Stochastic Analysis for Finance
金融随机分析学习笔记

Julique Zhang

2023 年 2 月 24 日

目录

第一章　概率空间与数学期望

1.1 Infinite Probability Spaces

1. (Example) Toss a coin infinitely many times, then a generic element of Ω_∞ will be denoted $\omega = \omega_1\omega_2\cdots$, where ω_n indicates the result of nth coin toss. The sample spaces listed like Ω_∞ are not only infinite but are *uncountably infinite*.

2. (Definition) The definition of σ-algebra or σ-field.

3. (Definition) The definition of probability measure $\mathbb{P}$ (A function, namely $\mathbb{P}:\mathcal{F}\to\mathbb{R}$) and probability space.

4. (Example) We toss a coin infinitely many times and let Ω_∞ denote the set of possible outcomes. We assume the probability of head on each toss is $p>0$, the probability of tail is $q=1-p>0$, and the different tosses are independent. We want to construct a probability measure corresponding to this random experiment.

   (a) We first define $\mathbb{P}(\varnothing)=0$ and $\mathbb{P}(\Omega)=1$. These $2^{\left(2^0\right)}=2$ sets form a σ-algebra, which we call $\mathcal{F}_0$:
   $$\mathcal{F}_0=\{\varnothing,\Omega\}$$

   (b) We next define $\mathbb{P}$ for the two sets:
   $$A_H := \{\omega:\omega_1=H\}$$
   $$A_T := \{\omega:\omega_1=T\}$$
   by setting $\mathbb{P}(A_H=p)$, $\mathbb{P}(A_T)=q$. Then we have now defined $\mathbb{P}$ for $2^{\left(2^1\right)}=4$ sets, which form a σ-algebra, which we call $\mathcal{F}_1$:
   $$\mathcal{F}_1=\{\varnothing,\Omega,A_H,A_T\}$$

   (c) We next define $\mathbb{P}$ for $A_{HH},A_{HT},A_{TH},A_{TT}$ by setting
   $$\mathbb{P}(A_{HH})=p^2,\mathbb{P}(A_{HT})=\mathbb{P}(A_{TH})=pq,\mathbb{P}(A_{TT})=q^2$$
   Based on the properties of $\mathbb{P}$, we also set
   $$\mathbb{P}(A_{HH}^c),\mathbb{P}(A_{HT}^c),\mathbb{P}(A_{TH}^c),\mathbb{P}(A_{TT}^c),\mathbb{P}(A_{HH}\cup A_{HT})$$
   $$\mathbb{P}(A_{HH}\cup A_{TH}),\mathbb{P}(A_{HH}\cup A_{TT}),\mathbb{P}(A_{HT}\cup A_{TH}),\mathbb{P}(A_{HT}\cup A_{TT}),\mathbb{P}(A_{TH}\cup A_{TT})$$

The number of events is easy to count, namely, $2^{\left(2^2\right)} = 16$. These sets form a σ-algebra, which we call $\mathcal{F}_2$:

$$\mathcal{F}_2 = \left\{ \begin{array}{c} \varnothing, \Omega, A_{HH}, A_{HT}, A_{TH}, A_{TT}, A_{HH}^c, A_{HT}^c, A_{TH}^c, A_{TT}^c, A_{HH} \cup A_{HT}, \\ A_{HH} \cup A_{TH}, A_{HH} \cup A_{TT}, A_{HT} \cup A_{TH}, A_{HT} \cup A_{TT}, A_{TH} \cup A_{TT} \end{array} \right\}$$

and $A_{HH} \cup A_{HT} = A_H, A_{TH} \cup A_{TT} = A_T$, $\mathcal{F}_2$ can be written as

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

(d) Likely, we can construct $\mathcal{F}_3$, which will have $2^{\left(2^3\right)} = 256$ sets.

(e) Finally, we will create $\mathcal{F}_\infty$. Consider the set A of sequences $\omega = \omega_1\omega_2\cdots$ for which

$$\lim_{n\to\infty} \frac{H_n\left(\omega_1\omega_2\cdots\omega_n\right)}{n} = \frac{1}{2}$$

where $H_n\left(\omega_1\omega_2\cdots\omega_n\right)$ denotes the number of Hs in the first n tosses.

In other words, A is the set of sequences of heads and tails for which the long-run average number is $\frac{1}{2}$, and $A \in \mathcal{F}_\infty$. For fixed positive integers m and n, we define the set

$$A_{n,m} = \left\{ \omega : \left| \frac{H_n\left(\omega_1\omega_2\cdots\omega_n\right)}{n} - \frac{1}{2} \right| < \frac{1}{m} \right\}$$

Based on the definition of limit, we can rewrite A in the form consisted by $A_{n,m}$:

$$A = \bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} A_{n,m}$$

The Strong Law of Large Numbers asserts that $\mathbb{P}(A) = 1$ if $p = \frac{1}{2}$, $\mathbb{P}(A) = 0$ if $p \neq \frac{1}{2}$.

5. (Definition) The definition of almost surely.

   e.g. In the example of tossing coin, the probability of $A_{HHH\cdots}$ is $\lim_{n\to\infty} p^n = 0$, in another words, we will get at least one tail *almost surely.*

1.2 Random Variables and Distributions

1. (Definition) The definition of random variable X and its distribution measure $\mu_X : \mathcal{F} \mapsto \mathcal{B}$.

2. (Remark) Some points concerning Borel set:

   (a) Every open interval is a Borel set.

   (b) Furthermore, every open set is a Borel set.

   (c) Borel σ-algebra of $\mathbb{R}$: $\mathcal{B}(\mathbb{R})$.

3. (Example) Two different random variables can have the same distribution.

   A single random variable can have two different distribution. If $X \sim U[0,1]$, then the distribution of $Y = 1 - X$ is also uniform on $[0,1]$.

$$\mu_X[a,b] = \mathbb{P}\{\omega : a \leqslant X(\omega) \leqslant b\} = \mathbb{P}[a,b] = b - a, 0 \leqslant a \leqslant b \leqslant 1$$

$$\begin{aligned}\mu_Y[a,b] &= \mathbb{P}\{\omega : a \leqslant Y(\omega) \leqslant b\} = \mathbb{P}\{\omega : a \leqslant 1 - X(\omega) \leqslant b\} \\ &= \mathbb{P}[1-b, 1-a] = b - a = \mu_X[a,b], 0 \leqslant a \leqslant b \leqslant 1\end{aligned}$$

In another way, we can define another probability measure $\tilde{\mathbb{P}}$ on $[0,1]$:

$$\tilde{\mathbb{P}}[a,b] = \int_a^b 2\omega \mathrm{d}\omega = b^2 - a^2, 0 \leqslant a \leqslant b \leqslant 1$$

Under $\tilde{\mathbb{P}}$, the random variable X no longer has the uniform distribution. Denoting the distribution measure of X under $\tilde{\mathbb{P}}$ by $\tilde{\mu}_X$, we have

$$\begin{aligned}\tilde{\mu}_X[a,b] &= \tilde{\mathbb{P}}\{\omega : a \leqslant X(\omega) \leqslant b\} = \tilde{\mathbb{P}}[a,b] = b^2 - a^2, 0 \leqslant a \leqslant b \leqslant 1 \\ \tilde{\mu}_Y[a,b] &= \tilde{\mathbb{P}}\{\omega : a \leqslant Y(\omega) \leqslant b\} = \tilde{\mathbb{P}}\{\omega : a \leqslant 1 - X(\omega) \leqslant b\} \\ &= \tilde{\mathbb{P}}[1-b, 1-a] = (1-b)^2 - (1-a)^2, 0 \leqslant a \leqslant b \leqslant 1\end{aligned}$$

4. (Definition) Cumulative distribution function: $F(x) = \mathbb{P}\{X \leqslant x\} = \mu_X(-\infty, x], x \in \mathbb{R}$.

$$\begin{aligned}\mu_X[a,b] &= \mu_X\left(\bigcap_{n=1}^{\infty}\left(a - \frac{1}{n}, b\right]\right) = \lim_{n\to\infty}\mu_X\left(a - \frac{1}{n}, b\right] \\ &= F(b) - \lim_{n\to\infty}F\left(a - \frac{1}{n}\right) = F(b) - F(a-)\end{aligned}$$

注：极限可以放在 μ_X 外，是利用了测度的下连续性。

5. (Definition) Probability Density function: a nonnegative function defined $f(x)$ for $x \in \mathbb{R}$ such that

$$\mu_X[a,b] = \mathbb{P}\{a \leqslant X \leqslant b\} = \int_a^b f(x)\,\mathrm{d}x, -\infty < a \leqslant b < \infty$$

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \lim_{n\to\infty}\int_{-n}^{n} f(x)\,\mathrm{d}x = \lim_{n\to\infty}\mathbb{P}\{-n \leqslant X \leqslant n\} = \mathbb{P}\{X \in \mathbb{R}\} = \mathbb{P}(\Omega) = 1$$

Remark: Not consider random variables that can take the value $\pm\infty$.

6. (Definition) Probability Mass Function: Define $p_i = \mathbb{P}\{X = x_i\} \geqslant 0$, we have $\sum_i p_i = 1$. The mass assigned to a Borel set $B \subset \mathbb{R}$ by the distribution measure of X is

$$\mu_X(B) = \sum_{\{i : x_i \in B\}} p_i, B \in \mathcal{B}(\mathbb{R})$$

Remark: For discrete situation.

7. (Example) A special example for uniform distribution:

$$Y_n(\omega) := \begin{cases} 1, & \text{if} \quad \omega_n = H, \\ 0, & \text{if} \quad \omega_n = T. \end{cases} \Longrightarrow X := \sum_{n=1}^{\infty}\frac{Y_n}{2^n} \sim U[0,1]$$

Proof. If n=2, then

X	0	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{3}{4}$
μ_X	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

For all $k, m \in \mathbb{N}$, if $0 \leqslant k < m \leqslant 2^n$,

$$\begin{aligned}\mu_X\left[\frac{k}{2^n}, \frac{m}{2^n}\right) &= \mu_X\left(\bigcup_{i=k}^{m-1}\left[\frac{i}{2^n}, \frac{i+1}{2^n}\right)\right) = \sum_{i=k}^{m-1}\mu_X\left[\frac{i}{2^n}, \frac{i+1}{2^n}\right) = \sum_{i=k}^{m-1}\mu_X\left\{X = \frac{i}{2^n}\right\} \\ &= \sum_{i=k}^{m-1}\frac{1}{2^n} = \frac{m-k}{2^n}\end{aligned}$$

For all $a, b \in \mathbb{R}$, if $0 \leqslant a < b \leqslant 1$,

$$\mu_X[a, b) = b - a$$

Specially, X can have the value of 1, because if $\forall n \in \mathbb{N}, Y_n = 1$, then $X = \sum_{n=1}^{\infty}\frac{1}{2^n} = 1$, but $\forall x \in [0,1]$, $\mu_X\{X = x\} = 0$, so it has no influence on $\mu_X[a, b]$. □

1.3 Expectations

1. (Remark) Expectations on different Ω.

   (a) Finite: $\mathbb{E}X = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\omega)$.

   (b) Countably infinite: $\mathbb{E}X = \sum_{k=1}^{\infty} X(\omega_k)\mathbb{P}(\omega_k)$.

   (c) Uncountably infinite: think in terms of integrals.

2. (Theorem) Some basic properties of Lebesgue integral (based on the random variable):

   (a) If X takes only finitely many values $y_0, y_1, y_2, \cdots, y_n$, then

   $$\int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \sum_{k=1}^{n} y_k\mathbb{P}\{X = y_k\}$$

   (b) (Integrability) The random variable X is integrable if and only if

   $$\int_\Omega |X(\omega)|\,\mathrm{d}\mathbb{P}(\omega) < \infty$$

   (c) (Comparison) If

   $$X \leqslant Y \text{ a.s.}$$

   and $\int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega)$,$\int_\Omega Y(\omega)\,\mathrm{d}\mathbb{P}(\omega)$ are defined, then

   $$\int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega) \leqslant \int_\Omega Y(\omega)\,\mathrm{d}\mathbb{P}(\omega)$$

   (d) (Linearity) If α and β are real constants and X, Y are integrable, or if α and β are nonnegative constants and X and Y are nonnegative, then

   $$\int_\Omega (\alpha X(\omega) + \beta Y(\omega))\,\mathrm{d}\mathbb{P}(\omega) = \alpha\int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega) + \beta\int_\Omega Y(\omega)\,\mathrm{d}\mathbb{P}(\omega)$$

3. (Definition) Expectation: Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the expectation of X is defined to be

$$\mathbb{E}X = \int_{\Omega} X(\omega) \, \mathrm{d}\mathbb{P}(\omega)$$

This definition makes sense if X is integrable, i.e.; if

$$\mathbb{E}|X| = \int_{\Omega} |X(\omega)| \, \mathrm{d}\mathbb{P}(\omega) < \infty$$

or if $X \geqslant 0$ a.s. ($\mathbb{E}X$ might be ∞) .

注：关于积分更详细的定义请参考实变函数和测度论的相关内容。

4. (Theorem, **Jensen's inequality**) If φ is a convex, real-valued function defined on $\mathbb{R}$, and if $\mathbb{E}|X| < \infty$, then

$$\varphi(\mathbb{E}X) \leqslant \mathbb{E}\varphi(X)$$

Proof. A convex function φ is the maximum of all linear functions ℓ that lie below it; i.e., for every $x \in \mathbb{R}$,

$$\varphi(x) = \sup\{\ell(x) : \ell \text{ is linear and } \ell(y) \leqslant \varphi(x), \forall x \in \mathbb{R}\}$$

We have

$$\mathbb{E}\varphi(X) \geqslant \mathbb{E}\ell(X) = \ell[\mathbb{E}X]$$

which holds for every $x \in \mathbb{R}$, and thus

$$\mathbb{E}\varphi(X) \geqslant \sup\{\ell \text{ is linear and } \ell(\mathbb{E}X) : \ell(y) \leqslant \varphi(y), \forall x \in \mathbb{R}\} = \varphi(\mathbb{E}X)$$ □

5. (Definition) **Lebesgue Measure**: Let $\mathcal{B}(\mathbb{R})$ be the σ-algebra of Borel subsets of $\mathbb{R}$. The Lebesgue measure on $\mathbb{R}$, which we denote by $\mathcal{L} : \mathcal{B} \mapsto [0, \infty)$ or $[0, \infty) \cup \{\infty\}$ such that

   (a) (Nonnegative) $\forall a, b \in \mathbb{R}, a \leqslant b$, then $\mathcal{L}[a, b] = b - a$

   (b) (Countable Additivity) if $\{B_n, n \geqslant 1\}$ is a sequence of disjoint sets in $\mathcal{B}$, then we have

$$\mathcal{L}\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mathcal{L}(B_n)$$

6. (Remark and Theorem) Comparison of Riemann and Lebesgue integrals.

7. (Definition) The definition of almost everywhere.

1.4 Convergence of Integrals

1. (Definition) Almost surely convergence: Let $X_1, X_2, X_3, \cdots$ be a sequence of random variables, all defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let X be another random variable defined on this space. We say that $X_1, X_2, X_3, \cdots$ converges to X almost surely, and write

$$\lim_{n \to \infty} X_n = X, \text{a.s.}$$

if the set of $\omega \in \Omega$ for which the sequence of numbers $X_1(\omega), X_2(\omega), \cdots$ has limit $X(\omega)$ is a set with probability one. Equivalently, the set of $\omega \in \Omega$ for which the sequence of numbers $X_1(\omega), X_2(\omega), \cdots$ does not converge to $X(\omega)$ is a set with probability zero.

Remark: An intuitively appealing case of almost surely convergence is the *Strong Law of Large Numbers.*

2. (Definition) Almost everywhere convergence.

3. (Example) Consider a sequence of normal densities, each with mean zero and the n^{th} having variance $\frac{1}{n}$:
$$f_n(x) = \sqrt{\frac{n}{2\pi}} e^{-\frac{nx^2}{2}}$$
The sequence $f_1, f_2, f_3, \ldots$ converges ***everywhere*** to the function
$$f^*(x) = \begin{cases} 0, & \text{if } x \neq 0 \\ \infty, & \text{if } x = 0 \end{cases}$$
and converges to $f(x) = 0$ ***almost everywhere***. Therefore,
$$\int_{\mathbb{R}} \lim_{n\to\infty} f_n(x) = 0 \neq \lim_{n\to\infty} \int_{\mathbb{R}} f_n(x) = 1$$
But if we define
$$g_n(x) = \begin{cases} f_n(x), & \text{if } x \neq 0 \\ 0, & \text{if } x = 0 \end{cases}$$
then g_n converges to 0 ***everywhere***, whereas
$$\int_{\mathbb{R}} \lim_{n\to\infty} g_n(x) = \lim_{n\to\infty} \int_{\mathbb{R}} g_n(x) = 1$$

4. (Example) Define f_n to be the normal density with mean zero and variance n, i.e.,
$$f_n = \frac{1}{\sqrt{2\pi n}} e^{-\frac{x^2}{2n}}$$
Note that
$$\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \frac{1}{\sqrt{2n\pi}} e^{-\frac{x^2}{2n}} = 0$$
and
$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)\,\mathrm{d}x = \lim_{n\to\infty} 1 = 1 \neq \int_{-\infty}^{\infty} \lim_{n\to\infty} f_n(x)\,\mathrm{d}x$$

5. (Remark) Any 2 functions that differ only on a set of zero Lebesgue measure must have the same Lebesgue integral.

6. (**Monotone Convergence Theorem**) Let $X_1, X_2, X_3, \cdots$ be a sequence of random variables converging almost surely to another random variable X. If
$$0 \leqslant X_1 \leqslant X_2 \leqslant X_3 \leqslant \cdots, \text{a.s.}$$
then
$$\lim_{n\to\infty} \mathbb{E}X_n = \mathbb{E}X$$

Let $f_1, f_2, f_3, \cdots$ be a sequence of Borel-measurable functions on $\mathbb{R}$ converging almost everywhere to a function f. If

$$0 \leqslant f_1 \leqslant f_2 \leqslant f_3 \leqslant \cdots, \text{a.e.}$$

then

$$\lim_{n\to\infty}\int_{-\infty}^{+\infty} f_n(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty}\lim_{n\to\infty} f_n(x)\,\mathrm{d}x$$

注：（**单调收敛定理**，测度与概率）给定 $(\Omega, \mathcal{F}, \mu)$，$\{f_n, n\in\mathbb{N}\}$ 是非负可测函数列，且 $0 \leqslant f_n \uparrow f$, a.e.，则

$$\lim_{n\to\infty}\int f_n = \int \lim_{n\to\infty} f_n$$

7. (Corollary) Suppose a nonnegative random variable X takes countably many value $x_0, x_1, x_2, \cdots$. Then

$$\mathbb{E}X = \sum_{k=0}^{\infty} x_k \mathbb{P}\{X = x_k\}$$

注：即为初等函数的积分定义。

e.g. (St. Peters-burg paradox) $\mathbb{E}X$ can be ∞, even though X is finite a.s.. Define

$$\mathbb{P}\left(X = 2^k\right) = \frac{1}{2^k}$$

Therefore,

$$\mathbb{P}(X=\infty) = \lim_{k\to\infty}\frac{1}{2^k} = 0$$

But

$$\mathbb{E}X = \sum_{k=1}^{\infty}\mathbb{P}\left(X=2^k\right)\cdot 2^k = \sum_{k=1}^{\infty} 1 = \infty$$

8. (**Dominated Convergence Theorem**) Let $X_1, X_2, X_3, \cdots$ be a sequence of random variables converging almost surely to another random variable X. If there is another random variable Y such that $\mathbb{E}Y < \infty$, and $|X_n| \leqslant Y$ almost surely for every n, then

$$\lim_{n\to\infty}\mathbb{E}X_n = \mathbb{E}X$$

Let $f_1, f_2, f_3, \cdots$ be a sequence of Borel-measurable functions on $\mathbb{R}$ converging almost everywhere to a function f. If there is another function g such that $\int_{-\infty}^{+\infty} g(x)\,\mathrm{d}x < \infty$ and $|f_n| < g$ almost everywhere for every n, then

$$\lim_{n\to\infty}\int_{-\infty}^{+\infty} f_n(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty}\lim_{n\to\infty} f_n(x)\,\mathrm{d}x$$

注：（**控制收敛定理**，实变函数、测度与概率）给定 $(\Omega, \mathcal{F}, \mu)$，$g, h$ 为实可积函数.

(a) 若 $\{f_n, n\in\mathbb{N}\}$ 是实可测函数列，当 $g \leqslant f_n \leqslant h$ a.e.,$\forall n\in\mathbb{N}$,$f_n \to f$ a.e. 时，有

$$\int f_n \to \int f$$

(b) 若 $\{f_n, n\in\mathbb{N}\}$ 是实或复可测函数列，当 $|f_n| \leqslant g$, a.e., $\forall n\in\mathbb{N}$, $f_n \to f$,a.e. 时，有

$$\int |f_n - f| \to 0$$

因而

$$\int f_n \to \int f$$

1.5 Computation of Expectations

1. (Definition) $\mathbb{E}X = \int_{\Omega} X(\omega)\,\mathrm{d}\mathbb{P}(\omega)$.

2. (Theorem) Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let g be a Borel-measurable function on $\mathbb{R}$. Then

$$\mathbb{E}\left|g(X)\right| = \int_{\mathbb{R}} \left|g(x)\right| \mathrm{d}\mu_X(x)$$

and if this quantity is finite, then

$$\mathbb{E}g(X) = \int_{\mathbb{R}} g(x)\,\mathrm{d}\mu_X(x)$$

Proof. Standard machine method for $g(x)$. □

注：“Standard Machine Method” 是常用的证明方法。我们如果相对一般可测函数证明某性质，则可以从简单函数开始，到非负可测函数，最后到一般可测函数的证明方法。前者使用到非负简单函数单调收敛到非负可测函数的性质，后者将一般可测函数表示为正部与负部的差。

3. (Definition) If there is a nonnegative, Borel-measurable function f defined on $\mathbb{R}$ such that

$$\mu_X(B) = \mathbb{P}(X \in B) = \int_B f(x)\,\mathrm{d}x, \forall B \in \mathcal{B}$$

we call X has a density function f.

4. (Theorem) Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let g be a Borel-measurable function on $\mathbb{R}$. Suppose that X has a density f. Then

$$\mathbb{E}\left|g(X)\right| = \int_{-\infty}^{+\infty} \left|g(x)\right| f(x)\,\mathrm{d}x$$

and if this quantity is finite, then

$$\mathbb{E}g(X) = \int_{-\infty}^{+\infty} g(x) f(x)\,\mathrm{d}x$$

Proof. Standard machine method for $g(x)$. □

1.6 Change of Measure

1. (Review) 考虑一般有限样本空间 Ω 上的两个概率测度 $\mathbb{P}$ 和 $\tilde{\mathbb{P}}$，假定两个概率测度下，对 Ω 中的每个元素均给出正的概率，于是

$$Z(\omega) = \frac{\tilde{\mathbb{P}}(\omega)}{\mathbb{P}(\omega)}$$

Z 是随机变量，因它依赖于随机试验的结果 ω。Z 被称为 $\tilde{\mathbb{P}}$ 关于 $\mathbb{P}$ 的 Radon-Nikodým 导数，随机变量 Z 具有重要性质，即下面的定理.

2. (Theorem) 设 $\mathbb{P}$ 和 $\tilde{\mathbb{P}}$ 是有限样本空间 Ω 上的两个概率测度，对 $\forall\omega \in \Omega, \mathbb{P}(\omega) > 0, \tilde{\mathbb{P}}(\omega) > 0$，定义随机变量 Z，满足 $Z(\omega) = \frac{\tilde{\mathbb{P}}(\omega)}{\mathbb{P}(\omega)}$。那么

   (a) $\mathbb{P}(Z > 0) = 1$;

(b) $\mathbb{E}Z = 1$;

(c) 对任意随机变量 Y，有 $\tilde{\mathbb{E}}Y = \mathbb{E}[ZY]$。

3. (Remark)

(a) Change notation: When Ω is uncountably infinite and $\mathbb{P}(\omega) = \tilde{\mathbb{P}}(\omega) = 0$ for every $\omega \in \Omega$, it no longer make sense to write $Z(\omega) = \frac{\tilde{\mathbb{P}}(\omega)}{\mathbb{P}(\omega)}$, we should rewrite this equation as

$$Z(\omega)\,\mathbb{P}(\omega) = \tilde{\mathbb{P}}(\omega)$$

It is a meaningful equation, with both sides equal to zero, but the equation tells us nothing about the relationship among $\mathbb{P}, \tilde{\mathbb{P}}$, and Z. Because when $\mathbb{P}(\omega) = \tilde{\mathbb{P}}(\omega) = 0$, the value of $Z(\omega)$ could be anything and this equation still holds.

(b) The nature of Measure Change: $Z(\omega)\,\mathbb{P}(\omega) = \tilde{\mathbb{P}}(\omega)$ does capture the spirit of what we would like to accomplish. To change from $\mathbb{P}$ to $\tilde{\mathbb{P}}$, we need to reassign probabilities in Ω using Z to tell us where in Ω we should revise the probability upward (where $Z > 1$) and where we should revise the probability downward (where $Z < 1$). However, we should do this ***set by set***, rather than ω-by-ω.

4. (Theorem) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let Z be an almost surely nonnegative random variable with $\mathbb{E}Z = 1$. For $A \in \mathcal{F}$, define

$$\tilde{\mathbb{P}}(A) = \int_A Z(\omega)\,\mathrm{d}\mathbb{P}(\omega)$$

Then $\tilde{\mathbb{P}}$ is a probability measure. Furthermore, if X is a nonnegative random variable, then

$$\tilde{\mathbb{E}}X = \mathbb{E}[ZX]$$

If $Z > 0$ a.s., we also have

$$\mathbb{E}Y = \tilde{\mathbb{E}}\left[\frac{Y}{Z}\right]$$

for every nonnegative random variable Y.

The $\tilde{\mathbb{E}}$ is expectation under the probability measure $\tilde{\mathbb{P}}$, i.e.,

$$\tilde{\mathbb{E}}X = \int_\Omega X(\omega)\,\mathrm{d}\tilde{\mathbb{P}}(\omega)$$

Remark: Suppose random variable X or Y can take both positive and negative values. We may apply $\tilde{\mathbb{E}}X = \mathbb{E}[ZX]$ and $\mathbb{E}Y = \tilde{\mathbb{E}}\left[\frac{Y}{Z}\right]$ respectively to its parts X^+ and X^- (Y^+, Y^-), and then subtract the resulting equations to see this equation holds for this X (Y) as well, provided the subtraction does not result in an $\infty - \infty$ situation.

Proof. To check that $\tilde{\mathbb{P}}$ is a probability measure, we must verify that $\tilde{\mathbb{P}}(\Omega) = 1$ and that $\tilde{\mathbb{P}}$ is countably additive. We have by assumption

$$\tilde{\mathbb{P}}(\Omega) = \int_\Omega Z(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \mathbb{E}Z = 1$$

For countable additivity, let $A_1, A_2, \cdots$ be a sequence of disjoint sets in $\mathcal{F}$, and define $B_n = \bigcup_{k=1}^n A_k, B_\infty = \bigcup_{k=1}^\infty A_k$, then we have

$$\mathbb{I}_{B_1} \leqslant \mathbb{I}_{B_2} \leqslant \cdots \leqslant \mathbb{I}_{B_k} \leqslant \cdots$$

and

$$\lim_{n\to\infty} \mathbb{I}_{B_n} = \mathbb{I}_{B_\infty}$$

By the Monotone Convergence Theorem, we have

$$\begin{aligned}\tilde{\mathbb{P}}\left(\bigcup_{k=1}^{\infty} A_k\right) &= \tilde{\mathbb{P}}\left(B_\infty\right) = \int_{B_\infty} Z\left(\omega\right) \mathrm{d}\mathbb{P}\left(\omega\right) \\ &= \int_\Omega \mathbb{I}_{B_\infty}\left(\omega\right) Z\left(\omega\right) \mathrm{d}\mathbb{P}\left(\omega\right) = \lim_{n\to\infty}\int_\Omega \mathbb{I}_{B_n}\left(\omega\right) Z\left(\omega\right) \mathrm{d}\mathbb{P}\left(\omega\right)\end{aligned}$$

But

$$\mathbb{I}_{B_n}\left(\omega\right) = \sum_{k=1}^{n} \mathbb{I}_{A_k}\left(\omega\right)$$

And so

$$\int_\Omega \mathbb{I}_{B_n}\left(\omega\right) Z\left(\omega\right) \mathrm{d}\mathbb{P}\left(\omega\right) = \int_\Omega \sum_{k=1}^{n} \mathbb{I}_{A_k}\left(\omega\right) Z\left(\omega\right) \mathrm{d}\mathbb{P}\left(\omega\right) = \sum_{k=1}^{n}\int_\Omega \mathbb{I}_{A_k}\left(\omega\right) Z\left(\omega\right) \mathrm{d}\mathbb{P}\left(\omega\right) = \sum_{k=1}^{n} \tilde{\mathbb{P}}\left(A_k\right)$$

Putting these two equation together, we obtain the countable additivity property

$$\tilde{\mathbb{P}}\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{n\to\infty}\sum_{k=1}^{n} \tilde{\mathbb{P}}\left(A_k\right) = \sum_{k=1}^{\infty} \tilde{\mathbb{P}}\left(A_k\right)$$

Now suppose X is a nonnegative random variable. If X is an indicator function $X = \mathbb{I}_A$, then

$$\tilde{\mathbb{E}}X = \tilde{\mathbb{P}}\left(A\right) = \int_\Omega \mathbb{I}_A\left(\omega\right) Z\left(\omega\right) \mathrm{d}\mathbb{P}\left(\omega\right) = \mathbb{E}\left[\mathbb{I}_A Z\right] = \mathbb{E}\left[XZ\right]$$

By using the **standard machine**, we can finish the proof.

When $Z > 0$ almost surely, $\frac{Y}{Z}$ is defined and we may replace X by $\frac{Y}{Z}$ to obtain $\mathbb{E}Y = \tilde{\mathbb{E}}\left[\frac{Y}{Z}\right]$. □

5. (Definition) Equivalent probability measure: Let Ω be a nonempty set and $\mathcal{F}$ a σ-algebra of subsets of Ω. Two probability measures $\mathbb{P}$ and $\tilde{\mathbb{P}}$ on $(\Omega, \mathcal{F})$ are said to be equivalent if they agree which sets in $\mathcal{F}$ have probability zero.

   Remark: Specially, $\mathbb{P}$ and $\tilde{\mathbb{P}}$ in this section are equivalent. Because $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are equivalent, we do not need to specify which measure we have in mind when we say an event occurs *almost surely.*

   注：可参考的定义。

   (a)（定义，测度与概率）设 $(\Omega, \mathcal{F}, \mu)$ 是测度空间，φ 是 $\mathcal{F}$ 上的一集函数。$\forall A \in \mathcal{F}$，若 $\mu\left(A\right) = 0$，有 $\varphi\left(A\right) = 0$，则称 φ 为 μ 连续，记作 $\varphi \ll \mu$。

   (b)（定义，金融观点下的随机分析基础）设 $\mathbb{P}$ 和 $\mathbb{Q}$ 是定义在 σ 代数 $\mathcal{F}$ 上的两个概率测度，若存在一个非负函数 f，使得

   $$\mathbb{Q}\left(A\right) = \int_A f\left(\omega\right) \mathrm{d}\mathbb{P}\left(\omega\right), \forall A \in \mathcal{F}$$

   则称 f 为 $\mathbb{Q}$ 关于 $\mathbb{P}$ 的密度，且称 $\mathbb{Q}$ 关于 $\mathbb{P}$ 绝对连续，记作 $\mathbb{Q} \ll \mathbb{P}$。

   (c)（定义，金融观点下的随机分析基础）设 $\mathbb{P}$ 和 $\mathbb{Q}$ 是定义在 σ 代数 $\mathcal{F}$ 上的两个概率测度，若 $\mathbb{Q}$ 关于 $\mathbb{P}$ 绝对连续且 $\mathbb{P}$ 关于 $\mathbb{Q}$ 绝对连续，即 $\mathbb{Q} \ll \mathbb{P}$ 且 $\mathbb{P} \ll \mathbb{Q}$，则称 $\mathbb{P}$ 和 $\mathbb{Q}$ 是等价的概率测度。

6. (Definition) Radon-Nikodým derivative: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\tilde{\mathbb{P}}$ be another probability measure on $(\Omega, \mathcal{F})$ that is equivalent to $\mathbb{P}$, and let Z be an almost surely positive random variable that relates $\mathbb{P}$ and $\tilde{\mathbb{P}}$ via $\tilde{\mathbb{P}}(A) = \int_A Z(\omega)\, \mathrm{d}\mathbb{P}(\omega)$.

   Then Z is called Radon-Nikodým derivative of $\tilde{\mathbb{P}}$ with respect to $\mathbb{P}$, and we write

$$Z(\omega) = \frac{\mathrm{d}\tilde{\mathbb{P}}(\omega)}{\mathrm{d}\mathbb{P}(\omega)}$$

7. (**Radon-Nikodým Theorem**) Let $\mathbb{P}$ and $\tilde{\mathbb{P}}$ be equivalent probability measures defined on $(\Omega, \mathcal{F})$. Then there exists an almost surely positive random variable Z such that $\mathbb{E}Z = 1$ and for $\forall A \in \mathcal{F}$, we have

$$\tilde{\mathbb{P}}(A) = \int_A Z(\omega)\, \mathrm{d}\mathbb{P}(\omega)$$

   注：$\mathbb{E}Z = 1$ 是很自然的结论，因为 $\tilde{\mathbb{P}}(\Omega) = \int_\Omega Z \mathrm{d}\mathbb{P} = \mathbb{E}Z = 1$。下面给出测度论中可参考的内容。

   (a)（**Radon-Nikodým 定理**，测度与概率）设 μ 是 Ω 中的 σ 代数 $\mathcal{F}$ 上的 σ 有限测度，φ 为 $\mathcal{F}$ 上 σ 有限、μ 连续的符号测度，则 φ 可以表示为

$$\varphi(A) = \int_A f \mathrm{d}\mu, A \in \mathcal{F}$$

   其中 f 是某一 a.e. 有限 $\mathcal{F}$ 可测函数（即表示为某一 a.e. 有限 $\mathcal{F}$ 可测函数的不定积分），且 f 由 φ 关于 μ 几乎唯一决定（即若 g 也满足上述条件，则 $f = g, \mu$ a.e.）。

   (b)（**Radon-Nikodým 定理的推广**，测度与概率）设 μ 是 Ω 中的 σ 代数 $\mathcal{F}$ 上的 σ 有限测度，φ 为 $\mathcal{F}$ 上 μ 连续的符号测度，则 φ 可以表示为

$$\varphi(A) = \int_A f \mathrm{d}\mu, A \in \mathcal{F}$$

   其中 f 是某一 $\mathcal{F}$ 可测函数（即表示为某一 $\mathcal{F}$ 可测函数的不定积分），且 f 由 φ 关于 μ 几乎唯一决定。

   (c)（定义，测度与概率）Radon-Nikodým 导数：设 $(\Omega, \mathcal{F}, \mu)$ 是 σ 有限测度空间，φ 是 $\mathcal{F}$ 上 μ 连续的符号测度，则依照 Radon-Nikodým 定理的陈述，μ 几乎唯一决定的函数 f 称为 φ 关于 μ 的 Radon-Nikodým 导数，记作 $f = \frac{\mathrm{d}\varphi}{\mathrm{d}\mu}$。

   (d) 例：概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 下，连续随机变量的概率分布为 $\mathbb{P}(A) = \int_A f \mathrm{d}x$，其中 f 是概率密度函数，$\mathrm{d}x$ 代表的是 Lebesgue 测度 $\mathcal{L}$，而 $\mathcal{L}$ 是 σ 有限测度，根据 Radon-Nikodým 定理，f 由 $\mathbb{P}$ 关于 $\mathcal{L}$ 几乎唯一确定。

8. (Example) Let $\Omega = [0, 1]$, $\mathbb{P}$ is the Lebesgue uniform measure, and

$$\tilde{\mathbb{P}}[a, b] = \int_a^b 2\omega \mathrm{d}\omega = b^2 - a^2, 0 \leqslant a \leqslant b \leqslant 1$$

   or we can rewrite as

$$\tilde{\mathbb{P}}[a, b] = \int_{[a,b]} 2\omega \mathrm{d}\mathbb{P}(\omega), 0 \leqslant a \leqslant b \leqslant 1$$

   for all Borel Measurable sets B,

$$\tilde{\mathbb{P}}(B) = \int_B 2\omega \mathrm{d}\mathbb{P}(\omega), B \in \mathcal{B}([0, 1])$$

which is with

$$Z(\omega) = 2\omega$$

and

$$\mathbb{E}Z = \int_{[0,1]} 2\omega \mathrm{d}\mathbb{P} = \int_0^1 2\omega \mathrm{d}\mathbb{P} = 1$$

For every nonnegative radon variable $X(\omega)$, we have

$$\int_{[0,1]} X \mathrm{d}\tilde{\mathbb{P}} = \int_{[0,1]} X(\omega) 2\omega \mathrm{d}\mathbb{P}(\omega)$$

which suggests the notation

$$\mathrm{d}\tilde{\mathbb{P}}(\omega) = 2\omega \mathrm{d}\omega = 2\omega \mathrm{d}\mathbb{P}$$

1.7 Selected Exercise

Exercise 1.1. The infinite coin-toss space Ω_∞ is *uncountably infinite*. In other words, we cannot list all its elements in a sequence. To see that this is impossible suppose there were such a sequential list of all elements of Ω_∞:

$$\begin{aligned}
\omega^{(1)} &= \omega_1^{(1)}\omega_2^{(1)}\omega_3^{(1)}\omega_4^{(1)}\cdots, \\
\omega^{(2)} &= \omega_1^{(2)}\omega_2^{(2)}\omega_3^{(2)}\omega_4^{(2)}\cdots, \\
\omega^{(3)} &= \omega_1^{(3)}\omega_2^{(3)}\omega_3^{(3)}\omega_4^{(3)}\cdots, \\
&\vdots
\end{aligned}$$

An element that does not appear in this list is the sequence whose first component is H if $\omega_1^{(1)} = T$ and is T if $\omega_1^{(1)} = H$, whose second component is H if $\omega_2^{(2)} = T$ and is T if $\omega_2^{(2)} = H$, etc. Thus, the list does not include every element of Ω_∞.

Now consider the set of sequences of coin tosses in which the outcome on each even-numbered toss matched the outcome of the toss preceding it, i.e.,

$$A := \{\omega = \omega_1\omega_2\omega_3\omega_4\cdots : \omega_1 = \omega_2, \omega_3 = \omega_4, \ldots\}.$$

1. Show that A uncountably infinite.

   Proof. Define a mapping $\varphi : A \to \Omega_\infty$, $\varphi(\omega_1\omega_3\omega_5\cdots) = \omega_1\omega_2\omega_3\cdots$, which is one-to-one correspondence. □

2. Show that, when $0 < p < 1$, we have $\mathbb{P}(A) = 0$.

   Proof. Let $A_n = \{\omega = \omega_1\omega_2\omega_3\cdots : \omega_1 = \omega_2, \cdots, \omega_{2n-1} = \omega_{2n}\}$, then $A_n \downarrow A$ as $n \to \infty$, and $\mathbb{P}(A_1) \leqslant 1 < \infty$, so we have

$$\mathbb{P}(A) = \mathbb{P}\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n)$$

while

$$\begin{aligned}\mathbb{P}(A_n) &= \mathbb{P}\{\omega_1 = \omega_2\} \cdot \mathbb{P}\{\omega_3 = \omega_4\} \cdots \mathbb{P}\{\omega_{2n-1} = \omega_{2n}\} \\ &= \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) \left[p^2\right]^k \left[(1-p)^2\right]^{n-k} \\ &= \left[p^2 + (1-p)^2\right]^n\end{aligned}$$

thus, we have

$$\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} \left[p^2 + (1-p)^2\right]^n \leqslant \lim_{n\to\infty} \left[2p(1-p)\right]^n \leqslant \lim_{n\to\infty} \left(\frac{1}{2}\right)^n = 0 \qquad \square$$

Exercise 1.2. The finite additivity property does not imply the countable additivity property. Consider the set function $\mathbb{P}$ defined for every subset of $[0,1]$ by the formula that $\mathbb{P}(A) = 1$ if A is a finite set and $\mathbb{P}(A) = \infty$ if A is an infinite set. Show that $\mathbb{P}$ satisfies the finite additivity property, but $\mathbb{P}$ does not have the countable additivity property.

Proof. The set function $\mathbb{P}$ is not a probability measure.

1. $\mathbb{P}(\varnothing) = 0$;
2. If A and B are both finite, then $A \cup B$ is also finite, which means $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$. If A is infinite but B is finite, then $A \cup B$ is infinite, the equation holds. If A and B are both infinite, the equation also holds.
3. $\mathbb{P}\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} \mathbb{P}(A_n)$ is the general form of $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$, even if A_ns are not disjoint.
4. Countable additivity property does not hold for $\mathbb{P}$. Let $A_n = \left\{\frac{1}{n}\right\}$, Then $A = \bigcup_{i=1}^{\infty} A_i$ is an infinite set and therefore $\mathbb{P}(A) = 1$. However, $\mathbb{P}(A_n) = 0$ for every n. $\square$

Exercise 1.3. (**Hints on Monte Carlo simulation**)

1. Construct a standard normal random variable Z on the probability space $(\Omega_\infty, \mathcal{F}_\infty, \mathbb{P})$ of the example of tossing coin under the assumption that the probability for head is $p = \frac{1}{2}$

   Solution. Give 2 solutions here:

   (a) Based on the Central-Limit Theorem, let $Z := \lim_{n\to\infty} \frac{\frac{1}{n}\sum_{i=1}^{n} X_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$, where X_n is a random variable such that $X_n = 1$ if $\omega_n = H$, $X_n = 0$ if $\omega_n = T$, and $\mu = \frac{1}{2}, \sigma = \sqrt{\frac{1}{2}\left(1 - \frac{1}{2}\right)} = \frac{1}{2}$.

   (b) we can define a random variable $X := \sum_{n=1}^{\infty} \frac{\mathbb{I}_{\{\omega_n = H\}}}{2^n} \sim U[0,1]$. Then using the *inverse transition method* to get a norm random variable $Z = N^{-1}(X)$, where $N(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t$. $\square$

2. Define a sequence of random variables $\{Z_n\}_{n=1}^{\infty}$ on Ω_∞ such that

$$\lim_{n\to\infty} Z_n(\omega) = Z(\omega), \forall \omega \in \Omega_\infty$$

   and, for each n, Z_n depends only on the first n coin tosses.

Solution. Correspondingly, we have

(a) $Z_n := \frac{\frac{1}{n}\sum_{i=1}^{n} X_n - \mu}{\frac{\sigma}{\sqrt{n}}} \to Z \sim N(0,1)$

(b) $Z_n := N^{-1}\left(\sum_{k=1}^{n} \frac{\mathbb{I}_{\{\omega_k = H\}}}{2^k}\right) \to Z \sim N(0,1)$ □

Exercise 1.4. When dealing with double Lebesgue integrals, just as with double Riemann integrals, the order of integration can be reversed. The only assumption required is that the function being integrated be either nonnegative or integrable. Here is an application of this fact.

Let X be a nonnegative random variable with CDF $F(x) = \mathbb{P}\{X \leqslant x\}$. Show that

$$\mathbb{E}X = \int_0^\infty (1 - F(x))\,\mathrm{d}x$$

by showing that

$$\int_\Omega \int_0^\infty \mathbb{I}_{[0,X(\omega)]}(x)\,\mathrm{d}x\mathrm{d}\mathbb{P}(\omega)$$

is equal to both $\mathbb{E}X$ and $\int_0^\infty (1 - F(x))\,\mathrm{d}x$.

Proof. By the information given by the problem, we have

$$\int_\Omega \int_0^\infty \mathbb{I}_{[0,X(\omega))}(x)\,\mathrm{d}x\mathrm{d}\mathbb{P}(\omega) = \int_0^\infty \int_\Omega \mathbb{I}_{[0,X(\omega))}(x)\,\mathrm{d}\mathbb{P}(\omega)\,\mathrm{d}x$$

while

$$\begin{aligned}\int_\Omega \int_0^\infty \mathbb{I}_{[0,X(\omega))}(x)\,\mathrm{d}x\mathrm{d}\mathbb{P}(\omega) &= \int_\Omega \int_0^{X(\omega)} \mathrm{d}x\mathrm{d}\mathbb{P}(\omega)\\ &= \int_\Omega X(\omega)\mathrm{d}\mathbb{P}(\omega)\\ &= \mathbb{P}(X)\\ &= \mathbb{E}X\end{aligned}$$

and

$$\begin{aligned}\int_0^\infty \int_\Omega \mathbb{I}_{[0,X(\omega))}(x)\,\mathrm{d}\mathbb{P}(\omega)\,\mathrm{d}x &= \int_0^\infty \int_\Omega \mathbb{I}_{\{0\leqslant x<X(\omega)\}}(x)\,\mathrm{d}\mathbb{P}(\omega)\,\mathrm{d}x\\ &= \int_0^\infty \int_\Omega \mathbb{I}_{\{x<X(\omega)\}}(x)\,\mathrm{d}\mathbb{P}(\omega)\,\mathrm{d}x\\ &= \int_0^\infty \mathbb{P}\{x < X(\omega)\}\mathrm{d}x\\ &= \int_0^\infty [1 - \mathbb{P}\{x \geqslant X(\omega)\}]\mathrm{d}x\\ &= \int_0^\infty [1 - F(x)]\mathrm{d}x\end{aligned}$$

which is what need to be show. □

Exercise 1.5. (**Moment-generating function**) Suppose the random variable X can take both positive and negative values and $\varphi(t) = \mathbb{E}\left[e^{tX}\right] < \infty$ and $\mathbb{E}\left[|X|\,e^{tX}\right] < \infty$ for every $t \in \mathbb{R}$. Show that

$$\varphi'(t) = \mathbb{E}\left[Xe^{tX}\right]$$

Hints:

1. Let X be a random variable, and assume that

$$\varphi(t) = \mathbb{E}e^{tX}$$

is finite for every $t \in \mathbb{R}$. Assume further that $\mathbb{E}\left[Xe^{tX}\right] < \infty$ for every $t \in \mathbb{R}$. The purpose of this exercise is to show that

$$\varphi'(t) = \mathbb{E}\left[Xe^{tX}\right]$$

and, in particular, $\varphi'(0) = \mathbb{E}[X]$.

2. Definition of derivative:

$$\varphi'(t) = \lim_{s\to t} \frac{\varphi(t) - \varphi(s)}{t-s} = \lim_{s\to t} \frac{\mathbb{E}e^{tX} - \mathbb{E}e^{ts}}{t-s} = \lim_{s\to t} \mathbb{E}\left[\frac{e^{tX} - e^{sX}}{t-s}\right]$$

The limit above is taken over a *continuous* variable s, but we can choose a sequence of numbers $\{s_n\}_{n=1}^{\infty}$ converging to t and compute

$$\lim_{s_n\to t} \mathbb{E}\left[\frac{e^{tX} - e^{s_nX}}{t-s_n}\right]$$

Define

$$Y_n := \frac{e^{tX} - e^{s_nX}}{t-s_n}$$

so we just need to compute $\lim_{n\to\infty} \mathbb{E}[Y_n]$

3. The *Mean Value Theorem* from calculus states that if f is differentiable, then for any two numbers s and t, $s \leqslant t$, there is a $\theta \in [s,t]$ such that

$$f(t) - f(s) = f'(\theta)(t-s)$$

If we fix $\omega \in \Omega$ and define $f(t) = e^{tX(\omega)}$, then

$$e^{tX(\omega)} - e^{sX(\omega)} = (t-s)X(\omega)e^{\theta(\omega)X(\omega)}$$

where $\theta(\omega) \in [s,t]$.

Proof. From the Mean Value Theorem, there exist a $\theta_n \in [s_n, t]$ (assume $t > s_n$), such that

$$|Y_n| = \left|\frac{e^{tX} - e^{s_nX}}{t-s_n}\right| = \left|X(\omega)e^{\theta_n(\omega)X(\omega)}\right| = X(\omega)e^{\theta_n(\omega)X(\omega)} \leqslant X(\omega)e^{tX(\omega)}$$

and by the assumption, we have $\mathbb{E}\left[Xe^{tX}\right] < \infty$. So by the Dominated Convergence Theorem,

$$\lim_{n\to\infty} \mathbb{E}Y_n = \mathbb{E}\left[\lim_{n\to\infty} Y_n\right] = \mathbb{E}\left[Xe^{tX}\right] \qquad \square$$

Exercise 1.6. Suppose X is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, $A \in \mathcal{F}, B \in \mathcal{B}$, if we have

$$\int_A \mathbb{I}_B(X(\omega))\,\mathrm{d}\mathbb{P}(\omega) = \mathbb{P}(A)\cdot\mathbb{P}(X\in B)$$

then we say X is *independent* of A.

Show that if X is independent of A, then

$$\int_A g(X(\omega))\,\mathrm{d}\mathbb{P}(\omega) = \mathbb{P}(A)\cdot\mathbb{E}g(X)$$

holds for every nonnegative, Borel-measurable function g.

Proof. (Standard Machine)

1. g is the form $\mathbb{I}_B(X(\omega))$, the definition is what we need to show.

2. $g = \sum_{i=1}^n \mathbb{I}_{B_i}(x)$, where each B_i is a Borel subset of $\mathbb{R}$, By the linearity of Lebesgue integral, the desired equality also holds for simple functions.

3. Since any nonnegative, Borel function g is the limit of an increasing sequence of simple functions, i.e., $g(x) = \lim_{n\to\infty} \varphi_n(x)$, where $\{\varphi_n\}$ is an increasing sequence of simple functions, the desired equality can be proved by the Monotone Convergence Theorem.

$$\begin{aligned}
\int_A g(X(\omega))\,\mathrm{d}\mathbb{P}(\omega) &= \int_A \lim_{n\to\infty} \varphi_n(X(\omega))\,\mathrm{d}\mathbb{P}(\omega) \\
&= \lim_{n\to\infty} \int_A \varphi_n(X(\omega))\,\mathrm{d}\mathbb{P}(\omega) \\
&= \lim_{n\to\infty} \mathbb{P}(A)\,\mathbb{E}[\varphi_n(X)] \\
&= \mathbb{P}(A)\,\mathbb{E}\left[\lim_{n\to\infty} \varphi_n(X)\right] \\
&= \mathbb{P}(A)\,\mathbb{E}[g(X)]
\end{aligned}$$

□

Exercise 1.7. (**Change of measure for a normal random variable**) X is a normal random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, and define $Y = X + \theta$ and $Z = e^{-\theta X - \frac{1}{2}\theta^2}$, where θ is a constant. and used Z as the Radon-Nikodým derivative to construct the probability measure $\tilde{\mathbb{P}}$ by

$$\tilde{\mathbb{P}}(A) = \int_A Z\,\mathrm{d}\mathbb{P}, \forall A \in \mathcal{F}$$

If we fix $\bar{\omega} \in \Omega$ and choose a *small* set A such that $\bar{\omega} \in A$, which ensures that $\mathbb{P}(A) \neq 0$.

The purpose of this exercise is to identify $Z(\bar{\omega})$ such that

$$\frac{\tilde{\mathbb{P}}(A)}{\mathbb{P}(A)} \approx Z(\bar{\omega}), \forall A \in \mathcal{F} \cap \{\bar{\omega}\} = \{F \cap \{\bar{\omega}\} : F \in \mathcal{F}\}$$

With $\bar{\omega}$ fixed, let $x = X(\bar{\omega})$. For $\varepsilon > 0$, we define $B(x, \varepsilon) := \left[x - \frac{\varepsilon}{2}, x + \frac{\varepsilon}{2}\right]$. Let $y = x + \theta$.

1. Show that
$$\frac{1}{\varepsilon}\mathbb{P}\{X \in B(x,\varepsilon)\} \approx \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{X^2(\bar{\omega})}{2}\right\}$$

   Proof.

$$\begin{aligned}
\frac{1}{\varepsilon}\mathbb{P}\{X \in B(x,\varepsilon)\} &= \frac{1}{\varepsilon}\mathbb{P}\left\{X \in \left[x - \frac{\varepsilon}{2}, x + \frac{\varepsilon}{2}\right]\right\} \\
&= \frac{1}{\varepsilon}\int_{x-\frac{\varepsilon}{2}}^{x+\frac{\varepsilon}{2}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \mathrm{d}x \\
&\approx \frac{1}{\varepsilon}\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \cdot \varepsilon \\
&= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{X^2(\bar{\omega})}{2}\right\}
\end{aligned}$$

□

2. Show that

$$\frac{1}{\varepsilon}\tilde{\mathbb{P}}\left\{Y \in B\left(y,\varepsilon\right)\right\} \approx \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{Y^2\left(\bar{\omega}\right)}{2}\right\}$$

Proof. Under $\tilde{\mathbb{P}}$, $Y \sim N(0,1)$, which is because

$$\begin{aligned}
\tilde{\mathbb{P}}\left\{Y \leqslant b\right\} &= \int_{\{\omega:Y(\omega\leqslant b)\}} Z\mathrm{d}\mathbb{P} \\
&= \int_{\Omega} \mathbb{I}_{\{Y(\omega)\leqslant b\}} \exp\left\{-\theta X\left(\omega\right) - \frac{1}{2}\theta^2\right\}\mathrm{d}\mathbb{P}\left(\omega\right) \\
&= \int_{\Omega} \mathbb{I}_{\{X(\omega)+\theta\leqslant b\}} \exp\left\{-\theta X\left(\omega\right) - \frac{1}{2}\theta^2\right\}\mathrm{d}\mathbb{P}\left(\omega\right) \\
&= \int_{-\infty}^{\infty} \mathbb{I}_{\{x\leqslant b-\theta\}} \cdot e^{-\theta x - \frac{1}{2}\theta^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \mathbb{I}_{\{x+\theta\leqslant b\}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x+\theta)^2}\mathrm{d}\left(x+\theta\right) \\
&= \int_{-\infty}^{\infty} \mathbb{I}_{\{y\leqslant b\}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\mathrm{d}y \\
&= \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\mathrm{d}y
\end{aligned}$$

注：上述过程中有一步需要用到如下的积分变换定理。

(a)（**积分变换定理**，测度与概率）：设 f 是 $(\Omega,\mathcal{F},\mu)$ 到 $(E,\mathcal{E})$ 上的可测映射，g 是 $(E,\mathcal{E})$ 上的可测函数，则

$$\int_{f^{-1}(B)} \left(gf\right)\mathrm{d}\mu = \int_B g\mathrm{d}\mu_f, \forall B \in \mathcal{E}$$

即左边积分存在当且仅当右边积分存在，且二者相等。

(b) **注**：积分变换定理可以将 $f^{-1}\left(B\right)$ 上的积分转换到 B 上。

(c) **例**：对概率空间 $(\Omega,\mathcal{F},\mathbb{P})$ 而言，若 $X:\mathcal{F}\mapsto\mathbb{R}$ 为其上的 r.v.，g 是 $(\mathbb{R},\mathcal{B})$ 上的可测函数，例如 $g\left(x\right)=x^2$，那么 $g\circ X = X^2$ 的期望为

$$\int_{\Omega}\left(g\circ X\right)\mathrm{d}\mathbb{P} = \int_{\mathbb{R}} g\mathrm{d}\mathbb{P}_X = \int_{\mathbb{R}} x^2\mathrm{d}\mathbb{P}_X$$

Thus, it is simple to proof (Similar to 1). □

3. Show that

$$\left\{X \in B\left(x,\varepsilon\right)\right\} = \left\{Y \in B\left(y,\varepsilon\right)\right\}$$

which we call $A\left(\bar{\omega},\varepsilon\right)$. $A\left(\bar{\omega},\varepsilon\right)$ contains $\bar{\omega}$ and is “small” when ε is small.

Proof. $\left\{X \in B\left(x,\varepsilon\right)\right\} = \left\{X \in B\left(y-\theta,\varepsilon\right)\right\} = \left\{X+\theta \in B\left(y,\varepsilon\right)\right\} = \left\{Y \in B\left(y,\varepsilon\right)\right\}$ □

4. Show that

$$\frac{\tilde{\mathbb{P}}\left(A\right)}{\mathbb{P}\left(A\right)} \approx \exp\left\{-\theta X\left(\bar{\omega}\right) - \frac{1}{2}\theta^2\right\}$$

Proof.

$$\begin{aligned}
\frac{\tilde{\mathbb{P}}(A)}{\mathbb{P}(A)} \approx \frac{\mathbb{P}\{Y \in B(y,\varepsilon)\}}{\mathbb{P}\{X \in B(x,\varepsilon)\}} &= \frac{\frac{1}{\varepsilon}\mathbb{P}\{Y \in B(y,\varepsilon)\}}{\frac{1}{\varepsilon}\mathbb{P}\{X \in B(x,\varepsilon)\}} \\
&= \frac{\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{Y^2(\bar{\omega})}{2}\right\}}{\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{X^2(\bar{\omega})}{2}\right\}} \\
&= \exp\left\{-\frac{[X(\bar{\omega})+\theta]^2}{2} + \frac{X^2(\bar{\omega})}{2}\right\} \\
&= e^{-\frac{\theta^2}{2}-\theta X(\bar{\omega})}
\end{aligned}$$

□

Exercise 1.8. (**Change of measure for a exponential random variable**) Let X be a nonnegative random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the exponential distribution, which is

$$\mathbb{P}\{X \leqslant x\} = 1 - e^{-\lambda x}, x \geqslant 0$$

where λ is a positive constant. Let $\tilde{\lambda}$ be another positive constant, and define

$$Z(\omega) = \frac{\tilde{\lambda}}{\lambda} e^{-(\tilde{\lambda}-\lambda)X(\omega)}$$

Define $\tilde{\mathbb{P}}$ by

$$\tilde{\mathbb{P}}(A) = \int_A Z \mathrm{d}\mathbb{P}, \forall A \in \mathcal{F}$$

1. Show that $\tilde{\mathbb{P}}(\Omega) = 1$

   Proof.

$$\begin{aligned}
\tilde{\mathbb{P}}(\Omega) = \int_\Omega Z(\omega)\,\mathrm{d}\mathbb{P}(\omega) &= \int_\Omega \frac{\tilde{\lambda}}{\lambda} e^{-(\tilde{\lambda}-\lambda)X(\omega)} \mathrm{d}\mathbb{P}(\omega) \\
&= \int_0^\infty \frac{\tilde{\lambda}}{\lambda} e^{-(\tilde{\lambda}-\lambda)x} \cdot \lambda e^{-\lambda x} \mathrm{d}x \\
&= \int_0^\infty \tilde{\lambda} \cdot e^{-\tilde{\lambda}x} \mathrm{d}x \\
&= 1
\end{aligned}$$

□

2. Compute the cumulative distribution function $\tilde{\mathbb{P}}(X \leqslant x)$ for $x \geqslant 0$, for the random variable X under the probability measure $\tilde{\mathbb{P}}$.

   Solution.

$$\begin{aligned}
\tilde{\mathbb{P}}\{X \geqslant a\} = \int_\Omega \mathbb{I}_{\{X \geqslant a\}} Z(\omega)\,\mathrm{d}\mathbb{P}(\omega) &= \int_{\{X \geqslant a\}} \frac{\tilde{\lambda}}{\lambda} e^{-(\tilde{\lambda}-\lambda)X(\omega)} \mathrm{d}\mathbb{P}(\omega) \\
&= \int_a^\infty \frac{\tilde{\lambda}}{\lambda} e^{-(\tilde{\lambda}-\lambda)x} \cdot \lambda e^{-\lambda x} \mathrm{d}x = \int_a^\infty \tilde{\lambda} e^{-\tilde{\lambda}x} \mathrm{d}x \\
&= 1 - e^{-\tilde{\lambda}a}
\end{aligned}$$

□

Exercise 1.9. Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and assume X has a density function $f(x)$ that is positive for every $x \in \mathbb{R}$. Let g be a strictly increasing, differentiable function satisfying

$$\lim_{x\to\infty} g(x) = \infty, \lim_{x\to-\infty} g(x) = -\infty$$

and define the random variable $Y = g(X)$.

Let $h(x)$ be an arbitrary nonnegative function satisfying

$$\int_{-\infty}^{+\infty} h(x)\,\mathrm{d}x = 1$$

We want to change the probability measure so that $h(x)$ is the density function for the random variable Y. To do this, we define

$$Z := \frac{h[g(X)]\,g'(X)}{f(X)}$$

1. Show that $\mathbb{E}Z = 1$.

Proof. $f > 0, h \geqslant 0, g' \geqslant 0 \Longrightarrow Z > 0$

$$\begin{aligned}
\mathbb{E}Z &= \int_\Omega \frac{h[g(X(\omega))]\,g'(X(\omega))}{f(X(\omega))}\mathrm{d}\mathbb{P}(\omega) \\
&= \int_{-\infty}^{\infty} \frac{h[g(x)]\,g'(x)}{f(x)} f(x)\,\mathrm{d}x \\
&= \int_{\lim_{n\to-\infty} g(x)}^{\lim_{x\to\infty} g(x)} h[g(x)]\mathrm{d}g(x) \\
&= \int_{-\infty}^{\infty} h(u)\mathrm{d}u \\
&= 1
\end{aligned}$$

□

2. Define $\tilde{\mathbb{P}}$ by

$$\tilde{\mathbb{P}}(A) = \int_A Z\mathrm{d}\mathbb{P}, \forall A \in \mathcal{F}$$

Show that Y has density h under $\tilde{\mathbb{P}}$.

Proof.

$$\begin{aligned}
\tilde{\mathbb{P}}\{Y \leqslant a\} = \tilde{\mathbb{P}}\{g(X) \leqslant a\} &= \tilde{\mathbb{P}}\left\{X \leqslant g^{-1}(a)\right\} \\
&= \int_{\{X \leqslant g^{-1}(a)\}} \frac{h[g(X(\omega))]\,g'(X(\omega))}{f(X(\omega))}\mathrm{d}\mathbb{P}(\omega) \\
&= \int_{-\infty}^{g^{-1}(a)} h[g(x)]\,g'(x)\,\mathrm{d}x \\
&= \int_{-\infty}^{g^{-1}(a)} h[g(x)]\,\mathrm{d}g(x) \\
&= \int_{-\infty}^{a} h(u)\,\mathrm{d}u
\end{aligned}$$

□

第二章　独立性与条件期望

2.1 Information and σ-algebras

1. (Definition) Filtration:
$$\mathcal{F}(s) \subset \mathcal{F}(t), s \leqslant t$$
e.g. Let $X_n, n \geqslant 0$ be a random process on $(\Omega, \mathcal{F}, \mathbb{P})$. Define $\{\mathcal{F}_n, n \geqslant 0\}$ by
$$\mathcal{F}_n = \sigma(X_1, X_2, \cdots, X_n)$$
, then $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_{n-1} \subset \mathcal{F}_n \subset \cdots$, which means $\{\mathcal{F}_n\}$ is a filtration.

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

   注:（定理，测度与概率）设 $\mathcal{C}$ 是 Ω 的一个子集类，则存在一个唯一的 Ω 的 σ 代数 $\sigma(\mathcal{C})$，它包含 $\mathcal{C}$ 而且被包含 $\mathcal{C}$ 的任一 σ 代数所包含，即 $\sigma(\mathcal{C})$ 是包含 $\mathcal{C}$ 的最小 σ 代数，也称为由 $\mathcal{C}$ 生成的 σ 代数。

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

   注：可参考的定义。

   (a)（定义，测度与概率）设 $(\Omega, \mathcal{F}, \mu)$ 是一测度空间，若函数 $f: \Delta \mapsto \bar{\mathbb{R}}$，$\Delta \in \mathcal{F}$，使得 $\forall B \in \bar{\mathcal{B}}$，有
   $$f^{-1}(B) := \{\omega \in \Delta : f(\omega) = B\} \in \Delta \cap \mathcal{F}$$
   其中 $\Delta \cap \mathcal{F} = \{\Delta \cap A : A \in \mathcal{F}\}$，则

      i. 称 f 为定义在 Δ 上的可测函数；当 μ 为概率测度时，该可测函数 f 称为在 Δ 上的广义随机变量。

      ii. 若 f 取值于 $\mathbb{R}$，即 $f: \Delta(\in \mathcal{F}) \mapsto \mathbb{R}$，则称 f 为定义在 Δ 上的有限实值可测函数；当 μ 为概率测度时，该可测函数 f 称为 Δ 上的有限实值随机变量，简称为随机变量，记作 r.v.。

   (b)（定义，应用随机过程）对随机过程 $\{X_n, n \geqslant 0\}$，如果 $\forall x \in \mathbb{R}, \{X_n \leqslant x\} \in \mathcal{F}_n$，那么称 $\{X_n, n \geqslant 0\}$ 是 $\mathcal{F}_n$ 可测的。

4. (Definition) Adapted stochastic process: Let Ω be a nonempty sample space equipped with a filtration $\mathcal{F}(t), 0 \leqslant t \leqslant T$. Let $X(t)$ be a collection of random variables indexed by $t \in [0, T]$.

We say this collection of random variables is an adapted stochastic process if, for each t, the random variable $X(t)$ is $\mathcal{F}(t)$-measurable.

注：对 $(\Omega, \mathcal{F}, \mathbb{P})$ 上的随机过程 $\{Y_n, n \geqslant 0\}$ 和 $\{X_n, n \geqslant 0\}$，定义 σ 代数流 $\{\mathcal{F}_n, n \geqslant 0\}$，其中 $\mathcal{F}_n = \sigma(Y_1, Y_2, \cdots, Y_n)$。如果想表达"$X_n$ 由 $(Y_1, Y_2, \cdots, Y_n)$ 决定"或"X_n 是 $(Y_1, Y_2, \cdots, Y_n)$ 的函数"，则指出 $(\Omega, \mathcal{F}, \mathbb{P})$ 是 $\{\mathcal{F}_n, n \geqslant 0\}$ 适应列即可。

2.2 Independence

1. (Remark) Independence can be affected by changes of probability measure, but *measurability* is not.

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

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B), \forall A \in \mathcal{G}, B \in \mathcal{H}$$

   (a) (Definition) Let X and Y be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We say X and Y are independent if $\sigma(X)$ and $\sigma(Y)$ are independent.

   (b) (Definition) We say the random variable X is independent of the σ-algebra of $\mathcal{G}$ if $\sigma(X)$ and $\mathcal{G}$ are independent.

注：可参考的内容。

(a)（定义，测度与概率）设 X_k 是 $(\Omega, \mathcal{F}, \mathbb{P})$ 上的 m_k 维实 r.v.，$k = 1, 2, \cdots, n$，若 $\forall x_k \in \bar{\mathbb{R}}^{m_k}$，有

$$\mathbb{P}\left(\bigcap_{k=1}^{n} \{X_k \leqslant x_k\}\right) = \prod_{k=1}^{n} \mathbb{P}(\{X_k \leqslant x_k\})$$

则称 $X_1, X_2, \cdots, X_n$ 独立。

注：

i. m_k 不一定相等，即 X_k 的维数可以不同。

ii. 不妨思考该定义的合理性。该定义仅将 X_k 用 x_k 控制，或者说仅研究区间 $(-\infty, x_k]$，而不是如 $B \in \mathcal{B}$ 这样更加一般的集合。下面两个定理解释了该定义的合理性。

(b)（**独立事件类的扩张定理**，测度与概率）给定概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$，$\mathcal{C}_k \subset \mathcal{F}$ 为 π 系，$\Omega \in \mathcal{C}_k, k = 1, 2, \cdots, n$，且 $\forall A_k \in \mathcal{C}_k, k = 1, 2, \cdots, n$，有下式

$$\mathbb{P}\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} \mathbb{P}(A_k)$$

则上式对 $\forall A_k \in \sigma(\mathcal{C}_k), k = 1, 2, \cdots, n$，都成立。

(c)（定理，测度与概率）设 X_k 是 $(\Omega, \mathcal{F}, \mathbb{P})$ 上的 m_k 维实 r.v.，$k = 1, 2, \cdots, n$，$X_1, X_2, \cdots, X_n$ 独立，则 $\forall B_k \in \mathcal{B}^{m_k}, k = 1, 2, \cdots, n$，有

$$\mathbb{P}\left(\bigcap_{k=1}^{n} \{X_k \in B_k\}\right) = \prod_{k=1}^{n} \mathbb{P}(\{X_k \in B_k\})$$

因而，若 $f_k : \mathbb{R}^{m_k} \mapsto \mathbb{R}^{n_k}$ 是 Borel 可测函数，则 $f_k(X_k), k = 1, 2, \cdots, n$ 独立。

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

   Proof. From the definition, we have

   $$\sigma(f(X)) = \{\{f(X) \in B\} : B \in \mathcal{B}\}, \sigma(X) = \{\{X \in B\} \subset \Omega : B \in \mathcal{B}\}$$

   Therefore, $\forall A \in \sigma(f(X))$, or $\forall B \in \mathcal{B}$, $A = \{\omega \in \Omega : f(X(\omega)) \in B\}$. And because f is Borel-measurable, we can define a Borel-measurable set D by $D := \{x \in \mathbb{R} : f(x) \in B\} \in \mathcal{B}$, then

   $$A = \{\{X \in D\} \subset \Omega : D \in \mathcal{B}\} \in \sigma(X)$$

   then we have $\sigma(f(X)) \subset \sigma(X)$. Similarly, $\sigma(g(Y)) \subset \sigma(Y)$.

   Because $\sigma(X)$ and $\sigma(Y)$ be independent, it is clear that $f(X)$ and $g(Y)$ are independent. □

4. (Definition) Joint distribution measure of (X, Y) is given by

   $$\mu_{X,Y}(C) = \mathbb{P}\{(X, Y) \in C\} \text{ for all Borel sets } C \subset \mathbb{R}^2$$

5. (Definition) Joint cumulative distribution function of (X, Y)

   $$F_{X,Y}(a, b) = \mu_{X,Y}((-\infty, a] \times (-\infty, b]) = \mathbb{P}\{X \leqslant a, Y \leqslant b\}, a \in \mathbb{R}, b \in \mathbb{R}$$

6. (Definition) Joint density: We say that a nonnegative, Borel-measurable function $f_{X,Y}(x, y)$ is a joint density for the pair of random variables (X, Y) if

   $$\mu_{X,Y}(C) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{I}_C(x, y) f_{X,Y}(x, y)\,\mathrm{d}x\mathrm{d}y \text{ for all Borel sets } C \subset \mathbb{R}^2$$

7. (Definition) Marginal cumulative distribution function: Firstly we can see the distribution measures of X and Y are

   $$\mu_X(A) = \mathbb{P}\{X \in A\} = \mu_{X,Y}(A \times \mathbb{R}), \forall A \in \mathcal{B}$$

   $$\mu_Y(B) = \mathbb{P}\{X \in B\} = \mu_{X,Y}(\mathbb{R} \times B), \forall B \in \mathcal{B}$$

   and thus, the marginal cumulative distribution functions are

   $$F_X(a) = \mu_X(-\infty, a] = \mathbb{P}\{X \leqslant a\}, \forall a \in \mathbb{R}$$

   $$F_Y(b) = \mu_Y(-\infty, b] = \mathbb{P}\{X \leqslant b\}, \forall b \in \mathbb{R}$$

8. (Definition) Marginal density: If the joint density $f_{X,Y}$ exists, then the marginal densities exist and are given by

   $$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,\mathrm{d}y, \forall x \in \mathbb{R}$$

   $$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,\mathrm{d}x, \forall y \in \mathbb{R}$$

   which are nonnegative, Borel-measurable functions that satisfy

   $$\mu_X(A) = \int_A f_X(x)\,\mathrm{d}\mathcal{L}, \forall A \in \mathcal{B}$$

$$\mu_Y(B) = \int_B f_Y(y)\,\mathrm{d}\mathcal{L}, \forall B \in \mathcal{B}$$

and thus,

$$F_X(a) = \int_{-\infty}^{a} f_X(x)\,\mathrm{d}x, \forall a \in \mathbb{R}$$

$$F_Y(b) = \int_{-\infty}^{b} f_Y(y)\,\mathrm{d}y, \forall b \in \mathbb{R}$$

9. (Theorem) Let X and Y be random variables. The following conditions are equivalent.

   (a) X and Y are independent.

   (b) The joint distribution measure factors:

   $$\mu_{X,Y}(A \times B) = \mu_X(A) \cdot \mu_Y(B) \text{ for all Borel sets } A \subset \mathbb{R}, B \subset \mathbb{R}$$

   Hints: $\mu_{X,Y}(A \times B) = \mathbb{P}\{X \in A, Y \in B\}$.

   (c) The joint cumulative distribution function factors:

   $$F_{X,Y}(a,b) = F_X(a) \cdot F_Y(b), \forall a \in \mathbb{R}, b \in \mathbb{R}$$

   (d) If there is a joint density, the joint density factors:

   $$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) \text{ for almost every } x \in \mathbb{R}, y \in \mathbb{R}$$

10. If X and Y are independent, then

   (a) The expectation factors:

   $$\mathbb{E}[XY] = \mathbb{E}X \cdot \mathbb{E}Y$$

   provided $\mathbb{E}|XY| < \infty$.

   (b) The joint moment-generating function factors:

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

   for all $u \in \mathbb{R}, v \in \mathbb{R}$ for which the expectations are finite.

   Hints: (Standard machine method) For every real-valued, Borel-measurable function $h(x,y) : \mathbb{R}^2 \to \mathbb{R}$, starting with the case when h is the indicator function of a Borel set of $\mathbb{R}^2$, we can proof that

   $$\mathbb{E}|h(X,Y)| = \int_{\mathbb{R}^2} |h(x,y)|\,\mathrm{d}\mu_{X,Y}(x,y)$$

   and if this value is finite, then

   $$\mathbb{E}h(X,Y) = \int_{\mathbb{R}^2} h(x,y)\,\mathrm{d}\mu_{X,Y}(x,y)$$

   If X and Y are independent, then

   $$\mu_{X,Y}(A \times B) = \mu_X(A) \cdot \mu_Y(B) \text{ for all Borel sets } A \subset \mathbb{R}, B \subset \mathbb{R}$$

   furthermore, we have

   $$\mathbb{E}h(X,Y) = \int_{\mathbb{R}^2} h(x,y)\,\mathrm{d}\mu_X(x)\,\mu_Y(y)$$

   Therefore, we can let $h(x,y) = e^{ux+uy}$ or $h(x,y) = xy$ and easily finish the proof.

11. (Definition) Some basic definitions:

    (a) Variance: $\operatorname{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}X)^2\right] = \mathbb{E}X^2 - (\mathbb{E}X)^2$.

    (b) Covariance: $\operatorname{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = \mathbb{E}[XY] - \mathbb{E}X\mathbb{E}Y$.

    (c) Correlation coefficient: $\rho(X, Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}$.

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

$$\mathbb{P}\{Z = 1\} = \frac{1}{2}, \mathbb{P}\{Z = -1\} = \frac{1}{2}$$

Define $Y = ZX$. Show that Y is standard normal. Furthermore, X and Y are uncorrelated, but they are not independent.

Proof. Firstly, we determine the distribution of Y.

$$\begin{aligned} F_Y(b) &= \mathbb{P}\{Y \leqslant b\} = \mathbb{P}(\{ZX \leqslant b\} \cup \Omega) \\ &= \mathbb{P}\{ZX \leqslant b, Z = 1\} + \mathbb{P}\{ZX \leqslant b, Z = -1\} \\ &= \mathbb{P}\{X \leqslant b, Z = 1\} + \mathbb{P}\{X \geqslant -b, Z = -1\} \\ &= \mathbb{P}\{X \leqslant b\} \cdot \frac{1}{2} + [1 - \mathbb{P}\{X < -b\}] \cdot \frac{1}{2} \\ &= \frac{1}{2} \cdot N(b) + \frac{1}{2} \cdot [1 - N(-b)] \\ &= N(b) \end{aligned}$$

where $N(x)$ is the normal CDF, and we used $N(-x) = 1 - N(x)$.

As for the correlation, we compute

$$\operatorname{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}X\mathbb{E}Y = \mathbb{E}\left[ZX^2\right] - 0 = \mathbb{E}Z \cdot \mathbb{E}\left[X^2\right] = 0$$

Furthermore, we compute

$$\mathbb{P}\{|X| \leqslant 1, |Y| \leqslant 1\} = \mathbb{P}\{|X| \leqslant 1, |ZX| \leqslant 1\} = \mathbb{P}\{|X| \leqslant 1\} = N(1) - N(-1)$$

But if X and Y are independent,

$$\mathbb{P}\{|X| \leqslant 1, |Y| \leqslant 1\} = \mathbb{P}\{|X| \leqslant 1\} \cdot \mathbb{P}\{|Y| \leqslant 1\} = [N(1) - N(-1)]^2$$

Therefore, X and Y are not independent. □

Remark: The pair (X, Y) does NOT have a joint density, but they have marginal densities and cumulative distribution function.

Proof. Since $|X| = |Y|$, the pair (X, Y) can only take values in

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

In other words, $\mu_{X,Y}(C) = 1$ and $\mu_{X,Y}(C^c) = 0$. But C has zero area, which means for any nonnegative function f, we must have

$$\int_{\mathbb{R}^2} f(x, y) \,\mathrm{d}\mu_{X,Y} = \int_C f(x, y) \,\mathrm{d}\mu_{X,Y} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{I}_C(x, y) f(x, y) \,\mathrm{d}y\mathrm{d}x = 0$$

If there is a joint density function f, it must satisfy $\int_{\mathbb{R}^2} f(x,y)\,\mathrm{d}\mu_{X,Y} = 1$. Therefore, there cannot be a joint density for (X,Y). However, X and Y have marginal densities because they are normal random variables. As for CDF, it is

$$\begin{aligned}
F_{X,Y}(a,b) &= \mathbb{P}\{X \leqslant a, Y \leqslant b\} \\
&= \mathbb{P}\{X \leqslant a, XZ \leqslant b\} \\
&= \mathbb{P}\{X \leqslant a, XZ \leqslant b, Z = 1\} + \mathbb{P}\{X \leqslant a, XZ \leqslant b, Z = -1\} \\
&= \mathbb{P}\{X \leqslant a, X \leqslant b, Z = 1\} + \mathbb{P}\{X \leqslant a, -X \leqslant b, Z = -1\} \\
&= \mathbb{P}\{X \leqslant \min(a,b)\} \cdot \frac{1}{2} + \mathbb{P}\{-b \leqslant X \leqslant a\} \cdot \frac{1}{2} \\
&= \frac{1}{2} N(\min(a,b)) + \frac{1}{2} \max(N(a) - N(-b), 0)
\end{aligned}$$

□

Here is another example for uncorrelated but dependent R.V.s.

Consider $Y = X^2$, where $X \sim N(0,1)$, then $\mathbb{E}Y = 1$, $\mathbb{E}X = 0$, thus,

$$\operatorname{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}Y\mathbb{E}X = 0 - 0 = 0$$

And they are dependent since $Y = X^2$.

13. (Definition) Two random variables X and Y are said to be jointly normal if they have the joint density

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

where $\sigma_1, \sigma 2 > 0$, $|\rho| < 1$, and $\mu_1, \mu_2 \in \mathbb{R}$.

More generally, a random column vector $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)^{\mathrm{T}}$ is jointly normal if it has joint density

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{C})}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\right\}$$

where $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ is a row vector of dummy variables, $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_n)$ is the row vectors of expectations, and $\boldsymbol{C}$ is the matrix of covariances.

注:（定理，概率论）n 维正态随机变量具有以下重要性质。

(a) 设 $(X_1, X_2, \cdots, X_n)$ 服从 n 维正态分布，则 $X_1, X_2, \cdots, X_n$ 相互独立的**充要条件**是 X_1, $X_2, \cdots, X_n$ 两两不相关。

(b) 设 $(X_1, X_2, \cdots, X_n)$ 服从 n 维正态分布，则 $X_i\,(i = 1, 2, \cdots, n)$ 都服从一维正态分布。

(c) 设 $X_1, X_2, \cdots, X_n$ 服从一维正态分布，且**相互独立**，则 $(X_1, X_2, \cdots, X_n)$ 服从 n 维正态分布。

(d) $(X_1, X_2, \cdots, X_n)$ 服从 n 维正态分布的**充要条件**是对任意不全为零的实数 $l_1, l_2, \cdots, l_n$，线性组合 $l_1X_1 + l_2X_2 + \cdots + X_nl_n$ 服从一维正态分布。

(e)（**线性变换不变性**）设 $(X_1, X_2, \cdots, X_n)$ 服从 n 维正态分布，设随机变量 $Y_1, Y_2, \cdots, Y_k$ 是 $X_1, X_2, \cdots, X_n$ 的线性函数，则 $(Y_1, Y_2, \cdots, Y_k)$ 服从 k 维正态分布。

例：若 $X \sim N(\mu, \sigma)$，则 $Y = \frac{X-\mu}{\sigma} \sim N(0,1)$，这说明一般的正态分布可以通过对随机变量变量线性函数变换为（或者说获得）标准正态分布的随机变量。

14. (Example) Let $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$, and

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

Show that X and W are independent and compute the joint density of (X, W).

Remark: What we have done in this example is having decomposed Y into the linear combination

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

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

Proof. To verify X and W are independent, we can firstly show that they are jointly normal and then have covariance zero. Clearly, W is a linear combination of X, Y, by the theorem given above, X and W are jointly normal. We next compute

$$\begin{aligned}
\operatorname{Cov}(W, X) &= \mathbb{E}[WX] - \mathbb{E}W\mathbb{E}X \\
&= \mathbb{E}\left[XY - \frac{\rho\sigma_2}{\sigma_1} X^2\right] - \mu_1 \mathbb{E}\left[Y - \frac{\rho\sigma_2}{\sigma_1} X\right] \\
&= [\operatorname{Cov}(X, Y) + \mathbb{E}X\mathbb{E}Y] - \frac{\rho\sigma_2}{\sigma_1}\left[\operatorname{Var}(X) + (\mathbb{E}X)^2\right] - \mu_1\mu_2 + \frac{\rho\sigma_2}{\sigma_1}\mu_1^2 \\
&= \rho\sigma_1\sigma_2 + \mu_1\mu_2 - \rho\sigma_2\sigma_1 - \frac{\rho\sigma_2}{\sigma_1}\mu_1^2 - \mu_1\mu_2 + \frac{\rho\sigma_2}{\sigma_1}\mu_1^2 \\
&= 0
\end{aligned}$$

and thus, X and W are independent. Moreover, we have $\mu_3 := \mathbb{E}W = \mu_2 - \frac{\rho\sigma_2}{\sigma_1}\mu_1$,

$$\begin{aligned}
\sigma_3^2 &:= \operatorname{Var}(W) = \mathbb{E}[W - \mathbb{E}W]^2 \\
&= \mathbb{E}\left[Y - \frac{\rho\sigma_2}{\sigma_1} X - \mu_2 + \frac{\rho\sigma_2}{\sigma_1}\mu_1\right]^2 = \mathbb{E}\left[(Y - \mu_2) - \frac{\rho\sigma_2}{\sigma_1}(X - \mu_1)\right]^2 \\
&= \sigma_2^2 - \frac{2\rho\sigma_2}{\sigma_1}\mathbb{E}[(X - \mu_1)(Y - \mu_2)] + \frac{\rho^2\sigma_2^2}{\sigma_1^2}\sigma_1^2 \\
&= \sigma_2^2 - \frac{2\rho\sigma_2}{\sigma_1}\rho\sigma_1\sigma_2 + \frac{\rho^2\sigma_2^2}{\sigma_1^2}\sigma_1^2 \\
&= \sigma_2^2 - \rho^2\sigma_2^2
\end{aligned}$$

Therefore, their joint density is

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

2.3 Conditional Expectations

1. (Remark) Random variable X and sub-σ-algebra $\mathcal{G}$:

(a) X is $\mathcal{G}$-measurable: the information in $\mathcal{G}$ is sufficient to determine the value of X.

(b) X is independent of $\mathcal{G}$: the information in $\mathcal{G}$ provides no help to determine the value of X.

(c) X is dependent of $\mathcal{G}$ but not $\mathcal{G}$-measurable: we can use the information in $\mathcal{G}$ to estimate the value of X.

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

(a) (Measurablity) $\mathbb{E}\left[X|\mathcal{G}\right]$ is $\mathcal{G}$-measurable;

(b) (Partial averaging)

$$\int_A \mathbb{E}\left[X|\mathcal{G}\right](\omega)\,\mathrm{d}\mathbb{P}(\omega) = \int_A X(\omega)\,\mathrm{d}\mathbb{P}(\omega), \forall A \in \mathcal{G}$$

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

Remark: $\mathbb{E}\left[X|\mathcal{G}\right]$ exists and is unique.

注：通过 Radon-Nikodým 导数来定义条件期望。

(a) (定义，测度与概率) 设 $(\Omega, \mathcal{F}, \mathbb{P})$ 是概率空间，σ 代数 $\mathcal{C} \subset \mathcal{F}$，$X$ 的期望存在，则对 $B \in \mathcal{C}$，$\mathbb{E}\left[X\mathbb{I}_B\right]$ 对 $\mathbb{P}_{\mathcal{C}}$ 的 Radon-Nikodým 导数称为 X 在 σ 代数 $\mathcal{C}$ 下（关于 $\mathbb{P}$）的条件期望，记作 $\mathbb{E}\left[X|\mathcal{C}\right]$。

注：$\nu(B) = \mathbb{E}(X\mathbb{I}_B) = \int_B X\mathrm{d}\mathbb{P} = \int_B X\mathrm{d}\mathbb{P}_{\mathcal{C}} \ll \mathbb{P}_{\mathcal{C}}$，显然 $\mathbb{P}_{\mathcal{C}}$ 是 σ 有限的，ν 是 σ 有限、$\mathbb{P}_{\mathcal{C}}$ 连续的，则根据 Radon-Nikodým 定理，存在由 ν 关于 $\mathbb{P}_{\mathcal{C}}$ 几乎唯一决定的 $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mathbb{P}_{\mathcal{C}}}$ 使得

$$\nu(B) = \mathbb{E}(X\mathbb{I}_B) = \int_B f\mathrm{d}\mathbb{P}_{\mathcal{C}}$$

(b) (定理，测度与概率) 设 Ω 为一集合，$(E, \mathcal{E})$ 为可测空间，$f: \Omega \mapsto E$，则 $\varphi: \Omega \mapsto \bar{\mathbb{R}}$ 为 $f^{-1}(\mathcal{E})$ 可测的充分必要条件是有 $(E, \mathcal{E})$ 上的可测函数 g 存在，使得 $\varphi = g \circ f$。如果 φ 有限（有界），则可取 g 有限（有界）。

注：证明需要应用到函数形式的单调类定理。

(c) (推论，测度与概率) 设 Ω 为一集合，随机变量 $Y: \Omega \mapsto \mathbb{R}$，则 $\varphi: \Omega \mapsto \mathbb{R}$ 为 $Y^{-1}(\mathcal{B})$ 可测的充分必要条件是有 $(\mathbb{R}, \mathcal{B})$ 上的可测函数 g 存在，使得 $\varphi = g \circ Y$。

(d) (定义，测度与概率) 在条件期望的定义中，若 $\mathcal{C} = \sigma(Y)$，Y 为随机变量，实际上 $\mathbb{E}\left[X|\,\sigma(Y)\right]$ 是 Ω 到 $\mathbb{R}$ 上 $\sigma(Y)$ 可测的函数，则根据上述定理可知存在一 Borel 可测函数 g 使得

$$\mathbb{E}\left[X|\,\sigma(Y)\right] = g(Y), \text{a.e.}$$

称之为 X 在 Y 之下的条件期望，记作 $\mathbb{E}\left[X|\,Y\right]$，于是

$$\mathbb{E}\left[X|\,Y = y\right] := g(y)$$

(e) (定义，测度与概率) $\mathbb{P}(A) = \mathbb{E}\left[\mathbb{I}_A\right]$，称 $\mathbb{E}\left[\mathbb{I}_A|\,\mathcal{C}\right]$ 为事件 $A \in \mathcal{F}$ 在 $\mathcal{C}$ 下的条件概率，记作 $\mathbb{P}(A|\,\mathcal{C})$。相应地，称 $\mathbb{E}\left[\mathbb{I}_A|\,Y\right]$ 为事件 $A \in \mathcal{F}$ 在 $\sigma(Y)$ 或 Y 下的条件概率，记作 $\mathbb{P}(A|\,Y)$。

(f) (定理，测度与概率) 条件期望的简单性质：

i. 若 $\mathbb{E}X$ 存在，则 $\mathbb{E}[\mathbb{E}[X|\mathcal{C}]]$ 存在，且 $\forall B\in\mathcal{C}$，有

$$\mathbb{E}[X\mathbb{I}_B] = \int_B \mathbb{E}[X|\mathcal{C}]\,\mathrm{d}\mathbb{P}_{\mathcal{C}} = \int_B \mathbb{E}[X|\mathcal{C}]\,\mathrm{d}\mathbb{P}$$

因而 $\mathbb{E}[\mathbb{E}[X|\mathcal{C}]] = \mathbb{E}X$；若 X 可积，则 $\mathbb{E}[X|\mathcal{C}]$ $\mathbb{P}_{\mathcal{C}}$ a.e. 有限。

ii. 若 $\mathcal{C}=\mathcal{F}$ 或 X 为 $\mathcal{C}$ 可测时，则 $\mathbb{E}[X|\mathcal{C}] = X, \mathbb{P}_{\mathcal{C}}$ a.e.。

(g) (定义，测度与概率) 给定可测空间 $(\Omega,\mathcal{F})$，称 $A\in\mathcal{F}$ 为 $\mathcal{F}$ 的非空原子，如果元素 A 除本身和空集外不包含其他 $\mathcal{F}$ 可测子集，即 $\forall B\in\mathcal{F}$，$A\cap B=A$或$\varnothing$。

(h) (定理，测度与概率) 定理：设 $(\Omega,\mathcal{F},\mathbb{P})$ 是概率空间，σ 代数 $\mathcal{C}\subset\mathcal{F}$，$X$ 的期望存在，则在 $\mathcal{C}$ 的每一非空原子 B 上，$\mathbb{E}[X|\mathcal{C}]$ 为常数，记为 $\mathbb{E}[X|B]$ 且

$$\int_B X\mathrm{d}\mathbb{P} = \mathbb{E}[X|\mathcal{C}]\cdot\mathbb{P}(B), \forall\omega\in B$$

注：若 $\mathbb{P}(B)>0$，则

$$\mathbb{E}[X|\mathcal{C}] = \frac{\int_B X\mathrm{d}\mathbb{P}}{\mathbb{P}(B)} = \frac{\mathbb{E}[X\mathbb{I}_B]}{\mathbb{P}(B)}$$

这说明 $\mathbb{E}[X|\mathcal{C}]$ 在 $\mathcal{C}$ 的原子上是 X 的平均值。

(i) (推论,测度与概率) 设 $\{B_n, n\in\mathbb{N}\}$ 是 Ω 的一个可数划分,则 B_n 是原子,$\mathcal{C}:=\sigma(B_n,n\in\mathbb{N})$,则对于任意期望存在的随机变量 X 有

$$\mathbb{E}[X|\mathcal{C}] = \sum_{n=1}^{\infty}\mathbb{E}[X|B_n]\,\mathbb{I}_{B_n}, \forall\omega\in\Omega$$

其中当 $\mathbb{P}(B_n)>0$，$\mathbb{E}[X|B_n] = \frac{1}{\mathbb{P}(B_n)}\int_{B_n}X\mathrm{d}\mathbb{P}$；当 $\mathbb{P}(B_n)>0$，$\mathbb{E}[X|B_n]$ 取任意常数。

3. (Theorem) Properties of Conditional Expectation: Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$.

(a) (**Linearity**) If X and Y are integrable random variable and c_1 and c_2 are constants, then

$$\mathbb{E}[c_1X + c_2Y|\mathcal{G}] = c_1\mathbb{E}[X|\mathcal{G}] + c_2\mathbb{E}[Y|\mathcal{G}]$$

This equation also holds if we assume that X and Y are nonnegative (rather than integrable) and c_1 and c_2 are positive, although both sides may be ∞.

(b) (**Taking out what is known**) If X and Y are integrable random variable, Y and XY are integrable, and X is $\mathcal{G}$-measurable, then

$$\mathbb{E}[XY|\mathcal{G}] = X\mathbb{E}[Y|\mathcal{G}]$$

This equation also holds if we assume that X is positive and Y is nonnegative (rather than integrable), although both sides may be ∞.

(c) (**Iterated conditioning**) If $\mathcal{H}\subset\mathcal{G}\subset\mathcal{F}$, and X is an integrable random variable, then

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

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

(d) (**Independence**) If X is integrable and independent of $\mathcal{G}$, then

$$\mathbb{E}\left[X|\mathcal{G}\right]=\mathbb{E}X$$

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

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

$$\mathbb{E}\left[\varphi\left(X\right)|\,\mathcal{G}\right]\geqslant\varphi\left(\mathbb{E}\left[X|\mathcal{G}\right]\right)$$

4. (Remark) $\mathbb{E}\left[X|\mathcal{G}\right]$ is an unbaised estimator of X: $\mathbb{E}\left(\mathbb{E}\left[X|\mathcal{G}\right]\right)=\mathbb{E}X$, because of the partial average property.

$$\int_{\Omega}\mathbb{E}\left[X|\mathcal{G}\right]\left(\omega\right)\mathrm{d}\mathbb{P}\left(\omega\right)=\int_{\Omega}X\left(\omega\right)\mathrm{d}\mathbb{P}\left(\omega\right)$$

5. (Theorem) Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Suppose the random variables $X_1,X_2,\cdots,X_K$ are $\mathcal{G}$-measurable and the random variables $Y_1,Y_2,\cdots,Y_L$ are independent of $\mathcal{G}$. Let $f\left(x_1,x_2,\cdots,x_K,y_1,y_2,\cdots,y_L\right)$ be the function of dummy variables $x_1,x_2,\cdots,x_K$ and $y_1,y_2,\cdots,y_L$, and define

$$g\left(x_1,x_2,\cdots,x_K\right):=\mathbb{E}\left[f\left(x_1,x_2,\cdots,x_K,Y_1,Y_2,\cdots,Y_L\right)\right]$$

Then

$$\begin{aligned}\mathbb{E}\left[f\left(X_1,X_2,\cdots,X_K,Y_1,Y_2,\cdots,Y_L\right)|\mathcal{G}\right]&=g\left(X_1,X_2,\cdots,X_K\right)\\&=\mathbb{E}\left[f\left(X_1,X_2,\cdots,X_K,Y_1,Y_2,\cdots,Y_L\right)\right]\end{aligned}$$

Remark: The proof needs more measure-theoretical ideas.

6. (Example) Let $(X,Y)\sim N\left(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho\right)$, and

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

We have $(X,W)\sim N\left(\mu_1,\mu_3,\sigma_1,\sigma_3,0\right)$, where $\mu_3=\mu_2-\frac{\rho\sigma_2}{\sigma_1}\mu_1$, $\sigma_3^2=\sigma_2^2\left(1-\rho^2\right)$. And thus X and W are independent. Compute $\mathbb{E}\left[Y|\,X\right]$ and $\mathbb{E}\left[f\left(X,Y\right)|\,X\right]$.

Solution.

$$\mathbb{E}\left[Y|\,X\right]=\mathbb{E}\left[W+\frac{\rho\sigma_2}{\sigma_1}X\middle|\,X\right]=\frac{\rho\sigma_2}{\sigma_1}X+\mathbb{E}W=\frac{\rho\sigma_2}{\sigma_1}\left(X-\mu_1\right)+\mu_2$$

which is $\sigma\left(X\right)$- measurable. In order to compute $\mathbb{E}\left[f\left(X,Y\right)|\,X\right]$, we firstly define

$$\begin{aligned}g\left(x\right)&=\mathbb{E}h\left(x,Y\right)=\mathbb{E}f\left(x,\frac{\rho_1\sigma_2}{\sigma_1}x+W\right)\\&=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}f\left(x,\frac{\rho\sigma_1}{\sigma_2}x+w\right)\exp\left\{-\frac{\left(w-\mu_3\right)^2}{2\sigma_3^2}\right\}\mathrm{d}w\end{aligned}$$

Then

$$\mathbb{E}\left[f\left(X,Y\right)|\,X\right]=g\left(X\right)$$

which is $\sigma\left(X\right)$- measurable, as it should be. □

7. (Definition) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $T > 0$. Let $\mathcal{F}(t), 0 \leqslant t \leqslant T$, be a filtration of sub-σ-algebras of $\mathcal{F}$. Consider an adapted stochastic process $\{M(t), 0 \leqslant t \leqslant T\}$.

   (a) If
   $$\mathbb{E}[M(t)|\mathcal{F}(s)] = M(s), 0 \leqslant s \leqslant t \leqslant T$$
   we say M is a martingale. It has no tendency to rise or fall.

   (b) If
   $$\mathbb{E}[M(t)|\mathcal{F}(s)] \geqslant M(s), 0 \leqslant s \leqslant t \leqslant T$$
   we say M is a sub-martingale. It has no tendency to fall, but may have a tendency to rise.

   (c) If
   $$\mathbb{E}[M(t)|\mathcal{F}(s)] \leqslant M(s), 0 \leqslant s \leqslant t \leqslant T$$
   we say M is a sup-martingale. It has no tendency to rise, but may have a tendency to fall.

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

   Remark: If we indicate the dependent on time rather than the sample point ω, we can rewrite $f(x)$, $g(x)$ as $f(t,x)$, $f(s,x)$. And thus,
   $$\mathbb{E}[f(t, X(t))|\mathcal{F}(s)] = f(s, X(s)), 0 \leqslant s \leqslant t \leqslant T$$
   which implies that f satisfies a partial differential equation. The PDE gives us a way to determine $f(s,x)$ if we know $f(t,x)$. The Black-Scholes-Merton partial differential equation is a special case of this.

2.4 Selected Exercise

Exercise 2.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and suppose a random variable X on this space is measurable with respect to the trivial σ-algebra $\mathcal{F}_0 = \{\varnothing, \Omega\}$. Show that X is not random, i.e., there is a constant c such that $\forall \omega \in \Omega, X(\omega) = c$. Such a random variable is called *degenerate*.

Proof. X is $\mathcal{F}_0$-measurable, then $\forall a \in \mathbb{R}, \{X \leqslant a\} \in \mathcal{F}_0 = \{\varnothing, \Omega\}$. So
$$F(x) = \mathbb{P}\{X \leqslant x\} = 0 \text{ or } 1$$
while
$$\lim_{n \to -\infty} F(x) = 0, \lim_{n \to +\infty} F(x) = 1$$
and because $F(x) = \mathbb{P}\{X \leqslant x\}$ is monotone increasing function, there exists x_0 such that
$$F(x) = \begin{cases} 0, & x \leqslant x_0 \\ 1, & x > x_0 \end{cases}$$

Let $A_n = \left\{X \geqslant x_0 - \frac{1}{n}\right\}$, $A = \{X > x_0\}$, so

$$A = \bigcap_{n=1}^{\infty} A_n$$

or $A_n \downarrow A, n \to \infty$, and since $\mathbb{P}(A_1) = \mathbb{P}\{X \geqslant x_0 - 1\} = 0 < \infty$, we have

$$\begin{aligned}
\mathbb{P}\{X = x_0\} &= \mathbb{P}\{X \geqslant x_0\} - \mathbb{P}\{X > x_0\} \\
&= \mathbb{P}\{X \geqslant x_0\} - \mathbb{P}\left\{\bigcap_{n=1}^{\infty}\left\{X \geqslant x_0 - \frac{1}{n}\right\}\right\} \\
&= \mathbb{P}\{X \geqslant x_0\} - \lim_{n\to\infty}\mathbb{P}\left\{X \geqslant x_0 - \frac{1}{n}\right\} \\
&= 1 - 0 \\
&= 1
\end{aligned}$$

Since $\{X = x_0\} \in \{\varnothing, \Omega\}$, we can conclude that $\{X = x_0\} = \Omega$, i.e., $\exists x_0 \in \mathbb{R}, \forall \omega \in \Omega, X(\omega) = x_0$. Therefore, X is degenerate. □

Exercise 2.2. (**Independence of random variables can be affected by change of measure**) To illustrate the point, consider the space of two coin tosses $\Omega_2 = \{HH, HT, TH, TT\}$, and let stock prices be given by

$$S_0 = 4, S_1(H) = 8, S_1(T) = 2, S_2(HH) = 16, S_2(HT) = S_2(TH) = 4, S_2(TT) = 1$$

Consider the probability measure $\mathbb{P}$ given by

$$\mathbb{P}(HH) = \frac{4}{9}, \mathbb{P}(HT) = \frac{2}{9}, \mathbb{P}(TH) = \frac{2}{9}, \mathbb{P}(TT) = \frac{1}{9}$$

and $\tilde{\mathbb{P}}$ given by

$$\tilde{\mathbb{P}}(HH) = \frac{1}{4}, \tilde{\mathbb{P}}(HT) = \frac{1}{4}, \tilde{\mathbb{P}}(TH) = \frac{1}{4}, \tilde{\mathbb{P}}(TT) = \frac{1}{4}$$

Define the random variable

$$X = \begin{cases} 1, & S_2 = 4 \\ 0, & S_2 \neq 4 \end{cases}$$

1. List all the sets in $\sigma(X)$.

   Solution. This is a simple case, we do not need to (or cannot) consider all Borel sets in $\mathbb{R}$.

   $$\begin{aligned}
   B_1 = \{0\} &\Longrightarrow \{S_2 \in B_1\} = \{HH, TT\} \\
   B_2 = \{1\} &\Longrightarrow \{S_2 \in B_2\} = \{HT, TH\} \\
   B_3 = \varnothing &\Longrightarrow \{S_2 \in B_3\} = \varnothing \\
   B_4 = \{0, 1\} &\Longrightarrow \{S_2 \in B_4\} = \Omega_2
   \end{aligned}$$

   Therefore, $\sigma(X) = \{\Omega_2, \varnothing, \{TH, HT\}, \{TT, HH\}\}$ □

2. List all the sets in $\sigma(S_1)$.

   Solution. Similar to 1, we have $\sigma(S_1) = \{\Omega_2, \varnothing, \{TH, TT\}, \{HT, HH\}\}$ □

	$\sigma(X)$	0	1	1/2	1/2
$\sigma(S_1)$		$\varnothing$	Ω	$\{HT, TH\}$	$\{HH, TT\}$
0	$\varnothing$	0	0	0	0
1	Ω	0	1	1/2	1/2
1/2	$\{HH, HT\}$	0	1/2	1/4	1/4
1/2	$\{TT, TH\}$	0	1/2	1/4	1/4

3. Show that $\sigma(X)$ and $\sigma(S_1)$ are independent under $\tilde{\mathbb{P}}$.

Proof. Compute as below.

Therefore, $\sigma(X)$ and $\sigma(S_1)$ are not independent under $\tilde{\mathbb{P}}$. □

4. Show that $\sigma(X)$ and $\sigma(S_1)$ are independent under $\mathbb{P}$.

Proof. Compute as below.

	$\sigma(X)$	0	1	4/9	5/9
$\sigma(S_1)$		$\varnothing$	Ω	$\{HT, TH\}$	$\{HH, TT\}$
0	$\varnothing$	0	0	0	0
1	Ω	0	1	4/9	5/9
6/9	$\{HH, HT\}$	0	6/9	$\frac{4}{9}\times\frac{6}{9}\neq\frac{8}{9}$	$\frac{5}{9}\times\frac{6}{9}\neq\frac{7}{9}$
3/9	$\{TT, TH\}$	0	3/9	$\frac{3}{9}\times\frac{4}{9}\neq\frac{5}{9}$	$\frac{5}{9}\times\frac{3}{9}\neq\frac{7}{9}$

Therefore, $\sigma(X)$ and $\sigma(S_1)$ are not independent under $\mathbb{P}$. □

5. Under $\mathbb{P}$, we have $\mathbb{P}\{S_1 = 8\} = \frac{2}{3}$ and $\mathbb{P}\{S_1 = 2\} = \frac{1}{3}$. Explain intuitively why, if you are told that $X = 1$, you would want to revise your estimate of the distribution of S_1.

Solution. Because S_1 and X are not independent under the probability measure $\mathbb{P}$, knowing the value of S_1 will affect our opinion on the distribution of S_1. And the conditional distribution of S_1 under $X = 1$ is $\mathbb{P}(S_1 = 8 | X = 1) = \mathbb{P}(S_1 = 2 | X = 1) = \frac{1}{2}$. Therefore, we have

$$\begin{aligned} \mathbb{E}S_1 &= \frac{2}{3}\cdot 8 + \frac{1}{3}\cdot 2 = 6 \\ \mathbb{E}[S_1 | S_2 = 4] &= \frac{1}{2}\cdot 8 + \frac{1}{2}\cdot 2 = 5 \end{aligned}$$

□

Exercise 2.3. (**Rotating the axes**) Let X and Y be independent standard normal random variables. Let θ be a constant, and define

$$V = X\cos\theta + Y\sin\theta \text{ and } W = -X\sin\theta + Y\cos\theta$$

Show that V and W are also independent standard normal random variables.

Proof. Because X and Y be independent standard normal random variables, so (X, W) and their linear combination are joint normal variable. Therefore, what we need to show is that they are uncorrelated.

$$\begin{aligned}
\mathbb{E}[VW] &= \mathbb{E}[(X\cos\theta + Y\sin\theta)(-X\sin\theta + Y\cos\theta)] \\
&= \mathbb{E}\left[-X^2\cos\theta\sin\theta + Y^2\cos\theta\sin\theta - XY\sin^2\theta + XY\cos^2\theta\right] \\
&= -\cos\theta\sin\theta + \cos\theta\sin\theta \\
&= 0
\end{aligned}$$

□

Exercise 2.4. Let X be a standard normal variable and let Z be independent of X and satisfy

$$\mathbb{P}\{Z=1\} = \frac{1}{2}, \mathbb{P}\{Z=-1\} = \frac{1}{2}$$

Define $Y = ZX$. Use moment-generating functions to show that Y is standard normal, X and Y are uncorrelated, but they are not independent.

1. Establish the joint moment-generating function formula, i.e., compute $\mathbb{E}e^{uX+vY}$.

   Solution. The computation needs to decompose expectation to many conditional expectations.

$$\begin{aligned}
\mathbb{E}e^{uX+vY} &= \mathbb{E}\left[e^{uX+vZX}\right] \\
&= \mathbb{E}\left[e^{uX+vZX}|Z=1\right]\cdot\mathbb{P}(Z=1) + \mathbb{E}\left[e^{uX+vZX}|Z=-1\right]\cdot\mathbb{P}(Z=-1) \\
&= \frac{1}{2}\mathbb{E}\left[e^{uX+vX}\right] + \frac{1}{2}\mathbb{E}\left[e^{uX-vX}\right] \\
&= \frac{1}{2}\left[e^{\frac{(u+v)^2}{2}} + e^{\frac{(u-v)^2}{2}}\right] \\
&= \frac{1}{2}\left[e^{\frac{u^2+v^2}{2}}\cdot e^{uv} + e^{\frac{u^2+v^2}{2}}\cdot e^{-uv}\right] \\
&= e^{\frac{u^2+v^2}{2}}\cdot\frac{e^{uv}+e^{-uv}}{2}
\end{aligned}$$

   □

2. Use the joint moment-generating function to show that Y is standard normal.

   Proof. Let $u=0$, we have $\mathbb{E}e^{vY} = e^{\frac{v^2}{2}}$, which means Y is standard normal. □

3. Show that X and Y are not independent.

   Proof.

$$\mathbb{E}e^{uX}\cdot\mathbb{E}e^{vY} = e^{\frac{u^2+v^2}{2}} \neq e^{\frac{u^2+v^2}{2}}\cdot\frac{e^{uv}+e^{-uv}}{2} = \mathbb{E}e^{uX+vY}$$

   □

Exercise 2.5. Let (X, W) be a pair of random variables with jointly density function

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

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

Proof. The density $f_X(x)$ can be obtained by

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y = \int_{-|x|}^{\infty} \frac{2|x|+y}{\sqrt{2\pi}} e^{-\frac{(2|x|+y)^2}{2}} \mathrm{d}y \\
&= \int_{|x|}^{\infty} \frac{u}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \mathrm{d}u = \int_{\frac{|x|^2}{2}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-v} \mathrm{d}v \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}
\end{aligned}
$$

Note that $\max\{x,0\} = -\min\{-x,0\}$, then

$$
\begin{aligned}
f_Y(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}x = \int_{-\infty}^{\infty} \mathbb{I}_{\{|x|\geqslant -y\}} \frac{2|x|+y}{\sqrt{2\pi}} e^{-\frac{(2|x|+y)^2}{2}} \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \mathbb{I}_{\{x\geqslant -y, x\geqslant 0\}} \frac{2x+y}{\sqrt{2\pi}} e^{-\frac{(2x+y)^2}{2}} \mathrm{d}x + \int_{-\infty}^{\infty} \mathbb{I}_{\{x\leqslant y, x<0\}} \frac{-2x+y}{\sqrt{2\pi}} e^{-\frac{(-2x+y)^2}{2}} \mathrm{d}x \\
&= \int_{\max\{-y,0\}}^{\infty} \frac{2x+y}{\sqrt{2\pi}} e^{-\frac{(2x+y)^2}{2}} \mathrm{d}x + \int_{-\infty}^{\min\{y,0\}} \frac{-2x+y}{\sqrt{2\pi}} e^{-\frac{(-2x+y)^2}{2}} \mathrm{d}x \\
&= \int_{\max\{-y,0\}}^{\infty} \frac{2x+y}{\sqrt{2\pi}} e^{-\frac{(2x+y)^2}{2}} \mathrm{d}x + \int_{-\min\{y,0\}}^{\infty} \frac{2u+y}{\sqrt{2\pi}} e^{-\frac{(2u+y)^2}{2}} \mathrm{d}u \\
&= 2\int_{\max\{-y,0\}}^{\infty} \frac{2x+y}{\sqrt{2\pi}} e^{-\frac{(2x+y)^2}{2}} \mathrm{d}x = \int_{|y|}^{\infty} \frac{t}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t \\
&= \int_{-\frac{|y|^2}{2}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-v} \mathrm{d}v \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}
\end{aligned}
$$

But

$$
f_{X,Y}(x,y) \neq f_X(x) \cdot f_Y(y)
$$

So, X and Y are not independent. Let

$$
F(x) = \int_{x}^{\infty} \frac{t^2}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t
$$

And we have

$$\begin{aligned}
\mathbb{E}[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathbb{I}_{\{|x|\geqslant -y\}}xy\frac{2|x|+y}{\sqrt{2\pi}}e^{-\frac{(2|x|+y)^2}{2}}\mathrm{d}x\mathrm{d}y\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathbb{I}_{\{|x|\geqslant -y\}}xy\frac{2|x|+y}{\sqrt{2\pi}}e^{-\frac{(2|x|+y)^2}{2}}\mathrm{d}y\mathrm{d}x\\
&= \int_{-\infty}^{\infty}x\mathrm{d}x\int_{-|x|}^{\infty}y\frac{2|x|+y}{\sqrt{2\pi}}e^{-\frac{(2|x|+y)^2}{2}}\mathrm{d}y\\
&= \int_{-\infty}^{\infty}x\mathrm{d}x\int_{|x|}^{\infty}(u-2|x|)\frac{u}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\mathrm{d}u\,(\text{Let } u=2|x|+y)\\
&= \int_{-\infty}^{\infty}x\mathrm{d}x\left(\int_{|x|}^{\infty}\frac{u^2}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\mathrm{d}u-\int_{|x|}^{\infty}\frac{2|x|u}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\mathrm{d}u\right)\\
&= \int_{-\infty}^{\infty}x\mathrm{d}x\left(\int_{|x|}^{\infty}\frac{u^2}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\mathrm{d}u-\frac{|x|}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right)\\
&= \int_{-\infty}^{\infty}x\mathrm{d}x\int_{|x|}^{\infty}\frac{u^2}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\mathrm{d}u+\int_{-\infty}^{\infty}x\left(-\frac{|x|}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right)\mathrm{d}x\\
&= \int_{0}^{\infty}x\mathrm{d}x\int_{x}^{\infty}\frac{u^2}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\mathrm{d}u+\int_{-\infty}^{0}x\mathrm{d}x\int_{-x}^{\infty}\frac{u^2}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\mathrm{d}u\\
&\quad -\int_{0}^{\infty}\frac{x^2}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\mathrm{d}x+\int_{-\infty}^{0}\frac{x^2}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\mathrm{d}x\\
&= \int_{0}^{\infty}x\mathrm{d}x\int_{x}^{\infty}\frac{u^2}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\mathrm{d}u+\int_{-\infty}^{0}x\mathrm{d}x\int_{-x}^{\infty}\frac{u^2}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\mathrm{d}u+0\\
&= \int_{0}^{\infty}xF(x)\mathrm{d}x+\int_{-\infty}^{0}xF(-x)\mathrm{d}x\\
&=0
\end{aligned}$$

And thus, X and Y are uncorrelated. □

Exercise 2.6. Consider a probability space Ω with four elements, which we call a, b, c, and d (i.e., $\Omega=\{a,b,c,d\}$). We define $\mathbb{P}$ by specifying that

$$\mathbb{P}\{a\}=\frac{1}{6},\mathbb{P}\{b\}=\frac{1}{3},\mathbb{P}\{c\}=\frac{1}{4},\mathbb{P}\{d\}=\frac{1}{4}$$

We next define random variables X, Y and Z by

$$X(\omega)=\begin{cases}1, & \omega\in\{a,b\}\\ -1, & \omega\in\{c,d\}\end{cases},Y(\omega)=\begin{cases}1, & \omega\in\{a,c\}\\ -1, & \omega\in\{c,d\}\end{cases}$$

And define

$$Z=X+Y$$

1. List the sets in $\sigma(X)$.

   Solution. $\sigma(X)=\{\varnothing,\Omega,\{a,b\},\{c,d\}\}$ □

2. Determine $\mathbb{E}[Y|X]$. Verify that the partial-averaging property is satisfied.

Solution. We firstly determine $\mathbb{E}\left[Y \middle| X\right]$

$$
\begin{aligned}
\mathbb{E}\left[Y|X\right] &= \sum_{\alpha\in\{-1,1\}} \mathbb{E}\left[Y|X=\alpha\right]\mathbb{I}_{\{X=\alpha\}} \\
&= \sum_{\alpha\in\{-1,1\}} \frac{\mathbb{E}\left[Y\cdot\mathbb{I}_{\{X=\alpha\}}\right]}{\mathbb{P}\left(X=\alpha\right)}\mathbb{I}_{\{X=\alpha\}} \\
&=\frac{\mathbb{E}\left[Y\cdot\mathbb{I}_{\{X=1\}}\right]}{\mathbb{P}\left(X=1\right)}\mathbb{I}_{\{X=1\}}+\frac{\mathbb{E}\left[Y\cdot\mathbb{I}_{\{X=-1\}}\right]}{\mathbb{P}\left(X=-1\right)}\mathbb{I}_{\{X=-1\}} \\
&=\frac{1\cdot\mathbb{P}\left(X=1,Y=1\right)-\mathbb{P}\left(X=1,Y=-1\right)}{\mathbb{P}\left(X=1\right)}\mathbb{I}_{\{X=1\}} \\
&\quad+\frac{1\cdot\mathbb{P}\left(X=-1,Y=1\right)-\mathbb{P}\left(X=-1,Y=-1\right)}{\mathbb{P}\left(X=-1\right)}\mathbb{I}_{\{X=-1\}} \\
&=\frac{\mathbb{P}\left\{a\right\}-\mathbb{P}\left\{b\right\}}{\mathbb{P}\left\{a,b\right\}}\mathbb{I}_{\{X=1\}}+\frac{\mathbb{P}\left\{c\right\}-\mathbb{P}\left\{d\right\}}{\mathbb{P}\left\{c,d\right\}}\mathbb{I}_{\{X=-1\}} \\
&=\frac{\frac{1}{6}-\frac{1}{3}}{\frac{1}{2}}\mathbb{I}_{\{X=1\}} \\
&=-\frac{1}{3}\mathbb{I}_{\{X=1\}}
\end{aligned}
$$

(Review: Partial averaging)

$$
\int_A \mathbb{E}\left[Y|X\right](\omega)\,\mathrm{d}\mathbb{P}(\omega)=\int_A Y(\omega)\,\mathrm{d}\mathbb{P}(\omega)\quad\text{for all } A\in\sigma(X)
$$

$$
\begin{aligned}
\mathbb{E}\left[\mathbb{E}\left[Y|X\right]\mathbb{I}_{\{X=1\}}\right] &= \mathbb{E}\left[-\frac{1}{3}\mathbb{I}_{\{X=1\}}\cdot\mathbb{I}_{\{X=1\}}\right]=-\frac{1}{3}\mathbb{E}\left[\mathbb{I}_{\{X=1\}}\right] \\
&=-\frac{1}{3}\cdot 1\cdot\mathbb{P}\left\{X=1\right\}=-\frac{1}{3}\cdot\mathbb{P}\left\{a,b\right\} \\
&=-\frac{1}{6} \\
\mathbb{E}\left[Y\mathbb{I}_{\{X=1\}}\right] &= \mathbb{P}\left\{X=1,Y=1\right\}-\mathbb{P}\left\{X=1,Y=-1\right\} \\
&=\mathbb{P}\left\{a\right\}-\mathbb{P}\left\{b\right\} \\
&=-\frac{1}{6}
\end{aligned}
$$

Thus,

$$
\mathbb{E}\left[\mathbb{E}\left[Y|X\right]\mathbb{I}_{\{X=1\}}\right]=\mathbb{E}\left[Y\mathbb{I}_{\{X=1\}}\right]
$$

Similarly,

$$
\mathbb{E}\left[\mathbb{E}\left[Y|X\right]\mathbb{I}_{\{X=-1\}}\right]=\mathbb{E}\left[Y\mathbb{I}_{\{X=-1\}}\right]=0
$$

$$
\mathbb{E}\left[\mathbb{E}\left[Y|X\right]\mathbb{I}_{\varnothing}\right]=\mathbb{E}\left[Y\mathbb{I}_{\varnothing}\right]=0
$$

$$
\mathbb{E}\left[\mathbb{E}\left[Y|X\right]\mathbb{I}_{\Omega}\right]=\mathbb{E}\left[Y\mathbb{I}_{\Omega}\right]=0
$$

In fact,

$$
\mathbb{E}\left[Y|X=1\right]=1\times\frac{1}{6}-1\times\frac{1}{3}=-\frac{1}{6}\mathbb{E}\left[Y|X=-1\right]=1\times\frac{1}{4}-1\times\frac{1}{4}=0
$$

$\square$

3. Determine $\mathbb{E}\left[Z \middle| X\right]$. Verify that the partial-averaging property is satisfied.

Solution.

$$Z(a)=2, Z(b)=Z(c)=0, Z(d)=-2$$

$$\begin{aligned}
\mathbb{E}[Z|X] = \sum_{\alpha\in\{-1,1\}} \mathbb{E}[Z|X=\alpha]\,\mathbb{I}_{\{X=\alpha\}} &= \sum_{\alpha\in\{-1,1\}} \frac{\mathbb{E}\left[Z\cdot\mathbb{I}_{\{X=\alpha\}}\right]}{\mathbb{P}(X=\alpha)}\mathbb{I}_{\{X=\alpha\}} \\
&=\frac{2\cdot\mathbb{P}(X=1,Z=2)-2\mathbb{P}(X=1,Z=-2)}{\mathbb{P}(X=1)}\mathbb{I}_{\{X=1\}} \\
&\quad+\frac{2\cdot\mathbb{P}(X=-1,Z=2)-2\mathbb{P}(X=-1,Z=-2)}{\mathbb{P}(X=-1)}\mathbb{I}_{\{X=-1\}} \\
&=\frac{2\cdot\mathbb{P}\{a\}-0}{\mathbb{P}\{a,b\}}\mathbb{I}_{\{X=1\}}+\frac{0-2\cdot\mathbb{P}\{d\}}{\mathbb{P}\{c,d\}}\mathbb{I}_{\{X=-1\}} \\
&=\frac{2\cdot\frac{1}{6}}{\frac{1}{2}}\mathbb{I}_{\{X=1\}}-\frac{2\cdot\frac{1}{4}}{\frac{1}{2}}\mathbb{I}_{\{X=-1\}} \\
&=\frac{2}{3}\mathbb{I}_{\{X=1\}}-\mathbb{I}_{\{X=-1\}}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}[Z|X]\,\mathbb{I}_{\{X=1\}}\right] &=\mathbb{E}\left[\left(\frac{2}{3}\mathbb{I}_{\{X=1\}}-\mathbb{I}_{\{X=-1\}}\right)\cdot\mathbb{I}_{\{X=1\}}\right]=\frac{2}{3}\mathbb{E}\left[\mathbb{I}_{\{X=1\}}\right] \\
&=\frac{2}{3}\cdot 1\cdot\mathbb{P}(X=1)=\frac{2}{3}\cdot\frac{1}{2} \\
&=\frac{1}{3} \\
\mathbb{E}\left[Z\mathbb{I}_{\{X=1\}}\right] &=2\cdot\mathbb{P}(X=1,Z=2)-2\cdot\mathbb{P}(X=1,Y=-2) \\
&=2\cdot\mathbb{P}\{a\}-0 \\
&=\frac{1}{3}
\end{aligned}$$

Verification of the partial-averaging property is skipped.

□

4. Determine $\mathbb{E}[Z|\,X]-\mathbb{E}[Y|\,X]$.

Solution. Give two ways for solution.

$$\mathbb{E}[Z|X]-\mathbb{E}[Y|X]=\frac{2}{3}\mathbb{I}_{\{X=1\}}-\mathbb{I}_{\{X=-1\}}+\frac{1}{3}\mathbb{I}_{\{X=1\}}=\mathbb{I}_{\{X=1\}}-\mathbb{I}_{\{X=-1\}}=X$$

$$\mathbb{E}[Z|X]-\mathbb{E}[Y|X]=\mathbb{E}[(Z-Y)\,|X]=\mathbb{E}[X|X]=X$$

□

Exercise 2.7. Let Y be an integrable random variable on a probability space $(\Omega,\mathcal{F},\mathbb{P})$ and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Based on the information in $\mathcal{G}$, we can form the estimate $\mathbb{E}[Y|\,\mathcal{G}]$ of Y and define the error of the estimation $\mathrm{Err}=Y-\mathbb{E}[Y|\,\mathcal{G}]$, which has expectation zero and some variance $\mathrm{Var}(\mathrm{Err})$. Let X be some other $\mathcal{G}$-measurable random variable, which we can regard as another estimate of Y. Show that

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

Hint: Let $\mu = \mathbb{E}(Y - X)$. Compute the variance of $Y - X$ as

$$\mathbb{E}\left[(Y - X - \mu)^2\right] = \mathbb{E}\left[((Y - \mathbb{E}[Y|\mathcal{G}]) + (\mathbb{E}[Y|\mathcal{G}] - X - \mu))^2\right]$$

Multiply out the right-hand side and use iterated conditioning to show the cross-term is zero.

Proof.

$$\begin{aligned}
\operatorname{Var}(Y - X) =& \mathbb{E}\left[(Y - X - \mu)^2\right] \\
=& \mathbb{E}\left[((Y - \mathbb{E}[Y|\mathcal{G}]) + (\mathbb{E}[Y|\mathcal{G}] - X - \mu))^2\right] \\
=& \operatorname{Var}(\mathrm{Err}) + \mathbb{E}\left[(\mathbb{E}[Y|\mathcal{G}] - X - \mu)^2\right] \\
& + 2\mathbb{E}\left[(Y - \mathbb{E}[Y|\mathcal{G}])(\mathbb{E}[Y|\mathcal{G}] - X - \mu)\right]
\end{aligned}$$

Let random variable $\xi := \mathbb{E}[Y - X - \mu|\mathcal{G}]$, which is $\mathcal{G}$-measurable, so the cross term satisfies

$$\begin{aligned}
& \mathbb{E}\left[(Y - \mathbb{E}[Y|\mathcal{G}])(\mathbb{E}[Y|\mathcal{G}] - X - \mu)\right] \\
=& \mathbb{E}\left[(Y - \mathbb{E}[Y|\mathcal{G}])(\mathbb{E}[Y - X - \mu|\mathcal{G}])\right] \\
=& \mathbb{E}\left[Y\xi - \mathbb{E}[Y|\mathcal{G}]\xi\right] \\
=& \mathbb{E}[Y\xi] - \mathbb{E}\left[\mathbb{E}[Y\xi|\mathcal{G}]\right] \\
=& 0
\end{aligned}$$

or note that $\mathbb{E}[Y|\mathcal{G}]$ is $\mathcal{G}$-measurable.

$$\begin{aligned}
& \mathbb{E}\left[(Y - \mathbb{E}[Y|\mathcal{G}])(\mathbb{E}[Y|\mathcal{G}] - X - \mu)\right] \\
=& \mathbb{E}\left[Y\mathbb{E}[Y|\mathcal{G}] - XY - \mu Y - (\mathbb{E}[Y|\mathcal{G}])^2 + X\mathbb{E}[Y|\mathcal{G}] + \mu\mathbb{E}[Y|\mathcal{G}]\right] \\
=& \mathbb{E}\left[Y\mathbb{E}[Y|\mathcal{G}]\right] - \mathbb{E}[XY] - \mu\mathbb{E}Y - \mathbb{E}\left[(\mathbb{E}[Y|\mathcal{G}])^2\right] + \mathbb{E}[XY] + \mu\mathbb{E}Y \\
=& \mathbb{E}\left[Y\mathbb{E}[Y|\mathcal{G}]\right] - \mathbb{E}\left[(\mathbb{E}[Y|\mathcal{G}])^2\right] \\
=& \mathbb{E}\left[Y\mathbb{E}[Y|\mathcal{G}]\right] - \mathbb{E}\left[\mathbb{E}\left[(Y\mathbb{E}[Y|\mathcal{G}])|\mathcal{G}\right]\right] \\
=& 0
\end{aligned}$$

And thus, $\operatorname{Var}(Y - X) = \operatorname{Var}(\mathrm{Err}) + \mathbb{E}\left[(\mathbb{E}[Y|\mathcal{G}] - X - \mu)^2\right] \geqslant \operatorname{Var}(\mathrm{Err})$. □

Exercise 2.8. This exercise is divided into 2 parts.

1. Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let W be a nonnegative $\sigma(X)$-measurable random variable. Show there exists a function g such that $W = g(X)$.

   Hint: Recall that every set in $\sigma(X)$ is of the form $\{X \in B\}$ for some Borel set $B \subset \mathbb{R}$. Suppose first that W is the indicator of such a set, and then use the standard machine.

   Proof. We can find a sequence $\{W_n\}$ of $\sigma(X)$-measurable simple functions such that $W_n \uparrow W$. Each W_n can be written as the form $\sum_{i=1}^{K_n} a_i^{(n)} \mathbb{I}_{A_i^{(n)}}$, where $A_i^{(n)} \in \sigma(X)$ and disjoint. So each $A_i^{(n)}$ can be written as $\left\{X \in B_i^{(n)}\right\}$ for some Borel subset $B_i^{(n)}$ of $\mathbb{R}$. i.e.,

$$W_n = \sum_{i=1}^{K_n} a_i^{(n)} \mathbb{I}_{\left\{X \in B_i^{(n)}\right\}} = \sum_{i=1}^{K_n} a_i^{(n)} \mathbb{I}_{B_i^{(n)}}(X) = g_n(X)$$

Let $g = \limsup_{n\to\infty} g_n$, then g is a Borel function. By taking upper limits to both sides, we get $W = g(X)$. □

2. Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let Y be a nonnegative random variable on this space. We do not assume that X and Y have a joint density. Nonetheless, show there is a function g such that $\mathbb{E}[Y|X] = g(X)$.

   Proof. $\mathbb{E}[Y|X]$ is $\sigma(X)$-measurable. □

第三章　随机游走和布朗运动

3.1 Scaled Random Walk

1. (Remark) The most important properties of Brownian Motion are that it is a martingale and that it accumulates quadratic variation is at rate one per unite time.

2. (Definition) Symmetric Random Walk: $M_0 = 0$, and

$$M_k = \sum_{i=1}^{k} X_i, k = 1, 2, \cdots$$

where

$$\mathbb{P}\{X_i = 1\} = \mathbb{P}\{X_i = -1\} = \frac{1}{2}$$

and $X_i, i \geqslant 1$ are independent.

3. (Theorem) Properties of scaled random walk.

(a) Independent increments.

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

(b) The variance of the symmetric random walk accumulates at the rate one per unit time.

$$\mathrm{Var}\left(M_{k_{i+1}} - M_{k_i}\right) = \sum_{j=k_i+1}^{k_{i+1}} \mathrm{Var}\left(X_j\right) = \sum_{j=k_i+1}^{k_{i+1}} 1 = k_{i+1} - k_i$$

(c) Martingale Property for the Symmetric Random Walk.

$$\begin{aligned}
\mathbb{E}\left[M_\ell | \mathcal{F}_k\right] &= \mathbb{E}\left[M_\ell - M_k + M_k | \mathcal{F}_k\right] \\
&= \mathbb{E}\left[M_\ell - M_k | \mathcal{F}_k\right] + \mathbb{E}\left[M_k | \mathcal{F}_k\right] \\
&= \mathbb{E}\left[\left.\sum_{j=k+1}^{\ell} X_j \right| \mathcal{F}_k\right] + M_k \\
&= \mathbb{E}\left[\sum_{j=k+1}^{\ell} X_j\right] + M_k \quad \text{(Independence)} \\
&= \sum_{j=k+1}^{\ell} \mathbb{E}\left[X_j\right] + M_k \\
&= M_k
\end{aligned}$$

(d) Quadratic Variation of the Symmetric Random Walk

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

注：（定义，随机过程）当 $\{t_i^n\}_{i=0}^n$ 遍取 $[0,t]$ 的分割，定义分割的模

$$\delta_n:=\max_{0\leqslant i\leqslant n-1}\left(t_{i+1}^n-t_i^n\right)$$

Brown 运动的二次变差 $B(t)$ 定义为依概率收敛意义下的极限

$$[B,B](t)=[B,B]([0,t])=\lim_{\delta_n\to 0}\sum_{i=0}^{n-1}\left|B\left(t_{i+1}^n\right)-B\left(t_i^n\right)\right|^2$$

变差即为 $\lim_{\delta_n\to 0}\sum_{i=0}^{n-1}\left|B\left(t_{i+1}^n\right)-B\left(t_i^n\right)\right|$。

3.2 Scaled Symmetric Random Walk

1. (Definition) Scaled symmetric random walk:

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

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

2. (Remark) Let $n\to\infty$, we get a Brownian motion $\lim_{n\to\infty}W^{(n)}(t)$.

3. (Theorem) Properties of Scaled symmetric random walk:

(a) If $0=t_0<t_1<\cdots<t_m$ are such that each nt_j is an integer, then the increments are independent. And if $0\leqslant s\leqslant t$ are such ns and nt are integers, then

$$\mathbb{E}\left[W^{(n)}(t)-W^{(n)}(s)\right]=0$$

$$\operatorname{Var}\left[W^{(n)}(t)-W^{(n)}(s)\right]=t-s$$

(b) Martingale property:

$$\begin{aligned}\mathbb{E}\left[W^{(n)}(t)\,|\mathcal{F}(s)\right]&=\mathbb{E}\left[W^{(n)}(t)-W^{(n)}(s)+W^{(n)}(s)\,|\mathcal{F}(s)\right]\\&=\mathbb{E}\left[W^{(n)}(t)-W^{(n)}(s)\,|\mathcal{F}(s)\right]+\mathbb{E}\left[W^{(n)}(s)\,|\mathcal{F}(s)\right]\\&=W^{(n)}(s)\end{aligned}$$

(c) Quadratic variation:

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

4. (Theorem) Limiting Distribution of the Scaled Symmetric Random Walk: Fix $t\geqslant 0$. As $n\to\infty$, the distribution of the scaled random walk $W^{(n)}(t)$ evaluated at time t converges to the normal distribution with mean zero and variance t.

Proof. We have the density function

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

The moment-generating function of normal distribution with zero mean is

$$\varphi(u) = \mathbb{E}\left(e^{uX}\right) = \int_{-\infty}^{\infty} e^{ux} f(x) \mathrm{d}x = e^{\frac{1}{2}ut^2}$$

The moment-generating function of $W^{(n)}(t)$ is

$$\begin{aligned}
\varphi_n(u) = \mathbb{E}e^{uW^{(n)}(t)} &= \mathbb{E}\exp\left\{\frac{u}{\sqrt{n}} M_{nt}\right\} = \mathbb{E}\exp\left\{\frac{u}{\sqrt{n}}\sum_{j=1}^{nt} X_j\right\} \\
&= \mathbb{E}\prod_{j=1}^{nt}\exp\left\{\frac{u}{\sqrt{n}}X_j\right\} \quad \text{(Independence)} \\
&= \prod_{j=1}^{nt}\mathbb{E}\exp\left\{\frac{u}{\sqrt{n}}X_j\right\} = \prod_{j=1}^{nt}\left(\frac{1}{2}e^{\frac{u}{\sqrt{n}}} + \frac{1}{2}e^{-\frac{u}{\sqrt{n}}}\right) \\
&= \left(\frac{1}{2}e^{\frac{u}{\sqrt{n}}} + \frac{1}{2}e^{-\frac{u}{\sqrt{n}}}\right)^{nt}
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\lim_{n\to\infty}\log\varphi_n(u) &= \lim_{n\to\infty} nt\log\left(\frac{1}{2}e^{\frac{u}{\sqrt{n}}} + \frac{1}{2}e^{-\frac{u}{\sqrt{n}}}\right) \quad \left(\text{Let } x = \frac{1}{\sqrt{n}}\right) \\
&= t\lim_{x\to 0}\frac{\log\left(\frac{1}{2}e^{ux} + \frac{1}{2}e^{-ux}\right)}{x^2} \\
&= t\lim_{x\to 0}\frac{\frac{1}{2}ue^{ux} - \frac{1}{2}ue^{-ux}}{\frac{1}{2}e^{ux} + \frac{1}{2}e^{-ux}}\cdot\frac{1}{2x} \\
&= \frac{t}{2}\cdot\lim_{x\to 0}\frac{1}{e^{ux}+e^{-ux}}\cdot\lim_{x\to 0}\frac{ue^{ux} - ue^{-ux}}{x} \\
&= \frac{t}{2}\cdot\frac{1}{2}\cdot\lim_{x\to 0}\frac{u^2e^{ux} + u^2e^{-ux}}{1} \\
&= \frac{u^2 t}{2}
\end{aligned}$$

Thus,

$$\lim_{n\to\infty}\varphi_n(u) = \frac{1}{2}u^2 t = \varphi(x)$$

Hence, the distribution of the scaled random walk $W^{(n)}(t)$ evaluated at time t converges to the normal distribution with mean zero and variance t. □

5. (Theorem) Log-Normal Distribution as the Limit of Binomial Model. As $n \to \infty$, the distribution of $S_n(t) = S(0)\, u_n^{H_{nt}} d_n^{T_{nt}}$ in binomial model converges to the distribution of

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

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

Proof. The arguments in binomial model are

$$r = 0, u_n = 1 + \frac{\sigma}{\sqrt{n}}, d_n = 1 - \frac{\sigma}{\sqrt{n}}$$

By the risk-neutral probability method, we get

$$\tilde{p} = \frac{1+r-d_n}{u_n-d_n} = \frac{\sigma/\sqrt{n}}{2\sigma/\sqrt{n}} = \frac{1}{2}$$

$$\tilde{q} = \frac{u_n-1-r}{u_n-d_n} = \frac{\sigma/\sqrt{n}}{2\sigma/\sqrt{n}} = \frac{1}{2}$$

The sum of the number of heads (ups) H_{nt} and number of tails (downs) T_{nt} in the first nt coin tosses is nt, thus,

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

Hence,

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

The stock price at time t is

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

Therefore,

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

Let $f\left(x\right) = \log\left(x\right)$, using Taylor's Expansion, we get

$$\log\left(1+x\right) = f\left(0\right) + f'\left(0\right)x + \frac{1}{2}f''\left(0\right)x^2 + o\left(x^3\right) = x - \frac{1}{2}x^2 + O\left(x^3\right)$$

Thus,

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

And $W^{(n)}\left(t\right) = \frac{1}{\sqrt{n}}M_{nt}$,

$$\log S_n\left(t\right) = \log S\left(0\right) - \frac{1}{2}\sigma^2 t + O\left(n^{-1/2}\right) + \sigma W^{(n)}\left(t\right) + O\left(n^{-1}\right)W^{(n)}\left(t\right)$$

Let $n\to\infty$, we have $\lim_{n\to\infty} W^{(n)}\left(t\right) = W\left(t\right)$, thus

$$\lim_{n\to\infty}\log S_n\left(t\right) = \log S\left(0\right) - \frac{1}{2}\sigma^2 t + \sigma W\left(t\right) + \lim_{n\to\infty} o\left(n^{-1}\right)W^{(n)}\left(t\right)$$

while

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}\left[O\left(n^{-1}\right) W^{(n)}(t)\right] &= \lim_{n\to\infty} O\left(n^{-3/2}\right) \mathbb{E}\left[\sum_{j=1}^{nt} X_j\right] \\
&= 0 \\
\lim_{n\to\infty} \mathrm{Var}\left[O\left(n^{-1}\right) W^{(n)}(t)\right] &= \lim_{n\to\infty} O\left(n^{-3}\right) \mathrm{Var}\left[\sum_{j=1}^{nt} X_j\right] \\
&= \lim_{n\to\infty} O\left(n^{-3}\right) nt \\
&= 0
\end{aligned}$$

Thus,

$$O\left(n^{-1}\right) W^{(n)}(t) \to 0, \text{a.s.}$$

Hence,

$$\lim_{n\to\infty} \log S_n(t) = \log S(0) - \frac{1}{2}\sigma^2 t + \sigma W(t)$$

□

3.3 Brownian Motion

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

$$W(t_1) - W(t_0), W(t_2) - W(t_1), \cdots, W(t_m) - W(t_{m-1})$$

are independent and each of these increments is normally distribution with mean zero and variance $t_{i+1} - t_i$.

2. (Remark) How to think of ω?

   (a) Think of ω as the Brownian motion path. A random experiment is performed, and its outcome is the path of Brownian motion.

   (b) Think of ω as something more primitive than the path itself, akin to the outcome of a sequence of coin tosses, although now the coin is being tossed "infinitely fast".

3. (Theorem) Covariance of Brownian motion:

   (a) Covariance of $W(s)$ and $W(t)$:

$$\mathrm{Cov}(W(s), W(t)) = \min\{s, t\}$$

   (b) Covariance matrix: For the m-dimensional random vector $[W(t_1), W(t_2), \cdots, W(t_m)]$, $t_1 \leqslant t_2 \leqslant \cdots \leqslant t_m$ is $\begin{pmatrix} t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & \cdots & t_2 \\ \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & \cdots & t_m \end{pmatrix}$.

4. (Theorem) Moment-generating function of $[W(t_1), W(t_2), \cdots, W(t_m)]$ is

$$\varphi(u_1, u_2, \cdots, u_m) = \exp\left\{\frac{1}{2}\sum_{i=1}^{m}\left[\left(\sum_{j=i}^{m} u_j\right)^2 \cdot (t_i - t_{i-1})\right]\right\}$$

where $t_0 = 0$.

Proof.

$$\begin{aligned}
&\varphi(u_1, u_2, \cdots, u_m) \\
=&\mathbb{E}\exp\{u_1 W(t_1) + u_2 W(t_2) + \cdots + u_m W(u_m)\} \\
=&\mathbb{E}\exp\{(u_1 + u_2 + \cdots + u_m) W(t_1) + (u_2 + \cdots + u_m)[W(t_2) - W(t_1)] \\
&\quad + \cdots + u_m [W(t_m) - W(t_{m-1})] \} \text{(Independence)} \\
=&\mathbb{E}\exp\{(u_1 + u_2 + \cdots + u_m) W(t_1)\} \cdot \mathbb{E}\exp\{(u_2 + \cdots + u_m)[W(t_2) - W(t_1)]\} \\
&\cdot \mathbb{E}\exp\{u_m [W(t_m) - W(t_{m-1})]\} \\
=&\exp\left\{\frac{1}{2}\left(\sum_{j=1}^{m} u_j\right)^2 t_1 + \frac{1}{2}\left(\sum_{j=2}^{m} u_j\right)^2 (t_2 - t_1) + \cdots + \frac{1}{2}u_m^2 (t_m - t_{m-1})\right\} \\
=&\exp\left\{\frac{1}{2}\sum_{i=1}^{m}\left[\left(\sum_{j=i}^{m} u_j\right)^2 \cdot (t_i - t_{i-1})\right]\right\} \text{(Let } t_0 = 0)
\end{aligned}$$

□

5. (Definition) Filtration for Brownian Motion: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which is defined a Brownian motion $W(t), t \geqslant 0$. A filtration for the Brownian Motion is a collection of σ-algebras $\mathcal{F}(t), t \geqslant 0$, such that:

   (a) (**Information accumulates**) For $0 \leqslant s < t$, $\mathcal{F}(s) \subset \mathcal{F}(t)$.

   (b) (**Adaptivity**) For each $t \geqslant 0$, $W(t)$ at time t is $\mathcal{F}(t)$-measurable.

   (c) (**Independence of future increments**) For $0 \leqslant t < u$, the increments $W(u) - W(t)$ is independent of $\mathcal{F}(t)$.

   Remark: The property leads to the efficient market hypothesis.

6. (Definition) Let $\Delta(t), t \geqslant 0$, be a stochastic process. We say that $\Delta(t)$ is adapted to the filtration $\mathcal{F}(t)$ if for each $t \geqslant 0$, the random variable $\Delta(t)$ is $\mathcal{F}(t)$-measurable.

7. (Theorem) Brownian motion is a martingale.

Proof. Let $0 \leqslant s < t$,

$$\begin{aligned}
\mathbb{E}[W(t) | \mathcal{F}(t)] &= \mathbb{E}[W(t) - W(s) + W(s) | \mathcal{F}(s)] \\
&= \mathbb{E}[W(t) - W(s) | \mathcal{F}(s)] + \mathbb{E}[W(s) | \mathcal{F}(s)] \\
&= 0 + W(s) \\
&= W(s)
\end{aligned}$$

□

3.4 Quadratic Variation

1. (Definition) First-order variation $\mathrm{FV}_T(f)$: First choose a partition $\Pi = \{t_0, t_1, \cdots, t_n\}$ of $[0, T]$, which is a set of times $0 = t_0 < t_1 < \cdots < t_n = T$. We do not require the partition points to to equally spaced. The maximum step size of the partition will be denoted $\|\Pi\| = \max_{j=0,1,\cdots n-1}(t_{j+1} - t_j)$. Define

$$\mathrm{FV}_T(f) = \lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} |f(t_{j+1}) - f(t_j)|$$

2. (Remark) For a ***differentiable*** function f, by the Mean Value Theorem, for each sub-interval $[t_j, t_{j+1}]$ there is a point t_j^* such that

$$\frac{f(t_{j+1}) - f(t_j)}{t_{j+1} - t_j} = f'\left(t_j^*\right)$$

or

$$f(t_{j+1}) - f(t_j) = f'\left(t_j^*\right)(t_{j+1} - t_j)$$

Thus,

$$\mathrm{FV}_T(f) = \lim_{\|\Pi\|\to 0} \sum_{j=1}^{n-1} \left|f'\left(t_j^*\right)(t_{j+1} - t_j)\right| = \int_0^T \left|f'(t)\right| \mathrm{d}t$$

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

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

where $\Pi = \{t_0, t_1, \cdots, t_n\}$ and $\|\Pi\| = \max_{j=0,1,\cdots n-1}(t_{j+1} - t_j)$.

4. (Remark) Suppose the function f has a ***continuous derivative***. Then

$$\sum_{j=0}^{n-1} [f(t_{j+1}) - f(t_j)]^2 = \sum_{j=0}^{n-1} \left[f'\left(t_j^*\right)(t_{j+1} - t_j)\right]^2 \leqslant \|\Pi\| \sum_{j=0}^{n-1} (t_{j+1} - t_j)\left[f'\left(t_j^*\right)\right]^2$$

Thus

$$\begin{aligned}
[f, f](T) &= \lim_{\|\Pi\|\to 0} \sum_{j=1}^{n-1} [f(t_{j+1}) - f(t_j)]^2 \\
&\leqslant \lim_{\|\Pi\|\to 0} \left[\|\Pi\| \sum_{j=0}^{n-1} (t_{j+1} - t_j)\left[f'\left(t_j^*\right)\right]^2\right] \\
&= \lim_{\|\Pi\|\to 0} \|\Pi\| \cdot \lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} (t_{j+1} - t_j)\left[f'\left(t_j^*\right)\right]^2 \\
&= \lim_{\|\Pi\|\to 0} \|\Pi\| \cdot \int_0^T \left[f'(t)\right]^2 \mathrm{d}t \\
&= 0
\end{aligned}$$

5. (Theorem) Let W be a Brownian motion. Then $[W, W](T) = T$ for all $T \geqslant 0$ almost surely.

   Remark:

(a) $\mathrm{d}W(t) \cdot \mathrm{d}W(t) = \mathrm{d}t$.

(b) Brownian motion accumulates quadratic variation at rate one per unite time.

$$[W,W](T_2) - [W,W](T_1) = T_2 - T_1$$

Proof.

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

Its expectation and variance are

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

and

$$\begin{aligned}
\mathrm{Var}[Q_\Pi] &= \mathrm{Var}\left[\sum_{j=0}^{n-1} [W(t_{j+1}) - W(t_j)]^2\right] \\
&= \sum_{j=0}^{n-1} \mathrm{Var}\left([W(t_{j+1}) - W(t_j)]^2\right) = \sum_{j=0}^{n-1} \mathbb{E}\left([W(t_{j+1}) - W(t_j)]^2 - (t_{j+1} - t_j)\right)^2 \\
&= \sum_{j=0}^{n-1} \mathbb{E}\left([W(t_{j+1}) - W(t_j)]^4 + (t_{j+1} - t_j)^2 - 2(t_{j+1} - t_j)[W(t_{j+1}) - W(t_j)]^2\right) \\
&= \sum_{j=0}^{n-1} 3(t_{j+1} - t_j)^2 + (t_{j+1} - t_j)^2 - 2(t_{j+1} - t_j)^2 = \sum_{j=0}^{n-1} 2(t_{j+1} - t_j)^2 \\
&\leqslant \sum_{j=0}^{n-1} 2\|\Pi\|(t_{j+1} - t_j) \\
&= 2\|\Pi\| T
\end{aligned}$$

In particular,

$$\lim_{\|\Pi\| \to 0} \mathrm{Var}[Q_\Pi] = 0$$

We conclude that

$$\lim_{\|\Pi\| \to 0} Q_\Pi = \mathbb{E}[Q_\Pi] = T$$

□

6. (Theorem) We can also compute the cross variation of $W(t)$ with t and the quadratic variation of t with itself, which are

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

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

Remark:

(a) $\mathrm{d}W(t)\cdot\mathrm{d}t=0$

(b) $\mathrm{d}t\cdot\mathrm{d}t=0$

Proof. By definition, we have

$$\begin{aligned}
&\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}\left[W(t_{j+1})-W(t_j)\right](t_{j+1}-t_j)\\
\leqslant&\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}\left|W(t_{j+1})-W(t_j)\right|(t_{j+1}-t_j)\\
\leqslant&\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}\left[\max_{0\leqslant k\leqslant n-1}\left|W(t_{k+1})-W(t_k)\right|(t_{j+1}-t_j)\right]\\
=&\lim_{\|\Pi\|\to 0}\left[\max_{0\leqslant k\leqslant n-1}\left|W(t_{k+1})-W(t_k)\right|\sum_{j=0}^{n-1}(t_{j+1}-t_j)\right]\\
=&\lim_{\|\Pi\|\to 0}\left[\max_{0\leqslant k\leqslant n-1}\left|W(t_{k+1})-W(t_k)\right|\right]\cdot T\\
=&0
\end{aligned}$$

And we have

$$\begin{aligned}
&\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}(t_{j+1}-t_j)^2\\
\leqslant&\lim_{\|\Pi\|\to 0}\left[\max_{0\leqslant k\leqslant n-1}(t_{k+1}-t_k)\sum_{j=0}^{n-1}(t_{j+1}-t_j)\right]\\
=&\lim_{\|\Pi\|\to 0}\|\Pi\|\cdot T\\
=&0
\end{aligned}$$

□

7. (Remark) The volatility of geometric Brownian motion. Geometric Brownian motion is

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

The log return rate is

$$\log\frac{S(t_{j+1})}{S(t_j)}=\sigma\left[W(t_{j+1})-W(t_j)\right]+\left(\alpha-\frac{1}{2}\sigma^2\right)(t_{j+1}-t_j)$$

And the realized volatility of return rate is

$$\begin{aligned}
\sum_{j=0}^{n-1}\left[\log\frac{S(t_{j+1})}{S(t_j)}\right]^2=&\sigma^2\sum_{j=0}^{n-1}\left[W(t_{j+1})-W(t_j)\right]^2+\left(\alpha-\frac{1}{2}\sigma^2\right)^2\sum_{j=0}^{n-1}(t_{j+1}-t_j)^2\\
&+2\sigma\left(\alpha-\frac{1}{2}\sigma^2\right)\sum_{j=0}^{n-1}\left[W(t_{j+1})-W(t_j)\right](t_{j+1}-t_j)
\end{aligned}$$

Then

$$\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}\left[\log\frac{S(t_{j+1})}{S(t_j)}\right]^2=\sigma^2 T$$

Hence, on the interval $[T_1, T_2]$, we have

$$\frac{1}{T_2-T_1}\sum_{j=0}^{n-1}\left[\log\frac{S(t_{j+1})}{S(t_j)}\right]^2\approx\sigma^2$$

8. (Theorem) Markov Property: Let $W(t), t\geqslant 0$ be a Brownian motion and let $\mathcal{F}(t), t\geqslant 0$ be a filtration for this Brownian motion. Then $W(t), t\geqslant 0$ is a Markov process, i.e., whenever $0\leqslant s\leqslant t$ and f, g are Borel-measurable function such that

$$\mathbb{E}\left[f(W(t))\,|\mathcal{F}(s)\right]=g(W(s))$$

or

$$\mathbb{E}\left[f(W(t),t)\,|\mathcal{F}(s)\right]=f(W(s),s)$$

Proof. We needs the theorem or the Independence Lemma below. Let's review.

Review: Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Suppose the random variables $X_1, X_2,\cdots,X_K$ are $\mathcal{G}$-measurable and the random variables $Y_1,Y_2,\cdots,Y_L$ are independent of $\mathcal{G}$. Let $f(x_1,x_2,\cdots,x_K,y_1,y_2,\cdots,y_L)$ be the function of dummy variables $x_1,x_2,\cdots,x_K$ and $y_1,y_2,\cdots,y_L$, and define

$$g(x_1,x_2,\cdots,x_K):=\mathbb{E}\left[f(x_1,x_2,\cdots,x_K,Y_1,Y_2,\cdots,Y_L)\right]$$

Then

$$\begin{aligned}\mathbb{E}\left[f(X_1,X_2,\cdots,X_K,Y_1,Y_2,\cdots,Y_L)\,|\mathcal{G}\right]&=g(X_1,X_2,\cdots,X_K)\\&=\mathbb{E}\left[f(X_1,X_2,\cdots,X_K,Y_1,Y_2,\cdots,Y_L)\right]\end{aligned}$$

If we write

$$\mathbb{E}\left[f(W(t))\,|\mathcal{F}(s)\right]=\mathbb{E}\left[f(W(t)-W(s)+W(s))\,|\mathcal{F}(s)\right]$$

And define $g(x)=\mathbb{E}f(W(t)-W(s)+x)$, then

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

The Independence Lemma states that if we now take the function $g(x)$ and replace the dummy variable x by random variable $W(s)$, then the equation $\mathbb{E}\left[f(W(t))\,|\mathcal{F}(s)\right]=g(W(s))$ holds, note that $W(s)$ is $\mathcal{F}(s)$-measurable. □

9. (Definition) Transition density $p(\tau,x,y)$: Let $\tau=t-s$ and $y=w+x$ in $g(x)$, then we define the transition density as

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

Remark: By the definition, we can rewrite

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

and

$$\mathbb{E}\left[f(W(t))\,|\mathcal{F}(s)\right]=\int_{-\infty}^{\infty}f(y)\,p(\tau,W(s),y)\,\mathrm{d}y$$

3.5 First Passage Time Distribution

1. (Theorem) Exponential martingale. Let $W(t), t \geqslant 0$, be a Brownian motion with a filtration $\mathcal{F}(t), t \geqslant 0$, and let σ be a constant. The process $Z(t), t \geqslant 0$, of

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

is a martingale.

Proof. For $0 \leqslant s \leqslant t$, we have

$$\begin{aligned}
&\mathbb{E}[Z(t)|\mathcal{F}(s)] \\
=&\mathbb{E}\left[\exp\left\{\sigma W(t) - \frac{1}{2}\sigma^2 t\right\}\middle|\mathcal{F}(s)\right] \\
=&\mathbb{E}\left[\exp\left\{\sigma W(t) - \sigma W(s) + \sigma W(s) - \frac{1}{2}\sigma^2 t\right\}\middle|\mathcal{F}(s)\right] \\
=&\mathbb{E}\left[\exp\{\sigma W(t) - \sigma W(s)\} \cdot \exp\left\{\sigma W(s) - \frac{1}{2}\sigma^2 t\right\}\middle|\mathcal{F}(s)\right] \\
=&\mathbb{E}[\exp\{\sigma W(t) - \sigma W(s)\}|\mathcal{F}(s)] \cdot \exp\left\{\sigma W(s) - \frac{1}{2}\sigma^2 t\right\} \\
=&\exp\left\{\frac{1}{2}(t-s)\sigma^2\right\} \cdot \exp\left\{\sigma W(s) - \frac{1}{2}\sigma^2 t\right\} \\
=&\exp\left\{\sigma W(s) - \frac{1}{2}s\sigma^2\right\}
\end{aligned}$$

□

2. (Definition) A stopping time is a function: $\tau : \Omega \to [0,\infty) \cup \{\infty\}$ such that τ is $\mathcal{F}_t$-measurable, i.e., for every $t \geqslant 0$, $\{\tau \leqslant t\} \in \mathcal{F}_t$.

3. (**Optional Sampling Lemma**) Let M be a martingale (sub-martingale or sup-martingale) and τ be a stopping time. Then the stopped process $M^\tau(t) := M(\tau \wedge t)$ is also a martingale (sub-martingale or sup-martingale). In particular, if M is a martingale,

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

And thus, we conclude that a martingale (sub-martingale or sup-martingale) stopped at a stopping time is a martingale (sub-martingale or sup-martingale, respectively).

注：有关停时与鞅的可参考内容。

(a)（定义，随机过程）设 $\{X_n, n \geqslant 0\}$ 是一个随机过程，称随机函数 T 是关于过程 $\{X_n, n \geqslant 0\}$ 的停时，如果

i. T 在 $\{0, 1, 2, \cdots, \infty\}$ 上取值；

ii. 对每一个 $n \geqslant 0$，$\{T = n\} \in \sigma(X_0, X_1, \cdots, X_n)$。

(b)（**有界停时定理**，随机过程）设 $\{M_n, n \geqslant 0\}$ 是一个关于 $\{X_n, n \geqslant 0\}$ 的鞅，T 是一个关于 $\{X_n, n \geqslant 0\}$ 的停时，并且 $\exists K > 0, T \leqslant K$（有界停时），设 $\mathcal{F}_n = \sigma(X_0, X_1, \cdots, X_n)$，则有

$$\mathbb{E}[M_T|\mathcal{F}_0] = M_0$$

等式两边取期望后，有

$$\mathbb{E}[M_T] = \mathbb{E}[M_0]$$

(c) (**鞅停时定理**, 随机过程) $\mathcal{F}_n = \sigma(X_0, X_1, \cdots, X_n)$, 设 $\{M_n, n \geqslant 0\}$ 是一个关于 $\{X_n, n \geqslant 0\}$ 的鞅（或者说 $\{M_n, n \geqslant 0\}$ 是一个关于 $\{\mathcal{F}_n = \sigma(X_0, X_1, \cdots, X_n)\}$ 的鞅），T 是一个关于 $\{X_n, n \geqslant 0\}$ 的停时，且满足

i. $\mathbb{P}\{T < \infty\} = 1$;

ii. $\mathbb{E}[|M_T|] < \infty$;

iii. $\lim_{n\to\infty} \mathbb{E}\left[|M_n| \mathbb{I}_{\{T>n\}}\right] = 0$;

则有

$$\mathbb{E}[M_T] = \mathbb{E}[M_0]$$

(d)（推论，随机过程）$\{M_n, n \geqslant 0\}$ 是一个关于 $\{\mathcal{F}_n = \sigma(X_0, X_1, \cdots, X_n)\}$ 的鞅，T 是一个关于 $\{X_n, n \geqslant 0\}$ 的停时，且满足

i. 鞅本身取值有界，即存在 $K \geqslant 0$ 使得 $|M_n| < K$;

ii. $\mathbb{P}\{T < \infty\} = 1$;

那么

$$\mathbb{E}[M_T] = \mathbb{E}[M_0]$$

(e)（推论，随机过程）设 $\{M_n, n \geqslant 0\}$ 是一个关于随机过程 $\{X_n, n \geqslant 0\}$ 的一致可积鞅，T 是关于 $\{X_n, n \geqslant 0\}$ 的停时，满足

i. $\mathbb{P}\{T < \infty\} = 1$;

ii. $\mathbb{E}[|M_T|] < \infty$;

$$\mathbb{E}[M_T] = \mathbb{E}[M_0]$$

(f) 两个一致可积的充分条件:

i. 假设 $X_0, X_1, \cdots, X_n$ 是一列随机变量，并且存在常数 $C < \infty$，使得 $\mathbb{E}\left[X_n^2\right] < C$ 对于所有的 n 成立，则该序列一致可积。

ii. 设 $\{M_n\}$ 是关于 $\{\mathcal{F}_n\}$ 的鞅，如果存在一个非负随机变量 Y，满足 $\mathbb{E}(Y) < \infty$ 且 $\forall n$, $|M_n| < Y$，则 $\{M_n\}$ 是一致可积鞅。

4. (Definition) First passage time to level m:

$$\tau_m = \inf\{t : t \geqslant 0, W(t) = m\}$$

If the Brownian motion never reaches the level m, we set $\tau_m = \infty$.

5. (Remark) A martingale that is stopped at a stopping time is still a martingale (it's clear that τ_m is a stopping time. And besides, $t \wedge \tau_m$ is a bounded stopping time), and thus must have a constant expectation. Hence,

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

We assume $\sigma > 0, m > 0$. In this case, the Brownian motion is always at or below level m for $t \leqslant \tau_m$ and so

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

There are two cases:

- If $\tau_m < \infty$, the term $\exp\left\{-\frac{1}{2}\sigma^2\left(t \wedge \tau_m\right)\right\}$ is equal to $\exp\left\{-\frac{1}{2}\sigma^2\tau_m\right\}$ for large enough t.
- If $\tau_m = \infty$, then the term $\exp\left\{-\frac{1}{2}\sigma^2\left(t \wedge \tau_m\right)\right\}$ is equal to $\exp\left\{-\frac{1}{2}\sigma^2 t\right\}$, and it converges to zero as $t \to \infty$.

Therefore,

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

Similarly, we have

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

Moreover,

$$\begin{aligned} 1 = \lim_{t\to\infty} 1 &= \mathbb{E}\left[\lim_{t\to\infty} \exp\left\{\sigma W\left(t \wedge \tau_m\right) - \frac{1}{2}\sigma^2\left(t \wedge \tau_m\right)\right\}\right] \\ &= \mathbb{E}\left[\mathbb{I}_{\{\tau_m<\infty\}} \exp\left\{\sigma m - \frac{1}{2}\sigma^2\tau_m\right\}\right] \\ &= \mathbb{E}\left[\mathbb{I}_{\{\tau_m<\infty\}} \exp\left\{-\frac{1}{2}\sigma^2\tau_m\right\}\right] \cdot e^{\sigma m} \end{aligned}$$

so

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

Let $\sigma \to 0$, we have

$$\mathbb{E}\left[\mathbb{I}_{\{\tau_m<\infty\}}\right] = 1$$

or

$$\mathbb{P}\left\{\tau_m < \infty\right\} = 1$$

we say τ_m is finite almost surely.

6. (Theorem) For $m \in \mathbb{R}$, the first passage time of Brownian motion to level m is finite almost surely, and the ***Laplace transform*** of its distribution is given by

$$\mathbb{E}e^{-\alpha\tau_m} = e^{-|m|\sqrt{2\alpha}}, \forall \alpha > 0$$

Proof. Since $\mathbb{E}\left[\mathbb{I}_{\{\tau_m<\infty\}}\right] = 1$ or $\sigma > 0$,, we have $\mathbb{I}_{\{\tau_m<\infty\}} \exp\left\{-\frac{1}{2}\sigma^2\tau_m\right\} = \exp\left\{-\frac{1}{2}\sigma^2\tau_m\right\}$ a.s., then we can drop the indicator function in the

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

so we get

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

Let $\frac{1}{2}\sigma^2 = \alpha$, then $\mathbb{E}e^{-\alpha\tau_m} = e^{-m\sqrt{2\alpha}}$. And because Brownian motion is symmetric, the first passage times τ_m and $\tau_{|m|}$ have the same distribution. □

7. (Remark) Differential of $\mathbb{E}e^{-\alpha\tau_m} = e^{-|m|\sqrt{2\alpha}}$ with respect to α is

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

Letting $\alpha \to 0$, we obtain $\mathbb{E}\tau_m = \infty$, as long as $m \neq 0$.

3.6 Reflection Principle

1. (Remark) Reflection Equality:

$$\begin{aligned}\mathbb{P}\left\{\tau_m \leqslant t, W\left(t\right) \leqslant w\right\} &= \mathbb{P}\left\{\tau_m \leqslant t, W\left(t\right) \geqslant 2m - w\right\} \\ &= \mathbb{P}\left\{W\left(t\right) \geqslant 2m - w\right\}, w \leqslant m, m > 0\end{aligned}$$

   It seems we have another Brownian motion, which reflects at m after τ_m (Compared with the old one).

2. (Theorem) For all $m \in \mathbb{R}$, the random variable τ_m has cumulative distribution function

$$\mathbb{P}\left\{\tau_m \leqslant t\right\} = \frac{2}{\sqrt{2\pi}} \int_{\frac{|m|}{\sqrt{t}}}^{\infty} e^{-\frac{y^2}{2}} \mathrm{d}y \ , t \geqslant 0$$

   and density

$$f_{\tau_m}\left(t\right) = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}\left\{\tau_m \leqslant t\right\} = \frac{|m|}{t\sqrt{2\pi t}} e^{-\frac{m^2}{2t}}, t \geqslant 0$$

   Proof. We first consider $m > 0$.

$$\begin{aligned}\mathbb{P}\left\{\tau_m \leqslant t\right\} =& \mathbb{P}\left\{\tau_m \leqslant t, W\left(t\right) \leqslant m\right\} + \mathbb{P}\left\{\tau_m \leqslant t, W\left(t\right) \geqslant m\right\} \\ =& \mathbb{P}\left\{\tau_m \leqslant t, W\left(t\right) \leqslant m\right\} + \mathbb{P}\left\{W\left(t\right) \geqslant m\right\} \\ =& 2\mathbb{P}\left\{W\left(t\right) \geqslant m\right\} \\ =& \frac{2}{\sqrt{2\pi t}} \int_m^{\infty} e^{-\frac{x^2}{2t}} \mathrm{d}x \ \left(\text{Let } y = \frac{x}{\sqrt{t}}\right) \\ =& \frac{2}{\sqrt{2\pi}} \int_{\frac{m}{\sqrt{t}}}^{\infty} e^{-\frac{y^2}{2}} \mathrm{d}y \ \left(m > 0\right)\end{aligned}$$

   It works similarly when $m \leqslant 0$,

$$\mathbb{P}\left\{\tau_m \leqslant t\right\} = \mathbb{P}\left\{\tau_{|m|} \leqslant t\right\} = \frac{2}{\sqrt{2\pi}} \int_{\frac{|m|}{\sqrt{t}}}^{\infty} e^{-\frac{y^2}{2}} \mathrm{d}y$$

   □

3. (Remark) For $m \in \mathbb{R}$, the first passage time of Brownian motion to level m is finite almost surely, and the ***Laplace transform*** of its distribution can be also given by

$$\mathbb{E}e^{-\alpha\tau_m} = \int_0^{\infty} e^{-\alpha\tau_m} f_{\tau_m}\left(t\right) \mathrm{d}t = \int_0^{\infty} \frac{|m|}{t\sqrt{2\pi t}} e^{-\alpha m - \frac{m^2}{2t}} \mathrm{d}t, \forall \alpha > 0$$

   Remark: Two different forms of Laplace transform for the density of τ_m, i.e., $\mathbb{E}e^{-\alpha\tau_m}$, are given, and those two forms must meet, i.e.,

$$\mathbb{E}e^{-\alpha\tau_m} = \int_0^{\infty} \frac{|m|}{t\sqrt{2\pi t}} e^{-\alpha m - \frac{m^2}{2t}} \mathrm{d}t = e^{-|m|\sqrt{2\alpha}}, \forall \alpha > 0$$

   which will be verified in the part Selected Exercise.

4. (Remark) We define maximum to date $M\left(t\right)$ by

$$M\left(t\right) = \max_{0 \leqslant s \leqslant t} W\left(s\right)$$

And we can rewrite the reflection equality: For positive m, we have

$$M(t) \geqslant m \Longleftrightarrow \tau_m \leqslant t$$

thus,

$$\begin{aligned}\mathbb{P}\{M(t) \geqslant m, W(t) \leqslant w\} =& \mathbb{P}\{\tau_m \leqslant t, W(t) \leqslant w\} \\ =& \mathbb{P}\{W(t) \geqslant 2m - w\}, w \leqslant m, m > 0\end{aligned}$$

5. (Theorem) For $t > 0$, the joint distribution of $(M(t), W(t))$ is

$$f_{M(t),W(t)}(m, w) = \frac{2(2m-w)}{t\sqrt{2\pi t}} e^{-\frac{(2m-w)^2}{2t}}, w \leqslant m, m > 0$$

Proof. Because

$$\mathbb{P}\{M(t) \geqslant m, W(t) \leqslant w\} = \int_m^\infty \int_{-\infty}^w f_{M(t),W(t)}(x, y)\, \mathrm{d}y\mathrm{d}x$$

and

$$\mathbb{P}\{W(t) \geqslant 2m - w\} = \int_{-\infty}^{2m-w} \frac{1}{\sqrt{2\pi t}} e^{-\frac{z^2}{2t}} \mathrm{d}z$$

thus,

$$\int_m^\infty \int_{-\infty}^w f_{M(t),W(t)}(x, y)\, \mathrm{d}y\mathrm{d}x = \int_{-\infty}^{2m-w} \frac{1}{\sqrt{2\pi t}} e^{-\frac{z^2}{2t}} \mathrm{d}z$$

We differentiate first with respect to m to obtain

$$-\int_{-\infty}^w f_{M(t),W(t)}(m, y)\, \mathrm{d}y = \frac{2}{\sqrt{2\pi t}} e^{-\frac{(2m-w)^2}{2t}}$$

We next differentiate with respect to w to see that

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

□

6. (Corollary) The conditional distribution of $M(t)$ given $W(t) = w$ is

$$f_{M(t)|W(t)}(m|w) = \frac{2(2m-w)}{t} \exp\left\{-\frac{2m(m-w)}{t}\right\}, w \leqslant m, m > 0$$

Proof.

$$\begin{aligned} f_{M(t)|W(t)}(m|w) &= \frac{f_{M(t),W(t)}(m, w)}{f_{W(t)}(w)} = \frac{\frac{2(2m-w)}{t\sqrt{2\pi t}} \exp\left\{-\frac{(2m-w)^2}{2t}\right\}}{\frac{1}{\sqrt{2\pi t}} \exp\left\{-\frac{w^2}{2t}\right\}} \\ &= \frac{2(2m-w)}{t} \exp\left\{-\frac{4m^2 + w^2 - 4mw}{2t} + \frac{w^2}{2t}\right\} \\ &= \frac{2(2m-w)}{t} \exp\left\{-\frac{2m(m-w)}{t}\right\} \end{aligned}$$

□

3.7 Selected Exercise

Exercise 3.1. (**Other variations of Brownian motion**)

1. Show that as the number n of partition points on $[0,T]$ approached infinity and the length of the longest sub-interval approaches zero, the sample first variation

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

for almost every path of the Brownian motion W.

Proof. (Proof by making a contradiction) Assume there exist $A \in \mathcal{F}$, such that $\mathbb{P}(A) > 0$ and for every $\omega \in A$,

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

Then,

$$\begin{aligned}
\sum_{j=0}^{n-1} [W(t_{j+1}) - W(t_j)]^2 &\leqslant \max_{0\leqslant k\leqslant n-1} |W(t_{k+1}) - W(t_k)| \cdot \sum_{j=0}^{n-1} |W(t_{j+1}) - W(t_j)| \\
&\leqslant \max_{0\leqslant k\leqslant n-1} |W(t_{k+1}) - W(t_k)| \cdot \limsup_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} |W(t_{j+1}) - W(t_j)| \\
&\to 0
\end{aligned}$$

since by uniform continuity of continuous functions over a closed interval,

$$\lim_{\|\Pi\|\to 0} \max_{0\leqslant k\leqslant n-1} |W(t_{j+1}) - W(t_j)|(\omega) = 0$$

However, we know

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

Therefore, $\sum_{j=0}^{n-1} |W(t_{j+1}) - W(t_j)|$ is not bounded. □

2. Show that as the number n of partition points on $[0,T]$ approaches infinity and the length of the longest sub-interval approaches zero, the sample cubic variation

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

for almost every path of the Brownian motion W.

Proof. We note by an argument similar to 1,

$$\begin{aligned}
\sum_{j=0}^{n-1} [W(t_{j+1}) - W(t_j)]^3 &\leqslant \max_{0\leqslant k\leqslant n-1} |W(t_{k+1}) - W(t_k)| \cdot \sum_{j=0}^{n-1} |W(t_{j+1}) - W(t_j)|^2 \\
&\to 0
\end{aligned}$$

as $n \to \infty$. □

Exercise 3.2. Let $W(t)$ be a Brownian motion and let $\mathcal{F}(t), t \geqslant 0$, be an associated filtration.

1. For $\mu \in \mathbb{R}$, consider the Brownian motion with drift μ:

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

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

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

satisfies $\mathbb{E}[f(X(t))|\mathcal{F}(s)] = g(X(s))$, and hence X has the Markov property. We may rewrite $g(x)$ as

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

where $\tau = t - s$ and

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

Proof. Note that $t + W(s)$ is $\mathcal{F}(s)$-measurable, and $W(t) - W(s)$ is independent of $\mathcal{F}(s)$.

$$\begin{aligned}
&\mathbb{E}[f(X(t))|\mathcal{F}(s)] \\
=&\mathbb{E}[f(X(t) - X(s) + X(s))|\mathcal{F}(s)] \\
=&\mathbb{E}[f(\mu t + W(t) - \mu s - W(s) + \mu s + W(s))|\mathcal{F}(s)] \\
=&\mathbb{E}[f(W(t) - W(s) + x)|\mathcal{F}(s)] \;\; (\text{Let } x = \mu t + W(s)\,) \\
=&\mathbb{E}[f(W(t) - W(s) + x)] \\
=&\int_{-\infty}^{\infty} f(x + \mu t + W(s)) \frac{1}{\sqrt{2\pi(t-s)}} \exp\left\{-\frac{x^2}{2(t-s)}\right\} \mathrm{d}x \\
=&\int_{-\infty}^{\infty} f(y) \frac{1}{\sqrt{2\pi(t-s)}} \exp\left\{-\frac{(y - \mu t - W(s) - \mu s + \mu s)^2}{2(t-s)}\right\} \mathrm{d}y \\
=&\frac{1}{\sqrt{2\pi(t-s)}} \int_{-\infty}^{\infty} f(y) \exp\left\{-\frac{(y - X(s) - \mu(t-s))^2}{2(t-s)}\right\} \mathrm{d}xy \\
=&g(X(s))
\end{aligned}$$

□

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

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

Set $\tau = t - s$ and

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

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

$$g(x) = \int_0^{\infty} f(y)\, p(\tau, x, y)\, \mathrm{d}y$$

satisfies $\mathbb{E}[f(S(t))|\mathcal{F}(s)] = g(S(s))$ and hence S has the Markov property and $p(\tau, x, y)$ is its density.

Proof. Let $\mu=\frac{\nu}{\sigma}$, then $\sigma W(t)+\nu t=\sigma\left[W(t)+\frac{\nu}{\sigma}t\right]=\sigma X(t)$, by what we have done in 1, we have

$$
\begin{aligned}
&\mathbb{E}\left[f\left(S\left(t\right)\right)|\mathcal{F}\left(s\right)\right]\\
=&\mathbb{E}\left[f\left(S\left(0\right)e^{\sigma X(t)}\right)|\mathcal{F}\left(s\right)\right]\\
=&\frac{1}{\sqrt{2\pi\left(t-s\right)}}\int_{-\infty}^{\infty}f\left(S\left(0\right)e^{\sigma y}\right)\exp\left\{-\frac{\left(y-X\left(s\right)-\mu\left(t-s\right)\right)^2}{2\left(t-s\right)}\right\}\mathrm{d}y\\
=&\frac{1}{\sqrt{2\pi\left(t-s\right)}}\int_{-\infty}^{\infty}f\left(z\right)\exp\left\{-\frac{\left(\frac{1}{\sigma}\ln\frac{z}{S(0)}-X\left(s\right)-\mu\left(t-s\right)\right)^2}{2\left(t-s\right)}\right\}\frac{1}{\sigma z}\mathrm{d}z\\
=&\frac{1}{\sqrt{2\pi\left(t-s\right)}}\int_{-\infty}^{\infty}f\left(z\right)\exp\left\{-\frac{\left(\ln\frac{z}{S(0)}-\sigma X\left(s\right)-\nu\left(t-s\right)\right)^2}{2\sigma^2\left(t-s\right)}\right\}\frac{1}{\sigma z}\mathrm{d}z\\
=&\frac{1}{\sqrt{2\pi\left(t-s\right)}}\int_{-\infty}^{\infty}f\left(z\right)\exp\left\{-\frac{\left(\ln\frac{z}{S(0)}-\ln\frac{S(s)}{S(0)}-\nu\left(t-s\right)\right)^2}{2\sigma^2\left(t-s\right)}\right\}\frac{1}{\sigma z}\mathrm{d}z\\
=&\frac{1}{\sqrt{2\pi\left(t-s\right)}}\int_{-\infty}^{\infty}f\left(z\right)\exp\left\{-\frac{\left(\ln\frac{z}{S(s)}-\nu\left(t-s\right)\right)^2}{2\sigma^2\left(t-s\right)}\right\}\frac{1}{\sigma z}\cdot\mathrm{d}z\\
=&\int_{-\infty}^{\infty}f\left(z\right)p\left(t-s,S\left(s\right),z\right)\mathrm{d}z\\
=&g\left(S\left(s\right)\right)
\end{aligned}
$$

□

Exercise 3.3. Let W be a Brownian motion. Fix $m>0$ and $\mu\in\mathbb{R}$. For $0\leqslant t<\infty$, define

$$X\left(t\right)=\mu t+W\left(t\right),\tau_m=\inf\left\{t\geqslant 0:X\left(t\right)=m\right\}$$

and set $\tau_m=\infty$ if $X(t)$ never reaches the level m. Let σ be a positive number and set

$$Z\left(t\right)=\exp\left\{\sigma X\left(t\right)-\left(\sigma\mu+\frac{1}{2}\sigma^2\right)t\right\}$$

1. Show that $Z(t), t\geqslant 0$, is a martingale.

Proof. For $t > s \geqslant 0$, we have

$$\begin{aligned}
\mathbb{E}\left[Z(t)\middle|\mathcal{F}(s)\right] &= \mathbb{E}\left[\exp\left\{\sigma X(t) - \sigma\mu t - \frac{1}{2}\sigma^2 t\right\}\middle|\mathcal{F}(s)\right] \\
&= \mathbb{E}\left[\exp\left\{\sigma\mu t + \sigma W(t) - \sigma\mu t - \frac{1}{2}\sigma^2 t\right\}\middle|\mathcal{F}(s)\right] \\
&= \mathbb{E}\left[\exp\left\{\sigma W(t) - \sigma W(s) + \sigma W(s) - \frac{1}{2}\sigma^2 t\right\}\middle|\mathcal{F}(s)\right] \\
&= \mathbb{E}\left[e^{\sigma W(t) - \sigma W(s)} \cdot \exp\left\{\sigma W(s) - \frac{1}{2}\sigma^2 t\right\}\middle|\mathcal{F}(s)\right] \\
&= e^{\frac{1}{2}\sigma^2(t-s)} \cdot \exp\left\{\sigma W(s) - \frac{1}{2}\sigma^2 t\right\} \\
&= \exp\left\{\sigma W(s) - \frac{1}{2}\sigma^2 s\right\} \\
&= Z(s)
\end{aligned}$$

And thus, $Z(t), t \geqslant 0$, is a martingale. □

2. Show that

$$\mathbb{E}\left[\exp\left\{\sigma X(t\wedge\tau_m) - \left(\sigma\mu + \frac{1}{2}\sigma^2\right)(t\wedge\tau_m)\right\}\right] = 1$$

Proof. By the Optional Sampling Lemma,

$$\mathbb{E}\left[Z(t\wedge\tau_m)\right] = \mathbb{E}\left[Z(0)\right] = \mathbb{E}\left[e^0\right] = 1$$ □

3. Now suppose $\mu \geqslant 0$. Show that, for $\sigma > 0$,

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

Use this fact to show $\mathbb{P}\{\tau_m < \infty\} = 1$ and to obtain the Laplace transform

$$\mathbb{E}e^{-\alpha\tau_m} = e^{m\mu - m\sqrt{2\alpha+\mu^2}}, \forall\alpha > 0$$

Proof. Note that

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

because $\lim_{t\to\infty} Z(t\wedge\tau_m)\,\mathbb{I}_{\{\tau_m=\infty\}} = \exp\left\{\sigma m - \left(\sigma\mu + \frac{1}{2}\sigma^2\right)\tau_m\right\}\mathbb{I}_{\{\tau_m=\infty\}} = 0$. Moreover, since $\mathbb{E}\left[Z(t\wedge\tau_m)\right] = 1$. Let $t\to\infty$, since $0 \leqslant Z(t\wedge\tau_m) \leqslant e^{\sigma m}$, by the Dominated Convergence Theorem,

$$\mathbb{E}\left[\mathbb{I}_{\{\tau_m<\infty\}}Z(\tau_m)\right] = \mathbb{E}\left[\lim_{t\to\infty} Z(t\wedge\tau_m)\right] = \lim_{t\to\infty}\mathbb{E}\left[Z(t\wedge\tau_m)\right] = 1$$

i.e.,

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

Let $\sigma\to 0$, $\mathbb{E}\left[\mathbb{I}_{\{\tau_m<\infty\}}\right] = 1$, i.e.,

$$\mathbb{P}\{\tau_m < \infty\} = 1$$

Therefore, we have

$$\mathbb{I}_{\{\tau_m<\infty\}}Z(\tau_m) = Z(\tau_m), \text{a.s.}$$

We can take the indicator function out of $\mathbb{E}\left[\mathbb{I}_{\{\tau_m<\infty\}} Z(\tau_m)\right]$, and get

$$\mathbb{E}[Z(\tau_m)] = \mathbb{E}\left[\mathbb{I}_{\{\tau_m<\infty\}} Z(\tau_m)\right] = e^{\sigma m} \cdot \mathbb{E}\left[\exp\left\{-\left(\sigma\mu + \frac{1}{2}\sigma^2\right)\tau_m\right\}\right] = 1$$

Let $\alpha = \sigma\mu + \frac{\sigma^2}{2} \Longrightarrow \sigma^2 + 2\mu\sigma - 2\alpha = 0 \Longrightarrow \sigma = -\mu + \sqrt{\mu^2 + 2\alpha}$, then we have

$$\mathbb{E}\left[e^{-\alpha\tau_m}\right] = e^{-\sigma m} = e^{-m\left(-\mu + \sqrt{\mu^2+2\alpha}\right)} = e^{m\mu - m\sqrt{\mu^2+2\alpha}}$$ □

4. Show that if $\mu > 0$, then $\mathbb{E}\tau_m < \infty$.

Proof. Differentiate the $\mathbb{E}\left[e^{-\alpha\tau_m}\right]$ with respect to α, we have

$$\frac{\partial}{\partial\alpha}\mathbb{E}\left[e^{-\alpha\tau_m}\right] = \mathbb{E}\left[-\tau_m e^{-\alpha\tau_m}\right] = \frac{\partial}{\partial\alpha} e^{m\mu - m\sqrt{\mu^2+2\alpha}} = \frac{-m}{\sqrt{\mu^2+2\alpha}} e^{m\mu - m\sqrt{\mu^2+2\alpha}}$$

Let $\alpha \to 0$,

$$\mathbb{E}[\tau_m] = \frac{m}{\sqrt{\mu^2}} = \frac{m}{\mu} < \infty$$ □

5. Now suppose $\mu < 0$. Show that, for $\sigma > -2\mu$,

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

Use this fact to show that $\mathbb{P}\{\tau_m < \infty\} = e^{-2m|\mu|} < 1$, and to obtain the Laplace transform

$$\mathbb{E}e^{-\alpha\tau_m} = e^{m\mu - m\sqrt{2\alpha+\mu^2}}, \forall \alpha > 0$$

Proof. By $\sigma > -2\mu > 0$, we get $\sigma\mu + \frac{1}{2}\sigma^2 > -2\mu^2 + \frac{4\mu^2}{2} = 0$. Then $0 \leqslant Z(t\wedge\tau_m) \leqslant e^{\sigma m}$ still holds. And

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

also holds, because $\lim_{t\to\infty} Z(t\wedge\tau_m)\,\mathbb{I}_{\{\tau_m=\infty\}} = \exp\left\{\sigma m - \left(\sigma\mu + \frac{1}{2}\sigma^2\right)\tau_m\right\}\mathbb{I}_{\{\tau_m=\infty\}} = 0$. And thus,

$$\begin{aligned}\mathbb{E}\left[\exp\left\{\sigma m - \left(\sigma\mu + \frac{1}{2}\sigma^2\right)\tau_m\right\}\mathbb{I}_{\{\tau_m<\infty\}}\right] &= \mathbb{E}\left[\lim_{t\to\infty} Z(\tau_m)\,\mathbb{I}_{\{\tau_m<\infty\}}\right] \\ &= \lim_{t\to\infty}\mathbb{E}[Z(t\wedge\tau_m)] \\ &= 1\end{aligned}$$

Let $\sigma \to -2\mu$, we get $\mathbb{E}\left[\mathbb{I}_{\{\tau_m<\infty\}}\right] = e^{2\mu m} = e^{-2m|\mu|} < 1$, i.e.,

$$\mathbb{P}\{\tau_m < \infty\} = e^{-2m|\mu|} < 1$$

Let $\alpha = \sigma\mu + \frac{\sigma^2}{2} > 0 \Longrightarrow \sigma = -\mu + \sqrt{\mu^2 + 2\alpha}$, and we can see that $e^{-\alpha\tau_m} = e^{-\alpha\tau_m}\mathbb{I}_{\{\tau_m<\infty\}}$ a.s., because $e^{-\alpha\tau_m}\mathbb{I}_{\{\tau_m=\infty\}} = 0$. Then we get

$$\mathbb{E}\left[e^{-\alpha\tau_m}\mathbb{I}_{\{\tau_m<\infty\}}\right] = e^{-\sigma m} = e^{m\mu - m\sqrt{\mu^2+2\alpha}}$$ □

Exercise 3.4. This problem presents the convergence of the distribution of stock prices in a sequence of binomial models to the distribution of geometric Brownian motion.

Let $\sigma > 0$ and $r \geqslant 0$ be given. For each positive integer n, we consider a binomial model taking n steps per unit time. In this model, the interest rate per period is $\frac{r}{n}$, the up factor is $u_n = e^{\sigma/\sqrt{n}}$, and the down factor is $d_n = e^{-\sigma/\sqrt{n}}$. The risk-neutral probabilities are

$$\tilde{p}_n = \frac{\frac{r}{n} + 1 - e^{-\frac{\sigma}{\sqrt{n}}}}{e^{\frac{\sigma}{\sqrt{n}}} - e^{-\frac{\sigma}{\sqrt{n}}}}, \tilde{q}_n = \frac{e^{-\frac{\sigma}{\sqrt{n}}} - \frac{r}{n} - 1}{e^{\frac{\sigma}{\sqrt{n}}} - e^{-\frac{\sigma}{\sqrt{n}}}}$$

Let t be an arbitrary positive rational number, and for each positive integer n for which nt is integer, define

$$M_{nt,n} = \sum_{k=1}^{nt} X_{k,n}$$

where $X_{1,n}, X_{2,n}, \cdots, X_{nt,n}$ are i.i.d. random variables with

$$\tilde{\mathbb{P}}\left\{X_{k,n} = 1\right\} = \tilde{p}_n, \tilde{\mathbb{P}}\left\{X_{k,n} = 1\right\} = \tilde{q}_n, k = 1, 2, \cdots nt$$

The stock price at time t in this binomial model, which is the result of nt steps from the initial time, is given by

$$\begin{aligned}
S_n(t) =& S(0)\, u_n^{\frac{1}{2}(nt+M_{nt,n})} d_n^{\frac{1}{2}(nt-M_{nt,n})} \\
=& S(0) \exp\left\{\frac{\sigma}{2\sqrt{n}}(nt + M_{nt,n})\right\} \exp\left\{-\frac{\sigma}{2\sqrt{n}}(nt - M_{nt,n})\right\} \\
=& S(0) \exp\left\{\frac{\sigma}{\sqrt{n}} M_{nt,n}\right\}
\end{aligned}$$

This problem shows that as $n \to \infty$, the distribution of the sequence of random variables $\frac{\sigma}{\sqrt{n}} M_{nt,n}$ appearing in the exponent above converges to the normal distribution with mean $\left(r - \frac{1}{2}\sigma^2\right)t$ and variance $\sigma^2 t$. Therefore, the limiting distribution of $S_n(t)$ is the same as the distribution of the geometric Brownian motion $S(0)\exp\left\{\sigma W(t) + \left(r - \frac{1}{2}\sigma\right)t\right\}$ at time t.

1. Show that the moment-generating function $\varphi_n(u)$ of $\frac{1}{\sqrt{n}} M_{nt,n}$ is given by

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

Proof. By the definition of moment-generating function, we have

$$\begin{aligned}
\varphi_n(x) =& \tilde{\mathbb{E}}\left[\exp\left\{\frac{u}{\sqrt{n}} M_{nt,n}\right\}\right] = \tilde{\mathbb{E}}\left[\exp\left\{\frac{u}{\sqrt{n}} \sum_{k=1}^{nt} X_{k,n}\right\}\right] \\
=& \tilde{\mathbb{E}}\left[\prod_{k=1}^{nt} \exp\left\{\frac{u}{\sqrt{n}} X_{k,n}\right\}\right] = \prod_{k=1}^{nt} \tilde{\mathbb{E}}\left[\exp\left\{\frac{u}{\sqrt{n}} X_{k,n}\right\}\right] \\
=& \left(\tilde{\mathbb{E}}\left[\exp\left\{\frac{u}{\sqrt{n}} X_{1,n}\right\}\right]\right)^{nt} = \left(e^{\frac{u}{\sqrt{n}}}\tilde{p}_n + e^{-\frac{u}{\sqrt{n}}}\tilde{q}_n\right)^{nt} \\
=& \left[e^{\frac{u}{\sqrt{n}}}\left(\frac{\frac{r}{n} + 1 - e^{-\frac{\sigma}{\sqrt{n}}}}{e^{\frac{\sigma}{\sqrt{n}}} - e^{-\frac{\sigma}{\sqrt{n}}}}\right) - e^{-\frac{u}{\sqrt{n}}}\left(\frac{\frac{r}{n} + 1 - e^{\frac{\sigma}{\sqrt{n}}}}{e^{\frac{\sigma}{\sqrt{n}}} - e^{-\frac{\sigma}{\sqrt{n}}}}\right)\right]^{nt}
\end{aligned}$$

□

2. We want to compute

$$\lim_{n\to\infty} \varphi_n(u) = \lim_{x\to 0} \varphi_{\frac{1}{x^2}}(u)$$

where we have made the change of variable $x = \frac{1}{\sqrt{n}}$. To do this, we will compute $\log \varphi_{\frac{1}{x^2}}(u)$ and then take the limit as $x \to 0$. Show that

$$\log \varphi_{\frac{1}{x^2}}(u) = \frac{t}{x^2} \log \left[\frac{\left(rx^2+1\right)\sinh ux + \sinh(\sigma - u)x}{\sinh \sigma x} \right]$$

where $\sinh x = \frac{e^x - e^{-x}}{2}$, $\cosh x = \frac{e^x + e^{-x}}{2}$. And use the formula

$$\sinh(A - B) = \sinh A \cosh B - \cosh A \sinh B$$

to rewrite this as

$$\log \varphi_{\frac{1}{x^2}}(u) = \frac{t}{x^2} \log \left[\cosh ux + \frac{\left(rx^2 + 1 - \cosh ux\right)\sinh ux}{\sinh \sigma x} \right]$$

Proof. By 1, we have

$$\varphi_n(x) = \left[e^{\frac{u}{\sqrt{n}}} \left(\frac{\frac{r}{n} + 1 - e^{-\frac{\sigma}{\sqrt{n}}}}{e^{\frac{\sigma}{\sqrt{n}}} - e^{-\frac{\sigma}{\sqrt{n}}}} \right) - e^{-\frac{u}{\sqrt{n}}} \left(\frac{\frac{r}{n} + 1 - e^{\frac{\sigma}{\sqrt{n}}}}{e^{\frac{\sigma}{\sqrt{n}}} - e^{-\frac{\sigma}{\sqrt{n}}}} \right) \right]^{nt}$$

Let $x = \frac{1}{\sqrt{n}}$, we have

$$\varphi_{\frac{1}{x^2}}(x) = \left[e^{ux} \left(\frac{rx^2 + 1 - e^{-\sigma x}}{e^{\sigma x} - e^{-\sigma x}} \right) - e^{-ux} \left(\frac{rx^2 + 1 - e^{\sigma x}}{e^{\sigma x} - e^{-\sigma x}} \right) \right]^{\frac{t}{x^2}}$$

By the definition and properties of sinh and cosh, we get

$$\begin{aligned}
\log \varphi_{\frac{1}{x^2}}(x) =& \frac{t}{x^2} \cdot \log \left[e^{ux} \left(\frac{rx^2 + 1 - e^{-\sigma x}}{e^{\sigma x} - e^{-\sigma x}} \right) - e^{-ux} \left(\frac{rx^2 + 1 - e^{\sigma x}}{e^{\sigma x} - e^{-\sigma x}} \right) \right] \\
=& \frac{t}{x^2} \cdot \log \left[\frac{\left(rx^2+1\right) e^{ux} - e^{ux - \sigma x} - \left(rx^2+1\right) e^{-ux} + e^{\sigma x - ux}}{e^{\sigma x} - e^{-\sigma x}} \right] \\
=& \frac{t}{x^2} \cdot \log \left[\frac{\left(rx^2+1\right)\sinh ux + \sinh(\sigma - u)x}{\sinh \sigma x} \right] \\
=& \frac{t}{x^2} \cdot \log \left[\frac{\left(rx^2+1\right)\sinh ux + \sinh \sigma x \cdot \cosh ux - \cosh \sigma x \cdot \sinh ux}{\sinh \sigma x} \right] \\
=& \frac{t}{x^2} \cdot \log \left[\cosh ux + \frac{\left(rx^2 + 1 - \cosh \sigma x\right) \cdot \sinh ux}{\sinh \sigma x} \right]
\end{aligned}$$

□

3. Use the Taylor's expansions of $\sinh x$ and $\cosh x$, we get

$$\cosh x = 1 + \frac{1}{2}x^2 + O\left(x^4\right), \sinh x = x + O\left(x^3\right)$$

to show that

$$\cosh ux + \frac{\left(rx^2 + 1 - \cosh ux\right)\sinh ux}{\sinh \sigma x} = 1 + \frac{1}{2}u^2x^2 + \frac{ru}{\sigma}x^2 - \frac{1}{2}u\sigma x^2 + O\left(x^4\right)$$

Proof. By the Taylor's expansions, we get

$$\begin{aligned}
&\cosh ux + \frac{\left(rx^2 + 1 - \cosh \sigma x\right) \cdot \sinh ux}{\sinh \sigma x} \\
=&1 + \frac{1}{2}u^2x^2 + O\left(x^4\right) + \frac{\left[rx^2 + 1 - 1 - \frac{1}{2}\sigma^2x^2 + O\left(x^4\right)\right] \cdot \left(ux + O\left(x^3\right)\right)}{\sigma x + O\left(x^3\right)} \\
=&1 + \frac{1}{2}u^2x^2 + \frac{urx^3 - \frac{1}{2}\sigma^2ux^3 + O\left(x^5\right)}{\sigma x + O\left(x^3\right)} + O\left(x^4\right) \\
=&1 + \frac{1}{2}u^2x^2 + \frac{urx^3 - \frac{1}{2}\sigma^2ux^3 + O\left(x^5\right)}{\sigma x\left[1 + O\left(x^2\right)\right]} + O\left(x^4\right) \\
=&1 + \frac{1}{2}u^2x^2 + \frac{ur}{\sigma}x^2 - \frac{1}{2}\sigma ux^2 + O\left(x^4\right)
\end{aligned}$$

□

4. Use the Taylor's expansion

$$\log(1 + x) = x + O\left(x^2\right)$$

to compute $\lim_{x\to 0} \log \varphi_{\frac{1}{x^2}}(u)$. Now explain how you know that the limiting distribution for $\frac{\sigma}{\sqrt{n}}M_{nt,n}$ is the normal distribution with mean $\left(r - \frac{1}{2}\sigma^2\right)t$ and variance $\sigma^2 t$.

Proof. By the Taylor's expansion of $\log x$, we get

$$\begin{aligned}
\lim_{x\to 0} \log \varphi_{\frac{1}{x^2}}(x) &= \lim_{x\to 0} \frac{t}{x^2} \cdot \log\left[\cosh ux + \frac{\left(rx^2 + 1 - \cosh \sigma x\right) \cdot \sinh ux}{\sinh \sigma x}\right] \\
&= \lim_{x\to 0} \frac{t}{x^2} \cdot \log\left[1 + \frac{1}{2}u^2x^2 + \frac{ur}{\sigma}x^2 - \frac{1}{2}\sigma ux^2 + o\left(x^4\right)\right] \\
&= \lim_{x\to 0} \frac{t}{x^2} \cdot \left[\frac{1}{2}u^2x^2 + \frac{ur}{\sigma}x^2 - \frac{1}{2}\sigma ux^2 + o\left(x^4\right)\right] \\
&= \frac{1}{2}tu^2 + \frac{t}{\sigma}\left(r - \frac{\sigma^2}{2}\right)u
\end{aligned}$$

Therefore,

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}}M_{nt,n} \sim N\left(\frac{t}{\sigma}\left(r - \frac{\sigma^2}{2}\right), t\right) \Longrightarrow \lim_{n\to\infty} \frac{\sigma}{\sqrt{n}}M_{nt,n} \sim N\left(t\left(r - \frac{\sigma^2}{2}\right), t\sigma^2\right)$$

□

Exercise 3.5. (**Laplace transform of first passage density**) Let $m > 0$ be given and define

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

which is the density in the variable t of the first passage time $\tau_m = \inf\{t \geqslant 0; W(t)\}$, where W is a Brownian motion. Let

$$g(\alpha, m) = \int_0^\infty e^{-\alpha t}f(t, m)\,\mathrm{d}t, \alpha > 0$$

be the Laplace transform of the density $f(t, m)$. The problem verifies that

$$g(\alpha, m) = \mathbb{E}e^{-\alpha\tau_m} = \int_0^\infty \frac{m}{t\sqrt{2\pi t}}e^{-\alpha m - \frac{m^2}{2t}}\,\mathrm{d}t = e^{-m\sqrt{2\alpha}}$$

where $\alpha > 0$, $m > 0$.

1. For $k \geqslant 1$, define
$$a_k(m) = \frac{1}{\sqrt{2\pi}} \int_0^\infty t^{-\frac{k}{2}} \exp\left\{-\alpha t - \frac{m^2}{2t}\right\} \mathrm{d}t$$
so $g(\alpha, m) = m a_3$. Show that
$$\begin{aligned} g_m(\alpha, m) &= a_3(m) - m^2 a_5(m) \\ g_{mm}(\alpha, m) &= -3m a_5(m) + m^3 a_7(m) \end{aligned}$$

Proof. Note that
$$\begin{aligned} \frac{\partial a_k}{\partial m} =& \frac{1}{\sqrt{2\pi}} \int_0^\infty t^{-\frac{k}{2}} \cdot \frac{\partial}{\partial m} \exp\left\{-\alpha t - \frac{m^2}{2t}\right\} \mathrm{d}t \\ =& \frac{1}{\sqrt{2\pi}} \int_0^\infty t^{-\frac{k}{2}} \cdot \left(-\frac{2m}{2t}\right) \exp\left\{-\alpha t - \frac{m^2}{2t}\right\} \mathrm{d}t \\ =& \frac{-m}{\sqrt{2\pi}} \int_0^\infty t^{-\frac{k+2}{2}} \exp\left\{-\alpha t - \frac{m^2}{2t}\right\} \mathrm{d}t \\ =& -m a_{k+2}(m) \end{aligned}$$

Therefore, we can easily get
$$\frac{\partial g}{\partial m} = a_3 + m\frac{\partial a_3}{\partial m} = a_3 - m^2 a_5$$
and
$$\frac{\partial^2 g}{\partial m^2} = 2\frac{\partial a_3}{\partial m} + m\frac{\partial^2 a_3}{\partial m^2} = -2m a_5 + m\frac{\partial}{\partial m}(-m a_5) = -3m a_5 + m^3 a_7$$
□

2. Use integration by parts to show that
$$a_5(m) = -\frac{2\alpha}{3} a_3(m) + \frac{m^2}{3} a_7(m)$$

Proof.
$$\begin{aligned} a_k(m) =& \frac{1}{\sqrt{2\pi}} \int_0^\infty t^{-\frac{k}{2}} \exp\left\{-\alpha t - \frac{m^2}{2t}\right\} \mathrm{d}t \\ =& -\frac{2}{k-2} \cdot \frac{1}{\sqrt{2\pi}} \int_0^\infty \exp\left\{-\alpha t - \frac{m^2}{2t}\right\} \mathrm{d}\left(t^{-k/2+1}\right) \\ =& -\frac{2}{k-2} \cdot \frac{1}{\sqrt{2\pi}} \int_0^\infty \exp\left\{-\alpha t - \frac{m^2}{2t}\right\} \mathrm{d}\left(t^{-k/2+1}\right) \\ =& -\frac{2}{k-2} \cdot \left[\frac{1}{\sqrt{2\pi}} t^{-\frac{k-2}{2}} e^{-\alpha t - \frac{m^2}{2t}}\Big|_0^\infty - \frac{1}{\sqrt{2\pi}} \int_0^\infty t^{-\frac{k-2}{2}} \mathrm{d}e^{-\alpha t - \frac{m^2}{2t}}\right] \\ =& \frac{2}{k-2} \frac{1}{\sqrt{2\pi}} \int_0^\infty t^{-\frac{k-2}{2}} \exp\left\{-\alpha t - \frac{m^2}{2t}\right\} \left(-\alpha + \frac{m^2}{2t^2}\right) \mathrm{d}t \\ =& \frac{2}{k-2} \frac{1}{\sqrt{2\pi}} \left[-\alpha \int_0^\infty t^{-\frac{k-2}{2}} \exp\left\{-\alpha t - \frac{m^2}{2t}\right\} \mathrm{d}t + \int_0^\infty \frac{m^2}{2t^2} t^{-\frac{k-2}{2}} \exp\left\{-\alpha t - \frac{m^2}{2t}\right\} \mathrm{d}t\right] \\ =& \frac{2}{k-2} \left[-\alpha a_{k-2}(m) + \frac{m^2}{2} a_{k+2}(m)\right] \end{aligned}$$

Plug $k = 5$, we obtain $a_5(m) = -\frac{2\alpha}{3} a_3(m) + \frac{m^2}{3} a_7(m)$. □

3. Use 1 and 2 to show that g satisfies the second-order ordinary differential equation

$$g_{mm}(\alpha, m) = 2\alpha g(\alpha, m)$$

Proof. Note that $g = ma_3$

$$g_{mm} = -3ma_5 + m^3a_7 = -3m\left[-\frac{2\alpha}{3}a_3 + \frac{m^2}{3}a_7\right] + m^3a_7 = 6\alpha ma_3 = 2\alpha g(\alpha, m)$$

□

4. (**General solution of second-order ODE**) For a second-order ODE

$$ay''(m) + by'(m) + cy(m) = 0$$

the general solution is

$$y(m) = A_1e^{\lambda_1 m} + A_2e^{\lambda_2 m}$$

where λ_1 and λ_2 are roots of the characteristic equation

$$a\lambda^2 + b\lambda + c = 0$$

Here we are assuming that these roots are distinct. Find the general solution of the equation

$$g_{mm}(\alpha, m) = 2\alpha g(\alpha, m), \alpha > 0$$

The solution has two undetermined parameters A_1 and A_2, and these may depend on α.

Solution. The corresponding characteristic equation is

$$\lambda^2 = 2\alpha$$

So the two characteristic solutions are $\lambda_1 = \sqrt{2\alpha}$ and $\lambda_2 = -\sqrt{2\alpha}$, and thus the general solution is

$$g(\alpha, m) = A_1e^{m\sqrt{2\alpha}} + A_2e^{-m\sqrt{2\alpha}}$$

□

5. Derive the bound

$$g(\alpha, m) \leqslant \frac{m}{\sqrt{2\pi}}\int_0^m \sqrt{\frac{m}{t}} \cdot t^{-\frac{3}{2}}e^{-\frac{m^2}{2}}\mathrm{d}t + \frac{1}{\sqrt{2\pi m}}\int_m^\infty e^{-\alpha t}\mathrm{d}t$$

and use it to show that, for every $\alpha > 0$,

$$\lim_{m\to\infty} g(\alpha, m) = 0$$

Use this fact to determine one of the parameters in the general solution to the equation

$$g_{mm}(\alpha, m) = 2\alpha g(\alpha, m), \alpha > 0$$

Solution. We have

$$\begin{aligned}
g(\alpha,m) =& \frac{m}{\sqrt{2\pi}}\int_0^\infty t^{-\frac{3}{2}}\exp\left\{-\alpha t-\frac{m^2}{2t}\right\}\mathrm{d}t\\
=& \frac{m}{\sqrt{2\pi}}\left[\int_0^m t^{-\frac{3}{2}}\exp\left\{-\alpha t-\frac{m^2}{2t}\right\}\mathrm{d}t+\int_m^\infty t^{-\frac{3}{2}}\exp\left\{-\alpha t-\frac{m^2}{2t}\right\}\mathrm{d}t\right]\\
\leqslant& \frac{m}{\sqrt{2\pi}}\left[\int_0^m e^{-\alpha t}t^{-\frac{3}{2}}\exp\left\{-\frac{m^2}{2t}\right\}\mathrm{d}t+\int_m^\infty m^{-\frac{3}{2}}e^{-\alpha t}\mathrm{d}t\right]\\
\leqslant& \frac{m}{\sqrt{2\pi}}\left[\int_0^m \sqrt{\frac{m}{t}}t^{-\frac{3}{2}}\exp\left\{-\frac{m^2}{2t}\right\}\mathrm{d}t+\int_m^\infty m^{-\frac{3}{2}}e^{-\alpha t}\mathrm{d}t\right]\\
=& \frac{m}{\sqrt{2\pi}}\int_0^m \sqrt{\frac{m}{t}}t^{-\frac{3}{2}}\exp\left\{-\frac{m^2}{2t}\right\}\mathrm{d}t+\frac{1}{\sqrt{2\pi m}}\int_m^\infty e^{-\alpha t}\mathrm{d}t
\end{aligned}$$

Note that $\alpha>0$, $m>0$. So

(a) $\exp\left\{-\frac{m^2}{2t}\right\}<e^0=1$.

(b) If $0<t\leqslant m$, we get $e^{-\alpha t}<1\leqslant\sqrt{\frac{m}{t}}$.

(c) If $0<m\leqslant t$, we get $t^{-3/2}\leqslant m^{-3/2}$.

Let $m\to\infty$, we have

$$\begin{aligned}
\lim_{m\to\infty}g(\alpha,m)\leqslant& \lim_{m\to\infty}\left[\frac{m}{\sqrt{2\pi}}\int_0^m\sqrt{\frac{m}{t}}t^{-\frac{3}{2}}\exp\left\{-\frac{m^2}{2t}\right\}\mathrm{d}t+\frac{1}{\sqrt{2\pi m}}\int_m^\infty e^{-\alpha t}\mathrm{d}t\right]\\
=&\lim_{m\to\infty}\frac{m^{3/2}}{\sqrt{2\pi}}\int_0^m t^{-2}\exp\left\{-\frac{m^2}{2t}\right\}\mathrm{d}t+0\\
=&\lim_{m\to\infty}\frac{m^{3/2}}{\sqrt{2\pi}}\int_0^m \exp\left\{-\frac{m^2}{2t}\right\}\mathrm{d}\left(-\frac{1}{t}\right)\\
=&\lim_{m\to\infty}\frac{m^{3/2}}{\sqrt{2\pi}}\int_{-\infty}^{-\frac{1}{m}} \exp\left\{\frac{m^2}{2}x\right\}\mathrm{d}x\\
=&\lim_{m\to\infty}\frac{m^{3/2}}{\sqrt{2\pi}}\left(\frac{2}{m^2}\exp\left\{\frac{m^2}{2}x\right\}\Big|_{-\infty}^{-\frac{1}{m}}\right)\\
=&\lim_{m\to\infty}\frac{m^{3/2}}{\sqrt{2\pi}}\left(\frac{2}{m^2}\exp\left\{-\frac{m}{2}\right\}-0\right)\\
=&\lim_{m\to\infty}\frac{2}{\sqrt{2\pi m}e^{m/2}}\\
=&0
\end{aligned}$$

Note that $\lim_{m\to\infty}\frac{1}{\sqrt{2\pi m}}\int_m^\infty e^{-\alpha t}\mathrm{d}t=0$. Therefore,

$$\lim_{m\to\infty}g(\alpha,m)=\lim_{m\to\infty}\left[A_1e^{m\sqrt{2\alpha}}+A_2e^{-m\sqrt{2\alpha}}\right]=0$$

which must require $A_1=0$ (necessary condition). And thus, we have

$$g(\alpha,m)=A_2e^{-m\sqrt{2\alpha}}$$

□

6. Using first the change of variable $s=\frac{t}{m^2}$ and then the change of variable $y=\frac{1}{\sqrt{s}}$, show that

$$\lim_{m\to 0}g(\alpha,m)=1$$

Use this fact to determine the other parameter in the general solution to the equation

$$g_{mm}(\alpha, m) = 2\alpha g(\alpha, m), \alpha > 0$$

Solution.

$$\begin{aligned}
g(\alpha, m) =& \frac{m}{\sqrt{2\pi}} \int_0^{\infty} t^{-\frac{3}{2}} \exp\left\{-\alpha t - \frac{m^2}{2t}\right\} \mathrm{d}t \\
=& \frac{m}{\sqrt{2\pi}} \int_0^{\infty} \left(sm^2\right)^{-\frac{3}{2}} \exp\left\{-\alpha sm^2 - \frac{m^2}{2sm^2}\right\} \mathrm{d}\left(sm^2\right) \text{ (Let } s = \frac{t}{m^2}) \\
=& \frac{m^3}{\sqrt{2\pi}} \int_0^{\infty} s^{-\frac{3}{2}} m^{-3} \exp\left\{-\alpha sm^2 - \frac{1}{2s}\right\} \mathrm{d}s \\
=& \frac{1}{\sqrt{2\pi}} \int_0^{\infty} s^{-\frac{3}{2}} \exp\left\{-\alpha sm^2 - \frac{1}{2s}\right\} \mathrm{d}s \\
=& \frac{1}{\sqrt{2\pi}} \int_{\infty}^{0} y^3 \exp\left\{-\frac{\alpha m^2}{y^2} - \frac{y^2}{2}\right\} \mathrm{d}\left(\frac{1}{y^2}\right) \text{ (Let } y = \frac{1}{\sqrt{s}}) \\
=& \frac{2}{\sqrt{2\pi}} \int_0^{\infty} \exp\left\{-\frac{\alpha m^2}{y^2} - \frac{y^2}{2}\right\} \mathrm{d}y
\end{aligned}$$

Moreover, because $0 < \exp\left\{-\frac{\alpha m^2}{y^2} - \frac{y^2}{2}\right\} \leqslant 1$, then by the Dominated Convergence Theorem, we have

$$\begin{aligned}
\lim_{m \to 0} g(\alpha, m) =& \lim_{m \to 0} \frac{2}{\sqrt{2\pi}} \int_0^{\infty} \exp\left\{-\frac{\alpha m^2}{y^2} - \frac{y^2}{2}\right\} \mathrm{d}y \\
=& \frac{2}{\sqrt{2\pi}} \int_0^{\infty} \lim_{m \to 0} e^{-\frac{\alpha m^2}{y^2}} \cdot \mathrm{d}y \\
=& \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} \mathrm{d}y \\
=& 1
\end{aligned}$$

Therefore,

$$\lim_{m \to 0} g(\alpha, m) = \lim_{m \to 0} \left[A_2 e^{-m\sqrt{2\alpha}}\right] = A_2 = 1$$

And hence,

$$g(\alpha, m) = e^{-m\sqrt{2\alpha}}$$

which is just what we want to proof. □

第四章　随机积分

4.1 Itô's Integral for Simple Integrands

1. (Definition) Define $I(t)=\int_0^T \Delta(t)\,\mathrm{d}W(t)$ for simple process $\Delta(t)=\sum_{k=0}^{n}\mathbb{I}_{[t_k,t_{k+1})}\Delta(t_k)$.

   (a) Partition: $\Pi=\{t_0,t_1,\cdots,t_n\}$ of $[0,T]$.

   (b) Calculate

$$
\begin{aligned}
0 \leqslant &t \leqslant t_1, I(t)=\Delta(t_0)\left[W(t_1)-W(t_0)\right]=\Delta(0)\,W(t_1)\\
t_1 \leqslant &t \leqslant t_2, I(t)=\Delta(0)\,W(t_1)+\Delta(t_1)\left[W(t_2)-W(t_1)\right]\\
t_2 \leqslant &t \leqslant t_3, I(t)=\Delta(0)\,W(t_1)+\Delta(t_1)\left[W(t_2)-W(t_1)\right]+\Delta(t_2)\left[W(t_3)-W(t_2)\right]\\
&\vdots
\end{aligned}
$$

   In general, for $t_k \leqslant t \leqslant t_{k+1}$, then

$$
\begin{aligned}
I(t) &= \sum_{j=0}^{k-1}\Delta(t_j)\left[W(t_{j+1})-W(t_j)\right]+\Delta(t_k)\left[W(t)-W(t_k)\right]\\
&=\int_0^t \Delta(u)\,\mathrm{d}W(u)
\end{aligned}
$$

2. (Theorem) Itô's Integral of simple process is a martingale.

   Proof. Assume that $0\leqslant s\leqslant t\leqslant T$, $s\in[t_\ell,t_{\ell+1})\,,t\in[t_k,t_{k+1})$. Then

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

   Take conditional expectation respect to $\mathcal{F}(s)$ on both sides, we get

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

   And

$$
\begin{aligned}
\mathbb{E}\left[\Delta(t_\ell)\left[W(t_{\ell+1})-W(t_k)\right]\middle|\,\mathcal{F}(s)\right]&=\Delta(t_\ell)\,\mathbb{E}\left[W(t_{\ell+1})-W(t_\ell)\,|\mathcal{F}(s)\right]\\
&=\Delta(t_\ell)\left[W(s)-W(t_\ell)\right]
\end{aligned}
$$

Note that $t_j \geqslant t_{\ell+1} > s$, thus $\mathcal{F}(s) \subset \mathcal{F}(t_{\ell+1}) \subset \mathcal{F}(t_j)$, so

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

Likewise,

$$\mathbb{E}\left[\mathit{\Delta}(t_\ell)[W(t) - W(t_k)] \middle| \mathcal{F}(s)\right] = \mathbb{E}\left[\mathbb{E}\left[\mathit{\Delta}(t_\ell)[W(t) - W(t_k)] \middle| \mathcal{F}(t_k)\right] \middle| \mathcal{F}(s)\right] = 0$$

Thus,

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

□

3. (**Itô isometry**) The Itô integral of simple process satisfies

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

Proof.

$$\begin{aligned}
I^2(t) =& \left(\sum_{j=0}^{k} \mathit{\Delta}(t_j)[W(t_{j+1}) - W(t_j)]\right)^2 \\
=&\sum_{j=0}^{k} \mathit{\Delta}^2(t_j)[W(t_{j+1}) - W(t_j)]^2 \\
&+2\sum_{i=0}^{k}\sum_{j=i+1}^{k} \mathit{\Delta}(t_i)\mathit{\Delta}(t_j)[W(t_{i+1}) - W(t_i)][W(t_{j+1}) - W(t_j)]
\end{aligned}$$

Set $D_j = W(t_{j+1}) - W(t_j)$ for $j = 0, 1, \cdots, k-1$ and $D_k = W(t) - W(t_k)$, then

$$I(t) = \sum_{j=0}^{k} \mathit{\Delta}(t_j) D_j$$

and

$$\mathbb{E} I^2(t) = \mathbb{E}\left[\sum_{j=0}^{k} \mathit{\Delta}^2(t_j) D_j^2\right] + 2\mathbb{E}\left[\sum_{i=0}^{k}\sum_{j=i+1}^{k} \mathit{\Delta}(t_i)\mathit{\Delta}(t_j) D_i D_j\right]$$

Because for $i < j$, the random variable $\Delta(t_i)\Delta(t_j)D_j$ is $\mathcal{F}(t_j)$-measurable, the second term is

$$
\begin{aligned}
2\mathbb{E}\left[\sum_{i=0}^{k}\sum_{j=i+1}^{k}\Delta(t_i)\Delta(t_j)D_iD_j\right] =& 2\sum_{i=0}^{k}\sum_{j=i+1}^{k}\mathbb{E}\left[\Delta(t_i)\Delta(t_j)D_iD_j\right] \\
=& 2\sum_{i=0}^{k}\sum_{j=i+1}^{k}\mathbb{E}\left[\mathbb{E}\left[\Delta(t_i)\Delta(t_j)D_iD_j\right]\middle|\mathcal{F}(t_j)\right] \\
=& 2\sum_{i=0}^{k}\sum_{j=i+1}^{k}\mathbb{E}\left[\Delta(t_i)\Delta(t_j)D_i\mathbb{E}\left[D_j|\mathcal{F}(t_j)\right]\right] \\
=& 0
\end{aligned}
$$

while the first term

$$
\begin{aligned}
\mathbb{E}\left[\sum_{j=0}^{k}\Delta^2(t_j)D_j^2\right] =& \sum_{j=0}^{k}\mathbb{E}\left[\Delta^2(t_j)D_j^2\right] \\
=& \sum_{j=0}^{k}\mathbb{E}\left[\mathbb{E}\left[\Delta^2(t_j)D_j^2|\mathcal{F}(t_j)\right]\right] \\
=& \sum_{j=0}^{k}\mathbb{E}\left[\Delta^2(t_j)\mathbb{E}\left[D_j^2|\mathcal{F}(t_j)\right]\right] \\
=& \sum_{j=0}^{k}\mathbb{E}\left[\Delta^2(t_j)(t_{j+1}-t_j)\right] \\
=& \mathbb{E}\int_0^t \Delta^2(u)\,\mathrm{d}u
\end{aligned}
$$

□

4. (Theorem) The quadratic variation accumulated up to time t by the Itô integral of simple process is

$$[I, I](t) = \int_0^t \Delta^2(u)\,\mathrm{d}u$$

Proof. On one of the sub-intervals $[t_j, t_{j+1}]$, we choose a partition

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

and consider

$$\sum_{i=0}^{m-1}[I(s_{i+1}) - I(s_i)]^2 = \sum_{i=0}^{m-1}[\Delta(t_j)[W(s_{i+1}) - W(s_i)]]^2 = \Delta^2(t_j)\sum_{i=0}^{m-1}[W(s_{i+1}) - W(s_i)]^2$$

As $m \to \infty$, $[W, W](s_i, s_{i+1}) = s_{i+1} - s_i$, so

$$[I, I](t_j, t_{j+1}) = \Delta^2(t_j)\sum_{i=0}^{m-1}(s_{i+1} - s_i) = \Delta^2(t_j)(t_{j+1} - t_j) = \int_{t_j}^{t_{j+1}}\Delta^2(u)\,\mathrm{d}u, t_j \leqslant u \leqslant t_{j+1}$$

Add all sub-intervals up, we get the theorem. □

5. (Remark) For a simple process $\Delta(t)$, we have following notations.

   (a) $I(t) = \int_0^t \Delta(u)\,\mathrm{d}W(u)$, and we set $I(0) = 0$. Then we get $\mathrm{d}I(t) = \Delta(t)\,\mathrm{d}W(t)$ or $I(t) = I(0) + \int_0^t \Delta(u)\,\mathrm{d}W(u)$.

   (b) $\mathrm{d}I(t)\,\mathrm{d}I(t) = \Delta^2(t)\,\mathrm{d}W(t)\,\mathrm{d}W(t) = \Delta^2(t)\,\mathrm{d}t$

4.2 Itô's Integral for General Integrands

1. (Definition) Assume general integrand $\Delta^2(t)$ satisfies the square-integrability condition:

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

Define *the Itô integrals for general integrand* $\Delta^2(t)$ by choosing a series of simple processes $\Delta_n(t)$, such that $\Delta_n(t)$ "converges" to $\Delta(t)$, i.e.,

$$\lim_{n\to\infty}\mathbb{E}\int_0^T |\Delta_n(t)-\Delta(t)|^2\,\mathrm{d}t = 0$$

we define Itô integral for the continuously varying integrand $\Delta(t)$ by the formula

$$\int_0^t \Delta(u)\,\mathrm{d}W(u) = \lim_{n\to\infty}\int_0^t \Delta_n(u)\,\mathrm{d}W(u), 0\leqslant t\leqslant T$$

For each t, the limit exists because $I_n(t)=\int_0^t \Delta_n(t)\,\mathrm{d}W(u)$ is a Cauchy sequence in $L_2(\Omega,\mathcal{F},\mathbb{P})$. This is because of Itô isometry, which yields

$$\mathbb{E}\left[I_n(t)-I_m(t)\right]^2 = \mathbb{E}\int_0^t |\Delta_n(u)-\Delta_m(u)|^2\,\mathrm{d}u \to 0$$

as $m,n\to\infty$.

2. (Theorem) Let T be a positive constant and let $\Delta(t), 0\leqslant t\leqslant T$, be an adapted stochastic process that satisfied square-integrability condition. Then Itô integral of $\Delta(t)$ has the following properties.

   (a) (**Continuity**) As a function of the upper limit of integration t, the paths of $I(t)$ are continuous.

   (b) (**Adaptivity**) For each t, $I(t)$ is $\mathcal{F}(t)$-measurable.

   (c) (**Linearity**) Let $I(t)=\int_0^t \Delta(u)\,\mathrm{d}W(u)$ and $J(t)=\int_0^t \Gamma(u)\,\mathrm{d}W(u)$, then

   $$I(t)\pm J(t) = \int_0^t \left[\Delta(u)\pm\Gamma(u)\right]\mathrm{d}W(u)$$

   Furthermore, for every constant c,

   $$cI(t) = \int_0^t c\Delta(u)\,\mathrm{d}W(u)$$

   (d) (**Martingale**) $I(t)$ is a martingale.

   (e) (**Itô isometry**)

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

   (f) (**Quadratic variation**)

   $$[I,I](t) = \int_0^t \Delta^2(u)\,\mathrm{d}u$$

注：详细证明见随机过程的内容。

3. (Example) Compute $\int_0^T W(t)\,\mathrm{d}W(t)$.

Solution. We choose a large integer n and approximate the integrand $\Delta(t) = W(t)$ by the simple process

$$\Delta_n(t) = \sum_{k=0}^{n-1} \mathbb{I}_{[t_k, t_{k+1})} W(t_k) = \begin{cases} W(0), & 0 \leqslant t < \frac{T}{n} \\ W\left(\frac{T}{n}\right), & \frac{T}{n} \leqslant t < \frac{2T}{n} \\ \vdots & \\ W\left(\frac{n-1}{n}T\right), & \frac{n-1}{n}T \leqslant t \leqslant T \end{cases}$$

Then

$$\lim_{n\to\infty} \int_0^T |\Delta_n(t) - W(t)|^2 \, \mathrm{d}t = 0$$

By definition,

$$\begin{aligned} \int_0^T W(t) \, \mathrm{d}W(t) &= \lim_{n\to\infty} \int_0^T \Delta_n(t) \, \mathrm{d}t \\ &= \lim_{n\to\infty} \sum_{k=0}^{n-1} W\left(\frac{k}{n}T\right) \left[W\left(\frac{k+1}{n}T\right) - W\left(\frac{k}{n}T\right) \right] \end{aligned}$$

To simplify notation, we denote $W_k = W\left(\frac{k}{n}T\right)$, and we first work to get

$$\begin{aligned} \frac{1}{2}\sum_{k=0}^{n-1} (W_{k+1} - W_k)^2 =& \frac{1}{2}\sum_{k=0}^{n-1} W_{k+1}^2 - \sum_{k=0}^{n-1} W_k W_{k+1} + \frac{1}{2}\sum_{k=0}^{n-1} W_k^2 \\ =& \frac{1}{2}W_n^2 + \frac{1}{2}\sum_{k=0}^{n-1} W_k^2 - \sum_{k=0}^{n-1} W_k W_{k+1} + \frac{1}{2}\sum_{k=0}^{n-1} W_k^2 \\ =& \frac{1}{2}W_n^2 + \sum_{k=0}^{n-1} W_k^2 - \sum_{k=0}^{n-1} W_k W_{k+1} \\ =& \frac{1}{2}W_n^2 + \sum_{k=0}^{n-1} W_k (W_k - W_{k+1}) \end{aligned}$$

And thus, we have

$$\sum_{k=0}^{n-1} W_k (W_k - W_{k+1}) = \frac{1}{2}\sum_{k=0}^{n-1} (W_{k+1} - W_k)^2 - \frac{1}{2}W_n^2$$

Therefore, we get

$$\begin{aligned} \int_0^T W(t) \, \mathrm{d}W(t) &= \lim_{n\to\infty} \sum_{k=0}^{n-1} W\left(\frac{k}{n}T\right) \left[W\left(\frac{k+1}{n}T\right) - W\left(\frac{k}{n}T\right) \right] \\ &= \lim_{n\to\infty} \left[\frac{1}{2}W^2\left(\frac{n}{n}T\right) - \frac{1}{2}\sum_{k=0}^{n-1} \left[W\left(\frac{k+1}{n}T\right) - W\left(\frac{k}{n}T\right) \right]^2 \right] \\ &= \frac{1}{2}W^2(T) + \frac{1}{2}T \end{aligned}$$

□

Remark: The extra term $-\frac{1}{2}T$ in Itô integral comes from the nonzero quadratic variation of Brownian motion.

(a) Itô integral:

$$\int_0^T W(t)\,\mathrm{d}W(t) = \frac{1}{2}W^2(T) - [W,W](T) = \frac{1}{2}W^2(T) - \frac{1}{2}T$$

(b) Riemann integral: Note that $g(0) = 0$,

$$\int_0^T g(t)\,\mathrm{d}g(t) = \int_0^T g(t)\,g'(t)\,\mathrm{d}t = \frac{1}{2}g^2(t)\Big|_0^T = \frac{1}{2}g^2(T)$$

4.3 Itô-Doeblin Formula

4.3.1 Formula for Brownian Motion

1. (Theorem) Let $f(t,x)$ be a function for which the partial derivatives $f_t(t,x)$, $f_x(t,x)$ and $f_{xx}(t,x)$ are defined and continuous, and let $W(t)$ be a Brownian motion. Then, for every $T \geqslant 0$,

$$f(T, W(T)) = f(0, W(0)) + \int_0^T f_t(t, W(t))\,\mathrm{d}t + \int_0^T f_x(t, W(t))\,\mathrm{d}W(t) + \frac{1}{2}\int_0^T f_{xx}(t, W(t))\,\mathrm{d}t$$

2. (Remark) Another form of Itô-Doeblin Formula.

$$\begin{aligned} \mathrm{d}f(t, W(t)) = & f_t(t, W(t))\,\mathrm{d}t + f_x(t, W(t))\,\mathrm{d}W(t) + \frac{1}{2}f_{xx}(t, W(t))\,\mathrm{d}W(t)\,\mathrm{d}W(t) \\ & + f_{tx}(t, W(t))\,\mathrm{d}t\mathrm{d}W(t) + \frac{1}{2}f_{tt}(t, W(t))\,\mathrm{d}t\mathrm{d}t \end{aligned}$$

But

$$\mathrm{d}W(t)\,\mathrm{d}W(t) = \mathrm{d}t, \mathrm{d}t\mathrm{d}W(t) = \mathrm{d}W(t)\,\mathrm{d}t = 0, \mathrm{d}t\mathrm{d}t = 0$$

Thus

$$\mathrm{d}f(t, W(t)) = f_t(t, W(t))\,\mathrm{d}t + f_x(t, W(t))\,\mathrm{d}W(t) + \frac{1}{2}f_{xx}(t, W(t))\,\mathrm{d}t$$

4.3.2 Formula for Itô Process

1. (Remark) Almost all stochastic processes, except those that have jumps, are Itô process.

2. (Definition) Let $W(t), t \geqslant 0$, be a Brownian motion, and let $\mathcal{F}(t)$, $t \geqslant 0$, be an associated filtration. An *Itô process* is a stochastic process of the form

$$X(t) = X(0) + \int_0^t \Delta(u)\,\mathrm{d}W(u) + \int_0^t \Theta(u)\,\mathrm{d}u$$

where $X(0)$ is nonrandom and $\Delta(u)$ and $\Theta(u)$ are adapted stochastic processes. And we may assume that, for every $t > 0$, $\mathbb{E}\int_0^t \Delta^2(u)\,\mathrm{d}u < \infty$, $\mathbb{E}\int_0^t |\Theta(u)|\,\mathrm{d}u < \infty$, so that their Itô integrals are well defined.

3. (Lemma) The quadratic variation of the Itô process is

$$[X, X](t) = \int_0^t \Delta^2(u)\,\mathrm{d}u$$

Proof. Let $I(t) = \int_0^t \Delta(u)\,\mathrm{d}W(u), R(t) = \int_0^t \Theta(u)\,\mathrm{d}u$, both of which are continuous in their upper limit of integration t. We choose a partition $\Pi = \{t_0, t_1, \cdots, t_n\}$ of $[0, t]$, and we write the sampled quadratic variation

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

As $\|\Pi\| \to 0$, the first term on the right-hand side converges to the quadratic variation of I on $[0, t]$, which is $[I, I](t) = \int_0^t \Delta^2(u)\,\mathrm{d}u$. The second term

$$\begin{aligned}\sum_{j=0}^{n-1} [R(t_{j+1}) - R(t_j)]^2 \leqslant& \max_{0\leqslant j\leqslant n-1} |R(t_{j+1}) - R(t_j)| \cdot \sum_{j=0}^{n-1} |R(t_{j+1}) - R(t_j)| \\ =& \max_{0\leqslant j\leqslant n-1} |R(t_{j+1}) - R(t_j)| \cdot \sum_{j=0}^{n-1} \left| \int_{t_j}^{t_{j+1}} \Theta(u)\,\mathrm{d}u \right| \\ \leqslant& \max_{0\leqslant j\leqslant n-1} |R(t_{j+1}) - R(t_j)| \cdot \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} |\Theta(u)|\,\mathrm{d}u \\ =& \max_{0\leqslant j\leqslant n-1} |R(t_{j+1}) - R(t_j)| \cdot \int_0^t |\Theta(u)|\,\mathrm{d}u\end{aligned}$$

and as $\|\Pi\| \to 0$, $\max\limits_{0\leqslant j\leqslant n-1} |R(t_{j+1}) - R(t_j)| \to 0$, because $R(t)$ is continuous, and $\int_0^t |\Theta(u)|\,\mathrm{d}u < \infty$. The third term

$$\begin{aligned}&2\sum_{j=0}^{n-1} [I(t_{j+1}) - I(t_j)][R(t_{j+1}) - R(t_j)] \\ \leqslant& 2\max_{0\leqslant j\leqslant n-1} |I(t_{j+1}) - I(t_j)| \sum_{j=0}^{n-1} |R(t_{j+1}) - R(t_j)| \\ \leqslant& 2\max_{0\leqslant j\leqslant n-1} |I(t_{j+1}) - I(t_j)| \cdot \int_0^t |\Theta(u)|\,\mathrm{d}u\end{aligned}$$

and as $\|\Pi\| \to 0$, $\max\limits_{0\leqslant j\leqslant n-1} |I(t_{j+1}) - I(t_j)| \to 0$, because $I(t)$ is continuous, while $\int_0^t |\Theta(u)|\,\mathrm{d}u < \infty$. □

4. (Remark) The differential notation of Itô process.

$$\mathrm{d}X(t) = \Delta(t)\,\mathrm{d}W(t) + \Theta(t)\,\mathrm{d}t$$

we can compute quadratic variation by this form as following.

$$\mathrm{d}X(t) \cdot \mathrm{d}X(t) = \Delta^2(t)\,\mathrm{d}W(t)\,\mathrm{d}W(t) + \Theta^2(t)\,\mathrm{d}t\mathrm{d}t + 2\Delta(t)\,\Theta(t)\,\mathrm{d}W(t)\,\mathrm{d}t = \Delta^2(t)\,\mathrm{d}t$$

5. (Definition) Let $X(t), t \geqslant 0$, be an Itô process, and let $\Gamma(t), t \geqslant 0$, be an adapted process. We define the *integral with respect to an Itô process* as

$$\int_0^t \Gamma(u)\,\mathrm{d}X(u) = \int_0^t \Gamma(u)\,\Delta(u)\,\mathrm{d}W(u) + \int_0^t \Gamma(u)\,\Theta(u)\,\mathrm{d}u$$

6. (Theorem) Let $X(t), t \geqslant 0$, be an Itô process, and let $f(t,x)$ be a function for which the partial derivatives $f_t(t,x)$, $f_x(t,x)$, and $f_{xx}(t,x)$ are defined and continuous. Then, for every $T \geqslant 0$,

$$\begin{aligned} f(T, X(T)) =& f(0, X(0)) + \int_0^T f_t(t, X(t))\,\mathrm{d}t + \int_0^T f_x(t, X(t))\,\mathrm{d}X(t) \\ &+ \frac{1}{2}\int_0^T f_{xx}(t, X(t))\,\mathrm{d}[X,X](t) \\ =& f(0, X(0)) + \int_0^T f_t(t, X(t))\,\mathrm{d}t + \int_0^T f_x(t, X(t))\,\varDelta(t)\,\mathrm{d}W(t) \\ &+ \int_0^T f_x(t, X(t))\,\varTheta(t)\,\mathrm{d}t + \frac{1}{2}\int_0^T f_{xx}(t, X(t))\,\varDelta^2(t)\,\mathrm{d}t \end{aligned}$$

7. (Remark) Differential notation.

$$\begin{aligned} \mathrm{d}f(t, X(t)) =& f_t(t, X(t))\,\mathrm{d}t + f_x(t, X(t))\,\mathrm{d}X(t) + \frac{1}{2}f_{xx}(t, X(t))\,\mathrm{d}X(t)\,\mathrm{d}X(t) \\ =& f_t(t, X(t))\,\mathrm{d}t + f_x(t, X(t))\,\varDelta(t)\,\mathrm{d}W(t) + f_x(t, X(t))\,\varTheta(t)\,\mathrm{d}t \\ &+ \frac{1}{2}f_{xx}(t, X(t))\,\varDelta^2(t)\,\mathrm{d}t \end{aligned}$$

4.3.3 Examples

1. (Example) Generalized geometric Brownian motion: Define the Itô process

$$X(t) = \int_0^t \sigma(s)\,\mathrm{d}W(s) + \int_0^t \left[\alpha(s) - \frac{1}{2}\sigma^2(s)\right]\mathrm{d}s$$

Then

$$\mathrm{d}X(t) = \sigma(t)\,\mathrm{d}W(t) + \left[\alpha(t) - \frac{1}{2}\sigma^2(t)\right]\mathrm{d}t$$

and

$$\mathrm{d}X(t)\cdot\mathrm{d}X(t) = \sigma^2(t)\,\mathrm{d}t$$

Consider an asset price process given by

$$S(t) = S(0)\,e^{X(t)} = f(X(t))$$

then

$$\begin{aligned} \mathrm{d}S(t) =& \mathrm{d}f(X(t)) \\ =& f'(X(t))\,\mathrm{d}X(t) + \frac{1}{2}f''(X(t))\,\mathrm{d}X(t)\cdot\mathrm{d}X(t) \\ =& S(0)\,e^{X(t)}\left[\sigma(t)\,\mathrm{d}W(t) + \left[\alpha(t) - \frac{1}{2}\sigma^2(t)\right]\mathrm{d}t\right] + \frac{1}{2}S(0)\,e^{X(t)}\sigma^2(t)\,\mathrm{d}t \\ =& S(0)\,e^{X(t)}\sigma(t)\,\mathrm{d}W(t) + S(0)\,e^{X(t)}\alpha(t)\,\mathrm{d}t \\ =& S(t)\,\sigma(t)\,\mathrm{d}W(t) + S(t)\,\alpha(t)\,\mathrm{d}t \end{aligned}$$

This example includes all possible models of an asset price process that is always positive, has no jumps, and is driven by a single Brownian motion. If α and σ are constant, we have the usual geometric Brownian motion model, and the distribution of $S(t)$ is log-normal. Moreover, if $\alpha = 0$, we first have

$$\begin{aligned} \mathrm{d}X(t) &= \sigma(t)\,\mathrm{d}W(t) - \frac{1}{2}\sigma^2(t)\,\mathrm{d}t \\ \mathrm{d}S(t) &= \sigma(t)\,S(t)\,\mathrm{d}W(t) \end{aligned}$$

Integration of both sides yields

$$S(t) = S(0) + \int_0^t \sigma(s) S(s) \,\mathrm{d}W(s)$$

The right-hand side is the nonrandom constant plus a Itô integral, which is a martingale, and hence

$$S(t) = S(0) e^{X(t)} = \exp\left\{\int_0^t \sigma(s) \,\mathrm{d}W(s) - \int_0^t \frac{1}{2}\sigma^2(s) \,\mathrm{d}s\right\}$$

is a martingale.

2. (Theorem) Let $W(s), s \geqslant 0$ be a Brownian motion, and let $\Delta(s)$ be a nonrandom function of time. Define $I(t) = \int_0^t \Delta(s) \,\mathrm{d}W(s)$. For each $t \geqslant 0$, the random variable $I(t)$ is normally distributed with expected value zero and variance $\int_0^t \Delta^2(s) \,\mathrm{d}s$.

Proof. Since $I(t)$ is a martingale and $I(0) = 0$, we must have $\mathbb{E}I(t) = I(0) = 0$, and Itô's isometry implies that

$$\mathrm{Var}[I(t)] = \mathbb{E}I^2(t) = \int_0^t \Delta^2(s) \,\mathrm{d}s$$

Now we need to show that $I(t)$ is normally distributed. In other words, we need to show that $I(t)$ has the moment-generating function of normal random variable with mean zero and variance $\int_0^t \Delta^2(s) \,\mathrm{d}s$, which is

$$\mathbb{E}e^{uI(t)} = e^{\frac{1}{2}u^2 \int_0^t \Delta^2(s)\mathrm{d}s}, u \in \mathbb{R}$$

which may be rewritten as (we need to show)

$$\mathbb{E}e^{uI(t) - \frac{1}{2}u^2 \int_0^t \Delta^2(s)\mathrm{d}s} = \mathbb{E}\exp\left\{uI(t) - \frac{1}{2}\int_0^t [u\Delta(s)]^2 \,\mathrm{d}s\right\} = 1$$

Note that the process $\exp\left\{uI(t) - \frac{1}{2}\int_0^t [u\Delta(s)]^2 \,\mathrm{d}s\right\}$ is a martingale, since it is a generalized geometric Brownian motion with mean rate of return $\alpha = 0$, $\sigma(s) = u\Delta(s)$. And thus, this process takes value 1 at $t = 0$, and hence its expectation is always 1, which is exactly what we need to show, □

Remark: $\mathbb{E}\exp\left\{uI(t) - \frac{1}{2}\int_0^t [u\Delta(s)]^2 \,\mathrm{d}s\right\} = 1$ always holds, regardless of whether $\Delta(s)$ is random. However, when $\Delta(s)$ is random, there is no reason that the distribution of $\int_0^t \Delta(s) \,\mathrm{d}W(s)$ should be normal.

3. (Example) Vasicek interest rate model: Let $W(t), t \geqslant 0$, be a Brownian motion. The Vasicek model of the interest rate process $R(t)$ is

$$\mathrm{d}R(t) = (\alpha - \beta R(t)) \,\mathrm{d}t + \sigma \mathrm{d}W(t)$$

where $\alpha, \beta, \text{and} \sigma$ are positive constants. The solution to the stochastic differential equation can be determined in closed form and is

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

Verify. Let $X(t)=\int_0^t e^{\beta s}\mathrm{d}W(s)$, and

$$f(t,x)=e^{-\beta t}R(0)+\frac{\alpha}{\beta}\left(1-e^{-\beta t}\right)+\sigma e^{-\beta t}x$$

so

$$\begin{aligned}
f_t&=-\beta e^{-\beta t}R(0)+\alpha e^{-\beta t}-\sigma\beta e^{-\beta t}x=\alpha-\beta f(t,x)\\
f_x&=\sigma e^{-\beta t}\\
f_{xx}&=0
\end{aligned}$$

Then we compute the differential of $f(t,X(t))$, which is

$$\begin{aligned}
\mathrm{d}f(t,X(t))=&f_t(t,X(t))\,\mathrm{d}t+f_x(t,X(t))\,\mathrm{d}X(t)+\frac{1}{2}f_{xx}(t,X(t))\,\mathrm{d}X(t)\cdot\mathrm{d}X(t)\\
=&\left[\alpha-\beta f(t,X(t))\right]\mathrm{d}t+\sigma e^{-\beta t}\mathrm{d}X(t)\\
=&\left[\alpha-\beta f(t,X(t))\right]\mathrm{d}t+\sigma\mathrm{d}W(t)
\end{aligned}$$

This shows that $f(t,X(t))$ satisfies the stochastic differential equation in Vasicek model. Thus,

$$f(t,X(t))=R(t),t\geqslant 0$$

□

4. (Remark) Moreover,

$$X(t)=\int_0^t e^{\beta s}\mathrm{d}W(s)\sim N\left(0,\int_0^t e^{2\beta s}\mathrm{d}s\right)$$

while,

$$\int_0^t e^{2\beta s}\mathrm{d}t=\frac{1}{2\beta}\left(e^{2\beta t}-1\right)$$

Therefore, $R(t)$ is normally distributed,

$$\begin{aligned}
R(t)=&e^{-\beta t}R(0)+\frac{\alpha}{\beta}\left(1-e^{-\beta t}\right)+\sigma e^{-\beta t}\int_0^t e^{\beta s}\mathrm{d}W(s)\\
\sim&N\left(e^{-\beta t}R(0)+\frac{\alpha}{\beta}\left(1-e^{-\beta t}\right),\frac{\sigma^2}{2\beta}\left(1-e^{-2\beta t}\right)\right)
\end{aligned}$$

In particular, whatever the parameters are chosen, there is positive probability that $R(t)$ is negative, which is a undesirable property for an interest rate model.

5. (Remark) Mean-reverting property of Vasicek model:

   (a) When $R(t)=\frac{\alpha}{\beta}$, the drift term is zero, and $\mathbb{E}R(t)=\frac{\alpha}{\beta}$ for all $t\geqslant 0$.

   (b) If $R(0)\neq\frac{\alpha}{\beta}$, then $\lim_{t\to\infty}\mathbb{E}R(t)=\frac{\alpha}{\beta}$.

6. (Remark) Cox-Ingersoll-Ross (CIR) interest rate model:

$$\mathrm{d}R(t)=(\alpha-\beta R(t))\,\mathrm{d}t+\sigma\sqrt{R(t)}\mathrm{d}W(t)$$

where α,β and σ are positive constants.

CIR model does not have a closed form solution, but the interest rate in CIR model does not become negative. Because when $R(t)$ reaches zero, the term multiplying $\mathrm{d}W(t)$ vanishes and the positive drift term $\alpha\mathrm{d}t$ drives the interest rate back into positive territory. Let

$$f(t,x) = e^{\beta t}x$$

then

$$\begin{aligned}\mathrm{d}f(t,R(t)) =& \mathrm{d}\left(e^{\beta t}R(t)\right)\\ =& f_t(t,R(t))\,\mathrm{d}t + f_x(t,R(t))\,\mathrm{d}R(t) + \frac{1}{2}f_{xx}(t,R(t))\,\mathrm{d}R(t)\cdot\mathrm{d}R(t)\\ =& \beta e^{\beta t}R(t)\,\mathrm{d}t + e^{\beta t}\mathrm{d}R(t)\\ =& \beta e^{\beta t}R(t)\,\mathrm{d}t + e^{\beta t}(\alpha-\beta R(t))\,\mathrm{d}t + \sigma e^{\beta t}\sqrt{R(t)}\mathrm{d}W(t)\\ =& \alpha e^{\beta t}\mathrm{d}t + \sigma e^{\beta t}\sqrt{R(t)}\mathrm{d}W(t)\end{aligned}$$

Integration of both sides,

$$e^{\beta t}R(t) - R(0) = \frac{\alpha}{\beta}\left(e^{\beta t}-1\right) + \sigma\int_0^t e^{\beta s}\sqrt{R(s)}\mathrm{d}W(s)$$

Thus,

$$\mathbb{E}R(t) = \frac{R(0)+\frac{\alpha}{\beta}\left(e^{\beta t}-1\right)}{e^{\beta t}} = e^{-\beta t}R(0) + \frac{\alpha}{\beta}\left(1-e^{-\beta t}\right)$$

which is the same expectation as in Vasicek model. To compute the variance of $R(t)$, we set

$$X(t) = e^{\beta t}R(t)$$

we have

$$\begin{aligned}\mathrm{d}X(t) &= \alpha e^{\beta t}\mathrm{d}t + \sigma e^{\beta t}\sqrt{R(t)}\mathrm{d}W(t)\\ &= \alpha e^{\beta t}\mathrm{d}t + \sigma e^{\frac{\beta t}{2}}\sqrt{X(t)}\mathrm{d}W(t)\\ \mathbb{E}X(t) &= R(0) + \frac{\alpha}{\beta}\left(e^{\beta t}-1\right)\end{aligned}$$

we compute the differential of $\mathrm{d}\left(X^2(t)\right)$ by setting $f(x)=x^2$,

$$\begin{aligned}\mathrm{d}\left(X^2(t)\right) =& 2X(t)\,\mathrm{d}X(t) + \mathrm{d}X(t)\,\mathrm{d}X(t)\\ =& 2\alpha e^{\beta t}X(t)\,\mathrm{d}t + 2\sigma e^{\frac{\beta t}{2}}[X(t)]^{\frac{3}{2}}\,\mathrm{d}W(t) + \sigma^2 e^{\beta t}X(t)\,\mathrm{d}t\end{aligned}$$

Integration of both sides,

$$X^2(t) - X^2(0) = \left(2\alpha+\sigma^2\right)\int_0^t e^{\beta s}X(s)\,\mathrm{d}s + \int_0^t 2\sigma e^{\frac{\beta s}{2}}[X(s)]^{\frac{3}{2}}\,\mathrm{d}W(s)$$

Thus,

$$\begin{aligned}\mathbb{E}X^2(t) =& X^2(0) + \left(2\alpha+\sigma^2\right)\int_0^t e^{\beta s}\mathbb{E}X(s)\,\mathrm{d}s\\ =& R^2(0) + \left(2\alpha+\sigma^2\right)\int_0^t e^{\beta s}\left[R(0)+\frac{\alpha}{\beta}\left(e^{\beta s}-1\right)\right]\mathrm{d}s\\ =& R^2(0) + \frac{2\alpha+\sigma^2}{\beta}\left(R(0)-\frac{\alpha}{\beta}\right)\left(e^{\beta t}-1\right) + \frac{2\alpha+\sigma^2}{\beta}\cdot\frac{\alpha}{\beta}\left(e^{2\beta t}-1\right)\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{E}R^2(t) =&e^{-2\beta t}\cdot\mathbb{E}X^2(t)\\ =&e^{-2\beta t}R^2(0)+\frac{2\alpha+\sigma^2}{\beta}\left(R(0)-\frac{\alpha}{\beta}\right)\left(e^{-\beta t}-e^{-2\beta t}\right)+\frac{2\alpha+\sigma^2}{\beta}\cdot\frac{\alpha}{\beta}\left(1-e^{-2\beta t}\right)\end{aligned}$$

Finally,

$$\begin{aligned}\mathrm{Var}\left[R(t)\right] =&\mathbb{E}R^2(t)-\left[\mathbb{E}R(t)\right]^2\\ =&e^{-2\beta t}R^2(0)+\frac{2\alpha+\sigma^2}{\beta}\left(R(0)-\frac{\alpha}{\beta}\right)\left(e^{-\beta t}-e^{-2\beta t}\right)\\ &+\frac{2\alpha+\sigma^2}{\beta}\cdot\frac{\alpha}{\beta}\left(1-e^{-2\beta t}\right)-e^{-\beta t}R(0)-\frac{\alpha}{\beta}\left(1-e^{-\beta t}\right)\\ =&\frac{\sigma^2}{\beta}R(0)\left(e^{-\beta t}-e^{-2\beta t}\right)+\frac{\alpha\sigma^2}{2\beta^2}\left(1-2e^{-\beta t}+e^{-2\beta t}\right)\end{aligned}$$

In particular,

$$\lim_{t\to\infty}\mathrm{Var}\left[R(t)\right]=\frac{\alpha\sigma^2}{2\beta^2}$$

4.4 Black-Scholes-Merton Equation and Application

4.4.1 Black-Scholes-Merton Equation

1. Geometric Brownian motion:

$$\mathrm{d}S(t)=\alpha S(t)\,\mathrm{d}t+\sigma S(t)\,\mathrm{d}W(t)$$

2. Portfolio: the portfolio value at each time t due to two factors, the capital gain $\Delta(t)\,\mathrm{d}S(t)$ on stock position and the interest earnings $X(t)-\Delta(t)\,S(t)\,\mathrm{d}t$ on cash position.

$$\begin{aligned}\mathrm{d}X(t) =&\Delta(t)\,\mathrm{d}S(t)+r\left(X(t)-\Delta(t)\,S(t)\right)\mathrm{d}t\\ =&\Delta(t)\left[\alpha S(t)\,\mathrm{d}t+\sigma S(t)\,\mathrm{d}W(t)\right]+r\left(X(t)-\Delta(t)\,S(t)\right)\mathrm{d}t\\ =&rX(t)\,\mathrm{d}t+\Delta(t)(\alpha-r)\,S(t)\,\mathrm{d}t+\Delta(t)\,\sigma S(t)\,\mathrm{d}W(t)\end{aligned}$$

The tree terms have different meaning:

   (a) $rX(t)\,\mathrm{d}t$: an average underlying rate of return r on portfolio,

   (b) $\Delta(t)(\alpha-r)\,S(t)\,\mathrm{d}t$: a risk premium $\alpha-r$ for investing in the stock,

   (c) $\Delta(t)\,\sigma S(t)\,\mathrm{d}W(t)$: a volatility term proportional to the size of the stock investment.

3. The differential of the discounted stock price is

$$\begin{aligned}\mathrm{d}\left(e^{-rt}S(t)\right)&=\mathrm{d}\left(f\left(t,S(t)\right)\right)\\ &=-re^{-rt}S(t)\,\mathrm{d}t+e^{-rt}\mathrm{d}S(t)\\ &=(\alpha-r)\,e^{-rt}S(t)\,\mathrm{d}t+\sigma e^{-rt}S(t)\,\mathrm{d}W(t)\end{aligned}$$

4. The differential of the discounted portfolio value is

$$\begin{aligned}\mathrm{d}\left(e^{-rt}X(t)\right) =&\mathrm{d}f(t,X(t))\\ =&-re^{-rt}X(t)\,\mathrm{d}t+e^{-rt}\mathrm{d}X(t)\\ =&\Delta(t)(\alpha-r)e^{-rt}S(t)\,\mathrm{d}t+\Delta(t)\sigma e^{-rt}S(t)\,\mathrm{d}W(t)\\ =&\Delta(t)\,\mathrm{d}\left(e^{-rt}S(t)\right)\end{aligned}$$

5. Call option price $c(t,S(t))$:

$$\begin{aligned}\mathrm{d}c(t,S(t)) =&c_t(t,S(t))\,\mathrm{d}t+c_x(t,S(t))\,\mathrm{d}S(t)+\frac{1}{2}c_{xx}(t,S(t))\,\sigma^2S^2(t)\,\mathrm{d}t\\ =&c_t(t,S(t))\,\mathrm{d}t+c_x(t,S(t))\left[\alpha S(t)\,\mathrm{d}t+\sigma S(t)\,\mathrm{d}W(t)\right]\\ &+\frac{1}{2}c_{xx}(t,S(t))\,\sigma^2S^2(t)\,\mathrm{d}t\\ =&\left[c_t(t,S(t))+c_x(t,S(t))\,\alpha S(t)+\frac{1}{2}c_{xx}(t,S(t))\,\sigma^2S^2(t)\right]\mathrm{d}t\\ &+c_x(t,S(t))\,\sigma S(t)\,\mathrm{d}W(t)\end{aligned}$$

6. The differential of discounted option price $e^{-rt}c(t,S(t))$:

$$\begin{aligned}&\mathrm{d}\left[e^{-rt}c(t,S(t))\right]\\ =&\mathrm{d}f(t,c(t,S(t)))\\ =&f_t(t,c(t,S(t)))\,\mathrm{d}t+f_x(t,c(t,S(t)))\,\mathrm{d}c(t,S(t))\\ &+\frac{1}{2}f_{xx}(t,c(t,S(t)))\,\mathrm{d}c(t,S(t))\cdot\mathrm{d}c(t,S(t))\\ =&-re^{-rt}c(t,S(t))\,\mathrm{d}t+e^{-rt}\mathrm{d}c(t,S(t))\\ =&e^{-rt}\left[c_t(t,S(t))+c_x(t,S(t))\,\alpha S(t)+\frac{1}{2}c_{xx}(t,S(t))\,\sigma^2S^2(t)-rc(t,S(t))\right]\mathrm{d}t\\ &+e^{-rt}c_x(t,S(t))\,\sigma S(t)\,\mathrm{d}W(t)\end{aligned}$$

7. Portfolio is equivalent with a call option:

$$\mathrm{d}\left(e^{-rt}X(t)\right)=\mathrm{d}\left(e^{-rt}c(t,S(t))\right),\forall t\in[0,T)$$

Thus,

$$\begin{aligned}&\Delta(t)(\alpha-r)S(t)\,\mathrm{d}t+\Delta(t)\sigma S(t)\,\mathrm{d}W(t)\\ =&\left[c_t(t,S(t))+c_x(t,S(t))\,\alpha S(t)+\frac{1}{2}c_{xx}(t,S(t))\,\sigma^2S^2(t)-rc(t,S(t))\right]\mathrm{d}t\\ &+c_x(t,S(t))\,\sigma S(t)\,\mathrm{d}W(t)\end{aligned}$$

which implies

$$\begin{cases}\Delta(t)=c_x(t,S(t))\\ c_t(t,S(t))+c_x(t,S(t))\,\alpha S(t)+\frac{1}{2}c_{xx}(t,S(t))\,\sigma^2S^2(t)-rc(t,S(t))=\Delta(t)(\alpha-r)S(t)\end{cases}$$

Furthermore, we have

$$c_t(t,S(t))+c_x(t,S(t))\,tS(t)+\frac{1}{2}c_{xx}(t,S(t))\,\sigma^2S^2(t)=rc(t,S(t))$$

In conclusion, we should seek a continuous function $c(t,x)$ that is a solution to BSM partial differential equation

$$c_t(t,x) + rxc_x(t,x) + \frac{1}{2}\sigma^2 x^2 c_{xx}(t,x), t \in [0,T), x \geqslant 0$$

and that satisfies the ***terminal condition***

$$c(T,x) = (x-K)^+$$

8. Solution: when $x=0$, the Black-Scholes-Merton PDE becomes

$$c_t(t,0) = rc(t,0)$$

which is an ODE with solution

$$c(t,0) = e^{rt}c(0,0)$$

Substituting $T=t$ in to this equation and use the fact that $c(T,0) = (0-K)^+ = 0$, we see that $c(0,0)=0$ and hence

$$c(t,0) = 0, \forall t \in [0,T]$$

This is the ***boundary condition*** at $x=0$.

As $x \to \infty$, the function $c(t,x)$ grows without bound. In such a case, we give the boundary condition at $x=\infty$ by specifying the rate of growth. One way to specify a ***boundary condition*** at $x=\infty$ for the European call is

$$\lim_{x\to\infty}\left[c(t,x) - \left(x - e^{-r(T-t)}\right)K\right] = 0, \forall t \in [0,T]$$

With a terminal condition and two boundary conditions, we can get the solution of the PDE.

$$\begin{cases} c(t,x) = xN(d_+(T-t,x)) - Ke^{-r(T-t)}N(d_-(T-t,x)) \\ d_\pm(\tau,x) = \frac{1}{\sigma\sqrt{\tau}}\left[\log\frac{x}{K} + \left(r \pm \frac{\sigma^2}{2}\right)\tau\right] \end{cases}$$

4.4.2 Greeks and Hedging

1. (Definition) For a European call option, we define

   (a) delta:

   $$\varDelta = \frac{\partial c}{\partial x} = c_x(t,x) = N(d_+)$$

   (b) theta:

   $$\varTheta = \frac{\partial c}{\partial t} = c_x(t,x) = -rKe^{-r(T-t)}N(d_-) - \frac{\sigma x}{2\sqrt{T-t}}N'(d_+)$$

   (c) gamma:

   $$\varGamma = \frac{\partial^2 c}{\partial x^2} = c_{xx}(t,x) = N'(d_+)\frac{\partial d_+}{\partial x} = \frac{N'(d_+)}{\sigma x\sqrt{T-t}}$$

   (d) rho:

   $$\rho = \frac{\partial c}{\partial r} = -Ke^{-r(T-t)}(T-t)[1-N(d_-)]$$

   (e) vega:

   $$V = \frac{\partial c}{\partial \sigma} = x\sqrt{T-t}N(d_+)$$

2. (Remark) The hedging portfolio value is $c = xN(d_+) - Ke^{-r(T-t)}N(d_-)$ at time t, and since $xc_x(t,x)$ of this value is invested in stock, the amount invested in the money market must be

$$c(t,x) - xc_x(t,x) = -Ke^{-r(T-t)}N(d_-)$$

which is a negative number. Therefore, to hedge a short position in a call option, one must borrow money. To hedge a long position in a call option, one does the opposite. In other words, to hedge a long call position one should hold $-c_x$ shares of stock (short position) and invest $Ke^{-r(T-t)}N(d_-)$ in the money market account.

3. (Remark) Note that delta and gamma are positive, and thus, for a fixed t, $c(t,x)$ is increasing and convex in the variable x.

4. (Remark and Definition) Suppose at time t the stock price is x_1 and we wish to take a long position in the option and hedge it. We do this by purchasing 1 unit of the option for $c(t,x_1)$ and shorting $c_x(t,x_1)$ shares of stock, which generates income $x_1c_x(t,x_1)$,and investing

$$M = x_1c_x(t,x_1) - c(t,x_1)$$

in the money market account. The initial portfolio value

$$c(t,x_1) - x_1c_x(t,x_1) + M$$

is zero at moment t, and it has 3 parts:

(a) Long potion: $c(t,x_1)$;

(b) Short stock: $-x_1c_x(t,x_1)$;

(c) Long money market account: M.

We wish to consider the sensitivity to stock price changes of the portfolio.

If the stock price were to instantaneously fall to x_0, then our total portfolio value would be

$$c(t,x_0) - x_0c_x(t,x_0) + M = c(t,x_0) - c(t,x_1) - (x_0 - x_1)c_x(t,x_0)$$

which is the difference at x_0 between

(a) The curve $y = c(t,x)$;

(b) The straight line $y = c(t,x_1) + c_x(t,x_0)(x - x_1)$.

And this difference is positive, our portfolio benefits from an instantaneous drop in the stock price. Otherwise, if the stock price were to instantaneously fall to x_2, our portfolio would also benefit from it. The portfolio we have set up is said to be

(a) *delta-neutral*, it refers to the fact the the straight line is tangent to the curve $y = c(t,x)$;

(b) *long gamma*, it refers to the convexity of $c(t,x)$ for a given t. If there is a instantaneous rise or fall in x, the value of the portfolio increases. A long gamma portfolio is profitable in times of high stock volatility.

If the straight line were steeper than the option price curve $y = c(t,x)$ at the start point x_1, then we would be *short delta*.

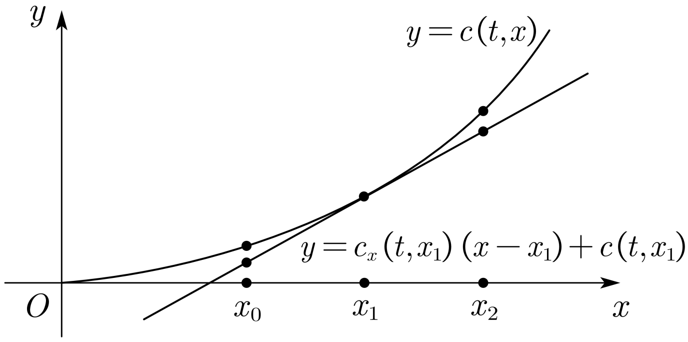

Fig. 4.1. Delta-neutral position with t fixed

(a) An upward move in the stock price would hurt the portfolio;

(b) A downward move would increase the portfolio.

If t changes, the drawback is that theta $c_t(t,x)$ is negative. As we move forward in time, the curve $y = c_t(t,x)$ is shifting downward. Therefore, we need to continuously re-balance our portfolio, i.e., with dynamic adjustment.

5. (Remark) The mean rate of stock return α does not appear in the BSM equation. From the point of view of no-arbitrage pricing, it is irrelevant how likely the stock is to go up and down because a delta-neutral position is a hedge against both up and down.

4.4.3 Put-Call Parity

1. (Definition) A forward contract with delivery price K obligates its holder to buy one share of stock at expiration time T in exchange for payment K.

2. (Remark) The value of the forward contract at earlier times $t \in [0,T]$ is

$$f(t,x) = x - Ke^{-r(T-t)}$$

3. (Definition) The forward price of a stock at time t is defined to be the value of K that causes the forward contract at time t to have value zero, i.e., it is the value of K that satisfies the equation

$$S(t) - Ke^{-r(T-t)} = 0$$

. Hence, the forward price at time t is $S(t)e^{r(T-t)}$.

4. (**Put-Call Parity**) Consider a forward contract that one can buy 1 share of stock in a price $K = e^{rt}S(0)$ at time t. The value of this forward contract is zero at time $t = 0$, but as soon as time begins to move forward, the value changes. Its value at time t is

$$f(t,S(t)) = S(t) - e^{rt}S(0)$$

At time T, we have

$$\begin{aligned} f(T,S(T)) =& S(T) - e^{rT}S(0) \\ =& (S_T - K)^+ - (K - S_T)^+ \\ =& c(T,S(T)) - p(T - S(T)) \end{aligned}$$

where c and p are European call and put option prices respectively.

Now we hold a portfolio that is long one call and short one put, because of non-arbitrage, we must have

$$f(t,x) = x - e^{-r(T-t)}K = c(t,x) - p(t,x), x \geqslant 0, t \in [0,T]$$

Otherwise, one could at some time t either sell or buy the portfolio that is long the forward, short the call and long the put, realizing an instant profit without any liability upon expiration of the contracts.

5. (Remark) From Black-Scholes-Merton European call formula, we can derive the put formula

$$\begin{aligned} p(t,x) =& c(t,x) - x + e^{r(T-t)}K \\ =& xN(d_+) - x - Ke^{-r(T-t)}N(d_-) + e^{r(T-t)}K \\ =& x[N(d_+) - 1] - Ke^{-r(T-t)}[N(d_-) - 1] \\ =& Ke^{-r(T-t)}N(-d_-) - xN(-d_+) \end{aligned}$$

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

4.5 Multivariable Stochastic Calculus

1. (Definition) A d-dimensional Brownian motion is a process $W(t) = (W_1(t), W_2(t), \cdots, W_d(t))$ with following properties.
   (a) Each $W_i(t)$ is a one-dimensional Brownian motion.
   (b) If $i \neq j$, then the process $W_i(t)$ and $W_j(t)$ are independent.

2. (Theorem) We also have a filtration $\mathcal{F}(t)$ associated with d-dimensional Brownian motion, such that the following holds.
   (a) (**Information accumulates**) For $0 \leqslant s < t$, $\mathcal{F}(s) \subset \mathcal{F}(t)$.
   (b) (**Adabtivity**) For each $t \geqslant 0$, the random vector $W(t)$ is $\mathcal{F}(t)$-measurable.
   (c) (**Independence of future increments**) For $0 \leqslant t < u$, the vector of increments $W(u) - W(t)$ is independent of $\mathcal{F}(t)$.

3. (Remark) Quadratic variation formula:

$$[W_i, W_i](t) = t \Longrightarrow \mathrm{d}W_i(t)\,\mathrm{d}W_i(t) = \mathrm{d}t$$

$$[W_j, W_j](t) = 0 \Longrightarrow \mathrm{d}W_i(t)\,\mathrm{d}W_j(t) = 0$$

4. (Remark) Let $X(t)$ and $Y(t)$ be Itô processes, that is

$$\begin{aligned} X(t) =& X(0) + \int_0^t \Theta_1(u)\,\mathrm{d}u + \int_0^t \sigma_{11}(u)\,\mathrm{d}W_1(u) + \int_0^t \sigma_{12}(u)\,\mathrm{d}u \\ Y(t) =& Y(0) + \int_0^t \Theta_2(u)\,\mathrm{d}u + \int_0^t \sigma_{21}(u)\,\mathrm{d}W_1(u) + \int_0^t \sigma_{22}(u)\,\mathrm{d}u \end{aligned}$$

or

$$\begin{aligned}
\mathrm{d}X(t) &= \Theta_1(t)\,\mathrm{d}t + \sigma_{11}(t)\,\mathrm{d}W_1(t) + \sigma_{12}(t)\,\mathrm{d}W_2(t) \\
\mathrm{d}Y(t) &= \Theta_2(t)\,\mathrm{d}t + \sigma_{21}(t)\,\mathrm{d}W_1(t) + \sigma_{22}(t)\,\mathrm{d}W_2(t)
\end{aligned}$$

By the quadratic variation formula, we have

$$\begin{aligned}
\mathrm{d}X(t)\,\mathrm{d}X(t) &= \left(\sigma_{11}^2(t) + \sigma_{12}^2(t)\right)\mathrm{d}t \\
\mathrm{d}Y(t)\,\mathrm{d}Y(t) &= \left(\sigma_{21}^2(t) + \sigma_{22}^2(t)\right)\mathrm{d}t \\
\mathrm{d}X(t)\,\mathrm{d}Y(t) &= \left(\sigma_{11}(t)\,\sigma_{21}(t) + \sigma_{12}(t)\,\sigma_{22}(t)\right)\mathrm{d}t
\end{aligned}$$

5. (Two-dimensional Itô-Doeblin Formula) Let be a function whose partial derivatives f_t, f_x, f_y, f_{xx}, f_{xy}, f_{yy}, f_{yx} are defined and are continuous. Let $X(t)$ and $Y(t)$ be Itô process.

$$\begin{aligned}
&\mathrm{d}f(t, X(t), Y(t)) \\
=&f_t(t, X(t), Y(t))\,\mathrm{d}t + f_x(t, X(t), Y(t))\,\mathrm{d}X(t) + f_y(t, X(t), Y(t))\,\mathrm{d}Y(t) \\
&+ \frac{1}{2}f_{xx}(t, X(t), Y(t))\,\mathrm{d}X(t)\,\mathrm{d}X(t) + f_{xy}(t, X(t), Y(t))\,\mathrm{d}X(t)\,\mathrm{d}Y(t) \\
&+ \frac{1}{2}f_{yy}(t, X(t), Y(t))\,\mathrm{d}Y(t)\,\mathrm{d}Y(t)
\end{aligned}$$

6. (Itô's product rule)

$$\mathrm{d}[X(t)\,Y(t)] = X(t)\,\mathrm{d}Y(t) + Y(t)\,\mathrm{d}X(t) + \mathrm{d}X(t)\,\mathrm{d}Y(t)$$

Proof. Let $(t, x, y) = xy$, we have $f_t = 0$, $f_x = y$, $f_y = x$, $f_{xx} = 0$, $f_{yy} = 0$, $f_{xy} = f_{yx} = 1$, then use the Two-dimensional Itô-Doeblin Formula. □

7. (Lévy Theorem, one dimension) Let $M(t), t \geqslant 0$, be a martingale relative to a filtration $\mathcal{F}(t)$, $t \geqslant 0$. Assume

   (a) $M(0) = 0$;

   (b) $M(t)$ has continuous paths;

   (c) $[M, M(t)] = t$ for all $t \geqslant 0$.

   Then $M(t)$ is a Brownian motion.

Idea of the proof. We need to proof that $M(t)$ has normally distributed increments. By Itô-Doeblin Formula, for any function $f(t, x)$ whose derivatives exist and are continuous, we have

$$\mathrm{d}f(t, M(t)) = f_t(t, M(t))\,\mathrm{d}t + f_x(t, M(t))\,\mathrm{d}M(t) + \frac{1}{2}f_{xx}(t, M(t))\,\mathrm{d}t$$

In integrated form, we have

$$f(t, M(t)) = f(0, M(0)) + \int_0^t \left[f_t(s, M(s)) + \frac{1}{2}f_{xx}(s, M(s))\right]\mathrm{d}s + \int_0^t f_x(s, M(s))\,\mathrm{d}M(s)$$

Note that the stochastic process $I(t) = \int_0^t f_x(s, M(s))\,\mathrm{d}M(s)$ is a martingale because $M(t)$ is a martingale (A special case will be seen at Selected Exercise in this chapter; the general case is upon passage to the limit). Therefore, we have

$$\mathbb{E}I(t) = \mathbb{E}\int_0^t f_x(s, M(s))\,\mathrm{d}M(s) = \mathbb{E}I(0) = 0$$

Taking expectations in the integrated form, we have

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

For a given constant u, let

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

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

$$f_t + \frac{1}{2}f_{xx} = 0$$

Hence,

$$\mathbb{E}\exp\left\{uM(t) - \frac{1}{2}u^2 t\right\} = \mathbb{E}e^{uM(0)} = 1$$

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

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

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

8. (Lévy Theorem, two dimensions) Let $M_1(t), M_2(t), t \geqslant 0$, be a martingale relative to a filtration $\mathcal{F}(t)$, $t \geqslant 0$. Assume that for $i = 1, 2$, we have

   (a) $M_i(0) = 0$;

   (b) $M_i(t)$ has continuous paths;

   (c) $[M_i, M_i](t) = t$ for all $t \geqslant 0$;

   (d) $[M_1, M_2](t) = 0$ for all $t \geqslant 0$.

   Then $M_1(t)$ and $M_2(t)$ are independent Brownian Motion.

   Idea of the proof. From one-dimensional Lévy Theorem, we have M_1 and M_2 are Brownian motions. We only need to show they are independence. Let $f(t, x, y)$ be a function whose derivatives are defined and continuous. By Itô-Doeblin Formula, we have

$$\begin{aligned}
\mathrm{d}f(t, M_1, M_2) =& f_t\mathrm{d}t + f_x\mathrm{d}M_1 + f_y\mathrm{d}M_2 + \frac{1}{2}f_{xx}\mathrm{d}t + f_{xy}\mathrm{d}M_1\mathrm{d}M_2 + \frac{1}{2}f_{yy}\mathrm{d}t \\
=& f_t\mathrm{d}t + f_x\mathrm{d}M_1 + f_y\mathrm{d}M_2 + \frac{1}{2}f_{xx}\mathrm{d}t + \frac{1}{2}f_{yy}\mathrm{d}t
\end{aligned}$$

   In integrated form, we have

$$\begin{aligned}
f(t, M_1(t), M_2(t)) =& f(0, M_1(0), M_2(0)) + \int_0^t \left(f_t + \frac{1}{2}f_{xx} + \frac{1}{2}f_{yy}\right)\mathrm{d}s \\
&+ \int_0^t f_x\mathrm{d}M_1(s) + \int_0^t f_y\mathrm{d}M_2(s)
\end{aligned}$$

Note that $\int_0^t f_x \mathrm{d}M_1(s)$, $\int_0^t f_y \mathrm{d}M_2(s)$ are martingales, and hence having expectation zero. Therefore,

$$\mathbb{E}f(t, M_1(t), M_2(t)) = \mathbb{E}f(0, M_1(0), M_2(0)) + \mathbb{E}\int_0^t \left(f_t + \frac{1}{2}f_{xx} + \frac{1}{2}f_{yy}\right) \mathrm{d}s$$

For fixed constants u_1 and u_2, let

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

Then

$$f_t + \frac{1}{2}f_{xx} + \frac{1}{2}f_{yy} = -\frac{1}{2}\left(u_1^2 + u_1^2\right) f + \frac{1}{2}u_1^2 f + \frac{1}{2}u_2^2 f = 0$$

Therefore, $\mathbb{E}f(t, M_1(t), M_2(t)) = \mathbb{E}f(0, M_1(0), M_2(0)) = 1$, and hence

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

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

9. (Example) (Correlated stock prices) Suppose

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

where W_1 and W_2 must be independent Brownian motions and $\sigma_1 > 0$, $\sigma_2 > 0$, $\rho \in [-1, 1]$ are constant. To analyze the second stock price process, we define

$$W_3(t) = \rho W_1(t) + \sqrt{1-\rho^2} W_2(t)$$

Then W_3 is a continuous martingale, and

$$\mathrm{d}W_3 \mathrm{d}W_3 = \rho^2 \mathrm{d}t + \left(1-\rho^2\right) \mathrm{d}t + 2\rho\sqrt{1-\rho^2}\mathrm{d}W_1 \mathrm{d}W_2 = \mathrm{d}t$$

In other words,

$$[W_3, W_3](t) = t$$

According to Lévy Theorem, W_3 is a Brownian motion. And we can rewrite the differential of S_2 as

$$\frac{\mathrm{d}S_1(t)}{S_2(t)} = \alpha_2 \mathrm{d}t + \sigma_2 \mathrm{d}W_3(t)$$

then S_2 is a geometric Brownian motion with mean rate of return α_2 and volatility σ_2. According to Itô's product rule,

$$\begin{aligned}\mathrm{d}[W_1 W_3] =& W_1 \mathrm{d}W_3 + W_3 \mathrm{d}W_1 + \mathrm{d}W_1 \mathrm{d}W_3 \\ =& W_1 \mathrm{d}W_3 + W_3 \mathrm{d}W_1 + \rho \mathrm{d}t\end{aligned}$$

Integrating, we get

$$W_1(t) W_3(t) = \int_0^t W_1(s) \mathrm{d}W_3(s) + \int_0^t W_3(s) \mathrm{d}W_1(s) + \rho t$$

Taking the expectation, we obtain

$$\mathbb{E}[W_1(t) W_3(t)] = \rho t$$

Therefore, the correlation between W_1 and W_3 is $\frac{\mathbb{E}[W_1(t)W_3(t)]}{\sqrt{t}\cdot\sqrt{t}} = \rho$.

4.6 Brownian Bridge

1. (Definition) A *Gaussian process* $X(t), t \geqslant 0$, is a stochastic process that the random variables $X(t_1), X(t_2), \cdots, X(t_n)$ are jointly normally distributed for any times $0 < t_1 < t_2 < \cdots < t_n$.

2. (Example) Brownian motion $W(t)$, is a Gaussian process, and

   (a) Mean function: $m(t) = \mathbb{E}W(t) = 0$.

   (b) Self-covariance function: $c(s,t) = \mathbb{E}[(W(t) - m(t))(W(s) - m(s))] = \min\{s,t\} = s \wedge t$.

3.

4.

5.

6.

7.

8.

9.

10.

11.

4.7 Selected Exercise

Exercise 4.1. *Proof.* □

第五章　风险中性定价

Risk-Neutral Measure

1.

2.

3.

4.

5.

6.

7.

8.

9.

10.

11.

5.1　Introduction

1.

2.

3.

4.

5.

6.

7.

8.

9.

10.

11.

5.2 Selected Exercise

Exercise 5.1. *Proof.* □