19

STOCHASTIC PROCESS

19.1 Stochastic process

19.1.1 Basic definition and concepts

Definition 19.1.1 (stochastic process). *[1] A stochastic process X is a collection of random variables $\{X_t\}_{t\in T}$ on some fixed probability triple $(\Omega, \mathcal{F}, P)$, indexed by a subset T of the real numbers.*

*If the index set is the positive integers, we call X a **discrete-time stochastic process**. If the T is an open interval on $\mathbb{R}$, it is called a **continuous-time stochastic process***

Definition 19.1.2 (sample path). *[2, p. 45] For a discrete-time stochastic process, the sequence of numbers $X_1(\omega), X_2(\omega), \ldots$ for any fixed $\omega \in \Omega$ is called a sample path. For continuous stochastic process, the mapping*

$$t \in T \to X_t(\omega) \in \mathbb{R}$$

is a sample path.

Remark 19.1.1 (interpretation). A stochastic process involves two variables, $t \in T, \omega \in \Omega$. For each fixed t, the mapping

$$\omega \in \Omega \to X_t(\omega) \in \mathbb{R}$$

is a random variable, and for each fixed ω, the mapping

$$t \in T \to X_t(\omega) \in \mathbb{R}$$

is a sample path.

Remark 19.1.2 (sample path examples).

- One trivial case is that $X_1, X_2, \ldots$ are the same mapping from sample space, then the sample path associated with a ω will be a horizontal line. However, if X_1, X_2 are different mapping from the sample space, then the sample path will not be a horizontal line.
- For a non-trivial case: consider $X_t(\omega) = Z(\omega)\sin(t)$. If $Z(\omega) = 0.5$, then $X_t = 0.5\sin(t)$
- For another non-trivial case: consider $X_t(\omega) = \omega^t$ assuming $\Omega = [0, 1]$

Note 19.1.1 (interpretation on sample space and σ algebra). [3, p. 97] Use random walk as example.

- Let $\omega \in \Omega$. One way to think of ω is as the random sample path. A random experiment is performed, and its outcome is the path of the random walk of horizon T. This random experiment outcome can be thought as a long sequence coin-toss outcome such that we map this long sequence coin-toss outcome to a random walk path, a function parameterized by time. See Figure 19.1.1.
- If time index is from 0 to T, then total number of sample points in Ω is 2^T.
- Some example random events in Ω are: (1) coin-toss sequences starting with H; (2) coin-toss sequence starting with HT.
- Then the σ-algebra is the σ-algebra on the sample-path space such that some 'suitable' subsets of all possible paths can be evaluated. For example, we can evaluate $P(W_t < 0.5)$ for some $t \geq 0$.

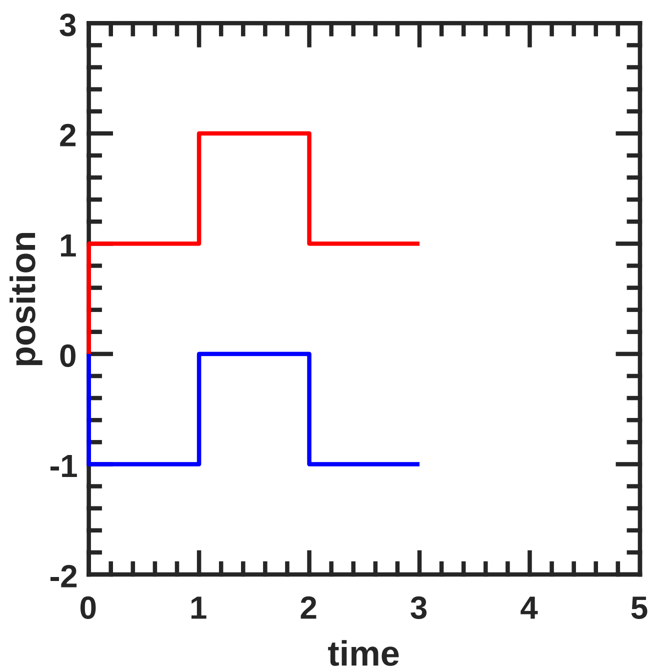


Figure 19.1.1: An illustration of a random walk mapping a sample point, ω, to a trajectory parameterized by time, where red trajectory sample point HHT, and blue trajectory has sample point THT.

19.1.2 Filtration and adapted process

Definition 19.1.3 (filtration). *The collection* $\{\mathcal{F}_t, t \geq 0\}$ *of* σ*-field on sample space* Ω *is called a filtration if*

$$\mathcal{F}_s \subseteq \mathcal{F}_t, \forall 0 \leq s \leq t.$$

Remark 19.1.3. A filtration represents an increasing stream of information.

Definition 19.1.4 (adapted process). *[4, p. 77]Consider a stochastic process* $\{X_t\}_{t\in I}$ *with a filtration* $\{\mathcal{F}_t\}_{t\in I}$ *on its* σ *field. The process is said to be* ***adapted to the filtration*** $\{\mathcal{F}_t\}_{t\in I}$ *if the random variable* X_t *is* $\mathcal{F}_t$ *measurable for all* $t \in I$*, or equivalently,* $\sigma(X_t) \subseteq \mathcal{F}_t$*.*

Remark 19.1.4. [5]

- Examples of 'non-adapted' process. Consider a stochastic process X with $I = \{0,1\}$. Let $\mathcal{F}_0, \mathcal{F}_1$ be σ field generated by X_0, X_1. And $\mathcal{F}_0$ and $\mathcal{F}_1$ are independent to each other, i.e. $\mathcal{F}_0 \nsubseteq \mathcal{F}_1$.
- For a discrete stochastic process $\{X_n\}$, let $F_n = \sigma(X_0, X_1, ..., X_n)$, then $\{X_n\}$ is an adapated process. Here $\sigma(X_0, X_1, ..., X_n)$ is the smallest σ algebra on Ω such that $X_0, X_1, ..., X_n$ is measurable.

Remark 19.1.5.

19.1.3 Natural filtration of a stochastic process

Definition 19.1.5 (natural filtration generated by a stochastic process). *Let* (S, Σ) *be a measurable space. Let* X_t *be a stochastic process such that* $X : I \times \Omega \to S$*, then natural filtration of* $\mathcal{F}$ *with respect to* X *is the filtration* $\{\mathcal{F}_t\}_{t\in I}$ *given by*

$$\mathcal{F}_t = \sigma(X_t^{-1}(A) | s \in I, s \leq t, A \in \Sigma)$$

here σ *is the* σ *field generation operation. Or equivalently, we write*

$$\mathcal{F}_t = \sigma(X_s, s \leq t).$$

Remark 19.1.6.

- In discrete setting, we have $\mathcal{F}_n = \sigma(X_0, X_1, ..., X_n)$.
- **Any stochastic process** X_t **is an adapted process with respect to its natural filtration** $\mathcal{F}_t$ **because** $\sigma(X_t) \subseteq \mathcal{F}_t$.

Remark 19.1.7 (interprete natural filtration). [6, p. 43]

- Let the symbol $\mathcal{F}_t^X$ denotes the σ-algebra (i.e., information) generated by X_t on the interval $[0, t]$, or alternatively 'what has happened to X over the interval $[0, t]$'. Note that $\mathcal{F}_t^X$ is one element in the natural filtration.
- (**interpretation of adaptivity**) Informally, if, based upon observations of the trajectory $\{X(s); 0 \leq s \leq t\}$, it is possible to decide whether a given **event** A has occurred or not, then we write this as $\sigma(A) \in F_t^X$, or say that 'A is F_t^X-measurable'.
- If the value of a given **random variable** Z can be completely determined by given observations of the trajectory $\{X(s); 0 \leq s \leq t\}$, then we also write $\sigma(Z) \in F_t^X$.
- If Y_t is a stochastic process such that we have $\sigma(Y(t)) \in \mathcal{F}_t^X, \forall t \geq 0$, then we say that Y is adapted to the filtration $\{\mathcal{F}_t^X, t \geq 0\}$.

We have the following simple examples:

- If we define the event A by $A = \{X(s) \leq 3.14, \forall s \leq 9\}$, then we have $A \in \mathcal{F}_9^X$.
- For the event $A = \{X(10) > 8\}$, we have $A \in \mathcal{F}_{10}^X$ but not $A \notin \mathcal{F}_9^X$ since it is impossible to decide A has occured or not based on the trajectory of X_t over the interval $[0, 9]$.
- For the random variable Z defined by

$$Z = \int_0^5 X(s) ds,$$

we have $\sigma(Z) \in \mathcal{F}_5^X$.

Example 19.1.1 (Trivial adaptive process: single Bernoulli experiment). Consider a stochastic process $\{X_n\}$ represents a single toss experiment. We then have a trivial adapted process by defining $\mathcal{F}_1 = \mathcal{F}_2 = ... = \mathcal{F}_n = \mathcal{F} = \sigma(X_1)$. For this filtration, the stochastic process $Z_n = \sum_{i=1}^n X_i$ is not adapted to it.

Example 19.1.2 (Infinite coin toss process(infinite Bernoulli experiments)). Consider the probability space for tossing a coin infinitely many time. We can define the sample space as Ω_∞ = the set of infinite sequences of Hs and Ts. A generic element of Ω_∞ will be denoted as $\omega = \omega_1\omega_2....$, where ω_n indicates the result of the nth coin toss.

We can define a stochastic process $\{X_n\}, X_n = f(W_1, W_2, ..., W_n)$, and its filtration $\mathcal{F}_n = \sigma(W_1, W2, ..., W_n)$. Then every X_n is $\mathcal{F}_n$ measurable. A simple event in $\mathcal{F}_n$ is the random experiment value of $W_1, W_2, ..., W_n$. Note that as n increase, $\mathcal{F}_n$ becomes finer and finer, and $\mathcal{F}_n$ can measure any previous $X_m, m < n$.

Remark 19.1.8 (σ algebra for a stochastic process)**.** From 19.1.1, we know that $\mathcal{F}$ is the σ algebra for the set of all possible sample paths. And $\mathcal{F}_t$ can be viewed as the σ algebra for the set of all possible sample paths upto t.

19.1.4 Continuity of sample path

Definition 19.1.6 (continuity of sample path). *A stochastic process with almost all sample paths continuous is called a continuous process. Similarly, a stochastic process is said to be right-continuous if almost all of its sample paths are right-continuous functions.*

Example 19.1.3.

- The Brownian motion is a stochastic process with continuous sample path.
- The white noise process has discontinuous sample path.
- The Poisson process has discontinuous sample path.

Definition 19.1.7 (right-continuous with left limit, cadlag).

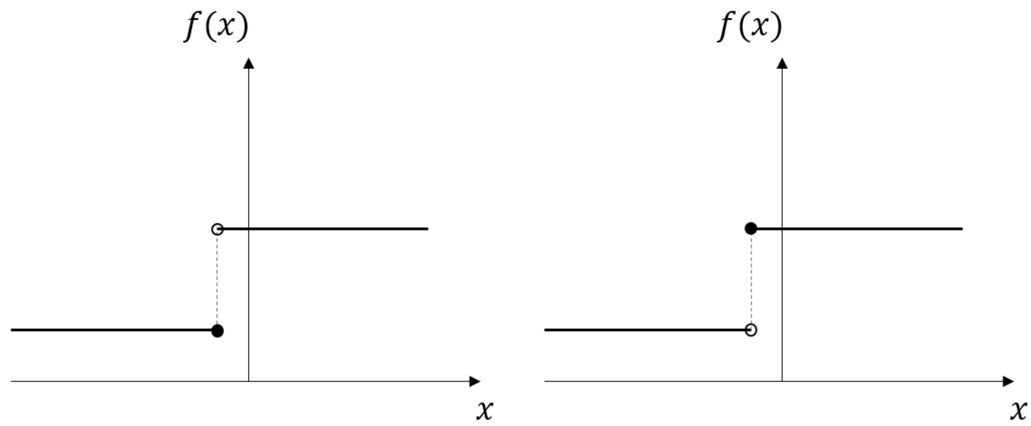
- *A sample path* $X_t : [0, \infty) \to \mathbb{R}$ *is called a right-continuous with left limit if for every* $t \in [0, \infty)$ *if*
  - *the left limit* $\lim_{s \to t^-} X(s)$ *exists;*
  - *the right limit* $\lim_{s \to t^+} X(s)$ *exists and* $\lim_{s \to t^+} X(s) = f(t)$.
- *A stochastic process with almost all sample paths being right-continuous with left limit is called a cadlag stochastic process.*

Example 19.1.4. The sample path of a Poisson process is cadlag.

19.1.5 Predictable process

Definition 19.1.8 (predictable process).

- *Given a probability space* $(\Omega, \mathcal{F}, P)$ *and a filtration* $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$*, a discrete-time stochastic process* $\{X_n\}_{n \in \mathbb{N}}$ *is* ***predictable*** *if* X_{n+1} *is measurable with respect to* $\mathcal{F}_n$ *for each* n.
- *Consider a probability space* $(\Omega, \mathcal{F}, P)$ *and a filtration* $\{\mathcal{F}_t\}_{t \geq 0}$. *Let* $\mathcal{P}$ *denote the predictable/previsible* σ*-algebra, i.e. the* σ*-algebra generated by all left-continuous process adapted to* $\{\mathcal{F}_t\}_{t \geq 0}$. *If a process* X_t *is measurable with respect to* $\mathcal{P}$ *for all* t*, then* X_t *is a predictable process.*



(a) A left continuous function.

(b) A right continuous function.

Figure 19.1.2: An illustration of left and right continuous functions.

Example 19.1.5.

- Every deterministic process is a predictable process.
- Every continuous-time stochasti process with continuous smaple path (e.g., Brownian motion) is adapted to and predictable with respect to its natrual filtration.

Example 19.1.6 (trading strategies as predictable processes).

19.2 Stationary process

19.2.1 Stationarity concepts

Definition 19.2.1 (strictly stationary process). *[7, p. 231][8, p. 30] A random process $\{X_t\}_{t\in T}$ over probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$ is called strictly stationary if for any $t_1, t_2, ..., t_k \in T$ and $A_1, ..., A_k \in \mathcal{B}(\mathbb{R})$ the probabilities*

$$P(X_{t_1+t} \in A_1, ..., X_{t_k+t} \in A_k)$$

do not depend on arbitrary t, where $t \in T$.

Example 19.2.1 (a sequence of iid random variables). A sequence of independent identically distributed random variables is a strictly stationary process.

Example 19.2.2 (A Markov chain starting from stationary distribution is strictly stationary process). [7, p. 231] Consider a finite state, irreducible and aperiodic Markov chain characterized by matrix P. Let the initial state distribution be π_0. If the stationary distribution $\pi = \pi_0$, then the Markov chain P is a strictly stationary process. Note that the stationary distribution will exist for the Markov chain [Theorem 22.4.4]. By iteration, we know that the distribution at every time step is π.

Definition 19.2.2 (weakly stationary process). *[7, p. 209][8, p. 32] A random process $\{X_t\}_{t\in T}$ is called a weakly stationary process if there exist a constant m and $b(t), t \in T$ function, such that*

$$E[X_t] = m, Var[X_{t_1}] = \sigma^2, cov(X_{t_1}, X_{t_2}) = r(t_1 - t_2), \forall t_1 \neq t_2 \in T,$$

where $b(0) = \sigma^2$.

That is, the mean and the covariance structure a weakly stationary process can be fully characterized by a constant mean parameter and a covariance function $r : \mathbb{R} \to \mathbb{R}$.

Lemma 19.2.1 (properties of a covariance function). *[8, p. 35] For a weakly stationary stochastic process, the covariance function $r(t_1 - t_2) \triangleq Cov(X(t_1), X(t_2))$ has the following properties:*

- $r(0) = Var[X(t)] \geq 0, \forall t$
- $Var[X(t+h) \pm X(t)] = E[(X(t+h) \pm X(t))^2] = 2(r(0) \pm r(h))$

- *(even function)*$r(\tau) = r(-\tau)$.
- $|r(\tau)| \leq r(0)$.
- *If* $|r(\tau)| = r(0)$, *for some* $\tau \neq 0$, *then* r *is periodic. In particular,*
  - *If* $r(\tau) = r(0)$, *then* $X(t+\tau) = X(t), \forall t$.
  - *If* $r(\tau) = -r(0)$, *then* $X(t+\tau) = -X(t) = X(t-\tau), \forall t$*(periodicity of* 2π*).*
- *If* $r(\tau)$ *is continuous for* $\tau = 0$, *then* $r(\tau)$ *is continuous everywhere.*

Proof. (1)(2)(3) Straight forward.

(4) Use Cauchy inequality for random variables(Theorem 12.10.4)

$$E[|X(t)-\mu||X(t+\tau)-\mu|] \leq \sqrt{Var[X(t)]Var[X(t+\tau)]} = \sqrt{r(0)^2} = r(0).$$

(5)

(a) If $r(\tau) = r(0)$, then from (2) we have $E[(X(t+\tau) - X(t))^2] = 0$. It can be showed via contradiction that having $E[(X(t+\tau) - X(t))^2] = 0$ implies $X(t+\tau) = X(t)$(that is the two maps are exactly the same). (b) If $r(\tau) = -r(0)$, then from (2) we have $E[(X(t+\tau) + X(t))^2] = 0$. It can be showed via contradiction that having $E[(X(t+\tau) + X(t))^2] = 0$ implies $X(t+\tau) = -X(t)$(that is the two maps are exactly the same). (6) For any t, consider

$$\begin{aligned}(r(t+h) - r(t))^2 &= (Cov(X(0), X(t+h)) - Cov(X(0), X(t)))^2 \\ &= (Cov(X(0), X(t+h) - X(t)))^2 \\ &\leq Var[X(0)]Var[X(t+h) - X(t)] = 2r(0)(r(0) - r(h))\end{aligned}$$

If $h \to 0$, then $r(0) - r(h) \to 0^+$ due to the continuity of $r(\tau)$ at $\tau(0)$,which implies $r(t+h) \to r(t)$(that is, $r(t)$ is continuous for any t). □

Lemma 19.2.2 (a strictly stationary process is a weakly stationary process). *A strictly stationary process* X_t *will be a weakly stationary process.*

Proof. For mean, use translation invariant property of of marginal distribution. For covariance, use translational invariant property of two variable joint distribution. □

Lemma 19.2.3. *A weakly stationary Gaussian process is a strictly stationary Gaussian process.*

Proof. To a process is strictly stationary, we need to show $(X_{t_1+\tau}, X_{t_2+\tau}, ..., X_{t_n+\tau}), \forall \tau \in \mathbb{R}$ has the same distribution as $(X_{t_1+\tau}, X_{t_2+\tau}, ..., X_{t_n+\tau})$. Because the full distribution of a multivariate Gaussian can be constructed from its pair distribution(Lemma 13.1.10), we only need to show that $(X_{t_1+\tau}, X_{t_2+\tau}), \tau \in \mathbb{R}$ has the same distribution as (X_{t_1}, X_{t_2}). From weak stationarity, we know that the mean vector and covariance matrix are the same; that is, their joint distribution are the same. Note that for a Gaussian distribution, mean and covaraince matrix fully determines the joint distribution. □

19.2.2 Random phase and amplitude

Definition 19.2.3 (random harmonic function). *[8, p. 35] Let $A > 0$ and ϕ be independent random variable, with ϕ uniform over $[0, 2\pi]$. The stochastic process $X(t)$ defined by*

$$X(t) = A\cos(2\pi f_0 t + \phi).$$

Lemma 19.2.4 (random harmonic function). *[8, p. 36] The random harmonic function $X(t) = A\cos(2\pi f_0 t + \phi)$ is a* ***stationary process*** *if A and ϕ are independent and ϕ is uniformly distributed in $[0, 2\pi]$. Then*

- $E[X(t)] = 0$
- $Var[X(t)] = \frac{1}{2}E[A^2] = \sigma^2$
- $r(\tau) = \sigma^2\cos(2\pi f_0 \tau)$

Proof. (1)

$$E[X(t)] = E[A\cos(2\pi f_0 t + \phi)] = E[A]E[\cos(2\pi f_0 t + \phi)]$$

$$\begin{aligned} E[\cos(2\pi f_0 t + \phi)] &= \int_0^{2\pi} \cos(2\pi f_0 t + x)\frac{1}{2\pi}dx \\ &= \frac{1}{2\pi}(\sin(2\pi f_0 t + 2\pi) - \sin(2\pi f_0 t + x)) \\ &= 0. \end{aligned}$$

(2)

$$Var[X(t)] = E[X(t)^2] = E[A^2\cos(2\pi f_0 t + \phi)^2] = E[A^2]E[\cos(2\pi f_0 t + \phi)^2].$$

Note that

$$\begin{aligned} E[\cos(2\pi f_0 t + \phi)^2] &= \int_0^{2\pi} \cos(2\pi f_0 t + x)^2\frac{1}{2\pi}dx \\ &= \frac{1}{2\pi}(\pi) \\ &= 1/2. \end{aligned}$$

(3)
$$Cov(X(t), X(s)) = E[X(t)X(s)] = E[A^2]E[\cos(2\pi f_0 t + \phi)\cos(2\pi f_0 s + \phi)].$$
Note that
$$\begin{aligned} E[\cos(2\pi f_0 t + \phi)\cos(2\pi f_0 s + \phi)] &= \frac{1}{2\pi}\int_0^{2\pi}\frac{1}{2}(\cos(2\pi f_0(s+t) + 2x) + \cos(2\pi f_0(s-t)))dx \\ &= \frac{1}{2}\cos(2\pi f_0(s-t)) \end{aligned}$$

□

Lemma 19.2.5 (linear supposition of random harmonic function). *[8, p. 38] The linear supposition of random harmonic functions*
$$X(t) = A_0 + \sum_{k=1}^{n} A_k \cos(2\pi f_k t + \phi_k)$$
with deterministic constants $A_0, A_1, ..., A_k$ and independent uniform random variables $\phi_1, \phi_2, ..., \phi_k$ has the following properties:

- $E[X(t)] = A_0$.
- $r(\tau) = \sigma_0^2 + \sum_{k=1}^n \sigma_k^2 \cos(2\pi f_k \tau)$.

Proof. (1)
$$E[A_0 + \sum_{k=1}^{n} A_k \cos(2\pi f_k t + \phi_k)] = E[A_0] + 0.$$
(2) Note that
$$\begin{aligned} &E[\sum_{k=1}^{n} A_j \cos(2\pi f_k t + \phi_k) \sum_{k=1}^{n} A_k \cos(2\pi f_k (t+\tau) + \phi_k)] \\ &= E[\sum_{j=1}^{n} A_j \cos(2\pi f_j t + \phi_j) \sum_{k=1}^{n} A_k \cos(2\pi f_k (t+\tau) + \phi_k)] \\ &= \sum_{k=1}^{n} E[A_k^2]E[\cos(2\pi f_k t + \phi_k)\cos(2\pi f_k (t+\tau) + \phi_k)] \\ &= \sum_{k=1}^{n} \sigma_k^2 \cos(2\pi f_k \tau) \end{aligned}$$

□

Lemma 19.2.6 (power and average power). *[8, p. 39] Consider the linear supposition of random harmonic functions*

$$X(t) = A_0 + \sum_{k=1}^{n} A_k \cos(2\pi f_k t + \phi_k)$$

with deterministic constants $A_0, A_1, ..., A_k$ and independent uniform random variables $\phi_1, \phi_2, ..., \phi_k$.

Then the average power is given by

$$E_T = \lim_{T\to\infty} \frac{1}{T}\int_0^T X(t)^2 dt = A_0^2 + \frac{1}{2}\sum_{k=1}^{n} A_k^2.$$

Proof. Note that

$$\begin{aligned} E_T &= \frac{1}{T}\int_0^T X(t)^2 dt \\ &= \frac{1}{T}\{TA_0^2 + 2\sum_{k=1}^{n} A_0 A_k \int_0^T \cos(2\pi f_k t + \phi_k)dt \\ &+ \sum_{k=1}^{n} A_k^2 \int_0^T \cos^2(2\pi f_k t + \phi_k)dt \\ &+ 2\sum_{k=1}^{n}\sum_{l=k+1}^{n} A_k A_l \int_0^T \cos(2\pi f_k t + \phi_k)\cos(2\pi f_l t + \phi_l)dt\}. \end{aligned}$$

Further note that:

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \cos(2\pi f_k t + \phi_k)dt = 0$$

since $\int_0^T \cos(2\pi f_k t + \phi_k)dt$ is bounded.

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \cos(2\pi f_k t + \phi_k)\cos(2\pi f_l t + \phi_l)dt = 0$$

since $\int_0^T \cos(2\pi f_k t + \phi_k)\cos(2\pi f_l t + \phi_l)dt$ is bounded.

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \cos^2(2\pi f_k t + \phi_k)dt = \frac{1}{2}$$

since $\int_0^T \cos^2(2\pi f_k t + \phi_k)dt = \frac{T}{2}$ is bounded.

□

Remark 19.2.1 (average power is independent of sample path). Note that the integral is independent of $\phi_1, ..., \phi_k$. That is, no matter what values $\phi_1, ..., \phi_k$ are taking in a sample path, we get the same average power.

19.3 Gaussian process and Finite dimension distributions

19.3.1 One-dimensional Gaussian process

19.3.1.1 *Definitions and properties*

Definition 19.3.1 (One-dimensional Gaussian process). *A stochastic process $\{X_t\}_{t\in T}$ is Gaussian process if for any $t_1, t_2, ..., t_n \in T$, the joint distribution on $(X_{t_1}, ..., X_{t_n})$ is Gaussian, i.e.,*

$$p(x_1, x_2, ..., x_n) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|det\Sigma|^{1/2}} \exp(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)).$$

Definition 19.3.2 (One-dimensional Gaussian process, alternative). *[9, p. 19] A stochastic process $\{X_t\}_{t\in T}$ is Gaussian process if every linear combination*

$$S = \sum_k a_k x(t_k), t_k \in T, a_k \in \mathbb{R}$$

has a Gaussian distribution.

Remark 19.3.1 (equivalence of two definitions)**.**

- The first definition can imply the second by affine transformation Theorem 15.1.1.
- The second can imply the first by using the multivariate Gaussian definition via linear combination(Lemma 13.1.8).

Lemma 19.3.1 (affine transformation of a Gaussian process is a Gaussian process). *Let X_t be a Gaussian process, then $aX_t + b, a, b \in \mathbb{R}$ is also a Gaussian process.*

Proof. Directly from definitions. □

Note 19.3.1 (Gaussian processes can be stationary or non-stationary)**.** A Gaussian process can be stationary(white noise process) or non-stationary(Wiener process).

19.3.1.2 *Stationarity*

Lemma 19.3.2. *If $X(t)$ is a stationary Gaussian process with mean m and covariance function $r(\tau)$. Then*

- *For all t, $X(t) \sim N(m, r(0)^2)$.*
- *For all t_1, t_2, $(X(t_1), X(t_2)) \sim MN(\mu, \Sigma)$, where*

$$\mu = \begin{bmatrix} m \\ m \end{bmatrix}, \Sigma = \begin{bmatrix} r(0) & r(t_1 - t_2) \\ r(t_1 - t_2) & r(0) \end{bmatrix}$$

Proof. Straight forward. □

19.3.1.3 *Examples*

Example 19.3.1 (white noise process). A white noise process W_t is a Gaussian process with zero mean and $cov(W_t, W_s) = \sigma^2\delta(s-t)$.

Example 19.3.2 (a discrete random walk is not a Gaussian process). A random walk B_n is not a Gaussian process. For example, B_1 is a Bernoulli distribution, not a Gaussian.

Example 19.3.3 (Ornstein-Uhlenbeck process is stationary Gaussian process in the long run). An OU process is a Gaussian process. As $t \to \infty$, the OU process becomes a stationary Gaussian process(Lemma 20.3.1).

Example 19.3.4 (a Wiener process(Brownian motion) is a Gaussian process). From Lemma 19.5.1, a Wiener process is the integral of a white noise Gaussian process. It is not stationary, but it has stationary increments.

Example 19.3.5 (a geometric Brownian motion is not a Gaussian process). Let X_t be a geometric Brownian motion process, then X_t is not Gaussian, thus not a Gaussian process.

Example 19.3.6 (a stable AR(1) process). A stable AR(1) process of X_k can be written as

$$X_k = \sum_{i=0}^{\infty} \beta^i W_{k-i},$$

where $W_k = w(t_k)$ is the discrete sampling of wiener process $w(t)$.

Because any linear combination of samples of a Gaussian process $w(t)$ is a normal random variable, X_k has a normal distribution.

19.3.2 finite dimensional distribution

Definition 19.3.3. *The finite dimensional distribution of a stochastic process X is the joint distribution of $X_{t_1}, X_{t_2}, ..., X_{t_k}$, where $t_1, t_2, ..., t_k \in \mathcal{I}$, we represent the probability measure as*

$$\mu_{t_1,t_2,...,t_k}(F_1 \times F_2 \times ...F_k) = P(X_{t_1} \in F_1 ... X_{t_k} \in F_k)$$

where F_i are measurable events/subsets in $\mathbb{R}$.

Remark 19.3.2 (purpose). We usually characterize a stochastic process by studying the joint distribution of a finite set of random variables in the stochastic process.

consistence of finite-dimensional distributions
Given a set of finite-dimensional distributions on different index set, to make sure that these fdd are derived from the same stochastic process, **we require these fdd to be consistent based on two criteria:**[10, lec 4]

- Invariant under permutations, i.e., two probability measure on two different set of $X_1, X_2, ..., X_k$ and $X_{\pi(1)}, .., X_{\pi(k)}$, then the measure should be the same given the same measurable subset.
- Marginal distribution are consistent,i.e., marginalizing different probability measure to the marginal measure to the same subset of random variables should be the same.

Theorem 19.3.1 (Kolmogorov Extension theorem). *If the family of measures $\{\mu_{t_1,...,t_k}\}$ satisfies the consistent condition, then there exists a stochastic process with the corresponding finite-dimensional distribution.*

19.3.3 Gaussian process generated by Brownian motin

Lemma 19.3.3 (Gaussian process stochastic differential equation). *A stochastic process X_t governed by*

$$dX_t = a(t)dt + b(t)dW_t,$$

where W_t is a Wiener process, is a Gaussian process.

Proof. It can be showed that $X(t) \sim N(\mu(t), \int_0^t b(s)^2 ds)$(Lemma 20.2.9). Also note that the increment is independent and Gaussian; that is $X(t_1) - X(t_2)$ is independent of $X(t_2) - X(t_3), t_1 > t_2 > t_3$. Therefore, the random vector $(X(t_1), X(t_2), ..., X(t_n))$ is multivariate normal since it can be constructed by affine transformation of $(X(t_1) - X(t_2), X(t_2) - X(t_3), ..., X(t_n)$(Theorem 15.1.1). □

Theorem 19.3.2 (linear combination of multiple Brownian-motion-generated Gaussian processes is a Gaussian process). *Consider N stochastic processes generated by M Brownian motions, given by*

$$dX_i(t) = \mu_i(t)dt + \sum_{j=1}^{M} \sigma_{ij}(t)dW_j,$$

where $W_1, W_2, ..., W_M$ are independent Brownian motion, $\mu_i(t), \sigma_{ij}(t)$ are state-independent deterministic function of t.

Then

- *the joint distribution of $X_1, X_2, ..., X_N$ is multivariate Gaussian.*
- *for any linear combination of $X_1(t), X_2(t), ..., X_N(t)$, given by*

$$Y(t) = \sum_{i=1}^{M} a_i X_i(t), a_i \in \mathbb{R},$$

$Y(t)$ is a Gaussian process.

Proof. (1) We only show a zero drift 2 by 2 case. Consider

$$X_1 = \int_0^t \sigma_{11}(s)dW_1 + \int_0^t \sigma_{12}(s)dW_2, X_2 = \int_0^t \sigma_{21}(s)dW_1 + \int_0^t \sigma_{22}(s)dW_2.$$

Denote

$$A = \int_0^t (\lambda_1\sigma_{11}(s) + \lambda_2\sigma_{21}(s))dW_1(s)$$

$$B = \int_0^t (\lambda_1\sigma_{12}(s) + \lambda_2\sigma_{22}(s))dW_2(s).$$

And we can see immediately that

$$E[A] = E[B] = 0, Var[A] = \int_0^t (\lambda_1\sigma_{11}(s) + \lambda_2\sigma_{21}(s))^2 ds, Var[B] = \int_0^t (\lambda_1\sigma_{12}(s) + \lambda_2\sigma_{22}(s))^2 ds.$$

More important, $A+B$ is a Gaussian random variable.

To show (X_1, X_2) is joint Gaussian, we can check its mgf, given by

$$\begin{aligned}\phi(\lambda_1, \lambda_2) &= E[\exp(\lambda_1 X_1 + \lambda_2 X_x)] \\ &= E[\exp(A+B)] \\ &= E[\exp(E[A+B] + \frac{1}{2}Var[A+B])] \\ &= E[\exp(\frac{1}{2}(Var[A] + Var[B]))] \\ &= E[\exp(\frac{1}{2}(\int_0^t (\lambda_1\sigma_{11}(s) + \lambda_2\sigma_{21}(s))^2 ds + \int_0^t (\lambda_1\sigma_{12}(s) + \lambda_2\sigma_{22}(s))^2 ds))]\end{aligned}$$

where we eventually will get a quadratic form of λ_1 and λ_2. Then using Lemma 13.1.7, we can show (X_1, X_2) are joint normal.

We can similarly prove cases containing multiple variables and drifting terms.

(2) Directly use affine transformation of multivariate Gaussian vector. □

Corollary 19.3.2.1. *Consider N stochastic processes generated by M Brownian motions, given by*

$$dX_i(t) = \mu_i(t)dt + \sigma_i(t)dW_j,$$

*where $W_1, W_2, ..., W_M$ are **correlated Brownian motions** such that $E[dW_i dW_j] = \rho_{ij}dt$, $\mu_i(t), \sigma_{ij}(t)$ are state-independent deterministic function of t.*

Then

- *the joint distribution of $X_1, X_2, ..., X_N$ is multivariate Gaussian.*
- *for any linear combination of $X_1(t), X_2(t), ..., X_N(t)$, given by*

$$Y(t) = \sum_{i=1}^{M} a_i X_i(t), a_i \in \mathbb{R},$$

$Y(t)$ is a Gaussian process.

Proof. use the fact that the two forms are equivalent(20.2.1). □

Remark 19.3.3 (Caution!). Note that if $X_1, X_2, ..., X_N$ are more general Gaussian processes not generated by Brownian motion, then Y is not necessarily Gaussian, since $X_1, X_2, ..., X_N$ are not necessarily joint normal.

19.4 Markov process

Definition 19.4.1 (Markov process). *[3]Let $(\Omega, \mathcal{F}, P)$ be a probability space, let T be a fixed positive number, and let $\mathcal{F}(t)$ be a filtration of $\mathcal{F}$. Consider an adapated stochastic process $X(t)$, assume that for all $0 \leq s \leq t$ and for every nonnegative, Borel-measurable function f, there is another Borel-measurable function g such that*

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

then we say $X(t)$ is a Markov process.

Definition 19.4.2 (Alternative definition). *[11]Let $(\Omega, \mathcal{F}, P)$ be a probability space, let T be a fixed positive number, and let $\mathcal{F}(t)$ be a filtration of $\mathcal{F}$. Consider an adapated stochastic process $X(t) : \Omega \to S$,for all $0 \leq s \leq t$, for each $A \in \mathcal{S}$, then*

$$P(X_t \in A|\mathcal{F}(s)) = P(X_t \in A|X_s)$$

A Markov process is stochastic process satisfies above Markov property with respect to its natural filtration.

Remark 19.4.1.

- The above two definition emphasizes that in a Markov process, only *the most immediate history* is useful when we make predictions on the future based on the past history.
- For discrete-time Markov chains, we have alternative definition as

$$P(X_n = x_n|X_{n-1} = x_{n-1}, ..., X_0 = x_0) = P(X_n = x_n|X_{n-1} = x_{n-1}).$$

- A martingale is not necessarily a Markov process. See link.
- Examples of non-Markovian include long memory auto-regressive processes.

Remark 19.4.2 (Gaussian processes, stationary processes and Markov processes are different characterization of stochastic process.)**.** We have:

- A Gaussian process is not necessarily a Markov process.
- A stationary is not necessarily a Markov process.

19.5 Wiener process (Brownian motion)

19.5.1 Basics

Definition 19.5.1 (Brownian motion). *A stochastic process $W(t)$ is a called a Wiener process or a Brownian motion if:*

- $W(0) = 0$
- *each sample path is continuous almost surely*
- $W(t) \sim N(0, t)$
- *for all* $0 < t_1 < t_2 < ...$ *the random variables:*

$$W(t_1), W(t_2) - W(t_1), ...$$

are independent and have $W(t_2) - W(t_1) \sim N(0, t_2 - t_1)$

Lemma 19.5.1 (basic properties of one-D Brownian motion). *Let $W(t)$ be a Brownian motion, then we have:*

- $E[W(t)] = 0;$
- $Var[W(t)] = t;$
- $cov(W(s), W(t)) = \min(s, t);$
- $$\rho(t, s) = \sqrt{1 - \frac{\tau}{t}}, t \geq s, \tau = t - s$$

  ; therefore, $W(t)$ is a nonstationary Gaussian process.

Proof. (1)(2) directly from definition. (3) Let $s < t$, and $cov(W(s), W(t)) = cov(W(s), W(t) - W(s) + W(s)) = cov(W(s), W(s)) = \min(s, t)$. (4) The joint distribution of $W(s), W(t), t > s$ can be constructed from joint distribution of $W(s), W(t) - W(s)$, which are Multivariate Gaussian,via affine transformation. We can similarly extend to arbitrary joint distributions. It is nonstationary because the autocorrelation function depends both on t and τ. □

Definition 19.5.2 (multi-dimensional independent Brownian motion). *A stochastic process $W(t) = (W_1(t), W_2(t), ..., W_n(t))$ is a called a n-dimensional Wiener process or Brownian motion if:*

- *Each $W_i(t)$ is a Wiener process.*
- *If $i \neq j$, then $W_i(t)$ and $W_j(t)$ are independent.*

Lemma 19.5.2 (basic properties of multidimensional independent Brownian motion). *Consider the vector $W(t) = (W_1(t), W_2(t), ..., W_m(t))^T$ denotes an $m-$dimensional independent Brownian motion/Wiener process, each component is uncorrelated with other components for all values of time t. We have*

$$cov(W_i(s)W_j(t)) = \delta_{ij}\min(s,t)$$

and

$$cov(dw_i(t_i)dw_j(t_j)) = \sigma_i^2\delta_{ij}dt_j = \sigma_i^2\delta(t_i - t_j)dt_i dt_j\delta_{ij}$$

where $dw(t) = w(t+dt) - w(t)$,Dirac delta function can be viewed as having a value of $1/dt$.

Definition 19.5.3 (multi-dimensional correlated Brownian motion). *A stochastic process $W(t) = (W_1(t), W_2(t), ..., W_n(t))$ is a called a n-dimensional Wiener process or Brownian motion with constant instantaneous correlation matrix ρ if:*

- *Each $W_i(t)$ is a Wiener process.*
- $$cov(W_i(s)W_j(t)) = \rho_{ij}\min(s,t).$$
  or in matrix form
  $$Cov(W(s), W(t)) = \rho\min(s,t).$$

Remark 19.5.1 (interpreting correlation Brownain motion). Let $X(t) = (X_1(t), X_2(t), ..., X_n(t))$ be a n-dimensional Brownian motion with constant instantaneous correlation matrix ρ. Then we can view $X(t)$ as the solution to the following SDE:

$$dX_i(t) = dW_i(t), X_i(0) = 0, i = 1, 2, ..., n,$$

where W_is are Brownian motions and

$$E[dW_i(t)dW_j(s)] = \rho_{ij}dt\delta(s-t).$$

19.5.2 Filtration for Brownian motion

Definition 19.5.4 (Brownian motion filtration). *[3][10]Let W_t be a Brownian motion, the filtration for the Brownian motion can be defined as $\mathcal{F}_t = \sigma(\{\mathcal{F}_s\}_{s\leq t})$.*

Remark 19.5.2.

- This filtration is also the natural filtration.
- W_t is $\mathcal{F}_t$ adapted, but W_{2t} is not $\mathcal{F}_t$ adapted.
- Any stochastic process S_t as solution of Ito SDE is also $\mathcal{F}_t$ adapted, which, intuitively,means that given the values of $\{W_s\}_{s\leq t}$, S_t is known.

Remark 19.5.3 (stochastic process adapted to Brownian filtration). Let $B_t, t \geq 0$ be a Brownian motion and $\{\mathcal{F}_t\}$ be the Brownian filtration.

- The stochastic process $X_t = f(t, B_t), t \geq 0$, where f is a function of two variables, are adapted to the Brownian filtration.
  - $X_t = B_t$, $X_t = B_t^2 - t$
  - $X_t = \max_{0\leq s\leq t} B_s$ and $X_t = \max_{0\leq s\leq t} B_s^2$
- Examples that are not adapted to the Brownian motion filtration are: $X_t = B_{t+1}$ and $X_t = B_t + B_T, T > 0$.

19.5.3 Quadratic variation

Remark 19.5.4 (purpose). We introduce the concept of **quadratic variation** to measure how jagged the paths of a Brownian motion are.

Definition 19.5.5 (quadratic variation). *[12, p. 101] The quadratic variation of a function $f : [0, T] \to \mathbb{R}$is defined to be*

$$Q^* = \lim_{l(\Delta)\to 0} \sum_{i=0}^{n-1} (f(t_{i+1} - f_i))^2$$

where Δ is a partition of the interval $[0, T]$ with $0 = t_0 < t_1 ... < t_n = T$, and $l(\Delta) = \max_i(t_{i+1} - t_i)$.

Theorem 19.5.1 (continuously differentiable functions have zero quadratic variations). *[12, p. 101]Given a continuously differentiable function on a closed interval, then its quadratic variation is zero.*

Proof. Using the mean value theorem, $f(t_{i+1}) - f(t_i) = f'(x)(t_{i+1} - t_i)$ for some $x \in (t_i, t_{i+1})$. Because $|f'(x)| \leq M$, then

$$(f(t_{i+1}) - f(t_i))^2 \leq M^2(t_{i+1} - t_i)^2$$

As $l(\Delta) \to 0$, we have $Q^* = 0$. □

Theorem 19.5.2 (Brownian motion quadratic variation). *[12, p. 102] The Brownian motion W on the interval $[0, T]$ has quadratic variation of T in the sense of convergence in mean square.[3]*

Proof. We first prove that

$$EQ(\Delta) = E\sum_{i=0}^{n-1}(W_{i+1} - W_i))^2 = \sum_{i=0}^{n-1}(t_{i+1} - t_i) = T$$

Then we can show $VarQ(\Delta) = 0$, as $l(\Delta) \to 0$. □

Lemma 19.5.3. *Let $W(t)$ be a Brownian motion, then*

$$dW(t)dW(t) = dt, dt \to 0$$

by which we mean

$$E[W(t + dt) - W(t))(W(t + dt) - W(t))] = dt$$

and

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

(that is, the variance will vanish as $dt \to 0$).

Proof. Let $X = dW(t) = W(t + dt) - W(t)$. Then $X \sim N(0, dt)$. Therefore,

$$E[X^2] = dt, Var[X^2] = E[X^4] - E[X^2]^2 = 2(dt)^2,$$

where we use the moment property of Gaussian random variable(subsection 13.1.2). Note that $dW(t)dW(t)$ is just a random variable with mean dt, and variance approaches 0 . □

Remark 19.5.5 (implications).

- Here the conclusion holds in statistical sense, not every sample path holds.

19.5.4 Symmetries and scaling laws

Lemma 19.5.4. *Let $W(t)$ be a standard Brownian motion, then the following are also Brownian motions:*

1. $-W(t)$
2. $W(t+s) - W(s)$
3. $aW(t/a^2)$
4. $tW(1/t)$

Proof. (3) $Var(a(W(s/a^2) - W(t/a^2))) = a^2(t-s)/a^2 = (t-s)$; (4)Assume $t > s$, then $Var(tW(1/t) - sW(1/s))) = Var((t-s)W(1/t) + s(W(1/t) - W(1/s)))) = (t-s)^2/t - s^2/t + s = t - s$. □

19.5.5 Non-differentiability and unbounded variation of path

Theorem 19.5.3. *The Brownian path $W(t)$ is almost surely not differentiable at $t \geq 0$.*

Proof. (informal) consider the differential at $t = 0 : \lim_{t\to 0} W(t)/t = \lim_{s\to\infty} sW(1/s) = \lim_{t\to\infty} \tilde{W}_t = \infty$, where $sW(1/s)$ is another Brownian motion. (For formal proof for non-differentiability anywhere, check [10]). □

Theorem 19.5.4. *[4, p. 189] For almost all Brownian sample path,*

$$\sup_{\tau} \sum_{i=1}^{n} \left| B_{t_i}(\omega) - B_{t_{i-1}}(\omega) \right| = \infty$$

where the supremum is taken over all possible partitions

Remark 19.5.6. Here use almost all is because there is some path, e.g. a path that $B_t(\omega) = const$, that variation will be zero; however, such path has zero probability measure.

19.5.6 The reflection principle

19.5.6.1 *Driftless case*

Lemma 19.5.5 (reflection principle). *[12, p. 208] Let W_t be a Brownian motion. Let m_T denote the minimum value of W_t over the interval $[0,T]$(the minimum value might occur at any time between $[0,T]$). Then*

$$P(W_T \geq x, m_T \leq y) = P(W_T \leq 2y - x),$$

where $x \geq y$ and $y < 0$. Moreover,

$$P(W_T \geq x, m_T \geq y) = P(W_T \geq x) - P(W_T \leq 2y - x)$$

Proof. Consider all trajectories hitting y at some time $\tau \in [0,T]$ and finally reaching $[x, x+dx]$. There are same number of trajectories that hit y at some time $\tau \in [0,T]$ and finally reaching $[2y-x, 2y-x+dx]$, that is

$$P(W_T \geq x, m_T \leq y) = P(W_T \leq 2y - x, m_T \leq y).$$

Note that $W_T \leq 2y - x \implies W_T \leq y$ since $x \geq y$. Therefore,

$$P(W_T \geq x, m_T \leq y) = P(W_T \leq 2y - x).$$

□

Remark 19.5.7 (interpretation). Let y be a barrier level, then

- $P(W_T \geq x, m_T \leq y)$ represents the probability that a random walker hitting the barrier and finally reaching above x.
- $P(W_T \geq x, m_T \leq y)$ represents the probability that a random walker **successfully avoid the barrier** and finally reaching above x.

Lemma 19.5.6 (path excursion distribution). *Let W_t be a driftless Brownian motion. Let m_T denote the minimum value of W_t over the interval $[0,T]$(the minimum value might occur at any time between $[0,T]$). Then*

$$P(m_T \leq y) = 2P(W_T \leq y) = 2N(\frac{y}{\sigma\sqrt{T}}), y \leq 0,$$

$$P(m_T \geq y) = 1 - 2N(\frac{y}{\sigma\sqrt{T}})$$

where W_T is zero mean Gaussian with variance $\sigma^2 T$. In particular, if $T \to \infty$, the $P(m_T \leq y) \to 1$; that is, the Brownian motion will hit any level y with probability 1.

Proof. Use reflection principle(Lemma 19.5.5), we have

$$\begin{aligned} P(m_T \leq y) &= P(m_T \leq y, W_T \leq y) + P(m_T \leq y, W_T \geq y) \\ &= P(m_T \leq y, W_T \leq y) + P(m_T \leq y, W_T \leq y) = 2P(m_T \leq y, W_T \leq y) = 2P(W_T \leq y). \end{aligned}$$

□

Remark 19.5.8 (interpretation & maximum excursion).

- Given a time T, this lemma gives the probability distribution of the excursion of a trajectory during time T.
- It is not possible to know exactly maximum excursion for all possible trajectories. We only know that the larger the excursion, the smaller the probability.

19.5.6.2 *Drifting case*

Lemma 19.5.7 (reflection principle with drift). *[12, p. 213] Let stochastic process Z_t be governed by*

$$dZ_t = vdt + \sigma dW_t,$$

where W_t is the Brownian motion. Let m_t denotes the minimum value of Z_t up to time t. Then, for $y < 0$ and $x > y$, we have

$$P(Z_t \geq x, m_t \leq y) = e^{2vy/\sigma^2} P(Z_t \leq 2y - x + 2vt)$$

Lemma 19.5.8 (path excursion with drift). *[12, p. 213] Let stochastic process Z_t be governed by*

$$dZ_t = vdt + \sigma dW_t,$$

where W_t is the Brownian motion. Let m_t denotes the minimum value of Z_t up to time t. Then, for $y < 0$, and we have

$$P(m_t \leq y) = P(Z_t \leq y) + P(m_t \leq y, Z_t \geq y) = P(Z_t \leq y) + e^{2vy/\sigma^2} P(Z_t \leq 2y - x + 2vt).$$

Proof. Use the fact that

$$P(m_t \leq y) = P(m_t \leq y, Z_t \leq y) + P(m_t \leq y, Z_t \geq y)$$

and note $P(m_t \leq y, Z_t \leq y) = P(Z_t \leq y)$ □

19.5.7 Asymptotic behaviors

Theorem 19.5.5 (law of iterated logarithms). *As $t \to \infty$, we have with probability 1 (i.e. almost surely):*

- $\lim_{t\to\infty} W_t/t = 0$

- $\limsup_{t\to\infty} W_t/\sqrt{t} = \infty$
- $\limsup_{t\to\infty} W_t/\sqrt{2t\log(\log t)} = 1$
- $\liminf_{t\to\infty} W_t/\sqrt{2t\log(\log t)} = 1$

Proof. See [10] for proofs. □

Corollary 19.5.5.1 (unboundedness of Brownian motion). *With probability 1 (i.e. almost surely)*

$$\limsup_t |W_t| = \infty$$

Proof. Use contradiction. If it does not hold, then the law of iterated logarithm cannot hold. □

Corollary 19.5.5.2 (first passage time). *Define $T_a = \inf\{t : W_n > a\}$. $T_a <= \infty$ almost surely (but the mean first passage time will be infinite).*

Remark 19.5.9. Lemma 19.8.5 also shows that the hitting probability is 1 given infinite amount of time.

19.5.8 Levy characterization of Brownian motion

Lemma 19.5.9. *Let $M(t), t \geq 0$ be a stochastic process adapted to a filtration $\{\mathcal{F}_t\}$. Assume that*

- *$M(t), t \geq 0$, be a martingale with respect to a filtration $\mathcal{F}_t$.*
- *$M(0) = 0$, $M(t)$ has continuous paths.*
- *quadratic variation $[M, M](t) = t, \forall t \geq 0$, or $dM_t dM_t = dt$*

Then for all $0 \leq s < t$ and a C^2 function f, we have

$$E[f(X_t)|\mathcal{F}_s] = X_s + \frac{1}{2}\int_s^t E[f''(X_u)|\mathcal{F}_u]$$

Theorem 19.5.6 (Levy characterization, one dimension). *[13, p. 87][3, p. 168] Let $M(t), t \geq 0$ be a stochastic process adapted to a filtration $\{\mathcal{F}_t\}$. Assume that*

- *$M(t), t \geq 0$, be a martingale with respect to a filtration $\mathcal{F}_t$.*
- *$M(0) = 0$, $M(t)$ has continuous paths.*

- *quadratic variation* $[M,M](t) = t, \forall t \geq 0$, *or* $dM_t dM_t = dt$

Then $M(t)$ *is a Brownian motion.*

Proof. (idea) Calculate the moment generating function and show $M(t) - M(s) \sim N(0, t - s)$.

Consider the function $\exp(\lambda M_t)$. We have (from Ito lemma)

$$d\exp(\lambda M_t) = \exp(\lambda M_t)\lambda dM_t + \frac{1}{2}\exp(\lambda M_t)\lambda^2 dt,$$

where we use the assumption $dM_t dM_t = dt$.

Therefore,

$$\exp(\lambda M_T) = 1 + \int_0^T \exp(\lambda M_t) dM_t + \frac{1}{2}\lambda^2 \int_0^T \exp(\lambda M_t) dt.$$

Take expectation, we have

$$E[\exp(\lambda M_T)] = 1 + \frac{1}{2}\lambda^2 \int_0^T E[\exp(\lambda M_t)] dt,$$

where we use the fact that $E[\int_0^T \exp(\lambda M_t) dM_t] = 0$ from Lemma 20.1.1.

Set $g(T) = E[\exp(\lambda M_T)]$, we have $dg = \frac{1}{2}\lambda g(t) dt$, therefore, $g(t) = \exp(\frac{1}{2}\lambda^2 t), g(0) = 1$.

That is, $M_t \sim N(0, t)$.

Similarly, we can show $M_t - M_s \sim N(0, t - s)$. □

Remark 19.5.10 (implication).

- **continuous path requirement is essential**, because Brownian motion has continuous paths.
- Given a stochastic process $X(t)$, Levy characterization enables to test whether $X(t)$ is a Brownian motion.
- Normality can be resulted even no Gaussian distribution is explicitly involved.

Theorem 19.5.7 (Levy characterization, multiple dimension). *[3, p. 168] Let* $M_1(t), ..., M_n(t), t \geq 0$ *be stochastic processes. Assume that*

- $M_i(t), t \geq 0, \forall i$, *be a martingale with respect to a filtration* $\mathcal{F}_t$.
- $M_i(0) = 0$, $M_i(t)$ *has continuous paths for* i.
- *quadratic variation* $[M_i, M_j](t) = t\delta_{ij}, \forall t \geq 0$, *or* $dM_{it} dM_{jt} = dt\delta_{ij}$

Then $[M_1(t), ..., M_n(t)]$ *is a multi-dimensional Brownian motion.*

19.5.9 Discrete-time approximations

Lemma 19.5.10 (discrete-time approximation of white noise). *Consider a white noise* $w(t)$ *satisfying*

$$E[w(t)] = 0, E[w(t)w(\tau)] = \sigma^2\delta(t-\tau)$$

Then its discrete-time approximation white noise process $\{w_1, w_2, ...,\}$ *is given as*

$$E[w_i] = 0, E[w_i w_j] = \frac{1}{\Delta t}\sigma^2\delta_{ij}$$

where w_i *approximate the* $w(t), t \in [t_0 + k\Delta t, t_0 + (k+1)\Delta t]$. *Note that* $\delta(x)$ *is the Dirac delta function, whereas* δ_{ij} *is the Kronecker delta function.*

Moreover, the random walk

$$S_N = \sum_{i=1}^{N} w_i,$$

where $N = \frac{T}{\Delta t}$, *has the distribution of* $N(0, T\sigma^2)$, *which is the same as the Brownian motion distribution at time* T, *given as* $B(T) \sim N(0, \sigma^2 T)$.

Proof. Note that as $\Delta t \to 0$, we recover the covariance for the while noise process. For the distribution S_N, use $N(0, N\frac{1}{\Delta t}\sigma^2) = N(0, T\sigma^2)$ and central limit theorem. □

Remark 19.5.11 (implications). The approximation scheme is important in simulating stochastic differential equations.

19.6 Poisson process

19.6.1 Basics

Definition 19.6.1 (Poisson process). *Let $\lambda > 0$ be fixed. The stochastic process $\{N(t), t \in [0, \infty)\}$ is called a Poisson process with rates λ if all the following conditions hold:*

- $N(0) = 0$.
- $N(t)$ *has independent increments.*
- *The number of arrivals in any interval of length* $N(t_2) - N(t_1) \sim Poisson(\lambda(t_2 - t_1))$.

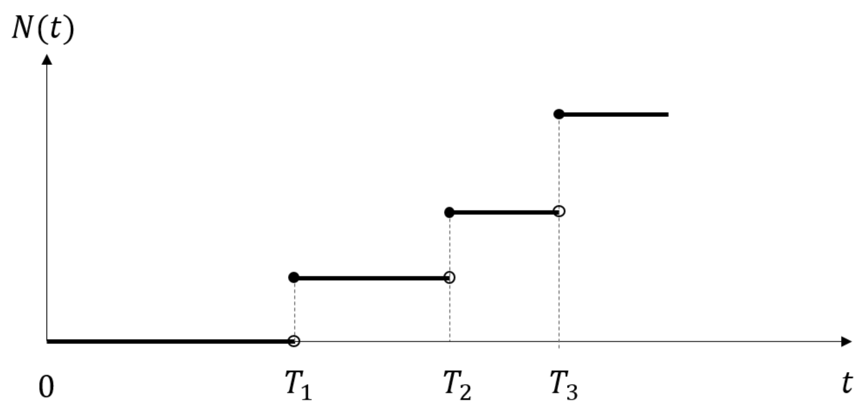


Figure 19.6.1: A typical realized trajectory from the Poisson process with jumps at T_1, T_2, and T_3.

Lemma 19.6.1 (basic properties of Poisson process). *Let $N(t)$ be a Poisson process with rate λ, then:*

- $N(t) \sim Poission(\lambda t)$, *that is*
$$P(N(t) = k) = \frac{e^{\lambda t}(\lambda t)^k}{k!}$$
- $N(t_2) - N(t_1) = N(t_2 - t_1) \sim Poisson(\lambda(t_2 - t_1))$
- $$E[N(t)] = \lambda t, Var[N(t)] = \lambda t, M_{N(t)}(s) = \exp(\lambda t(e^s - 1))$$

- *Jump probability within* $[t, t+\Delta t]$*: let* $\Delta N = N(t+\Delta t) - N(t)$*, we have*

$$Pr(\Delta N = n) = \frac{(\lambda \Delta t)^n}{n!}\exp(-\lambda \Delta t) = \frac{(\lambda \Delta t)^n}{n!}(1 - \lambda \Delta t + \frac{(\lambda \Delta t)^2}{2} + \cdots),$$

or explicitly

$$Pr(\Delta N = n) = \begin{cases} 1 - \lambda \Delta t + O((\Delta t)^2, n = 0 \\ \lambda \Delta t + O((\Delta t)^2, n = 1 \\ O((\Delta t)^2, n \geq 2 \end{cases}.$$

Proof. Directly from definition and the sum property of independent Poisson distribution(Lemma 13.1.21)and basic property of Poisson distribution(Lemma 13.1.20). □

Lemma 19.6.2 (additivity of Poisson process). *Let* $N_1(t)$ *and* $N_2(t)$ *be independent Poisson processes with rate* λ_1 *and* λ_2*, then* $N(t) = N_1(t) + N_2(t)$ *is a Poisson process with rate* $\lambda_1 + \lambda_2$*.*

Proof. Use the moment generating function for $N_1(t)$ and $N_2(t)$.(Lemma 13.1.21). □

19.6.2 Arrival and Inter-arrival Times

Lemma 19.6.3 (waiting time distribution). *Let* $N(t)$ *be a Poisson process with rate* λ*. Let* X_1 *be the time of the first arrival. Then*

$$P(X_1 > t) = \exp(-\lambda t), f_{X_1}(t) = \lambda \exp(-\lambda t)$$

Similarly, let X_n *be the waiting time between the arrival of* n *after the* $n-1$ *arrival, then*

$$P(X_n > t) = \exp(-\lambda t)$$

Proof. (1)From the definition of Poisson process, the $N(t) - N(0) \sim Poisson(\lambda t)$. Then

$$P(X_1 > t) = P(N(t) - N(0) = 0) = (\lambda t)^0 e^{-\lambda t}/0! = e^{-\lambda t}$$

(2) Using the independent increment property of Poisson process. □

Remark 19.6.1. Note that the waiting time distribution is an exponential distribution with parameter λ, whose mean is $1/\lambda$.

Lemma 19.6.4 (Arrival times for Poisson processes). *If $N(t)$ is a Poisson process with rate λ, then the arrival time $T_1, T_2, ...$ have $T_n \sim Gamma(n, \lambda)$ distribution:*

$$f_{T_n}(t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}$$

Moreover, we have $E[T_n] = n/\lambda, Var[T_n] = n/\lambda^2$.

Proof. Let random variables $X_1, X_2, ...$ be the interarrival time, then

$$\begin{aligned} T_1 &= X_1 \\ T_2 &= X_1 + X_2 \\ T_2 &= X_1 + X_2 + X_3 \\ &\cdots\cdots \end{aligned}$$

Since X_i has exponential distribution(which is $Gamma(1, \lambda)$), the T_n will be $Gamma(n, \lambda)$ distribution(which can be showed that the nth power of mgf of exponential function equal to the mgf of Gamma distribution.) Also see property of Gamma distribution(Theorem 13.1.3.) □

Methodology 19.6.1 (Simulating a Poisson process). *We first generate iid random variables $X_1, X_2, X_3, ...$, where $X_i \sim Exp(\lambda)$. Then the arrival times are given as*

$$\begin{aligned} T_1 &= X_1 \\ T_2 &= X_1 + X_2 \\ T_2 &= X_1 + X_2 + X_3 \\ &\cdots\cdots \end{aligned}$$

19.7 Martingale theory

19.7.1 Basics

Definition 19.7.1 (martingale). *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{\mathcal{F}_t\}$ be a filtration on $\mathcal{F}$. Let X_t be a stochastic process. X_t is called a $\mathcal{F}_t$-martingale, if*

- *X_t is adapted to $\{\mathcal{F}_t\}$;*
- *$E[|X(t)|] < \infty, \forall t$;*
- *$E[X_t|\mathcal{F}_s] = X_s$ almost surely, for all $0 \leq s \leq t$.*

Definition 19.7.2 (supermartingale, submartingale). *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{\mathcal{F}_t\}$ be a filtration on $\mathcal{F}$. Let X_t be a stochastic process.*

- *X_t is called a $\mathcal{F}_t$-supermartingale, if X_t is adapted to $\{\mathcal{F}_t\}$;$E[|X(t)|] < \infty, \forall t$; $E[X_t|\mathcal{F}_s] \leq X_s$ almost surely, for all $0 \leq s \leq t$.*
- *X_t is called a $\mathcal{F}_t$-submartingale, if X_t is adapted to $\{\mathcal{F}_t\}$;$E[|X(t)|] < \infty, \forall t$; $E[X_t|\mathcal{F}_s] \geq X_s$ almost surely, for all $0 \leq s \leq t$.*

Remark 19.7.1.

- Martingale is always an adapted process with respect to some filtration.
- Note that for discrete setting, we have $E[X_n|\mathcal{F}_{n-1}] = X_{n-1}$.

Definition 19.7.3 (discrete-time martingale). *[2, p. 49] A sequence $X_1, X_2, ...$ of random variables is called a martingale with respect to a filtration $\mathcal{F}_1, \mathcal{F}_2, ...$ if*

1. *$E[|X_n|] < \infty$;*
2. *$X_1, X_2, ...$ is adapted to $\mathcal{F}_1, \mathcal{F}_2, ...$;*
3. *$E[X_{n+1}|\mathcal{F}_n] = X_n$*

Example 19.7.1 (Sum of independent zero-mean RVs as martingale). Let $X_1, X_2, ...$ be a sequence of independent integrable RVs with $E[|X_k|] < \infty$, and

$$E[X_k] = 0, \forall k.$$

Define

$$S_n = \sum_{i=1}^{n} X_i,$$

such that

$$E[|S_n|] = E[|X_1 + X_2 + ... + X_n|] \leq E[|X_1| + E[|X_2|] + ... + E[|X_n|]] < \infty;$$

and

$$\mathcal{F}_n = \sigma(X_1, X_2, ..., X_n), \mathcal{F}_0 = \{\emptyset, \Omega\}$$

Then the sequence $S_1, S_2, ..., S_n$ is a martingale with respect to $\mathcal{F}_1, \mathcal{F}_2, ...$ Note that a simple event in $\mathcal{F}_n$ should specify the value of $X_1, X_2, ..., X_n$, otherwise we cannot measure S_n.

Definition 19.7.4 (continous-time martingale). *[2, p. 49] A sequence stochastic process X_t is called a martingale with respect to a filtration $\{\mathcal{F}_t\}$ if*

1. $E[|X_t|] < \infty$;
2. $\{X_t\}$ *is adapted to* $\{\mathcal{F}_t\}$;
3. $E[X_t|\mathcal{F}_s] = X_s, s \leq t$.

Lemma 19.7.1 (martingales have constant expectation).

- *A discrete-time martingale X_n has the property that its expectation $E[X_t]$ is constant $E[X_1]$.*
- *A continuous-time martingale X_t has the property that its expectation $E[X_t]$ is constant $E[X_0]$.*

Proof. From property (2), using iterated expectation($E[E[X|\mathcal{F}]] = E[X]$,subsection 12.8.4), we can have $E[X_{n+1}] = E[E[X_{n+1}|\mathcal{F}_n]] = E[X_n] = ... = E[X_1]$. □

Theorem 19.7.1 (conditional expectation process as Martingale). *Let $(\Omega, P, \mathcal{F})$ be a probability space, and let $\{\mathcal{F}_t\}$ be a filtration on $(\Omega, P, \mathcal{F})$. Let Z be a random variable defined on $(\Omega, P, \mathcal{F})$.*

Define $Z(t) = E[Z|\mathcal{F}_t]$, then $Z(t)$ is a martingale with respect to $\mathcal{F}_t$.

Proof.

$$E[Z(t)|\mathcal{F}_s] = E[E[Z|\mathcal{F}_t]|\mathcal{F}_s] = Z(s).$$

□

19.7.2 Exponential martingale

Lemma 19.7.2 (Exponential martingale).

- *Let $W(t)$ be the Wiener process, define $Z(t) = \exp(\sigma W(t) - \frac{1}{2}\sigma^2 t)$. Then $Z(t)$ is martingale; moreover, $E[Z(t)] = E[Z(0)] = 1$.*
- *Let $X \sim N(0, \sigma^2)$, then*
$$E[\exp(\pm X - \frac{1}{2}\sigma^2)] = 1.$$
- *Let $X \sim MN(0, \Sigma), \Sigma \in \mathbb{R}^{n\times n}, t \in \mathbb{R}^n$, then*
$$E[\exp(\pm t^T X - \frac{1}{2}t^T \Sigma t)] = 1.$$
- *Let $W(t)$ be a n-dimensional correlated Wiener process with $dW(t)dW(t)^T = \rho(t)dt \in \mathbb{R}^{n\times n}$. Let $\{\mathcal{F}_t\}$ be the filtration generated by $W(t)$. Let $\theta(t) \in \mathbb{R}^n$ be a process adapted to $\mathcal{F}_t$. It follows that*
$$Z(T) = \exp(\pm \int_0^T \theta(u)^T dW(u) - \frac{1}{2}\int_0^T \theta(u)^T \rho(u)\theta(u)du)$$
is martingale such that
$$Z(t) = E[Z(T)|\mathcal{F}_t], E[Z(t)] = E[Z(T)] = 1.$$

Proof. (1) (a)

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

where we use by the fact that

$$E[\exp(\sigma(W(t) - W(s))|\mathcal{F}_s] = \int \exp(\sigma x)f(x)dx = \exp(\frac{1}{2}\sigma^2(t-s)), X \sim N(0,(t-s)).$$

To calculate the expectation, we have

$$E[Z(t)] = \exp(-1/2\sigma^2 t)E[\exp(\sigma W(t))] = \exp(-1/2\sigma^2 t)M_X(\sigma\sqrt{t}) = 1$$

where M_X is the moment generating function of standard normal random variable X. (b) We can also use conclusion from (2). Note that $\sigma W(t) \sim N(0, \sigma^2 t)$.

(2)

$$\begin{aligned}E[\exp(-X-\frac{1}{2}\sigma^2)] &= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{x^2}{2\sigma^2})\exp(-x-\frac{1}{2}\sigma^2)dx\\ &= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-\sigma^2)}{2\sigma^2})dx)\\ &= 1\end{aligned}$$

(3) Note that $t^TX \sim N(0, t^T\Sigma t)$(Theorem 15.1.1) (4) Introduce

$$Y_t \triangleq \pm\int_0^t \theta(u)^T dW(u) - \frac{1}{2}\int_0^t \theta(u)^T\rho(u)\theta(u)du, Z_t = \exp(Y_t).$$

Then

$$dY_t = -\frac{1}{2}\theta(t)^T\rho(t)\theta(t)dt \pm \theta(t)^T dW(t),$$

and

$$dZ_t = \exp(Y_t)dY_t + \frac{1}{2}\exp(Y_t)(\theta(t)^T\theta(t))dt = \pm Z_t\theta(t)^T dW(t).$$

Note that $Z_0 = 1$, Z_t can be written in integral form as

$$Z_t = 1 \pm \int_0^t Z_s\theta(s)^T dW(s).$$

Because the expectation of Ito integral is zero, we have

$$E[Z_T] = E[Z_t] = 1.$$

To show $Z_t = E[Z_T|\mathcal{F}_t]$, we have

$$\begin{aligned}&E[Z_T|\mathcal{F}_t]\\ =&E[Z_t \pm \int_t^T Z_s\theta(s)^T dW(s)|\mathcal{F}_t]\\ =&Z_t\end{aligned}$$

□

Note 19.7.1 (understanding $\theta(t)$ process). Note that we said $\theta(t)$ is adapted to $\mathcal{F}_t$ generally means that it is governed by SDE

$$d\theta(t) = \mu(\theta, t)dt + \Sigma(\theta)dW(t).$$

This representation includes the following cases.

- $\theta(t)$ is a deterministic process. In this case, we can prove in this way:Note that $\int_0^T \theta(u)^T dW(u) \sim (0, \int_0^T \theta^T(u)\rho(u)\theta(u)du)$ from Lemma 20.2.9. Note that

$$E[(\int_0^T \theta^T dW(u))(\int_0^T \theta^T dW(v))^T] = \int_0^T \int_0^T \theta^T(u)\rho(u)\delta(u-v)\theta(u)dudv.$$

- $\theta(t)$ is a stochastic process driven by $W(t)$. Note that

$$\begin{aligned} dY_t &= -\frac{1}{2}\theta(t)^T\rho(t)\theta(t)dt \pm \theta(t)^T dW(t) - \int_0^t (\rho(t)\theta(t))^T d(\theta)dt \pm \int_0^t [d\theta(t)]^T dW(t) \\ &= -\frac{1}{2}\theta(t)^T\rho(t)\theta(t)dt \pm \theta(t)^T dW(t) \end{aligned}$$

where we use the fact that $d\theta(t) \sim O((dt)^{1/2})$ to ignore the higher order terms.

Example 19.7.2 (application in finance). Under risk-neutral measure, the stock price is given as

$$S_t = S_0 \exp((r - \sigma^2/2)t + \sigma W_t)$$

where r is risk-free rate, σ is the volatility and W_t is the Brownian motion. It can be showed that $\exp(-rt)S_t = \exp(\sigma W_t - \sigma^2 t/2)$ is an martingale(exponential martingale).

19.7.3 Martingale transformation

Definition 19.7.5 (Predictable/previsible process). *Let $\{Y_t\}$ be a sequence random variables adapted to filtration $\{\mathcal{F}_t\}$. The sequence Y_t is said to be predictable if for every $t \geq 1$, the random variable Y_t is $\mathcal{F}_{t-1}$ measurable, or equivalently, $\sigma(Y_t) \subseteq \mathcal{F}_{t-1}$*

Definition 19.7.6 (Martingale transform). *[4, p. 83] Let $\{X_t\}$ be a martingale, let $\{Y_t\}$ be a predictable sequence. The martingale transform $\{(Y \cdot X)_t\}$ is the*

$$(Y \cdot X)_t = X_0 + \sum_{j=1}^{t} Y_j(X_j - X_{j-1})$$

Lemma 19.7.3 (Martingale transformation is a martingale). *Assume that $\{X_t\}$ is an adapted sequence and $\{Y_t\}$ a predictable sequence, both relative to a filtration $\{\mathcal{F}_t\}$. If $\{X_t\}$*

is a martingale, then the martingale transform $\{(Y_t \cdot X_t)\}$ *is a martingale with respect to* $\{\mathcal{F}_t\}$ *if* $E[X_j^2] < \infty, \forall j$

Proof. $E[(Y \cdot X)_t - (Y \cdot X)_{t-1} | cF_{t-1}] = E[Y_t(X_t - X_{t-1})|\mathcal{F}_{t-1}] = 0$ □

Lemma 19.7.4 (connection to Ito integral). *Let* $S_n = X_1 + ... + X_n$ *be a random walk, then the new random process*

- $Y_n = \sum_{i=1}^n X_{i-1}(X_i - X_{i-1})$ *is a martingale. Moreover,* $E[Y_n] = 0$.
- $Z_n = \sum_{i=1}^n f(X_{i-1})(X_i - X_{i-1})$ *is a martingale for any function* $f(y)$*.Moreover,* $E[Z_n] = 0$.

Proof. It is easy to see that X_{i-1} is measurable respect to $\mathcal{F}_i$. Therefore they are martingale transformation and they are martingales. □

Remark 19.7.2 (interpretation as discrete version of Ito integral). Later we will see these two examples are discrete version of Ito integral of $\int_0^t W_t dW_t, \int_0^t f(W_t) dW_t$.

19.8 Stopping time

Definition 19.8.1 (stopping time, continuous version). *Let* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in I}, P)$, $I = [0, \infty)$ *be a filtered probability space. Then a random variable* $\tau : \Omega \to I$ *is called a* $\mathcal{F}_t$ *stopping time if*

$$\{\omega \in \Omega : \tau(\omega) \leq t)\} \in \mathcal{F}_t,$$

that is, the subset of Ω, $\{\omega \in \Omega : \tau(\omega) \leq t)\}$ *is measurable respect to* $\mathcal{F}_t$.

Definition 19.8.2 (stopping time, discrete version). *[14] Let* $X = \{X_n, n \geq 0\}$ *be a stochastic process. A stopping time* τ *with respect to* X *is a discrete random variable on the same probability space of* X*, taking values in the set* $\{0, 1, 2, ...\}$*, such that for each* $n \geq 0$*, the event* $\{\tau = n\}$ *is completed determined by the information up to* n*,i.e., the values of* $\{X_0, X_1, ..., X_n\}$*, or equivalently, the subset in* Ω*:* $\{\omega \in \Omega : \tau(\omega) \leq n\}$ *is* $\mathcal{F}_n$ *measurable.*

Remark 19.8.1. If X_n denote the price of the stock at time n, τ denotes the time at which we will sell it. If our selling decision is based on past information, then τ will be a function of past 'states' characterized by $\{X_0, X_1, X_2, ..., X_{\min(\tau,n)}\}$. Moreover, the amount of past information it depends on is restricted by τ.

19.8.1 Stopping time examples

19.8.1.1 *First passage time*

Let stochastic process X has a discrete state space, and let i be a fixed state, then the first passage time defined as[14]

$$\tau = \min\{n \geq 0 : X_n = i\}$$

is stopping time. At first, τ is a random variable; second, the event $\{\tau = n\}$ is completely determined by the value of $\{X_0, X_1, ..., X_n\}$, i.e., the information up to n. Therefore, it is a stopping time.

19.8.1.2 *Trivial stopping time*

Let X be any stochastic process, and let τ be a deterministic function. The real world example is that a gambler decides that he will only play 10 games regardless of the outcome. τ is a stopping time.

19.8.1.3 *Counter example: last exit time*

Consider the rat in a open maze, a stochastic process X, taking discrete values representing states. Let τ denote the last time the rat visits state i:

$$\tau = \max\{n \geq 0 : X_n = i\}$$

Clearly, we need to know the future to determine the value of τ.

19.8.2 Wald's equation

Theorem 19.8.1 (Wald's equation). *If τ is a stopping time with respect to an iid sequence $\{X_n : n \geq 1\}$, and if $E[\tau] < \infty, E\|X_n|] < \infty$, then*

$$E[\sum_{n=1}^{\tau} X_n] = E[\tau]E[X_1]$$

Proof.

$$E[\sum_{n=1}^{\tau} X_n] = E[\sum_{n=1}^{\infty} X_n I(\tau > n-1)] = E[\sum_{n=1}^{\infty} X_n I(\tau > n-1)] = E[X_1]E[\tau]$$

where $I(\tau > n-1)$ is an indicator function. Note that the event $\{\tau > n-1\}$ only depends on the values of $\{X_1, X_2, ..., X_{n-1}\}$ since its complement event $\{\tau \leq n-1\}$ only depends on the values of $\{X_1, X_2, ..., X_{n-1}\}$. And we have

$$\begin{aligned}
E[I(\tau > n-1)] &= \sum_{n=1}^{\infty} P(\tau > n-1) \\
&= \sum_{n=0}^{\infty} P(\tau > n) \\
&= \sum_{n=0}^{\infty} \sum_{i=n+1}^{\infty} P(\tau = i) \\
&= \sum_{i=0}^{\infty} \sum_{n=0}^{i} P(\tau = i) \\
&= \sum_{i=0}^{\infty} iP(\tau = i) = E[\tau]
\end{aligned}$$

□

19.8.3 Optional stopping

Theorem 19.8.2 (optional stopping theorem). *Let $X = \{X_n, n \geq 0\}$ be a martingale, let τ be a stopping time with respect to X. Define a stochastic process $\bar{X} = \{X_{n\wedge\tau}\}$, then $\bar{X}$ is a martingale.*

Proof. Let $\mathcal{F}_n = \sigma(X_0, X_1, ...X_n)$,we can rewrite $\bar{X}_{n+1} = \bar{X}_n + I(\tau > n)(X_{n+1} - X_n)$ (this can be verified by consider events of $\{\tau > n\}$ and $\{\tau \leq n\}$),then $E[\bar{X}_{n+1}|\mathcal{F}_n] = \bar{X}_n + 0 = \bar{X}_n$. □

Remark 19.8.2 (stopping time strategy in fair game is still fair). Since $\bar{X}_0 = X_0, E[\bar{X}_n] = X_0$, the implication is using any stopping time as a gambling strategy yields on average, no benefit; the game is still fair.

19.8.4 Stopping time analysis of Wiener processes

19.8.4.1 *Minimum and maximum of a Wiener process*

Lemma 19.8.1. *[15, pp. 88, 214] Let W_t be a Wiener process and let $M_t = \max_{0\leq s\leq t} W_s$ be the maximum level reached by W_t during hte time interval $[0,t]$ and let $m_t = \min_{0\leq s\leq t} W_s$ be the minimum level reached by W_t during the time interval $[0,t]$.*

Then for all $t \geq 0$, we have

- $$P(M_t \leq a, W_t \leq x) = \begin{cases} \Phi(\frac{x}{\sqrt{t}}) - \Phi(\frac{x-2a}{\sqrt{t}}), a \geq 0, x \leq a \\ \Phi(\frac{x}{\sqrt{t}}) - \Phi(-\frac{a}{\sqrt{t}}), a \geq 0, x \geq a \\ 0, a \leq 0 \end{cases},$$

  the associated joint density function is

  $$f_{M_t,W_t}(a,x) = \begin{cases} \frac{2(2a-x)}{t\sqrt{2\pi t}} \exp(-\frac{1}{2}(\frac{(2a-x)^2}{t})), a \geq 0, x \leq a \\ 0, \; otherwise \end{cases}.$$

- $$P(m_t \geq b, W_t \geq x) = \begin{cases} \Phi(-\frac{x}{\sqrt{t}}) - \Phi(\frac{2b-x}{\sqrt{t}}), b \leq 0, x \geq b \\ \Phi(-\frac{b}{\sqrt{t}}) - \Phi(-\frac{b}{\sqrt{t}}), b \leq 0, x \leq b \\ 0, b \geq 0 \end{cases},$$

  the associated joint density function is

  $$f_{m_t, W_t}(b, x) = \begin{cases} \frac{2(2b-x)}{t\sqrt{2\pi t}} \exp(-\frac{1}{2}(\frac{(2b-x)^2}{t})), b \leq 0, x \geq b \\ 0, \ otherwise \end{cases}.$$

- $$P(M_t \leq a) = \Phi(\frac{a}{\sigma\sqrt{t}}) - \Phi(\frac{-a}{\sigma\sqrt{t}}), a \geq 0.$$

- $$P(m_t \leq b) = \Phi(\frac{b}{\sigma\sqrt{t}}) + \Phi(\frac{-b}{\sigma\sqrt{t}}), b \leq 0.$$

Note 19.8.1 (common pitfall in extending to non-standard Wiener process).

- Consider a new process Y_t is defined by $Y_t = \mu t + W_t$. It is **wrong** to use $M_t^Y = \mu t + M_t$ to derive the distribution associated with the process Y_t. Becuase

  $$M_t^Y = \max_{0 \leq s \leq t} (\mu t + W_t) \neq \mu t + \max_{0 \leq s \leq t} W_t = \mu t + M_t.$$

  Instead, we need to use change of measure to derive the results associated with Y_t.
- However, for process Y_t defined by $Y_t = Y_0 + \sigma W_t$, we can use the relation $M_t^Y = Y_0 + M_t$ to derive the distribution associated with the process Y_t.

Lemma 19.8.2 (running min and maximum of a Wiener Process). *[15, p. 214] Let W_t be a Wiener process and let*

$$X_t = X_0 + \mu t + \sigma W_t.$$

let $M_t^X = \max_{0 \leq s \leq t} X_s$ be the maximum level reached by X_t during hte time interval $[0, t]$ and let $m_t^X = \min_{0 \leq s \leq t} W_s$ be the minimum level reached by X_t during the time interval $[0, t]$. Then

- $$P(M_t^X \leq x) = \Phi(\frac{x - X_0 - \mu t}{\sigma\sqrt{t}}) - \exp(\frac{2\mu(x - X_0)}{\sigma^2})\Phi(\frac{-x + X_0 - \mu t}{\sigma\sqrt{t}}), x \geq X_0.$$

- $$P(m_t^X \leq x) = \Phi(\frac{x - X_0 - \mu t}{\sigma\sqrt{t}}) + \exp(\frac{2\mu(x - X_0)}{\sigma^2})\Phi(\frac{-x + X_0 - \mu t}{\sigma\sqrt{t}}), x \leq X_0.$$

Lemma 19.8.3 (running min and maximum of a geometric Brownian Process). *Let W_t be a Wiener process and let*

$$S_t = S_0 \exp(\mu t + \sigma W_t).$$

Let $M_t^S = \max_{0 \leq u \leq t} S_u$ be the maximum level reached by S_t during hte time interval $[0, t]$ and let $m_t^X = \min_{0 \leq u \leq t} W_u$ be the minimum level reached by S_t during the time interval $[0, t]$. Then

- $$P(M_t^S \leq x) = \Phi(\frac{\ln(x/S_0) - \mu t}{\sigma\sqrt{t}}) - \exp(\frac{2\mu(\ln(x/S_0))}{\sigma^2})\Phi(\frac{-\ln(x/S_0) - \mu t}{\sigma\sqrt{t}}).,x \geq S_0.$$

- $$P(m_t^S \leq x) = \Phi(\frac{\ln(x/S_0) - \mu t}{\sigma\sqrt{t}}) + \exp(\frac{2\mu(\ln(x/S_0))}{\sigma^2})\Phi(\frac{-\ln(x/S_0) - \mu t}{\sigma\sqrt{t}}), x \leq S_0.$$

Proof. (1)Note that $S_t = S_0 \exp(X_t), X_t = \mu t + \sigma W_t$. Therefore

$$M_t^S = S_0 \exp(M_t^X),$$

and then use Lemma 19.8.2:

$$P(M_t^S \leq x) = P(M_t^X \leq \ln(\frac{x}{S_0})) = \Phi(\frac{\ln(x/S_0) - \mu t}{\sigma\sqrt{t}}) - \exp(\frac{2\mu(\ln(x/S_0))}{\sigma^2})\Phi(\frac{-\ln(x/S_0) - \mu t}{\sigma\sqrt{t}}).$$

(2) similar to (1). □

19.8.4.2 *Martingale method*

Lemma 19.8.4 (first hitting time in bounded region). *Let X_t be a Brownian motion with no drift. Consider two levels $\alpha > 0$ and $-\beta, \beta > 0$. Then*

- *The probability p_α hitting α before hitting $-\beta$ is $\frac{\beta}{\alpha+\beta}$; The probability p_β hitting $-\beta$ before hitting α is $\frac{\alpha}{\alpha+\beta}$*
- *the expected time to reach level α, or level β is $\alpha\beta$.*

Proof. (1) Let B_τ be process with τ being the stopping time hitting α or $-\beta$. B_τ is a martingale by optional stopping theorem(Theorem 19.8.2). Then we have

$$E[B_\tau] = p_\alpha \alpha + p_\beta(-\beta) = 0, p_\alpha + p_\beta = 1.$$

We can solve to get $p_\alpha = \beta/(\alpha+\beta)$,and $p_\beta = \alpha/(\alpha+\beta)$. (2) $E[B_t^2 - t] = 0 \implies E[\tau] = E[B_\tau^2] = p_\alpha \alpha^2 + p_\beta \beta^2 = \alpha\beta$. □

Lemma 19.8.5 (first hitting time of single level in unbounded region). *[3, p. 112] Let X_t be a Brownian motion with no drift. Consider one level $\alpha > 0$. Then*

- *The probability density of stopping time τ is*

$$\frac{\alpha \exp(-\alpha^2/2\tau)}{\tau\sqrt{2\pi\tau}}$$

- *The probability p_α hitting α before hitting $-\infty$ is 1*
- *The expected time to reach level α is ∞.*

Proof. Let τ be the stopping time, $\tau = \min\{t : X(t) = \alpha\}$. (1) $P(\tau < t) = P(\tau < t, X(t) \geq \alpha) + P(\tau < t, X(t) < \alpha) = 2P(\tau < t, X(t) \geq x) = 2P(X(t) \geq \alpha)$. Note that the event $X(t) \geq x$ already contains $\tau < \alpha$ (2) $P(\tau < \infty) = 1$. (3)

$$E[\tau] = \int_0^\infty \frac{\alpha \exp(-\alpha^2/2t)}{t\sqrt{2\pi t}} dt$$

and the integral will diverge. Another proof: We can use results from Lemma 19.8.4 and set $\beta = \infty$. □

19.8.4.3 *General method via Feyman Kac formula*

Lemma 19.8.6 (General method via Feyman Kac formula, fixed boundary and infinite time horizon). *Consider stochastic process given by*

$$dX(t) = mdt + \sigma dW(t), X(0) = 0$$

where $W(t)$ is the Brownian motion. Then given two levels $a > 0$ and $-b, b > 0$, and let the probability $P(t,x)$ denote the probability that the process starting at $X(t) = x$ hits a before hitting $-b$. Then we have

- *$P(t,x)$ is independent of time t.*

- *The governing equation for $P(x)$ is given by*

$$m\frac{\partial P}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 P}{\partial x^2} = 0,$$

with boundary condition $P(x = a, t) = 1, P(x = -b) = 0$.

Proof. (1) Note that this is a Markov process, therefore $P(t, x)$ will be depend on time. (2) Consider a value function $P(x, t) = E[P_T | X(t) = x]$ with final condition $P(x, T) = P_T$(P_T will take value 1 at target sites and take 0 elsewhere). Then from Feyman Kac theorem(Theorem 21.3.1), $P(x, t)$ is also the solution of

$$\frac{\partial P}{\partial t} + m\frac{\partial P}{\partial x} + \frac{1}{2}\sigma\frac{\partial^2 P}{\partial x^2} = 0,$$

with boundary condition $p(t, x) = P_T$. □

Remark 19.8.3 (generalization to high dimensional). This method can be easily generalized to high dimensional state space first passage time problem(Theorem 21.3.2).

Example 19.8.1. Let $a = 3, b = 5, \sigma = 1$, then we have

$$m\frac{\partial P}{\partial x} + \frac{1}{2}\frac{\partial^2 P}{\partial x^2}, P(3) = 1, P(-5) = 0.$$

The general solution gives

$$P(x) = c_1 e^{0x} + c_2 e^{-2mx} = c_1 + c_2 e^{-2mx}.$$

With boundary condition, we can solve c_1, c_2 therefore any $P(x)$ can be evaluated.

Example 19.8.2. Let $m = 0, \sigma = 1$. Then

$$P(x) = \frac{x + b}{a + b}, x \in [-b, a].$$

Example 19.8.3. Let $a = \infty, b = 1, m = \sigma = 1$, then we have

$$m\frac{\partial P}{\partial x} + \frac{1}{2}\frac{\partial^2 P}{\partial x^2}, P(\infty) = 0, P(-1) = 1.$$

The general solution gives

$$P(x) = c_1 e^{0x} + c_2 e^{-2mx} = c_1 + c_2 e^{-2mx}.$$

With boundary condition, we can solve c_1, c_2 as

$$c_1 + c_2 e^2 = 1, c_1 + c_2 e^{-\infty} = c_1 = 0$$

resulting $c_1 = 0, c_2 = e^{-2}, P(x) = e^{-2(x+1)}$.

Example 19.8.4 (boundary condition issue). **warning!**
Let $a = 1, b = \infty, m = \sigma = 1$, then we have

$$m\frac{\partial P}{\partial x} + \frac{1}{2}\frac{\partial^2 P}{\partial x^2}, P(1) = 1, P(-\infty) = ?.$$

Note that because the X is upward drifting, $P(-\infty)$ is not clear which value to set.

Lemma 19.8.7 (General method via Feyman Kac formula, fixed boundary, dynamic boundary and finite time horizon). *Consider stochastic process given by*

$$dX(t) = mdt + \sigma dW(t), X(0) = 0$$

where $W(t)$ is the Brownian motion. Then given one level $a(t)$ as a function of time. Let the probability $P(t,x)$ denote the probability that the process starting at $X(t) = x$ hits $a(t)$. Then the governing equation for $P(x,t)$ is given by

$$\frac{\partial P}{\partial t} + m\frac{\partial P}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 P}{\partial x^2} = 0,$$

with boundary condition $P(x(t) = a(t), t \leq T) = 1, P(x,T) = 0$.

Proof. Let τ be an arbitrary time satisfying $\tau \leq T$. Then,

$$P(x(\tau) = a(\tau)|x(t) = x) = E[\mathbf{1}_{x(\tau)=a(\tau)}|x(t) = x].$$

Then from Feyman Kac theorem(Theorem 21.3.1), $P(x,t)$ is also the solution of

$$\frac{\partial P}{\partial t} + m\frac{\partial P}{\partial x} + \frac{1}{2}\sigma\frac{\partial^2 P}{\partial x^2} = 0,$$

with boundary condition $P(\tau, x(\tau) = a(\tau), \tau \leq T) = 1$. Since we limit the time horizon to T, we have $P(x,T) = 0$. □

19.9 Notes on bibliography

For elementary level treatment on stochastic process, see [2][4][13] and intermediate level [16]. For general SDE, see [17][16]. For treatment on forward and backward SDE, see [18].

For numerical algorithm for SDE, see [19].

For comprehensive and advanced treatment of stochastic methods, see [20] and [21].

For martingale theory, see [22].

For stationary stochastic process, see [9][8].

A good source on simulating SDE with code is at [23].

BIBLIOGRAPHY

1. Dineen, S. *Probability theory in finance: a mathematical guide to the Black-Scholes formula* (American Mathematical Soc., 2013).
2. Brzezniak, Z. & Zastawniak, T. *Basic stochastic processes: a course through exercises* (Springer Science & Business Media, 1999).
3. Shreve, S. E. *Stochastic calculus for finance II: Continuous-time models* (Springer Science & Business Media, 2004).
4. Mikosch, T. *Elementary stochastic calculus with finance in view* (World scientific, 1998).
5. Williams, D. *Probability with martingales* (Cambridge university press, 1991).
6. Björk, T. *Arbitrage theory in continuous time* (Oxford university press, 2009).
7. Koralov, L. & Sinai, Y. G. *Theory of probability and random processes* (Springer Science & Business Media, 2007).
8. Lindgren, G., Rootzén, H. & Sandsten, M. *Stationary stochastic processes for scientists and engineers* (CRC press, 2013).
9. Lindgren, G. *Stationary Stochastic Processes: Theory and Applications* ISBN: 9781466557796 (Taylor & Francis, 2012).
10. Miranda, H.-C. *Applied Stochastic Analysis lecture notes* (NewYork University, 2015).
11. Wikipedia. *Markov process — Wikipedia, The Free Encyclopedia* [Online; accessed 28-June-2016]. 2016.
12. Joshi, M. S. *The concepts and practice of mathematical finance* (Cambridge University Press, 2003).
13. Wiersema, U. F. *Brownian motion calculus* (John Wiley & Sons, 2008).
14. Karl, S. *Stochastic modeling I lecture notes* (Columbia University, 2009).
15. Chin, E., Nel, D. & Ãlafsson, S. *Problems and Solutions in Mathematical Finance: Equity Derivatives* (John Wiley & Sons, 2017).
16. Klebaner, F. C. *Introduction to Stochastic Calculus with Applications* (Imperial College Press, 2005).
17. Oksendal, B. *Stochastic differential equations: an introduction with applications* (Springer Science & Business Media, 2013).
18. Ma, J. & Yong, J. *Forward-backward stochastic differential equations and their applications* **1702** (Springer Science & Business Media, 1999).

19. Higham, D. J. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM review* **43,** 525–546 (2001).

20. Gardiner, C. *Stochastic Methods: A Handbook for the Natural and Social Sciences* ISBN: 9783540707127 (Springer Berlin Heidelberg, 2009).

21. Gardiner, C. *Handbook of stochastic methods for physics, chemistry, and the natural sciences* ISBN: 9783540156079 (Springer, 1994).

22. Ladd, W. *Some Applications of Martingales to Probability Theory* tech. rep. (Technical Report, University of Chicago, 2011).

23. Iacus, S. M. *Simulation and inference for stochastic differential equations: with R examples* (Springer Science & Business Media, 2009).

20

STOCHASTIC CALCULUS

20.1 Ito integral

20.1.1 Construction of Ito integral

Definition 20.1.1 (simple process). *The stochastic process $C_t, t \in [0,T]$ is said to be simple if: there exista a partition*

$$\tau_n : 0 = t_0 < t_1 < ... < t_n = T$$

and a sequence of random variables $Z_i, i = 1,2,...,n$ such that

$$C_t = \begin{cases} Z_n, if\ t = T, \\ Z_i, if\ t_{i-1} \leq t < t_i, i = 1,2,...,n \end{cases}$$

The sequence Z_i is adapted to $\mathcal{F}_{t_{i-1}}$ and $E[Z_i^2] < \infty$,i.e., the sequence Z_i is a previsible process.

Definition 20.1.2 (Ito integral for simple processes). *Let C_t be a simple process on $[0,T]$, the Ito integral is defined as*

$$\int_0^T C_s dB_s = \sum_{i=1}^n C_{t_{i-1}}(B_{t_i} - B_{t_{i-1}})$$

Theorem 20.1.1 (Ito integral as martingale transform). *The sequence of Ito integral*

$$\int_0^{t_k} C_s dB_s, k = 0,1,2,...,n$$

of a simple process C_s is a martingale transform with respect to the Brownian filtration $\mathcal{F}_{t_k}$. And the stochastic process $I_t(C) = \int_0^t C_s dB_s$ is a martingale with respect to the Brownian filtration $\mathcal{F}_t$.

Proof. directly from martingale transformation in discrete time.(Lemma 19.7.3) □

Theorem 20.1.2 (Isometry for simple process). *The Ito stochastic integral satisfies the isometry property:*

$$E[(\int_0^t C_s dB_s)^2] = E[\int_0^T C_s^2 ds]$$

Proof. By definition, for simple process:

$$\int_0^t C_s dB_s = \sum_{i=1}^n Z_i \Delta_i B$$

where $\Delta_i B = B(t_i) - B(t_{i-1})$. Then

$$\begin{aligned} E[(\int_0^t C_s dB_s)^2] &= E[\sum_{i=1}^n \sum_{j=1}^n Z_i \Delta_i B Z_j \Delta_j B] = E[\sum_{j=1}^n (Z_i)^2 (\Delta_i B)^2] \\ &= \sum_{j=1}^n E[(Z_i)^2](t_i - t_{i-1}) = \int_0^t E[C_s^2] dt \end{aligned}$$

where we use $E[(\Delta_i B)^2] = t_i - t_{i-1}$ □

Corollary 20.1.2.1. *Consider the stochastic process X_t defined as the Ito integral of a simple process:*

$$X_t = \int_0^t C_s dB_s$$

Then we have

- $E[X_t] = 0$
- $E[X_t^2] = Var[X_t] = E[\int_0^2 C_s^2 dB_s]$

Theorem 20.1.3 (existence of approximating simple process). *Let C be a stochastic process that (1) adapted to Brownian filtration on $[0, T]$ and $\int_0^T E[C^2]dt$ is finite. Then there exist a sequence of simple process C_s^n such that*

$$\lim_{n\to\infty} \int_0^T E[(C_s^n - C)]dt = 0$$

Proof. see[1, p. 109]. □

Definition 20.1.3 (Ito integral of general process). *Let C be a stochastic process that (1) adapted to Brownian filtration on* $[0, T]$ *and (2)* $\int_0^T E[C^2]dt$ *is finite. Then*

$$\int_0^T CdB = \lim_{n\to\infty} \int_0^T C_s^n dB$$

where C_s^n *is a sequence of simple process approximating C in the mean square sense.*

20.1.2 Properties of Ito integral

Definition 20.1.4 (non-anticipating process). *A stochastic process* F_t *is said to be a non-anticipating process with respect to the Brownian motion* W_t *if* F_t *is independent of* $B_{t'} - B_t$ *with* $t' > t$.

Lemma 20.1.1 (basic properties).

- $\int_0^T cdW_t = cW_T$, *where c is a constant.*
- $I_s = \int_0^s W_t dW_t = 0.5(W_s^2 - s)$ *is a martingale, and* $E[I_s] = E[I_0] = 0$.
- *Let g be an square-integrable adapted process to the Brownian filtration* $\{\mathcal{F}_t\}$ *generated by Brownian motion* $W(s)$. *Then* $I(t) = \int_0^t g(s)dW(s)$ *is a continuous square-integrable martingale.* [a]

[a] an adapted process means either deterministic process or stochastic process represented by $dg_t = \mu(g(t), t)dt + \sigma(g(t), t)dW(t)$.

Proof. (1) recognize that this is a Wiener integral(Lemma 20.1.2) on the left, which will produce a normal distribution of $N(0, \int_0^T c^2 dt)$. The Right side has the exact same distribution. (2) Let $Y_t = W_t^2$, and then

$$dY_t = 2W_t dW_t + dt$$

Integrate both sides, we have

$$W_T^2 = 2\int_0^T W_t dW_t + T.$$

(3)

$$dI_t = g(t)dW(t) + \int_0^t dg(s)dW(s) = g(t)dW(t)$$

where we ignore $\int_0^t dg(s)dW(s)$ since it is of order $(O(t))$.

□

Theorem 20.1.4 (Properties of Ito integral). *[2, p. 100] Let $f(W_t,t), g(W_t,t)$ be nonanticipating processes and $c \in \mathbb{R}$, then we have*

1. *partition property:*
$$\int_S^T f dW_t = \int_S^u f dW_t + \int_u^T f dW_t$$
if $S < u < T$
2. *linearity:*
$$\int_S^T (cf + dg) dW_t = c \int_S^T f dW_t + d \int_S^T g dW_t$$
3. *zero mean:*
$$E[\int_S^T f(W_t,t) dW_t] = 0$$
4. *Isometry:*
$$E[(\int_a^b f(W_t,t) dW_t)^2] = E[\int_a^b f(W_t,t)^2 dt]$$
5. *Convariance:*
$$E[(\int_a^b f(W_t,t) dW_t)(\int_a^b g(W_t,t) dW_t)] = E[\int_a^b f(W_t,t) g(W_t,t) dt]$$

Remark 20.1.1.

- See [3][1]
- (3) does not hold for Stratonvich integral.
- all properties hold for both simple process and general process, since general process is defined via simple process.

20.1.3 Wiener integral and Riemman integral with Wiener process

Lemma 20.1.2 (Wiener integral). *[2, p. 112] Suppose $g : [0,\infty) \to \mathbb{R}$ is a bounded, piecewise continuous function in L^2. Let B_t be a Brownian motion, then*

$$\int_0^t g(s) dB_s$$

is a random variable which has a

$$N(0, \int_0^t g^2(s)ds)$$

distribution. This integral is also known as ***Wiener integral****. In particular,*

$$\int_0^t dB_s = B_t \sim N(0,t)$$

Proof. (informal) Directly from zero mean and Isometry properties(see Theorem 20.1.4). The resulting process is Gaussian can be derived from the sum of independent Gaussian random variables is Gaussian. □

Example 20.1.1. Consider stochastic process

$$X_t = \int_0^t \frac{1}{1-s} dW_s,$$

where W_t is the Wiener process. Then we have

- X_t is a Gaussian process.
- $E[X_t] = 0$.
- $Var[X_t] = E[X_t^2] - E[X_t]^2 = E[X_t^2]$, and

$$E[X_t^2] = \int_0^t \frac{1}{(1-s)^2} ds$$

  via Ito Isometry.

Lemma 20.1.3 (linearity of Wiener integral). *let $B(t)$ be the Wiener process, let $g, h, m : [0,\infty) \to \mathbb{R}$ be a bounded, piecewise continuous function in L^2. Then*

$$\alpha \int_0^t g(s)dB_s + \beta \int_0^t h(s)dB_s = \int_0^t \alpha g(s) + \beta h(s) dB_s.$$

In particular,

$$m(t)B_t - \int_0^t h(s)dB_s = \int_0^t (m(t) - h(s))dB_s \sim N(0, \int_0^t (m(t)-h(s))^2 ds)$$

Proof. The linearity can be derived directly from the linearity of Ito integral. □

caution!
We know that

$$Z_t = \int_0^t \alpha g(s) + \beta h(s) dB_s \sim N(0, \int_0^t (\alpha g(s) + \beta h(s))^2 dt)$$

however,

$$X_t = \alpha \int_0^t g(s) dB_s \sim N(0, \alpha^2 \int_0^t g(s)^2 ds), Y_t = \beta \int_0^t h(s) dB_s \sim N(0, \beta^2 \int_0^t h(s)^2 ds)$$

Note that X_t and Y_t are **not independent to each other because they are generated from the same Wiener process.**

Theorem 20.1.5 (Riemann integral with Wiener process). *Let g be a smooth function, and let $W(t)$ be the Wiener process, then*

$$\int_0^t g(W(s)) ds = g(W(t))t - \int_0^t s g'(W(s)) dW_s - \int_0^t \frac{1}{2} g''(W(s)) ds.$$

In particularly, if $g(x) = x$, then

$$\int_0^t W(s) ds = W(t)t - \int_0^t s dW_s \sim N(0, \int_0^t (t-s)^2 ds).$$

Proof. (1)Consider $f(W_t, t) = g(W_t)t$ and apply Ito rule (Lemma 20.2.1), we will get

$$d(g(W_t)t) = g(W_t)dt + tg'(W_t)dW_t + \frac{1}{2} tg''(W_t)dt$$

where $dW_t dW_t = dt$ is used. Then we integrate both sides. (2) To show

$$\int_0^t W(s) ds = W(t)t - \int_0^t s dW_s \sim N(0, \int_0^t (t-s)^2 ds),$$

note that $W(t) = \int_0^t dW_s$ and then use linearity of Wiener integral(Lemma 20.1.3). □

Corollary 20.1.5.1. *Let $W(t)$ be the Wiener process, then*

- $$\int_0^1 W(s) ds = W(1) - \int_0^1 s dW_s \sim N(0, \int_0^1 (1-s)^2 ds) = N(0, \frac{1}{3}).$$
- $$\int_0^T W(s) ds = \frac{T}{\sqrt{3}} W(T) \sim N(0, \frac{T^3}{3})$$

- $$\int_0^t g'(s)W_s ds \sim N(0, \int_0^t [g(t)-g(s)]^2 ds)$$
- $$\int_0^1 s^n W_s ds \sim N(0, \frac{2}{(2n+3)(n+2)}), n = 0,1,2,...$$

Proof. (2) Since
$$d(sW(s)) = W_s ds + s dW_s,$$
we have
$$TW_T) = \int_0^T W_s ds + \int_0^T s dW_s.$$
Rearrange, we have
$$\int_0^T W_s ds = \int_0^T (T-s) dW_s \sim N(0, \int_0^T (T-s)^2 ds) = N(0, \frac{T^3}{3}).$$

(3) Let $f(W_t, t) = g(t)W_t$.(4) from (2). □

20.1.4 Quadratic variations

Lemma 20.1.4 (Quadratic variations for Ito process). *Consider an Ito process given by $dX_t = a(t, X_t)dt + b(t, X_t)dB_t$, then the quadratic variation of a process given by*
$$\int_0^T dX_t dX_t = \int_0^T b^2 dt = b^2 t.$$

Proof. Directly compute $dX_t dX_t$ and ignore $o(dt)$ terms. □

20.2 Stochastic differential equations

20.2.1 Ito Stochastic differential equations

Definition 20.2.1 (Ito SDE). *[1, p. 137]An Ito stochastic differential equation is defined as*

$$dX_t = a(t, X_t)dt + b(t, X_t)dB_t$$

which could be interpret as

$$X_t = X_0 + \int_0^t a(s, X_s)ds + \int_0^t b(s, X_s)dB_s$$

where the first integral is Riemann integral, the second is Ito integral.

Remark 20.2.1. The integral equation is **Not** a solution, because it contains $X(t)$ itself.

Theorem 20.2.1 (existence). *[1, p. 138] Assume the initial condition X_0 has a finite second moment: $EX_0^2 < \infty$, and is independent of $(B_t, t \geq 0)$. Assume that, for all $t \in [0, T], x, y \in \mathbb{R}$, the coefficient functions $a(t, x)$ and $b(t, x)$ satisfy the following conditions:*

- *They are continuous*
- *They satisfy a Lipschitz condition with respect to the second variable:*

$$\left|a(t,x) - a(t,y)\right| + \left|b(t,x) - b(t,y)\right| \leq K\left|x - y\right|$$

Then the Ito stochastic differential equation has a unique solution X on $[0, T]$.

Theorem 20.2.2 (linear stochastic differential equation).

$$X_t = X_0 = \int_0^t (c_1 X_s + c_2)ds + \int_0^t (\sigma_1 X_s + \sigma_2)dB_s$$

for constants c_i and σ_i is called linear SDE. The Linear SDE has an unique solution.

Proof. It is easy to show that the continuous condition and Lipschitz condition are satisfied. □

20.2.2 Ito's lemma

20.2.2.1 *one-dimensional version*

Lemma 20.2.1. *[4, p. 79] Let $f(B_t, t)$ be a function of Brownian motion B_t, then*

$$df = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial B_t}dB_t + \frac{1}{2}\frac{\partial^2 f}{\partial B_t^2}dt$$

Lemma 20.2.2. *Let $f(X_t, t)$ be a function of stochastic process X_t governed by $dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dB_t$, then*

$$\begin{aligned} df &= \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial X_t}dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial X_t^2}(dX_t)^2 \\ &= (\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial X_t^2}\sigma^2 + \frac{\partial f}{\partial X_t}\mu)dt + \frac{\partial f}{\partial X_t}\sigma dB_t \end{aligned}$$

Example 20.2.1. $X_t = W_t^3$, then $dX_t = 3W_t^2 dW_t + 3W_t dW_t dW_t = 3W_t^2 dW_t + 3W_t dt$

Example 20.2.2. $Y_t = \ln(W_t)$, then $dY_t = dW_t/W_t - \frac{1}{2}dt/W_t^2$

20.2.2.2 *Multi-dimensional version*

Lemma 20.2.3. *Let $f(B_{1,t}, B_{2,t}, ..., B_{n,t}, t)$ be a function of Brownian motion $B_{1,t}, B_{2,t}, ..., B_{n,t}$, then*

$$df = \frac{\partial f}{\partial t}dt + \sum_{i=1}^{n}\frac{\partial f}{\partial B_{i,t}}dB_{i,t} + \sum_i\sum_j \frac{1}{2}\frac{\partial^2 f}{\partial B_{i,t}\partial B_{j,t}}D_{ij}dt$$

where we assume $E[dB_{i,t}dB_{j,t}] = D_{ij}dt$.

20.2.2.3 *Product rule and quotient rule*

Lemma 20.2.4 (product rule and quotient rule). *[4, p. 79] Consider*

$$\begin{aligned} dX_t/X_t &= r_1 dt + \sigma_1 dW_1 \\ dY_t/Y_t &= r_2 dt + \sigma_2 dW_2 \\ dW_1 dW_2 &= \rho dt \end{aligned}$$

It follows that

- *Given $Z_t = X_t Y_t$, we have*

$$\begin{aligned} dZ_t = d(X_t Y_t) &= X_t dY_t + Y_t dX_t + dX_t dY_t \\ &= X_t Y_t((r_1 + r_2 + \rho\sigma_1\sigma_2)dt + (\sigma_1 dW_1 + \sigma_2 dW_2)) \end{aligned}$$

- *Given $Z_t = X_t/Y_t$, we have*

$$\begin{aligned} dZ_t = d(X_t/Y_t) &= dX_t/Y_t - X_t dY_t/(Y_t)^2 - dX_t dY_t/(Y_t)^2 + X_t(dY_t)^2/(Y_t)^3 \\ &= (X_t/Y_t)((r_1 - r_2 - \rho\sigma_1\sigma_2 + \sigma_2^2)dt + (\sigma_1 dW_1 - \sigma_2 dW_2)) \end{aligned}$$

- *Given $Z_t = 1/X_t$, we have*

$$\begin{aligned} dZ_t = d(1/X_t) &= -dX_t/(X_t)^2 + (dX_t)^2/(X_t)^3 \\ &= (1/X_t)((-r_1 + \sigma_1^2)dt - \sigma_1 dW_1) \end{aligned}$$

Note that we have to calculate the Hessian for $f(x,y) = x/y$, and there are two terms for the cross-term.

Proof. (1)

$$dZ_t = \frac{\partial Z}{\partial X} dX + \frac{\partial Z}{\partial X} dX + \frac{1}{2}\frac{\partial^2 Z}{\partial X^2} dX dX + \frac{1}{2}\frac{\partial^2 Z}{\partial Y^2} dY dY + \frac{\partial^2 Z}{\partial X \partial Y} dX dY.$$

(2)(3) Same as (1). □

20.2.2.4 *Logorithm and exponential*

Lemma 20.2.5 (Ito lemma applied to logorithm and exponential). *Let $X(t)$ be an Ito stochasti process.*

- *If $Y_t = \exp(X(t))$, then*

$$dY_t = Y_t dX_t + \frac{1}{2} Y_t dX_t dX_t.$$

- *If* $Z_t = \ln(X(t))$, *then*

$$dZ_t = \frac{1}{X_t}dX_t - \frac{1}{2}\frac{1}{X_t^2}dX_t dX_t.$$

Proof. (1)

$$\begin{aligned} dY_t &= \exp(X_t)dX_t + \frac{1}{2}\exp(X_t)dX_t dX_t \\ &= Y_t dX_t + \frac{1}{2}Y_t dX_t dX_t. \end{aligned}$$

(2)

$$dZ_t = \frac{1}{X_t}dX_t - \frac{1}{2}\frac{1}{X_t^2}dX_t dX_t.$$

□

20.2.2.5 *Integrals of Ito process*

Lemma 20.2.6 (integrand is an Ito stochastic process). *Let* $r(t)$ *be an Ito stochasti process.*

- *If* $X_t = \int_0^t r(s)ds$, *then*
$$dX_t = r(t)dt.$$
- *If* $Y_t = \exp(X_t)$, *then*
$$dY_t = Y_t r(t)dt.$$

Proof. (1) Let Ω be the sample space associated with the stochastic process $r(t)$. Then for each sample path $\omega \in \Omega$, we have $X_t(\omega) = \int_0^t r(s,\omega)ds$ and $dX_t(\omega) = r(t,\omega)ds$. Since $dX_t(\omega) \triangleq \lim_{dt\to 0} X(t+dt,\omega) - X(t,\omega)$ and $r(t,\omega)ds$ are both random variables for fixed t, if they are equal for each $\omega \in \Omega$, we can write

$$dX_t = r(t)dt.$$

(2)

$$dY_t = \exp(X_t)dX_t = \exp(X_t)r(t)dt = Y_t r(t)dt.$$

□

Remark 20.2.2 (common pitfalls).

- It is worth noting that when $X_t = \int_0^t r(s)ds$ and $r(t)$ is an Ito stochastic process, X_t is not an Ito integral process.

- Similarly, for $Y_t = \exp(X_t)$, Y_t is not an Ito integral, and the Ito lemma does not apply.

Lemma 20.2.7 (Ito lemma applied to integral of Ito processes). *Let $X(t)$ be an Ito stochasti process. Let $r(t)$ be a deterministic function.*

- *If $Y_t = \int_0^t r(s)dX(s)$, then*
$$dY_t = r(t)dX(t).$$
- *If $Z_t = \exp(Y_t)$, then*
$$dZ_t = Z_t r(t)dX(t) + \frac{1}{2}Z_t r(t)^2 dX(t)dX(t).$$

Proof. (1) by definition. (2) Using Ito rule (Lemma 20.2.5), we have

$$\begin{aligned} dZ_t &= Z_t dY_t + \frac{1}{2}Z_t dY_t dY_t \\ &= Z_t r(t)dX(t) + \frac{1}{2}Z_t r(t)^2 dX(t)dX(t) \end{aligned}$$

□

20.2.2.6 *Ito Integral by parts*

Lemma 20.2.8. *[4, p. 79] Let $X(t), Y(t)$ be two Ito processes. Then*

$$\int_s^u Y(t)dX(t) = X(u)Y(u) - X(s)Y(s) - \int_s^u X(t)dY(t) - \int_s^u dX(t)dY(t)$$

Proof. From the product rule, we have

$$\begin{aligned} \int_s^u d[X(t)Y(t)] &= \int_s^u Y(t)dX(t) + \int_s^u X(t)dY(t) + \int_s^u dX(t)dY(t) \\ X(u)Y(u) - X(s)Y(s) &= \int_s^u Y(t)dX(t) + \int_s^u X(t)dY(t) + \int_s^u dX(t)dY(t) \\ \int_s^u Y(t)dX(t) &= X(u)Y(u) - X(s)Y(s) - \int_s^u X(t)dY(t) - \int_s^u dX(t)dY(t) \end{aligned}$$

Note that This integral-by-part formula is the same as Riemann integral except for the extra term $\int_s^u dX(t)dY(t)$. □

20.2.2.7 *Fundamental theorem of Ito stochastic calculus*

Theorem 20.2.3 (Fundamental theorem of Ito stochastic calculus). *[5, p. 79] Let $h(W_t)$ be a function on $W(t)$, then*

$$h(W_t) - h(W_0) = \int_0^t h'(W_s)dW_s + \frac{1}{2}\int_0^t h''(W_s)ds.$$

Proof. Note that

$$dh = h'dW_t + \frac{1}{2}h''dt.$$

□

Example 20.2.3. If $h(x) = 0.5x^2$, we have

$$\frac{1}{2}(W_t^2 - W_0^2) = \int_{t_0}^t W_s dW_s + \frac{1}{2}\int_{t_0}^t sds$$

20.2.3 Solutions to Ito stochastic differential equations

Definition 20.2.2 (strong solution to SDE). *[1, p. 136]A strong solution to the Ito SDE is a stochastic process X_t which satisfies the following conditions:*

- *X_t is adapted to the Brownian motion filtration, i.e., X_t is a function of $B_s, s \leq t$.*
- *X is a function of the underlying Brownian motion sample path and of the coefficient functions $a(t,x)$ and $b(t,x)$.*

Example 20.2.4 (solution to geometric Brownian motion). The solution to the GBM SDE:

$$X_t = X_0 + c\int_0^t X_s ds + \sigma \int_0^t X_s dB_s$$

or

$$dX_t = cX_t dt + \sigma X_t dB_t$$

is

$$X_t = X_0 \exp((c - 0.5\sigma^2)t + \sigma B_t)$$

Verification: Let $X_t = f(t, B_t)$, then

$$dX_t = f_t dt + f_{B_t} dB_t + \frac{1}{2} f_{B_t B_t} dB_t dB_t = (c - 0.5\sigma^2) X_t dt + \sigma X_t dB_t + \frac{1}{2}\sigma^2 X_t dt$$

Example 20.2.5 (solution to Ornstein-Uhlenbeck process). [1, p. 141] The Ornstein-Uhlenbeck process SDE is:

$$dX_t = cX_t dt + \sigma dB_t$$

and it has solution:

$$X_t = e^{ct} X_0 + \sigma \int_0^t \exp(c(t-s)) dB_s$$

Example 20.2.6 (solution to mean reversion with square-root diffusion). [4, p. 112] The SDE is

$$dr(t) = -\lambda(r(t) - \bar{r}) + \sigma\sqrt{r(t)} dB(t)$$

This SDE has no closed form expression, for its mean and variance property, we can refer to [4, p. 112].

20.2.4 Solution method to linear SDE

20.2.4.1 *State-independent linear arithmetic SDE*

Lemma 20.2.9 (state independent/general arithmetic SDE). *[2, p. 146][4, p. 116]The solution X_t of the stochastic differential equation*

$$dX_t = a(t)dt + b(t)dW(t)$$

is given by

$$X_t = X_0 + \int_0^t a(s)ds + \int_0^t b(s)dW(s),$$

which is a ***Gaussian distribution*** *with mean $X_0 + \int_0^t a(s)ds$ and variance $\int_0^t b^2(s)ds$.*

Moreover, X_t is a Gaussian process(Lemma 19.3.3).

Proof. The integral form is

$$X_t - X_0 = \int_0^t a(s)ds + \int_0^t b(s)dW(s).$$

X_t is a Gaussian because it is a deterministic term plus a Gaussian random process $\int_0^t b(s)dW(s)$. The mean is

$$E[X_t] = X_0 + \int_0^t a(s)ds$$

where the fact of expectation of Ito integral is zero is used. For the calculation of variance, we use

$$E[(\int_0^t b(s)dW(s))^2] = \int_0^t b^2(s)ds$$

via Ito isomery. □

20.2.4.2 *State-independent linear geometric SDE*

Lemma 20.2.10 (general geometric SDE). *[4, p. 116] Consider the SDE*

$$dX(t) = \mu(t)X(t)dt + \sigma(t)X(t)dB(t).$$

It follows that

- *It has the equivalent form*

$$Y_t = \ln X_t$$
$$dY_t = (\mu - \frac{1}{2}\sigma^2)dt + \sigma dB_t$$

- *The solution for $X(t)$ is given by*

$$X(t) = X(0)\exp(\int_0^t [\mu(s) - \frac{1}{2}\sigma(s)^2]ds + \int_0^t \sigma(s)dB(s)).$$

- *Particularly, if $\mu(t) = 0$ and $\sigma(t)$ is a constant, then*

$$X(t) = X(0)\exp(-\frac{1}{2}\sigma^2 t + \sigma B(t))$$

is a martingale.

Proof. (1)(2) use $Y_t = f(X_t)\ln(X_t)$ and Ito rule, we have

$$\begin{aligned} dY_t &= \frac{\partial f}{\partial X_t} dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial X_t^2} dX_t dX_t \\ &= \frac{1}{X_t} dX_t - \frac{1}{2}\frac{1}{X_t^2}(\sigma X_t)^2 dt \\ &= (\mu - \frac{1}{2}\sigma^2)dt + \sigma dB_t \end{aligned}$$

Then Y_t will have solution

$$Y_t = Y_t + \int_0^t (\mu - \frac{1}{2}\sigma^2)ds + \int_0^t \sigma dB_s.$$

(3) We want to prove $E[X(t)|\mathcal{F}_s] = X(s)$, where $\mathcal{F}_t$ is the filtration associated with Brownian motion. See Lemma 19.7.2. □

Corollary 20.2.3.1 (state independent geometric SDE, conversion to driftless SDE). *Consider SDE for X*

$$dX(t) = \mu X(t)dt + \sigma X(t)dB(t)$$

with constant μ, σ, *and let* $Y = \exp(-\mu t)X$, *then the SDE for* Y *is*

$$dY = \sigma Y(t)dB(t)$$

with solution of $Y(t)$ *being an exponential martingale as*

$$Y(t) = Y(0)\exp(-\frac{1}{2}\sigma^2 t + \sigma B(t)).$$

Then, $X(t)$ *is given by*

$$X(t) = X(0)\exp(\mu t - \frac{1}{2}\sigma^2 t + \sigma B(t))$$

Proof. From Ito lemma, we have

$$dY = -\mu\exp(-\mu t)Xdt + \exp(-\mu t)dX = \sigma Y(t)dB(t).$$

The rest can be proved using above lemma. □

Corollary 20.2.3.2 (mean and variance of a state-independent geometric SDE). *Consider SDE for X*

$$dX(t) = \mu X(t)dt + \sigma X(t)dB(t)$$

with constant μ, σ. Then,

- $E[X(t)] = X(0)e^{\mu t}$
- $Var[X(t)] = X(0)^2 e^{2\mu t}(e^{\sigma^2 t} - 1)$

Proof. Note that

$$\ln(\frac{X(t)}{X(0)}) = (\mu - \frac{\sigma^2}{2})t + \sigma W_t.$$

That is, $\frac{X(t)}{X(0)} \sim LN((\mu - \frac{\sigma^2}{2})t, \sigma^2 t)$.Then we can use Definition 13.1.6. □

Remark 20.2.3 (some implications).

- For geometric Brownian motion, once discounted with drift term, it becomes a martingale.
- This result is consistent with the finance in which the stock price S under martingale measure has the same drift r as risk free rate. As a consequence, when S is discounted with the risk-free rate, it becomes a martingale. In otherwise, martingale measure is the measure that make S a martingale when discounted.

20.2.4.3 *Integral of state-independent linear arithmetic SDE*

Lemma 20.2.11 (Integral of state independent arithmetic SDE). *Let X_t be governed by stochastic differential equation*

$$dX_t = a(t)dt + b(t)dW(t).$$

Further define a integral

$$I(t, T) = \int_t^T X(s)ds.$$

It follows that

- $$X_s = X_t + \int_t^s a(u)du + \int_t^s b(u)dW(u),$$

  which is a Gaussian distribution with mean $X_0 + \int_0^t a(s)ds$ and variance $\int_0^t b^2(s)ds$.
- *$I(t, T)$ has explicit form*

  $$I(t, T) = X_t(T - t) + \int_t^T (T - u)a(u)du + \int_t^T (T - u)b(u)dW(u).$$

- *$I(t,T)$ is a Gaussian distribution with mean and variance given by*

$$M(t,T) = X_t(T-t) + \int_t^T (T-u)a(u)du,$$

$$V(t,T) = \int_t^T (T-u)^2 b^2(u)du.$$

- *If $b(u) = b_0, a(u) = a_0$, then*

$$M(t,T) = X_t(T-t) + \frac{1}{2}a_0(T-t)^2,$$

$$V(t,T) = \frac{1}{3}b_0(T-t)^2.$$

Proof. (1) See Lemma 20.2.9. (2)

$$\begin{aligned}
&\int_t^T X_s ds \\
=&\int_t^T X_t ds + \int_t^T \int_t^s a(u)duds + \int_t^T \int_t^s b(u)dW(u)ds \\
=&X_t(T-t) + \int_t^T \int_u^T a(u)dsdu + \int_t^T \int_u^T a(u)dsdW(u) \\
=&X_t(T-t) + \int_t^T (T-u)a(u)du + \int_t^T (T-u)b(u)dW(u)
\end{aligned}$$

where we changed the order of integral. (3)(4) Use Lemma 20.2.9 again, we can see that $I(t,T)$ is actually a Gaussian process. □

Lemma 20.2.12 (Integral of sum of two state independent arithmetic SDE). *Let $X_1(t), X_2(t)$ be governed by stochastic differential equations*

$$\begin{aligned}
dX_1(t) &= a_1(t)dt + b_1(t)dW_1(t) \\
dX_2(t) &= a_2(t)dt + b_2(t)dW_2(t) \\
E[dW_1 dW_2] &= \rho dt
\end{aligned}$$

Further define a integral

$$I(t,T) = \int_t^T X_1(s) + X_2(s)ds.$$

It follows that

- $$X_1(s) + X_2(s) = X_1(t) + X_2(t) + \int_t^s a_1(u) + a_2(u)du + \int_t^s b_1(u) + b_2(u)dW(u),$$
- *$I(t,T)$ has explicit form*
$$I(t,T) = (X_1(t) + X_2(t))(T-t) + \int_t^T (T-u)a(u)du + \int_t^T (T-u)b(u)dW(u),$$
where
$$a(u) = a_1(u) + a_2(u), b(u) = \sqrt{b_1(u)^2 + b_2(u)^2 + 2\rho b_1(u)b_2(u)}$$
- *$I(t,T)$ is a Gaussian distribution with mean and variance given by*
$$M(t,T) = X_t(T-t) + \int_t^T (T-u)a(u)du,$$
$$V(t,T) = \int_t^T (T-u)^2 b^2(u)du.$$
- *If $b_1(u) = b_{10}, b_2(u) = b_{20}, a_1(u) = a_{10}, a_2(u) = a_{20}$, then*
$$M(t,T) = X_t(T-t) + \frac{1}{2}a_0(T-t)^2,$$
$$V(t,T) = \frac{1}{3}b_0(T-t)^2,$$
where
$$a_0 = a_{10} + a_{20}, b_0 = \sqrt{b_{10}^2 + b_{20}^2 + 2\rho b_{10}b_{20}}$$

Proof. Note that

$$\begin{aligned} d(X_1(t) + X_2(t)) &= (a_1(t) + a_2(t))dt + b_1(t)dW_1(t) + b_2(t)dW_2(t) \\ dZ(t) &= a(t)dt + \sqrt{b_1(u)^2 + b_2(u)^2 + 2\rho b_1(u)b_2(u)}dW_3 \end{aligned}$$

where $Z(t) \triangleq X_1(t) + X_2(t)$, the W_3 is a new Brownian motion. We arrive at

$$b_1(t)dW_1(t) + b_2(t)dW_2(t) = \sqrt{b_1(u)^2 + b_2(u)^2 + 2\rho b_1(u)b_2(u)}dW_3,$$

via the fact that two independent Gaussian random variable will sum to another Gaussian random variable. Then we use Lemma 20.2.11. □

Lemma 20.2.13 (Integral of sum of multiple state independent arithmetic SDE). *Let $X_1(t), X_2(t), ..., X_n$ be governed by stochastic differential equations*

$$\begin{aligned} dX_1(t) &= a_1(t)dt + b_1(t)dW_1(t) \\ dX_2(t) &= a_2(t)dt + b_2(t)dW_2(t) \\ &\cdots\cdots \\ dX_n(t) &= a_n(t)dt + b_n(t)dW_n(t) \\ E[dW_i dW_j] &= \rho_{ij}dt \end{aligned}$$

Further define a integral

$$I(t,T) = \int_t^T X_1(s) + X_2(s) + \cdots + X_n(s)ds.$$

It follows that

- $$X_1(s) + X_2(s) + \cdots + X_n(s) = \sum_{i=1}^n X_i(t) + \int_t^s \sum_{i=1}^n a_i(u)du + \int_t^s \sum_{i=1}^n b_i(u)dW(u),$$
- *$I(t,T)$ has explicit form*

  $$I(t,T) = (\sum_{i=1}^n X_i(t))(T-t) + \int_t^T (T-u)a(u)du + \int_t^T (T-u)b(u)dW(u),$$

  where

  $$a(u) = \sum_{i=1}^n a_i(u), b(u) = \sqrt{\sum_{i=1}^n b_i(u)^2 + 2\sum_{1\leq i<j\leq n} \rho_{ij}b_i(u)b_j(u)}$$
- *$I(t,T)$ is a Gaussian distribution with mean and variance given by*

  $$M(t,T) = X_t(T-t) + \int_t^T (T-u)a(u)du,$$

  $$V(t,T) = \int_t^T (T-u)^2 b^2(u)du.$$

20.2.4.4 *Multiple dimension extension*

Lemma 20.2.14 (multi-dimensional state independent/general arithmetic SDE). *[2, p. 146][4, p. 116]Consder a N dimensional stochastic differential equation(SDE) given by*

$$dX_i = a_i(t)dt + b_i(t)dW_i(t),$$

where $E[dW_i dW_j] = \rho_{ij}dt, E[dWdW^T] = \Sigma dt$*. It follows that*

- *The solution for* $X_i(t), i = 1, 2, ..., N$ *is given by*

$$X_i(t) = X_i(0) + \int_0^t a_i(s)ds + \int_0^t b_i(s)dW_i(s),$$

which is a ***Gaussian distribution*** *with mean* $X_i(0) + \int_0^t a_i(s)ds$ *and variance* $\int_0^t b_i^2(s)ds$.

- *The covariance structure for different* $X_i(t), X_j(s), s \geq t$ *is given by*

$$Cov(X_i(t), X_j(s)) = \int_0^t b_i(u)b_j(u)\rho_{ij}du.$$

Proof. (1) See Lemma 20.2.9. (2)

$$\begin{aligned} Cov(X_i(t), X_j(s)) &= \int_0^t \int_0^s b_i(u)b_j(v)dW_i(u)dW_j(v) \\ &= \int_0^t \int_0^s b_i(u)b_j(v)\rho_{ij}\delta(u-v)du \\ &= \int_0^t b_i(u)b_j(u)\rho_{ij}du \end{aligned}$$

□

Lemma 20.2.15 (general multi-dimensional geometric SDE). *[4, p. 116] Consider a N-dimensional SDE*

$$dX_i(t) = \mu_i(t)X_i(t)dt + \sigma_i(t)X_i(t)dW_i(t),$$

where It follows that

- *The solution for* $X_i(t), i = 1, 2, ..., N$*, is given by*

$$X_i(t) = X_i(0)\exp(\int_0^t [\mu_i(s) - \frac{1}{2}\sigma_i(s)^2]ds + \int_0^t \sigma_i(s)dW_i(s)).$$

- *Particularly, if* $\mu_i(t) = 0$ *and* $\sigma_i(t)$ *is a constant, then*

$$X_i(t) = X_i(0)\exp(-\frac{1}{2}\int_0^t \sigma_i^2(s)ds + \int_0^t \sigma_i(s)dW_i(s))$$

is a martingale.

- *The covariance structure for different* $X_i(t), X_j(s), s \geq t$ *is given by*

$$\begin{aligned}&Cov(X_i(t), X_j(s))\\ =&X_i(0)X_j(0)\exp(m_i(t) + m_j(s) + \frac{1}{2}(\Sigma_{ii}(t,t) + \Sigma_{jj}(s,s)))(\exp(\Sigma_{ij}(t,s)) - 1),\end{aligned}$$

where

$$m_i(t) = \int_0^t [\mu_i(u) - \frac{1}{2}\sigma_i(u)^2]du,$$

$$\Sigma_{ij}(t,s) = \int_0^t \sigma_i(u)\sigma_j(u)dt.$$

- *Particularly, if* $\mu_i(t) = 0$ *and* $\sigma_i(t)$ *is a constant, then*

$$Cov(X_i(t), X_j(s)) = X_i(0)X_j(0)\exp(\Sigma_{ij}(t,s)) - 1.$$

$$E(X_i(t), X_j(s)) = X_i(0)X_j(0)\exp(\Sigma_{ij}(t,s)),$$

Proof. (1)(2) See Lemma 20.2.10. (3) Lemma 13.1.17 (4) Note that when $\mu_i = 0$, we have $m_i(t) + \frac{1}{2}\Sigma_{ii}(t,t) = 0$. □

20.2.5 Exact SDE

Definition 20.2.3 (exact SDE). *[2, p. 151] The SDE*

$$dX_t = a(t, W_t)dt + b(t, W_t)dW_t$$

is called exact if there is a differentiable function $f(t, W_t)$ *such that*

$$a(t, W_t) = f_t + \frac{1}{2}f_{WW}, b(t, W_t) = f_W$$

Lemma 20.2.16. *The solution to an exact SDE is given as*

$$X_t = f(t, W_t) + C$$

Proof. Use Ito's lemma, we have

$$dX_t = df = f_t dt + f_W dW_t + \frac{1}{2} f_{WW} dt$$

□

Remark 20.2.4. Not every SDE is exact. With a, b given, we can try to first solve for f(not necessarily solvable). If we can get f then obtain an easy way to solve SDE.

Example 20.2.7. We have

$$dX_t = e^t(1 + W_t^2)dt + (1 + 2e^t W_t)dW_t$$

We can find $f(t, W_t) = W_t + e^t W_t^2$

Theorem 20.2.4 (exact SDE criterion,necessary condition). *[2, p. 152] If SDE is exact, then*

$$a_x = b_t + \frac{1}{2} b_{xx}$$

20.2.6 Calculation mean and variance from SDE

Theorem 20.2.5 (Fubini's theorem). *[6, p. 53]Let $X(t)$ be a stochastic process with continuous sample paths, then*

$$\int_0^T E[|X(t)|]dt = E[\int_0^T |X(t)| \, dt]$$

furthermore if this quantity is finite, then

$$\int_0^T E[X(t)]dt = E[\int_0^T X(t)dt]$$

Remark 20.2.5. This theorem gives us the condition to exchange expectation and integral.

Given a SDE

$$X_t = X_0 = \int_0^t a(X_s, s)ds + \int_0^t b(X_s, s)dW_s$$

we are just interested in the mean and variance of X_t. We can use the fact that **expectation of Ito integral is zero** to simplify our calculation:

$$E[X_t] = X_0 + \int_0^t E[a(X_s, s)]ds$$

where the integral of expectation and integral is justified by **Fubini's theorem**.

Using the fundamental theorem of calculus, we know that

$$\frac{dE[X_t]}{dt} = E[a(X_t, t)]$$

Lemma 20.2.17. *Let* $dX_t = a(t)X_t dt + c(t)dt + b(t)dW(t)$, *then*

$$E[X_t] = \Phi(t, 0)X_0 + \int_0^t \Phi(t, \tau)c(\tau)d\tau,$$

where

$$\Phi(t, s) = \exp(\int_s^t a(u)du)$$

Proof. It is easy to find the governing equation for $E[X_t]$ is

$$dE[X_t]/dt = a(t)E[X_t] + c(t),$$

then use solution methods in linear dynamical system to solve. □

Lemma 20.2.18. *Let* $dX_t = a(t)X_t dt + b(t)dW(t)$, *then*

$$E[X_t] = \Phi_1(t, 0)X_0$$

$$Var[X_t] = \int_0^t \Phi_2(t, \tau)b(\tau)^2 d\tau$$

where

$$\Phi_1(t, s) = \exp(\int_s^t a(u)du), \Phi_2(t, s) = \exp(\int_s^t 2a(u)du)$$

Proof. Let $Y_t = X_t^2$, then

$$dY_t = 2X_t dX_t + b(t)^2 dt = 2aY_t dt + b^2(t)dt + bdW(t),$$

use the above lemma, we have

$$E[Y_t] = E[X_t^2] = \Phi_2(t,0))X_0^2 + \int_0^t \Phi_2(t,\tau)b(\tau)^2 d\tau$$

Use $Var[X_t] = E[X_t^2] - E[X_t]^2 = \int_0^t \Phi_2(t,\tau)b(\tau)^2 d\tau$. □

Remark 20.2.6. We can also obtain the result using solutions to Ornstein-Uhlenbeck process Lemma 20.3.1.

20.2.7 Multi-dimensional Ito stochastic differential equations

Definition 20.2.4 (Multi-dimensional Ito SDE). *A system of N-dimensional Ito SDEs is defined as*

$$dx_i(t) = \mu_i(\boldsymbol{x}(t),t)dt + \sum_{j=1}^{M} \sigma_{ij}(\boldsymbol{x}(t),t)dw_j(t), \forall i = 1,...,d$$

where the dynamical system is driven by M Wiener process, and $dw_i(t_1)dw_j(t_2) = \delta_{ij}dt\delta(t_1 - t_2)$.

Note 20.2.1 (equivalence of two forms and redundancy of Brownian motion)**.**

- Consider the system of N-dimensional Ito SDEs:

  $$dx_i(t) = \mu_i(\boldsymbol{x}(t),t)dt + h_j(\boldsymbol{x}(t),t)dz_j(t), \forall i = 1,...,N$$

  where the dynamical system is driven by N Wiener process, and $dz_i(t_1)dz_j(t_2) = \rho_{ij}dt\delta(t_1 - t_2)$.
  To ensure the equivalence of the two forms, we have the following identity:

  $$\sigma\sigma^T = H\rho H^T$$

  If we know $H\rho H^T$, then we can calculate σ using Cholesky decomposition or eigendecomposition. If we know $\sigma\sigma^T$, then it is exactly the covariance matrix of N dimensional random vector dz. It is possible that the covariance matrix is non-singular.

- If $M > N$(more Brownian motions than Ito process) and assume σ has full row rank, then there are redundant Brownian motion, and we can restructure such that the system is driven by N independent Brownian motion.

20.3 Ornstein-Uhlenbeck(OU) process

20.3.1 OU process

20.3.1.1 *Constant coefficient OU process*

Definition 20.3.1 (Ornstein-Uhlenbeck process). *A stochastic process*

$$X_t = e^{-at}x_0 + \sigma \int_0^t e^{-a(t-s)} dB_s,$$

where a, σ, x_0 are constant parameters and B_t is the Brownian motion, is called Ornstein-Uhlenbeck process with parameter (a, σ) and initial value x_0.

The differential form of the OU process is given by

$$dX_t = \sigma dB_t - aX_t dt, X_0 = x_0.$$

Lemma 20.3.1 (OU process solution). *Consider the SDE*

$$dX_t = \sigma dB_t - aX_t dt$$

with $\sigma > 0, a > 0$, and initial condition $X_0 = x_0$.

It follows that

- *It has the solution*

$$X_t = x_0 e^{-at} + \int_0^t \exp(-a(t-s))\sigma dB_s.$$

- *X_t has Gaussian distribution, i.e.,*

$$X_t \sim N(x_0 e^{-at}, \frac{\sigma^2(1-e^{-2at})}{2a}).$$

- *X_t has the stationary distribution given by*

$$X_t \sim N(0, \frac{\sigma^2}{2a}).$$

Proof. (1)(2) Use $Y_t = X_t e^{at}$, then Ito rule gives

$$dY_t = aY_t + e^{at} dX_t = e^{at}\sigma dB_t$$

We have

$$Y_T - Y_0 = \int_0^T e^{at}\sigma dB_t \Leftrightarrow X_T = \exp(-aT)X_0 + \int_0^T e^{-a(T-t)}dB_t.$$

Use Lemma 20.1.2, we have

$$Y_T - Y_0 \sim N(0, \int_0^T (e^{at}\sigma)^2 dt).$$

Then

$$X^T \sim e^{-aT}N(X_0, \int_0^T (e^{at}\sigma)^2 dt)$$

simplifies to

$$X^T \sim N(X_0, e^{-2aT}\int_0^T (e^{at}\sigma)^2 dt).$$

(3) Take $t \to \infty$ will get the result. □

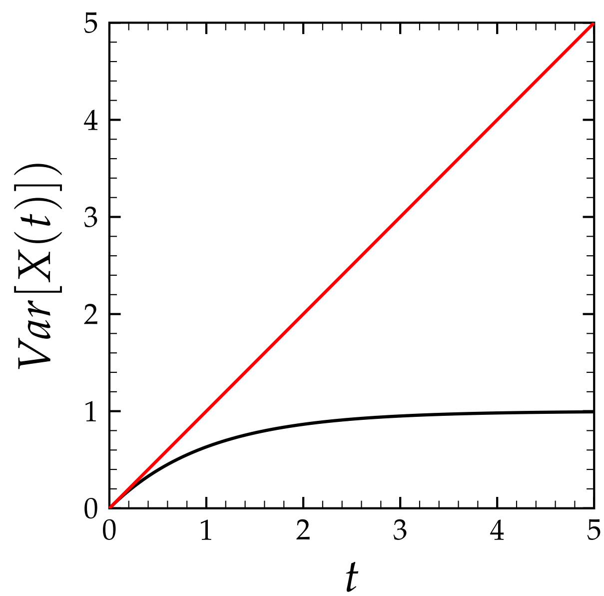


Figure 20.3.1: The variance function $Var[X(t)]$ for Brownian motion (red) and OU process(black) with $a = 0.5, \sigma = 1$.

Lemma 20.3.2 (constant shifted OU process). *Consider the constant shifted OU process*

$$dX_t = \sigma dB_t - a(X_t - \mu)dt$$

with $\sigma > 0, a > 0$, and initial condition $X_0 = x_0$.

- *It has the solution*

$$X_t \sim N((x_0-\mu)e^{-at}+\mu, \frac{\sigma^2(1-e^{-2at})}{2a})$$

and the stationary distribution is given as

$$X_t \sim N(\mu, \frac{\sigma^2}{2a}).$$

- *the constnat shifted OU process can be re-written as*

$$\begin{aligned} X_t &= Z_t + \mu \\ dZ_t &= \sigma dB_t - aZ_t dt \end{aligned}$$

Proof. (1) Use $Y_t = (X_t-\mu)e^{at}$. The rest is similar to Lemma 20.3.1. (2) Note that $dZ_t = dX_t + d\mu = dX_t$. Therefore

$$\begin{aligned} & dZ_t = \sigma dB_t - aZ_t dt \\ \implies & dX_t = \sigma dB_t - a(X_t-\mu)dt \end{aligned}$$

It can also be verified that:

$$X_t = \mu + Z_t, x_0 = \mu + z_0, Z_t \sim N(z_0 e^{-at}, \frac{\sigma^2(1-e^{-2at})}{2a})$$

gives

$$X_t \sim N(\mu + (x_0-\mu)e^{-at}, \frac{\sigma^2(1-e^{-2at})}{2a}).$$

□

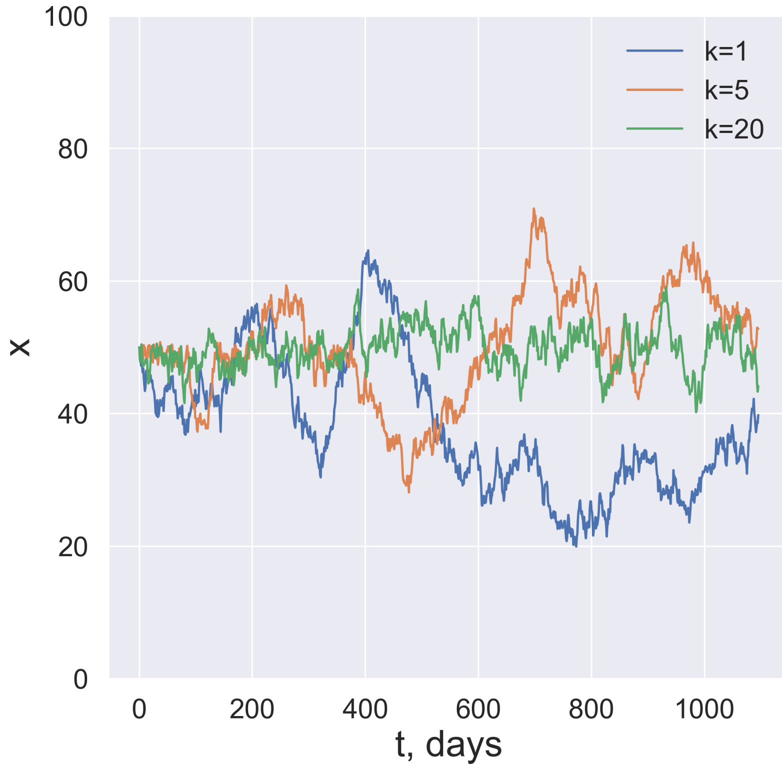


Figure 20.3.2: Representative trajectories from three OU processes with different k. k has the unit of inverse year. Mean level $\mu = 50$ and volatility $\sigma = 20$.

Lemma 20.3.3 (scaling property of OU process). *Consider the SDE*

$$dX(t) = \sigma dB(t) - a(X(t) - \mu)dt,$$

and let $X(t)$ be the solution. Then $Y(t) = \lambda X(mt)$ is the solution for

$$dY(t) = \sqrt{m}\lambda\sigma dB(t) - ma(Y(t) - \lambda\mu)dt.$$

Note that we interpret m as the time scaling factor and λ the spatial scaling factor.

Proof. Note that $X(mt)$ will satisfy

$$dX(mt) = \sigma dB(mt) - a(X(mt) - \mu)dmt,$$

or equivalently

$$dX(mt) = \sqrt{m}\sigma dB(t) - ma(X(mt) - \mu)dt.$$

Multiply both sides by λ, we have

$$d\lambda X(mt) = \sqrt{m}\lambda\sigma dB(t) - ma(\lambda X(mt) - \lambda\mu)dt.$$

Plug in $\lambda X(mt) = Y(t)$, we have

$$dY(t) = \sqrt{m}\lambda\sigma dB(t) - ma(Y(t) - \lambda\mu)dt.$$

□

Remark 20.3.1 (applications of scaling property). Suppose we have the dynamics of an asset with time unit day and value unit dollar, we can use the scaling property to find out the coefficients associated with time unit year and value unit JPY.

Lemma 20.3.4 (Stationary Gaussian process). *An Ornstein-Uhlenbeck process (a, σ) with Gaussian initial distribution $\eta \sim N(0, \sigma^2/2a)$(i.e., stationary distribution) is a strictly/weakly stationary Gaussian process.*

Proof. (1)

$$E[X_t] = E[e^{-at}\eta + \sigma\int_0^t e^{-a(t-s)}dW_s] = 0$$

since $E[\eta] = 0$ and $\int_0^t e^{-a(t-s)}dW_s$ is Ito integral(Theorem 20.1.4). (2) Let $s < t$, we have

$$\begin{aligned}
cov(X_t, X_s) = E[X_tX_s] &= e^{-a(s+t)}E[\eta^2] + \sigma^2 E[\int_0^s e^{-a(t-s)}dW_u \int_0^s e^{-a(t-m)}dW_m] \\
&= e^{-a(s+t)}\frac{\sigma^2}{2a} + \sigma^2\int_0^t e^{-2a(t-s)}dt \\
&= e^{-a(s+t)}\frac{\sigma^2}{2a} + \frac{\sigma^2}{2a}(e^{-2as} - 1) \\
&= e^{-a(s+t)}e^{-2as}\frac{\sigma^2}{2a} = \frac{\sigma^2}{2a}e^{-a(t-s)}
\end{aligned}$$

Note that a weakly stationary Gaussian process is strictly Gaussian process(Lemma 19.2.3).

□

20.3.1.2 *Time-dependent coefficient OU process*

Definition 20.3.2 (Time-dependent coefficient Ornstein-Uhlenbeck process). *A stochastic process with differential form*

$$dX_t = (\phi(t) - \lambda X_t)dt + \sigma dW_t,$$

where $\psi(t)$ is time dependent coefficient, a, σ, x_0 are constants, and W_t is Brownian motion., is called time-dependent coefficient Ornstein-Uhlenbeck process with parameter (a, σ) and initial distribution x_0.

Lemma 20.3.5. *Consider a stochastic process with differential form*

$$dX_t = (\psi(t) - \lambda X_t)dt + \sigma dW_t, X_t = x_0$$

where $\psi(t)$ is time dependent coefficient, a, σ are constants, and W_t is Brownian motion. It follows that

- *It has the equivalent form*

$$X_t = Y_t + \int_0^t \exp(-a(t-s))\psi(s)ds$$
$$dY_t = -aY_t dt + \sigma dW_t$$

- *It has solution*

$$X_t = x_0 \exp(-at) + \int_0^t \exp(-a(t-s))\psi(s)ds + \int_0^t \sigma \exp(-a(t-s))dW_t.$$

- *X_t has mean and covariance given by*

$$E[X_t] = x_0 \exp(-at) + \int_0^t \exp(-a(t-s))\psi(s)ds$$

$$Var[X_t] = \frac{\sigma^2(1-e^{-2at})}{2a}$$

- *X_t has Gaussian distribution at any t, we have*

$$X_t \sim N(x_0 \exp(-at) + \int_0^t \exp(-a(t-s))\psi(s)ds, \frac{\sigma^2(1-e^{-2at})}{2a})$$

- *$X_t, t \to \infty$ is generally not a stationary process since its mean depends on t.*

Proof. (1)Note that

$$\frac{d}{dt}\int_0^t \exp(-a(t-s))\psi(s)ds = -a\int_0^t \exp(-a(t-s)) + \psi(t);$$

Therefore,

$$dX_t = dY_t + \frac{d}{dt}\int_0^t \exp(-a(t-s))\psi(s)ds = (\psi(t) - aX_t)dt + \sigma dW_t.$$

(2)(3) Note that Y_t has solution and distribution

$$Y_t = x_0 e^{-at} + \int_0^t \exp(-a(t-s))\sigma dB_s,$$

$$Y_t \sim N(x_0 e^{-at}, \frac{\sigma^2(1-e^{-2at})}{2a}).$$

Then we use relation in (1). □

Lemma 20.3.6. *Let $\psi(t) = \phi_n, t_n \leq t < t_{n+1}, t_0 = 0$, then*

$$E[X(t_n)] \triangleq \mu(t_n) = x_0 \exp(-at_n) + \exp(-at_n) \sum_{k=1}^{n} \frac{\exp(at_k) - \exp(at_{k-1})}{a} \psi_{k-1}, n \geq 1$$

$$\exp(at)\mu_n = x(0) + \sum_{k=1}^{n} \frac{\exp(at_k) - \exp(at_{k-1})}{a} \psi_{k-1}, n \geq 1$$

$$\begin{aligned}
\exp(at)\mu_n &= x(0) + \sum_{k=1}^{n} \frac{\exp(at_k) - \exp(at_{k-1})}{a} \psi_{k-1}, n \geq 1 \\
a\exp(at)\mu_n &= ax(0) + \sum_{k=1}^{n} \exp(at_k) - \exp(at_{k-1})\psi_{k-1} \\
a\exp(at)\mu_{n-1} &= ax(0) + \sum_{k=1}^{n-1} \exp(at_k) - \exp(at_{k-1})\psi_{k-1}, \\
\implies \psi_{n-1} &= \lambda \frac{\exp(at_n)\mu_n - \exp(\lambda t_{n-1})\mu_{n-1}}{\exp(at_n) - \exp(at_{n-1})}
\end{aligned}$$

20.3.1.3 *Integral of OU process*

Lemma 20.3.7 (integral of OU process). *Consider an OU process given by*

$$dx(t) = -ax(t)dt + \sigma dW(t), x(0) = x_0$$

where $a\sigma$ are constants,W is a Brownian motion. For each t, T, the random variable

$$I(t,T) = \int_t^T x(u)du$$

conditioned to the σ field $\mathcal{F}_t$ is normally distributed with mean $M(t,T)$ and variance $V(t,T)$, respectively given by

$$M(t,T) = \frac{1-\exp(-a(T-t))}{a}x(t),$$

and variance

$$V(t,T) = \frac{\sigma^2}{a^2}(T-t+\frac{2}{a}\exp(-a(T-t)) - \frac{1}{2a}\exp(-2a(T-t)) - \frac{3}{2a}).$$

Proof. See the proof of Lemma 20.3.8. □

Lemma 20.3.8 (integral of sum of two OU process). *[7, p. 145][8, p. 64] Consider two OU processes given by*

$$\begin{aligned} dx_1(t) &= -a_1x_1(t)dt + \sigma_1 dW_1(t), x_1(0) = x_{10} \\ dx_2(t) &= -a_2x_2(t)dt + \sigma_2 dW_1(t), x_2(0) = x_{20} \end{aligned}$$

where $a_1, a_2, \sigma_1, \sigma_2$ are constant, $\psi(t)$ is a time-dependent function, and W_1, W_2 are correlated Brownian motion such that

$$dW_1(t)dW_2(t) = \rho dt.$$

For each t, T, the random variable

$$I(t,T) = \int_t^T (x_1(u) + x_2(u))du$$

conditioned to the σ field $\mathcal{F}_t$ is normally distributed with mean $M(t,T)$ and variance $V(t,T)$, respectively given by

$$M(t,T) = \frac{1-\exp(-a_1(T-t))}{a_1}x_1(t) + \frac{1-\exp(-a_2(T-t))}{a_2}x_2(t),$$

and variance

$$\begin{aligned}
V(t,T) = &\frac{\sigma_1^2}{a_1^2}(T - t + \frac{2}{a_1}\exp(-a_1(T-t)) - \frac{1}{2a_1}\exp(-2a_1(T-t)) - \frac{3}{2a_1}) \\
&+ \frac{\sigma_2^2}{a_2^2}(T - t + \frac{2}{a_2}\exp(-a_1(T-t)) - \frac{1}{2a_2}\exp(-2a_2(T-t)) - \frac{3}{2a_2}) \\
&+ \frac{2\rho\sigma_1\sigma_2}{a_1a_2}(T - t + \frac{\exp(-a_1(T-t)) - 1}{a_1} + \frac{\exp(-a_1(T-t)) - 1}{a_1} \\
&+ \frac{\exp(-(a_1+a_2)(T-t)) - 1}{a_1+a_2})
\end{aligned}$$

Proof. (1)Note that given the observation $x_1(t)$ at t, we have

$$x_1(u) = x_1(t)\exp(-a_1(u-t)) + \int_t^u \sigma\exp(-a_1(u-s))dW(s)$$

Therefore,

$$\begin{aligned}
&\int_t^T x_1(u)du \\
&= \int_t^T x_1(t)\exp(-a_1(u-t))du + \int_t^T\int_t^u \sigma\exp(-a_1(u-s))dW(s)du \\
&= x_1(t)\frac{1-\exp(-a_1(T-t))}{a_1} + \int_t^T\int_s^T \sigma\exp(-a_1(u-s))dudW(s) \\
&= x_1(t)\frac{1-\exp(-a_1(T-t))}{a_1} + \int_t^T \frac{\sigma_1}{a_1}(1-\exp(-a_1(T-s)))dW_1(s)
\end{aligned}$$

where we changed the order of integration. From this, we note that

$$E[\int_t^T x_1(u)du] = x_1(t)\frac{1-\exp(-a_1(T-t))}{a_1}.$$

Similarly, we can get eh expectation for $\int_t^T x_2(u)du$.

(2) To get the variance, we have

$$\begin{aligned}
&Var[\int_t^T x_1(u) + x_2(u)du] \\
=&Var[\int_t^T x_1(u)du] + Var[\int_t^T x_1(u)du] + 2Cov(\int_t^T x_1(u)du, \int_t^T x_1(u)du).
\end{aligned}$$

For $Var[\int_t^T x_1(u)du]$, we have

$$\begin{aligned}
&Var[\int_t^T x_1(u)du \\
&= E[\int_t^T \frac{\sigma_1}{a_1}(1-\exp(-a_1(T-s)))dW_1(s) \int_t^T \frac{\sigma_1}{a_1}(1-\exp(-a_1(T-s)))dW_1(s)] \\
&= \frac{\sigma_1^2}{a_1^2}(\int_t^T ds + \int_t^T \exp(-2a_1(T-s))ds - 2\int_t^T \exp(-a_1(T-s))ds) \\
&= \frac{\sigma_1^2}{a_1^2}(T-t+\frac{2}{a_1}\exp(-a_1(T-t)) - \frac{1}{2a_1}\exp(-2a_1(T-t)) - \frac{3}{2a_1}).
\end{aligned}$$

We can similarly evaluate $Var[\int_t^T x_2(u)du]$.

For $Cov[\int_t^T x_1(u)du, \int_t^T x_2(u)du]$, we have

$$\begin{aligned}
&Cov[\int_t^T x_1(u)du, \int_t^T x_2(u)du] \\
&= E[\int_t^T \frac{\sigma_1}{a_1}(1-\exp(-a_1(T-s)))dW_1(s) \int_t^T \frac{\sigma_1}{a_1}(1-\exp(-a_1(T-s)))dW_2(s)] \\
&= \frac{\rho\sigma_1\sigma_2}{a_1a_2}(\int_t^T (1-\exp(-a_1(T-s)-\exp(-a_1(T-s))+\exp(-(a_1+a_2)(T-s))))ds \\
&= \frac{2\rho\sigma_1\sigma_2}{a_1a_2}(T-t+\frac{\exp(-a_1(T-t))-1}{a_1}+\frac{\exp(-a_2(T-t))-1}{a_2} \\
&+ \frac{\exp(-(a_1+a_2)(T-t))-1}{a_1+a_2}).
\end{aligned}$$

□

Lemma 20.3.9 (integral of sum of multiple OU process). *[7, p. 145][8, p. 64] Consider n OU processes given by*

$$\begin{aligned}
dx_1(t) &= -a_1x_1(t)dt + \sigma_1 dW_1(t), x_1(0) = x_{10} \\
dx_2(t) &= -a_2x_2(t)dt + \sigma_2 dW_1(t), x_2(0) = x_{20} \\
&\cdots\cdots \\
dx_n(t) &= -a_nx_2(t)dt + \sigma_n dW_1(t), x_2(0) = x_{n0}
\end{aligned}$$

where $a_1, ..., a_n, \sigma_1, ..., \sigma_n$ are constants, and $W_1, W_2, ..., W_n$ are correlated Brownian motions such that

$$dW_i(t)dW_j(t) = \rho_{ij}dt.$$

For each t, T, the random variable

$$I(t,T) = \int_t^T (x_1(u) + x_2(u) + ... + x_n(u))du$$

conditioned on the σ field $\mathcal{F}_t$ is normally distributed with mean $M(t,T)$ and variance $V(t,T)$, respectively given by

$$M(t,T) = \sum_{i=1}^{n} \frac{1-\exp(-a_i(T-t))}{a_i} x_i(t),$$

and variance

$$V(t,T) = \sum_{i=1}^{n} \frac{\sigma_i^2}{a_i^2}(T-t+\frac{2}{a_i}\exp(-a_i(T-t)) - \frac{1}{2a_i}\exp(-2a_i(T-t)) - \frac{3}{2a_i})$$
$$+ \sum_{1\leq i<j\leq n} \frac{2\rho\sigma_i\sigma_j}{a_ia_j}(T-t+\frac{\exp(-a_i(T-t))-1}{a_i} + \frac{\exp(-a_j(T-t))-1}{a_j} + \frac{\exp(-(a_i+a_j)(T-t}{a_i+a_j}$$

20.3.2 Exponential OU process

Definition 20.3.3 (exponential constant coefficient Ornstein-Uhlenbeck process). *A stochastic process with differential form*

$$d(\ln X_t) = -a\ln X_t dt + \sigma dB_t, X_0 = x_0,$$

where a, σ, x_0 are constant parameters and B_t is the Brownian motion, is called exponential constant coefficient Ornstein-Uhlenbeck process with parameter (a,σ) and initial distribution x_0.

It also has the equivalent form

$$X_t = \exp(Y_t)$$
$$dY_t = -aY_t dt + \sigma dB_t, Y_0 = \ln x_0$$

Lemma 20.3.10 (exponential OU process solution). *Consider the SDE*

$$d(\ln X_t) = -a\ln X_t dt + \sigma dB_t, X_0 = x_0,$$

with $\sigma > 0, a > 0$, and initial condition $X_0 = x_0$.

It follows that

- *It has the solution*

$$X_t = \exp(Y_t), Y_t \sim\sim N(\ln(x_0)e^{-at}, \frac{\sigma^2(1-e^{-2at})}{2a}).$$

and the stationary distribution is given as

$$X_t = \exp(Y_t), Y_t \sim N(0, \frac{\sigma^2}{2a}).$$

- *(mean and variance property)*

$$E[X_t] = \exp(\mu_Y + \sigma_Y^2/2)$$
$$Var[X_t] = (\exp(\sigma_Y^2) - 1)\exp(2\mu_Y + \sigma_Y^2)$$

where

$$\mu_Y = \ln(x_0)e^{-at}, \sigma_Y^2 = \frac{\sigma^2(1-e^{-2at})}{2a}.$$

Proof. (1) Let $Y_t = \ln X_t$, then we have

$$dY_t = -aY_t dt + \sigma dB_t, Y_0 = \ln x_0.$$

From Lemma 20.3.1, we know that

$$Y_t \sim N(Y_0 e^{-at}, \frac{\sigma^2(1-e^{-2at})}{2a}).$$

(2) Use the property of log normal distribution(Lemma 13.1.12) □

Remark 20.3.2 (sanity check with Ito rule). Note that the SDE

$$d(\ln X_t) = -a \ln X_t dt + \sigma dB_t, X_0 = x_0,$$

will give the SDE for X_t via the equivalent form

$$X_t = f(Y_t) = \exp(Y_t)$$
$$dY_t = -aY_t dt + \sigma dB_t, Y_0 = \ln x_0.$$

Using Ito rule, we have

$$
\begin{aligned}
dX_t &= \frac{\partial f}{\partial Y_t} dY_t + \frac{1}{2}\frac{\partial^2 f}{\partial Y_t^2} dY_t dY_t \\
&= \exp(Y_t) dY_t + \frac{1}{2}\exp(Y_t)\sigma^2 dt \\
\implies dX_t/X_t &= dY_t + \frac{1}{2}\sigma^2 dt \\
dX_t/X_t &= (-a\ln X_t + \frac{1}{2}\sigma^2)dt + \sigma dB_t
\end{aligned}
$$

20.3.3 Parameter estimation for OU process

Note 20.3.1. The OU process

$$dX_t = k(\theta - X_t)dt + \sigma dW_t,$$

can be discretized at times $n\Delta t, n = 1, 2, ..., \infty$ which gives

$$X_{k+1} - X_k = k\theta\Delta t - kX_k\Delta t + \sigma(W_{k+1} - W_k),$$

or equivalently,

$$X_{k+1} = k\theta\Delta t - (k\Delta t - 1)X_k + \sigma\sqrt{\Delta t}\epsilon_k,$$

where $\epsilon_k \sim WN(0, 1)$.

The discrete-time form can be viewed as an AR(1) process, and least square method can be used to estimate k, θ, σ.

20.3.4 Multiple factor extension

Definition 20.3.4 (two-factor OU process). *The two-factor OU process is given by the following SDE*

$$
\begin{aligned}
r(t) &= x_1(t) + x_2(t) + \psi(t), r(0) = r_0 \\
dx_1(t) &= -a_1 x_1(t)dt + \sigma_1 dW_1(t), x_1(0) = x_{10} \\
dx_2(t) &= -a_2 x_2(t)dt + \sigma_2 dW_1(t), x_2(0) = x_{20}
\end{aligned}
$$

where $a_1, a_2, \sigma_1, \sigma_2, r_0$ are constant, $\psi(t)$ is a time-dependent function, and W_1, W_2 are correlated Brownian motion such that

$$dW_1(t)dW_2(t) = \rho dt.$$

Lemma 20.3.11 (basic properties). *Consider a The two-factor OU process is given by the following SDE*

$$\begin{aligned} r(t) &= x_1(t) + x_2(t) + \psi(t), r(0) = r_0 \\ dx_1(t) &= -a_1 x_1(t)dt + \sigma_1 dW_1(t), x_1(0) = x_{10} \\ dx_2(t) &= -a_2 x_2(t)dt + \sigma_2 dW_1(t), x_2(0) = x_{20} \end{aligned}$$

where $a_1, a_2, \sigma_1, \sigma_2, r_0$ are constant, $\psi(t)$ is a time-dependent function, and W_1, W_2 are correlated Brownian motion such that

$$dW_1(t)dW_2(t) = \rho dt.$$

It follows that

- *It has solution given by*

$$\begin{aligned} r(t) &= x_{10}\exp(-a_1 t) + x_{20}\exp(-a_2 t) \\ &\quad + \sigma_1 \int_0^t \exp(-a_1(t-s))dW_1(s) + \sigma_2 \int_0^t \exp(-a_2(t-s))dW_2(s) + \psi(t). \end{aligned}$$

-
$$E[r(t)] = x_{10}\exp(-a_1 t) + x_{20}\exp(-a_2 t) + \psi(t).$$

$$\begin{aligned} &Var[r(t)] \\ &= \frac{\sigma_1^2}{2a_1}(1-\exp(-2a_1 t)) + \frac{\sigma_2^2}{2a_2}(1-\exp(-2a_2 t)) + \frac{2\rho\sigma_1\sigma_2}{a_1+a_2}(1-\exp(-(a_1+a_2)t)). \end{aligned}$$

- *$r(t)$ has Gaussian distribution; that is,*

$$r(t) \sim N(E[r(t)], Var[r(t)]).$$

Proof. (1) From single factor OU process result(Lemma 20.3.1), we know that

$$x_1(t) = x_{10}\exp(-a_1 t) + \sigma_1 \int_0^t \exp(-a_1(t-s))dW_1(s).$$

Similarly, we can evaluate $x_2(t)$ and eventually $r(t)$. (2) The expectation can be easily evaluated based on the fact that Ito integral has zero mean. To evaluate the variance we have

$$
\begin{aligned}
Var[r(t)] &= Var[\sigma_1 \int_0^t \exp(-a_1(t-s))dW_1(s)] + Var[\sigma_2 \int_0^t \exp(-a_2(t-s))dW_2(s)] \\
&+ 2Cov(\sigma_1 \int_0^t \exp(-a_1(t-s))dW_1(s), \sigma_2 \int_0^t \exp(-a_2(t-s))dW_2(s)) \\
&= \int_0^t \sigma_1^2 \exp(-2a_1(t-s))ds + \int_0^t \sigma_2^2 \exp(-2a_1(t-s))ds + \int_0^t \sigma_1\sigma_2\rho \exp(-a_1(t-s)) \exp(-a_2(\\
&= \frac{\sigma_1^2}{2a_1}(1-\exp(-2a_1 t)) + \frac{\sigma_2^2}{2a_2}(1-\exp(-2a_2 t)) \\
&+ \frac{2\rho\sigma_1\sigma_2}{a_1+a_2}(1-\exp(-(a_1+a_2)t))
\end{aligned}
$$

where we use Ito isometry in the evaluation, for example,

$$
\begin{aligned}
&E[\sigma_1 \int_0^t \exp(-a_1(t-s))dW_1(s) \cdot \sigma_2 \int_0^t \exp(-a_2(t-s))dW_2(s)] \\
&= E[\sigma_1\sigma_2 \int_0^t \int_0^t \exp(-a_1(t-s)) \exp(-a_2(t-u))dW_1(s)dW_2(u)] \\
&= E[\sigma_1\sigma_2 \int_0^t \int_0^t \exp(-a_1(t-s)) \exp(-a_2(t-u))\rho dt\delta(u-s)] \\
&= E[\sigma_1\sigma_2\rho \int_0^t \exp(-(a_1+a_2)(t-s))\rho dt\delta(u-s)] \\
&= \frac{\rho\sigma_1\sigma_2}{a_1+a_2}(1-\exp(-(a_1+a_2)t))
\end{aligned}
$$

(3) The random variable $r(t) = x_1(t) + x_2(t)$ is a Gaussian process has been discussed in Theorem 19.3.2. □

20.4 Brownian motion variants

20.4.1 Brownian bridge

20.4.1.1 *Constructions*

Definition 20.4.1 (standard Brownian bridge). *A Brownian bridge is a stochastic process* $\{X_t, t \in [0,1]\}$ *with state space* $\mathbb{R}$ *that satisfies the following properties:*

- $X_0 = 0, X_1 = 0$ *almost surely.*
- X_t *is a Gaussian process.*
- $E[X_t] = 0$.
- $Cov(X_s, X_t) = \min(s,t) - st, \forall s,t \in [0,1]$.
- $Var[X_s] = s - s^2$.
- X_t *is almost surely continuous.*

Remark 20.4.1 (calculate covariance using conditional distribution). The joint distribution of (X_t, X_1) is a multivariate Gaussian with mean

$$\mu = (0,0)^T, \Sigma = \begin{bmatrix} t & t \\ t & 1 \end{bmatrix}$$

based on the property of standard Brownian motion(Lemma 19.5.1). Then

$$(X_t | X_1) \sim MN(0, t - t^2)$$

from Theorem 15.1.2. Similarly, the joint distribution of (X_s, X_t, X_1) is normal, and $(X_s, X_t | X_1) \sim MN(0, \min(s,t) - st)$.

Definition 20.4.2 (Brownian bridge, general state space). *A Brownian bridge is a stochastic process* $\{X_t, t \in [0,1]\}$ *with state space* $\mathbb{R}$ *that satisfies the following properties:*

- $X_0 = a, X_1 = b$ *almost surely.*
- X_t *is a Gaussian process.*
- $E[X_t] = (1-t)a + tb$.
- $Cov(X_s, X_t) = \min(s,t) - st, \forall s,t \in [0,1]$.
- X_t *is almost surely continuous.*

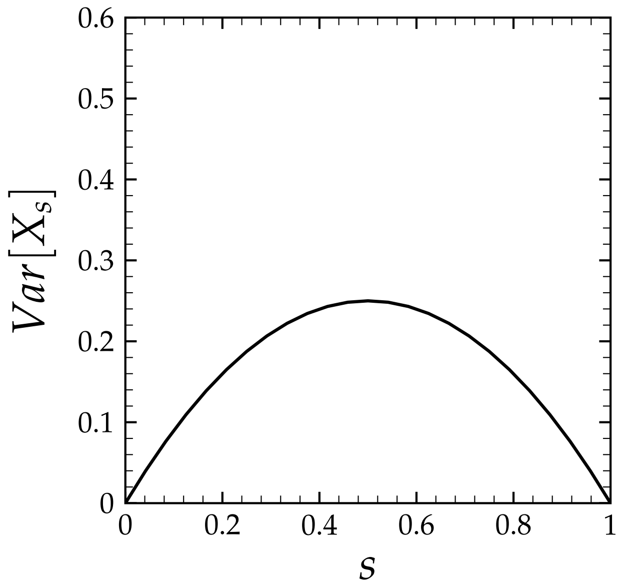


Figure 20.4.1: Variance of X_t in a Brownian bridge

Definition 20.4.3 (Brownian bridge, general temporal space). *A Brownian bridge is a stochastic process $\{X_t, t \in [p,q]\}$ with state space $\mathbb{R}$ that satisfies the following properties:*

- $X_p = 0, X_q = 0$ *almost surely.*
- X_t *is a Gaussian process.*
- $E[X_t] = 0$.
- $$Cov(X_s, X_t) = \min(s-p, t-p) - \frac{(s-p)(t-p)}{q-p}, \forall s,t \in [p,q]$$
  .
- X_t *is almost surely continuous.*

Definition 20.4.4 (Brownian bridge, general state space and temporal space). *A Brownian bridge is a stochastic process $\{X_t, t \in [p,q]\}$ with state space $\mathbb{R}$ that satisfies the following properties:*

- $X_p = a, X_q = b$ *almost surely.*
- X_t *is a Gaussian process.*
- $E[X_t] = (1 - \frac{t-p}{q-p})a + \frac{t-p}{q-p}b$.
- $$Cov(X_s, X_t) = \min(s-p, t-p) - \frac{(s-p)(t-p)}{q-p}, \forall s,t \in [p,q]$$
  .

- *X_t is almost surely continuous.*

Lemma 20.4.1 (construction of standard Brownian bridge).

- *Suppose Z_t is a standard Brownian motion. Let $X_t = Z_t - tZ_1, t \in [0,1]$.Then X_t is a Brownian bridge process.*
- *Suppose that $\{Z_t, t \in [0,\infty)\}$ is standard Brownian motions. Define $X_1 = 0$, and*

$$X_t = (1-t)\int_0^t \frac{1}{1-s} dZ_s, t \in [0,1).$$

Then X_t is a Brownian Bridge. Moreover, the stochastic process has the differential form as

$$dX_t = dZ_t - \frac{X_t}{1-t} dt.$$

Proof. (1)(a)$X_0 = X_1 = 0$. (b) The random vector $(X_{t_1}, X_{t_2}, ..., X_{t_n})$ can be constructed using affine transformation using random vector $(Z_{t_1}, Z_{t_2}, ..., Z_{t_n}, Z_1)$. Therefore, X_t is also a Gaussian process.(Theorem 15.1.1). (c) $E[X_t] = 0$. (d) $Cov(Z_s - sZ_1, Z_t - tZ_1) = \min(s,t) - st$.(e) X_t is continuous since Z_t, tZ_1 is continuous.

(2) (d)Note that X_t is a zero Gaussian process(Lemma 20.1.2,Lemma 19.3.3). Then

$$Cov(X_t, X_s) = (1-t)(1-s)\int_0^s \frac{1}{(1-u)^2} du = s - st.$$

To prove the differential form, we have

$$\begin{aligned}
X_t &= (1-t)\int_0^t \frac{1}{1-s} dZ_s \\
dX_t &= \int_0^t \frac{1}{1-s} dZ_s d(1-t) + (1-t)d(\int_0^t \frac{1}{1-s} dZ_s) \\
&= -\int_0^t \frac{1}{1-s} dZ_s + (1-t)\frac{1}{1-t} dZ_t \\
&= -\frac{X_t}{1-t} + dZ_t
\end{aligned}$$

□

Lemma 20.4.2 (construction generalized Brownian bridge). *Let $W(t)$ be a standard Brownian motion.*

- *Fix $a \in \mathbb{R}, b \in \mathbb{R}$. We can construct the Brownian bridge from a to b on $[0,1]$ to be the process*
$$Y(t) = a + (b-a)t + X(t),$$
where $X(t)$ is a standard Brownian bridge from 0 to 0 in time $[0,1]$.
- *Fix $p, q \in \mathbb{R}$. We can construct the Brownian bridge from 0 to 0 on $[p,q]$ to be the process*
$$Y(t) = X(\frac{t-p}{q-p}),$$
where $X(t)$ is a standard Brownian bridge from 0 to 0 in time $[0,1]$.
- *Fix $a, b, p, q \in \mathbb{R}$. We can construct the Brownian bridge from a to b on $[p,q]$ to be the process*
$$Y(t) = a + (b-a)\frac{t-p}{q-p} + X(\frac{t-p}{q-p}),$$
where $X(t)$ is a standard Brownian bridge from 0 to 0 in time $[0,1]$.

Proof. (1) straight forward.(2) □

20.4.1.2 *Simulation*

Remark 20.4.2 (simulation of Brownian bridge). We can simulate a Brownian bridge by first simulating a Wiener process W_t and then using
$$X_t = W_t - tW_1$$
to construct.[9, p. 27]

20.4.1.3 *Applications*

Remark 20.4.3 (general remarks). A Brownian bridge is used when you know the values of a Wiener process at the beginning and end of some time period, and want to understand the probabilistic behavior in between those two time periods.

Remark 20.4.4 (Brownian bridge as interpolation method). Suppose we have generated a number of points $W(0), W(1), W(2), W(3)$, etc. of a Wiener process path by computer simulation. We can use Brownian bridge simulation will interpolate path between $W(1)$ and $W(2)$.

Remark 20.4.5 (applications of Brownian bridge in bond). In the case of a long-term discount bond with known payoff at final term, we need to simulate values of the asset over a longer period of time such that the stochastic process is conditional on reaching a given final state. For example, take the case of a discount bond such as a 10 year Treasury bond. If we model a discount bond price as a stochastic process, then this process should be tied to the final state of the process.

20.4.2 Geometric Brownian motion

Definition 20.4.5 (geometric Brownian motion). *Suppose Z_t is standard Brownian motion and $\mu \in \mathbb{R}, \sigma > 0$, then*

$$X_t = X_0 \exp((\mu - \frac{1}{2}\sigma^2)t + \sigma Z_t), t \in [0, \infty)$$

is a stochastic process called geometric Brownian motion with drift μ and volatility parameter σ. Moreover, X_t is the solution to the Ito stochastic differential equation given as

$$dX_t = \mu X_t dt + \sigma X_t dZ_t$$

Lemma 20.4.3 (distribution). *The geoemtric Brownian motion has the lognormal distribution with parameter $(\mu - \frac{1}{2}\sigma^2)t$ and $\sigma\sqrt{t}$. The pdf is given as*

$$f_t(x) = \frac{1}{\sqrt{2\pi t}\sigma x} \exp(-\frac{(\ln(x/x_0) - (\mu - \sigma^2/2)t)^2}{2\sigma^2 t})$$

Lemma 20.4.4 (expectation and variance). *Let S_t be a geometric Brownian motion with initial condition S_0, then*

- $E[S_t] = S_0 e^{\mu t}$
- $Var[S_t] = S_0^2 e^{2\mu t}(e^{\sigma^2 t} - 1)$

Proof. See Definition 13.1.6. □

Remark 20.4.6 (martingale property).

- If $\mu \neq 0$, then Geometric Brownian motion is **not** a martingale since $E[X_t]$ is not a constant.
- If $\mu = 0$, then it is an exponential martingale(Lemma 19.7.2).

Remark 20.4.7. The geometric Brownian motion is **Not a Gaussian process.**

20.5 Notes on bibliography

For treatment on Stratonovich integral, see [1].

For treatment on calculating mean and variance from SDE, see [2].

For treatment on the techniques for solving SDE, see [2][1][4].

For finance related treatment, see [4].

See [10] for treatment on Girsanov theory and Feynman-Kac connection.

BIBLIOGRAPHY

1. Mikosch, T. *Elementary stochastic calculus with finance in view* (World scientific, 1998).
2. calin, O. *An Introduction to Stochastic Calculus with Applications to finance* (Eastern Michigan University, 2012).
3. Miranda, H.-C. *Applied Stochastic Analysis lecture notes* (NewYork University, 2015).
4. Wiersema, U. F. *Brownian motion calculus* (John Wiley & Sons, 2008).
5. Chang, F.-R. *Stochastic optimization in continuous time* (Cambridge University Press, 2004).
6. Klebaner, F. C. *Introduction to Stochastic Calculus with Applications* (Imperial College Press, 2005).
7. Brigo, D. & Mercurio, F. *Interest rate models-theory and practice: with smile, inflation and credit* (Springer Science & Business Media, 2007).
8. McInerney, D. & Zastawniak, T. *Stochastic Interest Rates* ISBN: 9781107002579 (Cambridge University Press, 2015).
9. Iacus, S. M. *Simulation and inference for stochastic differential equations: with R examples* (Springer Science & Business Media, 2009).
10. Steele, J. *Stochastic Calculus and Financial Applications* ISBN: 9780387950167 (Springer, 2001).

21

FOKKER-PLANCK EQUATION

21.1 Fokker-Planck equations

21.1.1 Formulations:one dimension

Lemma 21.1.1 (one dimensional Fokker-Planck equation). *[1, p. 121][2, p. 79] The Fokker-Planck equation is associated with an Ito SDE*

$$dx = \mu(x,t)dt + \sigma(x,t)dW_t,$$

where W_t is a Wiener process, is given as

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(\mu(x)p) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma(x,t)^2 p),$$

note that μ, σ can have dependence on both x and t.

Proof. Let $f(x)$ be an arbitrary smooth function of x, then

$$\int f(x)p(x,t)dx = \langle f \rangle \implies \int f(x)\frac{\partial p}{\partial t}dx = \left\langle \frac{df}{dt} \right\rangle.$$

Note that from Ito rule, we have

$$\begin{aligned} df &= \frac{df}{dx}dx + \frac{1}{2}\frac{d^2 f}{dx^2}\sigma^2 dt \\ df &= \frac{df}{dx}(\mu(x,t)dt + \sigma(x,t)dW_t) + \frac{1}{2}\frac{d^2 f}{dx^2}\sigma^2 dt \\ \implies \langle df \rangle &= (\frac{df}{dx}\mu(x,t) + \frac{1}{2}\frac{d^2 f}{dx^2}\sigma^2)dt \end{aligned}$$

Then

$$\begin{aligned} \int f(x)\frac{\partial p}{\partial t}dx &= \left\langle \frac{df}{dt} \right\rangle \\ &= \left\langle \frac{df}{dx}\mu(x,t) + \frac{1}{2}\frac{d^2 f}{dx^2}\sigma^2 \right\rangle \\ &= \int (\frac{df}{dx}\mu(x,t) + \frac{1}{2}\frac{d^2 f}{dx^2}\sigma^2)p(x,t)dx \end{aligned}$$

Integrating by parts(twice) and drop out the surface terms($p(x \to \infty) = 0$), we have

$$\int f(x)\frac{\partial p}{\partial t}dx = \int f(x)(-\frac{\partial \mu p}{\partial x} + \frac{1}{2}\frac{\partial^2 \sigma^2 p}{\partial x^2})dx$$

Because $f(x)$ is an arbitrary function, we have
$$\frac{\partial p}{\partial t} = -\frac{\partial \mu p}{\partial x} + \frac{1}{2}\frac{\partial^2 \sigma^2 p}{\partial x^2}$$
using fundamental lemma of calculus of variations. □

Remark 21.1.1. Fokker-Planck equation is also known as Kolmogorov forward equation.

Remark 21.1.2 (common boundary conditions). [3, p. 92]

- Adsorbing boundary condition. For example at interval $[0,1]$, we have $p(x = 0, t|x_0, t_0) = p(x = 1, t|x_0, t_0) = 0$.
- Reflecting boundary condition. For example at interval $[0,1]$, we have $\partial_x p(x = 0, t|x_0, t_0) = \partial_x p(x = 1, t|x_0, t_0) = 0$, which are interpreted as zero flux at the boundary.
- Periodic boundary condition. For example at interval $[0,1]$, we have $p(x = 0, t|x_0, t_0) = p(x = 1, t|x_0, t_0)$.

21.1.1.1 *Formulations: multiple dimension*

Lemma 21.1.2 (multi-dimensional Fokker-Planck equation). *[1, p. 121][2, p. 83] The Fokker-Planck equation is associated with a d dimensional Ito SDE*

$$dx_i = \mu_i(x,t)dt + \sum_{j=1}^{m} \sigma_{ij}^2(x,t)dW_j(t), i = 1, 2, ..., d$$

where $W_i(t), i = 1, ..., d$ is are multiple dimensional independent Wiener process, is given as

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{d} \frac{\partial}{\partial x_i}(\mu_i(x)p) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j}(\sum_{k=1}^{m} \sigma_{ik}\sigma_{jk} p).$$

Proof. Let G_t be a function of $x = (x_1, x_2, ..., x_d)$ given by $G_t = g(x_1, x_2, ..., x_d)$. From Ito rule, we have

$$\begin{aligned}
dG_t &= \sum_{i=1}^{d} \frac{\partial g}{\partial x_i} dx_i + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{\partial^2 g}{\partial x_i \partial x_j} dx_i dx_j \\
&= \sum_{i=1}^{d} \frac{\partial g}{\partial x_i}(\mu_i(x,t)dt + \sum_{j=1}^{m} \sigma_{ij}^2(x,t)dW_j(t)) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{\partial^2 g}{\partial x_i \partial x_j} \sum_{k=1}^{m} \sigma_{ik}\sigma_{jk} dt \\
&= (\sum_{i=1}^{d} \frac{\partial g}{\partial x_i}\mu_i(x,t) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{\partial^2 g}{\partial x_i \partial x_j} \sum_{k=1}^{m} \sigma_{ik}\sigma_{jk})dt + \sum_{i=1}^{d} \frac{\partial g}{\partial x_i} \sum_{j=1}^{m} \sigma_{ij}^2(x,t)dW_j(t).
\end{aligned}$$

Note that

$$\frac{d}{dt}E[G_t] = \frac{d}{dt}\int_{\mathbb{R}^d} p(x,t)g(x)dx = \int_{\mathbb{R}^d} \frac{d}{dt}p(x,t)g(x)dx.(*)$$

From another aspect, we have

$$\begin{aligned}
\frac{d}{dt}E[G_t] &= E[\frac{d}{dt}G_t] \\
&= E[(\sum_{i=1}^{d}\frac{\partial g}{\partial x_i}\mu_i(x,t) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\partial^2 g}{\partial x_i \partial x_j}\sum_{k=1}^{m}\sigma_{ik}\sigma_{jk})] \\
&= \int_{\mathbb{R}^d} p(x,t)(\sum_{i=1}^{d}\frac{\partial g}{\partial x_i}\mu_i(x,t) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\partial^2 g}{\partial x_i \partial x_j}\sum_{k=1}^{m}\sigma_{ik}\sigma_{jk})]dx \\
&= \int_{\mathbb{R}^d} p(x,t)(\sum_{i=1}^{d}\frac{\partial g}{\partial x_i}\mu_i(x,t) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\partial^2 g}{\partial x_i \partial x_j}\sum_{k=1}^{m}\sigma_{ik}\sigma_{jk})]dx \\
&= \int_{\mathbb{R}^d} [-\sum_{i=1}^{d}\frac{\mu_i(x,t)p(x,t)}{dx_i} + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\partial^2}{\partial x_i \partial x_j}\sum_{k=1}^{m}\sigma_{ik}\sigma_{jk}\sigma_{ik}\sigma_{jk}p(x,t)]g(x)dx
\end{aligned}$$

Combine with $(*)$, we have

$$\int_{\mathbb{R}^d}[\frac{d}{dt}p(x,t) - \sum_{i=1}^{d}\frac{\mu_i(x,t)p(x,t)}{dx_i} + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\partial^2}{\partial x_i \partial x_j}\sum_{k=1}^{m}\sigma_{ik}\sigma_{jk}\sigma_{ik}\sigma_{jk}p(x,t)]g(x)dx$$

holds for any function $g(x)$. Therefore

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{d}\frac{\partial}{\partial x_i}(\mu_i(x)p) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\partial^2}{\partial x_i \partial x_j}(\sum_{k=1}^{m}\sigma_{ik}\sigma_{jk}p).$$

□

21.1.2 Steady state & detailed balance

Lemma 21.1.3 (1d steady state solution). *Given a 1d Fokker-Planck*

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(\mu(x)p) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma(x,t)^2 p),$$

the steady state distribution p^{eq} will satisfy

$$-\frac{\partial}{\partial x}(\mu(x)p) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma(x,t)p) = 0.$$

Let $J = up - \frac{1}{2}\frac{\partial}{\partial x}(\sigma(x,t)^2 p)$, then the steady state solution is satisfying

$$\frac{\partial}{\partial x}J = 0.$$

In particular,

- *If domain $D = [a,b]$, and the boundary are reflecting, then we have stronger result of $J = 0$.*
- *If domain $D = [a,b]$, and the boundary are periodic, then we have stronger result of $J = const$, const might not be zero.*

Proof. On 1D, $\partial_x J = 0 \implies J = const$. □

Lemma 21.1.4 (2d/3d steady state solution). *Given a 2d/3d Fokker-Planck equation*

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{d}\frac{\partial}{\partial x_i}(\mu_i(x)p) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\partial^2}{\partial x_i \partial x_j}(\sum_{k=1}^{m}\sigma_{ik}\sigma_{jk}p),$$

the steady state solution is given as

$$-\sum_{i=1}^{d}\frac{\partial}{\partial x_i}(\mu_i(x)p) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\partial^2}{\partial x_i \partial x_j}(\sum_{k=1}^{m}\sigma_{ik}\sigma_{jk}p)$$

Let $Q_i = \sum_k \sigma_{ik}^{-1}(2\mu_k - 2\sum_j \frac{\partial}{\partial x_j}\sigma_{kj})$

21.1.3 Averages and adjoint operator

Lemma 21.1.5. *In the function space of square-integrable functions and the scalar product is defined by*

$$\langle f, g \rangle = \int_a^b f(x)\bar{g(x)}dx.$$

Consider a linear differential operator T given as

$$Tu = \sum_{k=0}^{n} a_k(x) \frac{d^k}{dx^k} u.$$

If f or g vanishes for $x \to a, b$, then the adjoint of T is

$$T^{\dagger} u = \sum_{k=0}^{n} (-1)^k \frac{d^k}{\overline{a_k(x)} dx^k} u$$

Proof. Use integration-by-parts. □

Example 21.1.1. The Sturm-Liouville operator is defined as

$$Lu = -(pu')' + qu.$$

It can be showed that $Lu = L^{\dagger} u, L = L^{\dagger}$.

Corollary 21.1.0.1 (adjoint operators).

-
$$Af = b(x) \frac{d}{dx} f(x), A^{\dagger} f = -\frac{d}{dx} b(x) f(x)$$
that is,
$$\int_{-\infty}^{\infty} g(x) b(x) \frac{d}{dx} f(x) dx = -\int_{-\infty}^{\infty} f(x) \frac{d}{dx} b(x) f(x) dx$$
-
$$Af = D(x) \frac{d^2}{dx^2} f(x), A^{\dagger} f = \frac{d^2}{dx^2} D(x) f(x)$$
-
$$Af = \frac{d}{dx} D(x) \frac{d}{dx} f(x), A^{\dagger} f = \frac{d}{dx} D(x) \frac{d}{dx} f(x) c$$

Proof. (1) use integration by parts, note that $\lim_{x\to\infty} f(x) = 0$.(2)(3) use integration by parts, note that $\lim_{x\to\infty} f(x) = 0$, $\lim_{x\to\infty} \frac{d}{dx} f(x) = 0$. □

Corollary 21.1.0.2 (Adjoint of Fokker-Planck operator). *Define*

$$Af = \sum_i b_i(x) \frac{\partial}{\partial x} f(x) + \frac{1}{2} \sum_{i,j} (\sigma(x)\sigma(x)')_{ij} \frac{\partial^2}{\partial x_i \partial x_j} f(x)$$

and its adjoint is

$$A^\dagger f = -\sum_i \frac{\partial}{\partial x} b_i(x) f(x) + \frac{1}{2}\sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} (\sigma(x)\sigma(x)')_{ij} f(x)$$

21.1.4 Backward equation

Lemma 21.1.6. *The backward Fokker-Planck equation is given as*

$$\frac{\partial}{\partial t_1} P(t_2, y_2|t_1, y_1) = -A(y_1, t_1)\frac{\partial}{\partial y_1} P(t_2, y_2|t_1, y_1) - \frac{1}{2} B(y_1, t_1) \frac{\partial^2}{\partial y_1^2} P(t_2, y_2|t_1, y_1),$$

where

$$\int (z - y_1) P(t_1, z|t - \Delta t, y_1) dz = \Delta t A(y_1, t_1 - \Delta t)$$

and

$$\int (z - y_1)^2 P(t_1, z|t - \Delta t, y_1) dz = \Delta t B(y_1, t_1 - \Delta t)$$

.

Proof. From Chapman-Kolmogorov equation, we have

$$P(t_2, y_2|t - \Delta t, y_1) = \int P(t_2, y_2|t_1, z) P(t_1, z|t_1 - \Delta t, y_1) dz.$$

Using Taylor expansion, we have

$$P(t_2, y_2|t_1, z) = P(t_2, y_2|t_1, y_1) + \frac{\partial}{\partial y_1} P(t_2, y_2|t_1, y_1)(z - y_1) + \frac{1}{2}\frac{\partial^2}{\partial y_1^2}(z - y_1)^2 + O(\int (z - y_1)^3.$$

Plug in, we have

$$\begin{aligned} P(t_2, y_2|t_1 - \Delta t, y_1) = {} & P(t_2, y_2|t_1, y_1) - \Delta t \frac{\partial}{\partial t_1} P(t_2, y_2|t_1, y_1) \\ & + \frac{\partial}{\partial y_1} P(t_2, y_2|t_1, y_1) \int (z - y_1) P(t_1, z|t - \Delta t, y_1) dz \\ & + \frac{1}{2}\frac{\partial^2}{\partial y_1^2} \int (z - y_1)^2 P(t_1, z|t_1 - \Delta t, y_1) dz + \\ & + O(\int (z - y_1)^3 P(t_1, z|t - \Delta t, y_1) dz) \end{aligned}$$

Let $\int (z-y_1)P(t_1,z|t-\Delta t, y_1)dz = \Delta t A(y_1, t_1-\Delta t)$ and $\int (z-y_1)^2 P(t_1,z|t-\Delta t, y_1)dz = \Delta t B(y_1,t_1-\Delta t)$, we have

$$\frac{\partial}{\partial t_1}P(t_2,y_2|t_1,y_1) = -A(y_1,t_1)\frac{\partial}{\partial y_1}P(t_2,y_2|t_1,y_1) - \frac{1}{2}B(y_1,t_1)\frac{\partial^2}{\partial y_1^2}P(t_2,y_2|t_1,y_1)$$

□

Remark 21.1.3 (interpretation).

- The backward Fokker-Planck equation describes the dependence of $P(y_2,t_2|y_1,t_1)$ on the initial condition (y_1,t_1).
- To be well-posed, the backward Fokker-Planck equation needs a final condition rather than an initial condition.

Lemma 21.1.7 (Backward equation using adjoint operator). *If*

$$\frac{\partial}{\partial t}P(x,t|x_0,t_0) = \mathcal{L}_x P,$$

then

$$\frac{\partial}{\partial t_0}P(x,t|x_0,t_0) = \mathcal{L}_{x_0} P,$$

Proof. From Chapman-Kolmogorov equation, we have

$$P(x_1,t_1|x_0,t_0) = \int P(x_1,t_1|x,t)P(x,t|x_0,t_0)dx,$$

where $t_0 < t < t_1$. Take the derivative with respect to t, then

$$\int [\frac{\partial}{\partial t}P(x_1,t_2|x,t)]P(x,t|x_0,t_0) + \int P(x_1,t_2|x,t)\frac{\partial}{\partial t}P(x,t|x_0,t_0) = 0$$
$$\int [\frac{\partial}{\partial t}P(x_1,t_2|x,t)]P(x,t|x_0,t_0) + \int P(x_1,t_2|x,t)\mathcal{L}_x P(x,t|x_0,t_0) = 0$$
$$\int [\frac{\partial}{\partial t}P(x_1,t_2|x,t) - \mathcal{L}_x^\dagger P(x_1,t_2|x,t)]P(x,t|x_0,t_0) = 0$$

where $\mathcal{L}_x^\dagger$ is the adjoint operator satisfying

$$\int P(x_1,t_1|x,t)\mathcal{L}_x P(x,t|x_0,t_0)dx = \int [\mathcal{L}_x^\dagger P(x_1,t_1|x,t)]P(x,t|x_0,t_0)dx.$$

Then

$$\frac{\partial}{\partial t}P(x_1,t_2|x,t) - \mathcal{L}_x^\dagger P(x_1,t_2|x,t) = 0.$$

□

Lemma 21.1.8 (Backward equation on expectation quantity). *Let* $u(t,x) = E[g(X_T)|X(t)=x]$. *Assume*

$$\int P(t,y|t-h,x)(y-x)dy = b(x,t)h + o(h)$$

and

$$\int P(t,y|t-h,x)(y-x)^2 dy = \sigma(x,t)h + o(h),$$

and higher moment vanish as $h \to 0$. *Then* u *satisfy*

$$\frac{\partial u}{\partial t} + b(x)\frac{\partial u}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2 f}{\partial x^2},$$

with final condition $u(T,x) = g(x)$.

Proof. Note that

$$u(t,x) = E[g(X_T)|X(t)=x] = \int g(X_T)P(T,X_T|t,x)dX_T$$

and

$$u(t-h,x) = \int E[g(X_T)|X(t)=y]P(t,y|t-h,x)dy$$
$$u(t,x) = \int E[g(X_T)|X(t)=x]P(t,y|t-h,x)dy$$

Therefore,

$$\begin{aligned}\frac{u(t)-u(t-h)}{h} &= \frac{\int P(t,y|t-h,x)(u(t,y)-u(t,x)}{h} \\ &= \frac{1}{h}\int P(t,y|t-h,x)[(y-x)\frac{\partial u}{\partial x} + \frac{1}{2}(y-x)^2\frac{\partial^2 u}{\partial x^2}]\end{aligned}$$

□

Remark 21.1.4. Backward equation can also be derived using Feyman Kac theorem by assuming the stochastic process of $X(t)$ (see Theorem 21.3.4).

Lemma 21.1.9 (Backward equation on expectation quantity using adjoint operator). *Let* $u(t,x) = E[g(X_T)|X(t)=x]$. *Assume*

$$\frac{\partial}{\partial t}P(x,t|x_0,t_0) = \mathcal{L}_x P(x,t|x_0,t_0).$$

Then

$$\frac{\partial u}{\partial t} = \mathcal{L}_x^\dagger u$$

with final condition $u(T,x) = g(x)$.

Proof. Note that

$$u(t,x) = E[g(X_T)|X(t) = x] = \int g(X_T)P(T, X_T|t,x)dX_T$$

and Therefore,

$$\begin{aligned}\frac{\partial u(t,x)}{\partial t} &= \int g(X_T)\frac{\partial}{\partial t}P(T, X_T|t,x)dX_T \\ &= \int g(X_T)\mathcal{L}_x^\dagger P dX_T = \mathcal{L}_x^\dagger \int g(X_T)PdX_T = \mathcal{L}_x^\dagger u(x,t).\end{aligned}$$

□

21.1.5 Mean first passage time problem

Definition 21.1.1 (survival probability). *The survival probability up to t in domain* Ω *is given as*

$$G(t|x_0,t_0) = \int_\Omega dx P(x,t|x_0,t_0).$$

Or equivalently,

$$G(t|x_0,0) = Pr(\tau > t) = \int_t^\infty \rho(\tau|x_0,0)$$

where $\rho(\tau|x_0,0)$ *is the probability density for the first passage time* τ *starting from* x_0 *at* $t = 0$.

Lemma 21.1.10 (survival probability governing equation). *Assume*

$$\frac{\partial}{\partial t}P(x,t|x_0,t_0) = -\mathcal{L}_x P(x,t|x_0,t_0).$$

Then

$$\frac{\partial}{\partial t}G(t|x,0) = -\mathcal{L}_x^\dagger G(t|x,0),$$

where $\mathcal{L}_x^\dagger$ *is the adjoint operator of* $\mathcal{L}_x$.

Proof. Use the theorem(Lemma 21.1.9) and define $g(X_T)$ as the indicator function. □

Lemma 21.1.11. *Define $T(x)$ be the mean first passage time, we have*

$$T(x) = -\int_0^\infty t\partial_t G(t|x_0 = x, t_0 = 0)dt = \int_0^\infty G(t|x,0)dt.$$

The governing equation for $T(x)$ is

$$\mathcal{L}^\dagger T = -1$$

with boundary condition $T(\partial D) = 0$.

Proof. Use integration by parts, we have

$$T(x) = \int_0^\infty G(t|x,0)dt.$$

Then

$$\begin{aligned}\mathcal{L}^\dagger T = \mathcal{L}^\dagger \int_0^\infty G(t|x,0) = \int_0^\infty \mathcal{L}^\dagger G(t|x,0) &= \int_0^\infty \partial_t G(t|x,0)dt \\ &= G(\infty|x,0) - G(0|x,0) = 0 - 1 = -1.\end{aligned}$$

□

Remark 21.1.5. For the mean first passage time calculation using adjoint Fokker-Planck equation, see [4].

21.2 Smoluchowski/advection-diffusion equation

Definition 21.2.1 (Smoluchowski/advection-diffusion equation). *The advection-diffusion for density $p(x,t), x \in \mathbb{R}^N$ is given as*

$$\frac{\partial p}{\partial t} = -\nabla \cdot (vp) + \nabla \cdot D\nabla p$$

where $v \in \mathbb{R}^N, D \in \mathbb{R}^{N\times N}$. Note that we can interpret the flux vector as $j = vp - D\nabla p$.

Remark 21.2.1 (Relation to computational fluid mechanics)**.** We can use the PDE to describe the solute transport in the flowing solutions. The v is the velocity field of the solution, which is further determined by the Navier-Stokes equation. When the velocity field describes an incompressible flow, i.e. $\nabla \cdot v = 0$, we can have simplification as $\nabla \cdot (vp) = v \cdot \nabla p$.

Remark 21.2.2 (Relation to advection equation)**.**

The advection equation for a conserved quantity described by a scalar field ϕ is expressed mathematically by a continuity equation

$$\phi_t + \nabla \cdot (\phi u) = 0$$

where u is the flow vector field. If we assume the flow is incompressible, we have

$$\phi_t + u \cdot \nabla\phi = 0.$$

In particular, if the flow is steady, we have

$$u \cdot \nabla\phi = 0.$$

Lemma 21.2.1 (Relation to Fokker-Planck equation). *The equation*

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial x}(\frac{D(x)}{kT}\frac{dU(x)}{dx}p) + \frac{\partial}{\partial x}D(x)\frac{\partial p}{\partial x}$$

has the following equilibrium solution

$$p \propto \exp(-\frac{U}{kT}).$$

This PDE can also be written as the form of Fokker-Planck equation as

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial x}(\frac{D(x)}{kT}\frac{dU}{dx}p - p\frac{\partial D(x)}{\partial x}) + \frac{\partial^2}{\partial x^2}(D(x)p)$$

with the associated Ito SDE as

$$dx(t) = (-\frac{D(x)}{kT}\frac{dU}{dx} + \frac{\partial D(x)}{\partial x})dt + \sqrt{2D(x)}dw(t)$$

Proof. Set $\frac{\partial p}{\partial t} = 0$, then we have

$$\begin{aligned}
0 &= \frac{\partial}{\partial x}(\frac{D(x)}{kT}\frac{\partial U}{\partial x}p) + \frac{\partial}{\partial x}D(x)\frac{\partial p}{\partial x} \\
0 &= \frac{\partial}{\partial x}(\frac{D(x)}{kT}\frac{\partial U}{\partial x}p + D(x)\frac{\partial p}{\partial x}) \\
0 &= (\frac{D(x)}{kT}\frac{dU}{dx}p + D(x)\frac{\partial p}{\partial x}) \\
\implies 0 &= \frac{D(x)}{kT}\frac{\partial U}{\partial x}p + D(x)\frac{\partial p}{\partial x} \\
\frac{d\ln p}{dx} &= -\frac{1}{kT}\frac{dU}{dx} \\
p &\propto \exp(-\frac{U}{kT})
\end{aligned}$$

□

21.3 Feyman Kac theorem and backward equation

21.3.1 Feyman Kac theorem

Theorem 21.3.1 (Feyman Kac theorem). *Consider the 1D parabolic*

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2(S,t)\frac{\partial^2 V}{\partial S^2} + m(S,t)\frac{\partial V}{\partial S} - rV = 0.$$

The solution is given as

$$V(S_t, t) = E_Q[e^{\int_t^T -r(\tau)d\tau} V(S_T, T)|\mathcal{F}_t]$$

where S_t is a stochastic process, and Q is probability measure, under which the dynamics of S is given as

$$dS = mdt + \sigma dW_t$$

where W_t is the Wiener process. Moreover,

$$e^{\int_0^t r(\tau)d\tau} V(S_t, t)$$

is a martingale under the measure Q.

Proof.

$$\begin{aligned} d(e^{\int_0^t r(\tau)d\tau} V(S_t,t)) &= d(e^{\int_0^t r(\tau)d\tau})V(t) + e^{\int_0^t r(\tau)d\tau} dV(t) + d(e^{\int_0^t r(\tau)d\tau})dV \\ &= -e^{\int_0^t -r(\tau)d\tau} r(t)Vdt + e^{-\int_0^t r(\tau)d\tau} dV. \end{aligned}$$

Use the fact that

$$dV = \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}dS + \frac{1}{2}(dS)^2\frac{\partial^2 V}{\partial S^2} = \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}mdt + \sigma\frac{\partial V}{\partial S}dW_t + \frac{1}{2}(\sigma)^2\frac{\partial^2 V}{\partial S^2}dt.$$

Then plug in dV, we have

$$\begin{aligned} d(e^{\int_0^t r(\tau)d\tau} V(S_t,t)) &= e^{\int_0^t -r(\tau)d\tau}(-r(t)Vdt + \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}mdt + \sigma\frac{\partial V}{\partial S}dW_t + \frac{1}{2}(\sigma)^2\frac{\partial^2 V}{\partial S^2}dt) \\ &= e^{\int_0^t -r(\tau)d\tau}\sigma\frac{\partial V}{\partial S}dW_t \end{aligned}$$

Therefore, $e^{\int_0^t r(\tau)d\tau}V(S_t,t))$ is a martingale. Then we can easily show using martingale property:

$$V(s) = E_Q[e^{-\int_s^t r(\tau)d\tau}V(S_t,t)|\mathcal{F}_s].$$

□

Remark 21.3.1 (interpretation and financial applications in path-independent derivatives).

- Feyman Kac theorem shows that certain types of parabolic equation can be solved using stochastic differential equation method(by simulating trajectories and take expectations.) **Note that in parabolic differential equation S is not a random variable, S is simply a variable.**

Remark 21.3.2 (special case of $r = 0$). When $r = 0$, the dynamics of S will not change(i.e. Q will not change), then $V(S_t,t)$ is a martingale. And the parabolic equation becomes Kologorov backward equation.

Theorem 21.3.2 (Feyman Kac theorem, multi-dimensional). *Consider the multidimensional parabolic*

$$\frac{\partial V}{\partial t} + \sum_{i=1}^N \mu_i \frac{\partial V}{\partial S} + \frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N \gamma_{ij}\frac{\partial^2 V}{\partial S_i \partial S_j} + -rV = 0$$

where $\gamma_{ij} = \sum_{k=1}^N \sigma_{ik}\sigma_{jk}$. The solution is given as

$$V(s,t) = E_Q[e^{\int_s^t r(\tau)d\tau}V(S_t,t)|\mathcal{F}_s]$$

where S_t is a N dimensional stochastic process, and Q is probability measure, under which the dynamics of S is given as

$$dS_i = \mu_i dt + \sum_{j=1}^N \sigma_{i,j} dW_j(t), i = 1,2,...,N$$

where W_t is the Wiener process. Moreover,

$$e^{\int_0^t r(\tau)d\tau}V(S_t,t)$$

is a martingale under the measure Q.

Proof. Similar to 1D case. □

Theorem 21.3.3 (Feyman-Kac formula for PDE with source term). *Consider the PDE*

$$\frac{\partial u(x,t)}{\partial t} + \mu(x,t)\frac{\partial u(x,t)}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 u(x,t)}{\partial x^2} - V(x,t)u(x,t) + f(x,t) = 0$$

with $x \in \mathbb{R}, t \in [0,T]$ and terminal condition of

$$u(x,T) = \psi(x)$$

then the solution to the PDE is

$$u(x,t) = E^Q[\int_t^T e^{-\int_t^r V(X_\tau,\tau} d\tau] f(X_r,r)dr - e^{-\int_t^T V(X_\tau,\tau} d\tau \psi(X_T)|X_t = x]$$

under the probability measure Q such that X is an Ito process driven by the equation

$$dX = \mu(X,t)dt + \sigma(X,t)dW^Q$$

with W^Q being the a Wiener process under Q.

Remark 21.3.3 (solving PDE by simulation).

- It offers a method of solving certain PDEs by simulating random paths of a stochastic process. The procedure for simulating a trajectory will be: simulate X_t from t to T using $dX = \mu(X,t)dt + \sigma(X,t)dW^Q$ with this trajectory we can evaluate

$$e^{-\int_t^r V(X_\tau,\tau} d\tau$$

  and

$$\int_t^T e^{-\int_t^r V(X_\tau,\tau)d\tau} f(X_r,r)dr$$

  to obtain one sample.
- The simulation evaluation approach clearly demonstrates the expectation will depends on $X_t = x$, i.e., x and t.

Remark 21.3.4. an important class of expectations of random processes can be computed by deterministic methods.

Corollary 21.3.3.1 (Black-Scholes equation). *The Black-Scholes equation given as*

$$\frac{\partial V(s,t)}{\partial t} + rs\frac{\partial V(s,t)}{\partial s} + \frac{1}{2}\sigma^2 s^2\frac{\partial^2 V(s,t)}{\partial s^2} - rV(s,t) = 0$$

with $s \in \mathbb{R}, t \in [0, T]$ and terminal condition of $V(s, T) = \psi(S_T)$ has the solution in conditional expectation form as

$$V(s,t) = E^Q[\psi(S_T)\exp(-r(T-t))|S_t = s]$$

under probability measure Q such that S_t is an Ito process given by

$$dS = rSdt + \sigma SdW.$$

Moreover, under probability measure Q, S has solution given as

$$S(T) = S(t)\exp((r - \frac{1}{2}\sigma^2)(T-t) + \sigma(B(T) - B(t)))$$

and therefore

$$V(s,t) = \int_{-\infty}^{\infty} \psi(S_T)\exp(-r(T-t))f(B(T) - B(t) = x)dx$$

Remark 21.3.5 (simulation approach to Black-Scholes equation)**.** We can evaluate the expectation by simulating trajectories of S_t. More precisely, starting with initial condition $S_t = s$ and use Euler algorithm to integrate

$$dS = rSdt + \sigma SdW.$$

21.3.2 Backward equation

Theorem 21.3.4 (Kolmogorov backward equation, expectation pricing). *Assume X_t is governed by the following SDE*

$$dX_t = \mu dt + \sigma dW_t.$$

Suppose we are given a payoff $V(X_T)$ at time T. Define $f(x,t) = E[V(X_T)|X_t = x]$ for all $t \leq T$. Then $f(x,t)$ is governed by the ***Kolmogorov backward equation*** *given as*

$$-\frac{\partial}{\partial t}f(x,t) = \mu\frac{\partial}{\partial x}f(x,t) + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2}f(x,t)$$

for $t \leq T$, subject to the final condition $f(x,T) = V(X_T)$.

Proof. Note that

$$E[dX_t|cF_t] = E[dX_t|X_t = x] = a(X_t,t)dt$$

and

$$E[dX_t^2|cF_t] = E[dX_t^2|X_t = x] = b^2(X_t, t)dt.$$

then we have (use the notation $E[f(X_s, s)|X_t = t] = E_{x,t}[f(X_s, s)]$)

$$\begin{aligned}
E_{x,t}[f(X_{t+dt}, t + dt)] &= E_{x,t}[f(x + dX_t, t + dt)] \\
&\approx E_{x,t}[f(x,t) + \partial_x f(x,t)dX_t + \frac{1}{2}\partial_x^2 f(x,t)dX_t^2 + \partial_t f(x,t)dt] \\
&= f(x,t) + \partial_x f(x,t)E_{x,t}[dX_t] + \frac{1}{2}\partial_x^2 f(x,t)E_{x,t}[dX_t^2] + \partial_t f(x,t)dt \\
&= f(x,t) + dt(\partial_t f(x,t) + \mu(x,t)\partial_x f(x,t) + \frac{\sigma^2(x,t)}{2}\partial_x^2 f(x,t))
\end{aligned}$$

$$E_{x,t}[f(X_{t+dt}, t + dt)] = f(x,t) \implies \partial_t f(x,t) + \mu(x,t)\partial_x f(x,t) + \frac{\sigma^2(x,t)}{2}\partial_x^2 f(x,t) = 0$$

□

Remark 21.3.6 (interpretation).

- We can interpret $f(x,t)$ as the price at t and the state is at x. When we cannot hedge the risk, we can price the asset by the expected value of payoff with respect to the real probability. Note that such pricing method does not take into account of the risk-aversion.
- The only difference to Black-Scholes is the extra source decreasing term.

Theorem 21.3.5 (Kolmogorov backward equation with discount). *Assume X_t is governed by the following SDE*

$$dX_t = \mu dt + \sigma dW_t.$$

Suppose we are given a payoff $V(X_T)$ at time T. Define $f(x,t) = E[e^{-\int_t^T r(tau)d\tau}V(X_T)|X_t = x]$ for all $t \leq T$. Then $f(x,t)$ is governed by the ***Kolmogorov backward equation*** *given as*

$$-\frac{\partial}{\partial t}f(x,t) = \mu\frac{\partial}{\partial x}f(x,t) + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2}f(x,t) - rf$$

for $t \leq T$, subject to the final condition $f(x,T) = V(X_T)$.

Proof. Note that

$$E[dX_t|cF_t] = E[dX_t|X_t = x] = a(X_t, t)dt$$

and

$$E[dX_t^2|cF_t] = E[dX_t^2|X_t = x] = b^2(X_t, t)dt.$$

then we have (use the notation $E[f(X_s,s)|X_t=t] = E_{x,t}[f(X_s,s)]$)

$$\begin{aligned}
E_{x,t}[(1-rdt)f(X_{t+dt},t+dt)] &= E_{x,t}[(1-rdt)f(x+dX_t,t+dt)] \\
&\approx (1-rdt)E_{x,t}[f(x,t)+\partial_x f(x,t)dX_t + \frac{1}{2}\partial_x^2 f(x,t)dX_t^2 + \partial_t f(x,t)dt] \\
&= f(x,t) - rf(x,t)dt + \partial_x f(x,t)E_{x,t}[dX_t] + \frac{1}{2}\partial_x^2 f(x,t)E_{x,t}[dX_t^2] + \partial_t f(x,t)dt \\
&= f(x,t) - rf(x,t)dt + dt(\partial_t f(x,t) + \mu(x,t)\partial_x f(x,t) + \frac{\sigma^2(x,t)}{2}\partial_x^2 f(x,t))
\end{aligned}$$

$$E_{x,t}[(1-rdt)f(X_{t+dt},t+dt)] = f(x,t)$$

$$\implies \partial_t f(x,t) + \mu(x,t)\partial_x f(x,t) + \frac{\sigma^2(x,t)}{2}\partial_x^2 f(x,t) - rf(x,t) = 0$$

□

Theorem 21.3.6 (Kolmogorov backward equation, multi-dimensional version). *Assume X_1, X_2 is governed by the following SDE*

$$dX_1(t) = \mu_1 dt + \sigma_1 dW_1(t).$$

and

$$dX_2(t) = \mu_2 dt + \sigma_2 dW_2(t).$$

Suppose we are given a payoff $V(X_1(T), X_2(T))$ at time T. Define

$$f(x_1,x_2,t) = E[e^{-\int_t^T r(\tau)d\tau}V(X_1(T),X_2(T)|X_1(t)=x_1, X_2(t)=x_2]$$

for all $t \leq T$. Then $f(x_1,x_2,t)$ is governed by the ***Kolmogorov backward equation*** *given as*

$$-\frac{\partial}{\partial t}f(x_1,x_2,t) = \mu_1\frac{\partial}{\partial x_1}f + \mu_2\frac{\partial}{\partial x_2}f + +\frac{1}{2}\sigma_1^2\frac{\partial^2}{\partial x_2^2}f + \frac{1}{2}\sigma_2^2\frac{\partial^2}{\partial x_2^2}f + \sigma_1\sigma_2\frac{\partial^2}{\partial x_1 \partial x_2}f - rf$$

for $t \leq T$, subject to the final condition $f(x_1,x_2,T) = V(X_1(T),X_2(T))$.

Proof. Similar to 1D case. □

Remark 21.3.7 (Kolmogorov forward and backward equation).

- the Kolmogorov forward equation addresses the following problem: We have information about the state x of the system at time t as $P(x,t)$, we want to know $P(x,s), s > t$.

- The Kolmogorov backward equation addresses the problem: at time t in whether at a future time s the system will be in a given subset of states B, sometimes called the target set. The target is described by a indicator function for the set B. We want to know for every state x at time $t, (t < s)$ what is the probability of ending up in the target set at time s (sometimes called the hit probability). In this case $u_s(x)$ serves as the final condition of the PDE, which is integrated backward in time, from s to t.

21.3.3 Application to first hitting probability

Lemma 21.3.1 (General method via Feyman Kac formula). *Consider stochastic process given by*

$$dX(t) = mdt + \sigma dW(t), X(0) = 0$$

where $W(t)$ is the Brownian motion. Then given two levels $a > 0$ and $-b, b > 0$, and let the probability $P(t,x)$ denote the probability that the process starting at $X(t) = x$ hits a before hitting $-b$. Then we have

- *$P(t,x)$ is independent of time t.*
- *The governing equation for $P(x)$ is given by*

$$m\frac{\partial P}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 P}{\partial x^2} = 0,$$

  with boundary condition $P(x=a) = 1, P(x=-b) = 0$.

Proof. (1) Note that this is a Markov process, therefore $P(t,x)$ will be depend on time. (2) Consider a value function $P(x,t) = E[P_T|X(t) = x]$ with final condition $P(x,T) = P_T(P_T$ will take value 1 at target sites and take 0 elsewhere). Then from Feyman Kac theorem(Theorem 21.3.1), $P(x,t)$ is also the solution of

$$\frac{\partial P}{\partial t} + m\frac{\partial P}{\partial x} + \frac{1}{2}\sigma\frac{\partial^2 P}{\partial x^2} = 0,$$

with boundary condition $p(t,x) = P_T$. □

21.4 Advanced analysis for Brownian motion

Lemma 21.4.1 (Kramers equation). *The following SDE for a particle subject to Brownian force*

$$dv = -\gamma v dt + \sigma dW_t$$
$$dx = v dt$$

is associated with the Fokker-Planck equation on $p(x,v,t)$, given as

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(vp) + \frac{\partial}{\partial v}(\gamma v p) + \frac{1}{2}\sigma\frac{\partial^2 p}{\partial v^2}.$$

Proof. Direct from Lemma 21.1.2. Note that the variance coefficient for x is zero. □

Theorem 21.4.1 (fluctuation-dissipation theorem). *[5, p. 3] Consider the SDE for velocity*

$$dV = -\gamma V dt + \sigma dW_t.$$

Under the assumption $\frac{1}{2}m\left\langle V(t)^2 \right\rangle = \frac{1}{2}kT$, the coefficient σ and γ are connected via

$$\sigma = \sqrt{2\gamma kT/m}.$$

Proof. From the solution of OU process(Lemma 20.3.1), we know that

$$\left\langle V(t)^2 \right\rangle = \frac{\sigma^2}{2\gamma} \implies \frac{\sigma^2}{2\gamma} = \frac{1}{m}kT.$$

□

Lemma 21.4.2 (auto-correlation of velocities). *[5, p. 34] Suppose v is governed by*

$$dv = -\gamma v dt + \sigma dW_t.$$

Then

$$\langle v(t_1)v(t_2) \rangle = v_0^2 e^{-\gamma(t_1+t_2)} + \frac{\sigma}{2\gamma}(e^{-\gamma|t_1-t_2|} - e^{-\gamma(t_1+t_2)}).$$

For large t_1, t_2 such that $\gamma t_1 \gg 1, \gamma t_2 \gg 1$, we have

$$\langle\, v(t_1)v(t_2) \,\rangle = \frac{\sigma}{2\gamma} e^{-\gamma|t_1 - t_2|}.$$

Proof. Note that from Lemma 21.1.2, then

$$v(t_1) = v_0 e^{-\gamma t_1} + \int_0^{t_1} e^{-\gamma(t_1 - t)} \sigma dW_t.$$

We can then proceed. □

Lemma 21.4.3 (mean displacement analysis and fluctuation-dissipation theorem for colloidal particles). *For a colloidal particle with hydrodynamic drag given as*

$$\gamma = \frac{6\pi\mu a}{m}$$

And the equation of motion is given as

$$\begin{aligned} dv &= -\frac{\gamma}{m} v dt + \frac{\sigma}{m} dW_t \\ dx &= v dt. \end{aligned}$$

Then, it can be showed that

$$\left\langle\, (x(t) - x(0)^2 \,\right\rangle = \frac{\sigma^2}{\gamma^2} t = 2Dt$$

where

$$D = \frac{kT}{\gamma} = \frac{kT}{6\pi\mu a}.$$

Proof. Using Theorem 21.4.1.

$$\begin{aligned} \left\langle\, (x(t) - x(0)^2 \,\right\rangle &= \int_0^t \int_0^t \langle\, v(t_1)v(t_2) \,\rangle dt_1 dt_2 \\ &= \left\langle\, v_0(t)^2 \,\right\rangle \frac{2}{\gamma} t = \frac{\sigma^2}{\gamma^2} = \frac{2kT}{\gamma} t = 2Dt \end{aligned}$$

□

Remark 21.4.1. There are two versions of fluctuation-dissipation theorem(Theorem 21.4.1,Lemma 21.4.3) Note that The variance of random forces is fundamentally different from the variance of the random displacements.

Lemma 21.4.4 (Maxwell distribution). *Consider the SDE for velocity*

$$dv = -\gamma v dt + \sigma dW_t.$$

Assume the fluctuation-dissipation theorem holds. The equilibrium distribution of the velocity is given as

$$p(v) = \sqrt{\frac{m}{2\pi kT}} \exp(-\frac{mv^2}{2kT}).$$

This distribution is also known as Maxwell distribution.

Proof. From the solution of OU process(Lemma 20.3.1), we know that

$$\left\langle V(t)^2 \right\rangle = \frac{1}{m}kT.$$

□

Theorem 21.4.2 (microscopic to continuum description,Green-Kubo formula). *[6, p. 177] From Lemma 21.4.3, we can claim that at the long time limit, the probability distribution of a Brownian particle is given by*

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2}.$$

From which we can show

$$\frac{d}{dt}\left\langle \|x(t)\|^2 \right\rangle = 2D$$

$$D = \frac{1}{2}\int_0^\infty \langle v(t)v(0) \rangle dt$$

Proof. (1)From SDE in Lemma 21.4.3, we know that the probability distribution of $x(t)$ is Gaussian, which is also the solution to this equation in this theorem. (2)

$$\begin{aligned}
\frac{d}{dt}\left\langle \|x(t)\|^2 \right\rangle &= \frac{d}{dt}\|x(t)\|^2 \rho(x,t)dx = D\int \|x\|^2 \nabla^2\rho(x,t)dx \\
&= D\int \nabla\cdot(\|x\|^2 \nabla\rho)dx - D\int (\nabla\|x\|^2)\cdot\nabla\rho dx \\
&= -2D\int x\cdot\nabla\rho dx \\
&= 2D\int (\nabla\cdot x)\rho dx = 2D
\end{aligned}$$

Using $x(t) = \int_0^t v(t_1)dt_1$, we have

$$\begin{aligned} \left\langle \|x(t)\|^2 \right\rangle &= \left\langle \int_0^t \int_0^t v(t_1)v(t_2)dt_1dt_2 \right\rangle \\ &= \int_0^t \int_0^t \langle v(t_1 - t_2)v(0) \rangle dt_1dt_2 \\ \implies \frac{d}{dt}\left\langle \|x(t)\|^2 \right\rangle &= \int_0^t \langle v(t)v(0) \rangle dt. \end{aligned}$$

□

Remark 21.4.2 (interpretation). This equation relates a macroscopic transport coefficient D to a quantity in microscopic system. Note that the velocity autocorrelation function $g(t) = \langle v(t)v(0) \rangle$ can be calculated using microscopic simulations such as molecular dynamics. Similar expressions can be found for viscosity, heat conductivity, etc.

21.4.1 Kramers problem: barrier escape

Lemma 21.4.5.

$$\frac{dx}{dt} = -\frac{1}{\gamma}U' + \frac{1}{\gamma}\xi$$

The corresponding Fokker-Planck equation is given as

$$\frac{\partial}{\partial t}P(x, t|x_0, 0) = \frac{\partial}{\partial x}\frac{1}{\gamma}U'P + \frac{\partial^2}{\partial x^2}DP.$$

Let $T(x)$ be the mean first passage time, we have

$$-\frac{1}{\gamma}U'\frac{d}{dx}T(x) + D\frac{d^2}{dx^2}T(x) = -1$$

with boundary condition given as $T(x_{target}) = 0$.

Proof. Multiply both sides by $\exp(-\beta U)$, we have

$$(\frac{d}{dx}e^{-\beta U})\frac{d}{dx}T + e^{-\beta U}\frac{d^2}{dx^2}T(x) = -\frac{1}{D}e^{-\beta U}$$

or

$$\frac{d}{dx}[e^{-\beta U}\frac{d}{dx}T(x)] = -\frac{1}{D}e^{-\beta U}.$$

Integrating from $-\infty$ to x we have

$$\frac{d}{dx}T(x) = -\frac{1}{D}e^{\beta U}\int_{-\infty}^{x} dze^{-\beta U(z)}.$$

Finally, integrating from x_{target} to x, we have

$$T(x) = \int_{x_{target}}^{x} \frac{1}{D}dye^{\beta U(y)}\int_{-\infty}^{x} dze^{-\beta U(z)}.$$

□

Remark 21.4.3 (approximate solution). In the first integral, the integrand is largest around the barrier top x_b, and we expand the external potential aournd this point as

$$U(y) = U_b - \frac{1}{2}mw_b^2(y-x_b)^2.$$

In the second integral, the integrand is largest around x_a and

$$U(z) = U_a + \frac{1}{2}mw_b^2(z-x_{target})^2.$$

Further assume D is a constant, then the mean time from a valley to another valley over the barrier is

$$\tau = \frac{1}{D}\int_{-\infty}^{\infty} dye^{\beta U_b}e^{-\frac{1}{2}\beta mw_b^2(y-x_b)^2}\int_{-\infty}^{\infty} e^{-\beta U_{target}}e^{-\frac{1}{2}\beta mw_{target}^2(z-x_{target})^2} = \frac{1}{D}\frac{2\pi k_B T}{mw_{target}w_b}e^{\beta(U_b-U_{target})}.$$

21.5 Notes on bibliography

For Fokker-Planck equation, see [5][7].

BIBLIOGRAPHY

1. Chirikjian, G. S. *Stochastic Models, Information Theory, and Lie Groups, Volume 1: Classical Results and Geometric Methods* (Springer Science & Business Media, 2011).
2. Clark, I. J. *Foreign exchange option pricing: a practitioner's guide* (John Wiley & Sons, 2011).
3. Pavliotis, G. A. *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations* (Springer, 2014).
4. Liu, C., Brennan, D. P., Bhattacharjee, A. & Boozer, A. H. Adjoint Fokker-Planck equation and runaway electron dynamics. *Physics of Plasmas* **23,** 010702 (2016).
5. Risken, H. & Frank, T. *The Fokker-Planck Equation: Methods of Solution and Applications* ISBN: 9783642615443 (Springer Berlin Heidelberg, 2012).
6. Weinan, E. *Principles of multiscale modeling* (Cambridge University Press, 2011).
7. Gardiner, C. *Stochastic Methods: A Handbook for the Natural and Social Sciences* ISBN: 9783540707127 (Springer Berlin Heidelberg, 2009).