Stochastic Process

Stochastic Process

Chapter 1 Introduction

1.1. Definition

Definition : A collection of $\mathbf{R}^n$-valued random variables $X = \{X(t)\}_{t\geq 0}$ defined on a measurable space $(\Omega, \mathcal{F})$

From the above, we can see t is just a parameter in function X , the whole space Ω is fixed, it aligns with my understanding of the subject. But we can also write $X(t)(\omega)$ as $X(t, \omega)$, which implied t has the same status as ω . The measurable space becomes $([0, \infty) \times \Omega, \mathcal{F})$. The reason why we do this is because this expression is important for filtration.

If ω is fixed, $X(t, \omega)$ is called a sample path of process X

Then we compare two stochastic processes: X, Y . We say the two are the same if $\forall t \in [0, \infty), \omega \in \Omega$ we have $X(t, \omega) = Y(t, \omega)$. This requirement is really strong, but we can weaken it after introducing measure.

1.2. Filtration

Definition : A filtration on a measurable space $(\Omega, \mathcal{F})$ is a collection $\{\mathcal{F}_t\}_{t\geq 0}$ of σ algebras satisfying

$$\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F},\ \forall\ 0 \leq s < t < \infty$$
$$\mathcal{F}_\infty = \sigma(\cup_{t\geq 0}\mathcal{F}_t)$$

Based on stochastic process, the reason why we need filtration is that our knowledge of the system will definitely expand as time goes by, which means we need a series of σ algebras to describe the information we have at our disposal at each moment.

Then if every moment, our information at hand is enough for us to deduce the value of $X(t, \omega)$, i.e. $X(t) \in \mathcal{F}_t$, we have the following definition.

Definition : Let $\{\mathcal{F}_t\}_{t\geq 0}$ be a filtration on a measurable space $(\Omega, \mathcal{F})$. A stochastic process $X = \{X(t)\}$ on $(\Omega, \mathcal{F})$ is called adapted to $\{\mathcal{F}_t\}$, or $\{\mathcal{F}_t\}$-adapted, if $X(t) \in \mathcal{F}_t$ for each $t \geq 0$

Observe the definition, we can see that even if it only states $X(t) \in \mathcal{F}_t$, since $\mathcal{F}_s \subset \mathcal{F}_t$, naturally $X(s) \in \mathcal{F}_t$, meaning the σ algebra $\mathcal{F}_t$ at time t would make every r.v before time t $\mathcal{F}_t$-measurable

Just like the discussion of random variable, after we have considered an arbitrary stochastic process and an arbitrary filtration, we wonder if the stochastic process is fixed, can we find the minimum filtration in some sense makes X adapted?

Definition : Given a stochastic process X , for all $t \in [0, \infty)$, set

$$\mathcal{F}_t^X = \sigma\{X(s), 0 \leq s \leq t\}$$

that is, $\mathcal{F}_t^X$ is the smallest σ algebra with respect to which all $X(s), s \in [0, t]$, are measurable. Obviously ${\mathcal{F}_t^X}_{t\geq 0}$ is a filtration, which is called the natural filtration of X, or, the filtration generated by X . If we want find a $\{\mathcal{F}_t\}$ that X is adapted on, the restriction we need to satisfy is $\mathcal{F}_t^X \subset \mathcal{F}_t, \forall t \geq 0$

1.3. Stopping Times

Definition : Let $\{\mathcal{F}_t\}_{t\geq 0}$ be a filtration on a measurable space $(\Omega, \mathcal{F})$. A mapping $\tau : \Omega \to [0, \infty]$ is called an $\{\mathcal{F}_t\}$-stopping time, or a stopping time with respect to $\{\mathcal{F}_t\}$, if $[\tau \leq t] \in \mathcal{F}_t$ for each $t \geq 0$

A stochastic process probably can go on forever, but in reality most of the time this is not the case. What should we do about it mathematically? The designer of stochastic process (maybe Kolmogorov?) didn't change the definition we have learned so far, but put a switch or an indicator on the process.

Here is the deal. Each $\omega \in \Omega$ is an experiment result who has its own life span. Again we state that $\mathcal{F}_t$ is the information we have at time t about the result ω , meaning it incorporates all my observations at or before time t . If ω stops at $s < t$, $\mathcal{F}_t$ should be aware. To put it mathematically

$$\{all\ \omega\ stop\ at\ or\ before\ t\} \subset \mathcal{F}_t$$

So how do we find out the life span of each ω ? We should know that the life span of each possible result is in itself, we can use a funtion to extract it, which is τ (so τ is a r.v.). We use $[\tau \leq t]$ to locate those ω that stops before t . Information of these ω of course is contained in $\mathcal{F}_t$, so we have

$$[\tau \leq t] \in \mathcal{F}_t$$

If the above is satisfied for every $t \geq 0$, we call τ stopping time.

My question is : when we study a system, the nature of each ω is fixed, meaning the life span of it is fixed, so τ should be unique. But the definition means as long as $[\tau \leq t] \in \mathcal{F}_t$, any form of τ is acceptable.

1.4. Types

- Discrete Time Markov chain: state is discrete and time is discrete
- Continuous Time Markov chain: state is discrete and time is continuous
- Markov process: it satisfies Markov properties, state and time are continuous, the most typical one is Brownian motion

Chapter 2 Random Walk

2.1. Random Walk

There is a series of iid random variables Z_i and we consider $X_n = \sum_{i=1}^{n} Z_i$.

- If the possible value of Z_i is $\{0, \pm 1, \pm 2, \dots\}$, it is called random walk
- If the possible value of Z_i is $\{-1, 0, 1, 2, \dots\}$, it is called left skip free random walk (from Sheldon Ross)
- If the possible value of Z_i is $\{-1, 0, 1\}$ or $\{-1, 1\}$, it is called simple random walk

We will discuss simple random walk with probability

$$P(Z_i = 1) = p, \;\; P(Z_i = -1) = q$$

$E[X_n] = n(p-q), Var[X_n] = 4npq$. When $p = q = \frac{1}{2}$, it is a special case of random walk and is called symmetric random walk. We have

$$E[X_n] = 0, \;\; Var[X_n] = n$$

the variance of random variable $X(t)$ in the special case can be understood as how many steps to get to $X(t)$.

Then we introduce the definition of quadratic variation:

$$\langle X_n, X_n \rangle = \sum_{i=1}^{n} (X_i - X_{i-1})^2$$

intuitive understanding of it is the increment of each step up to step n , so it is still a function of time. When X is random walk

$$\langle X_n, X_n \rangle = \sum_{i=1}^{n} (X_i - X_{i-1})^2 = \sum_{i=1}^{n} (Z_i)^2 = n$$

At each step the index changes from i to $i+1$, we can limit the time interval so the model can be more suitable to real world cases. But the possible increment of each step is still +1 or -1 which is too large. So the model is not accurate enough.

Definition : for a fixed integer n , we define the scaled symmetric random walk by

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

for all $t \geq 0$ such that nt is an integer; for all other nonnegative t we define $W^{(n)}(t)$ by linear interpolation

It uses n to shrink the step of each Z_i and manipulate time intervaval. Let $0 \leq s \leq t$ be such that both ns and nt are integers, then

$$E[W^{(n)}(t) - W^{(n)}(s)] = 0 \ \Rightarrow \ E[W^{(n)}(t)] = 0$$
$$Var[W^{(n)}(t) - W^{(n)}(s)] = t - s \ \Rightarrow \ Var[W^{(n)}(t)] = t$$
$$\langle W^{(n)} - W^{(n)} \rangle (t) = 1$$

From Central Limit Theorem, when we fix t , 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, i.e., for every $t \geq 0$

$$\lim_{n \to \infty} W^{(n)}(t) = N(0, t)$$

And at the same time, when n goes to infinity, scaled symmetric random walk becomes browanian motion.

Chapter 3 Markov Chain

3.1. Definition

In LLN and CLT, we consider a series of iid random variables, which is a very strong assumption. Here we downplay the independence but not too much, and divide timeline into three parts: past, present and future. We say status of the future only depends on the present but not the past, any process satisfy the requirement is called Markov Chain. The rough definition would be $P(X_t|\mathcal{F}_s) = P(X_t|X_s), t \geq s$

But this understanding is still troublesome because transition probability function is related to time s, meaning if $s_1 \neq s_2$ and $t_1 - s_1 = t_2 - s_2$, we can not say $P(X_{t_1}|X_{s_1}) = P(X_{t_2}|X_{s_2})$ for sure. So we have to introduce an simplification: temporally homogeneous (Durrett) or time-homogeneous, it means the transiton probability function structure doesn't depend on time s any more, only on the status of time s which is $X_s = x_s$ and the time gap $t - s$, naturally we have $P(X_t|X_s) = P(X_{t-s}|X_0)$

How to use Markov process? We often use its stationary distribution and the decision theory based on it. One example is MCMC algorithm which uses stationary distribution as a target distribution that is difficult to sample from directly. We can classify Markov process into few types considering state space and time. The classic example of the fourth is Brownian motion

1. discrete time, discrete state
2. continuous time, discrete state
3. discrete time, continuous state
4. continuous time, continuous state

Definition (Markov Process) : A function $p : S \times \mathcal{S} \to \mathbb{R}$ is said to be a transition probability if:

$$For\ each\ x \in S, A \to p(x, A)\ is\ a\ probability\ measure\ on (S, \mathcal{S})$$
$$For\ each\ A \in \mathcal{S}, x \to p(x, A)\ is\ a\ measurable\ function$$

We say X_n is a Markov chain (w.r.t. $\mathcal{F}_n$) with transition probability p if
$P(X_{n+1} \in B | \mathcal{F}_n) = p(X_n, B)$

3.2. Chapman–Kolmogorov Equations

It is the mathematical proof that we can use matrix product to calculate the probability of multiple-step probability. Since we can think in a intuitive way, CK equation is not important

But there is one thing should be noted: "manipulation of transition probability matrix targeting our task", that is example 4.12 and 4.13 in Sheldon Ross's book.

3.3. Classification of States

First we can classify states using communication

$$\begin{pmatrix} .5 & .5 & 0 & 0 & 0 \\ .5 & .5 & 0 & 0 & 0 \\ .2 & .2 & .2 & .2 & .2 \\ .2 & .2 & .2 & .2 & .2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$\{1, 2\}$, $\{3, 4\}$ and $\{5\}$ by definition. Then we can observe these classes, and notice that there are two types:

1. Class that once we go in, we can never go out. State is recurrent.
   $E[\sum_{n=0}^{\infty} I_n | X_0 = i] = \sum_{n=0}^{\infty} P_{ii}^n = \infty$
2. Class that once we leave, we never go back to. State is transient.
   $E[\sum_{n=0}^{\infty} I_n | X_0 = i] = \sum_{n=0}^{\infty} P_{ii}^n < \infty$

To put them in more mathematical terms, we define r.v. $\sum_{n=0}^{\infty} I_n$, it is how many times we go into state i in the future. Take state 1 as an example, it holds equal probability to stay and leave for state 2.

- If $\sum_{n=0}^{\infty} I_n = 1$, it goes to states 2 and never come back
- If $\sum_{n=0}^{\infty} I_n = 2$, it come back once and stay in state 2
- ……

Recurrent state means the average number of times that we come back is infinite, or $\sum_{n=0}^{\infty} I_n = \infty$ holds greater probability than any other cases. Same analysis goes for transient state

$\{1, 2\}$ and $\{5\}$ are the first, states $\{3, 4\}$ is the second. One interesting fact is that not all states in a finite Markov chain can be transient, but all states in a infinite Markov chain can be transient or recurrent (e.g. random walk).

3.4. Stationary Distribution

We want to divide recurrent state into two substates, but before doing that we have to introduce some definitions

$$N_j = \min\{n > 0 : X_n = j\}$$

It means we are currently in state j , we focus on the number of steps to come back for the first time anc call it N_j , it is an r.v. Then based on that, we define

$$m_j = E[N_j | X_0 = j]$$

It means how many steps to take on average to come back for the first time

1. $m_j < \infty$ is positive recurrent.
    a. Recurrent: we will come back to state j for infinite number of times
    b. Positive: we will be able to observe it for the first time in near future
2. $m_j = \infty$ is null recurrent
    a. Recurrent: we will come back to state j for infinite number of times
    b. Null: even though it will come back for infinite number of times, it will happen in distant future and we won't be able observe it, not even one time.

It can be calculated by

$$\pi = \pi P$$
$$\sum \pi_i = 1$$
$$\pi_i \geq 0$$

Intuitively π should only rely on P and nothing else, since it reflects the property of P . It is most of the time but sometimes it isn't. For a irreducible positive recurrent

3.5. Branching Processes

The core structure of general Markov Chain is transition probability, essentially a matrix most of the time. When studying such process, we do not mention closed form of expectation and variance, because

- The distribution is $q^{(0)}P^n$ at each step X_n , expectation and variance are really difficult to calculate at the presence of matrix P
- The stationary distribution is simple, and the calculation is easy even for an undergraduate

What differs Branching Process from general Markov Chain is how we phrase transition probability, making calculation of expectation and variance for each X_n an easy but interesting task.

3.6. Time Reversible Markov Chains

There are two events:

- Going from i to j forward in time, probability is P_{ij}
- Going from i to j backward in time, which equals to going from j to i forward in time, the related probability is Q_{ij} and P_{ji}

So first let see how Q_{ij} and P_{ji} relate to each other

$$
\begin{aligned}
Q_{ij} &= P\{X_m = j | X_{m+1} = i\} \\
P_{ji} &= P\{X_{m+1} = i | X_m = j\} \\
Q_{ij} &= \frac{\pi_j P_{ji}}{\pi_i} \Longrightarrow \pi_i Q_{ij} = \pi_j P_{ji}
\end{aligned}
$$

even Q_{ij} and P_{ji} describe the same thing, which is state j at step m and state i at step $m+1$, they are not the same and we have to consider the probability of initial state, which

- for Q is π_i , probability to land on i at step $m+1$
- for P is π_j , probability to land j at step m .

Without any restriction, Markov chain has the above property. Then we consider time reverse: the probability of jumping from i to j stays the same whether time goes forward or backward

$$
P_{ij} = P\{X_{m+1} = j | X_m = i\} = Q_{ij} = P\{X_m = j | X_{m+1} = i\}
$$

if Markov chain satisfies the above restriction, it is said to be time reversible. We then have

$$\pi_j P_{ji} = \pi_i Q_{ij} = \pi_i P_{ij}$$

the first equation is one event :

- jumping from j to i forward in time, jumping from i to j backward in time

the second equation links two events:

- jumping from i to j backward in time
- jumping from j to i forward in time

The final conclusion is that, if we can find solution to

$$x_j P_{ji} = x_i P_{ij}$$
$$\sum_i x_i = 1$$

Markov chain is time reversible and x is stationary distribution π

3.7. Markov Decision Process

It is also called Stochastic Dynamic Programming, and is widely used in fields like management and reinforcement learning. In standard Markov chain, there is only one thing that needs to be described: the objective state at each step, but if the process is used to characterize a person's fate, subjective action also needs to be considered: a_n is the action I take based on state X_n and these two together decide what would happen for me at X_{n+1}

$$P\{X_{n+1} = j | X_0, a_0, X_1, a_1, \ldots, X_n = i, a_n = a\} = P\{X_{n+1} = j | X_n = i, a_n = a\} = P_{ij}(a)$$

3.8. Hidden Markov Chains

Chapter 4 Poisson Process

4.1. Exponential Distribution

Failure (or hazard) rate function $r(t)$ is defined as follow, we can use it to study a distribution (not just exponential ones). $r(t)$ represents the conditional probability density that a t-year-old item will fail, it contains all the information we need for a distribution

$$r(t) = \frac{f(t)}{1-F(t)} \Longrightarrow P\{X \in (t, t+dt) | X > t\} = r(t)dt$$

$$F(t) = 1 - e^{-\int_0^t r(t)dt}$$

If we want to use it in exponential distribution,

$$F(x) = 1 - e^{-\lambda x}$$

comparing two formulas we can find out that failure rate function in exponential distribution is λ , which means the function is a constant

$$r(t) = \lambda$$

Sheldon Ross states that if an item follows exponential distribution, due to its memorylessness, its remaining conditiion should be the same as that of a new item, which means

$$P\{X \in (t, t+dt) | X > t\} = r(t)dt = P\{X \in (0, dt)\} = r(0)dt$$
$$r(t) = r(0) \Longrightarrow r(t) = C$$

so the fact "failure rate function is a constant in exponential distribution" roots directly in memorylessness.

Based on exponential distribution, there are two derived random variables: hyperexponential random variable, hypoexponential random variable

4.2. Poisson Process

The connection between exponential distribution and poisson process (distribution): if the times a event happens in interval $[0, t]$ follows poisson distribution

$$P(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

then the time gap T between two events follows exponential distribution

$$f(T) = \lambda e^{-\lambda T}$$

Chapter 5 Brownian Motion

5.1. Definition of Browanian Motion

Definition : Let $(\Omega, \mathcal{F}, P)$ be a complete probability space equipped with a complete filtration $\{\mathcal{F}_t\}_{t \geq 0}$. An n-dimensional $\{\mathcal{F}_t\}$-adapted stochastic process $\{B(t)\}_{t \geq 0}$ is called an n-dimensional standard $\{\mathcal{F}_t\}$-Brownian motion if it satisfies the following properties:

(1) $The\ path\ t \mapsto B(t,\omega)\ are\ continuous$

(2) $For\ every\ t > s \geq 0,\ B(t) - B(s)\ is\ independent\ of\ \mathcal{F}_s$

(3) $For\ every\ t > s \geq 0,\ B(t) - B(s)\ \sim N(0,(t-s)I_n),\ where\ I_n\ is\ the\ n \times n\ identity\ ma$

(4) $B(0) = 0$

It is also called Wiener process, so the notation can also be $W(t)$. The core idea is the function structure of random variable. I don't care what the specific form of P in $(\Omega, \mathcal{F}, P)$, as long as we choose a function family $B(t,\omega)$ based on that measure and it satisfies the above four restrictions, this stochastic process is called Brownian motion.

The second and the third properties mean that an experiment is performed and the result is ω which we don't know, all we have are the information we have on ω which is $\mathcal{F}_s$ and the collabration result of function structure of B and the measure P . If we are in time s , we can't know which event happens or doesn't happen in $\sigma(B(t) - B(s))$, and all we know is $P(B(t) - B(s)|\mathcal{F}_s) = N(0,(t-s)I_n)$. This property can derive what we call independent increment. To explain it in detail, If we have $0 \leq s_1 < t_1 \leq s_2 < t_2$, $B(t_1) - B(s_1)$ and $B(t_2) - B(s_2)$ are independent.

Then still we use $0 \leq s_1 < t_1 \leq s_2 < t_2$, the properties of Brownian motion :

$$Cov[B(t_1) - B(s_1), B(t_2) - B(s_2)] = 0$$

$$E[B(t_1 - s_1)B(t_2 - s_2)] = 0$$

$$E[W(t)] = 0$$

$$Var[W(t)] = t = E[W^2(t)]$$

$$if\ s < t,\ then\ Cov[W(s), W(t)] = E(W(s)W(t)) = min(s,t) = s$$

For a Brownian motion, if we do the followings transformations, the result would still be a Brownian motion. What a transformation do is to change the function structure of $B(t,\omega)$

$$reflection : X(t) = -B(t)$$

$$translation : X(t) = B(t+s) - B(s),\ \forall s \geq 0$$

$$rescaling : X(t) = \frac{1}{\sqrt{a}}B(at),\ \forall a > 0$$

$$inversion : X(t) = tB(\frac{1}{t}),\ t > 0\ and\ X(0) = 0$$

Then we discuss things related to continuity which I think maybe the most important feature of Brownian motion. Let's say we want to know how the instant increment at t

$$dW(t) = \lim_{\Delta t \to 0} W(t+\Delta t) - W(t) = W(t+dt) - W(t) \sim N(0, dt)$$

Then we want to calculate the derivative

$$\frac{dW(t)}{dt} = \lim_{\Delta t \to 0} \frac{W(t+\Delta t) - W(t)}{\Delta t}$$

Instant increment always exists, because it is just subtraction, but derivative is not the case, so we have to analyze,

$$E[\frac{W(t+\Delta t) - W(t)}{\Delta t}] = 0 \quad Var[\frac{W(t+\Delta t) - W(t)}{\Delta t}] = \frac{1}{\Delta t}$$

$$\Rightarrow \lim_{\Delta t \to 0} Var[\frac{W(t+\Delta t) - W(t)}{\Delta t}] = \infty$$

$\frac{dW(t)}{dt}$ is a random variable, the above two mean that the average of its possible values is zero and the value can reach infinity. In this case it indicates Brownian motion doesn't have derivative. Quadratic variation is related to derivative

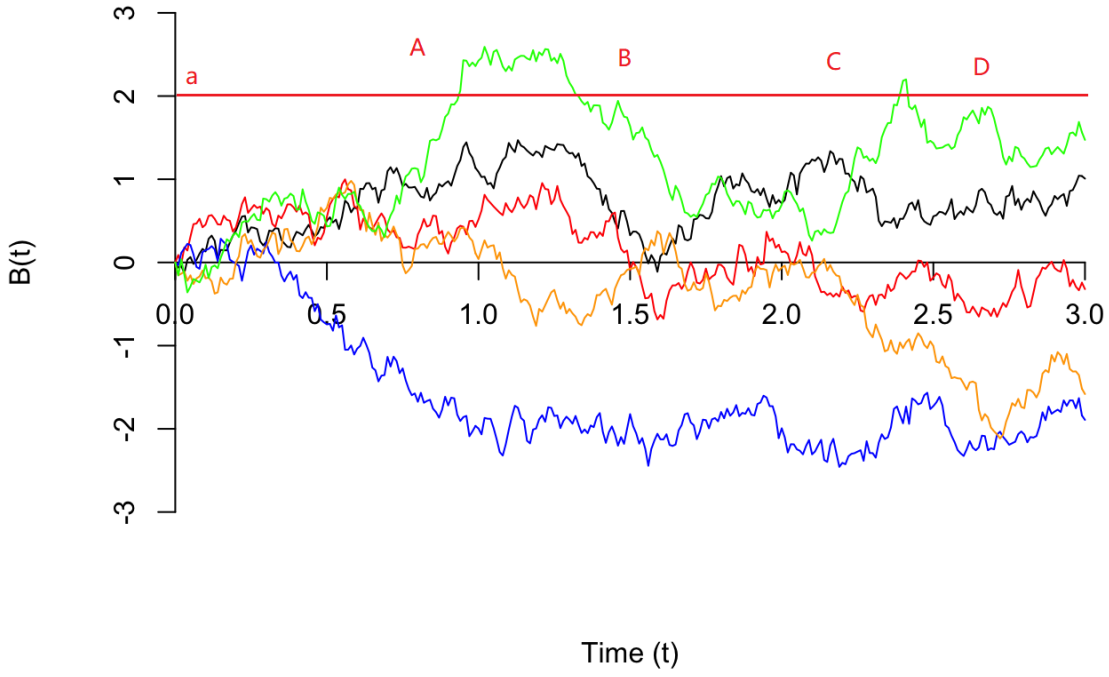
5.2. First Hitting Time

It is also called first passage time

Definition :

$$\tau_a = min\{t : t \geq 0, B(t) = a\}$$

The logic is that an experimental result ω is one try of Brownian motion which is the following figure



When we have a fixed a , we use it to find the earliest time at which the path of ω arrives at a , so it would be t_A in the figure. Essentially one ω means one path, and one path can find one time value, so what it really does is to gather a part of information about ω and project it on nonnegative number and that is the exact definition of stopping time.

Then we wonder what is its distribution like?

$$F_{\tau_a}(t) = P(\tau_a < t) = 2N(-\frac{|a|}{\sqrt{t}})$$

$$f_{\tau_a}(t) = \frac{|a|}{\sqrt{2\pi t^3}} exp[-\frac{a^2}{2t}],\ t > 0$$

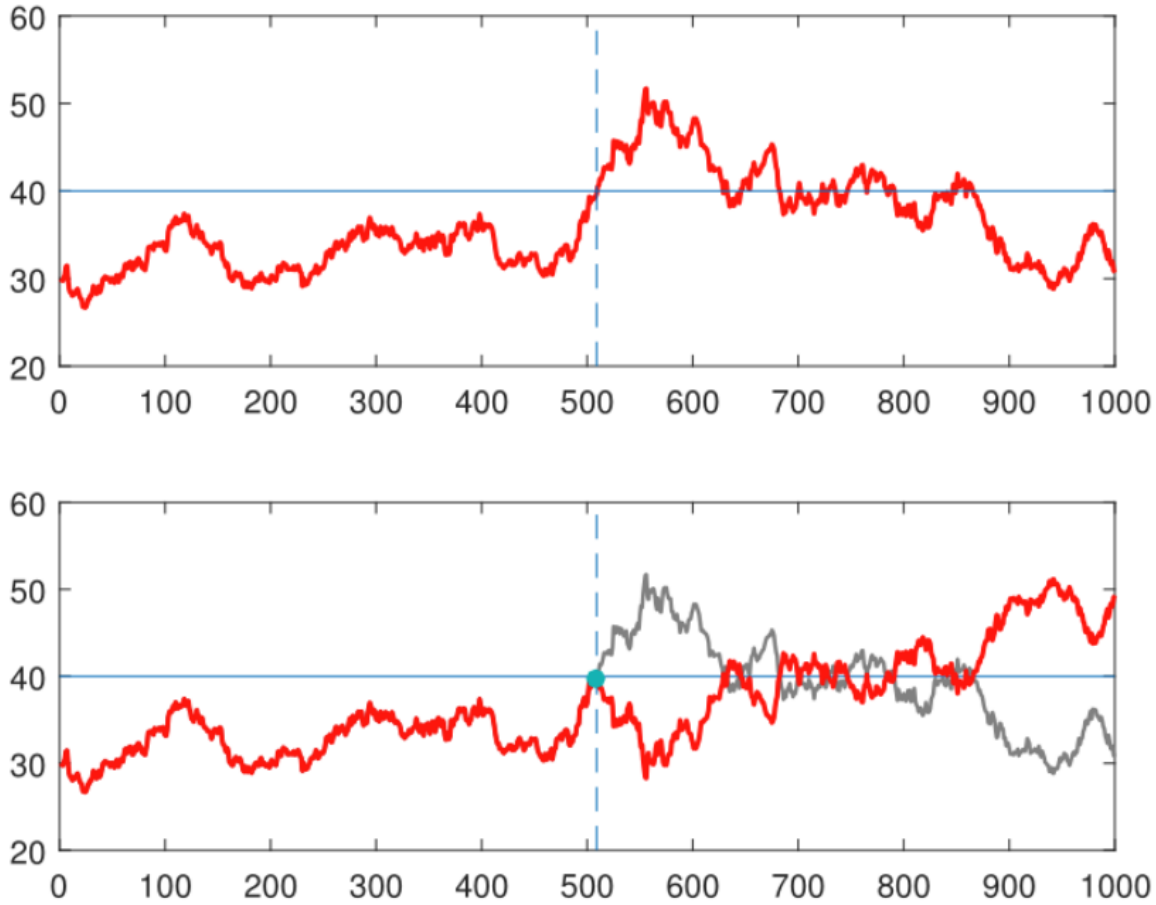
$$E(\tau_a) = \infty$$

$$P(\tau_a < \infty) = 1$$

5.3. Reflection Principle

Definition : $B(t)$ is a Brownian motion, and $\tilde{B}(t)$ is defined as

$$\tilde{B}(t) = \begin{cases} B(t) & t \in [0, \tau_a] \\ 2a - B(t) & t \in [\tau_a, \infty) \end{cases}$$



It essentially changes the function structure of $B(t)$ for ω . Based on the original $B(t)$, we derive function $\tau_a(\omega)$, each ω has value τ_a . Before τ_a the path for ω stays intact, but we reflect the path after it. The new function structure $\tilde{B}(t)$ is still a Brownian motion.

Definition : a Brownian motion $B(t)$, if $t \in [0, T]$ and we have

$$M_T = \max_{t \in [0,T]} B(t)$$

then M_T is maximum of Brownian motion in $[0, T]$

Each ω has a path and we can find the maximum on this path in $[0, T]$, so each ω corresponds to a maximum, which means M_T is a random variable.

The reason why we bring up M_T is that we can deduce the distribution function of it using reflection principle, we have

$$F_{M_t}(a) = \int_{-\frac{a}{\sqrt{t}}}^{-\frac{a}{\sqrt{t}}} \frac{1}{\sqrt{2\pi}} exp[-\frac{1}{2}x^2]dx$$

5.4. Three Variations of Brownian motion

Definition 1 of Brownian bridge : $B(t)$ is a Brownian motion, we let

$$B^*(t) = B(t) - tB(1), \quad t \in [0,1]$$

and call $B^*(t)$ Brownian bridge.

$B^*(0) = B^*(1) = 0$ and it is why $B^*(t)$ is named Brownian bridge. But the above definition only limits time in $[0,1]$, we can do more.

Definition 2 of Brownian bridge : $B(t)$ is a Brownian motion, we let

$$X(t) = B(t) - \frac{t}{T}B(T), \quad t \in [0,T]$$

We can extend it further

Definition 3 of Brownian bridge : $B(t)$ is a Brownian motion, we let

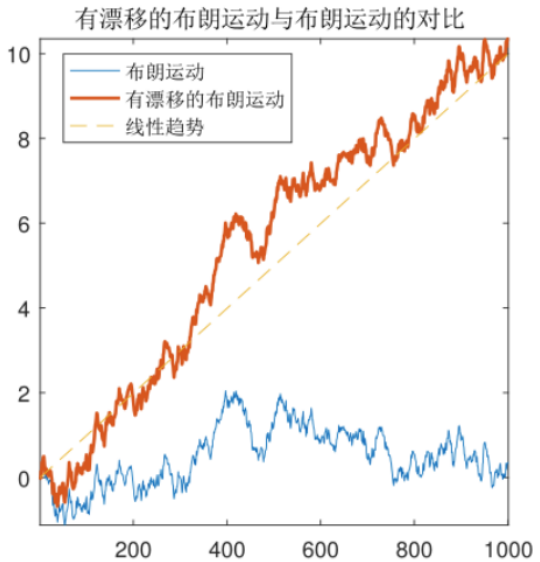
$$X(t) = B(t) - \frac{t}{T}B(T)$$
$$X^{a\to b}(t) = a + (b-a)\frac{t}{T} + X(t)$$
$$t \in [0,T]$$

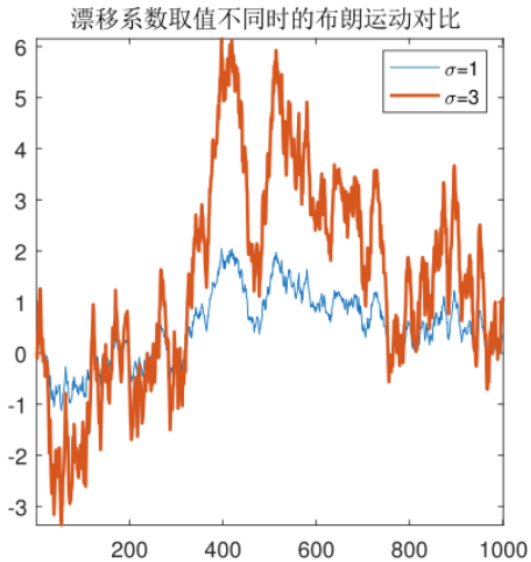
The first two definitions only uses $B(0) = 0$ as two anchors, however the third can use any anchors we want : $X^{a\to b}(0) = a$ and $X^{a\to b}(T) = b$

Definition : $B(t)$ is a Brownian motion

$$X(t) = \mu t + \sigma B(t), \quad t \geq 0$$

$X(t)$ is a stochastic process and is called Brownian motion with drift. μ is drift coefficient and σ is volatility.





For Brownian motion with drift

$$E(X(t)) = \mu t, \quad Var[X(t)] = \sigma^2 t$$

Definition :

$$X(t) = \mu t + \sigma B(t)$$
$$G(0) > 0$$
$$G(t) = G(0)\exp[X(t)], \quad t \geq 0$$

We call $G(t)$ geometric Brownian motion (GBM)

5.5. Quadratic Variation

Quadratic variation to a stochastic variable is variance to a random variable. By this logic, we can have "co-quadratic variation" since there is covariance. For detailed demonstration, see Girsanov theorem.

Chapter 6 Martingales

6.1. Conditional Expectation

What is conditional expectation? Intuitively, we can write the following formula:

$$E[X|A] = \frac{\int_A X\, dP}{P(A)}$$

What is $E[X|A]$? After a experiment is performed, I want to know which ω is it. So I pick a subset A of Ω and order the system "you have to tell whether ω is in A" . If the system tells me "it is in A" , based on the information I calculate the serial number of ω in $\sigma(X)$, in other words, I deduce which set in $\sigma(X)$ that ω belongs to. But there are chances that ω is not in A , so the above formula is a function only defined in $\omega \in A$. what should we do then?

The smart way is to pick a bunch sets and extend ω to the whole Ω , that's exactly what we do by $E[X|\mathcal{F}]$! Revisiting the section introducing σ algebra, we know that $\mathcal{F}$ essentially is the collection of many "A" , so we denote

$$E[X|\mathcal{F}](\omega) = E[X|A] \; if \; \omega \in A$$
$$E[X|\mathcal{F}](\omega) = E[X|A^c] \; if \; \omega \in A^c$$

Since $\mathcal{F}$ is a function structure to receive information from ω and the result of $\mathcal{F}$ (which set ω belong to) will help us to calculate the serial number of the set in $\sigma(X)$ at which ω is in.

But there is another issue has to be addressed. When ω comes in, there are going to be a lot of A that ω belongs to, which one do we choose? From Sherev book we know it's the smallest

$$E(X|\mathcal{F})(\omega) = E(X|A)$$
$$A\ is\ the\ smallest\ set\ that\ contains\ \omega$$

The formal definition from Durrett:

Definition : Given a probability space $(\Omega, \mathcal{F}_0, P)$, a σ field $\mathcal{F} \subset \mathcal{F}_0$, and a random variable $X \in \mathcal{F}_0$ with $E|X| < \infty$. We define the conditional expectation of X given $\mathcal{F}$, $E(X|\mathcal{F})$, to be any random variable Y that has

$$(1)\ Y \in \mathcal{F},\ that\ is,\ is\ \mathcal{F}\ measurable,\ and$$

$$(2)\ for\ all\ A \in \mathcal{F},\ \int_A X\,dP = \int_A Y\,dP$$

Any Y satisfying the above two is said to be a version of $E(X|\mathcal{F})$

The intuition for this definition is : $E(X|\mathcal{F})$ is a function and it's only difference compared with X is when ω_0 comes in, $X(\omega)$ takes value of $X(\omega_0)$ and $E(X|\mathcal{F})$ take the average of $X(\omega)$ in A which ω_0 belongs to. So why not define $E(X|\mathcal{F})$ as the function that can smooth the rough edges of $X(\omega)$.

6.2. Computing Expectation by Conditioning

In Sheldon Ross's book, he demonstrates how to use conditional expectation to calculate full expectation. The core formula is

$$E[X] = E[E[X|Y]]$$

$E[X|Y]$ itself is a random variable, when we took expectation, the result is $E[X]$. The reason why we use this method is that sometimes the pdf for X is just too complicated, and $E[X|Y]$ is much more easier to do.

Let's make things more interesting: what if what we want is $E[X|Y]$ and its pdf is complicated, can we add another variable to help? The answer is yes.

$$E[X|Y] = E[E[X|Y,W]|Y]$$

Similarly we can continue to expand $E[X|Y,W]$ if we do not like we have. Variance have similar formula

$$Var(X) = E[Var(X|Y)] + Var(E[X|Y])$$

6.3. Martingales

Definition : In a filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, P)$, if

$$(1)\ E\,|X_n| < \infty$$

$$(2)\ X_n\ is\ adapted\ to\ \mathcal{F}_n$$

$$(3)\ E(X_{n+1}|\mathcal{F}_n) = X_n\ for\ all\ n$$

Then process X is called a martingale.

So intuitive understanding is this: as n increases, $\mathcal{F}_n$ expands. Since $\mathcal{F}_{n-1} \subset \mathcal{F}_n$ and we have a funtion X_n on $\mathcal{F}_n$, we can find its conditional expectation on $\mathcal{F}_{n-1}$, surprisingly, it is X_{n-1} . We then call $X = (X_n : n \geq 0)$, essentially a process, a martingale.

In terms of gamble, what martingale is trying to say is that if I have capital X_n at time n and I want to continue to gamble, based on the information I currently have $\mathcal{F}_n$, my chances of winning and losing are equal, which means the expectation of capital at time $n+1$ would be X_n

supermartingale : $E(X_{n+1}|\mathcal{F}_n) \leq X_n$, meaning based on current information, I am going to lose money.

submartingale : $E(X_{n+1}|\mathcal{F}_n) \geq X_n$, meaning based on current information, I am going to win money.

Example 5.2.1 : Everything in this book is in $(\Omega, \mathcal{F}, P)$, it's the same deal here. ω is $(\omega_1, \omega_2, \dots)$ in which ω_1 means the situation at step 1 and index can extend to infinity

- $\xi_n = 1\ or\ -1$, what author really means is that the function ξ_n is applied to all the element in Ω , and what decides the function value for ω is the nth component.
- $X_n = \xi_1 + \dots + \xi_n$, it is still applied to Ω , but what decide the funtion value for ω is its first nth component.
- $\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n)$, each ξ is a σ algebra, what it does is to find $\sigma(\sigma(\xi_1), \dots, \sigma(\xi_n))$ that can cover all the information each $\sigma(\xi)$ has. Since X_n shrink the information of $\xi_1, \dots, \xi_n$ and $\mathcal{F}_n$ expands it, of course X_n is $\mathcal{F}_n$ measurable.

6.4. Local Martingale

After we define martingale, we can define local martingale

Definition: A process X is a local martingale if there exists a sequence of stopping times T_n which satisfies

1. T_n increases almost surely: $P\{T_n < T_{n+1}\} = 1$
2. T_n diverges almost surely: $P\{\lim_{n\to\infty} T_n = \infty\} = 1$

the process $Y_t = X_{min\{t,T_n\}}$ is a martingale for every n

Remark: first an experiment ω has happened, we have stoping time $T_n(\omega)$. Standing at t , we compare which is smaller, t or $T_n(\omega)$, let's denote it as t_{sm}. Then we use it to find a value which is $X(t_{sm}, \omega)$ and treat it as the value of a new process $Y(t, \omega)$ at t. For every T_n , we can find a such process. If we examine all these processes and find them all satisfy martingale property, we call the orginal X that generates all these derivative martingale, local martingale.

One thing should be noticed: there are no local martingales in discrete time; they are a continuous time phenomenon.

6.5. Semimartingale

A real valued stochastic process X is called a semimartingale if it can be decomposed as

1. the sum of a local martingale
2. a càdlàg adapted finite-variation process

Semimartingales are “good integrators”, forming the largest class of processes with respect to which the Itô integral and the Stratonovich integral can be defined. In other words, the purpose of semimartingale is to serve stochastic integral. The class of semimartingales is quite large (including, for example, all continuously differentiable processes, Brownian motion and Poisson processes). Submartingales and supermartingales together represent a subset of the semimartingales.

Examples of semimartingale

- Adapted and continuously differentiable processes are continuous finite variation processes, and hence semimartingales.
- Brownian motion is a semimartingale.
- All càdlàg martingales, submartingales and supermartingales are semimartingales.
- Itō processes, which satisfy a stochastic differential equation of the form $dX = \sigma dW + \mu dt$ are semimartingales. Here, W is a Brownian motion and σ, μ are adapted processes.
- Every Lévy process is a semimartingale.

Although most continuous and adapted processes studied in the literature are semimartingales, this is not always the case.

- Fractional Brownian motion with Hurst parameter $H \neq 1/2$ is not a semimartingale.

Chapter 7 Volterra process

The Volterra process is a class of processes that depend on the convolution of a function and their past sample path. It generalizes the classical Heston model, which is widely used for option pricing, as well as the more recent rough Heston model

The rough Heston model is a variant of the classical Heston model, where the variance process is replaced with a fractional square-root diffusion. This means that the variance process in the rough Heston model depends on its past self, making it non-Markovian in the current variance state. This also makes it more difficult to simulate using classical techniques.

In summary, the Volterra process is a generalization of both the classical and rough Heston models, while the rough Heston model is a variant of the classical Heston model with a different variance process. The two are connected in that they both belong to the class of affine Volterra processes.

Affine Volterra processes that are solutions of stochastic convolution equations with affine coefficients. This family includes a broad range of dynamics such as Ornstein–Uhlenbeck, Heston, and the Rough Heston processess

Chapter 8 Stochastic Calculus

8.1. Structure of Stochastic Calculus

Calculus contains two parts: integral and differential, one is the reverse of the other. Of course stochastic calculus contains these two parts too. We aim to study integral, but at the same time differential would inevitably show up. The only difference between regular calculus and stochastics calculus is that the latter replace dx with the increment of stochastic process $dX(t)$.

Now the question is which $X(t)$ do we choose? Since Brownian motion is wildly used in so many circumstances, we pick Brownian motion. So the task is to define:

$$\int_0^t f(s)dB(s)$$

We call it Ito integral.

After we accomplish this task, we have to consider other stochastic processes, which is similar to Brownian motion so that we can use the definition above but at the same time more general so that we can extend our definition. So we come up with the following stochastic process and it is called Ito process

$$\int_0^t dX(s) = \int_0^t F(s)ds + \int_0^t G(s)dB(s)$$

A less rigorous form is

$$dX(t) = F(t)dt + G(t)dB(t)$$

It actually represents a family of processes, because we can pick $F(t), G(t)$. Can we extend it further? The answer is yes! The form is

$$f(t, \omega) = f(t, X(t))$$

in which $X(t)$ is any process that contains Brownian motion (often it is Ito process) . In order for us to do integral, we have to use something like $df = dt + \dots$ and this is where we introduce Ito Lemma. It tells us Brownian motion and any process that contains Bm have different feature than regular function: their quadratic variation is not zero. We can't just simply do $df = ()dt + ()dX(t)$ and the right formula is

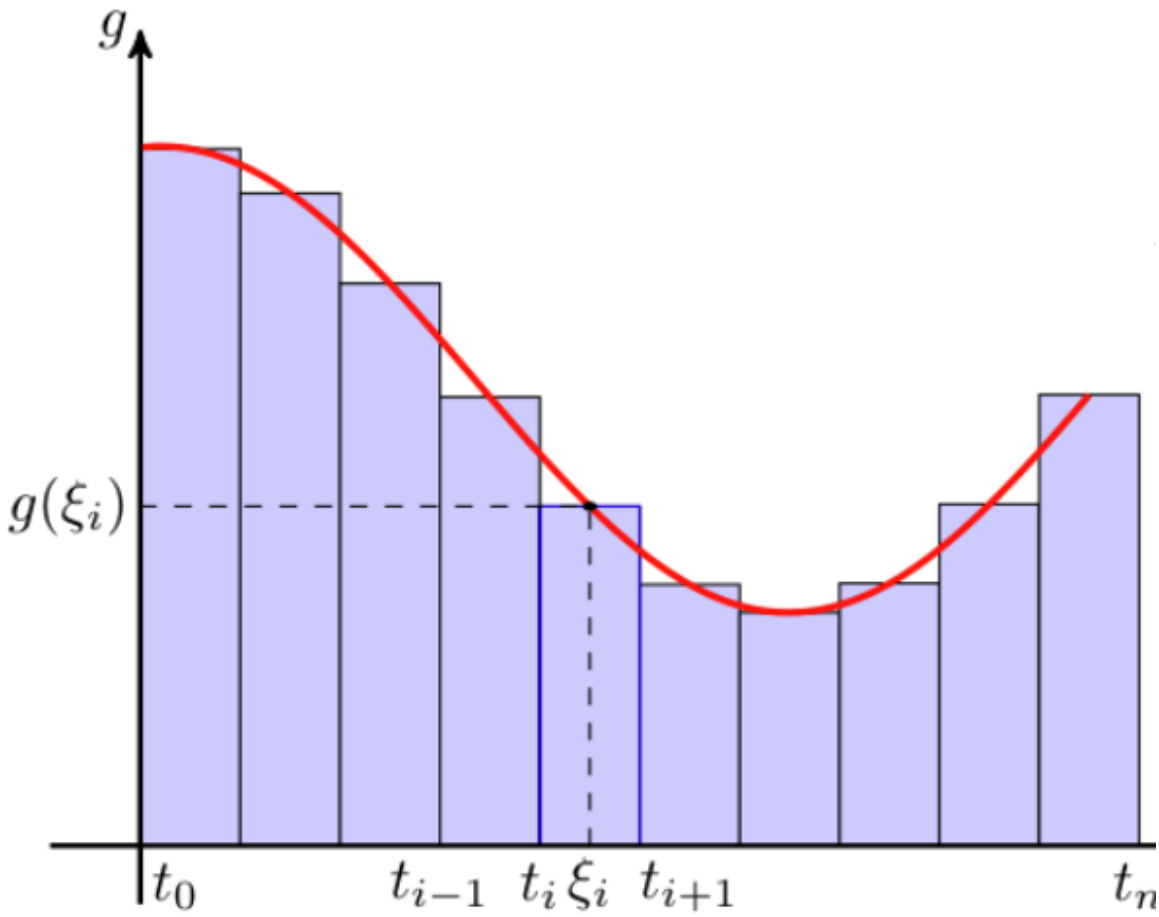
$$df = f_t dt + f_x dX(t) + \frac{1}{2} f_{xx} [dX(t)]^2$$

The things we have introduced so far have a name: Ito Calculus. It is just one type of stochastic calculus and there is another type: Malliavin calculus.

8.2. Review of Riemann Integral

When we metion integral, most of the time we mean Riemann integral.

$$\begin{aligned}
R(T) &= \int_0^T g(t)dt \\
R_1 &= \lim_{\Delta t \to 0} \sum_{i=0}^{n-1} g(t_i)(t_{i+1} - t_i) \\
R_2 &= \lim_{\Delta t \to 0} \sum_{i=0}^{n-1} g(\xi_i)(t_{i+1} - t_i), \quad \xi_i \in [t_1, t_{i+1}] \\
R(T) &= R_1 = R_2
\end{aligned}$$



We should note that since $g(t)$ is a smooth function, which means its quadratic variation is zero, so when we calculate the area of a small area, we can choose any function value from $g(t_i)$ to $g(t_{i+1})$ and the integral result would be the same.

Let's then expand it a little more, we call the following Riemann-Stieltjes integral

$$
\begin{aligned}
RS(T) &= \int_0^T g(t)df(t) \\
RS_1 &= \lim_{\Delta t \to 0} \sum_{i=0}^{n-1} g(t_i)[f(t_{i+1}) - f(t_i)] \\
RS_2 &= \lim_{\Delta t \to 0} \sum_{i=0}^{n-1} g(\xi_i)[f(t_{i+1}) - f(t_i)], \quad \xi_i \in [t_1, t_{i+1}] \\
RS(T) &= RS_1 = RS_2
\end{aligned}
$$

Again since $g(t)$ and $f(t)$ are smooth function and the quadratic variations of the two are zero, so we can choose any function value from $g(t_i)$ to $g(t_{i+1})$ and the integral result would be the same.

8.3. Stochastic Integral

Based on the definition of regular integral, we can write similar formula

$$I(T) = \int_0^T g(t)dB(t)$$

$B(t)$ is a $\mathcal{F}_t$ measurable standard Brownian motion and $g(t)$ is a $\mathcal{F}_t$ measurable stochastic process. What the formula means is that each ω has a function $g(t, \omega)$ and $B(t, \omega)$. For each ω we can do the above integral, so the result $I(T)$ is still a random variable.

So the question is how do we do the above integral? We can't treat it as a regular integral because the corner stone of Riemann integral is all functions are smooth, but here $B(t)$ is a continuous and non-differentiable function, i.e., not smooth. The direct symptoms are $B(t)$ doesn't have derivative and when we try

$$\lim_{\Delta t \to 0} \sum_{i=0}^{n-1} g(\xi_i)[B(t_{i+1}) - B(t_i)]$$

different value of ξ_i would lead to different result value. To address this issue we have the following two sub-sections.

8.3.1. Simple Integrands

For simple process $g(t)$ whose definition can be found in the book, we define its integral as follow.

Definition ： Ito Integral means

$$I(T) = \int_0^T g(t)dB(t) = \lim_{\Delta t \to 0} \sum_{i=0}^{n-1} g(t_i)[B(t_{i+1}) - B(t_i)]$$

we always choose the left boundary t_i to calculate the function value $g(t)$ in interval $[t_i, t_{i+1}]$

Definition : Stratonovich integral means

$$I(T) = \int_0^T g(t) \circ dB(t) = \lim_{\Delta t \to 0} \sum_{i=0}^{n-1} g(\frac{t_i + t_{i+1}}{2})[B(t_{i+1}) - B(t_i)]$$

we always choose the middle point $\frac{t_i + t_{i+1}}{2}$ to calculate the function value $g(t)$ in interval $[t_i, t_{i+1}]$

Example : In financial market, $g(t)$ and $B(t)$ mean the shares of a stock we hold at time t and the stock price

$$I(t) = \begin{cases} g(t_0)[B(t_1) - B(t_0)] = g(0)B(t), & t \in [t_0, t_1] \\ g(0)B(t_1) + g(t_1)[B(t) - B(t_1)], & t \in [t_1, t_2] \\ g(0)B(t_1) + g(t_1)[B(t_2) - B(t_1)] + g(t_2)[B(t) - B(t_2)], & t \in [t_2, t_3] \\ \cdots\cdots\cdots\cdots \end{cases}$$

What $I(t)$ means is profit of our asset (stock). When $t \in [t_0, t_1]$, our chares are $g(t_0)$, the price will change from $B(t_0)$ to $B(t)$, so the total value change (profit) is $g(t_0)[B(t_1) - B(t_0)]$.

When $t \in [t_1, t_2]$, the share changes from $g(t_0)$ to $g(t_1)$ and we think it just magically changes and doesn't effect our profit. First we sell the shares we hold at t_1 and make profit $g(t_0)[B(t_1) - B(t_0)]$ and then we have a total value change $g(t_1)[B(t) - B(t_1)]$, combination of the two is the profit we gain starting from t_0 .

Properties of Ito Integral :

- $I(t)$ is martingale
- $E[I(t)] = 0$

8.3.2. General Integrands

For a general stochastic process, we use a sequence of simple stochastic process to approach and define the limit of their integral value as the integral value of the general stochastic process.

8.4. Ito Lemma

It is also called Ito formula or Ito-Doeblin formula. To understand its origin, we have to go back to the basics. In traditional calculus, if there is a function $f(x)$, then the argument is x ; if there is a function $f(x, y)$, then the argument are x and y . In stochastic calculus, the arguments are t and $B_1(t), B_2(t), \ldots$

For a differentiable function $f(x)$, if we want to quantify the increment, we have the following taylor expansion

$$df(x) = f'(x)dx + \frac{1}{2}f''(x)[dx]^2 + \frac{1}{3!}f^{(3)}(x)[dx]^3 + \ldots$$

It holds trure for any function $f(x)$ and means the increment consists of infinite parts of $[dx]^n$. Notice the only restriction is that f is differentiable so that it can have derivatives. In undergraduate calculus we always wrote

$$df(x) = f'(x)dx$$

the reason is that variations $[dx]^n$ that have order higher than 1 are all zero, so we omit them. But now we have Brownian motion and we put it at the same level as argument x , so the above expansion becomes

$$df(B(t)) = f'(B(t))dB(t) + \frac{1}{2}f''(B(t))[dB(t)]^2 + \frac{1}{3!}f^{(3)}(B(t))[dB(t)]^3 + \dots$$

Since quadratic variation of $B(t)$ is not zero, in other words unlike x , $B(t)$ curve is not smooth, the second term of the expansion should stay, hence the following

$$df(B(t)) = f'(B(t))dB(t) + \frac{1}{2}f''(B(t))dt$$

It is the main difference between regular calculus and stochastic calculus. Obeserve the difference, we can see stochastic differential equation naturally has two variables - t and $B(t)$ and the latter can transform into the former by $[dB(t)]^2 = dt$

$$df(x) = f'(x)dx$$
$$df(B(t)) = f'(B(t))dB(t) + \frac{1}{2}f''(B(t))dt$$

Let's make things more complicated, If we let f be $f(t, x)$ and since we should keep all the second order terms

$$\begin{aligned} df(t, B(t)) &= f_t(t, B(t))dt + f_x(t, B(t))dB(t) + \frac{1}{2}f_{xx}(t, B(t))[dB(t)]^2 \\ &+ \frac{1}{2}f_{tt}(t, B(t))[dt]^2 + \frac{1}{2}f_{tx}(t, B(t))[dtdB(t)] + \frac{1}{2}f_{xt}(t, B(t))[dB(t)dt] \\ &= f_t(t, B(t))dt + f_x(t, B(t))dB(t) + \frac{1}{2}f_{xx}(t, B(t))dt \end{aligned}$$

Then we have the following theorem,

Theorem : Ito-Doeblin formula for Brownian motion, let $f(t, x)$ be a function for which the partial derivatives $f_t(t, x)$, $f_x(t, x)$ and $f_{xx}(t, x)$ are defined and continuous, and let $B(t)$ be a Brownian motion. Then for every $T \geq 0$,

$$f(T, B(T)) = f(0, B(0)) + \int_0^T f_t(t, B(t))dt + \int_0^T f_x(t, B(t))dB(t) + \frac{1}{2}\int_0^T f_{xx}(t, B(t))dt$$

We can pile more shit up!

Definition : If stochastic process $\{X(t)\}$ satisfies

$$dX(t) = F(t)dt + G(t)dB(t)$$

or

$$\int_0^T dX(t) = \int_0^T F(t)dt + \int_0^T G(t)dB(t)$$

$F(t)$ and $G(t)$ are $\mathcal{F}_t$ measurable. We call it Ito process.

The Instantaneous increment is divided into two parts, we call the first deterministic and the second stochastic. Because function t is the same for all elements in Ω , hence no matter which ω is there, the increment dt is the same. However $B(t) = B(t, \omega)$, the increment of it varies due to different ω .

If $F(t) = \mu$ and $G(t) = \sigma$ are constants, the process is Brownian motion with drift. If $F(t) = 0$ and $G(t) = 1$, the process is standard Brownian motion. So we can see Ito process is the generalization of Brown motion and we use it to repalce $B(t)$ in $f(t, B(t))$

$$\begin{aligned} df(t, X(t)) &= f_t dt + f_x dX(t) + \frac{1}{2} f_{xx} [dX(t)]^2 \\ &= f_t dt + f_x [F(t)dt + G(t)dB(t)] + \frac{1}{2} f_{xx} [F(t)dt + G(t)dB(t)]^2 \\ &= (f_t + f_x F + \frac{1}{2} f_{xx} G^2)\ dt + f_x G\ dB(t) \end{aligned}$$

Definition : Ito-Doeblin formula for an Ito process, it is the integral form of the above, see shreve II

For this formula, there is one more layer of function comprared with just Brownian motion

$$B(t) \Rightarrow X(t) \Rightarrow f(t, X(t))$$

Ito-Doeblin formula for an Ito process has generalized our understanding for differential equation, but when we encounter a process, how do we deduce the equation? The answer is fuck Ito-Doeblin formula and go to the basics. It means if we have

$$f = f(t, X(t)) \quad X(t) = g(t, B(t))$$

first we use taylor expansion for f targeting t and $X(t)$ keeping all the second order terms, then we do taylor expansion on $X(t)$, still keeping all the second order terms.

Chapter 9 Stochastic Differential Equation

9.1. Demonstration of Geometric Brownian Motion

Let's see a differential equation first

$$dS(t) = \mu S(t)dt$$
$$\frac{dS(t)}{S(t)} = \mu dt$$
$$d\ln\left(S(t)\right) = \mu dt$$
$$\ln\left(S(t)\right) - \ln\left(S(0)\right) = \mu t$$
$$S(t) = S(0)e^{\mu t}$$

The key to solve it is the transition from the second step to the third step. The reason why we can use $\frac{dS(t)}{S(t)} = d\ln\left(S(t)\right)$ is because $S(t)$ is a deterministic function with respect to t , so the regular $dy = f'dx$ is correct. But if we introduce a Brownian motion part

$$dS(t) = \mu S(t)dt + \sigma S(t)dB(t)$$
$$\frac{dS(t)}{S(t)} = \mu dt + \sigma dB(t)$$

we can't use $\frac{dS(t)}{S(t)} = d\ln\left(S(t)\right)$. However we can guess since the left side has this expression, it must somehow is related to $d\ln\left(S(t)\right)$ even though they are not equal, so we do

$$\begin{aligned} d\ln\left(S(t)\right) &= \frac{1}{S(t)}dS(t) - \frac{1}{2}\frac{1}{S(t)^2}\left[dS(t)\right]^2 \\ &= \frac{1}{S(t)}dS(t) - \frac{1}{2}\frac{1}{S(t)^2}\left[\mu S(t)dt + \sigma S(t)dB(t)\right]^2 \frac{dS(t)}{S(t)} = d\ln\left(S(t)\right) + \frac{\sigma^2}{2}dt \\ &= \frac{1}{S(t)}dS(t) - \frac{\sigma^2}{2}dt \end{aligned}$$

then

$$\frac{dS(t)}{S(t)} = d\ln\left(S(t)\right) + \frac{\sigma^2}{2}dt$$
$$d\ln\left(S(t)\right) + \frac{\sigma^2}{2}dt = \mu dt + \sigma dB(t)$$
$$d\ln\left(S(t)\right) = (\mu - \frac{\sigma^2}{2})dt + \sigma dB(t)$$
$$\ln\left(S(t)\right) - \ln\left(S(0)\right) = (\mu - \frac{\sigma^2}{2})t + \sigma B(t)$$
$$S(t) = S(0)e^{(\mu - \frac{\sigma^2}{2})t + \sigma B(t)}$$

So the key to solve a SDE is the guessing what the form of big $d()$ might be and deduce the relationship between what it is now and what might be