18

STOCHASTIC PROCESS

18.1 Stochastic process

18.1.1 Basic definition and concepts

From real-life applications to scientific research, we are often interested in multiple observations of random values over a period of time. Examples include

- Stock prices of a company over the past five years.
- Temperature over a period of time.
- Random movements of cells on a two-dimensional surface.
- Number of visitor to a restaurant from the opening.

These observations behave differently over time and there exist movement patterns like mean-reverting, fluctuating, diverging, and jumping. We model these observations and their evolving patterns through stochastic processes. The goal is at least two fold: first, to characterize the statistical properties of these sequential random observations, such as mean and covariance; second, to predict the evolution patterns and ultimately understand of underlying driving forces.

A **stochastic process**, or **random 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**

Example 18.1.1.

- A random walk process $\{X_n\}$ is generated by $X_n = X_{n-1} + Z$, Z is a random variable taking value in $\{-1, 1\}$ with equal probability.
- $M(t)$ describe the total value of money market account at time t after depositing one unit money at time 0. $M(t)$ is a random process since short-term interest rate is a random process.

For a discrete-time stochastic process, the sequence of numbers $X_1(\omega), X_2(\omega), ...$ 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.

A stochastic process X_t 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 (also called a realization or a trajectory) of X_t.

Example 18.1.2 (sample path examples).

- One trivial case is that $X_1, X_2, ...$ 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 $\Omega = [0, 1]$

Remark 18.1.1 (interpretation on sample space and σ algebra). [1, 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 18.1.1.
- 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 $t \geq 0$.

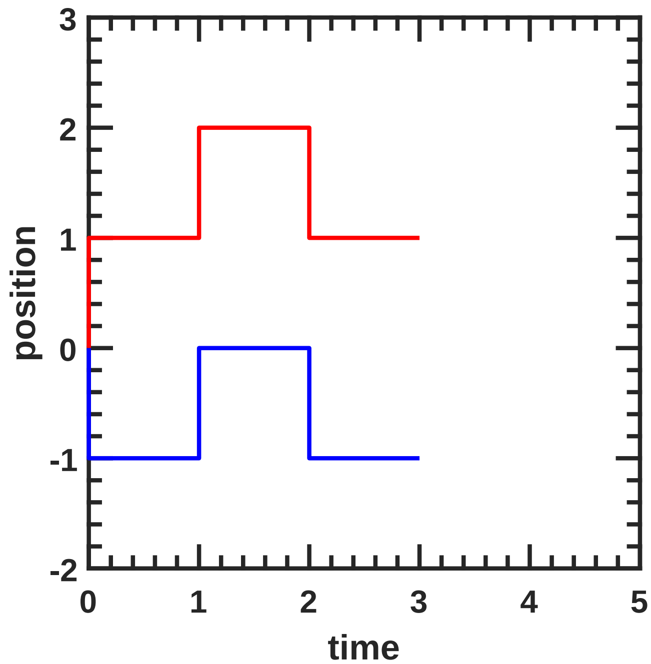

Figure 18.1.1: An illustration of a random walk mapping a sample point, ω, to a trajectory parameterized by time, where red trajectory sample point HHT, and blue trajectory has sample point THT.

18.1.2 Stationarity

One basic characterization of the evolution pattern is stationarity. A stochastic process $\{X_t\}$ is stationary if statistical properties of its sample path remain the same over time. In other words, a stationary process tends to repeat itself in the statistical sense. We first introduce strict stationarity, which specifies stationarity in the joint distribution.

Definition 18.1.1 (strictly stationary process). *A continuous-time stochastic process $\{X(t)\}$ is a strict stationary or simply stationary if, for all $t_1, t_2, \cdots, t_n \in \mathbb{R}$ and all $\Delta \in \mathbb{R}$, the joint cdf of $X(t_1), X(t_2), \cdots, X(t_N)$ has the same joint cdf as $X(t_1 + \Delta), X(t_2 + \Delta), \cdots, X(t_n + \Delta)$, That is,*

$$F_{X(t_1)X(t_2)\cdots X(t_n)}(x_1, x_2, \cdots, x_n) = F_{X(t_1+\Delta)X(t_2+\Delta)\cdots X(t_n+\Delta)}(x_1, x_2, \cdots, x_n).$$

In other words, all joint cdfs are translational invariant.[a]

[a] We can similarly define strict stationarity for a discrete-time process by requiring that t_i and $t_i + \Delta$ are both valid time indices.

Example 18.1.3 (a sequence of iid random variables). A sequence of iid random variables $\{X_1, X_2, ...\}$ is a strictly stationary process. We can compute the joint cdf via

$$P(X_1 \leq x_1, ..., X_n \leq x_n) = \prod_{i=1}^{n} P(X \leq x_i)$$

and

$$P(X_{1+\Delta} \leq x_1, ..., X_{n+\Delta} \leq x_n) = \prod_{i=1}^{n} P(X \leq x_i).$$

Example 18.1.4 (A Markov chain starting from stationary distribution is strictly stationary process). 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 20.4.4]. By iteration, we know that the distribution at every time step is π.

However, strict stationarity could be quite restrict for practical modeling applications since many interesting processes are not strictly stationary. Take a step back, we can define a weak stationarity, which specifies stationarity property in the the mean and covariance structure.

Definition 18.1.2 (weakly stationary process). *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 $r : \mathbb{R} \to \mathbb{R}$.

There are a number of useful properties regarding of the covariance function of a weakly stationary process.

Lemma 18.1.1 (properties of a covariance function). *[2, 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)*$r(\tau) = r(-\tau)$.

- $\left|r(\tau)\right| \leq r(0)$.
- *If* $\left|r(\tau)\right| = r(0)$, *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 11.9.4]

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

(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). □

We can also show that strict stationarity implies weak stationarity. The proof is straight forward: Strict stationarity offers translational invariance of two-variable joint cdf, therefore, mean and convariance, which derived from joint cdf, inherits such translational invariance.

Theorem 18.1.1 (a strictly stationary process is a weakly stationary process). *A strictly stationary process* X_t *will be a weakly stationary process.*

Example 18.1.5. For Gaussian process, a weakly stationary Gaussian process is a strictly stationary Gaussian process [Lemma 18.2.2].

18.2 Gaussian process

18.2.1 Basic Gaussian process

Definition 18.2.1 (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 follows a multivariate normal distribution [subsection 12.1.9].*

Example 18.2.1 (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 18.2.2 (a discrete random walk is not a Gaussian process). A discrete-time random walk B_n is not a Gaussian process. For example, B_1 is a Bernoulli distribution, not a Gaussian.

Example 18.2.3 (Brownian motion).

- A Brownian motion process W_t is the integral of a white noise Gaussian process. It is not stationary, but it has stationary increments [Lemma 18.3.1].
- A geometric Brownian motion process $X_t = \exp(W_t)$ is not a Gaussian process.

The affine transformation property Theorem 14.1.1 of Gaussian random variable allow us to construct new Gaussian processes via affine transformation.

Lemma 18.2.1 (construct new Gaussian processes via affine transformation).

- *Let* X_t *be a Gaussian process, then* $aX_t + b, a, b \in \mathbb{R}$ *is also a Gaussian process.*
- *Let* $X_1(t), X_2(t), ..., X_n(t)$ *be independent Gaussian processes. Then*

$$Y(t) = \sum_{i=1}^{n} \alpha_i X_i(t)$$

is a Gaussian process.

Example 18.2.4 (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 white noise 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.

18.2.2 Stationarity

A Gaussian process can be stationary (e.g., white noise process) or non-stationary (e.g., Brownian motion).An important property of Gaussian processes is that weak stationarity and strict stationary is equivalent. More specifically, we can state the following theorem.

Lemma 18.2.2. *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 12.1.16], 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. □

Example 18.2.5. 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

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

18.3 Brownian motion (Wiener process)

18.3.1 Definition and elementary properties

Brownian motion, also called Wiener process, is another continuous time stochastic process of fundamental importance. Its importance not only because its wide adoption in modeling critical stochastic dynamics in scientific research, engineering, and finance, but also because it is the building block to construct more complex stochastic process via stochastic calculus [chapter 19].

Example applications of Brownian motion including modeling motion of cells, stock price fluctuation, etc. Graphically [Figure 18.3.1], a Brownian motion trajectory is a continuous sample with jittery motions.

Figure 18.3.1: Sample trajectories of Brownian motion process.

A Brownian motion is a stochastic process described by following properties.

Definition 18.3.1 (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 any* $0 < t_1 < t_2$ *the random variables*

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

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

In the following, we prove a number of basic statistical properties for a one-dimensional Brownian motion. As proved, Brownian motion is a **nonstationary Gaussian process**.

Lemma 18.3.1 (basic properties of one-dimensional 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 the 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. □

Example 18.3.1 (Brownian motion with drift). Consider the process

$$X_t = \mu t + \sigma W_t, \quad t \geq 0$$

for constants $\sigma > 0$ and $\mu \in \mathbb{R}$. It is a Gaussian process with expectation and covariance functions given by

$$\mu_X(t) = \mu t \quad \text{and} \quad c_X(t, s) = \sigma^2 \min(t, s), \quad s, t \geq 0.$$

Example 18.3.2 (geometric Brownian motion). Consider the process

$$X_t = \exp(\mu t + \sigma W_t), \quad t \geq 0$$

for constants $\sigma > 0$ and $\mu \in \mathbb{R}$. It is known as geometric Brownain motion and it is not a Gaussian process.

18.3.2 Multi-dimensional Brownian motion

Definition 18.3.2 (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.*

Based on the results in one-dimensional Brownian motion, following properties can be straight forward.

Lemma 18.3.2 (basic properties of multidimensional independent Brownian motion). *Consider the vector $W(t) = (W_1(t), W_2(t), ..., W_m(t))^T$ representing 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 $dW(t) = W(t + dt) - W(t)$, $\delta(t)$ is the Dirac delta function.[a]

[a] Direct delta function can be viewed as having a value of $1/dt$.

By introducing correlation between components, we arrive at multi-dimensional correlated Brownian motion.

Definition 18.3.3 (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.*
- *For all i, j,*
$$cov(W_i(s) W_j(t)) = \rho_{ij} \min(s, t).$$

or in matrix form

$$Cov(W(s), W(t)) = \rho \min(s, t).$$

18.3.3 Asymptotic behaviors

In the following, we list a number of asymptotic behaviors of a Brownian motion.

Theorem 18.3.1 (law of iterated logarithms). *[3] As $t \to \infty$, we have with probability 1 (i.e. almost surely):*

- $\lim_{t\to\infty} W_t/t = 0$
- $\limsup_{t\to\infty} W_t/\sqrt{t} = \infty$
- $\limsup_{t\to\infty} W_t/\sqrt{2t\log(\log t)} = 1$
- $\liminf_{t\to\infty} W_t/\sqrt{2t\log(\log t)} = 1$

Corollary 18.3.1.1 (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 18.3.1.2 (first hitting time of a level). *Define $T_a = \inf\{t : W_n > a\}$. $T_a <= \infty$ almost surely (but the mean first passage time will be infinite).*

Remark 18.3.1. **??** also shows that the hitting probability is 1 given infinite amount of time.

Theorem 18.3.2. *[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 18.3.2. Here use almost all is because there is some path, e.g. a path that $W_t(\omega) = const$, that variation will be zero; however, such path has zero probability measure.

18.3.4 The reflection principle

Lemma 18.3.3 (reflection principle). *[5, 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 18.3.3 (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.

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.

Theorem 18.3.3 (path excursion distribution). *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*

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

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

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 18.3.3], we have

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

□

18.3.5 Quadratic variation

We introduce the concept of **quadratic variation** to measure how jagged the paths of a Brownian motion are.

Definition 18.3.4 (quadratic variation). *Consider a function $f : [0, T] \to \mathbb{R}$. Define*

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

where Δ is a partition of the interval $[0, T]$ with $0 = t_0 < t_1 ... < t_n = T$. Then the quadratic variation of f is defined to be

$$Q^* = \lim_{l(\Delta)\to 0} Q(\Delta),$$

where $l(\Delta) = \max_i(t_{i+1} - t_i)$.

One important property of quadratic variation is that any continuously differentiable function have zero quadratic variation.

Theorem 18.3.4 (continuously differentiable functions have zero quadratic variations). *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 $Q^* = 0$. □

Remark 18.3.4 (almost sure convergence). Here we only prove convergence in distribution. Actually, it can be further shown that the convergence is almost surely.

Brownian motion has almost surely continuous sample path. Its quadratic variation is given by the following theorem.

Theorem 18.3.5 (Brownian motion quadratic variation). *The Brownian motion W on the interval $[0,T]$ has a quadratic variation of T in the sense of convergence in mean square.*

Proof. We first prove that

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

and

$$\begin{aligned}
&E[Q(\Delta)Q(\Delta)] \\
=&E[\sum_{i=0}^{n-1}(W_{i+1}-W_i)^2\sum_{j=0}^{n-1}(W_{j+1}-W_j)^2] \\
=&E[\sum_{i=0}^{n-1}(W_{i+1}-W_i)^4] + E[\sum_{i=0}^{n-1}(W_{i+1}-W_i)^2]E[\sum_{j=0,j\neq i}^{n-1}(W_{j+1}-W_j)^2] \\
=&\sum_{i=0}^{n-1}3(t_{t+1}-t_i)^2 + \sum_{i=0}^{n-1}(t_{i-1}-t_i)(T-(t_{i-1}-t_i)) \\
=&\sum_{i=0}^{n-1}2(t_{t+1}-t_i)^2 + T^2.
\end{aligned}$$

Then

$$Var[Q(\Delta)] = \sum_{i=0}^{n-1}2(t_{t+1}-t_i)^2 \leq 2nl(\Delta)^2 \to 0$$

as $l(\Delta) \to 0$. Using mean convergence criterion [Theorem 11.10.3] we can get the final result. □

In applications involving Brownian motion, we often encounter the notation $dW(t)$, which is defined as $dW(t) = W(t+dt) - W(t)$. The following theorem gives its statistical properties.

Theorem 18.3.6 (differential forms of quadratic variation). *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 12.1.6]. Note that $dW(t)dW(t)$ is just a random variable with mean dt, and variance approaches 0 . □

18.3.6 Discrete-time approximations and simulation

The approximation scheme is important in simulating stochastic differential equations.

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 $B(T) \sim N(0, \sigma^2 T)$.

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.

18.4 Brownian motion variants

This section involves stochastic calculus [chapter 19].

18.4.1 Gaussian process generated by Brownian motion

Lemma 18.4.1 (Gaussian process stochastic differential equation). *Consider a stochastic process X_t governed by*

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

where W_t Brownian. It follows that

$$X(t) \sim N(\int_0^t a(s)ds, \int_0^t b(s)^2 ds)$$

and $X(t)$ is a Gaussian process.

Proof. See Corollary 19.1.4.1. □

Theorem 18.4.1 (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 12.1.13, 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. □

Remark 18.4.1 (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.

18.4.2 Brownian bridge

18.4.2.1 *Constructions*

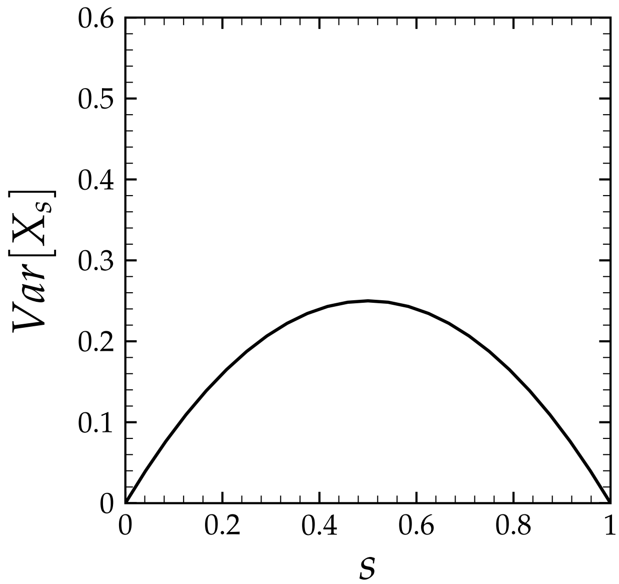

Figure 18.4.1: Variance of X_t in a Brownian bridge

Definition 18.4.1 (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.*

Particularly, we have can calcuate covariance using conditional distribution in the following way. 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 18.3.1]. Then

$$(X_t|X_1) \sim MN(0, t - t^2)$$

from Theorem 14.1.2. 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 18.4.2 (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.*

Definition 18.4.3 (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 18.4.4 (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 18.4.2 (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 14.1.1]. (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 [Theorem 19.1.4, Corollary 19.1.4.1]. 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 18.4.3 (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) □

Remark 18.4.2 (simulation of Brownian bridge). We can simulate a Brownian bridge by first simulating a Wiener process W_t and then using
$$X_t = W_t - tW_1.$$

18.4.2.2 *Applications*

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.

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 $W(2)$.

Example 18.4.1 (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.

18.4.3 Geometric Brownian motion

Definition 18.4.5 (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

$$dX_t = \mu X_t dt + \sigma X_t dZ_t.$$

Based on properties of the log-normal distribution, we can derive the following.

Lemma 18.4.4 (distribution). *The geometric Brownian motion has the lognormal distribution with parameter $(\mu - \frac{1}{2}\sigma^2)t$ and $\sigma\sqrt{t}$. The pdf is given as*

$$f_t(x) = \frac{1}{\sqrt{2\pi t}\sigma x} \exp(-\frac{(\ln(x/x_0) - (\mu - \sigma^2/2)t)^2}{2\sigma^2 t}).$$

Further, if S_t be a geometric Brownian motion with initial condition S_0, then

- $E[S_t] = S_0 e^{\mu t}$
- $Var[S_t] = S_0^2 e^{2\mu t}(e^{\sigma^2 t} - 1)$

18.5 Poisson process

18.5.1 Basics

Many interesting real-life applications involving a counting process, which counts the occurrences of a certain event over a period of time. For example, the number of customer visits to a store, $N(t)$, since a reference starting time 0. Other example counting processes include the the number of calls received by a medical emergency center, the arrival of buy and sell orders in electronic trading, etc. In this section, we introduce Poisson process, a widely used process with discrete state space, to model such counting process.

Poisson process offers different facets characterizing a counting process, ranging from basic statistical properties, arrival times, and inter-arrival waiting times.

Definition 18.5.1 (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 $N(t_2) - N(t_1) \sim Poisson(\lambda(t_2 - t_1))$.*

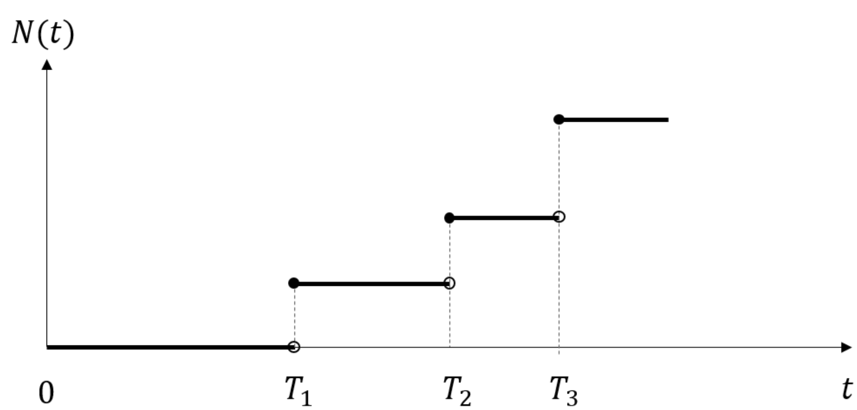

Figure 18.5.1: A typical realized trajectory from the Poisson process with jumps at T_1, T_2, and T_3.

Lemma 18.5.1 (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 12.1.3]and basic property of Poisson distribution [Lemma 12.1.2]. □

Lemma 18.5.2 (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 12.1.3]. □

18.5.2 Arrival and Inter-arrival Times

Lemma 18.5.3 (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 18.5.1. Note that the waiting time distribution is an exponential distribution with parameter λ, whose mean is $1/\lambda$.

Lemma 18.5.4 (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 12.1.3.) □

The investigation on the waiting time above givers a straight forward way to simulate a Poisson process.

Methodology 18.5.1 (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*

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

18.6 Martingale theory

18.6.1 Preliminaries: Filtration and adapted process

18.6.1.1 *Basic concepts in filtration*

Definition 18.6.1 (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 18.6.1. A filtration represents an increasing stream of information.

Definition 18.6.2 (adapted process). *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 18.6.2. [6]

- 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.

Definition 18.6.3 (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 18.6.3.

- 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 18.6.4 (interpret natural filtration). [7, 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 occurred 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 18.6.1 (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 18.6.2 (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 $\Omega_\infty =$ 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 18.6.5 (σ algebra for a stochastic process)**.** From Remark 18.1.1, 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.

18.6.1.2 *Filtration for Brownian motion*

Let W_t be a Brownian motion, the filtration for the Brownian motion can be defined as $\mathcal{F}_t = \sigma(\{\mathcal{F}_s\}_{s\leq t})$. This filtration is also the natural filtration. W_t is $\mathcal{F}_t$ adapted, but $W_s, s > t$ is not $\mathcal{F}_t$ adapted.

Also note that

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

18.6.2 Basics of martingales

We start with the definition of a martingale.

Definition 18.6.4 (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 $0 \leq s \leq t$.*

Note that Martingale is always an adapted process with respect to some filtration. Its discrete-time version is given below.

Definition 18.6.5 (discrete-time martingale). *[8, 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 18.6.3 (Brownian motion is a martingale). Let $W(t)$ be a Brownian motion process and $\{\mathcal{F}_t\}$ be its natural filtration. Then $W(t)$ is a martingale because

$$E[W_t|\mathcal{F}_s] = E[W_s + (W_t - W_s)|\mathcal{F}_s] = W_s.$$

Example 18.6.4 (Sum of independent zero-mean Random variables 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.

The first important property of a martingale is its constant mean property.

Lemma 18.6.1 (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 11.7.4], we can have $E[X_{n+1}] = E[E[X_{n+1}|\mathcal{F}_n]] = E[X_n] = ... = E[X_1]$. □

An important martingale widely used in financial modeling is the following expoential martingale.

Lemma 18.6.2 (Exponential martingale). *Let $W(t)$ be a Brownian motion process, define $Z(t) = \exp(\sigma W(t) - \frac{1}{2}\sigma^2 t)$. Then $Z(t)$ is martingale; moreover, $E[Z(t)] = E[Z(0)] = 1$.*

Proof. (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 $\sigma W(t) \sim N(0, \sigma^2 t)$. □

Example 18.6.5 (application in finance). In financial modeling, the stock price is modeled by

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

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).

Finally, we show that conditional expectation process is a martingale.

Theorem 18.6.1 (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.

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

□

18.6.3 Martingale transformation

Definition 18.6.6 (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 18.6.7 (Martingale transform). *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 18.6.3 (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 18.6.4 (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. □

18.7 Stopping time

Definition 18.7.1 (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 18.7.2 (stopping time, discrete version). *[9] 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 18.7.1. 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.

18.7.1 Stopping time examples

18.7.1.1 *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[9]

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

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.

18.7.1.2 *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.

18.7.1.3 *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.

18.7.2 Wald's equation

Theorem 18.7.1 (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

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

□

18.7.3 Optional stopping

Theorem 18.7.2 (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 18.7.2 (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.

18.7.4 martingale method for first hitting time

Lemma 18.7.1 (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 W_τ be process with τ being the stopping time hitting α or $-\beta$. W_τ is a martingale by optional stopping theorem [Theorem 18.7.2]. Then we have

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

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

18.8 Notes on bibliography

For elementary level treatment on stochastic process, see [8][4][10] and intermediate level [11]. For general SDE, see [12][11]. For treatment on forward and backward SDE, see [13].

For numerical algorithm for SDE, see [14].

For comprehensive and advanced treatment of stochastic methods, see [15] and [16].

For martingale theory, see [17].

For stationary stochastic process, see [18][2].

A good source on simulating SDE with code is at [19].

BIBLIOGRAPHY

1. Shreve, S. E. *Stochastic calculus for finance II: Continuous-time models* (Springer Science & Business Media, 2004).
2. Lindgren, G., Rootzén, H. & Sandsten, M. *Stationary stochastic processes for scientists and engineers* (CRC press, 2013).
3. Miranda, H.-C. *Applied Stochastic Analysis lecture notes* (NewYork University, 2015).
4. Mikosch, T. *Elementary stochastic calculus with finance in view* (World scientific, 1998).
5. Joshi, M. S. *The concepts and practice of mathematical finance* (Cambridge University Press, 2003).
6. Williams, D. *Probability with martingales* (Cambridge university press, 1991).
7. Björk, T. *Arbitrage theory in continuous time* (Oxford university press, 2009).
8. Brzezniak, Z. & Zastawniak, T. *Basic stochastic processes: a course through exercises* (Springer Science & Business Media, 1999).
9. Karl, S. *Stochastic modeling I lecture notes* (Columbia University, 2009).
10. Wiersema, U. F. *Brownian motion calculus* (John Wiley & Sons, 2008).
11. Klebaner, F. C. *Introduction to Stochastic Calculus with Applications* (Imperial College Press, 2005).
12. Oksendal, B. *Stochastic differential equations: an introduction with applications* (Springer Science & Business Media, 2013).
13. Ma, J. & Yong, J. *Forward-backward stochastic differential equations and their applications* **1702** (Springer Science & Business Media, 1999).
14. Higham, D. J. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM review* **43,** 525–546 (2001).
15. Gardiner, C. *Stochastic Methods: A Handbook for the Natural and Social Sciences* ISBN: 9783540707127 (Springer Berlin Heidelberg, 2009).
16. Gardiner, C. *Handbook of stochastic methods for physics, chemistry, and the natural sciences* ISBN: 9783540156079 (Springer, 1994).
17. Ladd, W. *Some Applications of Martingales to Probability Theory* tech. rep. (Technical Report, University of Chicago, 2011).
18. Lindgren, G. *Stationary Stochastic Processes: Theory and Applications* ISBN: 9781466557796 (Taylor & Francis, 2012).

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