let $a, b \in \mathbb{R}_{++}$.

1. $a + b \in \mathbb{R}_{++}$.
2. $ab \in \mathbb{R}_{++}$.
3. (Trichotomy Property) If $c \in \mathbb{R}$, then exactly one of the following statements holds.

$$c \in \mathbb{R}_{++}, \quad c = 0, \quad -c \in \mathbb{R}_{++}.$$

Remark A.5. $\mathbb{R}_+ = \mathbb{R}_{++} \cup \{0\}$.

Definition A.6. Let $a, b \in \mathbb{R}$. $a > b$ or $b < a$ if $a - b \in \mathbb{R}_{++}$. $a \geq b$ or $b \geq a$ if $a - b \in \mathbb{R}_+$.

Theorem A.7. Let $a, b, c \in \mathbb{R}$ and $a > b$.

1. If $b > c$, then $a > c$.
2. $a + c > b + c$.
3. If $c > 0$, then $ac > bc$. If $c < 0$, then $ac < bc$.

Definition A.8. The rational numbers are real numbers that can be written as $\frac{a}{b}$ where $a, b \in \mathbb{Z}$ and $b \neq 0$. The irrational numbers are the real numbers that are not rational.

Remark A.9.

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

Theorem A.10. (Triangle Inequality) If $a, b \in \mathbb{R}$, then $|a + b| \leq |a| + |b|$.

Definition A.11. Let S be a nonempty subset of \mathbb{R} and $s \in S$.

1. The set S is bounded above if there exists a number $u \in \mathbb{R}$ such that $s \leq u$ for all s . u is an upper bound. The set S is bounded below if there exists a number $l \in \mathbb{R}$ such that $l \leq s$ for all s . l is an lower bound.
2. A set is bounded if it is both bounded above and bounded below. A set is unbounded if it is not bounded.
3. If S is bounded above, then u is a supremum or least upper bound, denoted as $\sup S$, if u is an upper bound and $u \leq v$ if v is any upper bound. If S is bounded below, then l is a infimum or greatest lower bound, denoted as $\inf S$, if l is a lower bound and $k \leq l$ if k is any lower bound.

Theorem A.12. (The Completeness Property) Every nonempty set of real numbers that has an upper bound also has a supremum in \mathbb{R} .

Definition A.13. A sequence (x_n) converges to a limit x , denoted as $x_n \rightarrow x$, if for any $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that $|x_n - x| < \epsilon$ for all $n \geq N$. A sequence is convergent if it has a limit and divergent otherwise.

Remark A.14. It can be extended to the limit of a function at infinity.

Definition A.15. A sequence (x_n) is bounded if there exists $N > 0$ such that $|x_n| \leq N$ for all $n \in \mathbb{N}$.

Remark A.16. It can be extended to boundedness of a function.

Theorem A.17. Convergence implies boundedness. Boundedness does not imply convergence.

Theorem A.18. Let (x_n) and (y_n) be two sequences that converge to x and y respectively. Let $c \in \mathbb{R}$. Then,

1. $(x_n \pm y_n)$, $(x_n \cdot y_n)$ and (cx_n) converge to $x \pm y$, xy and cx respectively.
2. If $y \neq 0$, then $\left(\frac{x_n}{y_n}\right)$ converges to $\frac{x}{y}$.

Remark A.19. It can be extended to limits and continuities of functions.

Theorem A.20. (*Squeeze Theorem*) Suppose that (x_n) , (y_n) and (z_n) are sequences such that $x_n \leq y_n \leq z_n$ for all $n \in \mathbb{N}$ and $\lim x_n = \lim z_n$. Then, (y_n) is convergent and $\lim x_n = \lim y_n = \lim z_n$.

Remark A.21. It can be extended to limits of functions.

Definition A.22. A sequence (x_n) is increasing or strictly increasing if $x_{n+1} \geq x_n$ or $x_{n+1} > x_n$ for any $n \in \mathbb{N}$. A sequence (x_n) is decreasing or strictly decreasing if $x_{n+1} \leq x_n$ or $x_{n+1} < x_n$ for any $n \in \mathbb{N}$. A sequence is monotone if it is increasing or decreasing. A sequence is constant if it is both increasing and decreasing.

Theorem A.23. (*Monotone Convergence Theorem*) A bounded monotonic sequence is convergent.

Theorem A.24. The Euler's number is defined as

$$e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n. \quad (\text{A.1})$$

Theorem A.25. Any subsequence of a convergent sequence is convergent.

Theorem A.26. There exists a monotonic subsequence for any sequence.

Theorem A.27. (*The Bolzano-Weierstrass Theorem*) A bounded sequence has a convergent subsequence.

Theorem A.28. Let $x \in \mathbb{R}$. The sequence converges to x if any subsequence converges to x .

Definition A.29. Let (x_n) be a bounded sequence.

1. The limit superior of a_n , denoted as $\limsup a_n$, is the infimum of the set of $b \in \mathbb{R}$ such that $b < x_n$ for at most a finite number of $n \in \mathbb{N}$.
2. The limit inferior of a_n , denoted as $\liminf a_n$, is the supremum of the set of $b \in \mathbb{R}$ such that $x_n < b$ for at most a finite number of $n \in \mathbb{N}$.

Theorem A.30. A bounded sequence (x_n) is convergent if and only if $\limsup x_n = \liminf x_n$.

Definition A.31. A sequence (x_n) is Cauchy if, for any $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that $|x_n - x_m| < \epsilon$ for all $m > n \geq N$.

Theorem A.32. (*Cauchy Convergence Theorem*) A sequence is convergent if and only if it is Cauchy.

Definition A.33. (*Proper Divergence*) Let (x_n) be a sequence of real numbers. (x_n) tends to $\pm\infty$ if for any $r \in \mathbb{R}$ there exists a natural number N such that $x_n > r$ or $x_n < r$ if $n \geq N$.

Theorem A.34. A monotonic sequence of real numbers is properly divergent if and only if it is unbounded.

Remark A.35. It can be extended to proper divergence of functions.

Theorem A.36. (*Stolz-Cesaro Theorem*) If (a_n) and (b_n) are two sequences such that (b_n) is monotone and unbounded, then

$$\liminf_{n \rightarrow \infty} \frac{a_{n+1} - a_n}{b_{n+1} - b_n} \leq \liminf_{n \rightarrow \infty} \frac{a_n}{b_n} \leq \limsup_{n \rightarrow \infty} \frac{a_n}{b_n} \leq \limsup_{n \rightarrow \infty} \frac{a_{n+1} - a_n}{b_{n+1} - b_n}. \quad (\text{A.2})$$

Definition A.37. Let $A \subseteq \mathbb{R}$. A point $c \in \mathbb{R}$ is a cluster point of A if for any $\delta > 0$ there exists at least one point $x \in A$ and $x \neq c$ such that $|x - c| < \delta$.

Definition A.38. Let $A \subseteq \mathbb{R}$ and c be a cluster point of A . There exists a limit L of f at c if given any $\epsilon > 0$, there exists a $\delta > 0$ such that if $x \in A$ and $0 < |x - c| < \delta$, then $|f(x) - L| < \epsilon$.

Theorem A.39. There exists a limit if both left-handed and right-handed limits exist and are equal.

Definition A.40. Let $A \subseteq \mathbb{R}$ and $f : A \rightarrow \mathbb{R}$. Let $c \in A$. f is continuous at c if, for any $\epsilon > 0$, there is a $\delta > 0$ such that if $x \in A$ satisfy $|x - c| < \delta$, then $|f(x) - f(c)| < \epsilon$.

Theorem A.41. (Boundedness Theorem) Continuity on a closed interval implies boundedness.

Definition A.42. Let $A \subseteq \mathbb{R}$ and $f : A \rightarrow \mathbb{R}$. f is uniformly continuous on A if, for any $\epsilon > 0$, there is a $\delta > 0$ such that if $x, y \in A$ satisfy $|x - y| < \delta$, then $|f(x) - f(y)| < \epsilon$.

Theorem A.43. (Uniform Continuity Theorem) Let I be a closed bounded interval and let $f : I \rightarrow \mathbb{R}$ be continuous on I . Then, f is uniformly continuous on I .

Definition A.44. Let $A \subseteq \mathbb{R}$ and let $f : A \rightarrow \mathbb{R}$. If there exists a constant $k > 0$ such that $|f(x) - f(y)| \leq k|x - y|$ for any $x, y \in A$, then f is a Lipschitz function on A . Lipschitz functions are uniformly continuous.

Definition A.45. Let $I \subseteq \mathbb{R}$ and $f : I \rightarrow \mathbb{R}$. Let $c \in I$. There exists the derivative L of f at c if, for any $\epsilon > 0$, there exists $\delta > 0$ such that

$$\left| \frac{f(x) - f(c)}{x - c} - L \right| < \epsilon \quad (\text{A.3})$$

if $0 < |x - c| < \delta$. Then, f is differentiable at c and

$$f'(c) = \lim_{x \rightarrow c} \frac{f(x) - f(c)}{x - c} = L. \quad (\text{A.4})$$

Theorem A.46. Differentiability implies continuity. Continuity does not imply differentiability.

Theorem A.47. (Chain Rule) Let $I \subseteq \mathbb{R}$ and $J \subseteq \mathbb{R}$. Let $f : I \rightarrow \mathbb{R}$ and $g : J \rightarrow \mathbb{R}$ be functions such that $g(J) \subseteq I$. Let $c \in J$. If g is differentiable at c and f is differentiable at $g(c)$, then the composite function $f \circ g$ is differentiable at c and

$$(f \circ g)'(c) = (f' \circ g)(c) \cdot g'(c). \quad (\text{A.5})$$

Theorem A.48. Let $I \subseteq \mathbb{R}$ and $f : I \rightarrow \mathbb{R}$ be strictly monotone and continuous. Let $J \subseteq f(I)$ and $g : J \rightarrow \mathbb{R}$ be the strictly monotone, continuous and inverse to f . If f is differentiable at $c \in I$ and $f'(c) \neq 0$, then g is differentiable at $f(c)$ and

$$(g' \circ f)(c) = \frac{1}{f'(c)}. \quad (\text{A.6})$$

Theorem A.49. (Interior Extremum Theorem) Let $I \subseteq \mathbb{R}$ and $f : I \rightarrow \mathbb{R}$. If f has an interior extremum at $c \in I$, then $f'(c) = 0$ or $f'(c)$ does not exist.

Theorem A.50. (Mean Value Theorem) Suppose that f is continuous on $[a, b]$ and f has a derivative on (a, b) . Then, there exists at least one point c on (a, b) such that

$$f(b) - f(a) = f'(c)(b - a). \quad (\text{A.7})$$

Theorem A.51. (L'Hospital's Rule) Let an open interval $I \subseteq \mathbb{R}$ and $c \in I$. Let f and g be differentiable on I except possibly at c . If $g'(x) \neq 0$ and

$$(i) \lim_{x \rightarrow c} f(x) = \lim_{x \rightarrow c} g(x) = 0 \quad \text{or} \quad (ii) \lim_{x \rightarrow c} g(x) = \pm\infty$$

for all $x \in I \setminus \{c\}$, then

$$\lim_{x \rightarrow c} \frac{f(x)}{g(x)} = \lim_{x \rightarrow c} \frac{f'(x)}{g'(x)} = L \in \mathbb{R} \cup \{\pm\infty\}. \quad (\text{A.8})$$

Theorem A.52. (Taylor's Theorem) Let $f : [a, b] \rightarrow \mathbb{R}$ be such that f and its derivatives $f', f'', \dots, f^{(n)}$ are continuous on $[a, b]$ and that $f^{(n+1)}$ exists on (a, b) . Let $x_0 \in [a, b]$. For any $x \in [a, b]$, there exists a point c between x and x_0 such that

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k + \frac{f^{(n+1)}(c)}{(n+1)!} (x - x_0)^{n+1}. \quad (\text{A.9})$$

Corollary A.53. Let $k = k_1 + k_2 + \dots + k_n$ where $k_1, k_2, \dots, k_n \in \mathbb{N}$. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ such that f and its all partial derivatives up to n

$$D^k f = \frac{\partial^k f}{\partial^{k_1} x_1 \partial^{k_2} x_2 \dots \partial^{k_n} x_n} \quad (\text{A.10})$$

where $k \leq n$ are continuous and all partial derivatives $D^{n+1} f$ exist. Let $a \in \mathbb{R}^n$. For any $x \in \mathbb{R}^n$, there exists a point $c \in \mathbb{R}^n$ such that

$$f(x) = \sum_{k \leq n} \frac{D^k f(a)}{k!} (x - a)^k + \sum \frac{D^{n+1} f(c)}{(n+1)!} (x - a)^{n+1}. \quad (\text{A.11})$$

Definition A.54. A function $f : [a, b] \rightarrow \mathbb{R}$ is generalized Riemann integrable on $[a, b]$ and denoted as $\mathcal{R}^*[a, b]$ if there exists $L \in \mathbb{R}$ such that, for any $\epsilon > 0$, there exists a gauge $\delta > 0$ such that if \mathcal{P} is any δ -fine partition of $[a, b]$ with $[x_{i-1}, x_i] \subseteq [t_i - \delta(t_i), t_i + \delta(t_i)]$, then $|S(f; \mathcal{P}) - L| < \epsilon$.

Remark A.55. This is an extension to Riemann integrable ($\mathcal{R}[a, b]$). The gauge is more delicate than the norm ($\|\mathcal{P}\| < \delta$) since it provides more control in choosing subintervals.

Definition A.56. A function $f \in \mathcal{R}^*[a, b]$ such that $|f| \in \mathcal{R}^*[a, b]$ is Lebesgue integrable, denoted as $\mathcal{L}[a, b]$.

Theorem A.57. If $f \in \mathcal{R}^*[a, b]$, the following assertions are equivalent.

1. $f \in \mathcal{L}[a, b]$.
2. There exists $\omega \in \mathcal{L}[a, b]$ such that $f(x) \leq \omega(x)$ for all $x \in [a, b]$.
3. There exists $\alpha \in \mathcal{L}[a, b]$ such that $\alpha(x) \leq f(x)$ for all $x \in [a, b]$.

Definition A.58. A function f is (Lebesgue) measurable, denoted as $\mathcal{M}[a, b]$, if there exists a sequence of step functions (s_n) on $[a, b]$ and a null set $N \subseteq [a, b]$ such that $f(x) = \lim_{n \rightarrow \infty} s_n(x)$ for all $x \in [a, b] \setminus N$.

Theorem A.59. $f : [a, b] \rightarrow \mathbb{R} \in \mathcal{M}[a, b]$ if and only if there exists a sequence of continuous functions (f_n) and a null set $N \subseteq [a, b]$ such that $f = \lim_{n \rightarrow \infty} f_n$ for any $x \in [a, b] \setminus N$.

Theorem A.60. (Measurability Theorem) If $f \in \mathcal{R}^*[a, b]$, then $f \in \mathcal{M}[a, b]$.

Theorem A.61. (Properties of Generalized Riemann Integrals)

1. (Consistency) If $f \in \mathcal{R}[a, b]$, then $f \in \mathcal{R}^*[a, b]$.
2. (Uniqueness) If $f \in \mathcal{R}^*[a, b]$, then the value of the integral is uniquely determined.
3. (Countable Discontinuities) If $f \in \mathcal{R}^*[a, b]$ and $f(x) = g(x)$ except for a countable number of points in $[a, b]$, then $g \in \mathcal{R}^*[a, b]$ and $\int f = \int g$.
4. (Boundedness) If $f \in \mathcal{R}^*[a, b]$, then f is bounded on $[a, b]$.

Theorem A.62. (Cauchy Criterion) Let \mathcal{P} and \mathcal{Q} be any tagged partitions of $[a, b]$. A function $f : [a, b] \rightarrow \mathbb{R} \in \mathcal{R}^*[a, b]$ if and only if, for every $\epsilon > 0$, there exists $\delta > 0$ such that $|S(f; \mathcal{P}) - S(f; \mathcal{Q})| < \epsilon$.

Theorem A.63. (Squeeze Theorem) $f : [a, b] \rightarrow \mathbb{R} \in \mathcal{R}^*[a, b]$ if and only if, for every $\epsilon > 0$, there exist functions $\alpha(x)$ and $\omega(x) \in \mathcal{R}^*[a, b]$ with $\alpha(x) \leq f(x) \leq \omega(x)$ anywhere such that

$$\int_a^b (\omega(x) - \alpha(x)) dx < \epsilon. \quad (\text{A.12})$$

Theorem A.64. (Additivity Theorem) Let $f : [a, b] \rightarrow \mathbb{R}$ and $c \in (a, b)$. $f \in \mathcal{R}^*[a, b]$ if and only if $f \in \mathcal{R}^*[a, c]$ and $f \in \mathcal{R}^*[c, b]$. In this case,

$$\int_a^b f = \int_a^c f + \int_c^b f. \quad (\text{A.13})$$

Theorem A.65. (Fundamental Theorem of Calculus, Form 1) Suppose there is a countable set $E \subseteq [a, b]$ and functions $f, F : [a, b] \rightarrow \mathbb{R}$ such that F is continuous on $[a, b]$ and $F'(x) = f(x)$ for all $x \in [a, b] \setminus E$. Then, $f \in \mathcal{R}^*[a, b]$ and

$$\int_a^b f = F(b) - F(a). \quad (\text{A.14})$$

Theorem A.66. (Fundamental Theorem of Calculus, Form 2) Let $f \in \mathcal{R}^*[a, b]$ and let F be the indefinite integral of f defined by

$$F(x) = \int_a^x f(t) dt \quad (\text{A.15})$$

for $x \in [a, b]$. Then, F is continuous on $[a, b]$. There exists a null set N such that if $x \in [a, b] \setminus N$, then F is differentiable at x and $F'(x) = f(x)$. If f is continuous at a point $c \in [a, b]$, $F'(c) = f(c)$.

Corollary A.67. Let $f(x, t) \in \mathcal{R}^*[a, b] \times \mathcal{R}^*[\alpha, \beta]$ with $\frac{\partial f}{\partial t}(x, t) \in \mathcal{R}^*[a, b] \times \mathcal{R}^*[\alpha, \beta]$. Let $a(t)$ and $b(t)$ be differentiable on $[\alpha, \beta]$ with $a \leq a(t)$ and $b(t) \leq b$. Then,

$$\varphi(t) = \int_{a(t)}^{b(t)} f(x, t) dx \quad (\text{A.16})$$

is differentiable and

$$\varphi'(t) = \int_{a(t)}^{b(t)} \frac{\partial}{\partial t} f_t(x, t) dx + f(b(t), t)b'(t) - f(a(t), t)a'(t). \quad (\text{A.17})$$

Theorem A.68. If $f \in \mathcal{R}^*[a, b]$ and g is a monotone function on $[a, b]$, then $fg \in \mathcal{R}^*[a, b]$.

Theorem A.69. If $f, g \in \mathcal{R}[a, b]$, then $fg \in \mathcal{R}[a, b]$.

Theorem A.70. (Integration by Parts) Let f and g be differentiable on $[a, b]$. $f'g \in \mathcal{R}^*[a, b]$ if and only if $fg' \in \mathcal{R}^*[a, b]$. In this case,

$$\int_a^b f'g = fg|_a^b - \int_a^b fg'. \quad (\text{A.18})$$

Theorem A.71. (Lebesgue's Integrability Criterion) A bounded function $f : I \rightarrow \mathbb{R}$ is Riemann integrable if and only if it is continuous almost everywhere on I .

Theorem A.72. (Monotone Convergence Theorem) Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $\mathbb{X} \in \mathcal{F}$. Let $(f_k)_{k \in \mathbb{N}}$ be an increasing sequence of $(\mathcal{F}, \mathcal{B}(\mathbb{R}_+))$ -measurable positive functions and $f(x) = \lim_{k \rightarrow \infty} f_k(x)$ for any $x \in \mathbb{X}$. Then, f is $(\mathcal{F}, \mathcal{B}(\mathbb{R}_+))$ -measurable and

$$\lim_{k \rightarrow \infty} \int_{\mathbb{X}} f_k d\mu = \int_{\mathbb{X}} f d\mu. \quad (\text{A.19})$$

Theorem A.73. (Fatou's Lemma) Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $\mathbb{X} \in \mathcal{F}$. Let $(f_k)_{k \in \mathbb{N}}$ be an increasing sequence of $(\mathcal{F}, \mathcal{B}(\mathbb{R}_+))$ -measurable positive functions and $f(x) = \liminf_{k \rightarrow \infty} f_k(x)$ for any $x \in \mathbb{X}$. Then, f is $(\mathcal{F}, \mathcal{B}(\mathbb{R}_+))$ -measurable and

$$\liminf_{k \rightarrow \infty} \int_{\mathbb{X}} f_k d\mu \geq \int_{\mathbb{X}} f d\mu. \quad (\text{A.20})$$

Theorem A.74. (Dominated Convergence Theorem) Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $\mathbb{X} \in \mathcal{F}$. Let $(f_k)_{k \in \mathbb{N}}$ be a sequence of $(\mathcal{F}, \mathcal{B}(\mathbb{R}_+))$ -measurable functions and $f(x) = \lim_{k \rightarrow \infty} f_k(x)$ for any $x \in \mathbb{X}$. If there is an integrable function g such that $|f_k(x)| \leq g(x)$ for any $x \in \mathbb{X}$ and $n \in \mathbb{N}$, then f is $(\mathcal{F}, \mathcal{B}(\mathbb{R}_+))$ -measurable and

$$\lim_{k \rightarrow \infty} \int_{\mathbb{X}} f_k d\mu = \int_{\mathbb{X}} f d\mu. \quad (\text{A.21})$$

Theorem A.75. (Tonelli-Fubini Theorem) Let $(\mathbb{X}, \mathcal{S}, \mu)$ and $(\mathbb{Y}, \mathcal{T}, \nu)$ be σ -finite measure spaces. Let $f : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}$ be an $\mathcal{S} \otimes \mathcal{T}$ -measurable function that is either positive or integrable with respect to the product measure $\mu \times \nu$. Then,

1. $\int_{\mathbb{X}} f(x, y) \mu(dx)$ exists for ν -a.e. $y \in \mathbb{Y}$. $\int_{\mathbb{Y}} f(x, y) \nu(dy)$ exists for μ -a.e. $x \in \mathbb{X}$.

2. $g : x \rightarrow \int_{\mathbb{Y}} f(x, y) \mu(dy)$ is μ -a.e. measurable. $h : y \rightarrow \int_{\mathbb{X}} f(x, y) \mu(dx)$ is ν -a.e. measurable.

Thus,

$$\int_{\mathbb{X} \times \mathbb{Y}} f(x, y) (\mu \times \nu)(dx, dy) = \int_{\mathbb{X}} \left(\int_{\mathbb{Y}} f(x, y) \nu(dy) \right) \mu(dx) = \int_{\mathbb{Y}} \left(\int_{\mathbb{X}} f(x, y) \mu(dx) \right) \nu(dy). \quad (\text{A.22})$$

Theorem A.76. (Substitution Theorem) Let $(\mathbb{X}, \mathcal{S}, \mu)$ be any measure space and $(\mathbb{Y}, \mathcal{T})$ be any measurable space. Let $g : \mathbb{X} \rightarrow \mathbb{Y}$ be any \mathcal{T}/\mathcal{S} -measurable function and $f : \mathbb{Y} \rightarrow \mathbb{R}$ be any \mathcal{T} -measurable function. Then,

$$\int_{\mathbb{X}} (f \circ g) d\mu = \int_{\mathbb{Y}} f d(\mu \circ g^{-1}). \quad (\text{A.23})$$

Definition A.77. Let $A_0 \subseteq A \subseteq \mathbb{R}$. A sequence of functions $(f_n : A \rightarrow \mathbb{R})$ converges uniformly to a function $f : A_0 \rightarrow \mathbb{R}$ if, for any $\epsilon > 0$, there exists N such that $|f_n(x) - f(x)| < \epsilon$ for all $x \in A_0$ and $n > N$.

Theorem A.78. (Continuity Theorem) If a sequence of continuous functions (f_n) converges uniformly to a function f , then f is continuous.

Theorem A.79. (Differentiability Theorem) Let $I \subseteq \mathbb{R}$ be a bounded interval and let $(f_n : I \rightarrow \mathbb{R})$ be a sequence of functions. Suppose that there exists $x_0 \in I$ such that $f_n(x_0)$ converges and that the sequence of derivatives (f'_n) exists and converges uniformly to a function g . Then, (f_n) converges uniformly to f that has a derivative at any $x \in I$ and $f' = g$.

Theorem A.80. (Uniform Convergence Theorem) Let (f_n) be a sequence of functions in $\mathcal{R}^*[a, b]$ and suppose that f_n converges uniformly to f . Then, $f \in \mathcal{R}^*[a, b]$ and

$$\int_a^b f = \lim_{n \rightarrow \infty} \int_a^b f_n. \quad (\text{A.24})$$

Theorem A.81. (Dini Theorem) Let (f_n) be a monotone sequence of continuous functions that converges to a continuous function f . The convergence of (f_n) is uniform.

Theorem A.82. (Bounded Convergence Theorem) Let (f_n) be a sequence of functions in $\mathcal{R}^*[a, b]$ that converges to a function $f \in \mathcal{R}^*[a, b]$. Suppose that there exists $B > 0$ such that $|f_n(x)| \leq B$ for all $x \in [a, b]$ and $n \in \mathbb{N}$. Then,

$$\int_a^b f = \lim_{n \rightarrow \infty} \int_a^b f_n.$$

Definition A.83. Let (x_n) be a sequence. $\sum x_n$ is absolutely convergent if $\sum |x_n|$ is convergent. A series is conditionally convergent if it is convergent but not absolutely convergent.

Definition A.84. Regrouping is inserting parentheses that combines finite number of terms while rearrangement is changing the order of terms.

Remark A.85. Regrouping does not affect convergence. Rearrangement does not affect convergence of absolutely convergent series.

Theorem A.86. (Limit Comparison Test) Let (x_n) and (y_n) be nonzero sequences and suppose $r = \lim \left| \frac{x_n}{y_n} \right|$ exists. If $r \neq 0$, then $\sum x_n$ is absolutely convergent if and only if $\sum y_n$ is absolutely convergent. If $r = 0$ and $\sum y_n$ is absolutely convergent, then $\sum x_n$ is absolutely convergent.

Theorem A.87. (Root Test) Let (x_n) be a sequence and suppose $r = \lim |x_n|^{1/n}$ exists. $\sum x_n$ is absolutely convergent if $r < 1$ and divergent if $r > 1$.

Theorem A.88. (Ratio Test) Let (x_n) be a nonzero sequence and suppose $r = \lim \left| \frac{x_{n+1}}{x_n} \right|$ exists. $\sum x_n$ is absolutely convergent if $r < 1$ and divergent if $r > 1$.

Theorem A.89. (Integral Test) Let $f(t)$ be a positive and decreasing function on $[1, \infty)$. Then $\sum f(k)$ converges if and only if the improper integral $\int_1^\infty f(t)dt$ exists.

Theorem A.90. (Raabe's Test) Let (x_n) be a nonzero sequence. Let $a \in \mathbb{R}$ and $k \in \mathbb{N}$.

1. If there exist $a > 1$ such that $\left| \frac{x_{n+1}}{x_n} \right| \leq 1 - \frac{a}{n}$ for $n \geq k$, then $\sum x_n$ is absolutely convergent.
2. If there exist $a \leq 1$ such that $\left| \frac{x_{n+1}}{x_n} \right| \geq 1 - \frac{a}{n}$ for $n \geq k$, then $\sum x_n$ is not absolutely convergent.

Theorem A.91. (Alternating Series Test) Let (x_n) be a decreasing sequence of strictly positive numbers with $\lim x_n = 0$. Then, the alternating series $\sum (-1)^n x_n$ is convergent.

Lemma A.92. (Abel's Lemma) Let (x_n) and (y_n) be sequences. Let the partial sums of $\sum y_n$ be denoted by s_n with $s_0 = 0$. If $m > n$, then

$$\sum_{k=n+1}^m x_k y_k = (x_m s_m - x_{n+1} s_n) + \sum_{k=n+1}^{m-1} (x_k - x_{k+1}) s_k. \quad (\text{A.25})$$

Theorem A.93. (Dirichlet's Test) If (x_n) is a decreasing sequence with $\lim x_n = 0$ and $\sum y_n$ is bounded, then $\sum x_n y_n$ is convergent.

Theorem A.94. (Abel's Test) If (x_n) is a convergent monotone sequence and $\sum y_n$ is convergent, then $\sum x_n y_n$ is convergent.

Definition A.95. If the partial sum s_n of the sequence (f_n) is convergent to f on $\mathcal{D} \subseteq \mathbb{R}$, $\sum f_n$ is convergent on \mathcal{D} .

Theorem A.96. (Cauchy Criterion) Let $\mathcal{D} \subseteq \mathbb{R}$ and $(f_n : \mathcal{D} \rightarrow \mathbb{R})$ be a sequence of functions. The series $\sum f_n$ is uniformly convergent on D if and only if, for any $\epsilon > 0$, there exists N such that if $m > n > N$, then $|f_{n+1}(x) + \cdots + f_m(x)| < \epsilon$ for all $x \in \mathcal{D}$.

Theorem A.97. (Weierstrass M-Test) Let M_n be a sequence of positive real numbers such that $|f_n(x)| \leq M_n$ for $x \in \mathcal{D}$ and $n \in \mathbb{N}$. If the series $\sum M_n$ is convergent, then $\sum f_n$ is uniformly convergent on \mathcal{D} .

Theorem A.98. Let $\sum a_n x^n$ be a power series. Let $r = 1/\limsup(|a_n|^{1/n})$ be the radius of convergence of the power series. The series is absolutely and uniformly convergent if $|x| < r$ and divergent if $|x| > r$. Differentiation and integration do not change r . If $\sum a_n x^n$ and $\sum b_n x^n$ converge to the same function on the same interval, then $a_n = b_n$ for all $n \in \mathbb{N}$.

Definition A.99. The topological space is a set of open sets.

Definition A.100. \mathbb{X} is a metric space if there exists a metric $d : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}_+$ such that for $x, y, z \in \mathbb{X}$,

1. $d(x, x) = 0$.
2. $d(x, y) = d(y, x)$.
3. $d(x, y) \leq d(x, z) + d(z, y)$.

Definition A.101. Let X be a set and let $x, y, z \in \mathbb{X}$. A triple $(\mathbb{X}, +, \cdot)$ where $+: \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{X}, (x, y) \mapsto x + y$ and $\cdot: \mathbb{R} \rightarrow \mathbb{X}$ is a linear space if $(\mathbb{X}, +)$ is a commutative group and

1. $\alpha(\beta x) = (\alpha\beta)x$.
2. $1x = x$.
3. $\alpha(x + y) = \alpha x + \alpha y$ and $(\alpha + \beta)x = \alpha x + \beta x$.

Definition A.102. Let \mathbb{X} be linear space. Let $x, y \in \mathbb{X}$ and $\alpha \in \mathbb{R}$. $(\mathbb{X}, \|\cdot\|)$ where the norm $\|\cdot\|: \mathbb{X} \rightarrow \mathbb{R}, x \mapsto \|x\|$ is a normed space if

1. $\|x\| = 0 \iff x = \theta$.
2. $\|x\| \geq 0$.
3. $\|\alpha x\| = |\alpha| \|x\|$.
4. $\|x + y\| \leq \|x\| + \|y\|$.

Remark A.103. If only $\|\theta\| = 0$ holds, then $\|\cdot\|$ is a semi-norm. \mathbb{X} is a Banach space if every Cauchy sequence in \mathbb{X} is convergent.

Definition A.104. Given any measure space $(\mathbb{X}, \mathcal{S}, \mu)$ and any $p \in \mathbb{R}_{++}$, the space $\mathcal{L}^p(\mathbb{X}, \mathcal{S}, \mu)$ consists of all measurable functions $f: \mathbb{X} \rightarrow \mathbb{R}$ such that

$$\int_{\mathbb{X}} |f|^p d\mu < \infty. \quad (\text{A.26})$$

Remark A.105. \mathcal{L}^1 is the space of absolutely convergent sequences. \mathcal{L}^2 is the space of square-summable sequences or Hilbert space. If $p = \infty$, the space of bounded sequences.

Definition A.106. A measurable function is essentially bounded if $|f| \leq M$ μ -a.e. for some finite constant $M \in \mathbb{R}_{++}$. The smallest bound, called L^∞ -norm of f , is defined as

$$\|f\|_\infty = \inf\{M \in \mathbb{R}_{++} : |f| \leq M \text{ } \mu\text{-a.e.}\}. \quad (\text{A.27})$$

Remark A.107. The space \mathcal{L}^∞ consists of all measurable functions that are essentially bounded.

Theorem A.108. Let $(\mathbb{X}, \mathcal{S}, \mu)$ be a measure space and $f \in \bigcap_{1 \leq p \leq \infty} \mathcal{L}^p(\mu)$. Then,

$$\|f\|_\infty = \lim_{p \rightarrow \infty} \left(\int_{\mathbb{X}} |f|^p d\mu \right)^{1/p}. \quad (\text{A.28})$$

Theorem A.109. (Completeness of L^p) Given a measure space $(\mathbb{X}, \mathcal{S}, \mu)$ and $p \in [1, \infty]$, \mathcal{L}^p is a Banach space with the norm $\|\cdot\|_p$. If \mathbb{X} is a complete separable metric space (a Polish space) and \mathcal{S} is a Borel σ -field on \mathbb{X} , then $\mathcal{L}^p(\Omega, \mathcal{F}, P)$ is separable.

Theorem A.110. (*Hölder's Inequality*) Let $p, q \in [1, \infty]$ with $1/p + 1/q = 1$. Let $f \in \mathcal{L}^p(\mathbb{X}, \mathcal{S}, \mu)$ and $g \in \mathcal{L}^q(\mathbb{X}, \mathcal{S}, \mu)$. Then, $fg \in \mathcal{L}^1(\mathbb{X}, \mathcal{S}, \mu)$ and

$$\left| \int_{\mathbb{X}} fg \, d\mu \right| \leq \|fg\|_1 \leq \|f\|_p \|g\|_q. \quad (\text{A.29})$$

Remark A.111. The inequality becomes equality if and only if $\|f\|_p$ and $\|g\|_q$ are linearly dependent. The Cauchy-Bunyakovsky-Schwarz inequality is a special case with $p = q = 2$.

Theorem A.112. (*Hardy's Inequality*)

$$\sum_{n=1}^{\infty} \left(\frac{a_1 + a_2 + \cdots + a_n}{n} \right)^p \leq \left(\frac{p}{p-1} \right)^p \sum_{n=1}^{\infty} a_n \quad (\text{A.30})$$

where $\left(\frac{p}{p-1} \right)^p$ is best possible.

Proof. Let $A_k = \frac{a_1 + a_2 + \cdots + a_k}{k}$.

$$\begin{aligned} \sum_{k=1}^n A_k^{p-1} a_k &= \sum_{k=1}^n A_k^{p-1} (kA_k - (k-1)A_{k-1}) = \sum_{k=1}^n kA_k^p - \sum_{k=1}^n (k-1)A_k^{p-1} A_{k-1} \\ &\geq \sum_{k=1}^n kA_k^p - \sum_{k=1}^n (k-1) \left(\frac{p-1}{p} A_k^p + \frac{1}{p} A_{k-1}^p \right) \\ &= \frac{p-1}{p} \sum_{k=1}^{n-1} A_k^p + \left(n - \frac{(p-1)(n-1)}{p} \right) A_n^p \geq \frac{p-1}{p} \sum_{k=1}^n A_k^p. \end{aligned} \quad (\text{A.31})$$

Thus,

$$\sum_{k=1}^n A_k^p \leq \frac{p}{p-1} \sum_{k=1}^n A_k^{p-1} a_k \leq \frac{p}{p-1} \left(\sum_{k=1}^n A_k^p \right)^{\frac{p-1}{p}} \left(\sum_{k=1}^n a_k^p \right)^{\frac{1}{p}}. \quad (\text{A.32})$$

Let $a_k = k^{1-1/p} - (k-1)^{1-1/p}$. Then, $A_k = k^{-1/p}$.

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n A_k^p}{\sum_{k=1}^n a_k^p} &= \lim_{n \rightarrow \infty} \frac{A_n^p}{a_n^p} = \lim_{n \rightarrow \infty} \left(\frac{n^{-1/p}}{n^{1-1/p} - (n-1)^{1-1/p}} \right)^p \\ &= \lim_{n \rightarrow \infty} \left(\frac{n^{-1}}{1 - (1 - n^{-1})^{1-1/p}} \right)^p \\ &= \lim_{n \rightarrow \infty} \left(\frac{n^{-1}}{\left(1 - \frac{1}{p}\right) n^{-1} + O(n^{-2})} \right)^p = \left(\frac{p}{p-1} \right)^p. \end{aligned} \quad (\text{A.33})$$

□

Theorem A.113. (*Minkowsky-Riesz Inequality*) If $p \in [1, \infty]$ and $f, g \in \mathcal{L}^p(\mathbb{X}, \mathcal{S}, \mu)$, then $f + g \in \mathcal{L}^p(\mathbb{X}, \mathcal{S}, \mu)$ and

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Remark A.114. $\mathcal{L}^p(\mathbb{X}, \mathcal{S}, \mu)$ is a linear space and $\|\cdot\|_p$ is a seminorm for any $p \in [1, \infty]$. This inequality is triangle inequality in \mathcal{L}^p space.

Theorem A.115. (Jensen's Inequality) Let $(\Omega, \mathcal{F}, \mu)$ be a probability space. Let g be a μ -integrable real-valued function and φ be a convex function. Then,

$$\varphi\left(\int_{\Omega} g \, d\mu\right) \leq \int_{\Omega} \varphi \circ g \, d\mu.$$

Lemma A.116. (Markov Inequality) Let X be a nonnegative random variable and $a \in \mathbb{R}_{++}$. Then,

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Theorem A.117. (Bienaymé-Chebyshev Inequality) For any random variable X and $a \in \mathbb{R}_{++}$,

$$P(|X| \geq a) \leq \frac{\mathbb{E}[X^2]}{a^2}.$$

Remark A.118. If $E[|X|] < \infty$, $P(\mathbb{E}[X] - a < X < \mathbb{E}[X] + a) \geq 1 - \text{Var}(X)/a^2$.

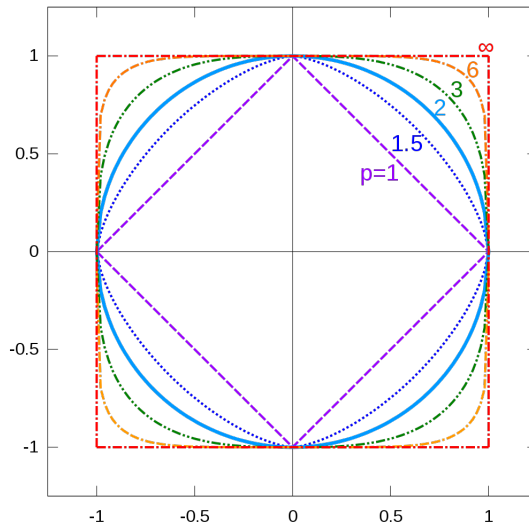
Theorem A.119. $\|X\|_p \leq \|X\|_{p(1+\epsilon)}$ for $p \in [1, \infty]$ and $\epsilon \in \mathbb{R}_+$.

Proof. $\|1\|_p = 1$. Let $X \in \mathcal{L}^0(\Omega, \mathcal{F}, P)$, $k \in [1, \infty[$, and $\epsilon \in \mathbb{R}_{++}$. By Hölder's inequality,

$$\begin{aligned} \|X\|_p &= (\mathbb{E}[|X|^p])^{1/p} \leq \left(\mathbb{E}[(|X|^p)^{1+\epsilon}]^{1/(1+\epsilon)} \cdot \mathbb{E}[1^{(1+\epsilon)/\epsilon}]^{\epsilon/(1+\epsilon)} \right)^{1/p} \\ &= \mathbb{E}[|X|^{p(1+\epsilon)}]^{1/(p(1+\epsilon))} = \|X\|_{p(1+\epsilon)}. \end{aligned} \quad (\text{A.34})$$

□

Remark A.120. $\|X\|_p$ is an increasing function of p and $\mathcal{L}^p(\Omega, \mathcal{F}, P) \subseteq \mathcal{L}^q(\Omega, \mathcal{F}, P)$ for $p \geq q \geq 1$. From the perspective of moment-generating, higher moments imply lower moments.



Definition A.121. Let X be a linear space. \mathbb{X} is a unitary space or an inner product space if there exists a scalar product $(x, y) : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}$ such that

1. $(x, x) = 0 \iff x = \theta$.
2. $(x, x) \geq 0$.
3. $(x, y) = (y, x)$.
4. $(\alpha x + y, z) = \alpha(x, z) + (y, z)$.

The vectors x and y are orthogonal if and only if $(x, y) = 0$. They are orthonormal if, in addition, $(x, x) = (y, y) = 1$.

Theorem A.122. A unitary space becomes a normed space if there exists a unitary norm $\|x\| = \sqrt{(x, x)}$.

Proof.

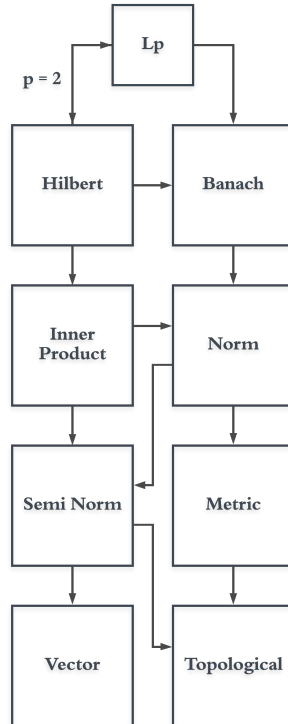
$$(x, x) \geq 0 \implies \|x\| = \sqrt{(x, x)} \geq 0 \quad (\text{A.35})$$

$$((x, x) = 0 \iff x = \theta) \implies (\|x\| = \sqrt{(x, x)} = 0 \iff x = \theta) \quad (\text{A.36})$$

$$\|\alpha x\| = \sqrt{(\alpha x, \alpha x)} = \sqrt{\alpha^2(x, x)} = |\alpha| \sqrt{(x, x)} = |\alpha| \|x\|. \quad (\text{A.37})$$

$$\|x + y\|^2 = (x + y, x + y) = (x, x) + 2(x, y) + (y, y) \leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2. \quad (\text{A.38})$$

□



Definition A.123. Let \mathbb{H} be a unitary space and $\|\cdot\|$ be the unitary norm. $(\mathbb{H}, \|\cdot\|)$ is a Hilbert space if it is a Banach space.

Definition A.124. Two functions $f(x)$ and $g(x)$ are orthogonal on the interval I with the weight function $\omega(x)$ if

$$\langle f(x)|g(x) \rangle = \int_I f(x)g(x)\omega(x) dx = 0. \quad (\text{A.39})$$

They are orthonormal if, in addition,

$$\int_I f^2(x)\omega(x) dx = \int_I g^2(x)\omega(x) dx = 1. \quad (\text{A.40})$$

Remark A.125. It can be viewed as the extension from linear vectors.

Theorem A.126. (Gram-Schmitt Orthogonalization Process) Given a finite, linearly independent functions $f_1(x), f_2(x), \dots, f_n(x)$, we can recursively derive

$$g_k(x) = \begin{cases} f_1(x), & k = 1 \\ f_k(x) - \sum_{i=1}^{k-1} h_i(x) \int_I h_i(x) f_k(x) dx, & 2 \leq k \leq n \end{cases} \quad (\text{A.41})$$

$$h_k(x) = \frac{g_k(x)}{\|g_k(x)\|} = \frac{g_k(x)}{\sqrt{\int_I g_k^2(x) dx}}. \quad (\text{A.42})$$

Then, $h_1(x), h_2(x), \dots, h_n(x)$ are orthogonal.

Lemma A.127. Let $p_n(x)$ be the orthogonal polynomials of degree n on the interval I . Then, $p_n(x)$ has n roots on I .

Definition A.128. An n -point Gaussian quadrature is a quadrature rule constructed to yield an exact result for orthogonal polynomials $p(x)$ up to the degree $2n - 1$ on the interval I by the optimal set of nodes (x_i) and their corresponding weights (λ_i) such that

$$\int_I \omega(x)f(x) dx = \sum_{i=1}^n \lambda_i f(x_i) + c(n)f^{(2n)}(\xi) \quad (\text{A.43})$$

for some c where the remainder only depends on n and $\xi \in I$ and

$$p_n(x_i) = \sum_{k=0}^n a_k x_i^k = 0, \quad \lambda_i = \int_I \omega(x) \prod_{\substack{1 \leq k \leq n \\ i \neq k}} \frac{x - x_k}{x_i - x_k} dx = \frac{a_n}{a_{n-1}} \cdot \frac{\int_I \omega(x) p_{n-1}^2(x) dx}{p'_n(x_i) p_{n-1}(x_i)}. \quad (\text{A.44})$$

Remark A.129. The remainder implies that the convergence rate of errors is $O(N^{-2N})$ if $f^{(2n)}$ does not grow exponentially. If $I = [a, b]$ and $\omega(x) = 1$, then

$$c(n) = \frac{(b-a)^{2n+1} (n!)^4}{(2n+1)((2n)!)^3}. \quad (\text{A.45})$$

I	$\omega(x)$	$p(x)$	λ_i
$(-1, 1)$	$(1-x)^\alpha(1+x)^\beta$	$P_n^{(\alpha, \beta)}(x)$	$-\frac{2n+\alpha+\beta+2}{n+\alpha+\beta+1} \cdot \frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{\Gamma(n+\alpha+\beta+1) \cdot (n+1)!} \cdot \frac{2^{\alpha+\beta}}{P_n^{(\alpha, \beta)}(x_i)P_{n+1}^{(\alpha, \beta)}(x_i)}$
$(-1, 1)$	$\frac{1}{\sqrt{1-x^2}}$	$T_n(x)$	$\frac{\pi}{n}$
$[-1, 1]$	$\sqrt{1-x^2}$	$U_n(x)$	$\frac{\pi}{n+1} \sin^2\left(\frac{i}{n+1}\pi\right)$
$[0, \infty)$	$x^\alpha e^{-x}$	$L_n^{(\alpha)}(x)$	$\frac{x_i}{(n+1)^2 L_{n+1}^2(x_i)}$ if $\alpha = 0$
$(-\infty, \infty)$	e^{-x^2}	$H_n(x)$	$\frac{2^{n-1}n!\sqrt{\pi}}{n^2 H_{n-1}^2(x_i)}$

Definition A.130. The Jacobi polynomials $P_n^{(\alpha, \beta)}(z)$ with $\alpha, \beta > -1$ can be expressed as follows. Note that they are reduced to the Legendre polynomials $P_n(z)$ if $\alpha = \beta = 0$.

$$P_0^{(\alpha, \beta)}(x) = 1. \quad P_1^{(\alpha, \beta)}(x) = (\alpha + 1) + (\alpha + \beta + 2) \frac{x-1}{2}. \quad (\text{A.46})$$

The recurrence relation of fixed α and β is

$$\begin{aligned} & 2k(k + \alpha + \beta)(2k + \alpha + \beta - 2)P_k^{(\alpha, \beta)}(x) \\ &= (2k + \alpha + \beta - 1) \left[(2k + \alpha + \beta)(2k + \alpha + \beta - 2)x + \alpha^2 - \beta^2 \right] P_{k-1}^{(\alpha, \beta)}(x) \\ & - 2(k + \alpha - 1)(k + \beta - 1)(2k + \alpha + \beta)P_{k-2}^{(\alpha, \beta)}(x). \end{aligned} \quad (\text{A.47})$$

The Rodrigues formula is

$$P_n^{(\alpha, \beta)}(x) = \frac{\frac{d^n}{dx^n} \left\{ (1-x)^\alpha (1+x)^\beta (1-x^2)^n \right\}}{2^n n! (1-x)^\alpha (1+x)^\beta}. \quad (\text{A.48})$$

Other explicit representations are

$$P_n^{(\alpha, \beta)}(x) = \frac{1}{2^n} \sum_{k=0}^n \binom{n+\alpha}{n-k} \binom{n+\beta}{k} (x-1)^k (x+1)^{n-k}. \quad (\text{A.49})$$

$$P_n^{(\alpha, \beta)}(x) = \frac{\Gamma(\alpha + n + 1)}{n! \Gamma(\alpha + \beta + n + 1)} \sum_{k=0}^n \binom{n}{k} \frac{\Gamma(\alpha + \beta + n + k + 1)}{\Gamma(\alpha + k + 1)} \left(\frac{x-1}{2} \right)^k. \quad (\text{A.50})$$

Definition A.131. The Chebyshev polynomials of the first kind $T_n(x)$ are defined by the recurrence

$$T_0(x) = 1. \quad T_1(x) = x. \quad T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x). \quad (\text{A.51})$$

The roots of $T_n(x)$ are

$$x_i = \cos \left(\frac{2i-1}{2n} \pi \right). \quad (\text{A.52})$$

The Chebyshev polynomials of the second kind $U_n(x)$ are defined by the recurrence

$$U_0(x) = 1. \quad U_1(x) = 2x. \quad U_k(x) = 2xU_{k-1}(x) - U_{k-2}(x). \quad (\text{A.53})$$

The roots of $U_n(x)$ are

$$x_i = \cos \left(\frac{i}{n+1} \pi \right). \quad (\text{A.54})$$

Theorem A.132. (Markov Inequality for Polynomials) Let $P_n(x)$ be a polynomial. For $k \in \mathbb{N}$, if $|P_n(x)| \leq 1$ for any $x \in [-1, 1]$, then

$$\left| P_n^{(k)}(x) \right| \leq \frac{n^2 (n^2 - 1^2) (n^2 - 2^2) \cdots (n^2 - (k-1)^2)}{1 \cdot 3 \cdot 5 \cdots (2k-1)}. \quad (\text{A.55})$$

The inequality holds if $|P(\cos \frac{k\pi}{n})| \leq 1$ for all $0 \leq k \leq n$.

Theorem A.133. (Bernstein Inequality) If $P_n \in \mathbb{R}$, then $\|P_n'\|_\infty \leq n^2 \|P\|_\infty$ where $\|P\|_\infty = \max_{|x| \leq 1} |P_n(x)|$. If $P_n \in \mathbb{C}$, then $\|P_n'\|_\infty \leq n \|P\|_\infty$ where $\|P\|_\infty = \max_{|z|=1} |P_n(z)|$.

Theorem A.134. If $P_n(z)$ has no zeros in $|z| < R$, then

1. ($R \geq 1$)

$$\|P_n^{(k)}\|_\infty \leq \frac{n(n-1) \cdots (n-k+1)}{1+R^k} \left(\|P_n\|_\infty - \min_{|z|=R} |P_n(z)| \right). \quad (\text{A.56})$$

2. ($R \leq 1$)

$$\|P_n'\|_\infty \leq \frac{n}{R^n + R^{n-1}} \|P_n\|_\infty. \quad (\text{A.57})$$

Theorem A.135. If $P_n(z)$ has all zeros in $|z| \leq R$, then

1. ($R \leq 1$)

$$\|P_n'\|_\infty \leq \frac{n}{1+R} \|P_n\|_\infty + \frac{n}{R^{n-1}(1+R)} \min_{|z|=R} |P_n(z)|. \quad (\text{A.58})$$

2. ($R \geq 1$)

$$\|P_n'\|_\infty \geq \frac{n}{1+R^n} \left(\|P_n\|_\infty + \min_{|z|=R} |P_n(z)| \right). \quad (\text{A.59})$$

Both the above inequalities are best possible. The equality is attained for $P_n(z) = (z+R)^n$ in the first case and $P_n(z) = z^n + R^n$ in the second case.

Definition A.136. The generalized Laguerre polynomials $L_n^{(\alpha)}(x)$ with $\alpha > -1$ can be expressed as follows.

$$L_0^{(\alpha)}(x) = 1, \quad L_1^{(\alpha)}(x) = \alpha + 1 - x. \quad (\text{A.60})$$

The recurrence relation of fixed α is

$$L_k^{(\alpha)}(x) = \frac{(2k + \alpha - 1 - x)L_{k-1}^{(\alpha)}(x) - (k + \alpha - 1)L_{k-2}^{(\alpha)}(x)}{k}. \quad (\text{A.61})$$

The Rodrigues formula is

$$L_n^{(\alpha)}(x) = \frac{x^{-\alpha} e^x}{n!} \cdot \frac{d^n}{dx^n} (e^{-x} x^{n+\alpha}). \quad (\text{A.62})$$

The closed form derived from the general Leibniz rule of the Rodrigues formula is

$$L_n^{(\alpha)}(x) = \sum_{k=0}^n (-1)^k \binom{n+\alpha}{n-k} \frac{x^k}{k!}. \quad (\text{A.63})$$

Definition A.137. The (physicists') Hermite polynomials $H_n(x)$ can be expressed as follows. The recurrence relation is

$$H_0(x) = 1. \quad H_k(x) = 2xH_{k-1}(x) - H'_k(x). \quad (\text{A.64})$$

The Rodrigues formula is

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2}. \quad (\text{A.65})$$

The explicit representation is

$$H_n(x) = n! \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(-1)^k}{k!(n-2k)!} (2x)^{n-2k}. \quad (\text{A.66})$$

Definition A.138. The optimization is referred as the maximization

$$\max_{x_1, x_2, \dots, x_n} \{f(x) \mid l_i \leq x_i \leq u_i, g(x) \leq c, h(x) = d\} \quad (\text{A.67})$$

or the minimization

$$\min_{x_1, x_2, \dots, x_n} \{f(x) \mid l_i \leq x_i \leq u_i, g(x) \geq c, h(x) = d\} \quad (\text{A.68})$$

Note that maximizing $f(x)$ is equivalent to minimizing $-f(x)$. $f(x)$ is the objective function. $g(x)$ and $h(x)$ are equal and unequal constraints respectively. The interval $[l_i, u_i]$ is a bound for each x_i . Every first partial derivative $\frac{\partial f}{\partial x_i}$ is a gradient while every second partial derivative $\frac{\partial^2 f}{\partial x_i \partial x_j}$ is a Hessian.

Theorem A.139. If $x^* = (x_1^*, x_2^*, \dots, x_n^*)$ is a local optimized point, then $\frac{\partial f}{\partial x_i^*} = 0$. Also, $\frac{\partial^2 f}{\partial x_i^* \partial x_j^*} < 0$ (or negative definite Hessian) for the local maximized point and $\frac{\partial^2 f}{\partial x_i^* \partial x_j^*} > 0$ (or positive definite Hessian) for the local minimized point.

Remark A.140. A local maximum is also a global maximum for concave functions. A local minimum is also a global minimum for convex functions.

SECTION B

Univariate Distributions

Distribution	Mean	Second Moment	Variance
Uniform (Discrete)	$\frac{a+b}{2}$	—	$\frac{(b-a+1)^2-1}{12}$
Binomial	np	—	$np(1-p)$
Geometric	$\frac{1-p}{p}$	—	$\frac{1-p}{p^2}$
Hypergeometric	$\frac{nK}{N}$	—	$\frac{nK}{N} \frac{N-K}{N} \frac{N-n}{N-1}$
Poisson	λ	—	λ
Uniform (Continuous)	$\frac{a+b}{2}$	$\frac{a^2+ab+b^2}{3}$	$\frac{(b-a)^2}{12}$
Exponential	$\frac{1}{c}$	$\frac{2}{c^2}$	$\frac{1}{c^2}$
Normal	a	$a^2 + \sigma^2$	σ^2
Cauchy	NA	NA	NA
Gamma	$\frac{p}{c}$	$\frac{p(p+1)}{c^2}$	$\frac{p}{c^2}$
Inverse Gamma	$\frac{c}{p-1}$	$\frac{c^2}{(p-1)(p-2)}$	$\frac{c^2}{(p-1)^2(p-2)}$
Beta	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Noncentral Chi-Square	$d + \lambda$	$d^2 + 2d + 2d\lambda + \lambda^2 + 4\lambda$	$2d + 4\lambda$

B.1 Error Function

$$\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^x e^{-t^2} dt. \quad (\text{B.1})$$

$$\operatorname{erf}(0) = 0. \quad \operatorname{erf}(\infty) = 1. \quad \operatorname{erf}(-\infty) = -1. \quad (\text{B.2})$$

B.2 Gamma Function

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx. \quad (\text{B.3})$$

$$\Gamma(1) = \int_0^\infty e^{-x} dx = -e^{-x} \Big|_0^\infty = 1. \quad (\text{B.4})$$

$$\Gamma(z+1) = \int_0^\infty x^z e^{-x} dx = (-x^{-z} e^{-x}) \Big|_0^\infty + z \int_0^\infty x^{z-1} e^{-x} dx = z\Gamma(z). \quad (\text{B.5})$$

B.3 Discrete Uniform Distribution

$$P(X = k) = \frac{1}{n+1}, \quad n \in \mathbb{N}, \quad k \in \{0, 1, \dots, n\}. \quad (\text{B.6})$$

B.4 Binomial Distribution

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad n \in \mathbb{N}_{++}, \quad p \in]0, 1[, \quad k \in \{0, 1, \dots, n\}. \quad (\text{B.7})$$

B.5 Geometric Distribution

$$P(X = k) = p(1-p)^k, \quad p \in]0, 1[, \quad k \in \mathbb{N}. \quad (\text{B.8})$$

B.6 Hypergeometric Distribution

$$P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}, \quad N \in \mathbb{N}, \quad n, K \in \mathbb{N}_{|N} \quad k \in \{\max(0, n+K-N), \dots, \min(n, K)\}. \quad (\text{B.9})$$

B.7 Poisson Distribution

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad \lambda \in \mathbb{R}_{++}, \quad k \in \mathbb{N}. \quad (\text{B.10})$$

B.8 Uniform Distribution

$$\mathbb{E}[X] = \int_a^b \frac{x}{b-a} dx = \frac{x^2}{2(b-a)} \Big|_a^b = \frac{a+b}{2}. \quad (\text{B.11})$$

$$\mathbb{E}[X^2] = \int_a^b \frac{x^2}{b-a} dx = \frac{x^3}{3(b-a)} \Big|_a^b = \frac{a^2 + ab + b^2}{3}. \quad (\text{B.12})$$

$$\text{Var}(X) = \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}. \quad (\text{B.13})$$

B.9 Exponential Distribution

$$\mathbb{E}[X] = \int_0^\infty x \cdot ce^{-cx} dx = -\frac{(cx+1)e^{-cx}}{c} \Big|_0^\infty = \frac{1}{c}. \quad (\text{B.14})$$

$$\mathbb{E}[X^2] = \int_0^\infty x^2 \cdot ce^{-cx} dx = -\frac{(c^2x^2 + 2cx + 2)e^{-cx}}{c^2} \Big|_0^\infty = \frac{2}{c^2}. \quad (\text{B.15})$$

$$\text{Var}(X) = \frac{2}{c^2} - \frac{1}{c^2} = \frac{1}{c^2}. \quad (\text{B.16})$$

B.10 Normal Distribution

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \left(\frac{a \cdot \operatorname{erf}\left(\frac{x-a}{\sqrt{2}\sigma}\right)}{2} - \frac{\sigma e^{-\frac{(x-a)^2}{2\sigma^2}}}{\sqrt{2\pi}} \right) \Bigg|_{-\infty}^{\infty} = a. \quad (\text{B.17})$$

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} \frac{x^2}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \left(\frac{(a^2 + \sigma^2) \cdot \operatorname{erf}\left(\frac{x-a}{\sqrt{2}\sigma}\right)}{2} - \frac{\sigma(x+a)e^{-\frac{(x-a)^2}{2\sigma^2}}}{\sqrt{2\pi}} \right) \Bigg|_{-\infty}^{\infty} = a^2 + \sigma^2. \quad (\text{B.18})$$

$$\operatorname{Var}(X) = (a^2 + \sigma^2) - a^2 = \sigma^2. \quad (\text{B.19})$$

B.11 Cauchy Distribution

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)} = \frac{\ln(1+x^2)}{2\pi} \Bigg|_{-\infty}^{\infty} = 0. \quad (\text{B.20})$$

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} \frac{x^2}{\pi(1+x^2)} = \pi(x - \arctan x) \Big|_{-\infty}^{\infty} = \infty. \quad (\text{B.21})$$

$$\operatorname{Var}(X) = \infty - 0 = \infty. \quad (\text{B.22})$$

B.12 Gamma Distribution

$$\mathbb{E}[X] = \int_0^{\infty} x \cdot \frac{c^p}{\Gamma(p)} x^{p-1} e^{-cx} dx = \frac{p}{c} \int_0^{\infty} \frac{c^{p+1}}{\Gamma(p+1)} x^{(p+1)-1} e^{-cx} dx = \frac{p}{c}. \quad (\text{B.23})$$

$$\Gamma(p+1) = p \cdot \Gamma(p). \quad (\text{B.24})$$

$$P(\Omega_{p+1,c}) = \int_0^{\infty} \frac{c^{p+1}}{\Gamma(p+1)} x^{(p+1)-1} e^{-cx} dx = 1. \quad (\text{B.25})$$

$$\mathbb{E}[X^2] = \int_0^{\infty} x^2 \cdot \frac{c^p}{\Gamma(p)} x^{p-1} e^{-cx} dx = \frac{p(p+1)}{c^2} \int_0^{\infty} \frac{c^{p+2}}{\Gamma(p+2)} x^{(p+2)-1} e^{-cx} dx = \frac{p(p+1)}{c^2}. \quad (\text{B.26})$$

$$\Gamma(p+2) = (p+1) \cdot \Gamma(p+1) = (p+1) \cdot p \cdot \Gamma(p). \quad (\text{B.27})$$

$$P(\Omega_{p+2,c}) = \int_0^{\infty} \frac{c^{p+2}}{\Gamma(p+2)} x^{(p+2)-1} e^{-cx} dx = 1. \quad (\text{B.28})$$

$$\operatorname{Var}(X) = \frac{p(p+1)}{c^2} - \left(\frac{p}{c}\right)^2 = \frac{p}{c^2}. \quad (\text{B.29})$$

B.13 Inverse Gamma Distribution

$$\mathbb{E}[X] = \int_0^\infty x \cdot \frac{c^p}{\Gamma(p)} x^{-p-1} e^{-\frac{c}{x}} dx = \frac{c}{p-1} \int_0^\infty \frac{c^{p-1}}{\Gamma(p-1)} x^{-(p-1)-1} e^{-\frac{c}{x}} dx = \frac{c}{p-1}. \quad (\text{B.30})$$

$$\Gamma(p) = (p-1) \cdot \Gamma(p-1). \quad (\text{B.31})$$

$$P(\Omega_{p-1,c}) = \int_0^\infty \frac{c^{p-1}}{\Gamma(p-1)} x^{-(p-1)-1} e^{-\frac{c}{x}} dx = 1. \quad (\text{B.32})$$

$$\begin{aligned} \mathbb{E}[X^2] &= \int_0^\infty x^2 \cdot \frac{c^p}{\Gamma(p)} x^{-p-1} e^{-\frac{c}{x}} dx = \frac{c^2}{(p-1)(p-2)} \int_0^\infty \frac{c^{p-2}}{\Gamma(p-2)} x^{-(p-2)-1} e^{-\frac{c}{x}} dx \\ &= \frac{c^2}{(p-1)(p-2)}. \end{aligned} \quad (\text{B.33})$$

$$\Gamma(p) = (p-1) \cdot \Gamma(p-1) = (p-1) \cdot (p-2) \cdot \Gamma(p-2). \quad (\text{B.34})$$

$$P(\Omega_{p-2,c}) = \int_0^\infty \frac{c^{p-2}}{\Gamma(p-2)} x^{-(p-2)-1} e^{-\frac{c}{x}} dx = 1. \quad (\text{B.35})$$

$$\text{Var}(X) = \frac{c^2}{(p-1)(p-2)} - \left(\frac{c}{p-1} \right)^2 = \frac{c^2}{(p-1)^2(p-2)}. \quad (\text{B.36})$$

B.14 Beta Distribution

$$\begin{aligned} \mathbb{E}[X] &= \int_0^1 x \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} dx \\ &= \frac{\alpha}{\alpha+\beta} \int_0^1 \frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+1)\Gamma(\beta)} x^{(\alpha+1)-1} (1-x)^{\beta-1} dx = \frac{\alpha}{\alpha+\beta}. \end{aligned} \quad (\text{B.37})$$

$$\Gamma(\alpha+\beta+1) = (\alpha+\beta) \cdot \Gamma(\alpha+\beta). \quad (\text{B.38})$$

$$P(\Omega_{\alpha+1,\beta}) = \int_0^1 \frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+1)\Gamma(\beta)} x^{(\alpha+1)-1} (1-x)^{\beta-1} dx = 1. \quad (\text{B.39})$$

$$\begin{aligned} \mathbb{E}[X^2] &= \int_0^1 x^2 \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} dx \\ &= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} \int_0^1 \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+2)\Gamma(\beta)} x^{(\alpha+2)-1} (1-x)^{\beta-1} dx \\ &= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}. \end{aligned} \quad (\text{B.40})$$

$$\Gamma(\alpha+\beta+2) = (\alpha+\beta+1) \cdot \Gamma(\alpha+\beta+1) = (\alpha+\beta+1) \cdot (\alpha+\beta) \cdot \Gamma(\alpha+\beta) \quad (\text{B.41})$$

$$P(\Omega_{\alpha+2,\beta}) = \int_0^1 \frac{\Gamma(\alpha + \beta + 2)}{\Gamma(\alpha + 2)\Gamma(\beta)} x^{(\alpha+2)-1} (1-x)^{\beta-1} dx = 1. \quad (\text{B.42})$$

$$\text{Var}(X) = \frac{\alpha(\alpha + 1)}{\alpha + \beta(\alpha + \beta + 1)} - \left(\frac{\alpha}{\alpha + \beta} \right)^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}. \quad (\text{B.43})$$

B.15 Noncentral Chi-Square Distribution

Let $Z \sim N(0, 1)$ and $X^2 \sim (Z + \mu)^2$. X^2 is the noncentral chi-square distribution with the degree of freedom 1.

$$\begin{aligned} \mathbb{E}[e^{t(z+\mu)^2}] &= \int_{-\infty}^{\infty} e^{t(z+\mu)^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= e^{\frac{\mu^2 t}{1-2t}} (1-2t)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}(1-2t)^{-\frac{1}{2}}} e^{-\frac{(z - \frac{2\mu t}{1-2t})^2}{2(1-2t)^{-1}}} dz = e^{\frac{\mu^2 t}{1-2t}} (1-2t)^{-\frac{1}{2}}. \end{aligned} \quad (\text{B.44})$$

Since the noncentral chi-square distribution with the degree of freedom d is the product of d noncentral chi-square distributions with the degree of freedom 1,

$$\begin{aligned} \mathbb{E}[e^{t\chi_{d,\mu}^2}] &= \mathbb{E}\left[\prod_{i=1}^d e^{t(z_i+\mu_i)^2}\right] = \prod_{i=1}^d \mathbb{E}\left[e^{t(z_i+\mu_i)^2}\right] \\ &= \prod_{i=1}^d (1-2t)^{-\frac{1}{2}} e^{\frac{\mu_i^2 t}{1-2t}} = (1-2t)^{-\frac{d}{2}} e^{\frac{t}{1-2t} \sum_{i=1}^d \mu_i^2} = (1-2t)^{-\frac{d}{2}} e^{\frac{\lambda t}{1-2t}}. \end{aligned} \quad (\text{B.45})$$

Then, we generate the first moment and the second moment at $t = 0$ which are equivalent to $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$ respectively.

$$\mathbb{E}[X] = \frac{d}{dt} \mathbb{E}\left[e^{t\chi_{d,\mu}^2}\right] \Big|_{t=0} = e^{\frac{\lambda t}{1-2t}} \left(d \cdot (1-2t)^{-\frac{d}{2}-1} + \lambda \cdot (1-2t)^{-\frac{d}{2}-2} \right) \Big|_{t=0} = d + \lambda. \quad (\text{B.46})$$

$$\begin{aligned} \mathbb{E}[X^2] &= \frac{d}{dt^2} \mathbb{E}\left[e^{t\chi_{d,\mu}^2}\right] \Big|_{t=0} \\ &= e^{\frac{\lambda t}{1-2t}} \left((d^2 + 2d) \cdot (1-2t)^{-\frac{d}{2}-2} + (2d\lambda + 4\lambda) \cdot (1-2t)^{-\frac{d}{2}-3} + \lambda^2 \cdot (1-2t)^{-\frac{d}{2}-4} \right) \Big|_{t=0} \\ &= d^2 + 2d(\lambda + 1) + \lambda(\lambda + 4). \end{aligned} \quad (\text{B.47})$$

$$\text{Var}(X) = d^2 + 2d(\lambda + 1) + \lambda(\lambda + 4) - (d + \lambda)^2 = 2d + 4\lambda. \quad (\text{B.48})$$

SECTION C

Multivariate Distributions

Definition C.1. A function $C : [0, 1]^d \rightarrow [0, 1]$ is a copula if it is a joint distribution function of a vector of d uniform random variables (U_1, U_2, \dots, U_d) with $U_i \in [0, 1]$.

$$C(u_1, u_2, \dots, u_d) = P(U_1 \leq u_1, U_2 \leq u_2, \dots, U_d \leq u_d). \quad (\text{C.1})$$

For $i \in \{1, 2, \dots, d\}$,

$$C(1, \dots, 1, u_i, 1, \dots, 1) = u_i. \quad (\text{C.2})$$

Remark C.2. For independent random variables, $C(u_1, u_2, \dots, u_d) = u_1 u_2 \dots u_d$.

Theorem C.3. (Sklar's Theorem) Let F be a joint distribution function with marginals F_1, F_2, \dots, F_d that are not necessarily uniform. Then, there exists a copula $C : [0, 1]^d \rightarrow [0, 1]$ such that

$$C(F_1(x_1), F_2(x_2), \dots, F_d(x_d)) = F(x_1, x_2, \dots, x_d). \quad (\text{C.3})$$

If the marginals are continuous, then C is unique. If C is a copula function and F_1, F_2, \dots, F_d are univariate distribution functions, then the function $H : \mathbb{R}^d \rightarrow [0, 1]$ defined by

$$H(x_1, x_2, \dots, x_d) = C(F_1(x_1), F_2(x_2), \dots, F_d(x_d)), \quad (\text{C.4})$$

is a joint distribution function with marginals F_1, F_2, \dots, F_d .

Definition C.4. Let \mathcal{N}_d be the d -dimensional cumulative distribution function for a standard normal distribution function with corresponding marginal \mathcal{N} , and covariance matrix Σ . The Gaussian copula is defined by

$$C(u_1, u_2, \dots, u_d) = \mathcal{N}_d(\mathcal{N}^{-1}(u_1), \mathcal{N}^{-1}(u_2), \dots, \mathcal{N}^{-1}(u_d)) = \frac{1}{\sqrt{\det \Sigma}} \exp \left(-\frac{1}{2} V(\Sigma - I)V^T \right) \quad (\text{C.5})$$

where $V = (\mathcal{N}^{-1}(u_1), \mathcal{N}^{-1}(u_2), \dots, \mathcal{N}^{-1}(u_d))$.

Definition C.5. An Archimedean copula $C : [0, 1]^d \rightarrow [0, 1]$ is a copula that can be represented in the form

$$C(u_1, \dots, u_d) = \phi^{-1} \left(\sum_{i=1}^d \phi(u_i) \right) \quad (\text{C.6})$$

for a continuous, strictly decreasing, convex function $\phi : [0, 1] \rightarrow [0, \infty)$ satisfying $\phi(1) = 0$ and $\phi(0_+) = \infty$. The function ϕ is called the generator of the copula.

SECTION D

The Formula of PI

D.1 Archimedes

$$a_0 = 2\sqrt{3}. \quad b_0 = 3. \quad a_{n+1} = \frac{2a_nb_n}{a_n + b_n}. \quad b_{n+1} = \sqrt{a_{n+1}b_n}. \quad \pi = \lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n. \quad (\text{D.1})$$

D.2 François Viète

$$k_0 = \sqrt{\frac{1}{2}}. \quad k_{n+1} = \sqrt{\frac{1}{2} + \frac{1}{2}k_n}. \quad \frac{2}{\pi} = \prod_{n=0}^{\infty} k_n. \quad (\text{D.2})$$

D.3 John Wallis

$$\frac{4}{\pi} = \prod_{n=1}^{\infty} \frac{(2n+1)(2n+1)}{2n \cdot (2n+2)}. \quad (\text{D.3})$$

D.4 James Gregory

$$\frac{\pi}{4} = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1}. \quad (\text{D.4})$$

D.5 Leonhard Euler

$$\frac{\pi^2}{6} = \sum_{n=1}^{\infty} \frac{1}{n^2}. \quad (\text{D.5})$$

D.6 James Glaisher

$$\frac{\pi}{2} = \sum_{n=0}^{\infty} \frac{\binom{2n}{n}}{2^{2n}(2n+1)} = \sum_{n=1}^{\infty} \frac{2^n}{\binom{2n}{n}n} = \sum_{n=2}^{\infty} \frac{2^n}{\binom{2n}{n}}. \quad (\text{D.6})$$

$$\frac{2\pi}{3\sqrt{3}} - 1 = \sum_{n=1}^{\infty} \frac{1}{\binom{2n}{n}(2n+1)}. \quad (\text{D.7})$$

$$\frac{2\pi}{9\sqrt{3}} + \frac{1}{3} = \sum_{n=1}^{\infty} \frac{1}{\binom{2n}{n}}. \quad (\text{D.8})$$

$$1 - \frac{4\pi}{9\sqrt{3}} = \sum_{n=2}^{\infty} \frac{2n}{\binom{2n}{n}(2n+1)}. \quad (\text{D.9})$$

$$\frac{\pi^2}{8} = \sum_{n=1}^{\infty} \frac{2^n}{\binom{2n}{n}n^2}. \quad (\text{D.10})$$

$$\frac{\pi^2}{4} = \sum_{n=2}^{\infty} \frac{2^{2n-1}}{\binom{2n}{n} n^2}. \quad (\text{D.11})$$

$$\frac{\pi^2}{18} - \frac{1}{2} = \sum_{n=2}^{\infty} \frac{1}{\binom{2n}{n} (2n+1)(2n+2)}. \quad (\text{D.12})$$

$$\frac{\pi^2}{18} + \frac{2\pi}{\sqrt{3}} - \frac{25}{6} = \sum_{n=2}^{\infty} \frac{1}{\binom{2n}{n} (2n+1)(2n+2)(2n+3)}. \quad (\text{D.13})$$

$$\frac{\pi^2}{8} + \pi - \frac{13}{3} = \sum_{n=2}^{\infty} \frac{2^n}{\binom{2n}{n} (n-1)^2 n}. \quad (\text{D.14})$$

D.7 Kurushima Kinai

$$k_0 = 2. \quad k_{n+1} = 2^{2n+1} - 2^{n+1} \sqrt{4^n - k_n}. \quad \frac{\pi}{2} = \lim_{n \rightarrow \infty} \sqrt{\frac{4}{3} k_{n+1} - \frac{1}{3} k_n}. \quad (\text{D.15})$$

D.8 David Bailey, Peter Borwein, Simon Plouffe

$$\pi = \sum_{n=0}^{\infty} \frac{1}{16^n} \left(\frac{4+8r}{8n+1} - \frac{8r}{8n+2} - \frac{4r}{8n+3} - \frac{2+8r}{8n+4} \right. \\ \left. - \frac{1+2r}{8n+5} - \frac{1+2r}{8n+6} + \frac{r}{8n+7} \right), \quad r \neq -1. \quad (\text{D.16})$$

$$\pi^2 = \frac{9}{8} \sum_{n=0}^{\infty} \frac{1}{64^n} \left(\frac{16}{(6n+1)^2} - \frac{24}{(6n+2)^2} - \frac{8}{(6n+3)^2} - \frac{6}{(6n+4)^2} + \frac{1}{(8n+5)^2} \right). \quad (\text{D.17})$$

$$\pi^2 = \sum_{n=0}^{\infty} \frac{1}{16^n} \left(\frac{16}{(8n+1)^2} - \frac{16}{(8n+2)^2} - \frac{8}{(8n+3)^2} - \frac{16}{(8n+4)^2} \right. \\ \left. - \frac{4}{(8n+5)^2} - \frac{4}{(8n+6)^2} + \frac{2}{(8n+7)^2} \right). \quad (\text{D.18})$$

D.9 Fabrice Bellard

$$\pi = \sum_{n=0}^{\infty} \frac{(-1)^n}{2^{2n-2}} \left(\frac{1}{2n+1} - \frac{1}{2^{8n+10}} \left(\frac{32}{4n+1} + \frac{8}{4n+2} + \frac{1}{4n+3} \right) \right). \quad (\text{D.19})$$

$$\pi = \sum_{n=0}^{\infty} \frac{(-1)^n}{2^{10n+6}} \left(-\frac{2^5}{4n+1} - \frac{1}{4n+3} + \frac{2^8}{10n+1} - \frac{2^6}{10n+3} \right. \\ \left. - \frac{2^2}{10n+5} - \frac{2^2}{10n+7} + \frac{1}{10n+9} \right). \quad (\text{D.20})$$

$$\pi = \frac{1}{740025} \left(\sum_{n=1}^{\infty} \frac{3p(n)}{\binom{7n}{2n} 2^{n-1}} - 20379280 \right),$$

$$p(n) = -885673181n^5 + 3125347237n^4 - 2942969225n^3 + 1031962795n^2 - 196882274n + 10996648. \quad (\text{D.21})$$

D.10 Victor Adamchik, Stan Wagon

$$\pi = \sum_{n=0}^{\infty} \frac{(-1)^n}{4^n} \left(\frac{2}{4n+1} + \frac{2}{4n+2} + \frac{1}{4n+3} \right). \quad (\text{D.22})$$

$$\pi = \sum_{n=0}^{\infty} \left(\frac{4+r}{4n+1} - \frac{3r}{4n+2} + \frac{r-4}{4n+3} + \frac{r}{4n+4} \right), \quad r \neq 0. \quad (\text{D.23})$$

$$\sqrt{2}\pi = \sum_{n=0}^{\infty} \frac{(-1)^n}{8^n} \left(\frac{4}{6n+1} + \frac{1}{6n+3} + \frac{1}{6n+5} \right). \quad (\text{D.24})$$

$$\pi^2 = \sum_{n=0}^{\infty} \frac{1}{n^3} \left(-\frac{12}{n+1} + \frac{384}{n+2} + \frac{45/2}{2n+1} - \frac{1215/2}{2n+3} \right). \quad (\text{D.25})$$

$$\pi = \sum_{n=0}^{\infty} \frac{1}{n^3} \left(-\frac{238}{n+1} + \frac{285/2}{2n+1} - \frac{667/32}{4n+1} - \frac{5103/16}{4n+3} + \frac{35625/32}{4n+5} \right). \quad (\text{D.26})$$

D.11 Kirby Urner

$$k_0 = 0, \quad h_0 = 2, \quad k_{n+1} = \sqrt{2+k_n}, \quad h_{n+1} = \sqrt{2-k_n}, \quad \pi = \sum_{n=0}^{\infty} h_n(1-k_n/2)2^n. \quad (\text{D.27})$$

D.12 John Machin

$$\frac{\pi}{4} = 4 \arctan \frac{1}{5} - \arctan \frac{1}{239}. \quad (\text{D.28})$$

D.13 Charles Hutton

$$\frac{\pi}{4} = 2 \arctan \frac{1}{3} + \arctan \frac{1}{7}. \quad (\text{D.29})$$

$$\frac{\pi}{4} = 2 \arctan \frac{1}{2} - \arctan \frac{1}{7}. \quad (\text{D.30})$$

$$\frac{\pi}{4} = \arctan \frac{1}{2} - \arctan \frac{1}{3}. \quad (\text{D.31})$$

D.14 Carl Friedrich Gauss

$$\frac{\pi}{4} = 12 \arctan \frac{1}{18} + 8 \arctan \frac{1}{57} - 5 \arctan \frac{1}{239}. \quad (\text{D.32})$$

D.15 Samuel Klingenstierna

$$\frac{\pi}{4} = 8 \arctan \frac{1}{10} - \arctan \frac{1}{239} - 4 \arctan \frac{1}{515}. \quad (\text{D.33})$$

D.16 Fredrik Carl Mülertz Størmer

$$\frac{\pi}{4} = 6 \arctan \frac{1}{8} + 2 \arctan \frac{1}{57} + \arctan \frac{1}{239}. \quad (\text{D.34})$$

$$\frac{\pi}{4} = 44 \arctan \frac{1}{57} + 7 \arctan \frac{1}{239} - 12 \arctan \frac{1}{682} + 24 \arctan \frac{1}{12943}. \quad (\text{D.35})$$

D.17 Ramanujan

$$\frac{1}{\pi} = \frac{2\sqrt{2}}{99^2} \sum_{k=0}^{\infty} \frac{(4k)!(26390k + 1103)}{(k!)^4 396^{4k}}. \quad (\text{D.36})$$

D.18 Chudonovsky

$$\frac{1}{\pi} = 12 \sum_{k=0}^{\infty} \frac{(-1)^k (6k)!(545140134k + 13591409)}{(3k)!(k!)^3 (640320)^{3k+3/2}}. \quad (\text{D.37})$$

D.19 Continued Fractions

$$\pi = 3 + \frac{1}{7+} \frac{1}{15+} \frac{1}{1+} \frac{1}{292+} \frac{1}{1+} \frac{1}{1+} \frac{1}{1+} \frac{1}{2+} \dots \quad (\text{D.38})$$

$$\frac{4}{\pi} = 1 + \frac{1^2}{2+} \frac{3^2}{2+} \frac{5^2}{2+} \frac{7^2}{2+} \dots \frac{(2n+1)^2}{2+} \dots \quad (\text{D.39})$$

$$\frac{4}{\pi} = 1 + \frac{1^2}{3+} \frac{2^2}{5+} \frac{3^2}{7+} \frac{4^2}{9+} \dots \frac{n^2}{(2n+1)+} \dots \quad (\text{D.40})$$

$$\pi = 3 + \frac{1^2}{6+} \frac{3^2}{6+} \frac{5^2}{6+} \frac{7^2}{6+} \dots \frac{(2n+1)^2}{6+} \dots \quad (\text{D.41})$$

$$\frac{\pi}{2} = 1 + \frac{2}{3+} \frac{3 \cdot 5}{4+} \frac{5 \cdot 7}{4+} \frac{7 \cdot 9}{4+} \dots \frac{(2n-1) \cdot (2n+1)}{4+} \dots \quad (\text{D.42})$$

$$\frac{\pi}{2} = \frac{1}{1+} \frac{1 \cdot 2}{1+} \frac{2 \cdot 3}{1+} \frac{3 \cdot 4}{1+} \dots \frac{n \cdot (n+1)}{1+} \dots \quad (\text{D.43})$$

- [1] A. LYASOFF. Stochastic Methods in Asset Pricing. 2017.
- [2] R. G. BARTLE. Introduction to Real Analysis. 2011.
- [3] E.REINER, M.RUBINSTEIN. Breaking down the barriers. Risk, 4:28–35, 1991.
- [4] A.CONZE, R.VISWANATHAN. Path dependent options: the case of lookback options. J. of Finance, 46:1893–1907, 1992.