Stochastic Process and Applications

Kazufumi Ito*

January 29, 2019

Abstract

In this monograph we cover the basic probability theory and stochastic analysis and its application in a wide class of science and engineering, including PDE theory, statistics, filtering, Data assimilation, parameter estimation, stochastic optimal control, game theory, and Financial mathematics. It is very essential that modeling of any process is analyzed using probability theory is stochastic at least in part. Stochastic systems and processes play a fundamental role in mathematical models of phenomena in many fields of science, engineering, and economics. The monograph is comprehensive and contains the basic probability theory, Markov process and the stochastic differential equations and advanced topics in nonlinear filtering, stochastic optimal control, backward stochastic differential equation —-

1 Probability Theory

In this section we discuss the basic concept and theory of the probability and stochastic process. The central objects of probability theory are to develop the mathematic tool to analyze random variables, stochastic processes, and random events. It provides the systematic and mathematical approach for analyzing a wide class of random phenomena.

1.1 Probability Triple

We introduce the probability triple $(\Omega, \mathcal{F}, P)$, which is the foundation of the probability analysis. Let Ω be a set and $\mathcal{F}$ be a collection of subsets of Ω. A point $\omega \in \Omega$ ia a sample and $A \in \mathcal{F}$ is an event. The probability measure P assigns $0 \leq P(A) \leq 1$ for each event $A \in \mathcal{F}$, i.e. the probability of event A occurs:

Definition (Probabilty Triple) The triple $(\Omega, \mathcal{F}, P)$ is the probability triple if
(1) $\mathcal{F}$ is σ-algebra, i.e.,

$$\Omega \in \mathcal{F}, \quad A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$$

$$F_n \in \mathcal{F} \Rightarrow \bigcup_n F_n \in \mathcal{F}$$

(2) P is σ-additive; for a sequence of disjoint events $\{A_n\}$ in $\mathcal{F}$,

$$P(\bigcup_n A_n) = \sum_{n=1}^{\infty} P(A_n)$$

and for $A \in \mathcal{F}$

$$P(\Omega) = 1, \quad P(A^c) = 1 - P(A).$$

*Department of Mathematics, North Carolina State University, Raleigh, North Carolina, USA

Since $(\bigcup_n F_n)^c = \bigcap_n F_n^c$, the countable intersection

$$\bigcap_n F_n \in \mathcal{F}.$$

Since $\Omega = A \cup A^c$=the disjoint union,

$$P(A) + P(A^c) = P(\Omega) = 1$$

The σ-additivity of the measure is equivalent to the monotone continuity of the measure:

Theorem (Monotone Convergence) The measure P is σ-additive if and only if for all sequence $\{A_k\}$ of nondecreasing events and $A = \bigcup_{k\geq 1} A_k$, $\lim_{n\to\infty} P(A_n) = P(A)$.

Proof: $\{A_k^c\}$ is a sequence of nonincreasing events and $A^c = \bigcap_{k\geq 1} A_k^c$. Since

$$\bigcap_{k\geq 1} A_k^c = A_1^c + (A_2^c \setminus A_1^c) + (A_3^c \setminus A_2^c) + \cdots$$

we have

$$P(A^c) = P(A_1^c) + P(A_2^c \setminus A_1^c) + P(A_3^c \setminus A_2^c) + \cdots$$

$$= P(A_1^c) + P(A_2^c) - P(A_1^c) + P(A_3^c) - P(A_1^c) + \cdots = \lim_{n\to\infty} P(A_n^c)$$

Thus,

$$P(A) = 1 - P(A^c) = 1 - (1 - \lim_{n\to\infty} P(A_n)) = \lim_{n\to\infty} P(A_n)$$

Conversely, let $A_1, A_2, \cdots \in \mathcal{F}$ be pairwise disjoint and let $\sum_{k=1}^\infty A_k \in \mathcal{F}$. Then

$$P(\sum_{k=1}^\infty A_k) = P(\sum_{k=1}^n A_k) + P(\sum_{k=n+1}^\infty A_k)$$

Since $\sum_{k=n+1}^\infty A_k \downarrow \varnothing$, we have

$$\lim_{n\to\infty} \sum_{k=n+1}^\infty A_k = \varnothing$$

and thus

$$P(\sum_{k=1}^\infty A_k) = \lim_{n\to\infty} \sum_{k=1}^n P(A_k) = \sum_{k=1}^\infty P(A_k).\square$$

Examples (σ-algebra)

$$\mathcal{F}_0 = \{\Omega, \varnothing\}, \quad \mathcal{F}^* = \text{all subsets of } \Omega.$$

Let A be a subset of Ω and σ-algebra generated by A is

$$\mathcal{F}_A = \{\Omega, \varnothing, A, A^c\}$$

Let $A,\ B$ be subsets of Ω and σ-algebra generated by $A,\ B$ is

$$\mathcal{F}_{A,B} = \{\Omega, \varnothing, A, A^c, B, B^c, A\cap B, A\cup B, A^c\cap B^c, A^c\cup B^c, A^c\cap B, A^c\cup B, A\cap B^c, A\cup B^c\}.$$

A finite set of subsets $A_1, A_2, \cdots, A_n$ of Ω which are pairwise disjoint and whose union is Ω. it is called a partition of Ω . It generates the σ-algebra: $\mathcal{A} = \{A = \bigcup_{j\in J} A_j\}$ where J runs over all subsets of $1, \cdots, n$. This σ-algebra has 2^n elements. For every finite σ-algebra is of this form, the smallest nonempty elements $\{A_1, \cdots, A_n\}$ of this algebra are called atoms.

Example (Countable measure) Let Ω has a countable decomposition $\{D_k\}$, i.e.,

$$\Omega = \sum D_k, \quad D_j \cap D_j = \varnothing, \ \ i \neq j.$$

Let $\mathcal{F} = \mathcal{F}^*$ and $P(D_k) = \alpha_k > 0$ and $\sum_k \alpha_k = 1$. For the Poisson random variable X

$$D_k = \{X = k\}, \quad P(D_k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

for $\lambda > 0$.

Example (Coin Tossing) If the cardinality of Ω is finite, then naturally we let $\mathcal{F} = \mathcal{F}^*$ and $P(\{\omega\})$, $\omega \in \Omega$ defines a measure on $(\Omega, \mathcal{F})$, i.e., $P(A) = \sum_{\omega \in A} P(\omega)$ for $A \in \mathcal{F}$. For example the case of coin tossing n-times independently is formulated as

$$\Omega = \{\omega = (b_1, \cdots, b_n),\ b_i = 0,\ 1\}$$

and $P(\omega) = p^{\sum a_i} q^{n-\sum a_i}$, where p is the probability of "Head" appears and q is the probability of "Tail" appears. The cardinality of Ω is 2^n in this case.

For the case of an infinite number of coin tossing Ω is the set of binary sequences;

$$\Omega = \{\omega = (b_1, b_2, \cdots),\ b_i = 0,\ 1\}.$$

Each number $x \in [0, 1)$ has the binary expression

$$x = \sum_{k=1}^{\infty} \frac{b_k}{2^k}$$

Thus, Ω has the cardinality of the continuum. Suppose $p = q = \frac{1}{2}$ and all samples $\omega \in \Omega$ have the same probability. Since the set $[0, 1)$ is uncountable, $P(\omega) = 0$ for each $\omega \in \Omega$.

The sets $[\frac{1}{2}, 1) = \{$"Head" appears at the first toss$\}$ and $[0, \frac{1}{2}) = \{$"Tail" appears at the first toss$\}$ should have the probability $\frac{1}{2}$. But, if we use $\mathcal{F} = \mathcal{F}^*$, then $P(A) = \sum_{\omega \in A} P(\omega) = 0$. This suggests $\mathcal{F}^*$ does not lead very far and $\mathcal{F}^*$ is too big. This leads to the concept of the Borel σ-algebra of a topological space. In summary for the probability space $(\Omega, \mathcal{F}, P)$, $\mathcal{F}$ must be closed with repeat to countable unions and intersections and complements and P must be assigned to all $A \in \mathcal{F}$ and is σ-additive.

Definition (σ-algebra) For any set C of subsets of Ω, we can define the σ-algebra $\sigma(C)$ by the smallest σ-algebra $\mathcal{A}$ which contains C. The σ-algebra $\sigma(C)$ is the intersection of all σ-algebras which contain C. It is again a σ-algebra, i.e.,

$$\mathcal{A} = \bigcap_{\alpha} \mathcal{A}_\alpha$$

where $\mathcal{A}_\alpha$ is all σ-algebras that contain C.

The following construction of the measure space and probability measure is essential for the triple $(\Omega, \mathcal{F}, P)$ on uncountable space Ω.

Construction of probability measure If $(E, \mathcal{O})$ is a topological space, where $\mathcal{O}$ is the set of open sets in E, then the σ-algebra $\mathcal{B}(E)$ generated by $\mathcal{O}$ is called the Borel σ-algebra of the topological space E and $(E, \mathcal{B}(E))$ defines the measure space and a set B in $\mathcal{B}(E)$ is called a Borel set. For example $(R^n, \mathcal{B}(R^d))$ is the the measure space and $\mathcal{B}(R^d)$ is the σ-algebra generated by open balls in R^d.

Caratheodory Theorem Let $\mathcal{B} = \sigma(\mathcal{A})$, the smallest σ-algebra containing an algebra $\mathcal{A}$ of subsets of E. Let μ_0 is a σ additive measure of on $(E, \mathcal{A})$. Then there exist a unique measure μ on $(E, \mathcal{B})$ which is an extension of μ_0, i.e., $\mu(A) = \mu_0(A), \quad A \in \mathcal{A}$.

Definition (Mesuerable functions) A map f from a measure space $(X, \mathcal{A})$ to an other measure space $(Y, \mathcal{B})$ is called measurable, if $f^{-1}(B) = \{x \in X : f(x) \in B\} \in \mathcal{A}$ for all $B \in \mathcal{B}$. If $f : (R^n, \mathcal{B}(R^n)) \to (R^m, \mathcal{B}(R^m))$ is measurable, we say f is a Borel function.

For example, for $f(x) = x^2$ on $(R, \mathcal{B}(R))$ one has $f^{-1}(([1,4]) = [1,2] \cup [-2,-1]$. In general every continuous function $R^n \to R^n$ is a Borel function since the inverse image of open sets in R^n are open in R^m.

Definition (Random Variable) A function $X : \Omega \to R$ is called a random variable, if it is a measurable map from $(\Omega, \mathcal{F})$ to $(R, \mathcal{B}(R))$, i.e., $X^{-1}(B) = \{\omega \in \Omega : (X(\omega) \in \mathcal{F}$ for all Borel set B. Every random variable X defines a σ-algebra $\mathcal{F}_X = \{X^{-1}(B) : B \in \mathcal{B}(R)\}$. which is called the σ-algebra generated by X.

Definition (Induced measure and Distribution) Let X be a random variable. Then we define the induced measure on on $(R, \mathcal{B}(R)$ by

$$\mu(B) = P(X^{-1}(B)), \quad B \in \mathcal{B}(R)$$

and the distribution function by

$$F(x) = P(X(\omega) \leq x), \quad x \in R.$$

Then, F satisfies that $x \in R \to F(x) \in R^+$ is nondecreasing, right continuous and the left limit exists everywhere and $F(-\infty) = \lim_{x\to-\infty} = 0$, $F(\infty) = \lim_{x\to\infty} = 1$. Such a function F is called a distribution function on R.

Example (Random Variable and Probability measure) Let $\Omega = R$ and $\mathcal{B}(R)$ be Borel σ-algebra. Note that

$$(a,b] = \bigcap_n (a, b+\frac{1}{n}), \quad [a,b] = \bigcap_n (a-\frac{1}{n}, b+\frac{1}{n}) \in \mathcal{B}(R).$$

Thus, $\mathcal{B}(R)$ coincides with the σ-algebra generated by the semi-closed intervals. Let $\mathcal{A}$ be the algebra of finite disjoint sum of semi-closed intervals $(a_i, b_i]$ and define P_0 by

$$P_0(\sum_{k=1}^n (a_k, b_k]) = \sum_{k=1}^n (F(b_k) - F(a_k))$$

where F is a distribution function on R. We have the measure P on $(R, \mathcal{B}(R))$ by the Caratheodory Theorem and thus a random variable $X(\omega) = \omega$ on $(\Omega, \mathcal{F}) = (R, \mathcal{B}(R))$. That is, a random variable X is uniquely identified with its distribution function F. For example if $F(x) = x$, P_0 is the Lesbegue measure dx on $(R, \mathcal{B}(R))$.

We now prove that P_0 is σ-additive on $\mathcal{A}$. By the monotone convergence theorem it suffices to prove that

$$P_0(A_n) \downarrow 0, \quad A_n \downarrow \varnothing, \quad A_n \in \mathcal{A}.$$

Without loss of the generality one can assume that $A_n \subset [-N, N]$. Since F is the right continuous, for each A_n there exists a set $B_n \in \mathcal{A}$ such that $\overline{B_n} \subset A_n$ and

$$P_0(A_n) - P_0(B_n) \leq \epsilon 2^{-n}$$

for all $\epsilon > 0$. The collection of sets $\{[-N,N] \setminus \overline{B_n}\}$ is an open covering of the compact set $[-N,N]$ since $\bigcap \overline{B_n} = \varnothing$. By the Heine-Borel theorem there exists a finite sub-covering;

$$\bigcup_{n=1}^{n_0} [-N,N] \setminus \overline{B_n} = [-N,N].$$

and thus $\cap_{n=1}^{n_0} \overline{B_n} = 0$. Thus,

$$P_0(A_{n_0}) = P_0(A_{n_0} \setminus \bigcap_{k=1}^{n_0} B_k) + P_0(\bigcap_{k=1}^{n_0} B_k) = P_0(A_{n_0} \setminus \bigcap_{k=1}^{n_0} B_k)$$

$$P_0(\bigcap_{k=1}^{n_0} (A_k \setminus B_k)) \leq \sum_{k=1}^{n_0} P_0(A_k \setminus B_k) \leq \epsilon.$$

Since $\epsilon > 0$ is arbitrary $P_0(A_n) \to 0$ as $n \to \infty$.

1.2 Exercises

Problem 1 Show that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
Problem 2 Show that $\cap_\alpha \mathcal{F}_\alpha$ is a σ-algebra.
Problem 3 Show that let X be a random variable, then $\{X^{-1}(B) : B \in \mathcal{B}(R)\}$ is a σ-algebra.
Problem 4 Show that X is a random variable if and only if $\{\omega : X(\omega) \le a\} \in \mathcal{F}$ for all $a \in R$.
Problem 5 Show that a distribution function has at most countably many discontinuities.
Problem 6 Show that if f is a Borel function and X is a random variable then $f(X)$ is a random variable.

1.3 Expectation

In this section we define the expectation of a random variable X on $(\Omega, \mathcal{F}, P)$.
Defintion (Simple random variable) A simple random variable X is defined by

$$X(\omega) = \sum^n x_i \, I_{A_k}(\omega)$$

where $\{A_k\}$ is a partition ofΩ, i.e, $A_k \in \mathcal{F}$ are disjoint and $\sum A_k = \Omega$. Then expectation of X is given by

$$E[X] = \sum x_k P(A_k).$$

Theorem (Approximation) For every random variable $X(\omega) \ge 0$ there exists a sequence of simple random variable $\{X_n\}$ such that $0 \le X_n(\omega) \le X(\omega)$ and $X_n(\omega) \uparrow X(\omega)$ for all $\omega \in \Omega$).

Proof: For $n \ge 1$, define a sequence of simple random variables by

$$X_n(\omega) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} I_{k,n}(\omega) + n I_{X(\omega)>n} \tag{1.1}$$

where $I_{k,n}$ is the indicator function of the set $\{\omega : \frac{k-1}{2^n} < X(\omega) \le \frac{k}{2^n}\}$. It is easy to verify that $X_n(\omega)$ is monotonically nondecreasing and $X_n(\omega) \le X(\omega)$ and thus $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$. $\square$

Definition (Expectation) For a nonnegative random variable X we define the expectation by

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

where the limit exists since $E[X_n]$ is an increasing number sequence.

Note that $X = X^+ - X^-$ with $X^+(\omega) = \max(0, X(\omega)),\ X^-(\omega) = \max(0, X(\omega))$. So, we can apply for Theorem and Definition for X^+ and X^-.

$$E[X] = E[X^+] - E[X^-]$$

If $E[X^+],\ E[X^-] < \infty$, X is integrable and

$$E[|X|] = E[X^+] + E[X^-].$$

Corollary Let μ_X is the induced distribution of a random variable X, i.e.,

$$\mu_X(x) = P(\{X(\omega) \le x\})$$

Then, for a Borel function $f : R \to R$ we have as $n \to \infty$

$$E[f(X_n)] = \sum^{n2^n} f(\frac{k-1}{2^n})(\mu(\frac{k}{2^n}) - \mu_X(\frac{k-1}{2^n})) + f(n)(1 - \mu_X(n)) \to \int_0^\infty f(x)\, d\mu_x(x),$$

which is the Lebesgue Stieljes integral with respect to measure $d\mu_x$. Thus,

$$E[f(X)] = \int_\Omega f(X(\omega))P(d\omega) = \int_{-\infty}^{\infty} f(x)\, d\mu(x).$$

Definition (Absolutely continuous) The measure Q is absolute continuous with respect to the measure P if

$$Q(A) = 0 \text{ if } P(A) = 0$$

for all $A \in \mathcal{F}$.

Theorem (Radon Nykodym derivative) If Q is Q is absolute continuous with respect to the measure P, then there exists a nonnegative random variable f such that

$$\int_A dQ(\omega) = \int_A f(\omega)dP(\omega)$$

for $A \in \mathcal{F}$, i.e.

$$\frac{dQ}{dP}(\omega) = f(\omega) \text{ for almost surely}$$

If μ_x is absolutely continuous with the Lebesgue measure dx, then here exists the probability density function $p(x) \geq$ of X exists, i.e.,

$$\frac{d\mu_x}{dx} = p(x) \quad (RadonNykodymderivative)$$

and

$$E[f(X)] = \int_\Omega f(X(\omega))P(d\omega) = \int_{-\infty}^{\infty} f(x)p(x)\, dx.$$

Definition (Independent Random variables) Random variables X and Y are independent if

$$P(\{X(\omega) \leq x\} \cap P(\{Y(\omega) \leq y\}) = P(\{X(\omega) \leq x\})P(\{Y(\omega) \leq y\})$$

for all $x,\ y$.

Since σ-algebra generated by semi closed intervals $(-\infty, x],\ x \in R$ equals to $\mathcal{B}(R)$, the definition is equivalent to

$$P(X^{-1}(B_1)\} \cap P(\{Y^{-1}(B_2)) = P(X^{-1}(B_1))P(Y^{-1}(B_2)$$

for all Borel sets $B_1,\ B_2$.

Theorem (Independent Random Variables) For independent random variables X and Y and Borel function $f,\ g$ if $f(X)$ and $g(Y)$ are integrable, then the product $f(X)g(Y)$ is integrable and

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)]$$

Proof: If X and Y are simple random variables, i.e.,

$$X = \sum x_k\, A_k(x), \quad Y = \sum y_j\, B_j(x),$$

where $\{A_k\}$ and $\{B_j\}$ are partitions of $(\Omega, \mathcal{F})$, then

$$E[f(X)g(Y)] = \sum f(x_k)g(y_j)P(X^{-1}(A_k))P(Y^{-1}(B_j))$$

$$= \sum_k f(x_k)P(X^{-1}(A_k)) \sum_j g(y_j)P(Y^{-1}(B_j)) = E[f(X)]E[g(Y)]$$

since

$$P(X^{-1}(A_k) \cap Y^{-1}(B_j)) = P(X^{-1}(A_k))P(Y^{-1}(B_j)).$$

In general we approximate $X,\ Y$ by a sequence of simple random variables by the approximation (**??**). $\square$

Theorem (Convolution) If X and Y are independent random variables, and let $F(x) = P(X \le x)$, and $G(y) = P(Y \le y)$, then

$$P(X+Y \le z) = \int F(z-y)dG(y).$$

The integral on the right-hand side is called the convolution of F and G and is denoted by $(F*G)(z)$.

Proof: Let $h(x,y) = I_{x+y\le z}$. Let μ and ν be the probability measures with distribution functions F and G. Since for fixed y

$$\int h(x,y)\mu(dx) = \int I_{(-\infty,z-y]}(x)\mu(dx) = F(z-y)$$

Thus,

$$P(X+Y \le z) = I_{x+y\le z}\mu(dx)\nu(dy) = \int F(z-y)\nu(dy) = \int F(z-y)dG(y).\square$$

For a random vector $X(\omega) = (X_1(\omega), \cdots, X_n(\omega)$ in R^n we define the mean m and variance R by

$$m = E[X] = \textstyle\int_{R^n} x\, d\mu_X(x)$$

$$R_{ij} = \mathrm{Var}(X) = E[(X_i-m_i)(X_j-m_j)] = \textstyle\int_{R^n} (x_i-m_i))x_j-m_j)\, d\mu_X(x).$$

The variance matrix R is symmetric and nonnegative definite since

$$\xi^t R\xi = E[|\sum_k \xi_k(X_k-m_k)|^2] \ge 0$$

for all $\xi \in R^n$.

Random Variable For $m \in R^n$ and $R > 0$ is symmetric positive definite matrix on R^n and

$$p(x) = \frac{1}{(2\pi)^{\frac{n}{2}}|R|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-m)^t R^{-1}(x-m)}$$

is the probability density function of a random vector X, where $|R| = det(R)$, the determinant of R, i.e.,

$$\mu(B) = \int_B p(x)\, dx, \quad B \in \mathcal{B}(R^n)$$

defines the Gaussian distribution on R^n. Thus, a Gaussian random variable is completely determined by its statistics (m, R).
Square integrable Random variables Define

$$L^2(\Omega, P) = \text{the space of square integrable random variable, } E[|X|^2] < \infty.$$

Then $L^2(\Omega,P)$ is a vector space, i.e.,

$$E[|\alpha\, X + \beta Y|^2] \le 2(\alpha^2 E[|X|^2] + \beta^2 E[|Y|^2])$$

since

$$|E[XY]| \le \sqrt{E|X|^2}\sqrt{E[|Y|^2]}.$$

In fact, if we define the inner product by

$$(X,Y)_{L^2} = E[XY]$$

then it is a complete inner product space, i.e., a Hilbert space.

1.4 Stochastic Process

A stochastic process is a parametric collections of random variables $\{X_t\}_{t\in T}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ and taking values in the state space S. The space T is either discrete time $T = 0, 1, \cdots$ or $T = [0, \infty)$. The state space S is a complete merit space. That is, for each $t \in T$

$$\omega \in \Omega \to X_t(\omega) \in S \text{ is a random variable.}$$

On the other hand, for each $\omega \in \Omega$

$$t \to X_t(\omega)$$

defines a sample path of X_t. Thus, $X_t(\omega)$ represents the value at time $t \in T$ of a sample $\omega \in \Omega$ and it may be regarded as a function of two variables:

$$(t, \omega) : T \times \Omega \to X(t, \omega) \in S$$

and we assume that $X(t, \omega)$ is jointly measurable on $(T \times \Omega, \mathcal{B}(T) \times \mathcal{F})$.

Let Ω be a subset of the product space S^T of functions $t \to X(t, \omega)$ from $t \in T$ to S. The σ-algebra $\mathcal{F}$ contains the σ-algebra $\mathcal{B}(S^T)$ generated by the cylinder sets of the form

$$\{\omega : X_{t_1} \in B_1, \cdots, X_{t_n} \in B_n\} \tag{1.2}$$

for all $t_1, \cdots, t_n \in T,\ n \in N$ and Borel sets B_k in $\mathcal{B}(S)$. Therefore, we adopt the point of view that a stochastic process is a probability measure on the measure space $(S^T, \mathcal{B}(S^T))$.

Definition (Finite dimensional distribution) The finite dimensional distribution of the stochastic process X_t are the measures defined $\mu_{t_1,\cdots,t_n}$ on S^n;

$$\mu_{t_1,\cdots,t_n}(F_1 \times \cdots \times F_n) = P(X_{t_1} \in F_1, \cdots, X_{t_n} \in F_n).$$

for all $t_k \in T,\ n \in N$ and Borel sets F_k of S. The family of finite dimensional distributions determines the statistical properties of the process X_t. Conversely, a given family of $\{\nu_{t_1,\cdots,t_n},\ t_k \in T,\ n \in N\}$ of probability measure on S^n with the two natural consistency conditions it follows from the Kolmogorov's extension theory we are able to construct a stochastic process;

Theorem (Kolmogorov's extension theory) For all $t_1, \cdots, t_n$, let $\nu_{t_1,\cdots,t_n}$ be the probability mesures on S^n satisfying

$$\nu_{t_{\pi(1)},\cdots,t_{\pi(n)}}(B_1 \times \cdots \times B_n) = \nu_{t_1,\cdots,t_n}(B_{\pi^{-1}(1)} \times \cdots \times B_{\pi^n(n)})$$

for all permutations π on $\{1, \cdots, n\}$ and

$$\nu_{t_1,\cdots,t_n}(B_1 \times \cdots \times B_n) = \nu_{t_1,\cdots,t_n,t_{n+1},\cdots t_{n+m}}(B_1 \times \cdots \times B_n \times S \cdots \times S)$$

there exits a probability space $(\Omega, \mathcal{F}, P)$ and a stochastic process X_t on Ω such that

$$\nu_{t_1,\cdots,t_n}(B_1 \times \cdots \times B_n) = P(X_{t_1} \in B_1, \cdots, X_{t_n} \in B_n),$$

for all $t_k \in T,\ n \in N$ and all Borel sets B_k.

Example (Brownian motion)

A stochastic process $B_t,\ t \geq 0$ is called a Brownian motion if it satisfies the following conditions:
i) For all $0 \leq t_1 < \cdots < t_n$ the increments $B_{t_n} - B_{t_{n-1}}, \cdots B_{t_2} - B_{t_1}$ are independent random variables.
ii) If $0 \leq s < t$, the increment $B_t - B_s$ has the normal distribution $N(0, t - s)$.
Based on the conditions we have

$$\nu_{t_1,\cdots,t_n}(F_1 \times \cdots \times F_n) = \int_{F_1 \times \cdots \times F_n} p(t_1, x, x_1) p(t_2 - t_1, x_1, x_2) \cdots p(t_n - t_{n-1}, x_{n-1}, x_n)\, dx_1 \cdots dx_n$$

where

$$p(t,x,y) = (2\pi t)^{-\frac{n}{2}} e^{-\frac{|x-y|^2}{2t}}.$$

We will discuss the properties of the Brownian motion in Chapter . It will be shown that $t \to B_t(\omega)$ is continuous.

Measure space $((C, \mathcal{B}(C))$ Let $T = [0,1]$ and let C be the space of continuous functions x_t on T. Let $\mathcal{B}_0$ be the σ-algebra generated by open sets of C with respect to the metric

$$d(x,y) = \max_{t \in [0,1]} |x_t - y_t|, \quad x, y \in C.$$

We show that $\mathcal{B}_0$ coincides with the σ-algebra generated by cylinder sets (**??**). Let $B = \{x \in C : x_{t_0} < b\}$ be a cylinder set. It is easy to see that B is an open set. Hence every cylinder set of the form $\{x : x_{t_1} < b_1, \cdots x_{t_n} < b_n\} \in \mathcal{B}_0$ and thus $\mathcal{B}(C) \subset \mathcal{B}_0$. Conversely, consider and open ball $S_\rho(x_0) = \{d(x,x_0) < \rho\}$. Since the function x_t is continuous,

$$B_\rho(x_0) = \{y \in C : \max_{t \in T} |y_t - x^0_t| < \rho\} = \cup_{t_k} \{y \in C : |y_{t_k} - x^0_{t_k}| < \rho\},$$

where $\{t_k\}$ are the rational points of $[0,1]$. Therefore, $\mathcal{B}_0 \subset \mathcal{B}(C)$. $\square$

1.5 Convergence of Stochastic Process

In this section we discuss the convergence of a sequence of Random variables $\{X_n\}$.

Theorem. If $\{X_n\}$ is a sequence of random variables, then so are

$$\inf_n X_n, \quad \sup_n X_n, \quad \limsup_n X_n, \quad \liminf_n X_n$$

Proof: The theorem follows from the facts;

$$\{\inf_n X_n < a\} = \cup_n \{X_n < a\} \in \mathcal{F}$$

$$\{\sup_n X_n > a\} = \cup_n \{X_n > a\} \in \mathcal{F}.$$

$$\liminf_n X_n = \sup_n(\inf_{m \geq n} X_m)$$

$$\limsup_n X_n = \inf_n(\inf_{m \geq n} X_m).\square$$

From Theorem, we see that

$$\{\omega : \lim_n X_n(\omega) \text{ exists}\} = \{\omega : \liminf_n X_n(\omega) = \limsup_n X_n(\omega)\}$$

is a measurable set.

Borel-Cantelli Lemma If $\sum P(A_n) < \infty$ then $P(A_n$ occurs infinitely manny time$) = 0$.
Proof: Note that

$$A_n \text{ occurs infinitely manny time}) = \limsup A_n = \cap_n^\infty \cup_{k \geq n}^\infty A_k.$$

Thus,

$$P(A_n \text{ occurs infinitely manny time}) = \lim_{n \to \infty} P(\cup_{k \geq n}^\infty A_k) \leq \lim_{n \to \infty} \sum_{k \geq n} P(A_k) = 0.$$

Fatou's Lemma Let $\{f_k\}$ be a sequence of non-negative measurable functions on a measure space $(E, \mathcal{B}, \mu)$. Define the function $f? : E \to [0, \infty]$ a.e. pointwise limit by

$$f(x) = \liminf_{n \to \infty} f_n(x), \quad x \in E.$$

Then $f?$ is measurable and

$$\int_E f\,d\mu \le \liminf_{n\to\infty} \int_E f_n\,d\mu.\square$$

Proof: For every k define pointwise the function

$$g_k = \inf_{n\ge k} f_n.$$

Then the sequence $\{g_k\}$ is increasing and converges pointwise to f. Since for all $k \le n$ we have $g_k \le f_n$,

$$\int_E g_k\,d\mu \le \int_E f_n\,d\mu,$$

and hence

$$\int_E g_k\,d\mu \le \inf_{n\ge k} \int_E f_n\,d\mu.$$

Using the monotone convergence theorem for the first equality, then the last inequality from above, and finally the definition of the limit inferior, it follows that

$$\int_E f\,d\mu = \lim_{k\to\infty} \int_E g_k\,d\mu \le \lim_{k\to\infty} \inf_{n\ge k} \int_E f_n\,d\mu = \liminf_{n\to\infty} \int_E f_n\,d\mu.\square$$

Theorem (Lebesgue's Dominated Convergence Theorem) Let $\{f_n\}$ be a sequence of real-valued measurable functions on a measure space $(E, \mathcal{B}, \mu)$. Suppose that the sequence converges pointwise to a function f a.e. and is dominated by some integrable function g in the sense that $|f_n(x)| \le g(x)$ a.e. Then f is integrable and

$$\lim_{n\to\infty} \int_E |f_n - f|\,d\mu = 0$$

which also implies

$$\lim_{n\to\infty} \int_E f_n\,d\mu = \int_E f\,d\mu$$

Proof: Since f is the pointwise limit of the sequence f_n of measurable functions that is dominated by g, it is also measurable and dominated by g, hence it is integrable. Note that

$$|f - f_n| \le |f| + |f_n| \le 2g$$

for all n and

$$\limsup_{n\to\infty} |f - f_n| = 0, \quad a.e.$$

Since $2g - |f - f_n| \ge 0$ it follows from the Fatou's lemma that

$$\limsup_{n\to\infty} \int_E |f - f_n|\,d\mu \le \int_E \limsup_{n\to\infty} |f - f_n|\,d\mu = 0,$$

which implies that

$$\lim_{n\to\infty} \int_E |f - f_n|\,d\mu = 0.\square$$

Let $\{X_n\}$ be a sequence of random variables. The we have the following convergence notions to a random variable X;

Definition (a) almost surely (a.s.) convergence:

$$X_n(\omega) \to X(\omega) \text{ except zero } P\text{-measure set.}$$

(b) L^p convergence: for $p \ge 1$ let $L^p(\Omega, P)$ be the Banach space of L^p integrable random variables with

$$|X|_p = E[|X|^p]^{\frac{1}{p}}.$$

X_n converges to X in L^p if
$$|X_n - X|_p \to 0 \text{ as } n \to \infty.$$
(c) convergence in probability: X_n converges in probability if

$$P(|X_n - X| > \epsilon) \to 0 \text{ as } n \to \infty$$

for all $\epsilon > 0$.
(d) convergence in distribution; X_n converges to X in distribution

$$E[f(X_n)] \to E[f(X)]$$

for all bounded continuous function f.

In the followings we present the relationships between the different convergence notions of a sequence of random variables $\{X_n\}$.

Theorem If X_n converges to X in probability implies the convergence in distribution.

Proof: Note that for random variables X, Y

$$\begin{aligned}
P(Y \leq a) &= P(Y \leq a,\ X \leq a+\epsilon) + P(Y \leq a,\ X > a+\epsilon) \\
&\leq P(X \leq a+\epsilon) + P(Y - X \leq a - X,\ a - X < -\epsilon) \\
&\leq P(X \leq a+\epsilon) + P(Y - X < -\epsilon) \\
&\leq P(X \leq a+\epsilon) + P(Y - X < -\epsilon) + P(Y - X > \epsilon) \\
&= P(X \leq a+\epsilon) + P(|Y - X| > \epsilon).
\end{aligned}$$

It can be shown that in order to prove convergence in distribution, one must show that the sequence of the distributions F_{X_n} of X_n converges to the one F_X of X at every point where F_X is continuous. Let a be such a point. For every $\epsilon > 0$, we have:

$$\begin{aligned}
P(X_n \leq a) &\leq P(X \leq a+\epsilon) + P(|X_n - X| > \epsilon) \\
P(X \leq a-\epsilon) &\leq P(X_n \leq a) + P(|X_n - X| > \epsilon)
\end{aligned}$$

and

$$P(X \leq a-\epsilon) - P(|X_n - X| > \epsilon) \leq P(X_n \leq a) \leq P(X \leq a+\epsilon) + P(|X_n - X| > \epsilon).$$

Taking the limit as $n \to \infty$, we obtain:

$$F_X(a-\epsilon) \leq \lim_{n\to\infty} P(X_n \leq a) \leq F_X(a+\epsilon),$$

Since F_x is continuous at a
$$\lim_{n\to\infty} F_{X_n}(a) = F_X(a). \square$$

The space $\mathcal{L}$ of random variables is a complete metric space with metric

$$d(X,Y) = E[(\frac{|X-Y|}{1+|X-Y|}]$$

In fact, $d(X,Y) = 0$ if and if $X = Y$ almost surely and since

$$\frac{|X+Y|}{1+|X+Y|} \leq \frac{|X|}{1+|X|} + \frac{|Y|}{1+|Y|},$$

d satisfies the triangle inequality.

Theorem (completeness) $(\mathcal{L}, d)$ is a complete metric space.

Proof: Let $\{X_n\}$ be a Cauchy sequence of $(\mathcal{L}, d)$. Select a subsequence X_{n_k} such that

$$\sum_{k=1}^{\infty} d(X_{n_k}, X_{n_{k+1}}) < \infty$$

Then,

$$\sum_k E[\frac{|X_{n_k} - X_{n_{k+1}}|}{1 + |X_{n_k} - X_{n_{k+1}}|}] < \infty$$

Since $|X| \leq \frac{2|X|}{1+|X|}$ for $|X| \leq 1$,

$$\sum^{\infty} |X_{n_k} - X_{n_{k+1}}| < \infty \text{ a.s.}$$

and $\{X_{n_k}\}$ almost surely converges to $X(\omega)$ for $X \in \mathcal{L}$. Moreover, $d(X_{n_k}, X) \to 0$ and thus $(\mathcal{L}, d)$ is complete.

Theorem X_n converges in probability to X if and only if X_n converges to X in d-metric.

Proof: For $X,\ Y \in \mathcal{L}$

$$d(X,Y) = \int_{|X-Y| \geq \epsilon} \frac{|X-Y|}{1+|X-Y|} dP + \int_{|X-Y|<\epsilon} \frac{|X-Y|}{1+|X-Y|} dP \leq P(|X-Y| \geq \epsilon) + \frac{\epsilon}{1+\epsilon}$$

holds for all $\epsilon > 0$. Thus, X_n converges in probability to X, then X_n converges to X in d-metric. Conversely, since

$$d(X,Y) \geq \frac{\epsilon}{1+\epsilon} P(|X-Y| \geq \epsilon)$$

if X_n converges to X in d-metric, then X_n converges in probability to X. □

Theorem (a) If either X_n converges to X in L^p or converges to a.s. to X, then X_n converges to X in probability.
(b) If X_n converges to X in probability, there exists a subsequence that converges a.s. to X.

Proof: Since $E[|X|] \leq (E|X|^p)^{\frac{1}{p}} E[1] = |X|_p$ and $d(X, 0) \leq |X|$, the L^p convergence implies the convergence in probability. If X_n converges to X a.s., then by the bounded convergence theorem $d(X_n, X) \to 0$ and thus X_n converges to X in probability. Conversely, if X_n converges to X in probability, then $d(X_n, X) \to 0$ and as shown in the proof of Theorem (completeness) there exists a subsequence that converges a.s. to X.

In summary we have the chain of implications between the various notions of convergence:

$$\begin{array}{ccccc} \xrightarrow{L^s} & \underset{s>r\geq 1}{\Rightarrow} & \xrightarrow{L^r} & & \\ & & \Downarrow & & \\ \xrightarrow{a.s.} & \Rightarrow & \xrightarrow{p} & \Rightarrow & \xrightarrow{d} \end{array}$$

1.6 Uniform integrable random variables

Definition (Uniform integrable) A sequence of random variables $\{X_n\}$ is uniformly integrable if

$$\sup_n \int_{|X_n| \geq c} |X_n| \, dP \to 0 \text{ as } c \to \infty.$$

Theorem (Uniform Integrable I) If $\{X_n\}$ is uniformly integrable, then

$$E[\liminf X_n] \leq \liminf E[X_n] \leq \limsup E[X_n] \leq E[\limsup X_n]$$

Theorem (Uniformly integrable II) A sequence $\{X_n\}$ of random variables converging in probability to X also converge to X in L^1 if and only if they are uniformly integrable.

Proof: Suppose $\{X_n\}$ is uniformly integrable. Since for $Y_n = X_n - X$

$$d(Y_n,0) \geq \int_{|Y_n \leq c} \frac{|Y_n|}{1+|Y_n|}\,dP \geq \frac{1}{1+c}\int_{|Y_n|\leq\epsilon} |Y_n|\,dP,$$

and thus

$$\int_\Omega |Y_n|\,dP \leq (1+c)d(X_n,0) + \int_{|Y_n|\geq c} |Y_n|\,dP.$$

For arbitrary $\epsilon > 0$ if we choose c such that $\sup_n \int_{|Y_n|\geq c} |Y_n| < \epsilon$, then

$$\int_\Omega |Y_n|\,dP \leq (1+c)\,d(Y_n,0) + \epsilon$$

Since $d(X_n,0) \to 0$, $\limsup_{n\to\infty} |Y_n| \leq c$ a.s. and thus $|X_n - X|_1 = |Y_n|_1 \to 0$.

Conversely, suppose $|X_n - X| \to 0$. For arbitrary $\epsilon > 0$ we choose n_0 such that $\sup_{n\geq n_0} E[|Y_n|] < \epsilon$. Then, for all $c > 0$

$$\sup_{n\geq n_0} \int_{|Y_n|>c} |Y_n|\,dP < \epsilon.$$

If for each $1 \leq k \leq n_0 - 1$ we choose c_k such that

$$\int_{|Y_k|>c_k} |Y_k|\,dP < \epsilon.$$

and let $c = \max 1 \leq k \leq n_0 - 1\, c_k$, we have

$$\sup_n \int_{|Y_n|>c} |Y_n|\,dP < \epsilon.\square$$

Lemma Let G be a nonnegative increasing function on R^+ such that $\lim_{t\to\infty} \frac{G(t)}{t} \to \infty$. If

$$\sup_n E[G(|X_n|)] < \infty$$

then $\{X_n\}$ is uniformly integrable.

Proof: For $\frac{G(t)}{t} > a,\ t > c$ we have

$$\sup_n \int_{|X_n|>c} |X_n|\,dP \leq \frac{1}{a}\sup_n \int_{|X_n|\geq c} G(|X_n|)\,dP \leq \frac{1}{a}\sup_n E[G(|X_n|)].\square$$

Theorem (Kolmogorov)

1.7 Conditional Expectation

Definition Let X be a random variable and $\mathcal{A}$ be a σ-algebra. The conditional expectation $E[X|\mathcal{A}]$ is a $\mathcal{A}$ random variable that satisfies

$$E[I_A\,E[X|\mathcal{A})] = E[I_A\,X] \tag{1.3}$$

for all $A \in \mathcal{A}$.

Note that $Q(A) = E[I_A\,X],\ A \in \mathcal{A}$ for a nonnegative random variable X defines a measure Q on $(\Omega, \mathcal{A})$ and if $P(A) = 0$ implies $Q(A) = 0$ (i.e. Q is absolutely continuous with respect to P). By the Radon-Nikodym theorem the conditional expectation exists as the Radon-Nikodym derivative $\frac{dQ}{dP} = E[X|\mathcal{A}]$.

Condition (**??**) is equivalent to the orthogonality condition;

$$E[Z\,(X - E[X|\mathcal{A}])] = 0 \text{ for all } \mathcal{A}\text{-measurable random variables } Z. \tag{1.4}$$

In fact, (**??**) follows from letting $Z = I_A$. Conversely, we approximate Z by a sequence of $\mathcal{A}$-measurable simple random variables

$$Z_n = \sum z_k\, I_{A_k}, \quad A_k \in \mathcal{A}.$$

Let $L^2(\Omega, \mathcal{F}, P)$ be a space of square integrable random variables and define the inner product by

$$(X, Y)_{L^2} = E[XY]$$

Then, $L^2(\Omega, \mathcal{F}, P)$ is a Hilbert space. Moreover $\hat{X} = E[X|\mathcal{A}]$ minimizes

$$E[|X - Z|^2] \text{ over all } \mathcal{A}\text{-measurable square integral random variables}$$

In fact,

$$E[|X - Z|^2] = E[|X - \hat{X}|^2 + 2(X - \hat{X})(\hat{X} - Z) + |Z - \hat{X}|^2] = E[|X - \hat{X}|^2] + E[|Z - \hat{X}|^2].$$

That is, $E[X|\mathcal{A}]$ is the orthogonal projection of X onto the subspace space of $\mathcal{A}$-measurable random variables of $L^2(\Omega, \mathcal{F}, P)$.
Example (1) If $\{G_k\}$ in $\mathcal{F}$ is a partition of Ω and $\mathcal{G}$ is the σ-algebra generated by $\{G_k\}$, then

$$E[I_A|\mathcal{G}] = P(A|\mathcal{G}) = \frac{P(A \cup G_k)}{P(G_k)} \quad \omega \in G_k.$$

If $\mathcal{G}_1$ and $\mathcal{G}_2$ be independent σ-algebras,

$$P(A|\mathcal{G}_2) = P(A)$$

Conversely, if the above holds for all $A \in \mathcal{G}_1$, then $\mathcal{G}_1$ and $\mathcal{G}_2$ are independent
(3) If X, Y are random variables

$$P(X \in B|Y = y) = \int_B \frac{p_{X,Y}(x, y)}{p_Y(y)}\, dx$$

where $p_{X,Y}$ is the joint density function of (X, Y) and and $p_Y(y)$ is the marginal density of Y. In fact,

$$P(X \in B|y-\delta < Y \leq y+\delta) = \frac{P(\{X \in B\} \cap \{y - \delta < Y \leq y + \delta\})}{P(y - \delta < Y \leq y + \delta)} = \frac{\int_B \int_{y-\delta}^{y+\delta} p_{X,Y}(x, y)\, dxdy}{\int_{y-\delta}^{y+\delta} p_Y(y)\, dy}$$

for $\delta > 0$.

Theorem (Property of Conditional Expectation)
(1) $E[E[X|\mathcal{H})|\mathcal{A}] = E[X|\mathcal{A}]$ for $\mathcal{A} \subseteq \mathcal{H}$.
(2) $E[X|\mathcal{A}] = E[X]$, if X is independent with $\mathcal{A}$.
(3) $E[Z\,X|\mathcal{A}] = Z\,E[X|\mathcal{A}]$ if Z is $\mathcal{A}$ measurable.
(4) $|E[X|\mathcal{A}]| \leq E[|X||\mathcal{A}]$.
(5) (Jensen's inequality) If f is convex function, the

$$\phi(E[X|\mathcal{A}]) \leq E[\phi(X)|\mathcal{A}].$$

Proof: (1) For all $H \in \mathcal{H}$

$$\int_H E[X|\mathcal{H}]dP = \int_H XdP = \int_H E[X|\mathcal{A}]dP.$$

(2) For all $A \in \mathcal{A}$ we have

$$\int_A E[X|\mathcal{A}]dP = \int I_A X dP = E[X]P(A) = \int_A [X]dP$$

(3) If we let $f = I_G,\ G \in \mathcal{A}$ we have

$$\int_A I_G E[X|\mathcal{H}]dP = \int_{A\cup G} E[X|\mathcal{A}]dP = \int_{A\cup G} X dP = \int_A I_G X = \int_G E[I_G X|\mathcal{A}]dP$$

for all $A \in \mathcal{A}$ and thus

$$E[I_X|\mathcal{A}] = I_G E[X|\mathcal{A}].$$

1.8 Characteristic Functions and Weak convergence

A sequence of random variables $\{X_n\}$ is converge weakly or converge in distribution to X if and only if distribution functions $F_n(x) = P(X_n \le x)$ converge to $F(x) = P(X \le x)$ for all x at which $F(x)$ is continuous and $\{F_n\}$ is tight, i.e., for arbitrary $\epsilon > 0$ thee exists C such that

$$|\int_{|x|\ge C} dF_n(x)| \le \epsilon. \tag{1.5}$$

To see that convergence at continuity points is enough to identify the limit, observe that F is right continuous and the discontinuities of F are at most a countable set.

Theorem (Weak Convergence) A sequence of random variables $\{X_n\}$ is converge weakly or converge in distribution to X if and only if $\{\mu_n\}$ tight and converge weakly to F.

Proof: By the bounded convergence theorem for g is a bounded continuous function

$$E[g(X_n)] \to E[g(X)]$$

if F_n converges to F weakly. Conversely, assume that $F(x)$ is continuous at x and define

$$g_\epsilon(y) = \begin{cases} 1 & y \le x \\ \dfrac{x+\epsilon-y}{\epsilon} & y \in [x, x+\epsilon] \\ 0 & y \ge x+\epsilon. \end{cases}$$

Since g_ϵ is continuous, and $g_\epsilon = 0$ for $y \ge x+\epsilon$,

$$\limsup P(X_n \le x) \le \limsup E[g_\epsilon(X_n)] = E[g_\epsilon(X_n)] \le P(X \le x+\epsilon).$$

Letting $\epsilon \to 0^+$, we have $\limsup P(X_n \le x) \le P(X \le x)$. Also,

$$\liminf P(X_n \le x) \ge \liminf E[g_\epsilon(X_n - \epsilon)] = E[g_\epsilon(X-\epsilon)] \ge P(X \le x-\epsilon).$$

Letting $\epsilon \to 0+$, we have $\liminf P(X_n \le x) \ge P(X \le x)$. Combining these, we have $F(x_n) \to F(x)$ as $n \to \infty$. $\square$

The characteristic function plays an important role in the weak convergence of random variables.

Definition For $X \in R^n$ is random vector the characteristic function of X is defined by

$$\varphi(\xi) = E[e^{i(\xi,X)}] = \int_{R^n} e^{i(\xi,x)}dF(x), \quad \xi \in R^n,$$

where F is the distribution of X.

Theorem The characteristic function of the random variable X $t \in R \to \varphi(t) = E[e^{itX}]$ satisfies;
(1) $|\varphi(t)| \le \varphi(0) = 1$.
(2) $\varphi(t)$ is uniformly continuous.
(3) $\varphi(t) = \overline{\varphi(-t)}$.
(4) $\varphi(t)$ is real-valued if and only if F is symmetric.
(5) If $E[|X|^n] < \infty$ for some $n \ge 1$, then $\varphi^{(n)}(t)$ exists for all $r \le n$,

$$\varphi^{(r)}(t) = \int_R (ix)^r e^{itx}\, dF(x), \quad (i)^r\, E[X^r] = \varphi^{(r)}(0),$$

and

$$\varphi(t) = \sum_{r=0}^n \frac{(it)^r}{r!}\, E[X^r] + \frac{(it)^n}{n!} \epsilon_n(t),$$

where $|\epsilon_n(t)| \le 3E[|X|^n]$ and $\epsilon_n(t) \to 0$ as $t \to 0$.
(6) If $\varphi^{(2n)}(0)$ exists and is finite, then $E[X^{2n}] < \infty$.
(7) If $E[|X|^n] < \infty$ for all $n \ge 1$ and $\limsup \frac{(E[|X|^n])^{\frac{1}{n}}}{n} = \frac{1}{eR} < \infty$, then

$$\varphi(t) = \sum_{n=0}^\infty \frac{(it)^n}{n!}\, E[|X|^n] \text{ for all } |t| < R.$$

Integrating by parts gives

$$\int_0^x (x-s)^n e^{is}\, ds = \frac{x^{n+1}}{n+1} + \frac{i}{n+1} \int_0^x (x-s)^{n+1} e^{is}\, ds \tag{1.6}$$

For $n = 0$

$$\int_0^x e^{is}\, ds = x + i \int_0^x (x-s) e^{is}\, ds$$

and thus

$$e^{ix} = 1 + i\,x - \int_0^x (x-s) e^{is}\, ds$$

By () with $n = 1$

$$e^{ix} = 1 + i\,x - \frac{x^2}{2} - i \int_0^x (x-s)^2 e^{is}\, ds$$

By induction

$$e^{ix} = \sum_{k=1}^n \frac{(ix)^k}{k!} + \frac{i^[n+1}{n!} \int_0^x (x-s)^n e^{is}\, ds$$

Since

$$E[|\int_0^{tX} (tX-s)^2 e^{is}\, ds|] \le t^2 E[(|t||X|^3 \wedge |X|^2].$$

we have

$$E[e^{itX}] = 1 + i\,tE[X] - \frac{t^2}{2} E[X^2] + o(t^2).$$

Theorem (Lévy continuity theorem) Let $\{F_n\}$ be a sequence of probability measures with characteristic function $\{\varphi_n\}$. (1) If F_n converges weakly to F, then $\varphi_n(t) \to \varphi(t)$ for all t. (2) If $\varphi_n(t)$ converges pointwise to φ that is continuous at 0, then the associated sequence of distributions $\{F_n\}$ is tight and converges weakly to the measure F with characteristic function φ.

Proof: For (1) since e^{its} is bounded continuos,

$$\varphi_n(t) = E[e^{itX_n}] \to E[e^{itx}] = \varphi(t).$$

For (2) note that

$$\int_{-c}^{c}(1-e^{itx})\,dt = 2(1-\frac{\sin(cx)}{x})$$

and thus

$$\frac{1}{c}\int_{-c}^{c}(1-\varphi_n(t))\,dt \le 2\int(1-\frac{\sin cx}{cx})\,dF_n(x) \ge 2\int_{|x|\ge\frac{2}{c}}(1-\frac{1}{cx})\,dF_n(x) \ge \mu_n(|x|\ge\frac{2}{c}).$$

Since $\varphi(t)\to 1$ sas $t\to 0$ there exists for c_ϵ such that $c\le c_\epsilon$

$$|\frac{1}{c}\int_{-c}^{c}(1-\varphi(t)\,dt|\le\epsilon.$$

But, since $\varphi(t)\to\varphi(t)$, by the bounded convergence theorem there exist N_ϵ such that for $n\ge N_\epsilon$

$$2\epsilon \ge \frac{1}{c}\int_{-c}^{c}(1-\varphi_n(t))\,dt \ge \mu_n(|x|\ge\frac{2}{c}).$$

and thus the sequence $\{\mu_n\}$ of measures is tight.

1.9 Law of Large numbers

The law of large numbers (LLN) describes the result of performing the same experiment a large number of times. According to the law, the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed.

Theorem (L^2 Weak Law of Large Number) Let $X_1, X_2, \cdots$ be uncorrelated random variables with $E[X_k]=\mu$ and $var(X_k)\le C$. If $S_n=\sum_{k=1}^n X_k$ then as $n\to\infty$ $\frac{S_n}{n}\to\mu$ in L_2 and in probability.

Proof:

$$E[|\frac{S_n}{n}-\mu|^2]=Var(\frac{S_n}{n})=\frac{1}{n^2}\sum Var(X_k)\le\frac{C}{n}\to 0.$$

By the Chebyshev inequality,

$$P(|\frac{S_n}{n}-\mu|>\epsilon)\le\frac{1}{\epsilon^2}E[|\frac{S_n}{n}-\mu|^2].$$

Theorem (Weak Law of Large numbers) Let $X_1, X_2, \cdots$ be i.i.d. random variables with $E[|X_k|]<\infty$. Then, $\frac{S_n}{n}\to\mu$ in probability.
Proof: By the Lebesgue dominated convergence theorem

$$xP(|X_1|>x)\le E[|X_1|I_{\{|X_1|>x\}}]\to 0 \text{ as } x\to\infty$$

$$\mu_n=E[X_1I_{\{|X_1|<n\}}\to E[X_1]=\mu \text{ as } n\to\infty.$$

Since $P(|\frac{S_n}{n}-\mu_n|>\epsilon)\to 0$.
Theorem (Strong law of large numbers) Let $X_1, X_2, \cdots$ be i.i.d. random variables with $E[|X_k|]<\infty$. Then, $\frac{S_n}{n}\to\mu$ a.s. as $n\to\infty$.

Proof: Let $Y_k=X_kI_{\{|X_k|\le k\}}$ and $T_n=\sum_{k=1}^n Y_k$. Since

$$\sum_k P(|X_k>k)\le\int^{\infty}P(X_1>t)\,dt=E[|X_1|]$$

by the Borel-Cantelli lemma, $P(\{\omega : X_k(\omega) = T_k(\omega) \text{ for infinitely many } k = 0\}) = 0$ and thus $\frac{S_n}{n} - \frac{T_n}{n} \to 0$ a.s. as $n \to \infty$. It suffices to show that $\frac{T_n}{n} \to \mu$ a.s. as $n \to \infty$. Let $Z_k = Y_k - E[Y_k]$, so $E[Z_k] = 0$. Now $Var(Z_k) = Var(Y_k) \leq E[|Y_k|^2]$ and

$$\sum_{k=1}^{\infty} \frac{Var(Z_k)}{k^2} \leq \sum_{k=1}^{\infty} \frac{E[|Y_k|^2}{k^2} < 4E[|X_1|].$$

We show that $\sum_{k=1}^{\infty} \frac{Z_k}{k}$ converges a.s. as $n \to \infty$. Let $Q_n = \sum_{k=1}^{n} \frac{Z_k}{k}$¿ Then,

$$P(\sup_{M \leq n \leq N} |Q_n - Q_M| > \epsilon) \leq \frac{1}{\epsilon^2} Var(Q_N - Q_M) \leq \frac{1}{\epsilon^2} \sum_{k \geq M+1}^{N} \frac{Var(Z_k)}{k^2}.$$

Letting $N \to \infty$ we have

$$P(\sup_{n \geq M} |Q_n - Q_M| > \epsilon) \leq\leq \frac{1}{\epsilon^2} \sum_{k > M} \frac{Var(Z_k)}{k^2} \to 0 \text{ as } M \to \infty.$$

Since

$$P(\sup_{n,\, m \geq M} |Q_n - Q_m| > 2\epsilon) \leq P(\sup_{n \geq M} |Q_n - Q_M| > \epsilon),$$

$\sup_{n,\, m \geq M} |Q_n - Q_m| \to 0$ a.s. and thus $\{Q_n(\omega)\}$ are Cauchy sequence in probability one. By the Kronecker lemma

$$\frac{1}{n} \sum_{k=1}^{n} (Y_k - E[Y_k]) \to 0, \ a.s. \text{ and thus } \frac{T_n}{n} - \frac{1}{n} \sum_{k=1}^{n} E[Y_k] \to 0 \ a.s.$$

Since by the Lebesgue dominated convergence theorem $E[Y_k] \to \mu$ as $k \to \infty$ it follows that $\frac{T_n}{n} \to \mu$. □

1.10 Central Limit Theory

The central limit theorem (CLT) states that, the arithmetic mean of a sufficiently large number of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed.

Theorem (Central Limit) Let $X_1, X_2, \cdots$ be i.i.d. random variables with with $E[X_k] = \mu$ and $Var(X_k) = \sigma^2$. If $S_n = \sum_k X_k$, then

$$\frac{\sqrt{n}(\frac{S_n}{n} - \mu)}{\sigma} \text{ converges in distribution to } N(0, 1),$$

the standard normal distribution.

Proof: Note that the characteristic function φ_n of $\frac{\sqrt{n}(\frac{S_n}{n} - \mu)}{\sigma}$ is given by

$$\varphi_n(t) = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \to e^{-t^2/2}, \quad n \to \infty.$$

But this limit is just the characteristic function of a standard normal distribution $N(0, 1)$, and the central limit theorem follows from the Lévy continuity theorem, which confirms that the convergence of characteristic functions implies convergence in distribution. □

By the law of large numbers, the sample averages $\frac{S_n}{n}$ converge in probability and almost surely to the expected value μ as $n \to \infty$. The central limit theorem describes the distribution of the random fluctuations of the sample average of a sequence of random variables around the maen μ. More precisely, it states that as sample size n gets larger, the distribution of the difference between the sample average $\frac{S_n}{n}$ and its limit μ, when multiplied by the factor $\sqrt{n}$ (that is, $\sqrt{n}(S_n-\mu)$, approximates the normal distribution $N(0,\sigma)$. For large enough n, the distribution of the sample average $\frac{S_n}{n}$ is close to the normal distribution with mean μ and variance $\frac{\sigma^2}{n}$. The usefulness of the theorem is that the distribution of $\sqrt{n}(S_n-\mu)$ approaches the normal distribution regardless of the distribution of the individual random variables X_k.

1.11 Large Deviations

The large deviations theory concerns with the asymptotic exponential rate of the probability measures of tail events. Let $X_1, X_2, \cdots$ be a sequence of i.i.d. random variables let $S_n = \sum_{k=1}^n X_k$. In this section, we will discuss the large deviations in terms of the rate function

$$I(x) = \lim_{n\to\infty} \log P(\frac{S_n}{n} \ge x)$$

for $x > \mu = E[X_k]$. If we let

$$\lambda(\theta) = \log E[e^{\theta X_k}]$$

and define

$$\lambda^*(x) = \sup_{\theta>0}\{\theta x - \lambda(\theta)\},$$

the Legendreenchel transformation of λ, then it will be shown that $I(x) = -\lambda^*(x)$.

For the existence of the limit if $\theta > 0$, then Chebyshev inequality implies

$$e^{\theta n x}\, P(\frac{S_n}{n} \ge x) \le E[e^{\theta\, S_n}] = e^{n\lambda(\theta)}.$$

and thus

$$P(\frac{S_n}{n} \ge x) \le e^{-n(\theta x - \lambda(\theta))}. \tag{1.7}$$

We assume that there exists $\theta > 0$ such that

$$E[e^{\theta X_k}] < \infty.$$

If we let $\pi_n = P(\frac{S_n}{n} \ge x)$, then

$$\pi_{m+n} \ge P(S_m \ge mx,\ S_{n+m} - S_m \ge nx) = \pi_m \pi_n$$

since S_m and $S_{n+m} - S_m$ are independent. Letting $I_n = \log\pi_n$ we have

Lemma If $I_{m+n} \ge I_m + I_n$, then as $n \to \infty$, $\frac{I_n}{n} \to \sup_m \frac{I_m}{m}$.

Proof: Since $\limsup \frac{I_n}{n} \le \sup_m \frac{I_m}{m}$, it suffices to prove that for any $\liminf \frac{I_n}{n} \ge \frac{I_m}{m}$. If we let $n = km + \ell$ with $0 \le \ell < m$ and from the assumption, we have $I_n \ge kI_m + I_\ell$. Dividing this by n, we obtain

$$\frac{I_n}{n} \ge \frac{km}{km+\ell}\frac{I_m}{m} + \frac{I_\ell}{n}$$

Letting $n \to \infty$, the desired result holds. □

From Lemma and (**??**) $\lim_{n\to\infty} \frac{1}{n}\log P(\frac{S_n}{n} \ge x) = I(x)$ exists and

$$P(\frac{S_n}{n} \ge x) \le e^{nI(x)}.$$

Since $\theta > 0$ is arbitrary, it follows from (**??**) $I(x) \le -\lambda^*(x)$ and

$$P(\frac{S_n}{n} \ge x) \le e^{-n\,\lambda^*(x)}.$$

Theorem $I(x) = -\lambda^*(x)$.

Proof: Let F be the distribution of X_k. Define

$$\phi(\theta) = \int_{-\infty}^{\infty} e^{\lambda x} dF(x)$$

and $\theta_+ = \sup\{\theta : \phi(\theta) < \infty\}$, $\theta_- = \inf\{\theta : \phi(\theta) < \infty\}$. Since for $\theta \in (\theta_1, \theta_+)$

$$F_\theta(x) = \frac{1}{\phi(\theta)} \int_{-\infty}^{x} e^{\theta\, y} dF(y)$$

and

$$\int_{-\infty}^{\infty} x dF_\theta(x) = \frac{\phi'(\theta)}{\phi(\theta)}$$

$$\int_{-\infty}^{\infty} x^2 dF_\theta(x) = \phi''(\theta)$$

we have

$$\frac{d}{d\theta}\left(\frac{\phi'}{\phi}\right) = \frac{\phi''}{\phi} - \left(\frac{\phi'}{\phi}\right)^2 = \int_{-\infty}^{\infty} x^2 dF_\theta(x) - (\int_{-\infty}^{\infty} x dF_\theta(x))^2 > 0$$

which is the variance of F_θ, and thus $\lambda = \log\phi(\theta)$ is convex. Thus, there exits $\bar{\theta}$ such that

$$x = \frac{\phi'(\bar{\theta})}{\phi(\bar{\theta})} \quad \text{and} \quad \bar{\theta} = \text{argmax}\{\theta x - \log\phi(\theta)\}.$$

Let $X_1^\theta, X_2^\theta, \cdots$ be i.i.d. random variables with distribution F_θ and let $S_n^\theta = \sum_{k=1}^n X_k^\theta$. By the definition the Radon-Nikodym derivative of the associated measures, it is immediate from the definition $\frac{dF}{dF_\theta} = e^{-\theta x}\phi(\theta)$. Then,

$$\frac{dF^n}{dF_\theta^n} = e^{-\theta x}\phi(\theta)^n. \tag{1.8}$$

In fact, by the induction we have

$$F^n(z) = (F^{n-1} * F)(z) = \int_{-\infty}^{\infty} dF_{n-1}(x) \int_{-\infty}^{z-x} dF(y)$$

$$= \int dF_\theta^{n-1}(x) \int dF_\theta(y) I_{x+y\le z} e^{-\theta(x-y)}\phi(\theta)^n$$

$$= E[I_{\{S_{n-1}^\theta + X_n^\theta \le z\}} e^{-\theta(S_{n-1}^\theta + X_n^\theta)}]\phi(\theta)^n$$

$$= \int_{-\infty}^{z} e^{-\theta x} dF_\theta^n(x)\phi(\theta)^n$$

If $c > x$, then from (**??**) and the monotonicity

$$P(\frac{S_n}{n} \ge x) = \int_{nx}^{nc} e^{-\theta y} dF_n^\theta(y) \ge e^{-nc\theta}(F_\theta^n(nc) - F_\theta^n(nx) \tag{1.9}$$

Since F_θ has mean $\frac{\phi'(\theta)}{\phi(\theta)}$, if we have $x < \frac{\phi'(\theta)}{\phi(\theta)} < c$, then the weak law of large numbers implies

$$F_\theta^n(nc) - F_\theta^n(nx) \to 1 \text{ as } n \to \infty.$$

Thus, we have

$$\liminf n \to \infty \frac{1}{n} \log P(\frac{S_n}{n} \ge x) \ge -\theta c + \log\phi(\theta)$$

Since $c > x$ and $\theta > \bar{\theta}$ are arbitrary, we obtain

$$I(x) = \liminf_{n\to\infty} \frac{1}{n}\log P(\frac{S_n}{n} \geq x) = -x\bar{\theta} + \log\phi(\bar{\theta}) = -\lambda^*(x)$$

In general we say that a large deviation principle holds for a sequence of probability measures P_n defined on the Borel subsets of a complete separable metric space X, with a rate function $H(x)$, if (1) $H(x) \geq 0$ is a lower semicontinuous function on X with the property that $K = \{x : H(x) \leq c\}$ is a compact set for every $c \geq 0$ and
(2) for any closed set $C \subset X$,

$$\limsup_{n\to\infty} \log P_n(C) \leq -\inf_{x\in C} H(x).$$

(3) for any open set $U \subset$

$$\liminf_{n\to\infty} \log P_n(U) \geq -\inf_{x\in U} H(x).$$

Rate functions with property (1) are referred to as good rate functions.

Theorem Let $\{P_n\}$ satisfy LDP on X with rate H and $F(x) : X \to R$ a bounded continuous function. Then

$$\lim_{n\to\infty} \frac{1}{n}\log\int e^{nF(x)}dP_n = \sup_x\{F(x) - H(x)\}.$$

Proof: We note that

$$\lim_{n\to\infty} \frac{1}{n}\log\sum_{k=1}^{n} e^{na_k} = \sup\, a_k.$$

For the upper bound, dividing the range of F into a finite number of intervals of size $\frac{1}{k}$ and denoting by $C_{j,k} = \{x : \frac{j-1}{k} \leq F(x) \leq \frac{j}{k}$. Then,

$$\int e^{nF(x)}dP_n \leq \sum_j \int_{C_{j,k}} e^{nF(x)}dP_n \leq \sum_j e^{\frac{nj}{k}} P(C_{j,k})$$

Thus, we obtain for any k, the bound

$$\lim_{n\to\infty} \frac{1}{n}\log\int e^{nF(x)}dP_n \leq sup_j\{\frac{j}{k} - \inf_{x\in C_{j,k}} H(x)\} \leq \sup_x\{F(x) - H(x)\} + \frac{1}{k}.$$

Letting $k \to \infty$ we obtain the upper bound. For the lower bound if we take any x_0 with $H(x_0) < \infty$, then in a neighborhood U of x_0, $F(x)$ is bounded below by $F(x_0) - \epsilon(U)$. By assumption $P_n(U) \geq exp(nH(x_0) + o(n))$. Since the integrand is nonnegative

$$\int_X e^{nF(x)}dP_n \geq \int_U e^{nF(x)}dP_n \geq exp(n(F(x_0) - H(x_0) - \epsilon(U)) + o(n))$$

Since x_0 is arbitrary and U can be shrunk to x_0, we obtain the lower bound.

Theorem (Puhalskii, O'Brien and Vervaat, de Acosta) Suppose that $\{X_n\}$ is exponentially tight, i.e., for each $a > 0$ there exists a compact set $K_a \subset S$ such that

$$\limsup_{n\to\infty} \frac{1}{n}\log P(X_n \notin K_a) \leq -a$$

Then there exists a subsequence $\{X_{n_k}\}$ for which the large deviation principle holds with a good rate function

2 Markov Chain

In this section consider the discrete time stochastic process. Let S be the state space, e.g., $S = Z = \{\text{integers}\}$, $S = \{0, 1, \cdots, N\}$ and $S = \{-N, \cdots, 0, \cdots, N\}$.

Definition We say that a stochastic process $\{X_n\},\ n \geq 0$ is a Markov chain with initial distribution π ($P(X_0) = \pi_i$) and (one-step) transition matrix P if for each n and $i_k,\ 0 \leq k \leq n-1$

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \cdots, X_0 = i_0) = P(X_{n+1} = j | X_n = i) = p_{ij} \tag{2.1}$$

with

$$\sum_{j \in S} p_{ij} = 1, \quad p_{ij} \geq 0.$$

Thus, the distribution of X_{n+1} depends only on the current state X_n and is independent of the past.

Example Consider the discrete-time dynamics:

$$X_{n+1} = f(X_n, w_n), \quad f : S \times R \to S,$$

where $\{w_n\}$ is independent identically distributed random variables. Then, we have

$$P(f(x, w) = j | x = i) = p_{ij}.$$

We have the following properties.

Theorem Let $P^n = \{p_{ij}^n\}$.

$$P(X_{n+2} = j | X_n = i) = \sum_{k \in S} p_{ik} p_{kj} = (P^2)_{ij} = p_{ij}^{(2)},$$

$$P(X_n = j) = (\pi P^n)_j = \sum \pi_i p_{i,,j}^{(n)}$$

and

$$p_{ij}^{(m+n)} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}^{(m)} \quad \text{(Chapman-Kolmogorov)}.$$

2.1 Classification of the States

In this section we analyze the asymptotic behavior of the Markov chain (**??**), e.g., including

Questions (1) The limits $\pi_j = \lim_{n \to \infty} p_{ij}^{(n)}$ exist and is independent of the initial state i.
(2) The limits $(\pi_1, \pi_2, \cdots)$ form a probability distribution, that is, $\pi \geq 0$ and $\sum \pi_i = 1$.
(3) The chain is ergotic, i.e., $\pi_i > 0$.
(4) There is one and only one stationary probability distribution π such that $\pi = \pi P$ (invariant).

The followings are the basic characterization of Markov Chains.

Definition (1) Communicate: $i \to j$ if $p_{ij}^{(n)} > 0$ for some $n \geq 0$. $i \leftrightarrow j$ (communicate) if $i \to j$ and $j \to i$.
(2) Communicating classes: $i \leftrightarrow j$ defines an equivalent relation, i.e., $i \leftrightarrow i$ (reflective), $i \leftrightarrow j \Leftrightarrow j \leftrightarrow i$ (symmetric) and $i \leftrightarrow j,\ j \leftrightarrow k \Leftrightarrow i \leftrightarrow k$ (transitive). Thus, the equivalent relation $i \leftrightarrow j$ defines equivalent classes of the states, i.e., the communicating classes. A communicating class is closed if the probability of leaving the class is zero, namely that if i is in an equivalent class C but j is not, then j is not accessible from i.
(3) Irreducible: A Markov chain is said to be irreducible if its state space is a single communicating class; in other words, if it is possible to get to any state from any state.

(4) Transient, Null and Positive recurrent: Let the random variable τ_i be the first return time to state i (the "hitting time"):

$$\tau_{ii} = \min\{n \geq 1 : X_n = i | X_0 = i\}.$$

The number of visits N_i to state i is defined by $N_i = \sum_{n=0}^{\infty} I_{\{X_n = i\}}$ and

$$E[N_i] = \sum_{n=0}^{\infty} P(X_n = i | X_0 = i) = \sum_{n=0}^{\infty} p_{ii}^{(n)},$$

where $I\{F\}$ is the indicator function of event F, i.e., $I\{F\}(\omega) = 1$ if $\omega \in F$ and $I\{F\}(\omega) = 0$ if $\omega \notin F$. If $\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$ state i is recurrent (return to the state infinitely may times). If $\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty$ state i is transit (return to the state finitely may times).

Define the probability of the first time return

$$f_{ii}^{(n)} = E[\tau_{ii} = n] = P(X_n = i,\ X_k \neq i | X_0 = i)$$

of state i. Let f_i be the probability of ever returning to state i given that the chain started in state i, i.e.

$$f_i = P(\tau_{ii} < \infty) = \sum_{n=1}^{\infty} f_{ii}^{(n)}.$$

Then, N_i has the geometric distribution, i.e.,

$$P(N_i = n) = f_i^{n-1}(1 - f_i)$$

and

$$E[N_i] = \frac{1}{1 - f_i}.$$

Thus, state i is recurrent if and only if $f_i = 1$ and state i is transit if and only if $f_i < 1$. The mean recurrence time of a recurrent state i is the expected return time μ_i:

$$\mu_i = E[\tau_{ii}] = \sum_{n=1}^{\infty} n f_{ii}^{(n)}.$$

State i is positive recurrent (or non-null persistent) if μ_i is finite; otherwise, state i is null recurrent (or null persistent).

(5) Period: State i has period $d = d(i)$ if (i) $p_{ii}^{(n)} > 0$ for values od $n = dm$, (ii) d is the largest number satisfying (i), equivalently d is the greatest common divisor of the numbers n for which $p_{ii}^{(n)} > 0$. Note that even though a state has period k, it may not be possible to reach the state in k steps. For example, suppose it is possible to return to the state in $\{6, 8, 10, 12, \dots\}$ time steps; k would be 2, even though 2 does not appear in this list. If $k = 1$, then the state is said to be aperiodic: returns to state i can occur at irregular times. Otherwise ($k > 1$), the state is said to be periodic with period k.

(6) Asymptotic: Let a Markov chain is irrecusable and aperiodic. Then, if either state i is transient and null recurrent $p_{ij}^{(n)} \to 0$ as $n\infty$ or if all state i is positive recurrent $p_{ij}^{(n)} \to \frac{1}{\mu_j}$ as $n \to \infty$.

(7) Stationary Distribution: The vector π is called a stationary distribution (or invariant measure) if its entries π_j are non-negative and $\sum_{j \in S} \pi_j = 1$ and if it satisfies

$$\pi = \pi P \Leftrightarrow \pi_j = \sum_{i \in S} \pi_i p_{ij}.$$

An irreducible chain has a stationary distribution if and only if all of its states are positive recurrent. In that case, it is unique and is related to the expected return time:

$$\pi_j = \frac{1}{\mu_j}.$$

Further, if the chain is both irreducible and aperiodic, then for any i and j,

$$\lim_{n\to\infty} p_{ij}^{(n)} = \frac{1}{\mu_j}.$$

Note that there is no assumption on the starting distribution; the chain converges to the stationary distribution regardless of where it begins. Such π is called the equilibrium distribution of the chain. If a chain has more than one closed communicating class, its stationary distributions will not be unique (consider any closed communicating class C_i in the chain; each one will have its own unique stationary distribution π_i. Extending these distributions to the overall chain, setting all values to zero outside the communication class, yields that the set of invariant measures of the original chain is the set of all convex combinations of the π_i's). However, if a state j is aperiodic, then

$$\lim_{n\to\infty} p_{jj}^{(n)} = \frac{1}{\mu_j}$$

and for any other state i, let f_{ij} be the probability that the chain ever visits state j if it starts at i,

$$\lim_{n\to\infty} p_{ij}^{(n)} = \frac{f_{ij}}{\mu_j}.$$

If a state i is periodic with period $d(i) > 1$ then the limit

$$\lim_{n\to\infty} p_{ii}^{(n)}$$

does not exist, although the limit

$$\lim_{n\to\infty} p_{ii}^{(dn+r)}$$

does exist for every integer r.

Theorem 1 Let C be a communicating class. Then either all states in C are transient or all are recurrent.

Proof: The theorem follows from

$$p_{ii}^{(n+r+m)} \geq p_{ij}^{(n)} p_{jj}^{(r)} p_{ji}^{(m)}.$$

Theorem 2 Every recurrent class is closed.
Proof: Let C be a class which is not closed. Then there exists $i \in C$, and $j \notin C$ and m with $P(X_m = j|X_0 = i) > 0$. Since we have

$$P(\{X_m = j\} \cap \{X_n = i \text{ for infinitely many } n\}|X_0 = i) = 0$$

this implies that

$$P(X_n = i \text{ for infinitely many } n|X_0 = i)) < 1,$$

so i is not recurrent, and so neither is C.

Theorem 3 Every finite closed class is recurrent.

Proof: Suppose C is closed and finite and that $\{X_n\}$ starts in C. Then for some $i \in C$ we have

$$0 < P(X_n = i \text{ for infinitely many } n) = P(X_n = i \text{ for some } n) P(X_n = i \text{ for infinitely many } n)$$

by the strong Markov property. This shows that i is not transient, so C is recurrent.

2.2 Stationary distribution

When the limits exist, let j denote the long run proportion of time that the chain spends in state j

$$\pi_j = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} I\{X_m = j | X_0 = i\} \text{ for all initial states } i. \tag{2.2}$$

Taking expected values if π_j exists then it can be computed alternatively by (via the bounded convergence theorem)

$$\pi_j = \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n} P(X_m = j | X_0 = i) \ = \frac{1}{n} \sum_{m=0}^{n} p_{ij}^{(m)} \text{ for all initial states } iquad\text{(Cesaro sense)},$$

or equivalently

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^m = \begin{pmatrix} \pi \\ \pi \\ \vdots \end{pmatrix} = \begin{pmatrix} \pi_0 & \pi_1 & \pi_2 & \cdots \\ \pi_0 & \pi_1 & \pi_2 & \cdots \\ & \vdots & & \end{pmatrix}. \tag{2.3}$$

<u>**Theorem 4**</u> If $\{X_n\}$ is a positive recurrent Markov chain, then a unique stationary distribution π_j exists and is given by $\pi_j = \frac{1}{E[\tau_{jj}]} > 0$ for all states $j \in S$. If the chain is null recurrent or transient then the limits in (**??**) are all 0 and no stationary distribution exits.
Proof: First, we immediately obtain the transient case result since by definition, each fixed state i is then only visited a finite number of times; hence the limit in (**??**) must be 0. Next, j is recurrent. Assume that $X_0 = j$. Let $t_0 = 0$, $t_1 = \tau_{jj}$, $t_2 = \min\{k > t_1 : X_k = j\}$ and in general $t_{n+1} = \min\{k > t_n : X_k = j\}$. These t_n are the consecutive times at which the chain visits state j. If we let $Y_n = t_n - t_{n-1}$ (the interevent times) then we revisit state j for the n-th time at time $t_n = Y_1 + \cdots + Y_n$. The idea here is to break up the evolution of the Markov chain into i.i.d. cycles where a cycle begins every time the chain visits state j. Y_n is the n-th cycle-length. By the Markov property, the chain starts over again and is independent of the past everytime it enters state j (formally this follows by the Strong Markov Property). This means that the cycle lengths Y_n, $n \geq 1$ form an i.i.d. sequence with common distribution the same as the first cycle length τ_{jj} . In particular, $E[Y_n] = E[\tau_{jj}]$ for all $n \geq 1$. Now observe that the number of revisits to state j is precisely n visits at time $t_n = Y_1 + \cdots + Y_n$, and thus the long-run proportion of visits to state j per unit time can be computed as

$$\pi_j = \lim_{m\to\infty} \frac{1}{m} \sum_{k=1}^{m} I\{X_k = j\} = \lim_{n\to\infty} \frac{n}{\sum_{i=1}^{n} Y_i} = \frac{1}{E[\tau_{jj}]}$$

where the last equality follows from the Strong Law of Large Numbers). Thus in the positive recurrent case, $\pi_j > 0$ for all $j \in S$, where as in the null recurrent case, $\pi_j = 0$ for all $j \in S$. Finally, if $X_0 = i \neq j$, then we can first wait until the chain enters state j (which it will eventually, by recurrence), and then proceed with the above proof. Uniqueness follows by the unique representation.

<u>**Theorem 5**</u> Suppose $\{X_n\}$ is an irreducible Markov chain with transition matrix P. Then $\{X_n\}$ is positive recurrent if and only if there exists a (non-negative, summing to 1) solution, π, to the set of linear equations $\pi = \pi P$, in which case π is precisely the unique stationary distribution for the Markov chain.

Proof: Assume the chain is irreducible and positive recurrent. Then we know from Theorem 5 that π exists and is unique. On the one hand, if we multiply (on the right) each side of Equation (5) by P, then we obtain

$$\lim \frac{1}{n} \sum_{m=1}^{n} P^{m+1} = \lim_{n\to\infty} \sum_{m=1}^{n} P^m + \lim_{n\to\infty} \frac{1}{n}(P^{n+1} - P) = \begin{pmatrix} \pi \\ \pi \\ \vdots \end{pmatrix},$$

which implies $\pi = \pi P$.

Conversely, assume the chain is either transient or null recurrent. From Theorem 4, we know that then the limits in (**??**) are identically 0, that is,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^m = 0$$

But if $\pi = \pi P$ then (by multiplying both right sides by P) $\pi = \pi P^2$ and more generally $\pi = \pi P^m,\ m \geq 1$ and so

$$\pi(\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^m) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \pi P^m = 0,$$

which implies $\pi = 0$, contradicting that π is a probability distribution. Having ruled out the transient and null recurrent cases, we conclude that the chain must be positive recurrent. For the uniqueness, suppose $\pi' = \pi' P$. Multiplying both sides of (**??**) (on the left) by π', we conclude that

$$\pi' = \pi' \begin{pmatrix} \pi \\ \pi \\ \vdots \end{pmatrix} = \pi' \begin{pmatrix} \pi_0 & \pi_1 & \pi_2 & \cdots \\ \pi_0 & \pi_1 & \pi_2 & \cdots \\ & \vdots & & \end{pmatrix}.$$

Since $\sum_{j\in S} \pi'_j = 1$, $\pi'_j = \pi_j$ for all $j \in S$.

2.3 Stopping Time

Let $\{\mathcal{F}_n,\ n \geq 0\}$ be an increasing family of σ-algebras and $\{X_n,\ n \geq 0\}$ be a $\{\mathcal{F}_n,\ n \geq 0\}$ adapted stochastic process.

Definition A stopping time with respect to $\{\mathcal{F}_n\}$ is a random variable such that $\{\tau = n\}$ is $\mathcal{F}_n$ measurable for all $b \geq 0$.

If $\mathcal{F}_n$ is the σ-algebra generated by $\{X_0, \cdots, X_n\}$, the event $\{\tau = n\}$ is completely determined by (at most) the total information known up to time n, $\{X_0, \cdots, X_n\}$.

For example the hitting time

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

of state i and

$$\tau_A = \min\{n \geq 0 : X_n \in A\}.$$

of closed set A are stopping times.

Wald's equation: We now consider the very special case of stopping times when $\{X_n. n \geq 1\}$ is an independent and identically distributed (i.i.d.) sequence with common mean $E[X]$. We are interested in the sum up to time: $\sum_{n=1}^{\tau} X_n$.

Theorem (Wald's Equation) If $\tau > 0$ is a stopping time with respect to an i.i.d. sequence $\{X_n,\ n \geq 1\}$ and if $E[\tau] < \infty$ and $E[|X|] < \infty$, then

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

Proof: Since

$$\sum_{n=1}^{\tau} X_n = \sum_{n+1}^{\infty} X_n I\{\tau > n-1\}$$

and X_n and $I\{\tau > n-1\}$ are independent, we have

$$E[\sum_{n=1}^{\tau} X_n] = E[X] \sum_{n=0}^{\infty} P(\{\tau > n\}) = E[X]E[\tau],$$

where the last equality is due to "integrating the tail" method for computing expected values of non-negative random variables.

Null recurrence of the simple symmetric random walk: Let R_n be the simple symmetric random walk: $R_n = \Delta_1 + \cdots + \Delta_n$ with $R_0 = 0$ where $\Delta_n,\ n \geq 1$ is i.i.d. with $P(\Delta = \pm 1) = 0.5$ and $E[\Delta] = 0$. This MC is recurrent but null recurrent. In fact we show that $E[\tau_{11}] = \infty$ By conditioning on the first step $i = 1$,

$$E[\tau_{11}] = (1 + E[\tau_{21}])\frac{1}{2} + (1 + E[\tau_{01}])\frac{1}{2} = 1 + 0.5E[\tau_{21}] + 0.5E[\tau_{01}].$$

Note that by definition, the chain at time $R_\tau = 1$ for $\tau = \tau_{01}$ and

$$1 = R_\tau = \sum_{n=1}^{\tau} \Delta_n$$

But from Wald's equation assuming $E[\tau] < \infty$, then we conclude that

$$1 = E[R_\tau] = E[\Delta]E[\tau] = 0$$

which yields the contradiction $1 = 0$ and thus $E[\tau_{01}] = E[\tau_{11}] = \infty$.

Theorem 6 Suppose $i \neq j$ are both recurrent. If i and j communicate and if j is positive recurrent ($E[\tau_j j] < \infty$), then i is positive recurrent ($E[\tau_{ii}] < \infty$) and also $E[\tau_{ij}] < \infty$. In particular, all states in a recurrent communication class are either all together positive recurrent or all together null recurrent.

Proof: Assume that $E[\tau_{jj}] < \infty$ and that i and j communicate. Choose the smallest $n \geq 1$ such that $p_{ji}^{(n)} > 0$. With $X_0 = j$, let $A = \{X_k \neq j; 1 \leq k \leq n,\ X_n = i\}$ and $P(A) > 0$. Then

$$E[\tau_{jj}] \geq E[\tau_{jj}|A]P(A) = (n + E[\tau_{ij}])P(A),$$

and hence $E[\tau_{ij}] < \infty$ (for otherwise $E[\tau_{jj}] = \infty$, a contradiction). With $X_0 = j$, let $\{Y_m,\ m \geq 1\}$ be i.i.d process as defined in the proof of Theorem 4. Thus the n-th revisit of the chain to state j is at time $t_n = Y_1 + \cdots + Y_n$, and $E[Y] = E[\tau_{jj}] < \infty$. Let

$$p = P(\text{the chain visits state } i \text{ before returning to state } j|X_0 = j),$$

then $p \geq P(A)$, where A is defined above. Every time the chain revisits state j, there is, independent of the past, this probability p that the chain will visit state i before revisiting state j again. Letting N denote the number of revisits the chain makes to state j until first visiting state i, we thus see that N has a geometric distribution with "success" probability p, and so $E[N] < \infty$. N is a stopping time with respect to the process $\{Y_m\}$, and

$$\tau_{ji} \leq \sum_{m=1}^{N} Y_m$$

and so by Wald's equation

$$E[\tau_{ji}] \leq E[N]E[Y] < \infty.$$

Finally, $E[\tau_{ii}] \leq E[\tau_{ij}] + E[\tau_{ji}] < \infty$. □

Strong Markov Chain property: If τ is a stopping time with respect to the Markov chain, then in fact, we get what is called the Strong Markov Property: Given the state X_τ at time τ (the present), the future $X_{\tau+1}, X_{\tau+2}, \cdots$ is independent of the past $X_0, \cdots, X_{tau-1}$. The point is that we can replace a deterministic time n by a stopping time τ and retain the Markov property. It is a stronger statement than the Markov property. This property easily follows since $\{\tau = n\}$ only depends on $X_0, \cdots, X_n$, the past and the present, and not on any of the future. Given the joint event $(\tau = n,\ X_n = i)$, the future $X_{n+1}, X_{n+2}, \cdots$ is still independent of the past:

$$P(X_{n+1} = j|\tau = n,\ X_n = i, \cdots, X_0 = i_0) = P(X_{n+1} = j|X_n = i, \cdots, X_0 = i_0) = p_{ij}.$$

2.4 Hitting Times and Absorption Probabilities

Let $\{X_n,\ n \geq 0\}$ be a Markov chain with transition matrix P. The hitting time of a subset A of S is the random variable H^A defined by

$$H^A = \inf\{n : X_n \in A\}$$

The probability starting from i that the chain ever hits A is then

$$h_i^A = P(H^A < \infty | X_0 = i)$$

When A is closed, h_i^A is called the absorption probability. The mean time taken for the chain to reach A; if $P(H^A < \infty | X_0 = i) = 1$, is given by

$$k_i^A = E[H^A | X_0 = i] = \sum_{n=0}^{\infty} n P(H^A = n | X_0 = i).$$

The vector of hitting probabilities $h_i^A = (h_i^A,\ i \in S)$ satisfies the linear system $h = Ph$;

$$h_i^A = 1 \text{ for } i \in A$$

$$h_i^A = \textstyle\sum_{j \in S} p_{ij} h_j^A \text{ for } i \notin A.$$

In fact, if $X_0 = i$ then $H^A = 0$ so $h_i^A = 0$. If $X_0 = i,\ i \notin\in A$, then $H^A \geq 1$, so by the Markov property

$$P(H^A < \infty | X_1 = j,\ X_0 = i) = P(H^A < \infty | X_0 = j) = h_j^A$$

and

$$h_i^A = P(H^A < \infty | X_0 = i) = \textstyle\sum_{j \in S} P(H^A < \infty,\ X_1 = j | X_0 = i)$$

$$= \textstyle\sum_{j \in S} P(H^A < \infty | X_1 = j) P(X_1 = j | X_0 = i) = \sum_{j \in S} p_{ij} h_j^A.$$

Similarly, the probability f_{ij} that the chain ever visits state j satisfies

$$f = Pf.$$

The vector of mean hitting times $k^A = (k_i^A,\ i \in S)$ satisfies the following system of linear equations, $k = 1 + Pk$;

$$k_i^A = 0 \text{ for } i \in A$$

$$k_i^A = 1 + \textstyle\sum_{j \notin A} p_{ij} k_j^A \text{ for } i \notin A$$

In fact, if $X_0 = i \in A$, then $H^A = 0$ so $k_i^A = 0$. If $X_0 = i \notin A$, then $H^A \geq 1$, so by the Markov property

$$E[H^A | X_1 = j,\ X_0 = i] = 1 + E[H^A | X_0 = j]$$

and

$$k_i^A = E[H^A | X_0 = i] = \textstyle\sum_{j \in S} E[H^A I\{X_1 = j\} | X_0 = i]$$

$$= \textstyle\sum_{j \in S} E[H^A | X_1 = j,\ X_0 = i] P(X_1 = j | X_0 = i) = 1 + \sum_{j \notin A} p_{ij} k_j^A.$$

Remark: The systems of these equations may have more than one solution. In this case, the vector of hitting probabilities h^A and the vector of mean hitting times k^A are the minimal non-negative solutions of these systems.

2.5 Examples

In this section we discusses examples of the Markov chains. First, consider the random walk, i.e, the transition probability P satisfies

$$p_{i,i-1} = q, \quad p_{i,i+1} = p, \quad p,\ q > 0 \text{ and } p + q = 1.$$

Example 1 (Simple Random Walk) The chain is irreducible and the period $d = 2$ with $p_{ii}^{(2n+1)} = 0$ and

$$p_{ii}^{(2n)} = \frac{(2n)!}{n!n!} p^n q^n \sim \frac{(4pq)^n}{\sqrt{2\pi n}},$$

by Stirlings formula. Thus, if $p = q$, then

$$\sum p_{ii}^{(n)} \sim \sum \frac{1}{\sqrt{2\pi n}} = \infty$$

and the chain is recurrent. If $p \neq q$, then $r = 4pq < 1$ and

$$\sum p_{ii}^{(n)} \sim \sum \frac{r^n}{\sqrt{2\pi n}} < \infty$$

and thus the chain is transient. If π is a stationary distribution, then

$$\pi_i = q\, p_{i-1} + p\, \pi_{i+1}$$

$$p\,(\pi_{i+1} - p_i) = q\,(\pi_i - \pi_{i-1})$$

Thus, for bounded solutions we must have $\pi_i = \pi_{i-1}$ and $\pi_0 = 0$. Hence $p = q$ the chain null recurrent.

Example 2 (Absorbing end $i = 0$) $S = \{0, 1, \cdots\}$ with the aborning state $i = 0$, i.e., $p_{00} = 1$. The chain two subclasses $C_0 = \{0\}$ and $C_1 = \{1, 2, \cdots\}$. C_0 is positive recurrent and C_1 is transient.$\pi = (1, 0, 0, \cdots)$ is a stationary distribution. The absorbing probability $\alpha_i = f_{i0}$ satisfies

$$\alpha_i = p\alpha_{i+1} + q\alpha_{i-1}$$

and

$$p(\alpha_{i+1} - \alpha_i) = q(\alpha_i - \alpha_{i-1})$$

Thus,

$$\alpha_i = A + B(\tfrac{q}{p})^i$$

For $\frac{q}{p} \geq 1$ since α is bounded, $B = 0$ and $\alpha_i = A = 1$. For $\frac{q}{p} < 1$, $\alpha_i = (\frac{q}{p})^i$ since $\alpha_0 = 1$ and $\alpha_\infty = 0$.

Example 3 (Absorbing ends $i = 0,\ N$) Let $S = \{0, 1, 2, \cdots, N\}$ and $p_{00} = 1$ and $p_{NN} = 1$. There are three subclasses $C_0 = \{0\}$, $C_1 = \{1, \cdots, N-1\}$ and $C_2 = \{N\}$. C_0, C_2 are positive recurrent and C_1 is transient.$\pi = (\alpha, 0, 0, \cdots, \beta)$ with $\alpha,\ \beta \geq 0$ and $\alpha + \beta = 1$ are statinary distributions. The absorbing probability $\alpha_i = f_{i0}$ satisfies

$$\alpha_i = p\alpha_{i+1} + q\alpha_{i-1}$$

Using the same arguments in Example 2,

$$\alpha_i = \begin{cases} \frac{(\frac{q}{p})^i - (\frac{q}{p})^N}{1 - (\frac{q}{p})^N}, & p \neq q \\ \\ 1 - \frac{i}{N} & p = q. \end{cases}$$

Example 4 (Reflecting rend $i = 0$) Let $S = \{0, 1, 2, \cdots\}$ and $p_{0,1} = 1$. The chain is irreducible with period $d = 2$. For $\frac{q}{p} < 1$, $f_{i1} = \alpha_i = (\frac{q}{p})^{i-1}$, $i > 1$ from Example 2. But, if the chain is recurrent, then $f_{i1} = 1$ for all $i > 1$. Thus, the chain ia transient $p_{ij}^n \to 0$ as $n \to \infty$.

Now, for $\frac{q}{p} \geq 1$ we have $f_{i1} = 1$ for $i > 1$ and $f_{11} = q + pf_{21} = 1$ and hence the chain is recurrent. If π is a stationary distribution,

$$\pi_0 = \pi_1 q$$

$$\pi_1 = \pi_0 + \pi_2 q$$

$$\pi_i = \pi_{i-1}p + \pi_{i+1}q, \quad i \geq 2$$

From the first two equations, $p\pi_1 = q\pi_2$. From the last equations, By induction in i we have $p\pi_i = q\pi_{i+1}$. If $p = q$, $\pi_i = \pi_0$ and consequently $\pi_0 = 0$ for all $i \geq 0$, which implies the chain is null recurrent.

Next, for $\frac{q}{p} > 1$ it follows from $\sum \pi_i = 1$

$$1 = \pi_1(q + \sum_{k=0}^{\infty} (\frac{p}{q})^k) = \pi_1(q + \frac{q}{q-p}).$$

Thus, $\pi_1 = \frac{q-p}{2q^2}$ and

$$\pi_0 = \frac{q-p}{2q}, \quad \pi_i = \pi_1 (\frac{p}{q})^{i-1} \text{ for } i \geq 1.$$

Therefore, for $\frac{q}{p} > 1$ the chain is positive recurrent.

Example 5 (Reflecting ends $i = 0,\ N$) Let $S = \{0, 1, \cdots, N\}$ and $p_{01} = 1$ and $p_{N,N-1} = 1$. The chain irreducible with period $d = 2$. As we did in Example 4, we have the stationary distribution

$$\pi_i = (\frac{p}{q})^{i-1} \sum_{k=0}^{N-2} (\frac{p}{q})^k, \quad 1 \leq i \leq N-1$$

and $\pi_0 = q\pi_1$ and $\pi_N = p\pi_{N-1}$ and thus the chain is positive recurrent.

Example 6 (Birth-and-death chain) Consider the Markov chain with state space $S = \{0, 1, 2, \cdots\}$ and transition probabilities $p_{00} = 1$ and $p_{i,i-1} = q_i,\ p_{i,i+1} = p_i$ for $i \geq 1$. As in Example 2, $C_0 = \{i = 0\}$ is positive recurrent and $C_1 = \{1, 2, \cdots\}$ is transient. We wish to calculate the absorption probability $\alpha_i = f_{i0}$. Such a chain may serve as a model for the size of a population, recorded each time it changes, p_i being the probability that we get a birth before a death in a population of size i.

$$\alpha_i = p_i \alpha_{i+1} + q_i \alpha_{i-1}$$

and

$$p_i(\alpha_{i+1} - \alpha_i) = q_i(\alpha_i - \alpha_{i-1})$$

Thus,

$$\alpha_{i+1} = 1 - \sum_{k=0}^{i} \Pi_{j=1}^{k} \frac{q_j}{p_j} (1 - \alpha_1)$$

There are two different cases:
(i) If $A = \sum_{k=0}^{\infty} \Pi_{j=1}^{k} \frac{q_j}{p_j} = \infty$, then $\alpha_1 = 1$ and $\alpha_i = 1$ for all $i \geq 0$.
(ii) If $A = \sum_{k=0}^{\infty} \Pi_{j=1}^{k} \frac{q_j}{p_j} < \infty$, then $1 - \alpha_1 = \frac{1}{A}$ and

$$1 - \alpha_{i+1} = \frac{\sum_{k=0}^{i} \Pi_{j=1}^{k} \frac{q_j}{p_j}}{\sum_{k=0}^{\infty} \Pi_{j=1}^{k} \frac{q_j}{p_j}},$$

so the population survives with positive probability.

2.6 Exercise

Problem 1 Show that the relation $\leftrightarrow$ is transitive
Problem 2 Show that for every Markov chain with countably many state,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} p_{ij}^{(m)} = \frac{f_{ij}}{\mu_j}.$$

(Hint: $p_{ij}^{(m)} = \sum_{k=1}^{m} f_{ij}^{(m-k)} p(k)_{jj}$).
Problem 3 Consider an irreducible chain with $\{0, 1, \cdots\}$. A necessary and sufficient condition for the chain to be transient is the system $u = Pu$ ($u_i = \sum_{j\in S} p_{ij} u_j$) has a bounded solution such that u_i is not a constant solution.
Problem 4 Complete the Example 5.
Problem 5 Consider a Markov chain with $S = \{0, 1, \cdots\}$ and transition probabilities:

$$p_{ij} = \begin{cases} p_i > 0, & j = i+1 \\ r_i \geq 0, & j = i \\ q_i > 0, & j = i-1 \\ 0 & \text{otherwise} \end{cases}$$

Let $\gamma_n = \Pi_{k=1}^{n} \frac{q_k}{p_k},\ n \geq 1$.
(1) Show that the chain is transient if and only if $\sum \gamma_n < \infty$ and the chain is recurrent if and only if $\sum \gamma_n = \infty$.
(2) Show that the chain is positive recurrent if and only if $\sum \frac{1}{\gamma_n p_n} < \infty$ and the chain is null recurrent if and only if $\sum \frac{1}{\gamma_n p_n} = \infty$.
Problem 6 Classify the states of a Markov chain

$$P = \begin{pmatrix} p & q & 0 & 0 \\ 0 & 0 & p & q \\ p & q & 0 & 0 \\ 0 & 0 & p & q \end{pmatrix}$$

where $p + q = 1$ and $p \geq 0,\ q \geq 0$.

3 Continuous time Markov Chain

A continuous-time Markov process (CTMC) is a stochastic process $\{X_t,\ t \geq 0\}$ that satisfies the Markov property and takes values from a set S called the state space; it is the continuous-time version of a Markov chain. For $s > t$

$$P(X_s = j|\sigma(X_t)) = P(X_s = j|\mathcal{F}_t),$$

where $\{\mathcal{F}_t,\ t \geq 0\}$ is an increasing family of σ-algebras, X_t is $\mathcal{F}_t$ measurable and $\sigma(X_t)$ is the σ-algebra generated by the random variable X_t. In effect, the state of the process at time s is conditionally independent of the history of the process before time t, given the state of the process at time t. The process is characterized by "transition rates" q_{ij} between states, i.e., q_{ij} (for $i \neq j$) measures how quickly that $i \to j$ transition happens. Precisely, after a tiny amount of time h, the probability the state is now at j is given by

$$P(X_{t+h} = j|X_t = i) = q_{ij}h + o(h), \quad i \neq j,$$

where $o(h)$ implies that $\frac{o(h)}{h} \to 0$ as $h \to 0^+$. Hence, over a sufficiently small interval of time, the probability of a particular transition (between different states) is roughly proportional to the duration of that interval. The q_{ij} are called transition rates because if we have a large ensemble of n systems in state i, they will switch over to state j at an average rate of nq_{ij} until n decreases appreciably.

The transition rates q_{ij} are given as the ij-th elements of the transition rate matrix Q. As the transition rate matrix contains rates, the rate of departing from one state to arrive at another should be positive, and the rate that the system remains in a state should be negative. The rates for a given state should sum to zero, yielding the diagonal elements to be

$$q_{ii} = -\sum_{j \neq i} q_{ij}.$$

With this notation, if let

$$P_{ij}(h) = P(X_h = j | X_0 = i)$$

be the transition probability, then

$$\lim_{h \to 0^+} \frac{P(h) - I}{h} = Q.$$

The transition probability satisfies the semigroup property

$$P(t+s) = P(t)P(s) \text{ for } t,\ s \geq 0 \text{ with } P(0) = I$$

Thus,

$$P(t+h) = P(t) = (P(h) - I)P(t), \quad P(t-h) - P(t) = (I - P(h))P(t-h)$$

for $t > 0,\ h > 0$ and hence

$$P'(t) = \lim_{\tau \to 0} \frac{P(t+\tau) - P(t)}{\tau} = QP(t).$$

Since

$$\lim_{t \to 0^+} \frac{e^{Qt} - I}{t} = Q,$$

where e^{Qt} is the matrix exponential defined by

$$e^{Qt} = \sum_{k=0}^{\infty} \frac{t^k}{k!} Q^k,$$

we obtain

$$P(t) = e^{Qt},$$

i.e., Q is the generator of $P(t)$. Thus, letting $p_j(t) = P(X_t = j)$, the evolution of a continuous-time Markov process is given by the first-order differential equation

$$\frac{d}{dt} p(t) = p(t)Q, \quad p(0) = \pi = \text{initial distribution}.$$

The probability that no transition happens in some time $r > 0$ is

$$P(X_s = i, \text{for all} s \in (t, t+r) \,|\, X_t = i) = e^{-q_i r}.$$

That is, the probability distribution of the waiting time until the first transition is an exponential distribution with rate parameter $q_i = -q_{ii}$, and continuous-time Markov processes are thus memoryless processes. Letting τ_n denote the time at which the n-th change of state (transition) occurs, we see that $Y_n = X_{\tau_n^+}$, the state right after the n-th transition, defines the underlying discrete-time Markov chain, called the embedded Markov chain. Y_n keeps track, consecutively, of the states visited right after each transition, and moves from state to state according to the one-step transition probabilities $\pi_{ij} = P(Y_{n+1} = j | Y_n = i)$. This transition matrix $\{\pi_{ij}\}$, together with the waiting-time rates q_i, completely determines the continuous time Markov chain, i.e.,

$$q_{ij} = q_i \pi_{ij} \quad \text{for all } j \neq i.$$

Hence,

$$Q = \Lambda(\Pi - I), \quad \Lambda = \text{diag}(q_0, q_1, \cdots).$$

Example (Poisson counting process) Let $N_t,\ t \geq$ be the counting process for a Poisson process at rate λ. Then N_t forms a continoous time Markov chain with $S = \{0, 1, 2, \cdots\}$ and $q_{i,j} = \lambda$ for $j = i+1$, otherwise 0, i.e. $\pi_{i,i+1} = 1$. This process is characterized by a rate parameter λ, also known as intensity, such that the number of events in time interval $(t, t+\tau]$ follows a Poisson distribution with associated parameter $\lambda\tau$, i.e.,

$$P(N_{t+\tau} - N_t = k) = \frac{e^{-\lambda\tau}(\lambda\tau)^k}{k!} \qquad k = 0, 1, \ldots,$$

where k is the number of jumps during $(t, t+\tau]$. That is,

$$p_k(\tau) = \frac{e^{-\lambda\tau}(\lambda\tau)^k}{k!}$$

satisfies

$$\frac{d}{dt} p_k(t) = -\lambda p_k(t) + \lambda p_{k-1}(t)$$

and thus $\frac{d}{dt} p(t) = Qp(t)$. The increment $N_{t+h} - N_t$ is independent of $\mathcal{F}_t$ and the gaps $\tau_1, \tau_2, \cdots$ between successive jumps are independent and identically distributed with exponential distribution;

$$P(\tau_i \geq t) = P(N(t) = 0) = e^{-\lambda t}, \quad t \geq 0.$$

Thus, a concrete construction of a Poisson process can be done as follows. Consider a sequence $\{\tau_n,\ n \geq 1\}$ be i.i.d. random variables with exponential law of parameter λ. Set $T_0 = 0$ and for $n \geq 1$, $T_n = \tau_1 + \cdots + \tau_n$. Note that $\lim_{n\to\infty} T_n = 1$ almost surely, because by the strong law of large numbers

$$\lim_{n\to\infty} \frac{T_n}{n} = E[\tau] = \frac{1}{\lambda}$$

$N_t,\ t \geq 0$ be the arrival process associated with the interarrival times T_n. That is

$$N_t = \sum_{n=0}^{\infty} nI\{T_n \leq t \leq T_{n+1}\}. \tag{3.1}$$

The characteristic function of N_t is given by

$$E[e^{iN_t\xi}] = \sum_{n=0}^{\infty} e^{in\xi} e^{-\lambda t} \frac{(\lambda t)^n}{n!} = e^{\lambda t(e^{i\xi}-1)}.$$

Thus,

$$E[N_t] = \lambda t.$$

and λ is the expected number of arrivals in an interval of unit length, or in other words, is the arrival rate. On the other hand, the expect time until a new arrival is $\frac{1}{\lambda}$.

$$Var(N_t) = \lambda t$$

and thus

$$E[|N_t - N_s|^2] = \lambda\,|t-s| + (\lambda\,|t-s|)^2$$

The Poisson process is continuous in mean of order 2 but the sample paths of the Poisson process are discontinuous with jumps of size 1.

Example (Sum of Poisson processes) Let $\{L_t,\ t \geq 0\}$ and $\{M_t,\ t \geq 0\}$ be two independent Poisson processes with respective rates λ and μ. The process $N_t = L_t + M_t$ is a Poisson process of rate $\lambda + \mu$.

Proof: Clearly, the process $N_t$has independent increments and $N_0 = 0$. Then, it suces to show that for each $0 < s < t$, the random variable $N_t - N_s$ has a Poisson distribution of parameter $(\lambda + \mu)(t - s)$.

$$P(N_t - N_s = n) = \sum_{k=0}^{n} P(L_t - L_s = k,\ M_t - M_s = n - k)$$

$$= \sum_{k=0}^{n} e^{-\lambda(t-s)} \frac{(\lambda(t-s))^k}{k!} e^{-\mu(t-s)} \frac{(\mu(t-s))^{n-k}}{(n-k)!} = e^{-(\lambda+\mu)(t-s)} \frac{((\lambda+\mu)(t-s))^n}{n!}.$$

Example (Compounded Poisson process) Let $\{X_n,\ n \geq 0\}$ be a Markov chain with transition probability Π and define the continuous Markov chain X_t by

$$X_t = X_{N_t}$$

Then,

$$p_{i,j}(t) = P(X_t = j | X_0 = i) = \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} \pi_{i,j}^{(k)}$$

or equivalently

$$P(t) = \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} \Pi^k = e^{\lambda(\Pi - I)t} = e^{Qt}$$

where $Q = \lambda\,(\Pi - I)$ is the generator of X_t.

In general, the construction of a continuous-time Markov chain with generator Q and initial distribution π is as follows. Consider a discrete-time Markov chain $X_n,\ n \geq 0$ with initial distribution π and transition matrix Π. The stochastic process $\{X_t,\ t \geq 0\}$ will visit successively the sates $Y_0, Y_1, Y_2, \cdots$ starting from $X_0 = Y_0$. Denote by $H_{Y_0}, \cdots, H_{Y_{n-1}}$ the holding times in the state Y_k. We assume the holding times $H_{Y_0}, \cdots, H_{Y_{n-1}}$ are independent exponential random variables of parameters $q_{Y_0}, \cdots, q_{Y_{n-1}}$, i.e., for $j \in S$

$$P(H_j \geq t) = e^{-q_j t}, \quad t \geq 0.$$

Let $T_n = H_{Y_0} + \cdots + H_{Y_{n-1}}$ and

$$X_t = Y_n, \quad \text{for } T_n \leq t < T_{n+1}$$

The random time

$$\zeta = \sum_{n=0}^{\infty} H_{Y_n}$$

is called the explosion time. We say that the Markov chain X_t is not explosive if $P(\zeta = \infty) = 1$.

Let $\{X_t,\ t \geq 0\}$ be an irreducible continuous-time Markov chain with generator Q. The following statements are equivalent:

(i) The jump chain Π is positive recurrent.

(ii) Q is not explosive and has an invariant distribution π.

Moreover, under these assumptions, we have

$$\lim p_{ij}(t) = \frac{1}{q_j \mu_j} \text{ as } t \to \infty,$$

where $\mu_j = E[\tau | X_0 = j] = E[\tau_{jj}]$ is the expected return time to the state j.

3.1 Explosion

When a state space S is infinite, it can happen that the process, through successive jumps, moves to state that have the shorter waiting time , i.e. have larger jump rates q_i. The waiting time at state i has the expected value $E[\tau_i] = \frac{1}{q_i}$.

Example (Birth process) A birth process $\{X_t,\ t \geq 0\}$ as generalization of the Poisson process in which the parameter λ is allowed to depend on the current state of the process. The data for a birth process consist of birth rates $q_i > 0$, where $i \geq 0$. Then, a birth process $\{X_t,\ t \geq 0\}$ is a continuous time Markov chain with state-space $S = \{0, 1, 2, \cdots\}$ and generator Q:

$$q_{i,i} = -q_i, \quad q_{i,j} = q_i \text{ for } j = 1, \quad q_{ij} = 0, \text{ otherwise.}$$

That is, conditional on $X_0 = i$, the holding times $H_i, H_{i+1}, \cdots$ are independent exponential random variables of parameters $q_i, q_{i+1}, \cdots$, respectively, and the jump chain is given by $Y_n = i + n$. Concerning the explosion time, two cases are possible:
(i) If $\sum_{j=0}^{\infty} \frac{1}{q_j} < \infty$, $\zeta < \infty$ a.s.
(ii) If $\sum_{j=0}^{\infty} \frac{1}{q_j} = \infty$, $\zeta = \infty$ a.s.
In fact, if $\sum_{j=0}^{\infty} \frac{1}{q_j} < \infty$, by the monotone convergence theory

$$E[\zeta|X_0 = i] = E[\sum_{n=0}^{\infty} \tau_n | X_0 = i] = \sum_{j=0}^{\infty} \frac{1}{q_{j+i}} < \infty,$$

$\zeta < \infty$ a.s.. If $\sum_{j=0}^{\infty} \frac{1}{q_{i+j}} = \infty$, then $\Pi_{j=0}^{\infty}(1 + \frac{1}{q_{i+j}}) = \infty$ and since τ_j is independent,

$$E[e^{-\sum_{n=0}^{\infty} \tau_n}] = \Pi_{n=0}^{\infty} E[e^{-\tau_n}] = \Pi_{j=1}^{\infty} \left(1 + \frac{1}{g_{i+j}}\right)^{-1} = 0,$$

so $\sum_{n=0}^{\infty} \tau_n = \infty$ a.s..

Particular case (Simple birth process): Consider a population in which each individual gives birth after an exponential time of parameter λ, all independently. If i individuals are present then the first birth will occur after an exponential time of parameter $i\lambda$. Then we have $i + 1$ individuals and, by the memoryless property, the process begins afresh. Then the size of the population performs a birth process with rates $q_i = i\lambda$, $i \geq 1$. Suppose $X_0 = 1$. Note that $\sum_{i=1}^{\infty} \frac{1}{i\lambda} = \infty$, so $\zeta = \infty$ a.s. and there is no explosion in finite time. However, the mean population size growths exponentially: $E[X_t] = e^{\lambda t}$: Indeed, let τ be the time of the first birth. Then if we let $\mu(t) = E[X_t]$, then

$$\mu(t) = E[X_t I\{\tau \leq t\}[+E[X_t I\{\tau > t\}] = \int_0^t 2\lambda e^{-\lambda s} \mu(t-s)\, ds + e^{-\lambda t}$$

By letting $r = t - s$ we have

$$e^{\lambda t} \mu(t) = 1 + 2\lambda \int_0^t e^{\lambda r} \mu(r)\, dr$$

and thus $\mu(t) = e^{\lambda t}$.

For the birth process with $q_i = (i+1)^2$ is explosive since

$$\sum_i \frac{1}{(i+1)^2} < \infty.$$

With bounded q_i the birth process is not explosive. If $q_i > 0$ is not bounded, the Q is no longer bounded.

Theorem (Explosive) The Markov chain corresponding to the transition rate matrix Q starting from i explodes in finite time if and only if there exists a nonnegative bounded sequence with $U_i > 0$ that satisfies

$$\sum q_{ij} U_j \geq \sigma U_i \text{ for all } i,$$

for some $\sigma > 0$.

Theorem (Non Explosive) If for some $\sigma > 0$, there exists a nonnegative U on S that satisfies

$$\sum q_{ij} U_j \leq \sigma U_i \text{ for all } i,$$

and $U_i \to \infty$ as $q_i \to \infty$, then the chain is not explosive.

3.2 Invariant distribution

A probability distribution (or, more generally, a measure) π on the state space S is said to be invariant for a continuous-time Markov chain $\{X_t,\ t \geq 0\}$ If $\pi P(t) = \pi$ for all $t \geq 0$. If we choose an invariant distribution π as initial distribution of the Markov chain $\{X_t,\ t \geq 0\}$, then the distribution of is π for all $t \geq 0$. If $\{X_t,\ t \geq 0\}$ is a continuous-time Markov chain irreducible and recurrent (that is, the associated jump matrix Π is recurrent) with generator Q, then, a measure π is invariant if and only if

$$\pi Q = 0,$$

and there is a unique (up to multiplication by constants) solution π which is strictly positive. On the other hand, if we set $\alpha_j = q_i \pi_j$, then it is equivalent to say that α is invariant for the jump matrix Π. In fact, we have $\alpha(\Pi - I)$ if and only if $\pi Q = 0$.

That is, to find the stationary probability distribution vector, we must next find α such that

$$\alpha(I - \Pi) = 0,$$

with α being a row vector, such that all elements in α are greater than 0 and $\sum_{j \in S} \alpha_j = 1$. From this, π may be found as

$$\pi_j = \frac{\alpha_j}{q_j}$$

and normalize π so that $\sum \pi_j = 1$.

A CTMC is called positive recurrent if it is irreducible and all states are positive recurrent. We define the limiting probabilities for the CTMC as the long-run proportion of time the chain spends in each state $j \in S$:

$$P_j = \lim_{t \to \infty} \frac{1}{t} \int_0^t I\{X_s = j | X_0 = i)\, ds, \quad w.p.1.,$$

which after taking expected values yields

$$P_j = \lim_{t \to \infty} \frac{1}{t} \int_0^t P_{ij}(s)\, ds.$$

When each P_j exists and $\sum P_j = 1$, then $P = (P_j,\ j \in S)$ (as a row vector) is called the limiting (or stationary) distribution for the Markov chain.

Proposition 1 If X_t is a positive recurrent CTMC, then the limiting probability distribution P exists, is unique, and is given by

$$P_j = \frac{E[H_j]}{E[\tau_{jj}]} = \frac{1}{q_j E[\tau_{jj}]}.$$

In words: The long-run proportion of time the chain spends in state j equals the expected amount of time spent in state j during a cycle divided by the expected cycle length (between visits to state j)". Moreover, the stronger mode of convergence (weak convergence) holds: $P_j = lim_{t\to\infty} P_{ij}(t)$. Finally, if the chain is either null recurrent or transient, then $P_j = 0,\ j \in S$, no limiting distribution exists.

Example (Birth-Death process) A birth-death chain is a continuous time Markov chain with state space $S = \{0, 1, 2, \cdots\}$ (representing population size) and transition rates:

$$q_{i,i+1} = \lambda_i, \quad q_{i,i-1} = \mu_i, \quad q_{i,i} = -\lambda_i - \mu_i$$

with $\mu_0 = 0$. Thus,

$$\pi_{i,i+1} = p_i, \quad \pi_{i,i-1} = 1 - p_i \quad \text{with } p_i = \frac{\lambda_i}{\lambda_i + \mu_i}.$$

The matrix Π is irreducible. Notice that

$$\frac{\sum \pi_{ii}^{(n)}}{\lambda_i + \mu_i}$$

is the expected time spent in state i. A necessary and sufficient condition for non explosion is that

$$\sum_{i=0}^{\infty} \frac{\sum \pi_{ii}^{(n)}}{\lambda_i + \mu_i} = \infty.$$

On the other hand, equation $\pi Q = 0$ satisfied by invariant measures leads to the system

$$\mu_1 \pi_1 = \lambda_0 \pi_0$$

$$\lambda_0 \pi_0 + \mu_2 \pi_2 = (\lambda_1 + \mu_1)\pi_1$$

$$\lambda_{i-1}\pi_{i-1} + \mu_{i+1}\pi_{i+1} = (\lambda_i + \mu_i)\pi_i, \quad i \geq 2.$$

So, π_i is an equilibrium if and only if

$$\lambda_i \pi_i = \mu_{i+1}\pi_{i+1}$$

and

$$\pi_i = \frac{\Pi_{k=0}^{i-1}\lambda_k}{\Pi_{j=1}^{i}\mu_k}\pi_0$$

Hence, an invariant distribution exists if and only if

$$c = \sum \frac{\Pi_{k=0}^{i-1}\lambda_k}{\Pi_{j=1}^{i}\mu_k} < \infty$$

and the invariant distribution is

$$\pi_0 = \frac{1}{1+c}, \quad \pi_i = \frac{\Pi_{k=0}^{i-1}\lambda_k}{\Pi_{j=1}^{i}\mu_k}\pi_0.$$

3.3 Dynkin's formula

Let τ_A is the exit time from A;

$$\tau_A = \inf\{t \geq 0 : X_t \notin A\}.$$

Theorem For $\lambda > 0$ the function

$$U_i = E[e^{-\lambda\,\tau_A} f(x_{\tau_A}) | X_0 = i] \tag{3.2}$$

is the unique solution to

$$(QU)_j = \lambda U_j, \ \ j \in A, \quad U_i = f(i), \ \ i \notin A. \tag{3.3}$$

Proof: First, note that if $i \notin A$, then $\tau_A = 0$ and $U_i = f_i$. Since $\frac{d}{dt}(e^{-\lambda t}P(t)) = (Q - \lambda I)e^{-\lambda t}P(t)$,

$$e^{\lambda t}P(t) = I + \int_0^t e^{-\lambda s}P(s)(Q - \lambda I)\,ds$$

Thus

$$M_t = e^{-\lambda t}f(X_t) - f(i) - \int_0^t e^{-\lambda s}(Q - \lambda I)f(X_s)\,ds \quad \text{is a martingale} \tag{3.4}$$

with respect $(\Omega, \mathcal{F}_t, P)$. In fact, $t \geq s$

$$E^i(M_t - M_s|\mathcal{F}_s) = e^{-\lambda s}E^i(e^{-\lambda(t-s)}f(X_t) - f(X_s) - \int_s^t e^{-\lambda(\sigma-s)}(Q - \lambda I)f(X_\sigma)\,d\sigma|\mathcal{F}_s)$$

$$= e^{-\lambda s}e^{-\lambda(t-s)}P(t-s)f(X_s) - f(X_s) - \int_s^t e^{-\lambda(\sigma-s)}P(\sigma - s)(Q - \lambda I)f(X_s)\,d\sigma = 0,$$

where we used

$$E^i(f(X_t)|\mathcal{F}_s) = P(t-s)f(X_s).$$

Thus, by the Doob's optional sampling theorem $E[M_\tau] = 0$ for a stooping time $\tau \geq 0$ and we have

$$E[e^{-\lambda\tau}\phi(X_\tau)|X_0 = i] = \phi(i) + E[\int_0^\tau e^{-\lambda s}(Q - \lambda I)\phi(X_s)\,ds|X_0 = i]. \tag{3.5}$$

Suppose U satisfies (**??**), letting $\phi = U$ and $\tau = \tau_A$,

$$E[e^{-\tau_A}U(X_{\tau_A})|X_0 = i] - U_i = 0,$$

which implies (**??**) holds.

Remark (1) Equation

$$\lambda U_j - (QU)_j = g_j, \ \ j \in A, \quad U_i = f(i), \ \ i \notin A. \tag{3.6}$$

has the unique solution of the form

$$U_i = E[e^{-\tau_A}f(x_{\tau_A}] + \int_0^{\tau_A} E[e^{-\lambda s}\,g(X_s)\,ds|X_0 = i]$$

(2) If $\lambda = 0$ it is required that $P(\tau_A < \infty) = 1$.
(3) If U satisfies $(QU)_j = 1, \ j \in A$ and $U_i = 0$ for $i \notin A$, then

$$E[\tau_A|X_0 = j] = U_j$$

3.4 Excises

Problem 1 Show that

$$E[N_t] = \lambda t \quad \text{and} \quad Var(N_t) = \lambda t.$$

Problem 2 The process defined by (**??**) is the Poisson process.
Problem 3 Construct a binary $S = \{0, 1\}$ continuous time Markov processes.
Problem 4 Let $\{L_t, \ t \geq 0\}$ and $\{M_t, \ t \geq 0\}$ be two independent Poisson processes with respective rates λ and μ. Show that the process $X_t = L_t - M_t$ is a continuous time Markov chain on $S = \{integers\}$ and find its generator. Let $P_n(t) = P(X_t = n|X_0 = 0)$. Show that

$$\sum_{n=-\infty}^{\infty} P_n(t)z^n = e^{-(\lambda+\mu)t}e^{\lambda zt + \mu z^{-1}t}, \quad |z| \neq 0$$

and

$$E[X_t] = (\lambda - \mu)t, \quad E[|X_t|^2] = (\lambda + \mu)t + (\lambda - \mu)^2\,t^2.$$

4 Markov Process

Let $(S,\mathcal{B})$ be a measurable space. A discrete time Markov process $\{X_n,\ n\geq 0\}$ is fully described by the one step transition probability $\Pi(x,A)$ defined for $x\in S$ and $A\in\mathcal{B}$, which is a probability measure on $(S,\mathcal{B})$ and

$$\Pi(x,A)=P(X_1\in A|X_0=x).$$

The multistep transition probability $\{\Pi^{(n)}(x,A)\}$ are determined by

$$\Pi^{(n+1)}(x,A)=\int_S \Pi^{(n)}(y,A)\Pi(x,dy).$$

The, they satisfies the Chapman-Kolmogorov equations;

$$\Pi^{(n+m)}(x,A)=\int_S \Pi^{(n)}(y,A)\Pi^{(m)}(x,dy).$$

In the continuous time Markov process $\{X_t,\ t\geq 0\}$ we use the transition probabilities $p(t,x,A)$ defined for $t\geq 0$, $x\in S$ and $A\in\mathcal{B}$ which is defined by

$$p(t,x,A)=P(X_t\in A|X_0=x).$$

They satisfy the Chapman-Kolmogorov equations

$$p(t+s,x,A)=\int_S p(s,y,A)p(t,x,dy).$$

Given transition probabilities, we define a consistent family of finite dimensional distributions on $(\Omega,\mathcal{F},P)$ by

$$F_{t_1,\cdots,t_n}(B_1\times\cdots\times B_n)=\int_{B_1}\int_{B_2}\cdots\int_{B_n} p(t_1,x,dy_1)p(t_2-t_1,y_1,dy_2)\cdots p(t_n-t_{n-1},y_{n-1},dy_n) \tag{4.1}$$

for the cylinder set, given arbitrary $0<t_1<\cdots<t_n$ and $B_j\in\mathcal{B}$. It reflects the fact that the increments $X_{t_j}-X_{t_{j-1}},\ 1\leq j\leq n$ are independent random variables. Conversely, such a consistent family of finite distributions by the Kolmogorov extension theory there exists a Markov process ω_t which satisfies

$$P(\bigcap_{j=1}^n\{\omega_{t_j}\in B_j\})=F_{t_1,\cdots,t_n}(B_1\times\cdots\times B_n)$$

4.1 Compouded Poisson Process

Suppose $\{Y_n,\ n\geq 1\}$ is i.i.d. random variables with distribution α. Let $S_n=Y_1+\cdots+Y_n$ and N_t is a Poisson process. We define s compound process $X_t=S_{N_t}$. Such a process inherits the independent increment property from N_t. The distribution of any increment $X_{t+h}-X_t$ is that of X_{N_t} and determined by the distribution of S_n where n is random variable and has a Poisson distribution with parameter λt; Assume Y_1 has the distribution α. Then,

$$E[e^{i(\xi,X_t)}]=\sum_{n=0}^\infty e^{-\lambda t}\frac{(\lambda t)^n}{n!}\hat{\alpha}(\xi)^n=e^{-\lambda t}e^{\lambda t\hat{\alpha}}=e^{\lambda t(\hat{\alpha}-1)}=e^{\lambda t\int_S(e^{i(\xi,x)}-1)d\alpha(x)},$$

where

$$E[e^{i(\xi,\sum_{k=1}^n Y_k)}]=E[\Pi_{k=1}^n e^{i(\xi,Y_k)}]=\hat{\alpha}(\xi)^n,\quad \hat{\alpha}(\xi)=\int_S e^{i\xi x}\,d\alpha(x).$$

In other words X_t has an infinitely divisible distribution with a Levy measure given by $\lambda t\,\alpha(x)$. If we let $M=\lambda\,\alpha$, we have

$$E[e^{i(\xi,X_t)}]=e^{t\int_S(e^{i(\xi,x)}-1)\,dM(x)}. \tag{4.2}$$

4.2 Infinite number of small jumps

A Poisson process cannot have an infinite number of jumps in a finite interval. But if we consider compounded Poisson processes we can, in principle by adding an infinite number of small jumps obtain a finite sum. That is, let $\{X_k(t)\}$ be a family of mutually independent compounded Poisson process with $M_k = \lambda_k \, \alpha_k$ and

$$X_t = \sum_k X_k(t)$$

If the sum exists then it is a process with independent increments. We may center these process with suitable constants $a_k \, t$ and we define

$$X_t = \sum_k (X_k(t) - a_k \, t)$$

We assume

$$\sum_k \int_{|x|>1} dM_k(x) < \infty \tag{4.3}$$

and

$$\sum_k \int_{|x|\leq 1} x^2 \, dM_k(x) < \infty \tag{4.4}$$

We decompose M_k as $M_k = M_k^{(1)} + M_k^{(2)}$ corresponding to jump of sizes $|x| \leq 1$ and $|x| > 1$. From

$$M^{(2)} = \sum_k M_k^{(2)}$$

sums to a finite measure and the corresponding process

$$X_t^{(2)} = \sum_k X_k^{(2)}(t)$$

exits. Since

$$\sum_k P(\sup_{0\leq s\leq t} |X_k(s)| \neq 0) \leq \sum_k (1 - e^{-tM_k^{(2)}(R)}) \leq \sum_k t \, M_k^{(2)}(R) < \infty$$

it follows from the Borel-Cantelli lemma, in any finite interval the sum is almost surely a finite sum. For the convergence of $\sum_k X_k^{(1)}(t)$ we let $a_k = \int_{|x|\leq 1} x \, dM_k(x)$ and we have

$$E[|X_k(t) - a_k \, t|^2] = t \int_{|x|\leq 1} x^2 \, dM_k(x)$$

From (**??**) and the two series theorem

$$\sum_k (X_k(t) - a_k \, t)$$

converges to $X_t^{(1)}$. A simple applications of Doob's inequality shows that in fact a.s. uniformly converges in finite time interval, i.e., define the tail

$$T_n(t) = \sum_{k\geq n} (X_k^{(1)}(t) - a_k t).$$

Since $E[X_k^{(1)}(t) - a_k \, t] = 0$, $T_n(t)$ is a martingale and by the Doob's martingale inequality

$$P(\sup_{0\leq t\leq T} \frac{1}{\delta^2} \sum_{k\geq n} V(X_k^{(1)}(t) - a_k \, t)) \to 0 \text{ as } n \to \infty.$$

If we now reassemble the pieces we obtain

$$E[e^{i\xi X_t}] = e^{t\int_{|x|\le 1}(e^{i\xi x}-1-i\xi\, x)dM(x)+t\int_{|x|>1}(e^{i\xi x}-1)\,dM(x)}, \tag{4.5}$$

which is the Levy-Kintchine representation of infinitely divisible distributions except for the missing Brownian motion term.

4.3 Feller semigroup

Let $B(S)$ be the Banach space of all essentially bounded functions $f(x) : S \to R$ with the norm

$$|f|_\infty = \sup_{x\in S} |f(x)|$$

Define a family of bounded linear operators $\{T(t),\ t \ge 0\}$ in $\mathcal{L}(B(S))$ by

$$(T(t)f)(x) = \int_S f(y)p(t,x,dy) = E^x[f(X_t)].$$

where

$$E^x[f(X_t)] = E[f(X_t)|X_0 = x]$$

The collection of $\{T(t),\ t \ge 0\}$ has the properties
(1) $T(t)$ maps nonnegative function on $(S, \mathcal{B})$ into nonnegative functions.
(2) $|T(t)f|_\infty \le |f|_\infty$ for all $f \in X$ and $T(t)1 = 1$. Thus, $\|T(t)\| = 1$.
(3) $T(0) = I$, $T(t+s) = T(t)T(s)$ (semigroup property) for $t,\ s \ge 0$.

Let $C_0(S)$ denote the space of all real-valued continuous functions on S that vanish at infinity, equipped with the sup-norm $|f| = |f|_\infty$. A Feller semigroup on $C_0(S)$ is a collection $\{T(t),\ t \ge 0\}$ of positive linear operators from $C_0(S)$ to itself such that
(1) $|T(t)f| \le |f|$ for all $t \ge 0$,
(2) the semigroup property: $T(t+s) = T(t)T(s)$ for all $s,\ t \ge 0$,
(3) $\lim_{t\to 0^+} |T(t)f - f| = 0$ for every f in $C_0(s)$) (strongly continuity at 0).

Thus, we let X be the subspace of $B(S)$ such that

$$X = \{f \in B(S) : \lim_{t\to 0^+} |T(t)f - f| \to 0\}$$

and the collection $\{T(t),\ t \ge 0\}$ forms the strongly continuous semigroup on X.

Let $\{X_n,\ n \ge 0\}$ be a discrete time Markov process with transition probability $\Pi(x, A)$. Define the bounded linear operator in X by

$$(\Pi f)(x) = \int_S f(y)\Pi(x,dy) = E[f(X_1)|X_0 = x]$$

Define a continuous time Markov process by $X_t = X_{N_t}$. Then,

$$T(t) = \sum_{n=0}^{\infty} e^{\lambda t}\frac{(\lambda t)^n}{n!}\Pi^n = e^{\lambda t(\Pi - I)} = e^{\mathcal{A}t}$$

where $\mathcal{A} = \lambda(\Pi - I)$.

In general we define the infinitesimal $\mathcal{A}$ of $\{T(t),\ t \ge 0\}$ by

$$\mathcal{A}f = s - \lim_{t\to 0^+} \frac{T(t)f - f}{t}$$

with domain

$$dom(\mathcal{A}) = \{f \in X : s - \lim_{t\to 0^+} \frac{T(t)f - f}{t} \text{ exists}\}.$$

If $\{X_t,\ t \geq 0\}$ is Markov process with stationary increments then we have a convolution semigroup

$$(T(t)f)(x) = \int_S f(x-y)\,\mu_t(dy),$$

where $\mu_{t+s} = \mu_t * \mu_s$ for $t,\ s \geq 0$ and

$$p(t,x,A) = \int_S 1_A(x+y)\,\mu_t(dy)$$

Then,

$$\mathcal{A}f = s - \lim_{t \to 0^+} \frac{\mu_t * f - f}{t}.$$

Theorem (C_0-semigroup) Let $u(t) = T(t)f = E^x[f(X_t)]$.
(1) If $u(t) = T(t)f \in C(0,T;X)$ for every $f \in X$.
(2) If $f \in dom\,(\mathcal{A})$, then $u \in C^1(0,T;X) \cap C(0,T;dom(\mathcal{A}))$ and

$$\frac{d}{dt}u(t) = \mathcal{A}u(t) = \mathcal{A}T(t)f.$$

(3) The infinitesimal generator $\mathcal{A}$ is closed and densely defined. For $f \in X$

$$T(t)f - f = \mathcal{A}\int_0^t T(s)f\,ds. \tag{4.6}$$

(4) $\lambda > 0$ the resolvent is given by

$$(\lambda\,I - \mathcal{A})^{-1} = \int_0^\infty e^{-\lambda s}T(s)\,ds \tag{4.7}$$

with estimate

$$|(\lambda\,I - \mathcal{A})^{-1}| \leq \frac{1}{\lambda}. \tag{4.8}$$

Proof: (1) follows from the semigroup property and the fact that for $h > 0$

$$u(t+h) - u(t) = (T(h) - I)T(t)f$$

and for $t - h \geq 0$

$$u(t-h) - u(t) = T(t-h)(I - T(h))f.$$

Thus, $x \in C(0,T;X)$ follows from the strong continuity of $S(t)$ at $t = 0$.
(2)–(3) Moreover,

$$\frac{u(t+h) - u(t)}{h} = \frac{T(h) - I}{h}T(t)f = T(t)\frac{T(h)f - f}{h}$$

and thus $T(t)f \in don(\mathcal{A})$ and

$$\lim_{h \to 0^+} \frac{u(t+h) - u(t)}{h} = \mathcal{A}T(t)f = \mathcal{A}u(t).$$

Similarly,

$$\lim_{h \to 0^+} \frac{u(t-h) - u(t)}{-h} = \lim_{h \to 0^+} T(t-h)\frac{T(h)f - f}{h} = S(t)\mathcal{A}f.$$

Hence, for $f \in dom(A)$

$$T(t)f - f = \int_0^t T(s)\mathcal{A}f\,ds = \int_0^t \mathcal{A}T(s)f\,ds = \mathcal{A}\int_0^t T(s)f\,ds \tag{4.9}$$

If $f_n \in don(\mathcal{A}) \to f$ and $\mathcal{A}f_n \to y$ in X, we have

$$T(t)f - f = \mathcal{A}\int_0^t T(s)y\,ds$$

Since

$$\lim_{t\to 0^+} \frac{1}{t}\int_0^t T(s)y\,ds = y$$

$f \in dom(\mathcal{A})$ and $y = \mathcal{A}f$ and hence $\mathcal{A}$ is closed. Since $\mathcal{A}$ is closed it follows from (**??**) that for $f \in X$

$$\int_0^t T(s)f\,ds \in dom(\mathcal{A})$$

and (**??**) holds. For $f \in X$ let

$$f_h = \frac{1}{h}\int_0^h T(s)f\,ds \in dom(\mathcal{A})$$

Since $f_h \to f$ as $h \to 0^+$, $dom(\mathcal{A})$ is dense in X.
(4) For $\lambda > 0$ define $R_t \in \mathcal{L}(X)$ by

$$R_t = \int_0^t e^{-\lambda s}T(s)\,ds.$$

Since $\mathcal{A} - \lambda\,I$ is the infinitesimal generator of the semigroup $e^{-\lambda t}T(t)$, from (**??**)

$$(\lambda\,I - \mathcal{A})R_t f = f - e^{-\lambda t}T(t)f \to f \text{ as } t \to \infty.$$

Since $\mathcal{A}$ is closed and $|e^{-\lambda t}T(t)| \to 0$ as $t \to \infty$, we have $R = \lim_{t\to\infty} R_t$ satisfies

$$(\lambda\,I - \mathcal{A})Rf = f.$$

Conversely, for $f \in dom(\mathcal{A})$

$$R(\mathcal{A} - \lambda\,I)f = \int_0^\infty e^{-\lambda s}T(s)(\mathcal{A} - \lambda\,I)f\,ds = \lim_{t\to\infty} e^{-\lambda t}T(t)f - f = -f$$

Hence

$$R = \int_0^\infty e^{-\lambda s}T(s)\,ds = (\lambda\,I - \mathcal{A})^{-1}$$

Since for $f \in X$

$$|Rf| \le \int_0^\infty |e^{-\lambda s}T(s)\phi| \le \int_0^\infty e^{(-\lambda)s}|\phi|\,ds = \frac{1}{\lambda}|f|,$$

we have

$$|(\lambda\,I - \mathcal{A})^{-1}| \le \frac{1}{\lambda}, \quad \lambda > 0l.$$

4.4 Infinitesimal generator

In this section we discuss examples of Markov process and the corresponding generators.

Example (Poisson process) For a Poisson process $\{N_t.\ t \ge 0\}$

$$\mathcal{A}f = \lambda(f(i+1) - f(i)), \quad i \in S = \{0, 1, \cdots\}$$

Example (Transport Process) For the shift (deterministic) process $x_t = ct$

$$T(t)f = E[f(X_t)|X_0 = x] = f(x + ct)$$

and

$$\mathcal{A}f = c\,f'(x) \text{ with } dom(\mathcal{A}) = \text{Lipschitz functions.}$$

Example (Levy process) Consider a process that has the Levy representation

$$E^x[e^{i\xi X_t}] = e^{t\int_R (e^{i\xi z}-1)\,dM(z)} e^{i\xi x}$$

with a finite Levy measure $M(dx)$. Then,

$$\mathcal{A}f = \int_R (f(x+z) - f(x))\,dM(z). \tag{4.10}$$

In fact we have for $f = e^{i\xi x}$

$$T(t)e^{i\xi x} = e^{t\int_R (e^{i\xi z}-1)\,dM(z)} e^{i\xi x}$$

and thus

$$\mathcal{A}e^{i\xi x} = \int_R (e^{i\xi(x+z)} - e^{ix\xi})\,dM(z).$$

Since for any f we have $f(x) = \frac{1}{2\pi}\int_R \hat{f} e^{i\xi x}\,d\xi$ by the inverse Fourier transform, (**??**) holds.

Example (Brownian Motion) A Brownian motion $\{B_t\ t \geq 0\}$ is a Markov process with the transition probability

$$p(t,x,A) = \int_A \frac{1}{\sqrt{2\pi t}\sigma} e^{-\frac{|y-x|^2}{2\sigma^2 t}}\,dy$$

and

$$E^x[e^{i\xi B_t}] = \frac{1}{\sqrt{2\pi t}\sigma}\int_R e^{i\xi y} e^{-\frac{|y-x|^2}{2\sigma^2 t}}\,dy = e^{-\frac{\sigma^2}{2}t|\xi|^2} e^{i\xi x}$$

Thus,

$$(\mathcal{A}f)(x) = -\frac{1}{2\pi}\int_R \frac{\sigma^2}{2}|\xi|^2 \hat{f}(\xi) e^{i\xi x}\,d\xi = -\frac{\sigma^2}{2} f''(x) \quad \text{with } dom(\mathcal{A}) = C_0^2(R).$$

Example (Levy-Kintchine process) For the process defined by (**??**) we have

$$(\mathcal{A}f)(x) = \frac{\sigma^2}{2} f''(x) + cf'(x) + \int_{|z|\leq 1} (f(x+z) - f(x) - zf'(x))\,dM(z) + \int_{|z|>1} (f(x+z) - f(x))\,dM(z).$$

Example (Cauchy Process) A Cauchy process $\{X_t\ t \geq 0\}$ is a Markov process with the transition probability

$$p(t,x,A) = \frac{1}{\pi}\int_A \frac{t}{t^2 + (y-x)^2}\,dy$$

and

$$E^x[e^{i\xi X_t}] = \frac{1}{\pi}\int_R e^{i\xi y}\frac{t}{t^2 + (y-x)^2}\,dy = e^{-t|\xi|} e^{i\xi x}$$

Thus,

$$(\mathcal{A}f)(x) = -\frac{1}{2\pi}\frac{1}{\pi}\int_R |\xi|\hat{f}(\xi) e^{i\xi x}\,d\xi = \frac{1}{\pi}\int_R \frac{f(z) - f(x)}{|x-z|^2}\,dz, \quad \text{with } dom(\mathcal{A}) = C_0^1(R).$$

In general, for the symmetric α-stable Levy process

$$E^x[e^{i\xi X_t}] = e^{-t|\xi|^\alpha} e^{i\xi x}.$$

Example (Gamma Process) A Gamma process $\{X_t,\ t \geq 0\}$ is a Markov process with the transition probability

$$p(t,x,A) = \int_A \frac{1}{\Gamma(t)} e^{-(y-x)}(y-x)^{t-1}\,dy$$

and

$$E^x[e^{i\xi X_t}] = \frac{1}{\Gamma(t)} \int_R e^{i\xi x} e^{-(1-i\xi)(y-x))} (y-x)^{t-1}\, dy = (1-i\xi)^{-t} e^{i\xi x}$$

Thus,

$$(\mathcal{A}f)(x) = \frac{1}{2\pi} \int_R e^{-(x-z)} \frac{f(x+z)-f(x)}{|x-z|}\, dz, \quad \text{with } dom(\mathcal{A}) = C_0^1(R).$$

4.5 Dynkin's formula

Theorem Let f be a bounded continuous function in $dom(\mathcal{A})$. Then

$$M_t = f(X_t) - f(x) - \int_0^t \mathcal{A}f(X_s)\, ds$$

is a martingale with respect $(\Omega, \mathcal{F}_t, P)$. Proof: For $t \geq s$

$$E^x[M_t - M_s|\mathcal{F}_s] = E^x[f(X_t) - f(X_s) - \int_s^t \mathcal{A}f(X_\sigma)\, d\sigma|\mathcal{F}_s] = T(t-s)f - f(X_s) - \int_s^t T(\sigma-s)\mathcal{A}f(X_s)\, d\sigma = 0.\square$$

Remark For $f \in dom(\mathcal{A})$

$$e^{\lambda t} f(X_t) - f(x) - \int_0^t e^{-\lambda s} \mathcal{A}f(X_s)\, ds \quad \text{for } \lambda \in R$$

$$f(X_t)\exp(-\int_0^t \frac{\mathcal{A}f(X_s)}{f(X_s)}\, ds) \quad \text{for uniformly positive } f$$

are martingales.

The characteristic operator $\mathcal{A}_c$ defined by

$$(\mathcal{A}_c f)(x) = \lim_{U \downarrow x} \frac{E^x[f(X_{\tau_U})] - f(x)}{E^x[\tau_U]},$$

where the sets U form a sequence of open sets U_k that decrease to the point x in the sense that

$$U_{k+1} \subseteq U_k \text{ and } \bigcap_{k=1}^{\infty} U_k = \{x\},$$

and

$$\tau_U = \inf\{t \geq 0 | X_t \notin U\}$$

is the exit time from U for X_t. $dom(\mathcal{A}_c)$ denotes the set of all f for which this limit exists for all $x \in S$ and all sequences $\{U_k\}$. If $E^x[\tau_U] = \infty$ for all open sets U containing x, define $\mathcal{A}_c f(x) = 0$. The characteristic operator is an extension of the infinitesimal generator, i.e., $dom(\mathcal{A}) \subset don(\mathcal{A}_c)$ and $\mathcal{A}_c f = \mathcal{A}f$ for $f \in dom(\mathcal{A})$.

Theorem (Dynkin's formula) Let f be a bounded continuous function in $dom(\mathcal{A}_c)$ and τ be a stoping time with $E[\tau] < \infty$. Then,

$$E^x[f(X_\tau)] = f(x) + E^x[\int_0^\tau \mathcal{A}_c f(X_s)\, ds].$$

4.6 Invariant measure

Let $T(t)f = E^x[f(X_t))] = \int_S f(y)p(t,x,y)\, dy$. Then for $f \in dom(\mathcal{A})$

$$\int_S f(y)p(t,x,y)\, dy = f(x) + \int_0^t \int_S p(s,x,y)\mathcal{A}f(y)\, dy.$$

Define the adjoint operator $\mathcal{A}^*$ of $\mathcal{A}$ is defined by

$$\int \mathcal{A}f\phi\,dy = \int f\mathcal{A}^*\phi\,dy \tag{4.11}$$

for all $f \in dom(\mathcal{A})$. Since $dom(\mathcal{A})$ is dense and there exists a unique closed linear operator $\mathcal{A}^*$ in X that satisfies (**??**). Thus, we have

$$\int_S (\int p(t,x,y) - \delta_x(y) - \mathcal{A}^* \int_0^t p(s,x,y)\,ds) f(y)\,dy = 0$$

for all $f \in \mathcal{A}$. Since $dom(\mathcal{A})$ is dense in X, it follows that

$$p(t,x,\cdot) = \delta_x + \mathcal{A}\int^t p(s,x,\cdot)\,ds.$$

Or, equivalently the transition probability p satisfies the Kolmogorov forward equation

$$\frac{\partial p}{\partial t} = \mathcal{A}^* p(t), \quad p(0) = \delta_x. \tag{4.12}$$

As we discussed in Section 4.1, if the state space S is countable, the invariant distribution π is defined as

$$\pi = \pi P(t), \quad t > 0$$

or equivalently

$$\pi Q = 0$$

where $P(t)$ is the transition probability matrix and Q is the transition rate matrix of the continuous time Markov chain $\{X_t,\ t \geq 0\}$. For the case S is continuum (e.g. $S = R$) we define the invariant measure μ (a bounded linear functional on $C(S)$) by

$$\mu(A) = \int_S p(t,x,A)\,d\mu(x)$$

for all $A \in \mathcal{B}$ and $t > 0$, or equivalently

$$\langle \mu, T(t)f\rangle = \langle \mu, f\rangle$$

for all $f \in C(S)$ and $t > 0$. One can state this as $T(t)^*\mu = \mu$ for all $t > 0$ or $A^*\mu = 0$, i.e.,

$$\langle \mu, Af\rangle = 0 \text{ for all } f \in dom(\mathcal{A}).$$

For the Ito's diffusion process

$$\mathcal{A}f = \frac{a(x)}{2}f'' + b(x)f'$$

and $d\mu = \phi\,dx$ satisfies

$$\mathcal{A}^*\phi = (\frac{a(x)}{2}\phi' + (-b(x) + \frac{a'(x)}{2})\phi)' = 0$$

and thus

$$\phi(x) = c\,e^{\int_0^x \frac{2b-a'}{a}\,dx}.$$

5 Martingale Process

In this section we consider a probability space $(\Omega, \mathcal{F}, P)$ and a nondecreasing sequence of σ-fields $\mathcal{F}_n$ contained in $\{\mathcal{F}_n,\ n \geq 0\}$.
Definition A sequence of real random variables $\{M_n\}$ is called a martingale with respect to the filtration $\{\mathcal{F}_n,\ n \geq 0\}$ if
(1) For each n, M_n is $\mathcal{F}_n$-measurable (that is, M_n is adapted to the filtration $\mathcal{F}_n$,
(2) For each n, $E[|X_n|] < \infty$,
(3) For each n, $E[M_{n+1}|\mathcal{F}_n] = M_n$.
The sequence $\{M_n\}$ is called a supermartingale (or submartingale) if property (iii) is replaced by

$$E[M_{n+1}|\mathcal{F}_n] \geq M_n \quad (\text{ or } E[M_{n+1}|\mathcal{F}_n] \leq M_n.$$

Notice that the martingale property implies that $E[M_n] = E[M_0]$ for all n. On the other hand, condition (iii) can also be written as

$$E[\Delta M_n|\mathcal{F}_{n-1}] = 0$$

for all n, where $\Delta M_n = M_n - M_{n-1}$.

Example 1 Suppose that ξ_n are independent centered random variables $(E[\xi_k] = 0,\ k \geq 1.$ Set $M_0 = 0$ and $M_n = \xi_1 \cdots + \xi_n$. Then M_n is a martingale with respect to the sequence of $\mathcal{F}_n = \sigma(\xi_1, \cdots, \xi_m),\ n \geq 1$.

Example 2 Suppose that $\{\xi_n,\ n \geq 1\}$ are independent random variable such that $\mathrm{P}(\xi_n = -1) = 1-p,\ P(\xi_n = 1) = p,\ 0 < p < 1$. Then $M_n = (\frac{1-p}{p})^{\xi_1 + \cdots + \xi_n}$ is a martingale with respect to the sequence of σ-fields $\sigma(\xi_1, \cdots, \xi_n),\ n \geq 1$. In fact,

$$E[M_{n+1}|\mathcal{F}_n] = E[(\tfrac{1-p}{p})^{\xi_{n+1}} M_n|\mathcal{F}_n] = E[(\tfrac{1-p}{p})^{\xi_{n+1}}]E[M_n|\mathcal{F}_n] = M_n.$$

Example 3 If M_n is a martingale and φ is a convex function such that $E[|\varphi(M_n)|] \leq < \infty$ for all n then $\varphi(M_n)$ is a submartingale. In fact, by Jensens inequality for the conditional expectation we have

$$E[\varphi(M_{n+1})|\mathcal{F}_n] \geq \varphi(E[M_{n+1}|\mathcal{F}_n]) = \varphi(M_n)$$

In particular, if $\{M_n\}$ is a martingale such that $E[|M_n|^p] < \infty$ for all n and for some $p \geq 1$, then $|M_n|^p$ is a submartingale.

Example 4 Suppose that $\{\mathcal{F}_n,\ n \geq 0\}$ is a given filtration. We say that $H_n,\ n \geq 1$ is a predictable sequence of random variables if for each n, H_n is $\mathcal{F}_{n-1}$-measurable. The martingale transform of a martingale M_n by a predictable sequence H_n as the sequence

$$(H \cdot M)_n = M_0 + \sum_{j=1}^{n-1} H_j \Delta M_j,$$

defines a martingale.

Example 5 (Likelihood Ratios) Let $\{Y_n,\ n \geq\}$ be i.i.d. random variables and let f_0 and f_1 be probability density functions. Define the sequence of probability ratios;

$$X_n = \frac{f_1(Y_0)f_1(Y_1)\cdots f_1(Y_n)}{f_0(Y_0)f_0(Y_1)\cdots f_0(Y_n)}$$

and let $\mathcal{F}_n = \sigma(Y_k,\ 0 \leq k \leq n)$. The, $\{X_n,\ n \geq 0\}$ is a martingale, i.e.,

$$E[X_{n+1}|\mathcal{F}_n)] = E[\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}X_n|\mathcal{F}_n] = E[\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}]X_n = X_n$$

where we used

$$E[\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}] = \int \frac{f_1(y)}{f_0(y)} f_0(y)\, dy = 1.$$

<u>Example 6 (Exponential Martingale)</u> Suppose that $\{Y_n\ n \geq 1\}$ are i.i.d. random variables with distribution $N(0, \sigma^2)$. Set $M_0 = 1$, and

$$M_n = e^{\sum_{k=1}^n Y_k - \frac{\sigma^2}{2}}$$

Then, $\{M_n\}$ is a nonnegative martingale. In fact

$$E[M_n|\mathcal{F}_{n-1}] = E[e^{Y_n - \frac{\sigma^2}{2}} M_{n-1}|\mathcal{F}_{n-1}] = E[e^{Y_n - \frac{\sigma^2}{2}}]M_{n-1} = M_{n-1}.$$

<u>Example 7 (Martingale induced by Eigenvector of Transition Matrix)</u> Let $\{Y_n,\ n \geq 0\}$ be a Poisson process with the transition probability P. Assume a bounded sequence $f(i) \geq 0$ satisfies

$$f(i) = \sum_j p_{ij} f(j).$$

Let $X_n = f(Y_n)$ and $\mathcal{F}_n = \sigma(Y_k,\ 0 \leq k \leq n)$. Then $\{X_n,\ n \geq 0\}$ is a martingale. In fact, $E[|X_n|] < \infty$ since f is bounded and

$$E[X_{n+1}|\mathcal{F}_n] = E[f(Y_{n+1})|\mathcal{F}_n[= E[f(Y_{n+1})|Y_n] = \sum_j p_{Y_n,j} f(j) = f(Y_n) = X_n.$$

<u>Example 8 (Radon-Nikodym derivatives)</u> Suppose Z be a uniformly distributed random variable on $[0, 1]$, define the sequence of random variables by setting

$$Y_n = \frac{k}{2^n}$$

for the unique k (depending on n and Z) that satisfies

$$\frac{k}{2^n} \leq Z < \frac{k+1}{2^n}.$$

That is, Y_n determines the first n bits of the binary representation of Z. Let f be a bounded function on $[0, 1]$ and form the finite difference quotient sequence

$$X_n = 2^n(f(Y_n + 2^{-n}) - f(Y_n)).$$

Then $\{X_n,\ n \geq 0\}$ is a martingale. In fact,

$$E[X_{n+1}|\mathcal{F}_n] = 2^{n+1}E[f(Y_{n+1} + 2^{-(n+1)}) - f(Y_{n+1})|\mathcal{F}_n]$$

$$= 2^{n+1}(\frac{1}{2}(f(Y_n + 2^{-(n+1)} - f(Y_n)) + \frac{1}{2}(f(Y_n + 2^{-n}) - f(Y_n + 2^{-(n+1)})))$$

$$= 2^n(f(Y_n + 2^{-n}) - f(Y_n)) = X_n.$$

where we used the fact that Z conditional on $\mathcal{F}_n$ has a uniform distribution $[Y_n, Y_n + 2^{-n})$ and thus Y_{n+1} is equally likely to be Y_n or $Y_n + 2^{-(n+1)}$.

<u>**Theorem (Martingale central limit theorem)**</u> Suppose $\{(M_n, \mathcal{F}_n)\}$ is a martingale sequence and let

$$V_k = E[|M_k - M_{k-1}|^2|\mathcal{F}_{n-1}]$$

If $\frac{V_n}{n} \to \sigma^2$ in probability and for all $\epsilon > 0$

$$\frac{1}{n}\sum_{k=1}^n \mathrm{E}\Big((M_k - M_{k-1})^2 I_{|M_k - M_{k-1}| > \epsilon\sqrt{n}}\Big) \to 0$$

as $n \to \infty$, then $\frac{M_n}{n}$ converges to $N(0, \sigma)$ in distribution.

5.1 Doob's decomposition

Theorem (Doob's Decomposition) Let $(X_n, \mathcal{F}_n)$ be a submartingale. The, there exit a unique Doob's decomposition of X_n such that for a martingale $(M_n, \mathcal{F}_n)$ and a predictable increasing sequence $(A_n, \mathcal{F}_{n-1})$ with $A_0 = 0$,

$$X_n = M_n + A_n.$$

Proof: Define

$$M_n = X_0 + \textstyle\sum_{j=1}^n (X_j - E[X_j|\mathcal{F}_{j-1}])$$

$$A_n = \textstyle\sum_{j=1}^n (E[X_j|\mathcal{F}_{j-1}] - X_{j-1}).$$

It is easy to see that (M_n, A_n) gives the desired decomposition. For the uniqueness, if we let $X_n = M_n' + A_n'$ the other decomposition, then

$$E[A_{n+1}' - A_n'|\mathcal{F}_n] = E[A_{n+1} - A_n) + (M_{n+1} - M_n) - (M_{n+1}' - M_n')|\mathcal{F}_n]$$

and thus we have

$$A_{n+1}' - A_n' = A_{n+1} - A_n$$

Since $A_0' = A_0$, this implies $A_n' = A_n$ for all $n \geq 1$ and hence the decomposition is unique.

The Doob's decomposition plays a key role in study of square integrable martingale $(M_n, \mathcal{F}_n))$, i.e., $E[M_n^2] < \infty$ for all $n \geq 0$. Since $\{\mathcal{M}_n^2,\ n \geq 0\}$ is a submartingale, from Theorem there exits a martingale m_n and a predictable increasing sequence $(\langle M \rangle_n, \mathcal{F}_{n-1})$ such that

$$M_n^2 = m_n + \langle M \rangle_n$$

The sequence $(\langle M \rangle_n, \mathcal{F}_{n-1})$ is called the quadratic variation of $\{M_n\}$ and is given by

$$\langle M \rangle_n = \sum_{j=1}^n E[(\Delta M_j)^2|\mathcal{F}_{j-1}],$$

where

$$E[(\Delta M_j)^2|\mathcal{F}_{j-1}] = E[M_j^2 - 2M_j M_{j-1} + M_{j-1}^2|\mathcal{F}_{j-1}] = E[M_j^2 - M_{j-1}^2|\mathcal{F}_{j-1}] = \langle M \rangle_j - \langle M \rangle_{j-1}$$

For $k \geq \ell$

$$E[(M_k - M_\ell)^2|\mathcal{F}_\ell)] = E[M_k^2 - M_\ell^2|\mathcal{F}_\ell] = E[\langle M \rangle_k - \langle M \rangle_\ell|\mathcal{F}_\ell].$$

In particular, if $M_0 = 0$, then $E[M_k^2] = E[\langle M \rangle_k]$.

If $(X_n, \mathcal{F}_n)$ and $(Y_n, \mathcal{F}_n)$ are square integrable martingales, we define

$$\langle X, Y \rangle_n = \frac{1}{4}(\langle X + Y \rangle_n - \langle X - Y \rangle_n).$$

It is easy to verify that

$$X_n Y_n - \langle X, Y \rangle_n \text{ is a martingale} \tag{5.1}$$

and for $k \geq \ell$

$$E[(X_k - X_\ell)(Y_k - Y_\ell)|\mathcal{F}_\ell] = E[\langle X, Y \rangle_k - \langle X, Y \rangle_\ell|\mathcal{F}_\ell]. \tag{5.2}$$

Moreover, we have

$$\langle X, Y \rangle_n = \sum_{j=1}^n E[\Delta X_j \Delta Y_j|\mathcal{F}_{j-1}] \tag{5.3}$$

In the case $X_n = \sum_{k=1}^n \xi_k$ and $Y_n = \sum_{k=1}^n \eta_k$, where $\{\xi_k\}$ and $\{\eta_k\}$ are sequences of independent square integrable random variables with $E[\xi_k] = E[\eta_k] = 0$, then

$$\langle X, Y \rangle_n = \sum_{j=1}^n E[\xi_j \eta_j].$$

Theorem For the martingale transform

$$(H \cdot M)_n = M_0 + \sum_{j=1}^{n} H_j \Delta M_j,$$

the quadratic variation is given by

$$\langle H \cdot M \rangle_n = \sum_{j=1}^{n} E[(H_j \Delta M_j)^2 | \mathcal{F}_{j-1})] = \sum_{j=1}^{n} |H_j|^2 E[|\Delta M_j|^2 | \mathcal{F}_{j-1}] = \sum_{j=1}^{n} |H_j|^2 \Delta \langle M \rangle_j.$$

5.2 Doob's Optional SamplingTheorem

Example 9 Let $\{\xi_k\},\ k \geq 1$ be an i.i.d sequence with Bernoulli random variables, $P(\xi_k = 1) = p,\ P(\xi_k = -1) = 1 - p$. Let $\mathcal{F}_n = \sigma(\eta_1, \cdot, \eta_n)$ and assume the player's stake V_n ($\mathcal{F}_{n-1}$-measurable) at the n-th turn. Then the player's gain X_n is

$$X_n = \sum_{k=1}^{n} V_k \xi_k$$

Then, $(X_n, \mathcal{F}_n))$ ia martingale if $p = \frac{1}{2}$. Consider the gambling strategy that doubles the stake after a loos and drops out the game immediately after a win, i.e, the stakes are

$$V_n = \begin{cases} 2^{n-1} & \text{if } \xi_1 = \cdots = \xi_{n-1} = -1 \\ 0 & \text{otherwise} \end{cases}$$

Then, if $\xi_1 = \cdots = \xi_{n-1} = -1$, the total loss after n turns is $\sum_{i=1}^{n} 2^{i-1} = 2^n - 1$. Thus, if $\xi_{n+1} = 1$, we have

$$X_{n+1} = X_n + V_{n+1} = -(2^n - 1) + 2^n = 1.$$

Let $\tau = \inf\{n \geq 1 : \xi_n = 1\}$. If $p = \frac{1}{2}$, the the game is fair and $P(\tau = n) = (\frac{1}{2})^n$, $P(\tau < \infty) = 1$ and $E[X_\tau] = 1$. Therefore, for a fair game, by applying this strategy, a player can in finite time complete the game successfully in increasing his capital by one unit ($E[X_\tau] = 1 > X_0 = 0$).

The following the basic theorem the typical case in which $E[X_\tau] = E[X_0]$ of a Markov time $\tau \geq 0$.

Theorem (Optional Sampling) Let $(X_n, \mathcal{F}_n)$ is a martingale (or submartingale), and $\tau_1 \leq \tau_2$ are stopping times. If

$$E[|X_{\tau_i}|] < \infty, \quad \liminf_{n \to \infty} E[|X_n| I\{\tau_i > n\}] = 0, \tag{5.4}$$

then

$$E[X_{\tau_2} | \mathcal{F}_{\tau_1}] = (\geq) X_{\tau_1} \text{ and } E[X_{\tau_2}] = (\geq) E[X_{\tau_1}].$$

Proof: It suffices to prove that for $A \in \mathcal{F}_{\tau_1}$,

$$\int_{A \cap \{\tau_2 \geq \tau_1\}} X_{\tau_2}\, dP = \int_{A \cap \{\tau_2 \geq \tau_1\}} X_{\tau_1}\, dP$$

for every $A \in \mathcal{F}_{\tau_1}$, or equivalently

$$\int_{B \cap \{\tau_2 \geq n\}} X_{\tau_2}\, dP = \int_{B \cap \{\tau_2 \geq n\}} X_{\tau_1}\, dP, \tag{5.5}$$

for $B = A \cap \{\tau_1 = n\}$ and all $n \geq 0$. Since

$$\int_{B \cap \{\tau_2 \geq n\}} X_n\, dP = \int_{B \cap \{\tau_2 = n\}} X_n\, dP + \int_{B \cap \{\tau_2 > n\}} E[X_{n+1} | \mathcal{F}_n]\, dP$$

$$= \int_{B \cap \{n \leq \tau_2 \leq n+1\}} X_{\tau_2}\, dP + \int_{B \cap \{\tau_2 \geq n+1\}} X_{n+2}\, dP$$

$$\cdots = \int_{B \cap \{n \leq \tau_2 \leq m\}} X_{\tau_2}\, dP + \int_{B \cap \{\tau_2 > m\}} X_m\, dP,$$

$$\int_{B\cap\{n\le\tau_2\le m\}} X_{\tau_2}\,dP + \int_{B\cap\{\tau_2\ge n\}} X_n\,dP = \int_{B\cap\{\tau_2>m\}} X_m\,dP.$$

Since $X_m = 2X_m^+ - |X_m|$, we have

$$\int_{B\cap\{\tau_2\ge n\}} X_{\tau_2}\,dP = \limsup_{m\to\infty}(\int_{B\cap\{\tau_2\ge n\}} X_n\,dP - \int_{B\cap\{\tau_2>m\}} X_m\,dP)$$

$$= \int_{B\cap\{\tau_2\ge n\}} X_n\,dP - \liminf_{m\to\infty}\int_{B\cap\{\tau_2>m\}} X_m\,dP = \int_{B\cap\{\tau_2\ge n\}} X_n\,dP,$$

which implies (**??**).
Example 9 (revisited)

$$\int_{\tau>n} |X_n|\,dP = (2^n-1)P(\tau>n) = (2^n-1)2^{-n} \to 1 \text{ as } n\to\infty.$$

and condition (**??**) is violated.

The followings are the versions of the Doob's optional sampling (optional stopping) theorem.

Corollary For some $N \ge 0$ such that $P(\tau_1 \le N) = P(\tau_2 \le N) = 1$, condition (**??**) holds and thus $E[X_\tau] = E[X_0]$.

Corollary If $\{X_n\}$ is uniformly integrable, then condition (**??**) holds and thus $E[X_\tau] = E[X_0]$.

Theorem Let $\{X_n\}$ be a martingale (or submartingale) and τ be a stopping time with respect to $\mathcal{F}_n = \sigma(X_k,\ k\le n)$. Suppose $E[\tau] < \infty$ and for all n and some constant C

$$E[|X_{n+1}-X_n||\mathcal{F}_n] \le C \quad (\{\tau\ge n\},\ P-a.s.)$$

Then,

$$E[|X_\tau|] < \infty \quad \text{and} \quad E[X_\tau] = (\ge)E[X_0].$$

Proof: Let $Y_0 = 0$ and $Y_j = |X_j - X_{j-1}|,\ j\ge 1$ Then, $|X_\tau| \le \sum_{j=0}^{\tau} Y_j$ and

$$E[|X_\tau|] \le E[\sum_{j=0}^{\tau} Y_j] = \sum_{n=0}^{\infty}\int_{\tau=n}\sum_{j=0}^{n} Y_j)\,dP$$

$$= \sum_{n=0}^{\infty}\sum_{j=0}^{n}\int_{\tau=n} Y_j\,dP = \sum_{j=0}^{\infty}\sum_{n=j}^{\infty}\int_{\tau=n} Y_j\,dP = \sum_{j=0}^{\infty}\int_{\tau\ge j} Y_j\,dP$$

Since $\{\tau\ge j\} = \Omega\setminus\{\tau<j\} \in \mathcal{F}_{j-1}$,

$$\int_{\tau\ge j} Y_j\,dP = \int_{\tau\ge j} E[Y_j|\mathcal{F}_{j-1}]\,dP \le C\,P(\tau\ge j)$$

and thus

$$E[|X_\tau|] \le E[\sum_{j=0}^{\tau} Y_j] \le E[|X_0|] + C\sum^{\infty} P(\tau\ge j) = E[|X_0|] + C\,P(\tau) < \infty$$

Moreover, if $\tau > n$ then

$$\sum_{j=0}^{n} Y_j \le \sum_{j=0}^{\tau} Y_j$$

and thus

$$\int_{\tau>n} |X_n|\,dP \le \int_{\tau>n} \sum_{j=0}^{\tau} Y_j\,dP.$$

Since $E[\sum_{j=0}^{\tau} Y_j] < \infty$ and $\{\tau > n\} \downarrow \varnothing$, it follows from the Lebesgue dominated convergence theorem that

$$\liminf_{n\to\infty} \int_{\tau>n} |X_n|\,dP \le \liminf_{n\to\infty} \int_{\tau>n} \sum_{j=0}^{\tau} Y_j\,dP = 0$$

Hence the theorem follows from Theorem (Optional Sampling).

Example (Wald's identities) Let $\{\xi_k,\ k \ge\}$ be i.i.d random variables with $E[|\xi_k|] < \infty$ and τ is a stopping time with respect to $\mathcal{F}_n = \sigma(\xi_k,\ k \le n)$. If $E[\tau] < \infty$,

$$E[\sum_{k=1}^{\tau} \xi_k] = E[\xi_1]\,E[\tau]$$

If moreover $E[|\xi_k|^2] < \infty$, then

$$E[|\sum_{k=1}^{\tau} \xi_k - \tau\,|\xi_1)|^2] = Var(\xi_1)\,E[\tau].$$

In fact,

$$X_n = \sum_{k=1}^{n} \xi_k - nE[\xi_1]$$

is a martingale and

$$E[|X_{n+1} - X_n||\mathcal{F}_n] = E[|\xi_{n+1} - E\xi_1)|\mathcal{F}_n] = E[|\xi_{n+1} - E[\xi_1]|) \le 2E[|\xi_1|] < \infty.$$

Thus, $E[X_\tau] = E[X_0] = 0$ and the claimed identity holds.

Example (Wald's fundamental identity) Let $\{\xi_k,\ k \ge\}$ be i.i.d random variables with and τ is a stopping time with respect to $\mathcal{F}_n = \sigma(\xi_k,\ k \le n)$. Define $S_n = \sum_{k=1}^{n} \xi_k$ assume $E[\tau] < \infty$ and $|S_n| \le C,\ (\tau > n,\ P - a.s.)$ (for example, $\tau = \{n \ge 0 : |S_n| \ge a\}$ for some $a > 0$)). Let $\phi(t) = E[e^{\xi_1 t}]$ and for some $t_0 \ne 0$, $\phi(t_0)$ exits and $\phi(t_0) \ge 1$. Then,

$$E[e^{t_0 S_\tau}\phi(t_0)^{-\tau}] = 1.$$

In fact, $X_n = e^{t_0 S_n}\phi(t_0)^{-n}$ is martingale and

$$E[|X_{n+1} - X_n||\mathcal{F}_n] = X_n E[|e^{t_0\xi_{n+1}}\phi(t_0)^{-1} - 1||\mathcal{F}_n] = X_n E[|e^{t_0\xi_{n+1}}\phi(t_0)^{-1} - 1|] < \infty.$$

The claimed identity follows from $E[X_1] = 1$.

5.3 Martingale Convergence

Theorem (Doob's Maximal Inequality) Suppose that $\{M_n\}$ is a submartingale. Then

$$P(\max_{k\le n} M_k \ge \lambda) \le \frac{1}{\lambda} E[M_n I\{\max_{k\le n} M_k \ge \lambda\}].$$

Proof: Define the stopping time $\tau = \min\{n \ge 0 : M_n \ge \lambda\} \wedge n$. Then, by the optional sampling theory,

$$E[M_n] \ge E[M_\tau] = E[M_\tau I\{\max_{k\le n} M_k \ge \lambda\}] + E[M_\tau|\{\max_{k\le n} M_k < \lambda\}]$$

$$\ge \lambda\,P(\max_{k\le n} M_k \ge \lambda) + E[M_n I\{\max_{k\le n} M_k < \lambda].$$

As a consequence, if $\{M_n\}$ is a martingale and $p \geq 1$, applying Doobs maximal inequality to the submartingale $\{|M_n|^p\}$ we obtain

$$P(\max_{0\leq n\leq N}|M_n| \geq \lambda) \leq \frac{1}{\lambda^p}E[|M_N|^p] \quad \text{for } p \geq 1, \tag{5.6}$$

which is a generalization of Chebyshev inequality.
Kolmogorov's Inequality Let $\{\xi_k,\ k \geq\}$ be i.i.d random variables with $E[\xi_k] = 0$ and $E[|\xi_1|^2] < \infty$. since $S_n = \sum_{k=1}^n \xi_k$ is a martingale with respect to $\mathcal{F}_n = \sigma(\xi_k,\ k \leq n)$,

$$P(\max_{k\leq n}|S_k| \geq \epsilon) \leq \frac{ES_n^2}{\epsilon^2}.$$

For $a < b$ let $\tau_0 = 1$ and

$$\tau_1 = \min\{n > 0; X_n \leq a\}, \quad \tau_2 = \min\{n > \tau_1; X_n \geq b\}, \cdots$$

$$\tau_{2n-1} = \min\{n > \tau_{2n-2}; X_n \leq a\}, \quad \tau_{2n} = \min\{n > \tau_{2n-1}; X_n \geq b\}, \cdots$$

Let $\beta_n(a,b) = \max\{m : \tau_{2m} \leq n\}$ be the upcrossing number of $[a,b]$ by the process $\{X_k,\ k \geq 1\}$.
Theorem (The Martingale Convergence Theorem) If $\{M_n\}$ is a submartingale such that $\sup_n E[M_n^+] < \infty$, then

$$M_n \to M \ \ a.s.,$$

where M is an integrable random variable.
Proof: First, since

$$E[M_n^+] \leq E[|M_n|] = 2\,E[M_n^+] - E[M_n] \leq 2\,E[M_n^+] - E[M_1],$$

we have $\sup_n\ E[|M_n|] < \infty$. Suppose that

$$A = \{\limsup M_n > \liminf M_n\} \quad \text{and} \quad P(A) > 0.$$

The since

$$A = \cup_{a<b}(\limsup M_n > b > a > \liminf M_n \ \text{ where } a,\ b \text{ are rational numbers}$$

for some rational numbers $a,\ b$

$$P(\{\limsup M_n > b > a > \liminf M_n\}) > 0 \tag{5.7}$$

Let $\beta_n(a,b)$ be the number of upcrossings of (a,b) by the sequence $M_1,\cdots,M_n$.

$$E[\beta_n(a,b)] \leq \frac{E[(M_n-a)^+]}{b-a} \leq \frac{E[M_n^+]+|a|}{b-a}$$

and thus

$$\lim_{n\to\infty} E[\beta_n(a,b)] \leq \frac{\sup_n\ E[M_n^+]+|a|}{b-a}$$

which contradicts to assumption (**??**). Hence $\lim_{n\to\infty} M_n = M$ exists and by Fatou' lemma

$$E|M| \leq \sup_n E|M_n| < \infty$$

Example 6 (revisited) Since $E[|M_n|] = 1$ and $\lim_{n\to\infty} M_n$ exists almost surely. By the law of large number $\frac{\sum_{k=1}^n Y_k}{n} \to 0$ in probability, we have $\lim_{n\to\infty} M_n = 0,\ a.s.$.
Theorem (P.Levy) Let ξ be an integrable random variable and $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$. Then,

$$E[\xi|\mathcal{F}_n] \to E[\xi|\mathcal{F}_\infty] \quad \text{a.s. and in } L^1.$$

Proof: Let $X_n = E[\xi|\mathcal{F}_n]$. For $a > 0$ and $b > 0$

$$\int_{\{|X_n|\geq a\}} |X_n|\, dP \leq \int_{\{|X_n|\geq a\}} E[|\xi||\mathcal{F}_n]\, dP = \int_{\{|X_n|\geq a\}} |\xi|\, dP$$

$$= \int_{\{\{|X_n|\geq a\}\cap\{|\xi|\leq b\}\}} |\xi|\, dP + \int_{\{\{|X_n|\geq a\}\cap\{|\xi|>b\}\}} |\xi|\, dP$$

$$\leq bP(|X_n| \geq a) + \int_{\{|\xi|\leq b\}} |\xi|\, dP$$

$$\leq \frac{b}{a}E[|X_n|] + \int_{\{|\xi|\leq b\}} |\xi|\, dP \leq \frac{b}{a}E[|\xi|] + \int_{\{|\xi|\leq b\}} |\xi|\, dP$$

Letting $a \to \infty$ and the $b \to \infty$ in this, we have

$$\lim_{a\to\infty} \sup_n \int_{\{|X_n|\geq a\}} |X_n|\, dP = 0,$$

i.e., $\{X_n\}$ is uniformly integrable. Thus, from Martingale Convergence theorem there exists a random variable X such that $X_n = E[\xi|\mathcal{F}_n] \to X$ a.s and in L^1. For the last assertion let $m \geq n$ and $A \in \mathcal{F}_n$. Then,

$$\int_A X_m\, dP = \int_A X_n\, dP = \int_A E[\xi|\mathcal{F}_n]\, dP = \int_A \xi\, dP.$$

Since $\{X_n\}$ is uniformly integrable, $E[I_A|X_m - X|] \to 0$ as $m \to \infty$ and

$$\int_A X\, dP = \int_A \xi\, dP$$

for all $A \in \mathcal{F}_n$ and thus for all $A \in \cup_n \mathcal{F}_n$. Since $E|X| < \infty$ and $E|\xi| < \infty$ the left and right hand side of the above inequalities define σ-additive measures on the algebra $\cup_n\mathcal{F}_n$. By Caratheodory's theorem there exists the unique extension on these measures to $\mathcal{F}_\infty = \sigma(\cup_n\mathcal{F}_n)$. Thus,

$$\int_A X\, dP = \int_A \xi\, dP = \int_A E[\xi|\mathcal{F}_\infty]\, dP.$$

Since X is $\mathcal{F}_\infty$-measurable, $X = E[\xi|\mathcal{F}_\infty]$.

Corollary (Doob Martingale) A $\{M_n\}$ is uniformly integrable martingale if and only if there exists an integrable random variable M such that $M_n = E[X|\mathcal{F}_n]$ for $n \geq 1$.

Proof: Since $\{M_n\}$ is uniformly integrable, $\sup_n E[|M_n|] < \infty$ and $M_n \to M$ in $L^1(\Omega, P)$ as $n \to \infty$. Since $\{M_n\}$ is a martingale, for $A \in \mathcal{F}_m$ and $n \geq m$,

$$\int_A E[M_n|\mathcal{F}_m]\, dP = \int_A M_m\, dP$$

But, we have

$$\int_A E[M_n|\mathcal{F}_m]\, dP = \int_A M_n\, dP$$

Hence

$$|\int_A (M_m - M)\, dP| = |\int_A (M_n - M)\, dP| \leq \int_\Omega |M_n - M|\, dP \to 0$$

as $n \to \infty$ and

$$\int_A M_m\, dP = \int_A M\, dP.$$

Corollary If $(M_n,\mathcal{F}_n)$ is submartingale, and for some $p>1$ $\sup_n\ E[|M_n|]<\infty$ then there exits an integrable random variable M such that

$$M_n=E[M|\mathcal{F}_n] \quad \text{and} \quad M_n\to M \text{ in } L^p.$$

Corollary If $(M_n,\mathcal{F}_n)$ is a martingale

$$\frac{M_n}{\langle M\rangle_n}\to 0 \ \ P-a.s.$$

Example 8 (revisited) Assume f is Lipschitz continuous,i.e. $|f(x)-f(y)|\le L\,|x-y|$. Then $|X_n|\le L$. Note that $\mathcal{F}=\mathcal{B}[0,1])=\sigma(\cup_n\mathcal{F}_n)$ there is $\mathcal{F}$-measurable function $g=g(x)$ such that $X_n\to g$ a.s. and

$$X_n=E[g|\mathcal{F}_n]$$

Thus, for $B=[0,k2^{-n}]$

$$f(k2^{-n})-f(0)=\int_0^{k2^{-n}}X_n\,dx=\int_0^{k2^{-n}}g\,dx.$$

Since n and k are arbitrary, we obtain

$$f(x)-f(0)=\int_0^x g(s)\,ds,$$

i.e., f is absolutely continuous and $\frac{d}{dx}f=g$ a.s.

5.4 Continuous time Martingale and Stochastic integral

Let $\{X_t,\ t\ge 0\}$ be a continuous time stochastic process on a probability space $(\Omega,\mathcal{F},P)$ and $\{\mathcal{F}_t,\ t\ge 0\}$ be a family of sub-σ algebras with $\mathcal{F}_s\subset\mathcal{F}_t$ for all $t>s\ge 0$. A random variable $\tau\ge 0$ is a Markov time with respect to the filteration $\mathcal{F}_t$ if for all $t\ge 0$, the event $\{\tau\le t\}$ is $\mathcal{F}_t$ measurable,i.e., the event is completely described by the information available up to time t. For continuous time process it is not sufficient to require $\{\tau=t\}$ is $\mathcal{F}_t$ measurable for all $t\ge 0$. If $\tau_1,\ \tau_2$ are Markov times, so are $\tau_1+\tau_2$, $\tau_1\wedge\tau_2=\min(\tau_1,\tau_2)$ and $\tau_1\vee\tau_2=\max(\tau_1,\tau_2)$. Thus, $\tau\wedge t$ is a Markov time. For example, let $\mathcal{F}_t=\sigma(X_s,\ s\le t)$ of a continuous process X_t. The exit time from an open set A;

$$\tau_A=\inf\{t: X_t\notin A\}$$

is a Markov process, i.e,

$$\{\tau>t\}=\bigcup_{k=1}^\infty\bigcap_{r\in Q,\ 0\le r\le t}\{dist(X_r,A^c)\ge\frac{1}{k}\}.$$

In general if X_t is not continuous τ_A is not necessary a Markov time. Suppose $t\to X_t(\omega)$ is continuous from the right and has a limit from the left, i.e., $X_t=\lim_{s\downarrow t}$ and $X_{t-}=\lim_{s\uparrow t}X_s$ exists for all $t\ge 0$. Let

$$\mathcal{F}_{t^+}=\bigcap_{s>t}\mathcal{F}_s$$

Then, $\mathcal{F}_{t^+}$ is a σ algebra, X_t is $\mathcal{F}_{t^+}$ and $\mathcal{F}^{t^+}\subset\mathcal{F}_{s^+}$ for $t<s$. Next, $\bar{\mathcal{F}}_{t^+}$ be the smallest σ algebra containing every set in $\mathcal{F}_{t^+}$ and every set A in $\mathcal{F}$ with $P(A)=0$, i.e, it consists of all events that are $P-a.s.$ equivalent to events in $\mathcal{F}_{t^+}$. Then, for every Borel set B, the arrival time

$$\tau_B=\begin{cases}\inf\{t\ge 0: X_t\in B\}, & X_t\in B \text{ for some } t\ge 0\\ \infty, & X_t\notin B \text{ for all } t\ge 0,\end{cases}$$

is a Markov time with respect to $\bar{\mathcal{F}}_{t^+}$.

Given a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$, then a continuous-time stochastic process $(X_t)_{t\geq 0}$ is a martingale (submartingale) if
(a) X_t is $\mathcal{F}_t$ measurable for all $t \geq 0$.
(b) $E[X_t^+] < \infty$.
(c) $X_t = (\geq) E[X_t|\mathcal{F}_s]$ for all $t \geq s \geq 0$.

Both the martingale optional sampling and convergence theorems hold for continuous time, i.e.,

$$E[X_0] \leq E[X_{\tau\wedge t}] \leq E[X_t]$$

for all Markov times τ. Here, the inequalities for a submartingale and the equalities for a martingale. If $P(\tau < \infty) = 1$ then $P-$a.s.

$$X_{\tau\wedge t} \to X_\tau \text{ as } t \to \infty.$$

Theorem (Optional Sampling) Let $\{X_t,\ t \geq 0\}$ be a martingale (submartingale) and τ is a Markov time with respect to $\mathcal{F}_t$. If $P(\tau < \infty)$ and the random variables $\{X_{t\wedge\tau}^+,\ t \geq 0\}$ are uniformly integrable, then $E[X_0] = (\leq) E[X_\tau]x$.

Corollary Let $\{X_t,\ t \geq 0\}$ is a martingale and τ is a Markov time with respect to $\mathcal{F}_t$. If $P(\tau < \infty)$ and $E[\sup_{t\geq 0} |X_t, t \geq 0\}] < \infty$, then $E[X_0] = E[X_\tau]$.

We use these results to derive a number of important proprieties of the Brownian motion in Chapter 7.

Example (Poisson Process) If $\{N_t,\ t \geq 0\}$ is a Poisson process with parameter λ, then

$$N_t - \lambda\, t,\ \ (N_t - \lambda t)^2 - \lambda\, t,\ \ e^{-\theta\, N_t + \lambda t\, (1-e^{-\theta})} \tag{5.8}$$

are martingales with respect to $\mathcal{F}_t = \sigma(N_s,\ s \leq t)$. Let a is a positive integer and $\tau_a = \inf\{t \geq 0 : N_t \geq a\}$ starting from $N_0 = 0$. With the observation $N_{\tau_a} = a$, we have

$$a = \lambda E[\tau_a],\ \ E[(\lambda\, \tau_a - a)^2] = \lambda\, E[\tau_a] = a,\ \ E[e^{-\beta\tau_a}] = e^{\theta a} = \left(\frac{\lambda}{\lambda+\beta}\right)^a \tag{5.9}$$

where $\beta = -\lambda\,(1 - e^{-\theta})$. The last equation is the Laplace transform of τ_a and it shows that τ_a has a gamma distribution with parameters a and λ.

Example (Birth Processes) Let $\{X_t.\ t \geq 0\}$ be a pure birth process having the birth rate $\lambda(i)$ for $i \geq 0$. If $X_t = 0$ then

$$Y_t = X_t - \int_0^t \lambda(X_s)\, ds, \quad V_t = e^{\theta X_t + (1-e^\theta)\int_0^t \lambda(X_s)\, ds}$$

are martingales with respect to $\mathcal{F}_t = \sigma(X_s,\ s \leq t)$.

Lemma 2.2 Suppose M_t is almost surely continuous martingale with respect to $(\Omega, \mathcal{F}_t, P)$ and A_t is a progressively measurable (predictable) function, which is almost surely continuous and of bounded variation in t. Then, under the assumption that $\sup_{0\leq s\leq t} |M_s| \mathrm{Var}_{[0,t]} A(\cdot,\omega)$ is integrable,

$$M_t A_t - M_0 A_0 - \int_0^t M(s)\, dA(s)$$

is a martingale.

Proof: The main step is to see why

$$E[M_t A_t - M_0 A_0 - \int_0^t M_s\, dA_s] = 0$$

Then the same argument, repeated conditionally will prove the martingale property.

$$E[M_tA_t - M_0A_0] = \lim \sum_j E[M_{t_j}A_{t_j} - M_{t_{j-1}}A_{t_{j-1}}]$$

$$= \lim \sum_j E[E[(M_{t_j} - M_{t_{j-1}}]A_{t_{j-1}}|\mathcal{F}_{t_{j-1}}] + M_{t_j}(A_{t_j} - A_{t_{j-1}}))$$

$$= \lim \sum_j E[M_{t_j}(A_{t_j} - A_{t_{j-1}}] = E[\int_0^t M_s\,dA_s].$$

where the assumption and the dominated convergence theorem. □

For

$$M_t = f(X_t) - f(X_0) - \textstyle\int_0^t \mathcal{A}f(X_s)\,ds$$

$$A_t = e^{-\int_0^t \frac{\mathcal{A}f(X_s)}{f(X_s)}\,ds}$$

we have

$$f(X_t)e^{-\int_0^t \frac{\mathcal{A}f(X_s)}{f(X_s)}\,ds} \tag{5.10}$$

is a martingale if f is uniformly positive. In fact

$$M_tA_t - M_0A_0 - \int_0^t M_s\,dA_s = f(X_t)A_t - f(X_0)A_0. \tag{5.11}$$

5.5 Excises

Problem 1 Show (**??**)–(**??**).

Problem 2 If $\{\xi_k,\ k \geq 1\}$ is a sequence of independent random variables with $E[\xi_k] = 1$. Show that $X_n = \Pi_{k=1}^n \xi_k$ is a martingale with respect to $\mathcal{F}_n = \sigma(\xi_k\ k \leq n)$. Consider the case $P(\xi_k = 0) = P(\xi_k = 2) = \frac{1}{2}$. Show that there is no an integrable random variable ξ such that $X_n = E[\xi|\mathcal{F}_n]$.

Problem 3 Let $\{\xi_k\}$ be a sequence of independent random variables with $E[\xi_k] = 0$ and $V(\xi_k) = \sigma_k^2$. Define $S_n = \sum_{k=1}^n \xi_k$ and $\mathcal{F}_n = \sigma(\xi_k,\ k \leq n)$. Show the following generalization of Wald's identities. If $E[\sum_{k=1}^\tau |\xi_k|] < \infty$ then $E[S_\tau] = 0$. If $E[\sum_{k=1}^\tau |\xi_k|^2] < \infty$ then $E[S_\tau^2] = E[\sum_{k=1}^\tau \sigma_k^2]$.

Problem 4 Show (**??**) and (**??**).

Problem 5 Suppose $\{X_n\}$ is a martingale satisfying some $p > 1$ $E[|X_n|^p] < \infty$. Show

$$(E[(\max_{0\leq k\leq n} |X_k|)^p])^{\frac{1}{p}} \leq \frac{p}{p-1} E[|X_n|^p]^{\frac{1}{p}}.$$

Hint: $E[|\xi|^p] = p\int_0^\infty t^{p-1}P(|\xi| > t)\,dt$. Now, we use the maximal inequality for the submartingale $|X_n|$.

Problem 6 Show (**??**)-(**??**).

Problem 7 Show that $X_t = e^{\lambda B_t - \frac{\lambda^2 t}{2}}$ satisfies $dX_t = \lambda X_t\,dB_t$. Find the generator of X_t.

6 Brownian Motion and Ito's Calculus

In 1827 Robert Brown observed the complex and erratic motion of grains of pollen suspended in a liquid. It was later discovered that such irregular motion comes from the extremely large number of collisions of the suspended pollen grains with the molecules of the liquid. In the 20s Norbert Wiener presented a mathematical model for this motion based on the theory of stochastic processes. The position of a particle at each time $t \geq 0$ is a d dimensional random vector B_t. The mathematical definition of a Brownian motion is the following Definition:

Definition (Brownian motion) A stochastic process $B_t,\ t \geq 0$ is called a Brownian motion if it satisfies the following conditions:
i) For all $0 = t_0 \geq t_1 < \cdots < t_n$ the increments $B_{t_n} - B_{t_{n-1}}, \cdots B_{t_1} - B_{t_0}$ are independent random variables.
ii) If $0 \leq s < t$, the increment $B_t - B_s$ has the normal distribution $N(0, t-s)$.

Theorem (Continuous Process) If X_t is a stochastic process on $(\Omega, \mathcal{F}, P)$ satisfying

$$E[|X_t - X_s|^\alpha] \leq C|t-s|^{1+\beta}$$

for some positive constants $\alpha,\ \beta$ and C, then if necessary, $X_t,\ t \geq 0$ can be modified for each t on a set of measure zero, to obtain an equivalent version that is almost surely continuous.

For the Brownian Motion, from (ii) an elementary calculation yields

$$E|B_t - B_s|^4 = 3|t-s|^2$$

so that Theorem with $\alpha = 3,\ \beta = 1$ and $C = 3$ applies.

Remark (1) With probability 1 Brownian paths satisfy a Holder condition with any exponent less than $\frac{1}{2}$. It is not hard to see that they do not satisfy a Holder condition with exponent $\frac{1}{2}$. The random variables $(B_t - B_s)/\sqrt{|t-s|}$ have standard normal distributions for any interval $[s,t]$ and they are independent for disjoint intervals. We can find as many disjoint intervals as we wish and therefore dominate the Holder constant from below by the supremum of absolute values of an arbitrary number of independent Gaussians.
(2) The mapping $\omega \to B_t(\omega) \in C([0,1);R)$ induces a probability measure P_B which is called the Wiener measure, on the space of continuous functions $C = C([0;1);R)$ equipped with its Borel-field $\mathcal{B}(C)$, generated by open balls in C. Then we can take as canonical probability space for the Brownian motion the space $(C, \mathcal{B}(C), P_B)$. In this canonical space, the random variables are the evaluation maps: $X_t(\omega) = \omega(t)$.

First, we will show that the quadratic variation $\sum_{k=1}^n |\Delta B_k|^2,\ \Delta B_k = B_{t_k} - B_{t_{k-1}}$ converges in mean square to the length of the interval as the length of the subdivision tends to zero, i.e.,

$$\sum_{k=1}^n |B_{t_k} - B_{t_{k-1}}|^2 \to t$$

as $\max_k |t_k - t_{k-1}| \to 0$.

$$E[(\textstyle\sum_{k=1}^n |\Delta B_k|^2 - t)^2] = \sum_{k,\ell} (|\Delta B_k|^2 - \Delta t_k)(|\Delta B_\ell|^2 - \Delta t_\ell)$$

$$= \sum_{k=1}^n E[(|\Delta B_k|^2 - \Delta t_k)^2] = \sum_k 3(\Delta t_k)^2 - 2(\Delta t_k)^2 + (\Delta t_k)^2$$

$$= \sum_k^n 2(\Delta t_k)^2 \leq 2t \max_k |\Delta t_k| \to 0.$$

where we used if $k \neq \ell$

$$E[(|\Delta B_k|^2 - \Delta t_k))(|\Delta B_\ell|^2 - \Delta t_\ell)] = 0$$

$$E[(|\Delta B_k|^2 - \Delta t_k)^2] = E[|\Delta B_k|^4 - 2E[|\Delta B_k|^2]\Delta t_k + (\Delta t_k)^2.$$

On the other hand, the total variation, defined by $V = sup\ \sum_{k=1}^n |\Delta B_k|$ over all partition $0 = t_0 < t_1 < \cdots < t_n = t$, is infinite with probability one. In fact, using the continuity of the trajectories of the Brownian motion, we have

$$\sum_{k=1}^n |\Delta B_k|^2 \leq sup_k |\Delta B_k| \sum_{k=1}^n |\Delta B_k| \leq V\, sup_k |\Delta B_k| \to 0$$

if $V < \infty$, which contradicts the fact that $\sum_{k=1}^n |\Delta B_k|^2$ converges in mean square to t.

6.1 Brownian motion and Martingale

If $\{B_t,\ t \geq 0\}$ is a Brownian motion and $\mathcal{F}_t$ is the filtration generated by B_t, then, the processes B_t, $|B_t|^2 - t$ and $e^{\lambda B_t - \frac{\lambda^2}{2}t}$ are martingales. In fact

$$E[e^{\lambda B_t - \frac{\lambda^2}{2}t}|\mathcal{F}_s] = E[e^{\lambda(B_t-B_s) - \frac{\lambda^2}{2}(t-s)}e^{\lambda B_s - \frac{\lambda^2}{2}s}|\mathcal{F}_s] = E[e^{\lambda(B_t-B_s) - \frac{\lambda^2}{2}(t-s)}]e^{\lambda B_s - \frac{\lambda^2}{2}s} = e^{\lambda B_s - \frac{\lambda^2}{2}s}$$

Consider the stopping time $\tau_a = \inf\{t \geq 0 : B_t = a\}$ for $a > 0$. Since the process $M_t = e^{\lambda B_t - \frac{\lambda^2}{2}t}$ is a martingale, $E[M_t] = E[M_0] = 1$. By the Optional Stopping theorem we obtain $E[M_{\tau_a \wedge N}] = 1$ for all $N \geq 1$. Note that

$$M_{\tau_a \wedge N} = e^{\lambda B_{\tau_a \wedge N} - \frac{\lambda^2}{2}\tau_a \wedge N} \leq e^{\lambda a}.$$

On the other hand,

$$\lim_{N\to\infty} M_{\tau_a \wedge N} = M_{\tau_a} \text{ if } \tau_a < \infty, \quad \lim_{N\to\infty} M_{\tau_a \wedge N} = 0 \text{ if } \tau_a = \infty,$$

and the dominated convergence theorem implies $E[I\{\tau_a < \infty\}M_{\tau_a}] = 1$, that is,

$$E[I\{\tau_a < \infty\}e^{-\frac{\lambda^2}{2}\tau_a}] = e^{-\lambda a}.$$

Letting $\lambda \to 0^+$, we obtain $P(\tau_a < \infty) = 1$ and consequently,

$$E[e^{-\frac{\lambda^2}{2}\tau_a}] = e^{-\lambda a} \tag{6.1}$$

With the change of variables $\frac{\lambda^2}{2} = \alpha$, we have

$$E[e^{-\alpha\tau_a}] = e^{-\sqrt{2\alpha}a}. \tag{6.2}$$

From this expression we can compute the distribution function of the random variable τ_a;

$$P(\tau_a \leq t) = \int_0^t \frac{ae^{-a^2/2s}}{\sqrt{2\pi s^3}}\,ds.$$

On the other hand, the expectation of τ_a can be obtained by computing the derivative of (**??**) with respect to the variable a:

$$E(\tau_a e^{-\alpha\tau_a}) = \frac{ae^{-2\sqrt{\alpha}a}}{\sqrt{2\alpha}}.$$

and letting $\alpha \to 0^+$ we obtain $P(\tau_a < \infty) = 1$.
(3) One can use the Martingale inequality in order to estimate the probability $P(\sup_{s\leq t}|B_s| \geq \ell)$. For $A > 0$, by Doob's inequality

$$P(\sup_{s\leq t} e^{\lambda B_s - \frac{\lambda^2}{2}s} \geq A) \leq \frac{1}{A}.$$

and thus

$$P(\sup_{s\leq t} B_s \geq \ell) \leq P(\sup_{s\leq t}|B_s - \tfrac{\lambda s}{2}| \geq \ell - \tfrac{\lambda}{2}t)$$

$$= P(\sup_{s\leq t}|\lambda B_s - \frac{\lambda^2 s}{2}| \geq \lambda\ell - \frac{\lambda^2}{2}t) \leq e^{\lambda\ell - \frac{\lambda^2}{2}t}$$

Optimizing over $\lambda > 0$ we obtain

$$P(\sup_{s\leq t} B_s \geq \ell) \leq e^{-\frac{\ell^2}{2t}}$$

and by symmetry

$$P(\sup_{s\le t}|B_s|\ge \ell)\le 2e^{-\frac{\ell^2}{2t}}$$

The estimate is not too bad because by reflection principle

$$P(\sup_{s\le t}|B_s|\ge \ell)\ge 2P(B_t\ge \ell)=\sqrt{\frac{2}{2\pi t}}\int_\ell^\infty e^{-\frac{x^2}{2t}}\,dx=\sqrt{\frac{2}{\pi}}\int_{\frac{\ell}{\sqrt{t}}}^\infty e^{-\frac{y^2}{2}}\,dy$$

and thus

$$\lim_{t\to\infty}P(\tau_\ell\le t)=1.$$

In particular, the one-dimensional Brownian motion starting from 0 will get up to any level ℓ at some time.

Theorem (Levy theorem) If P is a measure on $(C[0,1],\mathcal{B},P)$ such that $P(X_0=0)=1$ and the the functions X_t and $|X_t|^2-t$ are martingales with respect to $(C[0,T],\mathcal{B}_t,P)$ then P is the Wiener measure.

Proof: The proof is based on the observation that a Gaussian distribution is determined by two moments. But that the distribution is Gaussian is a consequence of the fact that the paths are almost surely continuous and not part of our assumptions. The actual proof is carried out by establishing that for each real number λ

$$X_\lambda(t)=e^{\lambda\,X_t-\frac{\lambda^2}{2}t} \tag{6.3}$$

is a martingale with respect to $(C[0;T];\mathcal{B}_t,P)$. Once this is established it is elementary to compute

$$E[e^{\lambda\,(X_t-X_s)}|\mathcal{B}_s]=e^{\frac{\lambda^2}{2}(t-s)} \tag{6.4}$$

which shows that we have a Gaussian process with independent increments with two matching moments $(0,t-s)$. The proof of (**??**) is more or less the same as proving the central limit theorem. In order to prove (**??**) we can assume with out loss of generality that $s=0$ and will show that

$$E[e^{\lambda X_t-\frac{\lambda^2}{2}t}]=1. \tag{6.5}$$

To this end let us define successively $\tau_{0,\epsilon}=0$ and

$$\tau_{k+1,\epsilon}=\min(\inf\{s\ge\tau_{k,\epsilon}:|X_s-X_{\tau_{k,\epsilon}}|\ge\epsilon\},t,\tau_{k,\epsilon}+\epsilon).$$

Then each $\tau_{k,\epsilon}$ is a stopping time and eventually $\tau_{k,\epsilon}=t$ by continuity of paths. The continuity of paths also guarantees that $|X_{\tau_{k+1,\epsilon}}-X_{\tau_{k,\epsilon}}|\le\epsilon$. We have

$$X_t=\sum_{k\ge0}(X_{\tau_{k+1,\epsilon}}-X_{\tau_{k,\epsilon}}),\quad t=\sum_{k\ge0}(\tau_{k+1,\epsilon}-\tau_{k,\epsilon}).$$

To establish (**??**) we calculate the left hand side as

$$\lim_{n\to\infty}E[e^{\sum_{0\le k\le n}\lambda\,(X_{\tau_{k+1,\epsilon}}-X_{\tau_{k,\epsilon}})-\frac{\lambda^2}{2}(\tau_{k+1,\epsilon}-\tau_{k,\epsilon})}]$$

and show that it is equal to 1. Let us consider the σ-algebra $\mathcal{F}_k=\mathcal{B}_{\tau_{k,\epsilon}}$ and let

$$q_k(\omega)=E[e^{\lambda\,(X_{\tau_{k+1,\epsilon}}-X_{\tau_{k,\epsilon}})-(\frac{\lambda^2}{2}+\delta)(\tau_{k+1,\epsilon}-\tau_{k,\epsilon})}|\mathcal{F}_k]$$

where $\delta=\delta(\epsilon,\lambda)$ is to be chosen later such that $0\le\delta(\epsilon,\lambda)\le1$ and $\delta(\epsilon,\lambda)\to0$ as $\epsilon\to0^+$ for every fixed λ. If z and τ are random variables bounded by ϵ such that

$$E[z]=E[z^2-\tau]=0,$$

then for any $0 \leq \delta \leq 1$

$$E[e^{\lambda z-(\frac{\lambda^2}{2}+\delta)\tau}] \leq E[1+(\lambda z-(\frac{\lambda^2}{2}+\delta)\tau]+\frac{1}{2}(\lambda z-(\frac{\lambda^2}{2}+\delta)\tau)^2+C_\lambda(|z|^3+\lambda^3) \leq E[1-\delta\tau+C_\lambda\epsilon\tau] \leq 1$$

provided that $\delta = C_\lambda\epsilon$. Clearly there is a choice of $\delta(\epsilon,\lambda) \to 0$ as $\epsilon \to 0^+$ such that $q_k(\omega) \leq 1$ for every k and almost all ω. In particular, by induction

$$E[e^{\sum_{0\leq k\leq n}\lambda\,(X_{\tau_{k+1,\epsilon}}-X_{\tau_{k,\epsilon}})-(\frac{\lambda^2}{2}+\delta)(\tau_{k+1,\epsilon}-\tau_{k,\epsilon})}] \leq 1$$

for every n and by Fatou's lemma

$$E[e^{\lambda(X_t-X_0)-(\frac{\lambda^2}{2}+\delta)t}] \leq 1.$$

Since $\epsilon > 0$ is arbitrary we have proved one half of (**??**). To prove the other half, we note that $X_\lambda(t)$ is a submartingale and from Doob's martingale inequality we can get a tail estimate

$$P(\sup_{0\leq s\leq t}|X_t - X_0| \geq \ell) \leq 2\,e^{-\frac{\ell^2}{2t}}.$$

Since this allows us to use the dominated convergence theorem and establish

$$E[e^{\sum_{0\leq k\leq n}\lambda\,(X_{\tau_{k+1,\epsilon}}-X_{\tau_{k,\epsilon}})-(\frac{\lambda^2}{2}-\delta)(\tau_{k+1,\epsilon}-\tau_{k,\epsilon})}] \geq 1.\square$$

6.2 Random walks and Brownian Motion

Let ξ_k be a sequence of independent identically distributed random variables with mean 0 and variance 1 The partial sums S_k are defined by $S_0 = 0$ and $S_k = \xi_1 + \cdots + \xi_k$ for $1 \leq k \leq n$. We define stochastic processes $X_n(t),\ t \in [01]$ by

$$X_n(\frac{k}{n}) = \frac{S_k}{\sqrt{n}}$$

for $0 \leq k \leq n$ and for $t \in [\frac{k-1}{n}, \frac{k}{n}]$

$$X_n(t) = (nt-k+1)X(\frac{k}{n}) + (k-nt)X_n(\frac{k-1}{n}).$$

Let P_n denote the distribution of the process $X_n(t,\omega)$ on $C[0,1]$ and P the distribution of Brownian Motion. We want to explore the sense in which $\lim_{n\to\infty} P_n = P$.

<u>Lemma</u> For any finite collection $0 < t_1 < \cdots < t_m \leq 1$ of m sample points, the joint distribution of $(X(t_1), \cdots X(t_m))$ under P_n converges, as $n \to \infty$, to the corresponding distribution under P.

Proof: We are dealing here basically with the central limit theorem for sums of independent random variables. Let us define $k_n^i = [nt_i]$ and the increments

$$\xi_n^i = \frac{S_{k_n^i} - S_{k_n^{i-1}}}{\sqrt{n}}$$

for $i = 1, \cdots m$. For each n, ξ_n^i are m mutually independent random variables and their distributions converge as $n \to \infty$ to Gaussians with 0 means and variances $t_i - t_{i-1}$, respectively. This is of course the same distribution for these increments under Brownian Motion. The interpolation is of no consequence, because the difference between the end points is exactly some $\frac{\xi_i}{\sqrt{n}}$. So it does not really matter if in the definition of $X_n(t)$ we take $k_n = [nt]$ or $k_n = [nt]+1$ or take the interpolated value. We can state this convergence in the form

$$\lim_{n\to\infty} E[f(X_n(t_1), X_n(t_2), \cdots X_n(t_m))] = E[f(B_{t_1}, \cdots, B_{t_m})],$$

for every m, any sample points $(t_1, \cdots t_m)$ and any bounded continuous function f on R^m. Equivalently, for a simple random walk

$$p_k^{(n+1)} = \frac{1}{2}p_{k-1}^{(n)} + \frac{1}{2}p_{k+1}^{(n)},$$

or

$$\frac{p_k^{(n+1)} - p_k^{(n)}}{\Delta t} = \frac{1}{2}\frac{p_{k-1}^{(n)} - 2p_k^{(n)} + \frac{1}{2}p_{k+1}^{(n)}}{\Delta x^2},$$

where $\Delta t = \delta x^2$. Letting $\Delta t \to 0$ we obtain

$$\frac{\partial}{\partial t}p(t,x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}p(t,x).$$

6.3 Stochastic Integral with respect to Brownian motion

In this section we define the Ito's stochastic intergral. For a deterministic $f(t) \in L^2(0,T)$

$$\int_0^t f(s)dB_s = \lim_{\Delta t \to 0^+} \sum_j f_j(B_{t_{j+1}} - B_{t_j})$$

defines the Wienner integral. Since

$$\sum_j (f_{j+1}B_{t_{j+1}} - f_j B_{t_j}) = \sum_j f_j(B_{t_{j+1}} - B_{t_j}) + \sum_j (f_{j+1} - f_j)B_{t_{j+1}},$$

for $f \in BV(0,T)$

$$\int_0^t f(s)dB_s = f(t)B_t - f(0)B_0 - \int_0^t B_s\, df(s).$$

Given the filtration $\{\mathcal{F}_t\}$, assume the Brownian motion is $\mathcal{F}_t$-adapted.

Definition For an elementary stochastic process

$$f_t(\omega) = \sum_{k=0}^n f_{k-1}(\omega)\chi_{[t_{k-1},t_k)}, \quad f_{k-1} \text{ is a random variable,}$$

we define the stochastic integral

$$\int_0^t f_s(\omega)\, dB_s(\omega) = \sum_{k=0} f_{k-1}(\omega)(B_{t_k}(\omega) - B_{t_{k-1}}(\omega)).$$

It defines the Ito's sum and is the generalization of the ordinary concept of a Riemanntieltjes integral.

Examples (Stochastic integral) (1) Let $f_{k-1} = B_{t_{k-1}}$ we have

$$\sum_k B_{t_{k-1}}(B_{t_k} - B_{t_{k-1}}) = \sum_k \frac{1}{2}(|B_{t_k}|^2 - |B_{t_{k-1}}|^2 - |B_{t_k} - B_{t_{k-1}}|^2)$$

$$\to \frac{1}{2}(|B_t|^2 - |B_0|^2) - \frac{1}{2}t.$$

(2) For $f_{k-1} = \frac{B_{t_{k-1}}+B_{t_k}}{2}$

$$\sum_k \frac{B_{t_{k-1}} + B_{t_k}}{2}(B_{t_{k-1}} - B_{t_{k-1}}) = \sum_k \frac{1}{2}(B_{t_k}^2 - B_{t_{k-1}}^2) = \frac{1}{2}(|B_t|^2 - |B_0|^2)$$

(3) For $f_{k-1} = B_{t_k}$

$$\sum_k B_{t_k}(B_{t_{k-1}} - B_{t_{k-1}}) = \sum_k \frac{1}{2}(|B_{t_k}|^2 - |B_{t_{k-1}}|^2 + |B_{t_k} + B_{t_{k-1}}|^2) \to \frac{1}{2}(B_t^2 - B_0^2) + \frac{t}{2}.$$

Examples show that unlike a Riemann–Stieltjes integral, it is very essential to select f_{k-1} on $[t_{k-1}, t_k)$ for an approximating sequence for elementary stochastic processes of continuous process $f_t(\omega)$. We use the non-anticipated one, i.e., f_{k-1} is $\mathcal{F}_{t_{k-1}}$-measurable. Then, we have the following properties.

Lemma (Isometry) For $\mathcal{F}_t$-adapted elementary process $f_t(\omega)$

$$E[\int_0^t f_s\,dB_s] = 0, \quad E[|\int_0^t f_s\,dB_s|^2] = \int_0^t E[|f_s(\omega)|^2]\,ds. \text{ (Isometry).}$$

Proof: Let $\Delta B_{k-1} = B_{t_k} - B_{t_{k-1}}$.

$$E[\int_0^t f_s\,dB_s] = \sum_k E[f_{k-1}\Delta B_{k-1}] = 0$$

since f_{k-1} and ΔB_{k-1} are independent.

$$E|\int_0^t f_s\,dB_s|^2] = \sum_{k,j} E[f_{k-1}f_{\ell-1}\Delta B_{k-1}\Delta B_{\ell-1}] = \sum_k E[|f_{k-1}|^2](t_k - t_{k-1}) = \int_0^t E[|f_s(\omega)|^2]\,ds.$$

where we use

$$E[f_{k-1}f_{\ell-1}\Delta B_{k-1}\Delta B_{\ell-1}] = \begin{cases} 0 & \text{if } k \neq \ell \\ \\ E[f_{k-1}|^2](t_k - t_{k-1}) & \text{if } k = \ell, \end{cases}$$

since $f_{k-1}f_{\ell-1}\Delta B_{k-1}$ and $\Delta B_{\ell-1}$ are independent if $k < \ell$. □

Define the class $\mathcal{V}[0, T]$ satisfying

(a) f is adapted and measurable (the mapping $(t, \omega) \to f_t(\omega)$ is measurable on the product space $[0, T] \times \Omega$ with respect to the product σ-algebra $\mathcal{B}[0, T] \times \mathcal{F}$).
(b) $E[\int_0^T |f_t|^2\,dt] < \infty$.

It follows from the isometry that if $\{f^n\}$ be a approximating sequence of $f \in V[0, T]$, then

$$E[|I_t^n - I_t^m|^2] = \int_0^t E[|f_s^n - f_s^m|^2]\,ds$$

for

$$I_t^n = \int_0^t f_s^n\,dB_s.$$

where $\mathcal{V}(0, T)$ is Hilbert space with inner product

$$(f, g)_{\mathcal{V}(0,T)} = E[\int^T (f(t, \omega), g(t, \omega)\,dt].$$

Lemma (Density) For $f \in \mathcal{V}[0, T]$ there exists a sequence of $\mathcal{F}_t$-adapted elementary stochastic processes such that

$$\int_0^T E[|f_s^n - f_s|^2]\,ds \to 0 \text{ as } n \to \infty. \tag{6.6}$$

Proof: It suffices to prove the theorem for the bounded process since if we define $h_n(t,\omega)$ by

$$h_n(t,\omega) = \begin{cases} -n & \text{if } f(t,\omega) < -n \\ f(t,\omega) & \text{if } |f(t,\omega)| \le n \\ n & \text{if } f(t,\omega) > -n \end{cases}$$

then

$$\int_0^T E[|f_s(\omega) - h_n(s,\omega)|^2]\, ds \to 0 \text{ as } n \to \infty$$

by the Lebesgue dominated convergence theorem. First, suppose $t \to f_t(\omega)$ is bounded and continuous. We let

$$f_t^n = \sum_k f_{t_{k-1}}(\omega)\chi_{[t_{k-1},t_k)}.$$

Since $f_t^n(\omega)$ converges to $f_t(\omega)$ a.e., and by the bounded convergence theory (**??**) holds. Next, If f_t is bounded, and let $g_t(\omega) = \frac{1}{h}\int_{t-h}^t f_s(\omega)\, ds$, for $h > 0$, then

$$\int_0^T E|f_s - g_s|^2\, ds \to 0 \text{ as } h \to 0^+.$$

Since $t \to g_t(\omega)$ is bounded and continuous, (**??**) holds. □

By this lemma, we define:

Definition (Ito stochastic Integral) The Ito stochastic integral is defined by

$$\int_0^t f_s\, dB_s = \lim_{|P|\to 0} \sum_k f_{k-1}^n\,(B_{t_k} - B_{t_{k-1}})$$

for class $\mathcal{V}[0,T]$, i.e., $\{I_t^n\}$ is a Cauchy sequence in $L^2(\Omega, \mathcal{F}_t, P)$ has a unique limit. If $t \to f_t(\omega)$ is continuous,

$$\int_0^t f_s\, dB_s = \lim_{|P|\to 0} \sum_k f(t_{k-1},\omega)\,, (B_{t_k} - B_{t_{k-1}}).$$

Theorem (Ito stochastic Integral) The process

$$t \to I_t(\omega) = \int_0^t f_s(\omega)\, dB_s$$

satisfies (i)

$$E[I_t] = 0, \quad E[|I_t|^2] = \int_0^t E[|f_s|^2]\, ds.$$

(ii) $t \in I_t(\omega)$ is $\mathcal{F}_t$ martingale, and
(iii) $t \in I_t(\omega)$ is continuos a.s..

Proof: For an elementary stochastic process f_t

$$E[T_t|\mathcal{F}_s] = I_s + \sum_{t_k > s} E[f_{k-1}\Delta B_{k-1}|\mathcal{F}_{t_{k-1}]}\mathcal{F}_s].$$

Since

$$E[E[f_{k-1}\Delta B_{k-1}|\mathcal{F}_{t_{k-1}}]|\mathcal{F}_s] = f_{t_{k-1}} E[\Delta B_{k-1}] = 0$$

we have $E[I_t|\mathcal{F}_s] = I_s,\ t \ge s.$
For (iii) we have

$$P(\sup_{0\le t\le T} |I_t^n - I_t^m| \ge \delta) \le \frac{1}{\delta} E[|I_T^n - I_T^m|^2] = \frac{1}{\delta}\int_0^T E[|f^n - f^m|^2]\, ds$$

by the Doob's martingale inequality. Hence we may choose a subsequence n_k such that

$$P(\sup_{0\le t\le T}|I_t^{n_{k+1}}-I_t^{n_k}|\ge 2^{-k})<2^{-k}$$

By the Borel-Cantelli lemma

$$P(\sup_{0\le t\le T}|I_t^{n_{k+1}}-I_t^{n_k}|>2^{-k}\text{ for infinitely many }k)=0.$$

Thus, for a.s ω there exits $k_1(\omega)$ such thet

$$\sup_{0\le t\le T}|I_t^{n_{k+1}}-I_t^{n_k}|\le 2^{-k}\text{ for all }k\ge k_1(\omega).$$

and therefore $I_t^{n_k}(\omega)$ converges uniformly to I_t a.s. on $[0,T]$, i.e., for arbitrary $\epsilon>0$ there exists $N(\omega)$ such that for $n\ge N(\omega)$

$$\sup_{0\le t\le T}|I^n(\omega)-I_t(\omega)|\le\frac{\epsilon}{3}.$$

Since $t\to I_t^N(\omega)$ is continuous there exists $\delta(\omega)$ such that for $|t-s|\le\delta(\omega)$

$$|I_t(\omega)-I_s(\omega)|\le|I_t^N(\omega)-I_t(\omega)|+|I_t^N(\omega)-I_s^N(\omega)|+|I_s^N(\omega)-I_s(\omega)|\le\epsilon.\square$$

One can extend the Ito stochastic integral replacing property (b) by the weaker assumption: (b') $P(\int_0^t|f_t|^2<\infty)=1$.

We denote by $W(0,T)$ the space of processes that verify properties (a) and (b'). Stochastic integral is extended to the space $W(0,T)$ by means of a localization argument. Suppose that u belongs to $W(0,T)$. For each $n\ge 1$ we define the stopping time

$$\tau_n(\omega)=\inf\{t\ge 0:\int_0^t|f_s(\omega)|^2\,ds\ge n\}$$

where, by convention, $\tau_n=T$ if $\int_0^T|f_s|^2\,ds<n$. In this way we obtain a nondecreasing sequence of stopping times such that $\tau_n\uparrow T$. Furthermore,

$$t<\tau_n\Leftrightarrow\int_0^t|f_s|^2\,ds<n$$

Set $f_t^{(n)}=f_tI_{[0,\tau_n]}(t)$ and then $f^{(n)}\in V(0,T)$. For $m\ge n$, on the set $\{t\le\tau_n\}$

$$\int_0^t u_s^{(m)}\,dB_s=\int_0^t u_s^{(n)}\,dB_s,$$

since

$$\int_0^t f_s^{(n)}\,dB_s=\int_0^t f_s^{(m)}I_{[0,\tau_n]}\,dB_s=\int_0^{t\wedge\tau_n}f_s^{(m)}\,dB_s.$$

As a consequence, there exists an adapted and continuous process denoted by $\int_0^t f_s\,dB_s$ such that for any $n\ge 1$ and $t\le\tau_n$

$$\int_0^t f_s^{(n)}\,dB_s=\int_0^t f_s\,dB_s.$$

The stochastic integral of processes in the space $W(0,T)$ is linear and has continuous trajectories. $t\to I_t=\int_0^t f_s\,dB_s$ is a local martingale, i.e., $t\to\int_0^t f_s^{(n)}\,dB_s$ is a martingale. However, it may have infinite expectation and variance. Instead of the isometry property, there is a continuity property in probability by the following proposition:

Proposition 4 Suppose that $f \in W(0,T)$. For all $K,\ \delta > 0$ we have:

$$P(|\int_0^T f_s\,dB_s| \geq K) \leq P(\int_0^T |f_s|^2\,ds \geq \delta) + \frac{\delta}{K^2}.$$

Proof: Consider the stopping time defined by

$$\tau = \inf\{t \geq 0 : \int_0^t |f_s|^2\,ds \geq \delta\}$$

Then, we have

$$P(|\int_0^T f_s\,dB_s| \geq K) \leq P(\int_0^T |f_s|^2\,ds \geq \delta) + P(\{|\int_0^T f_s\,dB_s| \geq K|\} \cap \{\int_0^T |f_s|^2\,ds \leq \delta\})$$

where

$$P(\{|\int_0^T f_s\,dB_s| \geq K|\} \cap \{\int_0^T |f_s|^2\,ds \leq \delta\}) = P(\{|\int_0^T f_s\,dB_s| \geq K|\} \cap \{\tau = T\})$$

$$= P(\{|\int_0^\tau f_s\,dB_s| \geq K|\} \cap \{\tau = T\}) \leq \frac{1}{K^2}E[|\int_0^\tau f_s^2\,dB_s|^2]\frac{1}{K^2}E[\int_0^\tau |f_s|^2\,ds] \leq \frac{\delta}{K^2}.\Box$$

As a consequence of the above proposition, if $f^{(n)}$ is a sequence of processes in the space $W(0,T)$ which converges to $f \in W(0,T)$ in probability:

$$P(\int_0^T |f_s^{(n)} - f_s|^2\,ds > \epsilon) \to 0 \text{ as } n \to \infty$$

for all $\epsilon > 0$, then

$$\int_0^T f_s^{(n)}\,dB_s \to \int_0^T f_s\,dB_s \text{ in probability.}$$

6.4 Stratonovich integral

The Stratonovich stochastic integral

$$\int_0^T X_t \circ \mathrm{d}B_t : \Omega \to \mathbb{R}$$

is defined to be the limit in probability of

$$\sum_{i=0}^{k-1} X_{\frac{t_{i+1}+t_i}{2}}(B_{t_{i+1}} - B_{t_i})$$

as the mesh of the partition $P = \{0 = t_0 < t_1 < \cdots < t_k = T\}$ of $[0,T]$ tends to 0.
Examples (Stratonovich's stochastic integral) Since

$$\sum_j B_{\frac{t_{j+1}+t_j}{2}}(B_{t_{j+1}} - B_{t_j}) = \sum_j B_{\frac{t_{j+1}+t_j}{2}}(B_{t_{j+1}} - B_{\frac{t_{j+1}+t_j}{2}} + B_{\frac{t_{j+1}+t_j}{2}} - B_{t_j})$$

$$= \frac{1}{2}(|B_{t_{j+1}}|^2 - |B_{t_j}|^2) + \frac{1}{2}\sum_j(|(B_{t_{j+1}} - B_{\frac{t_{j+1}+t_j}{2}}|^2 - |B_{\frac{t_{j+1}+t_j}{2}} - B_{t_j}|^2) \to \frac{1}{2}(|B_t|^2 - |B_0|^2),$$

we have

$$\int_0^t B_s \circ dB_s = \frac{1}{2}(|B_t|^2 - |B_0|^2).$$

Conversion between Ito and Stratonovich integrals
The conversion is performed using the formula

$$\int_0^t f(B_s)\circ dB_s = \frac{1}{2}\int_0^t f'(B_s)\,ds + \int_0^t f(B_s)\,dB_s, \tag{6.7}$$

where f is a continuously differentiable function and the last integral is the Ito integral. In fact, we have

$$f(B_{t_{j+\frac{1}{2}}}) - f(B_{t_j}) = f'(B_{t_j})(B_{t_{j+\frac{1}{2}}} - B_{t_j}) + o(|B_{t_{j+\frac{1}{2}}} - B_{t_j}|)$$

and

$$E[\sum_j (f'(B_{t_j})(B_{t_{j+\frac{1}{2}}} - B_{t_j})(B_{t_{j+1}} - B_{t_j})] = E[\frac{1}{2}\sum_j f'(B_{t_j})\Delta t_j].$$

Thus,

$$\sum_j (f'(B_{t_j})(B_{t_{j+\frac{1}{2}}} - B_{t_j})(B_{t_{j+1}} - B_{t_j}) \to \frac{1}{2}\int_0^t f'(B_s)ds \text{ in } L^2(\Omega,\mathcal{F},P).$$

since

$$\sum_j E[|f'(B_{t_j})((B_{t_{j+\frac{1}{2}}} - B_{t_j})(B_{t_{j+1}} - B_{t_{j+\frac{1}{2}}}) - \frac{\Delta t_j}{2})|^2] = \frac{1}{4}\sum_j E[|f'(B_{t_j})|^2]\Delta t_j|^2 \to 0.$$

Stratonovich Calculus Stratonovich integrals are defined such that the chain rule of ordinary calculus holds, i.e.,

$$f(X_t) - f(X_0) = \int_0^t f'(X_s)\circ dX_s.$$

Since

$$\begin{aligned}
&f'(X_{t_{j+\frac{1}{2}}})(X_{t_{j+1}} - X_{t_j}) \\
&= f(X_{t_{j+1}}) - f(X_{t_j}) + f(X_{t_{j+1}}) - f(X_{t_{j+\frac{1}{2}}}) - f'(X_{t_{j+\frac{1}{2}}})(X_{t_{j+1}} - X_{t_{j+\frac{1}{2}}}) \\
&-(f(X_{t_j}) - f(X_{t_{j+\frac{1}{2}}}) - f'(X_{t_{j+\frac{1}{2}}})(X_{t_j} - X_{t_{j+\frac{1}{2}}}))
\end{aligned}$$

and

$$f(X_{t_{j+1}}) - f(X_{t_{j+\frac{1}{2}}}) - f'(X_{t_{j+\frac{1}{2}}})(X_{t_{j+1}} - X_{t_{j+\frac{1}{2}}}) = \tfrac{1}{2}f''(X_{t_{j+\frac{1}{2}}})(X_{t_{j+1}} - X_{t_{j+\frac{1}{2}}})^2 + O(|X_{t_{j+1}} - X_{t_{j+\frac{1}{2}}}|^3)$$

$$f(X_{t_j}) - f(X_{t_{j+\frac{1}{2}}}) - f'(X_{t_{j+\frac{1}{2}}})(X_{t_j} - X_{t_{j+\frac{1}{2}}}) = \tfrac{1}{2}f''(X_{t_j})(X_{t_{j+\frac{1}{2}}} - X_{t_j})^2 + O(|X_{t_{j+1}} - X_{t_{j+\frac{1}{2}}}|^3)$$

Thus, if X_t is an Ito's process

$$\sum_j f'(X_{t_{j+\frac{1}{2}}})(X_{t_{j+1}} - X_{t_j}) \to f(X_t) - f(X_0) \quad \text{in } L^2(0,T,P)$$

as $\Delta t \to 0^+$.

6.5 Martingale representation

Martingale representation Theory Let $\mathcal{F}_t = \sigma(B_s,\ s \le t)$. For every square integrable, continuous $\mathcal{F}_t$ martingale there exits a unique $f \in \mathcal{V}(0,T) = \{$square integrable $\mathcal{F}_t$ adapted process$on(0,T)$ such that

$$M_t = E[M_0] + \int_0^t f(s,\omega)\,dB_t(\omega).$$

Proof: Step 1 Let $\{h_k(t)\}$ is the orthonormal basis of $L^2(0,T)$. Define the exponential martingale

$$Y_k(t) = e^{\int_0^t h_k(s)\,dB_s - \frac{1}{2}\int_0^t |h_k(s)|^2\,ds}.$$

That is, if $dX_k = k_k\,dB_t - \frac{|h_k|^2}{2}\,dt$, then it follows from Ito's rule that $Y_k = e^{X_k(t)}$ satisfies

$$dY_k(t) = Y_k(t)(dX_k + \frac{1}{2}|h_k|^2\,dt) = h_k(t)Y_k(t)\,dB_t$$

and

$$Y_k(t) = 1 + \int_0^t h_k(t)Y_k(t)\,dB_t.$$

Since

$$d((Y_kY_j) = dY_kY_j + Y_kdY_j + h_kh_jY_kY_j\,dt,$$

we have

$$E[Y_k(t)Y_j(t)] = 1 + \int_0^t h_k(s)h_j(s)E[Y_k(s)Y_j(s)]\,ds$$

Thus,

$$E[Y_k(T)Y_j(T)] = e^{\int_0^T h_k(s)h_j(s)\,ds} = 1$$

and

$$E[(|Y_k(T) - 1|^2] = e - 1.$$

since $\{h_k(t)\}$ are an orthonormal basis in $L^2(0,T)$.
Step 2: Next, we prove that $\{\frac{1}{\sqrt{T}}, \frac{Y_k(t)-1}{\sqrt{e-1}},\ k \geq 1\}$ are dense in $L^2(\mathcal{F}_T, P)$. For $h(t) = \sum_k h_k\chi_{[t_{k-1},t_k)}(t)$

$$\int_0^T h\,dB_t = \sum_k h_k(B_{t_k} - B_{t_{k-1}}) = \sum_k \lambda_k\,B_{t_k}$$

Suppose $g \in L^2\mathcal{F}_T, P)$ satisfies

$$G(\lambda) = E[ge^{\sum_k \lambda_k\,B_{t_k}}] = 0$$

for all λ. For arbitrary $\phi \in C_0^\infty(R^n)$

$$E[\phi(B_{t_1},\cdots,B_{t_n})g] = E[((2\pi)^{-\frac{n}{2}}\int_{-\infty}^\infty e^{-i\sum y_kB_{t_k}}\hat{\phi}(y)\,dy)g]$$

$$= (2\pi)^{-\frac{n}{2}}\int_{-\infty}^\infty \hat{\phi}(y)E[ge^{-i\sum y_kB_{t_k}}]\,dy = 0,$$

where $\hat{\phi}$ is the Fourier transform of ϕ;

$$\hat{\phi}(y) = (2\pi)^{-\frac{n}{2}}\int_{-\infty}^\infty e^{i\sum x_ky_k}\phi(x)\,dx.$$

Since a family of random variables $\{\phi(B_{t_1},\cdots,B_{t_n}),\ 0 = t_0 < t_1 < \cdots t_n \leq T,\ \phi \in C_0^\infty(R^n)\}$ is dense in $L^2(\mathcal{F}_T, P)$, thus the span of random variables $\{e^{\int_0^T h(s)\,dB_s - \frac{1}{2}\int_0^T |h(s)|^2\,ds},\ h \in L^2(0,T)\}$ is dense in $L^2(\mathcal{F}_T, P)$.

Hence, $\{\dfrac{1}{\sqrt{T}}, \dfrac{Y_k(t)-1}{\sqrt{e-1}},\ k \geq 1\}$ are an orthonormal basis in $L^2(\Omega, \mathcal{F}_T, P)$ and for every $\mathcal{F}_T$ measurable random variable F has

$$\begin{aligned} F &= \alpha_0 + \sum_{k=1}^\infty \alpha_k(Y_k(t) - 1) = E[F] + \sum_{k=1}^\infty \int_0^T \alpha_kh_k(s)Y_k(s)dB_s \\ &= E[F] + \int_0^T f(s,\omega)\,dB_s, \end{aligned} \tag{6.8}$$

where

$$f(s,\omega)=\sum_{k=1}^\infty \alpha_k h_k(s)Y_k(s) \text{ with } \alpha_k=\frac{1}{\sqrt{e-1}}E[(Y_k(t)-1)F].$$

By the isometry

$$E[|F|^2]=E[|F_0|^2]+E[\int_0^T |f(s,\omega)|^2\,ds].$$

and the representation (**??**) is unique.
<u>Step 3</u> By Step 2 for $t_1\le t_2$

$$M_{t_1}=E[M_{t_2}|\mathcal{F}_{t_1}]=E[M_0]+E[\int_0^{t_2} f^{(t_2)}(s,\omega)\,dB_s|\mathcal{F}_{t_1}]$$

$$=E[M_0]+\int_0^{t_1} f^{(t_2)}(s,\omega)\,dB_s=E[M_0]+\int_0^{t_1} f^{(t_1)}(s,\omega)\,dB_s.$$

Thus,

$$0=E[|\int_0^{t_1}(f^{(t_1)}(s,\sigma)-f^{(t_2)}(s,\omega))\,dB_s|^2]=\int_0^{t_1}E[|f^{(t_1)}(s,\omega)-f^{(t_2)}(s,\omega))|^2]\,ds.$$

and $f^{(t_1)}(s,\omega)=f^{(t_2)}(s,\omega)=f(s,\omega)$ almost surely. □

6.6 Ito's stochastic calculus

In this section we discuss the Ito's stochastic calculus.

<u>Ito's Lemma</u> Let X_t be an Ito's process, i.e.,

$$X_t=X_0+\int_0^t u(s,\omega)+\int_0^t v(s,\omega)\,dB_s$$

where $\mathcal{F}_t$-adapted processes $u,\ v$ satisfy

$$P(|u(t,\omega)|<\infty)=1,\quad P(|vv^t(t,\omega)|<\infty)=1.$$

For $f\in C^{1,2}([0,T]\times R^d)$

$$f(X_t)-f(X_0)=\int_0^t\frac{\partial}{\partial t}f(s,X_s)\,ds+\sum_{i=1}^n\frac{\partial f}{\partial x_j}(s,X_s)dX_s^j+\frac{1}{2}\sum_{i,j=1}^n\int_0^t\frac{\partial^2 f}{\partial x_i\partial x_j}(s,X_s)(vv^t)_{i,j}(s,\omega)\,ds.$$

Or, equivalently (increment form)

$$df(t,X_t)=\frac{\partial f}{\partial t}\,dt+\sum_{j=1}^n\frac{\partial f}{\partial x_j}dX_t^j+\frac{1}{2}\sum_{i,j=1}^n\frac{\partial^2 f}{\partial x_i\partial x_j}(t,X_t)(vv^t)_{i,j}(t,\omega)\,dt. \tag{6.9}$$

Proof: For a positive integer n we define a stopping time τ_n by

$$\tau_n=\inf\{t>0: |X_0+\int_0^t v\,dB_s|>n \text{ or } |\int_0^t u\,ds|>n\}.$$

Then $\tau_n\to\infty$ as $n\to\infty$ a.s.. Thus, it suffices to prove the formula for $X_{t\wedge\tau_n}$ and thus without loss of generality we can assume that $|X_0+\int_0^t v\,dB_s|,\ |\int_0^t u\,ds|$ are bounded and $f,\ \frac{\partial f}{\partial x_j},\ \frac{\partial^2 f}{\partial x_i\partial x_j}$ are bounded and uniformly continuous. Note that by the mean value theorem

$$f(X_t)-f(X_0)=\sum_{k=1}^N[\frac{\partial f}{\partial t}(t_k,X_{t_{k-1}})(t_k-t_{k-1})+\sum_{i=0}^n\frac{\partial f}{\partial x_j}(X_{t_{k-1}})(X_{t_k}-X_{t_{k-1}})$$

$$+\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n\frac{\partial^2 f}{\partial x_i\partial x_j}(\xi_{i,j})(X^i_{t_k}-X^i_{t_{k-1}})(X^j_{t_k}-X^j_{t_{k-1}})].$$

By the definition of the Ito stochastic integral the second term of RHS converges to

$$\sum_{j=1}^{n}\int_0^t \frac{\partial f}{\partial x_j}(X_s)\,dX_s^j.$$

Note that

$$(X_{t_k}^i-X_{t_{k-1}}^i)(X_{t_k}^j-X_{t_{k-1}}^j)=(u_{k-1}\Delta t_{k-1}+v_{k-1}\Delta B_{k-1})^i(u_{k-1}\Delta t_{k-1}+v_{k-1}\Delta B_{k-1})^j,$$

where we assumed $u,\ v$ are elementary processes.

6.7 Tanaka's formula

In this section we discuss the Tanaka's formula for the one dimensional Brownian motion. Let

$$g_\epsilon(x)=\begin{cases} |x|, & |x|\geq\epsilon \\ \\ \dfrac{x^2}{2\epsilon}+\dfrac{\epsilon}{2} & |x|\leq\epsilon. \end{cases}$$

By the Ito formula

$$g_\epsilon(B_t)=g_\epsilon(B_0)+\int_0^t g_\epsilon'(B_s)dB_s+\frac{1}{2\epsilon}\,m(s\in[0,t]:B_s\in(-\epsilon,\epsilon))$$

Since

$$\int_0^t (g_\epsilon'(B_s)-sign_0(B_s))dB_s=\int_0^t \frac{B_s}{\epsilon}I\{|B_s|\leq\epsilon\}dB_s$$

and by the isometry

$$E[|\int_0^t \frac{B_s}{\epsilon}I\{|B_s|\leq\epsilon\}dB_s|^2]=\int_0^t\frac{1}{\sqrt{2\pi s}}\int_{-\epsilon}^{\epsilon}\frac{x}{\epsilon}^2e^{-\frac{x^2}{2s}}\,dxdt\to 0 \text{ as } \epsilon\to 0^+,$$

we obtain

$$|B_t|=|B_0|+\int_0^t sign_0(B_s)dB_s(\omega)+L_t(\omega)$$

where $L_t(\omega)=$ the local time for the Brownian motion is defined by

$$L_t(\omega)=\lim_{\epsilon\to 0^+}\frac{1}{2\epsilon}\,m(s\in[0,t]:B_s\in(-\epsilon,\epsilon)) \text{ in } L^2(\Omega,\mathcal{F},P).$$

6.8 Feynman-Kac formula

The Feynmanac formula, named after Richard Feynman and Mark Kac, establishes a link between parabolic partial differential equations (PDEs) and stochastic processes. It offers a method of solving certain PDEs by simulating random paths of a stochastic process. Conversely, an important class of expectations of random processes can be computed by deterministic methods.

Theorem (Feynman-Kac Formula) For $f\in C_0^2(R^n)$, $v\in C^{1,2}([0,T],R^n)$ satisfies

$$\frac{\partial v}{\partial t}+\mathcal{A}v-q(x)v=0,\quad v(T,x)=f(x) \tag{6.10}$$

if and only if

$$v(t,x)=E^{t,x}[e^{-\int_t^T q(X_s)\,ds}f(X_T)]$$

Proof: Let

$$Z_s=e^{-\int_t^s q(X_s)\,ds}$$

and then

$$dZ_s = -q(X_s) Z_s \, ds.$$

By the Ito's formula

$$d(Z_s v(s, X_s)) = v \, dZ_s + Z_s(\frac{\partial v}{\partial s} + \mathcal{A}v) + b \cdot (\nabla v(s, X_t)\sigma(X_s) dB_s)$$

Thus,

$$E^{t,x}[Z_T v(T, X_T)] = v(t, x) + E^{t,x}[\int_t^T (Z_s(\frac{\partial v}{\partial s} + \mathcal{A}v - qv) \, ds]$$

which implies the representation. □

<u>Example</u> Note that (**??**) is backward in time with terminal condition at $t = T$. Let $u(t, x) = v(T - t, x)$. Then (**??**) is equivalent to the initial value problem for u

$$\frac{\partial u}{\partial t} = \mathcal{A}u - q(x)u, \quad u(0, x) = f(x).$$

Let $X_t = x + \sigma \, B_t$. Then

$$\frac{\partial u}{\partial t} = \frac{\sigma^2}{2} \Delta u - q(x)u, \quad u(0, x) = f(x).$$

has the solution representation

$$u(t, x) = E[e^{-\int_0^t q(X_t) \, dt} f(X_t)] = E[e^{-\int_0^t q(X_t) \, dt}] \frac{1}{(\sqrt{2\pi t}\sigma)^n} \int_{R^n} f(y) e^{-\frac{|y-x|^2}{2\sigma^2 t}} \, dy$$

6.9 Excises

<u>Problem 1</u> Show (**??**)
<u>Problem 2</u> Let B_t be a two-dimensional Brownian motion. Given $\rho > 0$, compute $P(|B_t| < \rho)$.
<u>Problem 3</u> Compute the mean and covariance of the geometric Brownian motion. Is it a Gaussian process?
<u>Problem 4</u> Let B_t be a Brownian motion. Find the law of B_t conditioned by B_{t_1}, B_{t_2} , and $(B_{t_1}; B_{t_2})$ assuming $t_1 < t < t_2$.
<u>Problem 5</u> Check if the following processes are martingales,

$$e^{\lambda B_t - \frac{\lambda^2 t}{2}}, \;\; e^{t/2} \cos(B_t), \;\; (B_t + t)e^{-B_t - \frac{t}{2}}, \; B_1(t)B_2(t)B_3(t)$$

where B_1, B_2 and B_3 are independent Brownian motions.
<u>Problem 6</u> Derive the formula

$$v(t, x) = E^{x,t}[f(X_T) + \int_t^T g(X_s) \, ds]$$

for $v_t + \mathcal{A}v + g(x) = 0$.
<u>Problem 7</u> Show that the solution to boundary value problem

$$-\mathcal{A}u = f, \quad u = g \text{ at the boundary } \partial\Omega$$

we have the solution representation

$$u(x) = E^x[g(X_{\tau_\Omega})] + E^x[\int^{\tau_\Omega} f(X_s) \, ds]$$

7 Diffusion Process

When we model a stochastic process in the continuous time it is almost impossible to specify in some reasonable manner a consistent set of finite dimensional distributions in general. The one exception is the family of Gaussian processes with specified means and covariances. It is much more natural and beneficial to take an evolutionary approach. For simplicity let us consider the one dimensional case where we are trying to define a real valued stochastic process X_t with continuous trajectories. The space $C[0,T]$ is the space on which we wish to construct the measure P. We have the σ-fields $\mathcal{F}_t = \sigma(X_s,\ 0 \le s \le i)$ defined for $t \le T$. Let $\mathcal{F} = \mathcal{F}_T$, the total σ-field. We try to specify the measure P by specifying the conditional distributions $P[X_{t+h} - X_t \in A | \mathcal{F}_t]$ of process $\{X_t\}$ by the infinitesimal evolution of their mean and variance:

$$
\begin{aligned}
E[X_{t+h} - X_t | \mathcal{F}_t] &= h\, b(t,\omega) + o(h) \\
E[|X_{t+h} - X_t|^2 | \mathcal{F}_t] &= h\, a(t,\omega) + o(h),
\end{aligned}
\tag{7.1}
$$

where for each $t \le T$ the drift $b(t,\omega)$ and the variance $a(t,\omega)$ are $\mathcal{F}_t$-measurable functions. Since we insist on continuity of paths, this will force the distributions to be nearly Gaussian and no additional specification should be necessary. Equations (**??**) are infinitesimal differential relations and the integrated forms are precise mathematical statements. We will discuss the approach by K. Ito that realizes the increments $X_{t+h} - X_t$ as

$$X_{t+h} - X_t \sim b(t, X_t)h + \sqrt{a(t,X_t)}(B_{t+h} - B_t)$$

and as $h \to 0$ X_t defines a solution to the stochastic differential equation

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

We need some definitions.

Definition (Progressively measurable) We say that a function $f : [0,T] \times \Omega \to R$ is progressively measurable if, for every $t \in [0,T]$ the restriction of f to $[0,t] \times \Omega$ is a measurable function of t and ω on $([0,t] \times \Omega, \mathcal{B}[0,t] \times \mathcal{F}_t)$.

The condition is somewhat stronger than just demanding that for each t, $f(t,\omega)$ is $\mathcal{F}_t$ is measurable. The following facts hold.
(1) If $f(t,x)$ is measurable function of t and x, then $f(t, X_t(\omega))$ is progressively measurable.
(2) If $f(t,\omega)$ is either left continuous (or right continuous) as function of t for every ω and if in addition $f(t,\omega)$ is $\mathcal{F}_t$ measurable for every t, then f is progressively measurable.
(3) There is a sub σ-field $\Sigma \subset \mathcal{B}[0,T] \times \mathcal{F}$ such that progressive measurability is just measurability with respect to Σ. In particular standard operations performed on progressively measurable functions yield progressively measurable functions.

We shall always assume that the functions $b(t,\omega)$ and $a(t,\omega)$ be progressively measurable. Let us suppose in addition that they are bounded functions. The boundedness will be relaxed at a later stage. We reformulate conditions (**??**) as

$$M_1(t) = X_t - X_0 - \int_0^t b(s,\omega)\, ds, \quad \text{and} \quad M_2(t) = M_1(t)^2 - \int_0^t a(s,\omega)\, ds$$

are martingales with respect to $(\Omega, \mathcal{F}_t, P)$. We can define a diffusion process corresponding to a, b as a probability measure P on $(\Omega, \mathcal{F})$ such that relative to $(\Omega, \mathcal{F}_t, P)$ $M_1(t)$ and $M_2(t)$ are martingales. If in addition we are given a probability measure μ as the initial distribution, i.e. $\mu(A) = P(X_0 \in A)$, then we can expect that P is determined by a, b and μ. We have seen that if $a = 1$ and $b = 0$, with $\mu = \delta_0$, we obtain the standard Brownian motion B_t. If $a = a(t, X_t)$ and $b = b(t, X_t)$, we expect that X_t is a Markov process, because the infinitesimal parameters depend only on the current position and not on the past history. If there is no explicit dependence on time, then the Markov process have the stationary transition probabilities. Finally, if $a(t,\omega) = a(t)$ is a deterministic function in t and $b(t,\omega) = b_1(t) + c(t)X_t$, then P^x is Gaussian, if μ is so.

Since X_t are continuous we can establish that

$$Z_\lambda(t) = e^{\lambda\, M_1(t) - \frac{\lambda^2}{2}\int_0^t a(s,\omega)\, ds} = e^{\lambda(X_t - X_0 - \int_0^t b(s,\omega)\, ds) - \frac{\lambda^2}{2}\int_0^t a(s,\omega)\, ds}$$

is a martingale with respect to $(\Omega, \mathcal{F}_t, P)$ for every real λ. We can also use for the definition of a diffusion process corresponding to a, b the condition that $Z_\lambda(t)$ be a martingale with respect to $(\Omega, \mathcal{F}_t, P)$ for every λ. If so, we do not have to assume that the paths were almost surely continuous. $(\Omega, \mathcal{F}_t, P)$ could be any probability space on which a stochastic process X_t such that $Z_\lambda(t)$ is a martingale for all λ. If C is an upper bound for a, it is easy to see that

$$E[e^{\lambda(M_1(t) - M_1(s))}] \le e^{\frac{C\lambda^2}{2}}.$$

The lemma of Garsia-Rodemich-Rumsey will guarantee that the paths can be chosen to be continuous.

In general, let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\mathcal{T}$ be the interval $[0,T]$ for some finite T or the infinite interval $[0,\infty)$ $\mathcal{F}_t,\ t \in \mathcal{T}$ be sub σ-algebras such that $\mathcal{F}_s \subset \mathcal{F}_t$ for $s < t$. We can assume without loss of generality that $\mathcal{F} = \bigcup_{t \in T} \mathcal{F}_t$. Let a stochastic process X_t with values in R^n be given. Assume that it is progressively measurable with respect to $(\Omega, \mathcal{F}_t)$. We generalize the definitions described in the above to diffusion processes with values in R^n. Given a positive semidefinite $n \times n$ matrix function $a = a_{i,j}$ and an n-vector $b = b_j$ function, we define the operator

$$(\mathcal{L}_{a,b} f)(x) = \frac{1}{2}\sum_{i,j} a_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} f + \sum_j b_j \frac{\partial}{\partial x_j} f.$$

If $a = a_{i,j}(t,\omega)$ and $b = b_j(t,\omega)$ are progressively measurable functions, we define

$$(L_{t,\omega} f)(x) = (\mathcal{L}_{a(t,\omega), b(t,\omega)} f)(x).$$

Then, we have the equivalent characterizations of the diffusion process:

Theorem 2 (Diffusion Process) The following definitions are equivalent. X_t is a diffusion process corresponding to bounded progressively measurable functions $a(t,\omega)$, $b(t,\omega)$ that take with values in the space of symmetric positive semidefinite $n \times n$ matrices, and n-vectors if

(1) X_t has an almost surely continuous version and

$$Y(t) = X_t - X_0 - \int_0^t b(s,\omega)\, ds, \quad Z_{i,j}(t) = Y_i(t,\omega) Y_j(t,\omega) - \int_0^t a_{i,j}(s,\omega)\, ds$$

are $(\Omega, \mathcal{F}_t, P)$ martingales.

(2) For every $\lambda \in R^n$

$$Z_\lambda(t,\omega) = e^{(\lambda, Y(t,\omega)) - \frac{1}{2}\int_0^t (\lambda, a(s,\omega)\lambda)\, ds} \quad \text{is an } (\Omega, \mathcal{F}_t, P) \text{ martingale.}$$

(3) For every $\lambda \in R^n$

$$X_\lambda(t,\omega) = e^{i(\lambda, Y(t,\omega)) + \frac{1}{2}\int_0^t (\lambda, a(s,\omega)\lambda)\, ds} \quad \text{is an } (\Omega, \mathcal{F}_t, P) \text{ martingale.}$$

(4) For every smooth bounded function f on R^n with two bounded continuous derivatives

$$f(X_t) - f(X_0) - \int_0^t (L_{s,\omega} f(X_s))\, ds \quad \text{is an } (\Omega, \mathcal{F}_t, P) \text{ martingale.}$$

(5) For every smooth bounded function ϕ on $T \times R^n$ with at least two bounded continuous x derivatives and one bounded continuous t derivative

$$\phi(t, X_t) - \phi(0, X_0) - \int_0^t (\frac{\partial}{\partial t} + L_{s,\omega})\phi(s, X_s))\, ds \quad \text{is an } (\Omega, \mathcal{F}_t, P) \text{ martingale.}$$

(6) For every smooth bounded function $\phi \in C^{1,2}(\mathcal{T}, R^n)$

$$\exp\left(\phi(t,X_t)-\phi(0,X_0)-\int_0^t(\frac{\partial}{\partial t}+L_{s,\omega}\phi(s,X_s))\,ds-\frac{1}{2}\int_0^t(\nabla_x\phi(s,X_s),a(s,\omega)\nabla_x\phi(s,X_s))\,ds\right)$$

is an $(\Omega, \mathcal{F}_t, P)$ martingale.
(7) Same as (6) except that ϕ is replaced by ψ of the form $\psi(t,x)=(\lambda,x)+\phi(t,x)$ where is ϕ as in (6) and $\lambda \in R^n$ is arbitrary.

Under any one of the above definitions, $Y(t,\omega)$ has an almost surely continuous version satisfying

$$P(\sup_{0\le s\le t}|Y(s,\omega)-Y(0,\omega)|\ge \ell)\le 2n\,e^{-\frac{\ell^2}{Ct}}.$$

for some constant C depending only on the dimension n and the upper bound for a.

Proof: (3) Since

$$dZ_\lambda(t)=((\lambda,dY_t)-\frac{1}{2}(\lambda,a\,\lambda)\,dt)Z_\lambda(t)+\frac{1}{2}(\lambda,a\,\lambda)Z_\lambda(t)\,dt=(\lambda,dY_t)Z_\lambda(t),$$

we have

$$Z_\lambda(t)-Z_\lambda(s)=\int_s^t Z_\lambda(\sigma)(\lambda,dY_\sigma)$$

and thus Z_t ia a martingale.
(4) Let us apply the above lemma with $M_t = X_\lambda(t)$ and

$$A_t=e^{\int_0^t i(\lambda,b_s)-\frac{1}{2}(\lambda,a_s\,\lambda)\,ds}$$

Then a simple computation yields

$$M_tA_t-M_0A_0-\int_0^t M_s dA_s=e_\lambda(X_t-X_0)-1-\int_0^t(\mathcal{L}_{s,\omega}e_\lambda)(X_s-X_0)\,ds,$$

where $e_\lambda(x)=e^{i(\lambda,x)}$. Multiplying this by $e_\lambda(X_0)$, which is essentially a constant, we conclude that

$$e_\lambda(X_t)-e_\lambda(X_0)-\int_0^t(\mathcal{L}_{s,\omega}e_\lambda)(X_s)\,ds$$

is a martingale. That is,

$$E[e^{i(\lambda,X_t-X_s}|\mathcal{F}_s]=\int_s^t E[(-i(\lambda,b(\sigma))+(\lambda,a(\sigma)\lambda))e^{i(\lambda,X_\sigma-X_s)}|\mathcal{F}_s]\,d\sigma.$$

If b and a are deterministic

$$E[e^{i(\lambda,X_t-X_s)}|\mathcal{F}_s]=e^{-\int_s^t i(\lambda,b(\sigma))+(\lambda,a(\sigma)\lambda)\,d\sigma}$$

and X_t is a Gaussian process if X_0 is so.
(5) Note that

$$E[\phi(t,X_t)-\phi(s,X_s)|\mathcal{F}_s]=E[\phi(t,X_t)-\phi(t,X_s)|\mathcal{F}_s]+E[\phi(t,X_s)-\phi(s,X_s)|\mathcal{F}_s]$$

$$=E[\int_s^t\mathcal{L}_{\sigma,\omega}\phi(\sigma,X_\sigma)\,d\sigma|\mathcal{F}_s]+E[\int_s^t\frac{\partial}{\partial t}\phi(\sigma,X_s)\,d\sigma|\mathcal{F}_s]$$

$$=E[\int_s^t(\frac{\partial}{\partial t}+\mathcal{L}_{\sigma,\omega})\phi(\sigma,X_\sigma)\,d\sigma|\mathcal{F}_s]+J$$

where

$$J = E[\int_s^t \mathcal{L}_{\sigma,\omega}(\phi(t, X-\sigma) - \phi(\sigma, X_\sigma)\, d\sigma|\mathcal{F}_t) + \int_s^t (\frac{\partial}{\partial t}\phi(\sigma, X_s) - \frac{\partial}{\partial t}\phi(\sigma, X_\sigma)\ d\sigma|\mathcal{F}_s]$$

$$= E[\int_s^t \int_u^t (\frac{\partial}{\partial t}\phi(v, X_u)\mathcal{L}_{u,\omega}\phi(v, x_u)\, dudv|\mathcal{F}_s)] - E[\int_s^t \int_s^u (\mathcal{L}_{v,\omega}\frac{\partial}{\partial t}\phi(u, X_v)\, dudv|\mathcal{F}_s]$$

$$= E[\int\int_{s\le u\le v\le t} \mathcal{L}_{u,\omega}\frac{\partial}{\partial t}\phi(v, X_u)\, dudv - \int\int_{s\le v\le u\le t} (\mathcal{L}_{v,\omega}\frac{\partial}{\partial t}\phi(u, X_v)\, dudv|\mathcal{F}_s)] = 0$$

where we used the fact that the last two integrals are symmetric with respect to (u, v).

7.1 Excises

Problem 1 Show that

$$M_t = u(t, X_t) - u(0, X_0) - \int_0^t (\frac{\partial}{\partial t} + \mathcal{L})u(s, X_s)\, ds$$

is a $\mathcal{F}_t$ martingale. If we assume

$$\frac{\partial u}{\partial t} + \mathcal{L}u(t, x) = 0, \quad u(T, x) = f(x)$$

then show that $u(t, x) = E^{t,x}[f(X_T)] = E[f(X_T)|X_t = x]$.
Problem 2 Show that

$$M_t = e^{-\int_0^t q(X_s)\, ds}u(t, X_t) - u(0, X_0) - \int_0^t e^{-\int_0^s q(X_\sigma)\, d\sigma}(\frac{\partial}{\partial t} + \mathcal{L} - q(X_s))u(s, X_s))\, ds$$

is a $\mathcal{F}_t$ martingale. If we assume

$$\frac{\partial u}{\partial t} + \mathcal{L}u(t, x) - q(x)u(t, x) = 0, \quad u(T, x) = f(x),$$

then show that $u(t, x) = E^{t,x}(e^{-\int_t^T q(X_s)\, ds}f(X_T))$ (Feynman-Kac formula).
Problem 3 Let

$$X_t = e^{r\, t + \sigma\, B_t - \frac{\sigma^2 t}{2}}$$

Show that

$$dX_t = rX_t\, dt + \sigma X_t\, dB_t$$

and

$$\mathcal{L}f = rx\, f' + \frac{\sigma^2 x^2}{2} f''.$$

If

$$u(t, x) = E^{t,x}(e^{-r(T-t)}f(X_T)),$$

then show that u satisfies Black-Scholes equation

$$\frac{\partial u}{\partial t} + rx\,\frac{\partial u}{\partial x} + \frac{\sigma^2 x^2}{2}\frac{\partial^2 u}{\partial x^2} - r\, u = 0, \quad u(T, x) = \max(0, K - x) = f(x).$$

(for the European call option)

8 Stochastic Differential equation

A stochastic differential equation models physical processes driven by random forces and random rate change and uncertainty in models and initial conditions. It has a wide class of applications in multidisplined sciences and engineering and provides a mathematical tool to analyze concrete stochastic dynamics and apply and develop probabilistic methods.

For example, consider the population growth mode

$$\frac{dN}{dt} = a(t)N(t) + f(t), \quad N(0) = N_0, \tag{8.1}$$

where $N(t)$ is the size of population at time t, $a(t)$ is the growth rate and $f(t)$ is the generation rate of the population at time t. We introduce the random environmental effects through;

$$a(t) = r(t) + \text{"}noise\text{"}, \quad f(t) = F(t) + \text{"}noise\text{"}$$

and random initial value N_0. In different applications (**??**) can be used to model the chemical concentration and the mathematical finance for example. We will formulate the random equations as the Ito's stochastic differential equation and develop the solution methods.

Consider the discrete dynamics for $X_k,\ k \geq 0$

$$X_{k+1} = X_k + b(X_k)\,\Delta t + \sigma(X_k)w_k, \quad X_0 = x \tag{8.2}$$

where $b(x)$ is the drift, $\sigma(x)$ is the standard deviation, and $w_k = B_{t_{k+1}} - B_{t_k}$ is independent, identically distributed Gaussian random variables with $N(0, \sqrt{\Delta t})$. Equivalently, we have

$$X_n = x + \sum_{k=1}^{n} b(X_{k-1})\Delta t + \sum_{0}^{n} \sigma(X_k)(B_{t_k} - B_{t_{k-1}}).$$

We will analyze the limit as the time-stepsize $\Delta \to 0$, i.e., let

$$X_t^{\Delta t} = X_n \text{ on } [t_n, t_{n+1})$$

then $\{X_t^{\Delta t}\}$ converges to the weak solution to the stochastic differential equations.

First, we establish the existence of the strong solution to the stochastic differential equation

$$dX_t = b(t, X_t) + \sigma(t, X_t)dB_t$$

under the conditions
H1) (Lipschitz)

$$|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq D\,|x - y|$$

H2) (Linear Growth)

$$|b(t,x))| + |\sigma(t,x)| \leq C(1 + |x|).$$

<u>**Ito's Lemma**</u> Let a square integrable random variable X_0 and $\mathcal{F}_t$-Brownian motion $B_t,\ t \geq 0$ be given and assume they are independent. Under conditions H1) and H2) there exists a unique almost surely continuous measurable processes$X_t(\omega)$ that satisfies

$$X_t(\omega) = X_0(\omega) + \int_0^t b(s, X_s(\omega))\,ds + \int_0^t \sigma(s, X_s(\omega))dB_s(\omega). \tag{8.3}$$

Proof: (Uniqueness) Suppose $X_t,\ \hat{X}_t$ be two solutions. Then, we have

$$E[|X_t - \hat{X}_t|^2] = E[|X_0 - \hat{X}_0 \int_0^t (b(s, X_s) - b(s, \hat{X}_s)\,ds + \int_0^t (\sigma(s, X_s) - \sigma(s, \hat{X}_s)dB_s|^2)]$$

$$\leq 3E[|(X_0 - \hat{X}_0|^2] + 3(1+t)D^2E[\int_0^t |X_s - \hat{X}_s|^2\,ds]$$

By Gronwall's inequality

$$E[|X_t - \hat{X}_t|^2] \le 3E[|(X_0 - \hat{X}_0|^2]e^{3D^2 t(1+\frac{t}{2}}.$$

(Existence) Consider the fixed point iterate

$$X_t^{k+1} = \Phi(t, X_t^k) \quad \text{with } X_t^0 = X_0$$

and

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

Then,

$$E[|X_t^{k+1} - X_t^k|^2] \le (1+t)D^2 \int_0^t E[|X_s^k - X_s^{k-1}|^2]\, ds$$

and

$$E[|X_t^1 - X_t^0|^2] \le 2C^2 t(1 + E[|X_0|^2])$$

By induction in k we have

$$E[|X_t^k - X_t^{k-1}|^2] \le \frac{A^k t^k}{k!} \tag{8.4}$$

on $t \in [0,T]$. Thus, $\{X_t^k\}$ is Cauchy a sequence in $L^2(\Omega, \mathcal{F}_t, P)$ has a unique limit $X_t(\omega) = \lim_{k\to\infty} X_t^k(\omega)$ uniformly on $[0,T]$. By the martingale inequality

$$\sup_{0\le s\le T} P(|X_t^{k+1} - X_t^k| \ge 2^{-k}) \le P(\int_0^T |b(s, X_s^{k+1}) - b(s, X^k)|^2 \ge 2^{-2k-2})$$

$$+2^{k+1} E[\int_0^T |\sigma(s, X_s^{k+1}) - \sigma(s, X^k)|^2]\, ds.$$

From (**??**) and by Borel-Cantelli lemma $X_t(\omega) = \lim_{k\to\infty} X_t^k(\omega)$ a.s., uniformly on $[0,T]$. □

Theorem (Ito's Rule) Let X_t be a Ito's diffusion process, i.e. X_t satisfies

$$dX_t = b(X_t)\, dt + \sigma(X_t) dB_t.$$

where b σ are Lipschitz continuous. Then,

$$f(x_t) = f(X_0) + \int_0^t \nabla f(X_s) \cdot (b(X_s)\, ds + \sigma(X_s) dB_s) + \int_0^t \frac{1}{2} a_{i,j}(X_s)(\frac{\partial^2 f}{\partial x_i \partial x_j})(X_s)\, ds,$$

where $a(x) = \sigma\sigma^t$. $\{X_t,\ t \ge 0\}$ is a Markov process and

$$f(X_t) - f(X_0) - \int_0^t \mathcal{A} f(X_s)\, ds$$

is an $\mathcal{F}_t$-martingale. The generator $\mathcal{A}$ of $\{X_t\}$ is given by

$$\mathcal{A} f = \sum_j b_j(x)(\frac{\partial f}{\partial x_j})(x) + \frac{1}{2} a_{i,j}(x)(\frac{\partial^2 f}{\partial x_i \partial x_j})(x).$$

with $dom(\mathcal{A}) = C_0^2(R^n)$.

Crollary

$$df(t, B_t) = (\frac{\partial f}{\partial t} + \frac{1}{2}\Delta f)(B_t)\, dt + \nabla f(B_t) \cdot dB_t$$

and thus $f(t, B_t)$ is a martingale if and only if $\frac{\partial f}{\partial t} + \frac{1}{2}\Delta f = 0$.

8.1 Weak solution

The solution $X_t(\omega)$ to (**??**) under Ito's condition is a strong solution in the sense that the sample path of Brownian motion $t \to B_t(\omega)$ is given in advance and then the sample solution is constructed as an $\mathcal{F}_t$ adapted stochastic process.

Definition (Weak Solution) A pair of stochastic processes $(\tilde{X}_t, \tilde{B}_t, \mathcal{H}_t)$ is a weak solution to (**??**):

$$\tilde{X}_t = X_0 + \int_0^t b(s, \tilde{X}_s)\, ds + \int_0^t \sigma(s, \tilde{X}_s)\, d\tilde{B}_s,$$

if $\mathcal{H}_t$ is an increasing family of σ-algebras, $\tilde{X}_t$ is $\mathcal{H}_t$-adapted and $\tilde{B}_t$ is an $\mathcal{H}_t$-Brownian motion.

The uniqueness of the weak solution implies that any two solutions are identical in law, i.e., have the same finite dimensional distributions. From the point of view of stochastic modelings the weak solution concept is more natural since we do not need to specify the Brownian motion sample beforehand. As will be shown next, there are stochastic differential equations that have no strong solution but a unique weak solution.

Example(Weak Solution) Consider the stochastic differential equation

$$dX_t = sign(X_t)\, dB_t$$

We show that it does not have a strong solution. Suppose it has a strong solution X_t, i.e.,

$$X_t = \int_0^t sign(X_s)\, dB_s \tag{8.5}$$

It will be shown that since $|sign(X_t)| = 1$, X_t is a Brownian motion. Since $dB_t = sign(X_t)\, dX_t$ we have

$$B_t = \int_0^t sign(X_s)\, dX_s$$

It follows from the Tanaka's formula

$$B_t = |X_t| - |X_0| - L_t(\omega)$$

where $L_t(\omega)$ is the local time of the Brownian motion X_t. Thus, if $\mathcal{H}_t$ is σ-algebra generated by $\{|X_s|,\ s \leq t\}$, then B_t is $\mathcal{H}_t$-measurable. But, since $\mathcal{H}_t$ is strictly contained in $\mathcal{F}_t$, the σ-algebra generated by $\{B_s,\ s \leq t\}$, it yields a contradiction.

A weak solution to (**??**) is any Brownian motion X_t. In fact

$$\tilde{B}_t = \int_0^t sign(X_t)\, dX_t$$

is a Brownian motion. Since $d\tilde{B}_t = sign(X_t) dX_t$, we have $dX_t = sign(X_t) d\tilde{B}_t$.

Yamada and Watanabe Theorem Suppose SDE has a weak solution and the solution is path wise unique. The equation has s strong solution.

Yamada and Watanabe Theorem II Assume b and σ satisfy

$$|b(t, x_1) - b(t_x 2)| + |\sigma(t, x_1) - \sigma(t, x_2)| \rho(|x_1 - x_2|$$

where ρ is continuous, $\rho(0) = 0$ increasing, convex and

$$\int^{\epsilon} \frac{1}{\rho(s)}\, ds = \infty$$

for some $\epsilon > 0$. Then, SDE has unique strong solution.

8.2 Markov Property of Ito's diffusion

Consider the SDE

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t$$

with Lipschitz continuous $b,\ \sigma$. It follows from Ito's Lemma that there exists a unique strong solution $X_t^{0,x}$ with $X_0 = x$ and

$$X_t^{0,x} = \Phi(t, x, \{B_s,\ s \le t\}). \tag{8.6}$$

is a unique pathwise solution that depends on the initial condition and the Brownian motion path. For $h > 0$

$$X_{t+h}^{0,x} = X_t^{0,x} + \int_t^{t+h} b(X_s^{0,x})\,ds + \int_t^{t+h} b(X_s^{0,x})\,dB_s$$

Let $\tilde{X}_v = X_{t+v}^{0,x}$, $\tilde{B}_v = B_{t+v} - B_t$ and $s = t + v$. Then

$$\tilde{X}_h = X_t^{0,x} + \int_0^h b(\tilde{X}_v)\,ds \int_0^h \sigma(\tilde{X}_v)\,d\tilde{B}_v,$$

i.e.,

$$\tilde{X}_h = \Phi(h, X_t^{0,x}, \{\tilde{B}_v,\ v \le h\}). \tag{8.7}$$

Define

$$E^x[g(X_t)] = E[g(X_t^{0,x}].$$

for bounded continuous function g.

Theorem (Markov property) If $\{X_t,\ t \ge 0\}$ is an Ito diffusion process, then for $h \ge 0$

$$E^x[g(X_{t+h}|calF_t] = E^{X_t(\omega}[g(X_h)].$$

Proof It follows from (**??**)–(**??**) that

$$E[X_{t+h}^{0,x}|\mathcal{F}_t] = E[X_h^{0,x}|x = X_t^{0,x}].\square$$

Define a linear operator H_t by

$$(H_t g)(x) = E^x[g(X_t)], \quad g \in X = C(R^n).$$

Then, we have **Corollary** For $t,\ s \ge 0$, $H_{t+s} = H_t H_s$.

Proof It follows from the Markov property that

$$H_{t+s}g = E^x[g(X_{t+s})] = E[E[g(X_{t+s}^{0,x}|\mathcal{F}_t] = E[E^{X_t^{0,x}}g(X_s)]] = E^x[(H_t g)(X_s)] = H_s(H_t g).\square$$

Theorem H_t is a strongly continuous semigroup on on the space X of bounded uniformly continuous functions on R^n.

Proof: From the Ito's lemma

$$E[|X_t^{0,x} - X_t^{0,y}|^2] \le M\,|x-y|^2.$$

Thus, for all $y_n \to x$ there exists a subsequence z_n of y_n such that $X_t^{z_n} \to X_t^x$ almost surely.

$$\lim_{n\infty} E[g(X_t^{0,z_n})] = E[g(X_t^x)].$$

Moreover,

$$|X_t^{0,x} - x| \to 0 \text{ almost surely as } t \to 0^+,$$

and thus $H_t g \to g$ in X as $t \to 0^+$.

Definition (Generator) Define the generator of Ito's diffusion process X_t by

$$(\mathcal{A}\phi)(x) = \lim_{t\ to0^+} \frac{H_t\phi - \phi}{t} = \lim_{t\ to0^+} \frac{E^x[\phi(X_t) - \phi(x)}{t}$$

with domain

$$dom(\mathcal{A}) = \{\phi \in X : s - \lim_{t\ to0^+} \frac{E^x[\phi(X_t) - \phi(x)}{t} \text{ exists}\}.$$

It will be shown that

Theorem (Generator)

$$\mathcal{A}\phi = \sum_j b_j(x)\,\frac{\partial}{\partial x_j}\phi + \frac{1}{2}\sum_{i,j}(\sigma(x)\sigma(x)')_{i,j}\,\frac{\partial^2}{\partial x_i \partial x_j}\phi$$

for $\phi \in C_0^2(R^n)$.

For SDE (**??**) we consider the augmented system

$$dX_0 = dt$$

$$dX_t = b(X_0, X_t) + \sigma(X_0, X_t)dB_t$$

Thus, the generator

$$\mathcal{A}\phi(t,x) = \frac{\partial}{\partial t}\phi(t,x) + \mathcal{A}_x\phi(t,x)$$

for $\phi \in C_0^{1,2}(R \times \Omega)$.

8.3 Dynkin's formula

In this section we discuss the Dynkin's formula.

Theorem (Dynkyin's Formula) Let X_t be an Ito's process

$$dX_t = u\,dt + v\,dB_t.$$

Assume a stopping τ satisfies $E[\tau] < \infty$. Then, for $f \in C_0^2(R^n)$

$$E^x[f(\tau, X_\tau)] = f(x) + E^x[\int_0^\tau (\frac{\partial}{\partial t} + L(s,\omega))f(s, X_s)\,ds]$$

where

$$L(t,\omega)f = \sum_j u_j \frac{\partial f}{\partial x_j}(t, X_t) + \frac{1}{2}\sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(t, X_t).$$

Proof: By the Ito's formula

$$f(X_\tau) = f(x) + \int_0^\tau (\frac{\partial}{\partial t} + L(t,\omega))f(s, X_s)\,ds + \int_0^\tau \nabla f(s, X_s)\cdot v\,dB_s.$$

Since

$$E^x[\int_0^{\tau\wedge k} \nabla f(s, X_s)\cdot v\,dB_s] = 0.$$

and

$$E^x[|\int_0^{\tau\wedge k} \nabla f(s, X_s)\cdot v\,dB_s - \int_0^\tau \nabla f(s, X_s)\cdot v\,dB_s|^2] = M^2\,E[\tau - \tau\wedge k]| \to 0$$

for some M as $k \to \infty$, we have

$$E^x[\int_0^\tau \nabla f(s, X_s) \cdot v \, dB_s] = 0. \square$$

Let Ω be a bounded open set in R^n and $\tau = \tau_\Omega$. Then the Dynkin's formula holds for $f \in C^2(R^n)$. Assume that there exits a function f such that

$$\mathcal{A} f = 0, \quad f|_{\partial\Omega} = 1$$

Then, for $\tau = \tau_\Omega$ and $x \in \Omega$ we have

$$E^x[f(X_\tau)] = f(x) + E^x[\int^\tau \mathcal{A} f(X_s) \, ds]$$

and thus

$$E[\tau] = 1 - f(x).$$

Example (Brownian Motion) Let $X_t = x + B_t$ in R^n and $f = |x|^2$. Define a stooping time τ by

$$\tau = \inf\{t \geq 0 : |X_t| = R\}, \quad |x| < R.$$

By the Dynkin's formula

$$E^x[f(X_{\tau \wedge k})] = f(x) + E^x[\int_0^{\tau \wedge k} \frac{1}{2} \Delta f(X_s) \, ds] = |x|^2 + n \, E^x[\tau \wedge k].$$

Thus, letting $k \to \infty$

$$E^x[\tau] = R^2 - |x|^2.$$

For $n = 2$ let $f(x) = -\log|x|$ and $|x| \geq R$. Since $\Delta f = 0$,

$$E^x[f(X_{\tau_k}] = f(x)$$

for $\tau_k = \inf\{t \geq 0 | X_t| = R$ or $|X_t| = 2^k R\}$. For $p_k = P^x(|X_{\tau_k}| = R)$ and $q_k = P^x(|X_{\tau_k}| = 2^k R)$.

$$-\log R \, p_k - (\log R + k \log 2) q_k = -\log|x|$$

Thus, $q_k \to 0$ as $k \to \infty$ and $P^x(\tau < \infty) = 1$. This implies the Brownian motion is recurrent. For $n > 2$ let $f(x) = |x|^{2-n}$. Since

$$R^{2-n} p_k + (2^k R)^{2-n} q_k = |x|^{2-n},$$

$$\lim_{k \to \infty} p_k = P^x(\tau < \infty) = (\frac{|x|}{R})^{2-n}$$

and the Brownian motion is transient.

8.4 Martingale Problem

For Ito's diffusion X_t we have it follows from Ito's rule

$$f(X_t) = f(X_s) + \int_0^t A f(X_s) \, ds + \int_s^t \nabla f(X_s)^t \sigma(X_s) \, dB_s$$

for $f \in C^2(R^n)$. Thus, we have

$$E^x[f(X_t) | \mathcal{F}_s] = f(X_s) + E^x[\int_s^t A f(X_r) \, dr | \mathcal{F}_s].$$

Define

$$M_t = f(X_t) - \int_0^t Af(X_s)\,ds.$$

Then, M_t is a $\mathcal{F}_t$-martingale. In fact,

$$E^x[M_t|\mathcal{F}_s] = f(X_s) + E^x[\int_s^t Af(X_r)\,dr|\mathcal{F}_s] - \int_0^t Af(X_r)\,dr$$

$$= f(X_s) - \int_0^s Af(X_r)\,dr = M_s.$$

Definition (Martingale Problem) Given a second order elliptic linear operator L:

$$Lu = \sum_{i,j} a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j}u + \sum_j b_j(x)\frac{\partial}{\partial x_j}u(x)$$

the martingale Problem for L consists in finding for each $x \in R^n$ a probability measure P^x over the space of all continuous functions $X_t(\omega) : [0,\infty) \to R^n$ such that $X_0 = x$ and for all $f \in C^2(R^n)$ we have that

$$f(X_t) - \int_0^t (Lf)(X_s)\,ds$$

is an $\mathcal{F}_t$ local martingale under P^x. If for any $x \in R^n$ there exists a unique P^x satisfying the above conditions we say that the martingale problem for L is well-posed.

Note that existence for the martingale problem implies existence of solutions to the Cauchy problem for the operator L with initial data in $C^(R^n)$, indeed, if $f \in C^2(R^n)$ then we can define

$$u(x,t) = E_{P^x}[f(X_t)].$$

Then it can be shown that $u(x,t)$ is a classical solution to the Cauchy problem

$$\frac{\partial u}{\partial t} = Lu \text{ in } R^n, \quad u(x,0) = f(x).$$

Thus at first glance it would seem that well-posedness for the martingale Problem for L is a stronger fact than that the well-posedness for the corresponding Cauchy problem, but often one can turn this around and show the well-posedness for the martingale problem after understanding the Cauchy problem for L well enough. If $a(x) = \sigma\sigma(x)^t$ and $\sigma,\ b$ and c are Lipschitz continuous, then the martingale is well-posed for L. The martingale problem for $\mathcal{A} = \sum_j b_j(x)\frac{\partial}{\partial x_j} + \frac{1}{2}\sum_{i,j}(\sigma\sigma^t)_{i,j}\frac{\partial^2}{\partial x_i \partial x_j}$ has a solution if and only if the stochastic differential equation

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t$$

has a weak solution. Moreover, this weak solution is a Markov process if and only if the martingale problem for $\mathcal{A}$ is well-posed. The case where the coefficients are merely continuous is much harder and it is an important result of Stroock and Varadhan that says still in this case the martingale problem is well-posed.

Theorem Let X_t is an Ito's diffusion given by

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t$$

and Y_t is an Ito's process given by

$$dY_t = u(t,\omega)\,dt + v(t,\omega)\,dB_t$$

Then $X_t \sim Y_t$ if and only if

$$E^x[u(t)|\mathcal{Y}_t] = b(Y_t) \text{ and } v(t,\omega)v^t(t,\omega) = \sigma(Y_t)\sigma(Y_t)^t$$

for a.a. (t, ω), where $\mathcal{Y}_t$ is σ-algebra generated by $\{Y_s,\ s \le t\}$.

Proof: Assume $\mathcal{A}$ is the generator ofX_t

$$\mathcal{A} = \sum_j b_j(x)\frac{\partial}{\partial x_j} + \frac{1}{2}\sum_{i,j}(\sigma(x)\sigma(x)^t)_{ij}\frac{\partial^2}{\partial x_i \partial x_j}$$

and define

$$Lf(t,\omega) = \sum_j u_j(t,\omega)\frac{\partial f}{\partial x_j}(Y_t) + \frac{1}{2}\sum_{i,j}(v(t,\omega)v(t,\omega)^t)_{ij}\frac{\partial^2 f}{\partial x_i \partial x_j}(Y_t)$$

It follows from the Ito's rule that

$$E[f(Y_t)|\mathcal{Y}_s] = f(Y_s) + E[\int_s^t Lf(r,\omega)\,dr|\mathcal{Y}_s] + E[\int_s^t \nabla f^t v\,dB_r|\mathcal{Y}_s]$$

$$= f(Y_s) + E[\int_s^t [Lf(r,\omega)|\mathcal{Y}_r]\,dr|\mathcal{Y}_s]$$

$$= f(Y_s) + E[\int_s^t \mathcal{A}f(Y_r)\,dr|\mathcal{Y}_s].$$

Thus, if we define

$$M_t = f(Y_t) - \int_0^t \mathcal{A}f(Y_r)\,dr,$$

then

$$E[M_t|\mathcal{Y}_s] = f(Y_s) + E[\int_s^t \mathcal{A}f(Y_r)\,dr|\mathcal{Y}_s] - E[\int_0^t \mathcal{A}f(Y_r)\,dr|\mathcal{Y}_s] = f(Y_s) - \int_0^s \mathcal{A}f(Y_r)\,dr = M_s$$

Hence M_t is a martingale with respect to the σ-algebra $\mathcal{Y}_t$. Since the martingale problem has the unique solution, X_t and Y_t have the same law. □

The specific case of Theorem is:

Corollary (Brownian Motion) Let be an Ito's process defined by

$$dY_t = v(t,\omega)\,dB_t$$

Then $Y_t \sim B_t$ if and only if

$$v(t,\omega)v^t(t,\omega) = I_n \text{ for a.a. } (t,\omega)$$

Example (Bessel's process) Let $R_t(\omega) = |B_t(\omega)|_2$ in $R^n,\ n \ge 2$. Then it satisfies

$$dR_t = \frac{B_t \cdot dB_t}{R_t} + \frac{n-1}{2R_t}\,dt. \tag{8.8}$$

From Corollary

$$\hat{B}_t = \int_0^t \frac{B_t \cdot dB_t}{R_t}$$

is Brownian motion. Thus, (**??**) is written as

$$dR_t = \frac{n-1}{2R_t}\,dt + d\hat{B}_t$$

which has the generator

$$Af(x) = \frac{1}{2}f''(x) + \frac{n-1}{2x}f'(x).$$

8.5 Girzanov Transform

In this section we discuss the Grizanov transform. In probability theory, the Girzanov theorem describes how the dynamics of stochastic processes change when the original measure is changed to an equivalent probability measure, i.e. one can shift the drift term in Ito's diffusion by the measure transform. The Grizanov theorem uses the measure change:

$$\frac{d\nu}{d\mu}(\omega) = f(\omega) \in L^1(\Omega) \text{ on } (\Omega, \mathcal{F}$$

on a measure space $(\Omega, \mathcal{F})$ and we have

Lemma (Measure Change) Assume

$$E_\nu[|X|] = \int_\Omega |X(\omega)| f(\omega)\, d\mu = E_\mu[fX] < \infty.$$

Then we have

$$E_\nu[X|\mathcal{H}]\, E_\mu[f|\mathcal{H}] = E_\mu[fX|\mathcal{H}].$$

Proof: The lemma follows from the following identities:

$$\int_H E_\nu[X|\mathcal{H}] f\, d\mu = \int_H E_\nu[X|\mathcal{H}]\, d\nu = \int_H X\, d\nu = \int_H Xf\, d\mu = E_\mu[fX|\mathcal{H}]$$

$$\int_H E_\nu[X|\mathcal{H}] f\, d\mu = E_\mu[E_\nu[X|\mathcal{H}] f I_H|\mathcal{H}] = E_\mu[I_H E_\nu[X|\mathcal{H}]\, E_\mu[f|\mathcal{H}]] = \int_H E_\nu[X|\mathcal{H}] E_\mu[f|\mathcal{H}]\, d\mu. \square$$

The standard form of the Grizanov theorem is:

Theorem I (Girzanov) Let Y_t be an Ito process defined by

$$dY_t = b(t,\omega)\, dt + dB_t$$

and M_t is an exponential martingale;

$$M_t = e^{-\int_0^t b(s,\omega)\, dB_s - \frac{1}{2}\int_0^t |b(s,\omega)|^2\, ds}.$$

Define the measure Q on $(\Omega, \mathcal{F}_T)$ by

$$dQ = M_t(\omega) dP \quad \text{on } \mathcal{F}_t$$

Then, Y_t is $(\mathcal{F}_t, Q)$-Brownian motion.

Proof: Since

$$dM_t = -bM_t\, dB_t, \quad d(M_tY_t) = M_t(b\, dt + dB_t) - Y_tM_t\, dB_t - bM_t\, dt = M_t(1 - Y_t)\, dB_t, \quad (8.9)$$

M_tY_t is a martingale. Since

$$d(M_tY_t^2) = dM_tY_t^2 + 2Y_tM_tdY_t + M_tdt - 2bM_tY_t\, dt$$

$$= (-bM_tdB_t)Y_t^2 + 2M_tY_t(b\, dt + dB_t) + M_tdt - 2bM_tY_t, dt = (-bM_tY_t^2 + 2M_tY_t)\, dB_t + M_t\, dt, \quad (8.10)$$

$$E[M_tY_t^2|\mathcal{F}_s] = M_sY_s^2 + (t-s)M_s$$

Hence Y_t and $Y_t^2 - t$ are $(\mathcal{F}_t, Q)$ martingale since

$$E_Q[Y_t|\mathcal{F}_s] = \frac{E[M_tY_t|\mathcal{F}_s]}{E[M_t|\mathcal{F}_s]} = \frac{M_sY_s}{M_s} = Y_s$$

and

$$E_Q[Y_t^2 - t|\mathcal{F}_s] = \frac{E[M_t(Y_t^2 - t)|\mathcal{F}_s]}{E[M_t|\mathcal{F}_s]} = \frac{M_s Y_s^2 - sM_s}{M_s} = Y_s^2 - s.$$

By the Levy characterization of Brownian motion Y_t is a $(\mathcal{F}_t, Q)$ Brownian motion.

Remark $M_T\, dP = M_t dP$ on $\mathcal{F}_t,\ t \le T$, i.e.,

$$\int f M_T\, dP = E[M_T f] = E[E[M_T f|\mathcal{F}_t]] = E[f E[M_T|\mathcal{F}_t]] = E[f M_t] = \int f M_t\, dP$$

for all $f \in \mathcal{F}_t$ -measurable.

The drift shift by the Grizanov theorem is:

Theorem II (Girzanov) Let $X_t,\ Y_t$ be Ito processes defined by

$$dX_t = \alpha(t,\omega)\, dt + \theta(t,\omega)\, dB_t \tag{8.11}$$

and

$$dY_t = \beta(t,\omega)\, dt + \theta(t,\omega)\, dB_t \tag{8.12}$$

Assume that there exists a $u(t,\omega)$ such that

$$\theta(t,\omega)u(t,\omega) = \beta(t,\omega) - \alpha(t,\omega)$$

and assume $E[e^{\frac{1}{2}\int_0^T |u(s,\omega)|^2\, ds}] < \infty$ (Novikov condition). Let M_t be an exponential martingale;

$$M_t = e^{-\int_0^t u(s,\omega)\, dB_s - \frac{1}{2}\int_0^t |u(s,\omega)|^2\, ds}.$$

Define the measure Q on $(\Omega, \mathcal{F}_T)$ by

$$dQ = M_t(\omega) dP \quad \text{on } \mathcal{F}_t$$

Then, $\hat{B}_t = \int_0^t u(s,\omega)\, ds + B_t$ is $(\mathcal{F}_t, Q)$-Brownian motion and

$$dY_t = \alpha_t\, dt + \theta(t,\omega) d\hat{B}_t \text{ on } (\mathcal{F}_T, Q),$$

i.e., $(Y_t, \hat{B}_t)$ is a weak solution to (**??**) on $(\mathcal{F}_t, Q)$.

Proof: The theorem follows from

$$dY_t = \beta(t,\omega)\, dt + \theta(t,\omega)(d\hat{B}_t - u(t,\omega)\, dt)$$

$$= (\beta(t,\omega) - \theta(t,\omega)u(t,\omega))\, dt + \theta(t,\omega) d\hat{B}_t = \alpha(t,\omega)\, dt + \theta(t,\omega) d\hat{B}_t.\square$$

The weak solution by the Grizanov theorem is:

Theorem III (Girzanov) Let Y_t be an Ito diffusion processes defined by

$$dY_t = b(t, Y_t) dt + \sigma(t, Y_t)\, dB_t.$$

Assume that there exists a $u(t,\omega)$ such that

$$\sigma(t, Y_t)u(t,\omega) = b(t, Y_t) - a(t, Y_t)$$

and assume $E[e^{\frac{1}{2}\int_0^T |u(s,\omega)|^2\, ds}] < \infty$ (Novikov condition). Let M_t be an exponential martingale;

$$M_t = e^{-\int_0^t u(s,\omega)\, dB_s - \frac{1}{2}\int_0^t |u(s,\omega)|^2\, ds}.$$

Define the measure Q on $(\Omega, \mathcal{F}_T)$ by

$$dQ = M_t(\omega) dP \quad \text{on } \mathcal{F}_t$$

Then,

$$\hat{B}_t = \int_0^t u(s,\omega)\,ds + B_t$$

is $(\mathcal{F}_t, Q)$-Brownian motion. Then $(Y_t, \hat{B}_t)$ satisfies

$$dY_t = a(t,Y_t)\,dt + \sigma(t,Y_t)d\hat{B}_t \text{ on } (\mathcal{F}_T, Q).$$

Example (Weak Solution) (1) One can construct a weak solution to

$$dX_t = a(t,x_t)\,dt + dB_t, \tag{8.13}$$

where $a(t,x)$ be bounded continuous. Let

$$Y_t = x + B_t$$

with $b = 0,\ \sigma = I$. Let $u = -a(Y_t)$ and

$$M_t = e^{\int_0^t a(s,x+B_s)dB_s - \frac{1}{2}\int_0^t |a(s,x+B_s)|^2\,ds}$$

and

$$\hat{B}_t = -\int_0^t a(s,x+B_s)\,ds + B_t.$$

Then, $(Y_t, \hat{B}_t)$ is a weak solution to

$$dX_t = a(t,X_t)\,dt + d\hat{B}_t \text{ on } (\mathcal{F}_T, Q).$$

Also, for any weak solution X_t to (**??**) we have

$$E[f(X^x_{t_1},\cdots,X^x_{t_k})] = E_Q[f(Y^x_{t_1},\cdots,Y^x_{t_k})] = E_P[M_t f(B^x_{t_1},\cdots,B^x_{t_k}]$$

for all bounded continuous function f and $0 \le t_1 \le \cdots \le t_n \le t$.
(2) Let Y_t is the geometrical Brownian motion:

$$Y_t = e^{\int_0^t \sigma(s)\,dB_s - \frac{1}{2}\int_0^t |\sigma(s)|^2\,ds}$$

i.e.,

$$dY_t = \sigma(t)Y_t\,dB_t \quad (b=0).$$

Let $u(t,Y_t) = -\dfrac{a(t,Y_t)}{\sigma(t)Y_t}$ and define

$$M_t = e^{-\int_0^t u(s,Y_s)dB_s - \frac{1}{2}\int_0^t |u(s,Y_s)|^2\,ds}$$

and

$$\hat{B}_t = \int_0^t u(s,Y_s)\,ds + B_t.$$

Then, $(Y_t, \hat{B}_t)$ is a weak solution to

$$dX_t = a(t,X_t)\,dt + \sigma(t)X_t\,d\hat{B}_t$$

on $(\mathcal{F}_t, Q)$.
(3) Let Y_t is the martingale

$$Y_t = x + \int_0^t \sigma(s)\,dB_s,$$

i.e.

$$dY_t = \sigma(t)\,dB_t \quad (b=0).$$

Let $u(t, Y_t) = -\dfrac{a(t, Y_t)}{\sigma(t)}$ and define

$$M_t = e^{\int_0^t u(s,Y_s)dB_s - \frac{1}{2}\int_0^t |u(s,Y_s)|^2\, ds}$$

and

$$\hat{B}_t = \int_0^t u(s, Y_s)\, ds + B_t.$$

Then, $(Y_t, \hat{B}_t)$ is a weak solution to

$$dX_t = a(t, X_t)\, dt + \sigma(t)\, d\hat{B}_t.$$

on $(\mathcal{F}_t, Q)$.

8.6 Excises

Problem 1 Check (**??**)–(**??**).
Problem 2 Consider the SDE

$$dX_t = f(t, X_t)\, dt + \sigma(t) X_t\, dB_t$$

(1) Define

$$F_t = e^{-\int_0^t \sigma(s)dB_s + \frac{1}{2}\int_0^t |\sigma(s)|^2\, ds}.$$

Show that $d(F_t X_t) = F_t f(t, X_t)\, dt$.
(2) Let $Y_t(\omega)$ be a solution to

$$\frac{d}{dt} Y_t(\omega) = F_t(\omega) f(t, F_t^{-1}(\omega) Y_t(\omega)).$$

Show that $X_t = F_t^{-1}(\omega) Y_t(\omega)$ defines a solution to the SDE.
(3) If $f(t, x) = r(t)x$, then

$$X_t = X_0 e^{\int_0^t \sigma(s)dB_s + \int_0^t (r(s) - \frac{1}{2}|\sigma(s)|^2)\, ds}$$

Derive a solution to

$$dX_t = X_t^\gamma\, dt + \sigma X_t\, dB_t.$$

9 Stochastic Integral with respect to Martingale Process

Let $(\Omega, \mathcal{F}, P)$ be the probability space and $\mathcal{F}_t$ be the right continuous increasing family of sub σ algebras (i.e. $\mathcal{F}_t = \cap_{s>t} \mathcal{F}_s$). Let M_t is a right continuous square integrable martingale. The process X_t is predictable if measurable with respect to the σ-algebra $\mathcal{F}_{t^-}$ for each time t. Every process that is left continuous is a predictable process. For every square integrable $\mathcal{F}_t$ adapted process there exists a predictable $\tilde{\Phi} \in \mathcal{L}_2$ such that $\tilde{\Phi}$ is a modification of Φ. For example, we may take

$$\tilde{\Phi}_t(\omega) = \limsup_{h \to 0^+} \frac{1}{h} \int_{t-h}^t \Phi_s(\omega)\, ds.$$

One can define the stochastic integral

$$X_t = \int_0^t H_s dM_s, \tag{9.1}$$

where $\{M_t,\ t \geq 0\}$ is a square integrable martingale and $\{H_t,\ t \geq 0\}$ is a predictable process.

Definition Let $\mathcal{L}_0$ be the set of bounded adapted process such that

$$H_t = H_j \text{ on } [t_j, t_{j+1}) \text{ and } H_j \text{ is } \mathcal{F}_{t_j} \text{ measurable},$$

with some partition $P = \{0 = t_0 < t_1 < \cdots\}$ of the interval $[0, T]$. For $H_t \in \mathcal{L}_0$

$$X_t = I(H_t) = \sum_{j=0}^{k-1} H_j(M_{t_{j+1}} - M_{t_j}) + H_{t_k}(M_t - M_{t_k}). \tag{9.2}$$

As the discrete time case, we define the quadratic variation of $\{M_t,\ t \geq 0\}$ by

$$E[\langle M\rangle_t - \langle M\rangle_s|\mathcal{F}_s] = E[(M_t - M_s)^2|\mathcal{F}_s],$$

then

$$|M_t|^2 - \langle M\rangle_t$$

is a martingale and $\langle M\rangle_t$ is naturally increasing predictable process. We can complete a space of predictable process by the norm

$$\int_0^T |H_t|^2\, d\langle M\rangle_t,$$

and the completion is called $L^2(\langle M\rangle)$. Note that I is a linear operator on the subspace $\mathcal{L}_0$ of simple predictable process of $L^2(\langle M\rangle)$ and it follows from Theorem for the martingale transform that

$$|X_t|^2 - \int_0^t |H_s|^2\, d\langle M\rangle_s$$

ia martingale and

$$\langle X\rangle_t = \int_0^t |H_s|^2\, d\langle M\rangle_s.$$

Proposition 1 The stochastic integral $\int_0^t f_s\, dM_s$ for $f \in \mathcal{L}_0$ is a square integrable martingale and satisfies

$$E[\textstyle\int_0^t f_s\, dM_s] = 0$$

$$\langle \textstyle\int_0^t f_s\, dM_s\rangle_t = \int_0^t f_s^2\, d\langle M\rangle_s$$

$$E[|\textstyle\int_0^t f_s\, dM_s|^2] = E[\int_0^t f_s^2\, d\langle M\rangle_s] = ||f||^2.$$

Proof: For $t > s$ (without loss of generality) we assume that t belongs to the partition P.

$$E[\textstyle\int_0^t f_s\, dM_s] = \sum_i E[E[f_{t_i}(M_{t_{i+1}} - M_{t_i})|\mathcal{F}_{t_i})]]$$

$$= \sum_i E[f_{t_i} E[(M_{t_{i+1}} - M_{t_i})|\mathcal{F}_{t_i})]] = 0$$

Next,

$$E[|\int_s^t f_\sigma\, dM_\sigma|^2|\mathcal{F}_s] = \sum_i E[E[f_{t_i}^2(M_{t_{i+1}} - M_{t_i})^2|\mathcal{F}_{t_i})|\mathcal{F}_s]]$$

$$+2\sum_{k>\ell} E[E[f_{t_k} f_{t_\ell}(M_{t_{k+1}} - M_{t_k})(M_{t_{\ell+1}} - M_{t_\ell})|\mathcal{F}_{t_\ell}]|\mathcal{F}_s].$$

Here,

$$E[f_{t_i}^2(M_{t_{i+1}} - M_{t_i})^2|\mathcal{F}_{t_i}] = f_{t_i}^2 E[(M_{t_{i+1}} - M_{t_i})^2|\mathcal{F}_{t_i}]$$

$$= f_{t_i}^2 E[M_{t_{i+1}}^2 - M_{t_i}^2|\mathcal{F}_{t_i}] = f_{t_i}^2 E[\langle M\rangle_{t_{i+1}} - \langle M\rangle_{t_i}|\mathcal{F}_{t_i}]$$

and

$$E[f_{t_k}f_{t_\ell}(M_{t_{k+1}}-M_{t_k})(M_{t_{\ell+1}}-M_{t_\ell})|\mathcal{F}_{t_\ell}]=E[f_{t_k}f_{t_\ell}E[M_{t_{k+1}}-M_{t_k}|\mathcal{F}_{t_k}](M_{t_{\ell+1}}-M_{t_\ell})|\mathcal{F}_{t_\ell}]=0.$$

Thus,

$$E[|\int_s^t f_\sigma\,dM_\sigma|^2|\mathcal{F}_s]=\sum_i E[f_{t_i}^2]E[\langle M\rangle_{t_{i+1}}-\langle M\rangle_{t_i}|\mathcal{F}_s]=E[\int_s^t |f_\sigma|^2\,d\langle M\rangle_\sigma|\mathcal{F}_s].$$

which implies the claim. □

Definition (Stochastic Integral with respect martingale) For $H\in L^2(\langle M\rangle)$

$$\int_0^t H_s\,dM_s=\lim\ X_t^n=\lim\int_0^t H_s^n\,dM_s$$

where $H_t^n\in\mathcal{L}_0$ and $||H^n-H||\to 0$ as $n\to\infty$. From Proposition 1

$$E[|X_T^n-X_T^m|^2]=||H^n-H^m||^2$$

and by the martingale inequality

$$E[\sup_{0\le s\le T}|X_s^n-X_s^m|^2]\le 4\,E[|X_T^n-X_T^m|^2].$$

Since $\mathcal{L}_0$ is dense in $L^2(\langle M\rangle)$ there exits a unique limit X_t of X_t^n in $L^2(\langle M\rangle)$ and $X_t^n,\ 0\le t\le T$ has a subsequence that converges uniformly a.s. to X_t (pathwise). Thus, the limit $Y_t,\ 0\le t\le T$ defines the stochastic integral $\int_0^t f_s\,dM_s$ and is right continuous. That is, I is a bounded linear operator on $\mathcal{L}_0$ and since $\mathcal{L}_0$ is dense in $L^2(\langle M\rangle)$ the stochastic integral (**??**) is the extension of (**??**) on $L^2(\langle M\rangle)$. Then, X_t is martingale and

$$E[X_t]=0$$

$$|X_t|^2-\int_0^t|H_s|^2\,d\langle M\rangle_s$$

is a martingale after the extension.

Remark (1) If M_t is continuous, then it is not necessary to assume that $\mathcal{F}_t$ is right continuous. If we let $\mathcal{F}_{t^+}=\cap_{s>t}\mathcal{F}_s$. Then if M_t is an $\mathcal{F}_t$ continuous martingale, M_t is also an $\mathcal{F}_{t^+}$ martingale. The corresponding natural increasing process $\langle M\rangle_t$ is $\mathcal{F}_{t^+}$ adapted, but since $\langle M\rangle_t$ is continuous $\langle M\rangle_t$ is $\mathcal{F}_t$ adapted. Hence $M_t^2-\langle M\rangle_t$ is an $\mathcal{F}_t$ continuous martingale.
(2) If M_t is continuous, then it is not necessary to assume that Φ_t is predictable, and $\int\Phi_s\,dM_s$ is a continuous $\mathcal{F}_t$ martingale for Φ_t is a square integrable $\mathcal{F}_t$ adapted process.
(3) $L^2(\langle M\rangle)$ is a Hilbert space with inner product

$$(f,g)=E[\int_0^T f_tg_t\,d\langle M\rangle_t].$$

If the original martingale M_t is almost surely continuous and so is X_t. This is obvious if H_t is simple by (**??**) and follows from the Doob's martingale inequality for general. That is,

$$P(\sup_{0\le s\le T}|X_s^m-X_s^n|\ge\epsilon)\le\frac{1}{\epsilon^2}||H^m-H^n||.$$

Choose a sequence n_k such that

$$P(\sup_{0\le s\le T}|X_s^m-X_s^n|\ge 2^{-k})\le 2^{-k}$$

and thus

$$\sum^{\infty} P(\sup_{0\le s\le T} |X_s^{n_{k+1}} - X_s^{n_k}| \ge 2^{-k}) < \infty.$$

By Borel-Cantelli lemma

$$P(\sup_{0\le s\le T} |X_s^{n_{k+1}} - X_s^{n_k}| \ge 2^{-k}) \text{ for infinitely many k}) = 0.$$

So, for almost surely ω, there exists $k \ge k_1(\omega)$ such that for all $k \ge k_1(\omega)$

$$\sup_{0\le s\le T} |X_s^{n_{k+1}} - X_s^{n_k}| \le 2^{-k}.$$

Hence, $\lim X_t(\omega) = \lim_{k\to\infty} X_t^{n_k}(\omega)$ is continues.

Example Let $M_t = N_t - t$ for Poisson process N_t. Then, M_t and $|M_t|^2 - t$ are martingales.

$$X_t = \int_0^t N_s\, dM_s = \sum_{\tau_j\le t} N((\tau_j)^-) - \int_0^t N_s\, ds.$$

9.1 Generalized Ito's differential rule

Let $X_t = X_0 + M_t + A_t$ where $M_t \in \mathcal{M}_c^2$ is a continuous (locally) square integrable martingale and A_t is an continuous process of bounded variation. Then we have the Ito's differential rule:

Theorem For $f \in C^{1,2}([0,T]\times R^d)$

$$f(X_t) - f(X_0) = \int_0^t f_t(s,X_s)\, ds + \sum_{i=1}^d f_{x_i}(s,X_s) dX_s^i + \frac{1}{2}\sum_{i,j=1}^d \int_0^t f_{x_i,x_j}(s,X_s)\, d\langle M^i, M^j\rangle_s$$

Or, equivalently (increment form)

$$df(t,X_t) = f_t\, dt + +f_{x_i}(X_t) dX_t^i + \frac{1}{2}\sum_{i,j=1}^d \int_0^t f_{x_i,x_j}(X_t)\, d\langle M^i, M^j\rangle_t. \tag{9.3}$$

Proof: For a positive integer n we define a stopping time τ_n by

$$\tau_n = \inf\{t > 0 : |X_0 + M_t| > n \text{ or } |A_t| > n\}$$

Then $\tau_n \to \infty$ as $n \to \infty$ a.s.. Thus, it suffices to prove the formula for $X_{t\wedge\tau_n}$ and thus without loss of generality we can assume that $|X_0+M_t|,\ |A_t|$ are bounded and $f,\ f_{x_i},\ f_{x_i,x_j}$ are bounded and uniformly continuous.

Note that by the mean value theorem

$$f(X_t)-f(X_0) = \sum_{k=0}^n \sum_{i=0}^d f_{x_i}(X_{t_k})(X_{t_{k+1}} - X_{t_k}) + \frac{1}{2}\sum_{k=0}^n \sum_{i=1}^d \sum_{j=1}^d f_{x_i,x_j}(\xi_{i,j})(X_{t_{k+1}}^i - X_{t_k}^i)(X_{t_{k+1}}^j - X_{t_k}^j).$$

By the definition of the stochastic integral the first term of RHS converges to

$$\sum_{i=1}^d \int_0^t (f_{x_i}(X_s)\, dM_s^i + f_{x_i}(X_s)\, dA_s^i)$$

The second term is a linear combination of forms

$$\sum_k g(\xi_k)(M_{t_{k+1}} - M_{t_k})(N_{t_{k+1}} - N_{t_k})$$

$$\sum_k g(\xi_k)(M_{t_{k+1}} - M_{t_k})(A_{t_{k+1}} - A_{t_k})$$

$$\sum_k g(\xi_k)(A_{t_{k+1}} - A_{t_k})(C_{t_{k+1}} - C_{t_k})$$

where $M_t,\ N_t \in \mathcal{M}_2^c$ and $A_t,\ C_t$ are continuous process of bounded variation. Here

$$|\sum_k g(\xi_k)(M_{t_{k+1}}-M_{t_k})(A_{t_{k+1}}-A_{t_k})| \leq \|g\| \sup_k |M_{t_{k+1}}-M_{t_k}|\, A_t \to 0$$

as $|P| \to 0$. In the following theorem it will be shown that the first term converges to $\int_0^t g(X_s)\, d\langle M,N\rangle_s$.

Lemma If $|M_s| \leq C$ for some C on $[0,t]$, then

$$E[|V_n|^2] \leq 12\, C^4 \quad \text{if} \quad V_n = \sum_{k=0}^n (M_{t_{k+1}}-M_{t_k})^2.$$

Proof: It is easy to see that

$$|V_n|^2 = \sum_{k=0}^n (M_{t_{k+1}}-M_{t_k})^4 + 2\sum_{k=1}^n (V_n - V_{k-1})(M_{t_{k+1}}-M_{t_k})^2$$

and

$$E[(V_n-V_{k-1})|\mathcal{F}_{t_k}] = E[\sum_{i=k}^n (M_{t_{i+1}}-M_{t_i})^2|\mathcal{F}_{t_k}] = E[(M_t - M_{t_k})^2|\mathcal{F}_{t_k}] \leq 4\, C^2$$

Thus,

$$E[\sum_{k=1}^n (V_n-V_{k-1})(M_{t_{k+1}}-M_{t_k})^2] \leq 4C^2 E[V_n] = 4C^2\, E[M(t)^2] \leq 4C^4.$$

Also,

$$E[\sum_{k=0}^n (M_{t_{k+1}}-M_{t_k})^4] \leq 4C^2\, E[V_n] \leq 4\, C^4. \square$$

Theorem Let M_t and N_t be bounded continuous martingale. For a bounded uniformly continuous function g

$$\sum g_k\, (M_{t_{k+1}}-M_{t_k})(N_{t_{k+1}}-N_{t_k}) \to \int_0^t g(X_s)\, d\langle M,N\rangle_s \quad \text{in } L^1(\Omega).$$

where $g_k = g(X_{t_k} + (1-\theta_k)(X_{t_{k+1}}-X_{t_k})$ with $\theta_k \in [0,1]$.

Proof: Let

$$I = \sum\ g(X_{t_k})\,[(M_{t_{k+1}}-M_{t_k})(N_{t_{k+1}}-N_{t_k}) - (\langle M,N\rangle_{t_{k+1}} - \langle M,N\rangle_{t_k})]$$

$$J = \sum\ (g_k - g(X_{t_k}))(M_{t_{k+1}}-M_{t_k})(N_{t_{k+1}}-N_{t_k})$$

$$K = \sum\ g(X_{t_k})[(\langle M,N\rangle_{t_{k+1}} - \langle M,N\rangle_{t_k}) - \int_0^t g(X_s) d\langle M,N\rangle_s$$

We show that $I,\ J,\ K \to 0$ as $|P| \to 0$. Clearly, $E[|K|] \to 0$ as $|P| \to 0$. Let

$$V_t = \sum_{t_{k+1}\leq t} (M_{t_{k+1}}-M_{t_k})^2, \quad W_t = \sum_{t_{k+1}\leq t} (N_{t_{k+1}}-N_{t_k})^2.$$

Since

$$|J| \leq \sup_k |g_k - g(X_{t_k})| (V_t W_t)^{1/2}$$

we have from Lemma

$$E|J| \leq E[\sup_k |g_k - g(X_{t_k})|^2]^{1/2} E[V_t^2]^{1/4} E[W_t^2]^{1/4} \leq \sqrt{12} C^2 E[\sup_k |g(\xi_k) - g(X_{t_k})|^2]^{1/2} \to 0$$

as $|P|\to 0$. For I let

$$I_i=\sum_{k=0}^{i-1} g(X_{t_k})\,[(M_{t_{k+1}}-M_{t_k})(N_{t_{k+1}}-N_{t_k})-(\langle M,N\rangle_{t_{k+1}}-\langle M,N\rangle_{t_k})]$$

Then $(I_i,\mathcal{F}_{t_i})$ is a discrete-time martingale. Thus from the same arguments as in the proof of Proposition 1

$$E[|I|^2]=\sum_{k=0}^{n} E[|g(X_{t_k})|^2\,((M_{t_{k+1}}-M_{t_k})(N_{t_{k+1}}-N_{t_k})-(\langle M,N\rangle_{t_{k+1}}-\langle M,N\rangle_{t_k}))^2$$

and therefore

$$E[|I|^2]\le 2\|g\|^2\sum_{k=0}^{n} E[(M_{t_{k+1}}-M_{t_k})^2(N_{t_{k+1}}-N_{t_k})^2]+2\|g\|^2\sum_{k=0}^{n} E[(\langle M,N\rangle_{t_{k+1}}-\langle M,N\rangle_{t_k})^2].$$

Here

$$\textstyle\sum_{k=0}^{n} E[(M_{t_{k+1}}-M_{t_k})^2(N_{t_{k+1}}-N_{t_k})^2]\le E[\sup_k\,(M_{t_{k+1}}-M_{t_k})^2\sum_k\,(N_{t_{k+1}}-N_{t_k})^2]$$

$$\le E[\sup_k\,(M_{t_{k+1}}-M_{t_k})^4]^{1/2}E[|W_t|^2]^{1/2}\to 0$$

as $|P|\to 0$. Since $\langle M,N\rangle_s,\ s\in[0,t]$ is bounded

$$\sum_{k=0}^{n} E[(\langle M,N\rangle_{t_{k+1}}-\langle M,N\rangle_{t_k})^2]\le E[\sup_k\,|\langle M,N\rangle_{t_{k+1}}-\langle M,N\rangle_{t_k}|\|\langle M,N\rangle\|_t]\to 0$$

as $|P|\to 0$. Thus $E[|I|^2]\to 0$. $\square$

Theorem (Levy Characterization of Brownian motion) Let X_t be a continuous $\mathcal{F}_t$ adapted process. Then the followings are equivalent
(1) X_t is an $\mathcal{F}_t-$ Brownian motion.
(2) X_t is a square integrable martingale and $\langle X^i,X^j\rangle_t=\delta_{i,j}\,t$.
Proof: It suffices to prove that

$$E[e^{i\,(\xi,X_t-X_s)}|\mathcal{F}_s]=e^{-\frac{1}{2}\,|\xi|^2(t-s)}.$$

Applying the Ito's formula for $e^{i\,(\xi,X_t)}$

$$e^{i\,(\xi,X_t)}-e^{i\,(\xi,X_s)}=\int_s^t(i\xi\,e^{i\,(\xi,X_\sigma)},dX_\sigma)-\frac{1}{2}\int_s^t|\xi|^2e^{i\,(\xi,X_\sigma)}\,d\sigma.$$

Since $X_t\in\mathcal{M}_2^c$

$$E[\int_s^t(i\xi\,e^{i\,(\xi,X_\sigma)},dX_\sigma)|\mathcal{F}_s]=0.$$

Multiplying the both sides of this by $e^{-i\,(\xi,X_s)}$, for $A\in\mathcal{F}_s$

$$E[e^{i\,(\xi,X_t-X_s)}\,\chi_A]-P(A)=-\frac{1}{2}\,|\xi|^2\int_s^t E[e^{i\,(\xi,X_\sigma-X_s)}\chi_A]\,d\sigma.$$

Thus, we obtain

$$E[e^{i\,(\xi,X_t-X_s)}\,\chi_A]=P(A)\,e^{-\frac{1}{2}|\xi|^2(t-s)}.\square$$

9.2 Semimartingale

A stochastic process $\{X_t,\ t \geq 0\}$ is called a semimartingale if it can be decomposed as the sum of a local martingale and an adapted finite-variation process. Semimartingales are "good integrators", forming the largest class of processes with respect to which the Ito-integral can be defined. The class of semimartingales is quite large (including, for example, all continuously differentiable processes, Brownian motion and Poisson processes). Submartingales and supermartingales together represent a subset of the semimartingales. As with ordinary calculus, integration by parts is an important result in stochastic calculus. The integration by parts formula for the Ito- integral differs from the standard result due to the inclusion of a quadratic covariation term. This term comes from the fact that Ito-calculus deals with processes with non-zero quadratic variation, which only occurs for infinite variation processes (such as Brownian motion). If X and Y are semimartingales then

$$X_tY_t = X_0Y_0 + \int_0^t X_{s-}\,dY_s + \int_0^t Y_{s-}\,dX_s + \langle X, Y\rangle_t$$

where $\langle X, Y\rangle$ is the quadratic covariance process. The result is similar to the integration by parts theorem for the RiemannStieltjes integral but has an additional quadratic variation term.

Ito's lemma is the version of the chain rule or change of variables formula which applies to the Ito stochastic integral. It is one of the most powerful and frequently used theorems in stochastic calculus. For a continuous d-dimensional semimartingale $X_t \in R^d$ and twice continuously differentiable function f from R^d to R, it states that $f(X_t)$ is a semimartingale and,

$$df(X_t) = \sum_{i=1}^d f_i(X_t)\,dX_t^i + \frac{1}{2}\sum_{i,j=1}^d f_{i,j}(X_t)\,d\langle X^i, X^j\rangle_t.$$

This differs from the chain rule used in standard calculus due to the term involving the quadratic covariation. The formula can be generalized to non-continuous semimartingales by adding a pure jump term to ensure that the jumps of the left and right hand sides agree.

9.3 Excises

Problem 1 Show (**??**)–(**??**).

Problem 2 If $\{\xi_k,\ k \geq 1\}$ is a sequence of independent random variables with $E[\xi_k] = 1$. Show that $X_n = \Pi_{k=1}^n \xi_k$ is a martingale with respect to $\mathcal{F}_n = \sigma(\xi_k\ \ k \leq n)$. Consider the case $P(\xi_k = 0) = P(\xi_k = 2) = \frac{1}{2}$. Show that there is no an integrable random variable ξ such that $X_n = E[\xi|\mathcal{F}_n]$.

Problem 3 Let $\{\xi_k\}$ be a sequence of independent random variables with $E[\xi_k] = 0$ and $V(\xi_k) = \sigma_k^2$. Define $S_n = \sum_{k=1}^n \xi_k$ and $\mathcal{F}_n = \sigma(\xi_k,\ k \leq n)$. Show the following generalization of Wald's identities. If $E[\sum_{k=1}^\tau |\xi_k|] < \infty$ then $E[S_\tau] = 0$. If $E[\sum_{k=1}^\tau |\xi_k|^2] < \infty$ then $E[S_\tau^2] = E[\sum_{k=1}^\tau \sigma_k^2]$.

Problem 4 Show (**??**) and (**??**).

Problem 5 Suppose $\{X_n\}$ is a martingale satisfying some $p > 1$ $E[|X_n|^p] < \infty$. Show

$$(E[(\max_{0\leq k\leq n} |X_k|)^p])^{\frac{1}{p}} \leq \frac{p}{p-1} E[|X_n|^p]^{\frac{1}{p}}.$$

Hint: $E[|\xi|^p] = p\int_0^\infty t^{p-1} P(|\xi| > t)\,dt$. Now, we use the maximal inequality for the submartingale $|X_n|$.

Problem 6 Show (**??**)-(**??**).

Problem 7 Show that $X_t = e^{\lambda B_t - \frac{\lambda^2 t}{2}}$ satisfies $dX_t = \lambda X_t\,dB_t$. Find the generator of X_t.

10 Filtering Theory

In this section we discuss the nonlinear filtering problem for a Markov process x_t. It is the problem of estimating the state x_t from the observation $\{y_s,\ s \le t\}$ defined by

$$y_t = \int_0^t h(X_s)\,ds + W_t.$$

10.1 Discrete time Filtering

We discuss the nonlinear filtering problem for the discrete-time signal system for R^d-valued process x_k:

$$x_k = f(x_{k-1}) + w_k \tag{10.1}$$

and the observation process $y_k \in R^p$ is given by

$$y_k = h(x_k) + v_k \tag{10.2}$$

where w_k and v_k are independent i.i.d. Gaussian random variables with covariances Q and R, respectively. We assume that the initial condition x_0 and $w_k,\ v_k$ are independent random variables. The optimal nonlinear filtering problem is to find the conditional expectation $E[x_k|Y_k]$ of the process x_k given the observation data $Y_k = \{y_j,\ 0 \le j \le k\}$. The probability density function $p_{k|k}$ of the conditional expectation $E[x_k|Y_k]$ is updated by the Bayes' formula:

$$p_{k|k-1}(x) = \int_{R^d} \frac{1}{((2\pi)^n det\, Q)^{1/2}}\, e^{-\frac{1}{2}(x-f(t))^T Q^{-1}(x-f(t))} p_{k-1|k-1}(t)\,dt \tag{10.3}$$

and

$$p_{k|k}(x) = c\, e^{-\frac{1}{2}(y-h(x))^T R^{-1}(y-h(x))} p_{k|k-1}(x) \tag{10.4}$$

where $p_{k|k-1}$ is the one-step prediction and is the probability density function of x_k conditioned on Y_{k-1} and its update is based on the dynamics (**??**). The second update is based on the observation process (**??**). That is, the recursive filter (**??**)–(**??**) consists of the prediction step (**??**) and the correction step (**??**).

Most well know case when (**??**)–(**??**) has the exact update is the Kalman filter for the linear Gaussian system. Consider the linear system

$$x_k = A_k x_{k-1} + b_k + v_k$$

and

$$y_k = H_k x_k + w_k$$

If x_0 is Gaussian, then $\{(x_k, y_k)\}$ are Gaussian system. That is, $p_{k-1|k-1}$ is a Gaussian with mean $x_{k-1|k-1}$ and covariance $\Sigma_{k-1|k-1}$. From the predictor step of (**??**) we have

$$p_{k|k-1}(t) = \frac{1}{((2\pi)^n \det \Sigma_{k-1|k-1})^{\frac{1}{2}}} e^{-\frac{1}{2}(t-x_{k-1|k-1})^t \Sigma^{-1}_{k-1|k-1}(t-x_{k-1|k-1})}\,dt.$$

We show that $p_{k|k-1}$ is a Gaussian with mean $x_{k|k-1}$ and covariance $\Sigma_{k|k-1}$ with

$$\begin{aligned} x_{k|k-1} &= A_k x_{k-1|k-1} + b_k \\ \Sigma_{k|k-1} &= A_k \Sigma_{k-1|k-1} A_k^t + Q_k. \end{aligned} \tag{10.5}$$

In fact,

$$x_{k|k-1} = E[x_k|Y_{k-1}] = E_{k|k-1}[Ax_{k-1} + b_k + v_k|Y_{k-1}] = Ax_{k-1|k-1} + b_k$$

and

$$\Sigma_{k|k-1} = E_{k|k-1}[(x_k - x_{k|k-1})(x_k - x_{k|k-1})^t]$$

$$= E_{k|k-1}[(A_k(x_{k-1} - x_{k-1|k-1})(x_{k-1} - x_{k-1|k-1})^t A_k^t|Y_{k-1}] + Q_k = A_k\Sigma_{k-1|k-1}A_k^t$$

since v_k is independent with Y_{k-1}.

For the corrector step we define the innovation process

$$I_k = y_k - E_{k|k-1}[y_k] = y_k - E_{k|k-1}[H_k x_k] = y_k - Hx_{k|k-1}.$$

We show that the corrector step of (**??**) is equivalent to $p_{k|k}$ is a Gaussian with mean $x_{k|k}$ and covariance $\Sigma_{k|k}$ updated by

$$\begin{aligned} x_{k|k} &= x_{k|k-1} + G_k(y_k - Hx_{k|k-1}) \\ \Sigma_{k|k} &= \Sigma_{k|k-1} - \Sigma_{k|k-1}H_k^t(H\Sigma_{k|k-1}H^t + R)^{-1}H\Sigma_{k|k-1} \end{aligned} \tag{10.6}$$

where

$$G_k = \Sigma_{k|k-1}H_k^t(R_k + H_k\Sigma_{k|k-1}H_k^t)^{-1}$$

is called the Kalman filter gain. In fact, since $I_k = H(x_k - E_{k|k-1}[x_k]) + w_k$, the innovation process I_k is independent with Y_{k-1} and thus

$$x_{k|k} = E[x_k|Y_k] = E[x_k|Y_{k-1}] + GI_k \text{ for some } G \in R^{d\times p}$$

and thus

$$GE[x_{k|k}I_k^*] = E[(x_{k|k} - x_{k|k-1})I_k^*] = E[x_k I_k^* t]$$

Since

$$E[x_k I_k^*] = E[x_k(w_k + H_k(x_k - x_{k|k-1})^t]$$

$$= E[x_k(x_k - x_{k|k-1})^*]H_k^t$$

$$= E[(x_k - x_{k|k-1})(x_k - x_{k|k-1})^*]H_k^t = \Sigma_{k|k-1}H_k^*$$

Similarly,

$$E[I_k I_k^t] = R_k + H_k P_{k|k-1}H_k^t$$

and thus

$$G = G_k = P_{k|k-1}H_k^t(R_k + H_k P_{k|k-1}H_k^* + R_k)^{-1}.$$

Now,

$$\Sigma_{k|k} = E[(x_k - x_{k|k})(x_k - x_{k|k})^*] = E[(x_k - x_{k|k-1} - GI_k)(x_k - x_{k|k-1} - GI_k)^*]$$

$$= E[(x_k - x_{k|k-1})(x_k - x_{k|k-1})^*] - G^*E[I_k I_k^*]G$$

$$= P_{k|k-1} - \Sigma_{k|k-1}H_k^*(R_k + H_k\Sigma_{k|k-1}H_k^*)^{-1}H_k\Sigma_{k|k-1}.$$

For the nonlinear case given density $p_{k|k-1}$ let

$$z_k = y_k - E_{k|k-1}[y_k] = h(x_k) - \hat{h}_k + w_k$$

where

$$\hat{h}_k = E_{k|k-1}[h(x_k)] = \int h(x)p_{k|k-1}(x)\,dx.$$

Then, since w_k is independent of Y_{k-1}

$$E_{k|k-1}[(z_k, \eta)] = E[(h(x_k) - \hat{h}_k, \eta)] = 0$$

for all $\eta \in Y_{k-1}$. Thus,

$$x_{k|k} = x_{k|k-1} + G z_k$$

defies the least square estimate. That is, if we let

$$P_{zz} = E_{k|k-1}[z_k z_k^*] = R_k + \int_{R^d} (h(x) - \hat{h}_k)(h(x) - \hat{h}_k)^* p_{k|k-1}(x)\, dx$$

$$P_{xz} = E_{k|k-1}[(x_k - x_{k|k-1}) z_k^*] = \int_{R^d} (x - x_{k|k-1})(h(x) - \hat{h}_k)^* p_{k|k-1}(x)\, dx,$$

then the gain G_k is given by $G_k = P_{xz} P_{zz}^{-1}$.

10.2 Gaussian Filter

We consider a Gaussian approximation of the optimal filter (**??**)–(**??**) and develop Gaussian filters. We use the assumed density for $p_{k-1|k-1}$ as a single Gaussian distribution $N(x_{k-1|k-1}, P_{k-1|k-1})$ with mean $x_{k-1|k-1}$ and covariance $P_{k-1|k-1}$. Then we construct the Gaussian approximation of $p_{k|k}$ in the following two steps. First, we consider the predictor step. We approximate $p_{k|k-1}$ by the Gaussian distribution that has the same mean and covariance as $p_{k|k-1}$. By Fubini's theorem the mean and covariance of $p_{k|k-1}$ are given by

$$\begin{aligned}
E_{k|k-1}[x] &= \int_{R^d} x p_{k|k-1}(x)\, dx \\
&= \int_{R^d} \left(\int_{R^d} \frac{1}{((2\pi)^n \det Q)^{1/2}}\, e^{-\frac{1}{2}(x-f(t))^t Q^{-1}(x-f(t))} x\, dx \right) p_{k-1|k-1}(t)\, dt \\
&= \int_{R^d} f(t) p_{k-1|k-1}(t)\, dt
\end{aligned}$$

and

$$\begin{aligned}
E_{k|k-1}[xx^t] &= \int_{R^d} x x^t p_{k|k-1}(x)\, dx \\
&= \int_{R^d} \left(\int_{R^d} \frac{1}{((2\pi)^n \det Q)^{1/2}}\, e^{-\frac{1}{2}(x-f(t))^t Q^{-1}(x-f(t))} x x^t\, dx \right) p_{k-1|k-1}(t)\, dt \\
&= Q + \int_{R^d} f(t) f(t)^t p_{k-1|k-1}(t)\, dt
\end{aligned}$$

Remark 1.1 To derive the mean and covariance of $p_{k|k-1}$, we also can use (**??**) and the independence of v_k and x_{k-1}. For example,

$$E_{k|k-1}[x] = E[f(x_{k-1} + v_k | Y_{k-1}] = E[f(x_{k-1})|Y_{k-1}],$$

which is precisely the same as the above, though differently expressed.

Thus, if $p_{k-1|k-1}$ is a Gaussian with mean $x_{k-1|k-1}$ and covariance $P_{k-1|k-1}$, then the Gaussian approximation of $p_{k|k-1}$ has mean $x_{k|k-1}$ and covariance $P_{k|k-1}$ defined by

$$x_{k|k-1} = \int_{R^d} f(t) \frac{1}{((2\pi)^n \det P_{k-1|k-1})^{1/2}}\, e^{-\frac{1}{2}(t-x_{k-1|k-1})^t P_{k-1|k-1}^{-1}(t-x_{k-1|k-1})}\, dt \tag{10.7}$$

and

$$\begin{aligned}
P_{k|k-1} = Q + \textstyle\int_{R^d} (f(t) - x_{k|k-1})(f(t) - x_{k|k-1})^t \frac{1}{((2\pi)^n \det P_{k-1|k-1})^{1/2}} \\
e^{-\frac{1}{2}(t-x_{k-1|k-1})^t P_{k-1|k-1}^{-1}(t-x_{k-1|k-1})}\, dt.
\end{aligned} \tag{10.8}$$

Next, we discuss the corrector step.

$$p_{k|k}(x) = c\,e^{-\frac{1}{2}(y-h(x))^t R^{-1}(y-h(x))} p_{k|k-1}$$

where c is the normalization constant and we assume that $p_{k|k-1}$ is given by the Gaussian approximation defined by (**??**)–(**??**). Define the innovation process

$$I_k = y_k - E_{k|k-1}[y_k] = y_k - E_{k|k-1}[h(x_k)].$$

Here, we again approximate the conditional distribution of $(x(k), h(x(k))$ given the observations up to time $k-1$ by a Gaussian. That is, we approximate $E_{k|k-1}[h(x_k)]$ by its Gaussian approximation z. This means that the probability density function of z is given by the Gaussian distribution with mean $\hat{z}$ and covariance P_{zz} defined by

$$\hat{z} = \int_{R^d} h(t) \frac{1}{((2\pi)^n det\, P_{k|k-1})^{1/2}}\, e^{-\frac{1}{2}(t-x_{k|k-1})^t P_{k|k-1}^{-1}(t-x_{k|k-1})}\, dt \tag{10.9}$$

and

$$P_{zz} = \int_{R^d} (h(t)-\hat{z})(h(t)-\hat{z}) \frac{1}{((2\pi)^n det\, P_{k|k-1})^{1/2}}\, e^{-\frac{1}{2}(t-x_{k|k-1})^t P_{k|k-1}^{-1}(t-x_{k|k-1})}\, dt. \tag{10.10}$$

Finally, we construct the Gaussian approximation of $p_{k|k}$ with mean $x_{k|k}$ and covariance defined by

$$x_{k|k} = x_{k|k-1} + L_k(y(k) - \hat{z}) \tag{10.11}$$

and

$$P_{k|k} = P_{k|k-1} - L_k P_{xz}^t \tag{10.12}$$

where the Kalman filter gain L_k is defined by

$$L_k = P_{xz}(R + P_{zz})^{-1} \tag{10.13}$$

and the covariance P_{xz} is defined by

$$P_{xz} = \int_{R^d} (t - x_{k|k-1})(h(t)-\hat{z})^t \frac{1}{((2\pi)^n det\, P_{k|k-1})^{1/2}}\, e^{-\frac{1}{2}(t-x_{k|k-1})^t P_{k|k-1}^{-1}(t-x_{k|k-1})}\, dt. \tag{10.14}$$

Example If f is a polynomial, then the updates (**??**)–(**??**) have the exact form, for example if f is quadratic, then

$$x_{k|k-1} = f(x_{k-1|k-1}) + f''\Sigma_{k-1|k-1}$$

and

$$P_{k|k-1} = Q + f'(x_{k-1|k-1})P_{k-1|k-1}f'(x_{k-1|k-1})^t + \frac{1}{2}f''P_{k-1|k-1}(f'')^t.$$

10.3 Central differenceFilter

In order to implement the Gaussian filter we must develop the approximation methods to evaluate integrals (**??**)–(**??**). That is, we discuss the approximation methods for the integral of the form

$$\int_{R^d} F(t) \frac{1}{((2\pi)^n det\, \Sigma)^{1/2}}\, e^{-\frac{1}{2}(t-\bar{x})^t \Sigma^{-1}(t-\bar{x})}\, dt, \tag{10.15}$$

where $F(t)$ is a given function. Especially our discussions include the Gauss-Hermite quadrature rule and finite difference approximation. If we assume $\Sigma = SS^t$ and change the coordinate of integration by $t = Sz + \bar{x}$, then

$$I = \int_{R^d} \tilde{F}(z) \frac{1}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}|z|^2}\, dz \tag{10.16}$$

with $\tilde{F}(z) = F(Sz + \bar{x})$.

First, we apply the Gauss-Hermite quadrature rule. The Gauss-Hermite quadrature rule is given by

$$\int_{-\infty}^{\infty} g(x)\, \frac{1}{(2\pi)^{1/2}}\, e^{-x^2}\, dx = \sum_{i=1}^{m} w_i g(x_i)$$

where the equality holds for all polynomials of degree up to $2m-1$ and the quadrature points x_i and the weights are determined (e.g., see [?]) as follows. Let J be the symmetric tri-diagonal matrix with zero diagonals and $J_{i,i+1} = \sqrt{i/2},\ 1 \le i \le m-1$. Then $\{x_i\}$ are the eigenvalues of J and w_i equal to $|(v_i)_1|^2$ where $(v_i)_1$ is the first element of the i-th normalized eigenvector of J. Thus, I is approximated by

$$I_m = \sum_{i_1=1}^{m} \cdots \sum_{i_n=1}^{m} \tilde{F}(q_{i_1}, q_{i_2}, \ldots, q_{i_n})\, w_{i_1} w_{i_2} \ldots w_{i_n} \tag{10.17}$$

where $q_i = \frac{x_1}{\sqrt{2}},\ 1 \le i \le m$ and I_m is exact for all polynomials of the form $z_1^{i_1} z_2^{i_2} \ldots z_n^{i_n}$ with $1 \le i_k \le 2m-1$. In order to evaluate I_m we need m^n-point function evaluations. For example $m = 3$ we have

$$q_1 = -\sqrt{3},\ q_2 = 0,\ q_3 = \sqrt{3} \quad \text{and} \quad w_1 = w_3 = \frac{1}{6},\ w_2 = \frac{2}{3}.$$

and I_3 requires 9-point function evaluations for $n = 2$.

Next, we consider the polynomial interpolation methods. We approximate $\tilde{F}(t)$ by the quadratic function P

$$P(z) = \tilde{F}(0) + \sum_{i=1}^{n} \frac{\tilde{F}(he_i) - \tilde{F}(-he_i)}{2h}\, z_i + \sum_{i=1}^{n} \frac{1}{2}\, H_{i,i} z_i^2 \tag{10.18}$$

Thus,

$$I \sim M = \int_{R^d} P(z)\, \frac{1}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}|z|^2}\, dz = \tilde{F}(0) + \sum_{i=1}^{n} \frac{1}{2}\, H_{i,i} \tag{10.19}$$

and

$$\begin{aligned} J_c &= \int_{R^d} (P^{(1)}(z) - M_1)(P^{(2)}(z) - M_2)\, \frac{1}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}|z|^2}\, dz \\ &= \sum_{i=1}^{n} \frac{\tilde{F}_1(he_i) - \tilde{F}_1(-he_i)}{2h}\, \frac{\tilde{F}_2(he_i) - \tilde{F}_2(-he_i)}{2h} + \tilde{F}_1(0)\tilde{F}_2(0) + \sum_{i=1}^{n} \frac{1}{2}\, H_{i,i}^{(1)} H_{i,i}^{(2)}. \end{aligned} \tag{10.20}$$

If we remove the second order correction term in (**??**)–(**??**), then this coincides with the extended Kalman filter with the central difference approximation of the Jacobian of $\tilde{F}$.

Central Difference Filter

Predictor Step Compute the factorization $P_{k-1|k-1} = SS^t$ and compute the central difference components

$$\begin{aligned} a_i &= \tfrac{f(x_{k-1|k-1} + h\, s_i) - f(x_{k-1|k-1} - h\, s_i)}{2h} \\ H_{i,i} &= \tfrac{f(x_{k-1|k-1} + h\, s_i) - 2f(x_{k-1|k-1}) + f(x_{k-1|k-1} - h\, s_i)}{h^2}. \end{aligned} \tag{10.21}$$

Update $p_{k|k-1} = N(x_{k|k-1}, P_{k|k-1})$ by

$$\begin{cases} x_{k|k-1} = f(x_{k-1|k-1}) + \sum_{i=1}^{d} \frac{1}{2}\, H_{i,i} \\ P_{k|k-1} = Q + \sum_{i=1}^{d} a_i a_i^t + \sum_{i=1}^{n} \frac{1}{2}\, H_{i,i} H_{i,i}^i \end{cases} \tag{10.22}$$

Corrector Step Compute the factorization $P_{k|k-1} = \tilde{S}\tilde{S}^t$ and compute the central difference components

$$\tilde{a}_i = \frac{h(x_{k-1|k-1}+h\,s_i)-h(x_{k-1|k-1}-h\,s_i)}{2h}$$

$$\tilde{H}_{i,i} = \frac{h(x_{k-1|k-1}+h\,s_i)-2h(x_{k-1|k-1})+h(x_{k-1|k-1}-h\,s_i)}{h^2}.$$

Update $p_{k|k} = N(x_{k|k}, P_{k|k})$ by

$$x_{k|k} = x_{k|k-1} + L_k(y(k) - z_k)$$

$$P_{k|k} = P_{k|k-1} - L_k P_{xz}^t$$

where

$$z_k = \tilde{h}(x_{k|k-1}) + \sum_{i=1}^n \frac{1}{2}\tilde{H}_{i,i}$$

$$P_{xz} = \sum_{i=1}^n s_i b_i^t$$

$$P_{zz} = \sum_{i=1}^n \tilde{a}_i \tilde{a}_i^t + \sum_{i=1}^n \frac{1}{2}\tilde{H}_{i,i}\tilde{H}_{i,i}$$

$$L_k = P_{xz}(R + P_{zz})^{-1}.$$

Remark (1) In the algorithm we avoid to calculate the derivatives of f and h. Instead, we use the central difference,which provides a better linearization. The followings are the operation counts in terms of function evaluations.

(2) If we use the quadrature rule based on the q-point Gauss-Hermite rule then the algorithm requires $(d+p)q^d$ function evaluations.

(3) If we use the central difference algorithm, the $(d+p)(2d+1)$ function evaluation are required.

(4) The update (**??**) is the form of square root filters and is very stable.

(5) One can use reduced order directions $\{s_i\}_{i=1}^m$ via the truncated singular value decomposition of $\Sigma_{k-1|k-1}$ using m dominant singular values in update (**??**)–(**??**) without loosing much accuracy.

(6) The stepsize $h > 0$ can be adjusted to obtain a nonlocal behavior of f and for the constraints for state x

10.4 Mixed Gaussian Filter

In this section we discuss the mixed Gaussian filter. We approximate $p_{k-1|k-1}$ by the linear combination of multiple Gaussian distributions, i.e.,

$$p_{k-1|k-1}(x) = \sum_{i=1}^m \alpha_{k-1}^{(i)} \frac{1}{((2\pi)^n det\, P_{k-1|k-1}^{(i)})^{1/2}}\, e^{-\frac{1}{2}(x-x_{k-1|k-1}^{(i)})^t (P_{k-1|k-1}^{(i)})^{-1}(x-x_{k-1|k-1}^{(i)})}.$$

Then, each Gaussian distribution $N(x_{k-1|k-1}^{(i)}, P_{k-1|k-1}^{(i)})$ is updated separately by the Gaussian filter (**??**)–(**??**). Each update is independent from the others and can be performed in a parallel manner. But, we update the weights $\alpha_{k-1|k-1}^{(i)}$ for the new update $p_{k|k-1}(x)$ We use the update via the collocation property:

$$\alpha_{k|k-1}^{(i)}\, N(x_{k|k-1}^{(i)}, P_{k|k-1}^{(i)})(x_{k|k-1}^{(i)}) = \sum_j \alpha_{k-1|k-1}^{(j)}\, N(x_{k-1|k-1}^{(j)}, P_{k-1|k-1}^{(j)})(x_{k|k-1}^{(i)}).$$

For the corrector step we discuss three different weights update for $\alpha_k^{(i)}$ for $p_{k|k}(x)$. First, by equating the zero moment of each Gaussian distribution we obtain

$$\alpha_k^{(i)} \int_{R^d} \frac{1}{((2\pi)^n det\, P_{k|k}^{(i)})^{1/2}}\, e^{-\frac{1}{2}(x-x_{k|k}^{(i)})^t (P_{k|k}^{(i)})^{-1}(x-x_{k|k}^{(i)})}\, dx$$

$$= \alpha_{k-1}^{(i)} \int_{R^d} \frac{1}{(2\pi)^n (det\, R\, det\, P_{k|k-1}^{(i)})^{1/2}}\, e^{-\frac{1}{2}(y-h(x))^t R^{-1}(y-h(x))+(x-x_{k|k-1}^{(i)})^t (P_{k|k-1}^{(i)})^{-1}(x-x_{k|k-1}^{(i)}))}\, dx.$$

Here we approximate the right hand side by the Gaussian distribution as in (**??**)–(**??**) and obtain

$$\alpha_k^{(i)} = \alpha_{k-1}^{(i)} \frac{1}{((2\pi)^n det\,(R+P_{zz}))^{1/2}}\, e^{-\frac{1}{2}(y-\hat{z})^t (R+P_{zz})^{-1}(y-\hat{z})} \tag{10.23}$$

where $\hat{z},\ P_{zz}$ are defined by (**??**)(2.8), which is the update formula discussed in [**?**].

Next, we apply the collocation condition at $x_{k|k}^{(i)}$:

$$\alpha_k^{(i)} \frac{1}{((2\pi)^n det\, P_{k|k}^{(i)})^{1/2}}$$

$$= \alpha_{k-1}^{(i)} \frac{1}{((2\pi)^n det\, P_{k|k-1}^{(i)})^{1/2}}\, e^{-\frac{1}{2}((y-h(x_{k|k}^{(i)}))^t R^{-1}(y-h(x_{k|k}^{(i)}))+(x_{k|k}^{(i)}-x_{k|k-1}^{(i)})^t (P_{k|k-1}^{(i)})^{-1}(x_{k|k}^{(i)}-x_{k|k-1}^{(i)}))}$$

to obtain the update

$$\alpha_k^{(i)} = \alpha_{k-1}^{(i)} \frac{(det\, P_{k|k}^{(i)})^{1/2}}{(det\, P_{k|k-1}^{(i)})^{1/2}} e^{-\frac{1}{2}((y-h(x_{k|k}^{(i)}))^t R^{-1}(y-h(x_{k|k}^{(i)}))+(x_{k|k}^{(i)}-x_{k|k-1}^{(i)})^t (P_{k|k-1}^{(i)})^{-1}(x_{k|k}^{(i)}-x_{k|k-1}^{(i)}))}. \tag{10.24}$$

Finally we discuss the simultaneous update of the weights. We determine the weights $\alpha_k^{(i)}$ by the L^2-projection, i.e., $\alpha_k^{(i)},\ 1 \le i \le m$ minimize

$$\int_{R^d} \left| p_{k|k}(x) - \sum_{i=1}^m \alpha_k^{(i)} \frac{1}{((2\pi)^n det\, P_{k|k}^{(i)})^{1/2}}\, e^{-\frac{1}{2}(x-x_{k|k}^{(i)})^t (P_{k|k}^{(i)})^{-1}(x-x_{k|k}^{(i)})} \right|^2 dx \tag{10.25}$$

over $(R^+)^m$, where $p_{k|k}$ is defined as in the corrector step (**??**) with

$$p_{k|k-1}(x) = \sum_{i=1}^m \alpha_{k-1}^{(i)} \frac{1}{((2\pi)^n det\, P_{k|k-1}^{(i)})^{1/2}}\, e^{-\frac{1}{2}(x-x_{k|k-1}^{(i)})^t (P_{k|k-1}^{(i)})^{-1}(x-x_{k|k-1}^{(i)})}.$$

In order to perform the minimization (**??**) we need to evaluate the integral of the form

$$\int_{R^d} e^{-\frac{1}{2}(y-h(x))^t R^{-1}(y-h(x))} \frac{1}{((2\pi)^n det\, \Sigma)^{1/2}}\, e^{-\frac{1}{2}(x-\bar{x})^t \Sigma^{-1}(x-\bar{x})}\, dx.$$

and it is relatively expensive. Hence we propose the minimization of the sum of collocation distances:

$$\sum_{i=1}^m \left| p_{k|k}(x_{k|k}^{(i)}) - \sum_{j=1}^m \alpha_k^{(j)} \frac{1}{((2\pi)^n det\, P_{k|k}^{(j)})^{1/2}}\, e^{-\frac{1}{2}(x_{k|k}^{(i)}-x_{k|k}^{(j)})^t (P_{k|k}^{(j)})^{-1}(x_{k|k}^{(i)}-x_{k|k}^{(j)})} \right|^2. \tag{10.26}$$

over $\alpha \in R^m$ satisfying $\alpha \ge \alpha_0 > 0$. A positive constant α_0 is chosen so that the likelihood of each Gaussian distribution is nonzero (e.g., $\alpha_0 = 0.001(1,\dots,1)^t \in R^m$). Problem (**??**) is formulated as the quadratic programming

$$\min \quad \frac{1}{2}\,\alpha^t A^t A\alpha - A^t b + \frac{\delta}{2}\,|\alpha|^2 \quad \text{subject to } \alpha \ge \alpha_0, \tag{10.27}$$

where $\delta > 0$ is chosen so that the singularity of the matrix A^tA is avoided and the matrices (A, b) are defined by

$$A_{i,j} = \frac{1}{((2\pi)^n det\, P^{(i)}_{k|k})^{1/2}}\, e^{-\frac{1}{2}(x^{(i)}_{k|k}-x^{(j)}_{k|k})^t (P^{(i)}_{k|k})^{-1}(x^{(i)}_{k|k}-x^{(j)}_{k|k})}$$

$$b_i = \sum_{j=1}^{m} \frac{1}{(2\pi)^n (det\, R\, det\, P^{(j)}_{k|k-1})^{1/2}} e^{-\frac{1}{2}((y-h(x^{(i)}_{k|k}))^t R^{-1}(y-h(x^{(i)}_{k|k}))+(x^{(i)}_{k|k}-x^{(i)}_{k|k-1})^t (P^{(j)}_{k|k-1})^{-1}(x^{(i)}_{k|k}-x^{(j)}_{k|k-1}))}.$$

Thus, we solve (**??**) to obtain the weights $\alpha_k^{(j)}$ at each corrector step by using the existing numerical optimization method (e.g., see [**?**]).

The theoretical foundation of the Gaussian sum approximation as above is that any probability density function can be approximated as closely as desired by a Gaussian sum. More precisely, we state the following error estimate, the proof of which is found in [**?**].

Theorem 5.1: Let $\mathcal{M}$ be a non-negative integer and $\mathcal{N} = 2\mathcal{M} + 2$. For any $\epsilon > 0$ there must exist $D > 0$ and a mask

$$\left\{ \alpha_j \mid j = (j_1, \ldots, j_n)^t \in Z^n,\, |j| = \sum_{s=1}^{n} j_s \leq \mathcal{M} \right\}$$

such that for any density function $p(x) \in C^{\mathcal{N}}(R^d) \cap W^{\mathcal{N}}_\infty(R^d)$ and all $h > 0$ the estimate

$$\|p(x) - p_h(x)\| \leq c_{\mathcal{N}} h^{\mathcal{N}} |p|_{W^{\mathcal{N}}_\infty(R^d)} + \epsilon \|p\|_{W^{\mathcal{N}-1}_\infty(R^d)} \tag{10.28}$$

holds, where $p_h(x)$ is a linear combination of Gaussian distributions given by

$$p_h(x) = \frac{1}{\sqrt{(2\pi)^n D h^2}} \sum_{k \in Z^n} p_k\, e^{-\frac{|x-hk|^2}{2Dh^2}} \tag{10.29}$$

with

$$p_k = \sum_{|j| \leq \mathcal{M}} \hat{\alpha}_j p(h(k-j)), \qquad \hat{\alpha}_j = hD^{-\frac{n-1}{2}} \alpha_j$$

and $c_{\mathcal{N}}$ is a constant independent of $p(x)$. Here $|\cdot|_{W^{\mathcal{N}}_\infty(R^d)}$ and $\|\cdot\|_{W^{\mathcal{N}-1}_\infty(R^d)}$ are defined by

$$|p|_{W^{\mathcal{N}}_\infty(R^d)} = \sum_{|j|=\mathcal{N}} \|\partial^j p\|_{C(R^d)}$$

and

$$\|p\|_{W^{\mathcal{N}-1}_\infty(R^d)} = \sum_{|j|=0}^{\mathcal{N}-1} \frac{|\partial^j p|}{j!}$$

respectively.

Roughly speaking, the estimate (**??**) shows that any probability density function $p(x) \in C^{\mathcal{N}}(R^d) \cap W^{\mathcal{N}}_\infty(R^d)$ can be approximated by a sum of Gaussian distributions each of whose components is given by

$$\frac{1}{\sqrt{(2\pi)^n\, D h^2}} e^{-\frac{|x-hk|^2}{2Dh^2}} \tag{10.30}$$

with order $O(h^{\mathcal{N}})$ for $h \to 0$.

10.5 Continuous-time optimal Nonlinear Filtering theory

In this section we discuss the optimal filtering theory fot the continuous time process. A signal X_t is an $\mathcal{F}_t$-adapted Markov process. Let Y_t be a R^p-valued observation process given by

$$Y_t = \int_0^t h(X_s)\,ds + V_t \tag{10.31}$$

where V_t is a p-dimensional $\mathcal{F}_t$-adapted Brownian motion and $h_t = h(X_t)$ is a predictable process that contains the information of the signal process X_t and satisfies $E[\int_0^T |h_t|^2\,ds] < \infty$. Let $\mathcal{Y}_t$ be the σ-algebra generated by the observation $\{Y_s,\ s \le t\}$. The conditional expectation $\hat{X}_t = E[X_t|\mathcal{Y}_t]$ is the least mean square estimate of the signal X_t. Define the conditional probability distribution π_t by

$$\pi_t(\phi) = E[\phi(X_t)|\mathcal{Y}_t] \quad \text{for } \phi \in C_b^2(R^d) \tag{10.32}$$

Then π_t satisfies the so-called Kushner-Stratonovich equation

$$\pi_t(\phi) = \pi_0(\phi) + \int_0^t \pi_t(\mathcal{A}\phi)\,ds + \int_0^t (\pi_s(h\,\phi) - \pi_s(h)\,\pi_t(\phi), dY_t - \pi_s(h)\,ds), \tag{10.33}$$

for $\phi \in dom(\mathcal{A}$, where $\mathcal{A}$ is the generator of the Markov process X_t. In this section we derive the filter equation (**??**) following Fujisaki-Kallianpur-Kunita [].

We define the innovation process I_t by

$$I_t = Y_t - \int_0^t E[h_s|\mathcal{Y}_s]\,ds. \tag{10.34}$$

Theorem 1 I_t is a $\mathcal{Y}_t$-adapted Brownian motion.

Proof: For $t \ge s$

$$E[I_t - I_s|\mathcal{Y}_s] = E[\int_s^t (h_\sigma - \hat{h}_\sigma)\,ds|\mathcal{Y}_s] + E[V_t - V_s|\mathcal{Y}_s] = 0.$$

Thus I_t is a $\mathcal{Y}_t$ martingale. Now, the quadratic variation of I_t coincides with the one of V_t since I_t is the sum of V_t and a process of bounded variation. That is,

$$\langle I^i, I^j\rangle_t = \langle V^i, V^j\rangle_t = t\,\delta_{i,j}.$$

Thus, the claim follows from the Levy's theorem. □

Since I_t is $\mathcal{Y}_t$ measurable, thus $\sigma(I_s,\ s \le t) \subset \mathcal{Y}_t$. If the two σ-algebras coincide for every $t \ge 0$, then I_t has the innovation property. It has been shown that the innovation property is valid if h_t is independent of V_t and satisfies $E[\int_0^T |h_t|^2\,dt] < \infty$. On the other hand, if h_t and V_t are not independent, the innovation property is not necessary valid. Suppose that the innovation property is satisfied. Then by the martingale representation theorem every $\mathcal{Y}_t$-square integrable martingale has the representation

$$M_t = M_0 + \int_0^t (\Phi_s, dI_s)$$

where $\Phi_s \in L^2(\langle I\rangle)$. We will show that this representation holds without assuming the innovation property. The claim is based on the Girsanov transformation.

Grisanov's Theorem Let B_t is a d-dimensional; Brownian motion on $(\Omega, \mathcal{F}, P)$ and $\mathcal{F}_t = \sigma(B_s,\ s \le t)$. Let g_t be a $\mathcal{F}_t$-predictable process with $\int_0^t |g_s|^2\,ds < \infty$ a.s.. Define

$$\alpha_t = exp(\int_s^t (g_s, dB_s) - \frac{1}{2}\int_0^t |g_s|^2\,ds)$$

and assume that $E[\alpha_t] = 1$ holds for all $t \geq 0$. Define the probability measure Q on $\mathcal{F}_T$ by

$$Q(A) = \int_A \alpha_t \, dP, \quad A \in \mathcal{F}_t.$$

Then, (1) If M_t is a $(\mathcal{F}_t, P)$-local martingale, then

$$\tilde{M}_t = M_t - \langle \int (g_s, dB_s), M \rangle_t$$

is a $(\mathcal{F}_t, Q)$-local martingale. In particular $\tilde{B}_t = B_t - \int_0^t g_s \, ds$ is a $(\mathcal{F}_t, Q)$ Brownian motion.
(2) Every square integrable $(\mathcal{F}_t, Q)$ martingale $\tilde{M}_t$ is represented by

$$\tilde{M}_t = \tilde{M}_0 + \int_0^t (\Psi_s, d\tilde{B}_s),$$

where Ψ_t is an $\mathcal{F}_t$ predictable process such that $E_Q[\int_0^T |\Psi_t|^2 \, dt] < \infty$.

Proof: Since $d\alpha_t = g_t \alpha_t \, dB_t$, by Ito's rule

$$\alpha_t \tilde{M}_t = \tilde{M}_0 + \int_0^t \tilde{M}_s d\alpha_s + \alpha_s d\tilde{M}_s + \int_0^t \alpha_s g_s d\langle B, M \rangle_s$$

$$= \tilde{M}_0 + \int_0^t g_s \alpha_s \tilde{M}_s dB_s + \int_0^t \alpha_s dM_s.$$

Thus, $\alpha_t \tilde{M}_t$ is a $(\mathcal{F}_t, P)$-local martingale and therefore $\tilde{M}_t$ is a $(\mathcal{F}_t, Q)$-local martingale.

To prove the second assertion, we observe

$$\alpha_t^{-1} = exp(-\int_0^t (g_s, d\tilde{B}_s) - \frac{1}{2} \int_0^t |g_s|^2 \, ds)$$

and thus $dP = \alpha_t^{-1} \, dQ$ on $\mathcal{F}_t$. Thus if $\tilde{M}_t$ is a $(\mathcal{F}_t, Q)$ martingale, then

$$M_t = \tilde{M}_t + \langle \int (g_s, d\tilde{B}_s), \tilde{M} \rangle_t$$

is $(\mathcal{F}_t, Q)$ martingale. By the martingale representation theorem

$$M_t = M_0 + \int_0^t (\Psi_s, dB_s)$$

and thus

$$\tilde{M}_t = \int_0^t (\Psi_s, d\tilde{B}_s) + \int_0^t (\Psi_s, g_s) \, ds - \langle \int (g_s, d\tilde{B}_s), \tilde{M} \rangle_t.$$

We note that the first term of the right hand side is a $(\mathcal{F}_t, P)$ martingale. Thus, the remaining term should be a martingale. But since it is a process of bounded variation and a continuous martingale with bounded variation is zero,it is zero. $\square$.

Corollary Let $h(X_t)$ be a $\mathcal{F}_t$-predictable process with $\int_0^t |h(X_s)|^2 \, ds < \infty$ a.s.. Define

$$\beta_t = exp(\int_s^t -(h(X_s), dB_s) - \frac{1}{2} \int_0^t |h(X_s)|^2 \, ds)$$

and assume that $E[\beta_t] = 1$ holds for all $t \geq 0$. Define the probability measure Q on $\mathcal{F}_T$ by

$$Q(A) = \int_A \beta_t \, dP, \quad A \in \mathcal{F}_t.$$

Then, (1) If M_t is a $(\mathcal{F}_t, P)$-local martingale, then $Y_t = V_t + \int_0^t h(X_s) \, ds$ is a $(\mathcal{F}_t, Q)$ Brownian motion.

Fujisaki-Kallianpur-Kunita Theorem Let $\mathcal{Y}_t$ be the σ-algebra generated by the observation process Y_t. Suppose that M_t is a square integrable $\mathcal{Y}_t$ martingale. Then there exists a $\mathcal{Y}_t$ predictable process Φ_t such that

$$M_t = M_0 + \int_0^t (\Phi_s, dI_s).$$

Proof: Let

$$\beta_t = exp(-\int_0^t (\hat{h}_s, dI_s) - \frac{1}{2}\int_0^t |\hat{h}_s|^2\, ds)$$

and define the probability measure by $dQ = \beta_t\, dP$ on $\mathcal{Y}_t$. Then $I_t + \int_0^t \hat{h}_s\, ds = Y_t$ is a $(\mathcal{Y}_t, Q)$ Brownian motion by Girsanov theorem. Hence every square integrable $(\mathcal{Y}_t, Q)$ martingale is represented as a stochastic integral by Y_t. Observe that

$$\alpha_t = \beta_t^{-1} = exp(\int_0^t (\hat{h}_s, dY_s) - \frac{1}{2}\int_0^t |\hat{h}_s|^2\, ds)$$

and $dP = \alpha_t\, dQ$. Then by Girsanov theorem I_t is the basis of $(\mathcal{Y}_t, P)$ martingale, i.e., every square integrable $(\mathcal{Y}_t, P)$ martingale is represented as a stochastic integral by I_t. $\square$

Let X_t be an Ito's process

$$X_t = X_0 + \int_0^t f_s\, ds + W_t$$

where W_t is a square integrable $(\mathcal{F}_t, P)$ martingale. Define

$$M_t = \hat{X}_t - \hat{X}_0 - \int_0^t \hat{f}_s\, ds \tag{10.35}$$

is $\mathcal{Y}_t$ martingale. In fact,

$$E[M_t - M_s|\mathcal{Y}_s] = E[\hat{X}_t - \hat{X}_s - \textstyle\int_s^t \hat{f}_\sigma\, d\sigma|\mathcal{Y}_s]$$

$$= E[X_t - X_s - \textstyle\int_s^t f_\sigma\, d\sigma|\mathcal{Y}_s] = E[W_t - W_s|\mathcal{Y}_s] = 0.$$

Thus, by FKK theorem we have $M_t = \int_0^t (\Phi_s, dI_s)$. We have the following representation of Φ_t.

Theorem 2

$$M_t = \int_0^t (E[X_s h(X_s)|\mathcal{Y}_s] - E[X_s|\mathcal{Y}_s]E[h(X_s)|\mathcal{Y}_s] + E[D_s|\mathcal{Y}_s], dY_s - E[h(X_s)|\mathcal{Y}_s]\, ds)$$

where D_t is the predictable process defined by $\displaystyle\int_0^t D_s\, ds = \langle W, V\rangle_t$.

Proof: If we let $\varphi_t = \int_0^t (f_s - \hat{f}_s)\, ds$, then

$$M_t = W_t + \varphi_t + (\hat{X}_t - X_t) + X_0 - \hat{X}_0$$

By FKK theorem for $Z_t \in \mathcal{M}_2$ there exists $g_s \in \mathcal{L}_2$ such that

$$Z_t = \int_0^t g_s\, dI_s = \int_0^t (g_s, dV_t) + \int_0^t (g_s, h_s - \hat{h}_s)\, ds = N_t + \psi_t.$$

Then

$$E[M_t Z_t] = E[(W_t + \varphi_t + X_0 - \hat{X}_0)Z_t].$$

By Ito's formula

$$(W_t+\varphi_t+X_0-\hat{X}_0)Z_t=\mathcal{F}_t\text{-martingale}+\int_0^t (W_t+\varphi_t+X_0-\hat{X}_0)\,d\psi_t+\int_0^t Z_s\,d\varphi_s+\langle W,V\rangle_t.$$

Since $W_t+\varphi_t+(X_0-\hat{X}_0)=X_t-\hat{X}_0+\int_0^t \hat{f}_s\,ds$,

$$E[\int_0^t (W_t+\varphi_t+X_0-\hat{X}_0)d\psi_s]=E\int_0^t [X_s\,(g_s,h_s-\hat{h}_s)]=\int_0^t E[(g_s,\widehat{X_sh_s}-\hat{X}_s\hat{h}_s]\,ds.$$

Since

$$E[\int_0^t Z_s d\varphi_s]=\int_0^t E[Z_s\,(g_s,f_s-\hat{f}_s)]\,ds=0,$$

it follows that

$$E[M_tZ_t]=E[\int_s^t (g_s,\widehat{X_sh_s}-\hat{X}_s\hat{h}_s+\hat{D}_s)\,ds$$

for all $Z_t\in\mathcal{M}_2$, which proves the theorem. □

Let X_t be Ito's diffusion process, i.e., $dX_t=f(X_t)\,dt+\sigma(X_t)\,dB_t$. By Ito's formula for $\phi\in C_b^2(R^d)$

$$\phi(X_t)-\phi(X_0)=\int_0^t \mathcal{A}\phi(X_s)\,ds+\tilde{W}_t$$

where

$$\tilde{W}_t=\int_0^t \phi_{x_i}(X_s)\sigma_{i,j}(X_s)\,dB_s$$

is a $(\mathcal{F}_t,P)$-martingale. It follows from Theorem 2 that for $\phi\in C_b^2(R^d)$

$$\begin{aligned}E[\phi(X_t)|\mathcal{Y}_t]&=E[\phi(X_0)]+\int_0^t E[\mathcal{A}\phi(X_s)|\mathcal{Y}_s]\,ds\\&\quad+\int_0^t (E[\mathcal{M}\,\phi(X_s)h(X_s)|\mathcal{Y}_s]-E[\phi(X_s)|\mathcal{Y}_s]E[h(X_s)|\mathcal{Y}_s],dY_s-E[h(X_s)|\mathcal{Y}_s]\,ds)\end{aligned}\tag{10.36}$$

where

$$\mathcal{M}\phi=h\,\phi+\sum_{i,j}\phi_{x_i}\,\sigma_{i,j}\,\langle W^j,V\rangle$$

since

$$\langle\tilde{W},V\rangle_t=\int_0^t \phi_{x_i}(X_s)\sigma_{i,j}(X_s)\,d\langle W^j,V\rangle_s.\square$$

From the theorem the optimal estimate $\hat{X}_t$ of the state X_t given observations $\mathcal{Y}_t$ satisfies the optimal filter equation

$$d\hat{X}_t=\widehat{f(X_t)}\,dt+G(t)(dY_t-\widehat{h(X_t)}\,dt)\tag{10.37}$$

where

$$\widehat{\phi(X_t)})=E[\phi(X_t)|\mathcal{Y}_t]=\int_{R^d}\phi(x)pi(t,x)\,dx$$

and

$$G(t)=\int_{R^d}(x-\hat{x}_t)(h(x)-\widehat{h(X_t)})^t\pi(t,x)\,dx,\tag{10.38}$$

where we assume $D_t=0$. If the conditional density function $\pi(t,x)$ exists, i.e.,

$$\pi_t(\phi)=\int_{R^d}\phi(x)\pi(t,x)\,dx,$$

then $\pi(t,x)$ satisfies the Kushner's equation

$$d\pi(t,x) = \mathcal{A}(t)^*\pi\,dt + (h-\pi(h))\pi(dY_t - \pi(h)\,dt). \tag{10.39}$$

Also, the error covariance $\Sigma(t)$ defined by

$$\Sigma(t) = E[(x_t-\hat{x}_t)(x_t-\hat{x}_t)^t|\mathcal{Y}_t] = \int_{R^d}(x-\hat{x}_t)(x-\hat{x}_t)^t\pi(t,x)\,dt$$

satisfies

$$d\Sigma(t) = (F(t)\Sigma(t)+\Sigma(t)F(t)^t - G(t)R^{-1}G(t)+Q)\,dt + C(t)\,dY_t, \tag{10.40}$$

where

$$F(t)\Sigma(t) = \textstyle\int_{R^d}(f(x)-\widehat{f(X_t)})(x-\hat{x}_t)^t\pi(t,x)\,dx,$$

$$C(t) = \textstyle\int_{R^d}(x-\hat{x}_t)(x-\hat{x}_t)^t(h(x)-\widehat{h(X_t)})\pi(t,x)\,dx.$$

In fact, by the Ito's rule

$$d(x-\hat{x}_t)(x-\hat{x}_t)^t\pi(t,x) = -(\widehat{f(X_t)}(x-\hat{x}_t)^t + \widehat{f(X_t)}^t(x-\hat{x}_t))\pi\,dt + (x-\hat{x}_t)(x-\hat{x}_t)^t\mathcal{A}^*\pi\,dt$$

$$-2(x-\hat{x}_t)(h(x)-\widehat{h(X_t)})G(t)^t\pi\,dt + G(t)G(t)^t\pi\,dt + (x-\hat{x}_t)(x-\hat{x})^t(h(x)-\widehat{h(X_t)})\pi\,dY_t$$

$$+((x-\hat{x}_t)G(t)(dY_t-\widehat{h(X_t)}\,dt)^t + G(t)(dY_t-\widehat{h(X_t)}\,dt)(x-\hat{x}_t)^t).$$

Since

$$\int_{R^d}(x-\hat{x}_t)(x-\hat{x}_t)^t\mathcal{A}^*\pi(t,x)\,dx = \int_{R^d}(f(x)(x-\hat{x}_t)^t+(x-\hat{x}_t)f(x)^t+Q)\pi(t,x)\,dx$$

from(**??**) we obtain (**??**).

For the linear Gaussian case: $f(t,x)=A(t)x+b(t)$, $h(t,x)=H(t)x$, $\pi_0=N(m,\Sigma_0)$, we have $\pi(t,x)=N(\hat{x}_t,\Sigma(t))$ and from (**??**) we obtain the Kalman-Bucy filter:

$$\begin{cases} d\hat{x}_t = (A(t)\hat{x}_t + b(t)\,dt + G(t)(dY_t - H(t)\hat{x}_t\,dt). \quad \hat{x}_0 = m \\ G(t) = \Sigma(t)H^t \\ \frac{d}{dt}\Sigma(t) = A(t)\Sigma(t)+\Sigma(t)A(t)^t - \Sigma(t)H^tR^{-1}H\Sigma(t)+Q(t). \quad \Sigma(0)=\Sigma_0. \end{cases} \tag{10.41}$$

Corrorary Let S be a complete metric space and $B(S)$ be the space of all bounded measurable functions on S. Let X_t be an S-valued Markov process X_t. That is,

$$P(X_t\in B|\sigma(X_r,\ r\le s)) = P(X_t\in B|\sigma(X_s))$$

for every Borel set B. Let $P(s,x,t,E)$ be the transition probability distribution of X_t and define

$$(P_s^tf)(y) = \int_S P(s,dx,t,y)f(x), \quad f\in B(S)$$

for all Borel sets B. Then P_s^t is a linear bounded operator on $B(S)$. Let $\mathcal{D}$ be a subspace of $B(S)$. A family of linear operators $\mathcal{U}_t$ defined on $\mathcal{D}$, satisfying

$$P_s^tf - f = \int_s^t P_s^\sigma\mathcal{U}_\sigma f\,d\sigma, \text{ for all } f\in\mathcal{D}$$

is called the generator of P_s^t. Then, for $f\in\mathcal{D}$

$$W_t(f) = f(X_t) - f(X_0) - \int_0^t \mathcal{U}_sf(X_s)\,ds$$

is a $\mathcal{F}_t$ martingale. In fact, for $t \geq s$

$$E[W_t(f) - W_s(f)|\mathcal{F}_s] = E[f(X_t) - f(X_s) - \textstyle\int_s^t \mathcal{U}_\sigma f(X_\sigma)\,d\sigma|\mathcal{F}_s]$$

$$= E[f(X_t)|\mathcal{F}_s] - f(X_s) - E[\int_s^t \mathcal{U}_\sigma f(X_\sigma)\,d\sigma|\mathcal{F}_s].$$

Since by the Markov property $E[f(X_t)|\mathcal{F}_s] = P_s^t f(X_s)$, thus

$$E[f(X_t)|\mathcal{F}_s] - f(X_s) = P_s^t f(X_s) - f(X_s) = \int_s^t P_s^\sigma \mathcal{U}_\sigma f(X_\sigma)\,d\sigma = \int_s^t E[\mathcal{U}_\sigma f(X_\sigma)\,d\sigma|\mathcal{F}_s]\,d\sigma.$$

and therefore $E[W_t(f) - W_s(f)|\mathcal{F}_s] = 0$.

Let $Y_t = \int_0^t h(s, X_s) + V_t$ and assume that the processes X_t and V_t are independent. Then we have

$$\widehat{f(X_t)} - \widehat{f(X_0)} = \int_0^t E[\mathcal{U}_s f(X_s)|\mathcal{Y}_s]\,ds + \int_0^t (\widehat{f_s\,h_s} - \hat{f}_s\,\hat{h}_s, dI_s)$$

where $f_s = f(X_s)$ and $h_s = h(s, X_s)$. Define $\pi_t(B) = E[\chi_B(X_t)]$ for $B \in \mathcal{B}(S)$. Now, suppose there exists a measure λ on $(S, \mathcal{B}(S))$ such that for every $t \geq 0$ $P(X_t \in dx)$ is absolutely continuous with respect to $\lambda(dx)$. Then π_t is a probability measure and absolute continuous with respect to λ a.s. and let $\rho_t(x,\omega)$ be the Radon-Nikodym derivative of π_t, i.e. $\pi_t(B) = \int_B \rho_t(x,\omega)\,\lambda(dx)$. We define the dual operator $\mathcal{U}_s^*$ of $\mathcal{U}_s$ by

$$\int_S \mathcal{U}_s^* g(x)\lambda(dx) = \int_S g(x)\mathcal{U}_s f(s)\lambda(dx).$$

It follows from (**??**) that if $\rho_t(x)$ is in domain of $\mathcal{U}_t^*$ for $t \geq 0$, then

$$\rho_t(x) = \rho_0(x) + \int_0^t \mathcal{U}_s^* \rho_s(x)\,ds + \int_0^t (\rho_s(x)(h(s,x) - \int_S h(s,y)\rho_s(y)\lambda(dy)), dI_s). \tag{10.42}$$

Examples (1) For Ito's diffusion process X_t

$$\mathcal{U}_t = a_{i,j}(t,x)\,\frac{\partial^2}{\partial x_i \partial x_j} + b_i(t,x)\,\frac{\partial}{\partial x_i}$$

and

$$\mathcal{U}_t^* \rho = \frac{\partial^2}{\partial x_i \partial x_j}(a_{i,j}\,\rho) - \frac{\partial}{\partial x_i}(b_i\,\rho).$$

(2) Let X_t be a finite state Markov chain process, i.e., $S = \{1, 2, \cdots, n\}$ and

$$\lim_{\Delta t \to 0^+} \frac{P_t^{t+\Delta t}(i,j) - \delta_{i,j}}{\Delta t} = q_{i,j}(t).$$

Then $P_s^t(i,j)$ satisfies

$$P_s^t(i,j) = \delta_{i,j} + \int_s^t P_s^\sigma(i,k) q_{k,j}(\sigma)\,d\sigma.$$

Thus, if we define

$$(\mathcal{U}_t f)(i) = \sum_{j \in S} q_{i,j}(t) f(j)$$

then $\mathcal{U}_t$ is the generator of the Markov process X_t. Moreover,

$$(\mathcal{U}_t^* q)(i) = \sum_{j \in S} \rho(j) q_{j,i}(t)$$

and hence $\rho_t(i) = P(X_t = i|\mathcal{Y}_s)$ satisfies

$$\rho_t(i) = \rho_0(i) + \sum_{j \in S} \int_0^t \rho_s(j)\,q_{j,i}(s)\,ds + \int_0^t (\rho_s(i)(h(s,i) - \sum_{j \in S} \rho_s(j)\,h(s,j), dI_s). \tag{10.43}$$

10.6 Zakai equation

The Kushner's equation is for the conditional probability density function $\pi = \pi(t,x)$:

$$\pi_t(\phi) = \pi_0(\phi) + \int_0^t \pi_s(\mathcal{A}(s)\phi)\,ds + \int_0^t \pi_s(\mathcal{M}(s)\phi) - \pi_s(\phi)\,\pi_s(h), dY_s - \pi_s(h)\,ds) \qquad (10.44)$$

for all $\phi \in C_0^2(R^d)$ and It is a Brownian motion Y_t driven stochastic PDE with the cubic nonlinearity on $(\Omega, \mathcal{Y}_t, Q)$. If X_t is an Ito's diffusion process X_t, for $\phi \in C_b^2(R^d)$

$$\mathcal{A}(t)\phi = \sum_{i,j} a_{i,j}(t,x)\,\frac{\partial^2\phi}{\partial x_i \partial x_j} + \sum_i b_i(t,x)\,\frac{\partial\phi}{\partial x_i}$$

and

$$\mathcal{M}(t)\phi = h(t,x)\,\phi + \sum_{i,j} \frac{\partial\phi}{\partial x_i}\,\sigma_{i,j}(t,x)\,\langle W^j, V\rangle.$$

We will transform it to a linear stochastic PDE, so-called Zakai equation.

Theorem (1) Suppose π_t is a solution to the Kushner's equation (**??**) and let

$$\alpha_t = exp(\int_0^t (\pi_s(h_s), dY_s) - \frac{1}{2}\int_0^t |\pi_s(h_s)|^2\,ds).$$

Then $\rho_t(\phi) = \pi_t(\phi)\alpha_t$ satisfies the Zakai equation

$$\rho_t(\phi) = \rho_0(\phi) + \int_0^t \rho_s(\mathcal{A}(s)\phi)\,ds + \int_0^t (\rho_s(\mathcal{M}(s)\phi), dY_s) \qquad (10.45)$$

for all $\phi \in dom(\mathcal{A}$.

(2) If ρ_t is a solution to the Zakai equation (**??**), then $\pi_t(\phi) = \dfrac{\rho_t(\phi)}{\rho_t(1)}$ satisfies the Kushner's equation.

Proof: (1) By Ito's formula

$$d(\pi_t(\phi)\alpha_t) = \pi_t(\phi)d\alpha_t + \alpha_t d\pi_t(\phi) + d\langle\alpha, \pi(\phi)\rangle_t.$$

Since

$$\pi_t(\phi)d\alpha_t = (\pi_t(\phi)\pi_t(h)\alpha_t, dY_t)$$

$$\alpha_t\,d\pi_t(\phi) = \pi_t(\mathcal{A}(t)\phi)\alpha_t\,dt + \alpha_t(\pi_t(\mathcal{M}(t)\phi) - \pi_t(\phi)\pi_t(h), dY_t - \pi_t(h)\,dt)$$

$$d\langle\alpha, \pi(\phi)\rangle_t = \alpha_t\,(\pi_t(h), \pi_t(\mathcal{M}(t)\phi) - \pi_t(\phi)\pi_t(h))\,dt$$

thus ρ_t satisfies

$$d\rho_t(\phi) = \rho_t(\mathcal{A}(t)\phi)\,dt + (\rho_t(\mathcal{M}(t)\phi), dY_t).$$

(2) From (**??**)

$$\rho_t(1) = \rho_0(1) + \int_0^t (\rho_s(h), dY_s) = \rho_0(1) + \int_0^t (\pi_s(h)\rho_s(1), dY_s).$$

where $\rho_t(\mathcal{M}(t)1) = \rho_t(h) = \pi_t(h)\rho_t(1)$. Hence $\rho_t(1)$ is given by

$$\rho_t(1) = \rho_0(1)\,exp(\int_0^t (\pi_s(h), dY_s) - \int_0^t |\pi_s(h)|^2\,ds).$$

Thus, $\beta_t = \rho_t(1)^{-1}$ satisfies

$$d\beta_t = \beta_t\,(\pi_t(h)), dY_t) + |\pi_t(h)|^2\beta_t\,dt.$$

By Ito's formula

$$d(\rho_t(\phi)\beta_t) = \rho_t(\phi) d\beta_t + \beta_t\, d\rho_t(\phi) + d\langle \beta, \rho(\phi)\rangle_t.$$

Since

$$\rho_t(\phi)\, d\beta_t = (\pi_t(\phi)\pi_t(h), dY_t) - |\pi_t(h)|^2\pi_t(\phi)\, dt$$

$$\beta_t\, d\rho_t(\phi) = \pi_t(\mathcal{A}(t)\phi)\, dt + (\pi_t(\mathcal{M}(t)\phi\, h), dY_t)$$

$$d\langle \beta, \rho(\phi)\rangle_t = -(\pi_t(\mathcal{M}(t)\phi\, h, \pi_t(h)\, dt.$$

π_t satisfies the Kushner's equation. □

Remark The Zakai equation is linear equation

$$d\rho(t,x) = \mathcal{A}^*(t)\rho\, dt + \mathcal{M}(t)\rho\, dY_t$$

and is a Brownian Motion driven Fokker Planck equation on $(\Omega, \mathcal{Y}_t, Q)$.

10.7 Relation to the Discrete Time Optimal Filter

Consider the discrete-time observation

$$y_k = h(X_{t_k}) + V_k$$

The optimal filter is given by

$$\begin{aligned} &\frac{\partial}{\partial t}\pi(t,x) = \mathcal{A}^*(t)\pi \text{ on } [t_k, t_{k+1}), \quad \pi(t_k) = \pi(t_k^+) \\ &\pi(t_{k+1}^+, x) = c\, e^{-\frac{(y_k - h(x))^t R^{-1}(y_k - h(x)}{2}}\pi(t_{k+1}, x). \end{aligned} \tag{10.46}$$

The predictor step (??) of the discrete-time optimal filter is the time splitting method of the Fokker-Planck predictor step (??) when $\sigma\sigma^t = Q$.

10.8 Gaussian filter

We consider a Gaussian approximation of the optimal nonlinear filter (??) by approximating the conditional probability density π by the normal distribution

$$N(t,x) = N(\hat{x}(t), P(t))(x) = \frac{1}{(2\pi)^n det P(t))^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\hat{x}(t))^t P(t)^{-1}(x-\hat{x}(t))}.$$

Thus, we have the filter equation

$$d\widehat{x(t)}^{(1)} = \widehat{f(x(t))}^{(1)} + L^{(1)}(t)(dy_t - \widehat{h(x(t))}^{(1)}\, dt), \tag{10.47}$$

where

$$\widehat{f(x(t))}^{(1)} = \int f(x)N(t,x)\, dx$$

$$\widehat{h(x(t))}^{(1)} = \int h(x)N(t,x)\, dx.$$

and

$$L^{(1)}(t) = \int (x - \widehat{x(t)}^{(1)})(h(x) - \widehat{h(x(t))}^{(1)}))^t N(t,x)\, dx.$$

In order to obtain the update for the covariance $P(t) = P^{(1)}(t) = E[(x(t) - \hat{x(t)}^{(1)})(x(t) - \hat{x(t)}^{(1)})^t]$ we note that

$$d(x(t) - \hat{x(t)}^{(1)}) = (f(x) - \widehat{f(x(t))}^{(1)})\,dt - L^{(1)}(t)(h(x(t)) - \widehat{h(x(t))}^{(1)})\,dt$$

$$+\sigma\,dw_1(t) - L^{(1)}(t)dw_2(t).$$

If we assume $h(x) = Hx$, then

$$L^{(1)}(t) = P^{(1)}(t)H^t.$$

Thus, it follows from (**??**) that

$$\frac{d}{dt}P^{(1)}(t) = F^{(1)}(\widehat{x(t)}^{(1)}, P^{(1)}(t)) - L^{(1)}(t)L^{(1)}(t)^t + Q, \tag{10.48}$$

where

$$F^{(1)}(\widehat{x(t)}^{(1)}, P^{(1)}(t)) = \int (f(x) - \widehat{f(x(t))}^{(1)})(x - \widehat{x(t)}^{(1)})^t N(t,x)\,dx$$

$$+\int (x - \widehat{x(t)}^{(1)})(f(x) - f\widehat{(x(t))^{(1)}})^t N(t,x)\,dx.$$

Note that $\widehat{f(x(t))}^{(1)}$ and $\widehat{h(x(t))}^{(1)}$ depend on $P(t) = P^1(t)$ and thus (**??**) and (**??**) are coupled. In order to the filter equation separated from the covariance update we consider

$$d\widehat{x(t)}^{(2)} = f(\widehat{x(t)}^{(2)}) + L^{(2)}(t)(dy_t - h(\widehat{x(t)}^{(2)})\,dt), \tag{10.49}$$

$$\frac{d}{dt}P^{(2)}(t) = F^{(2)}(\widehat{x(t)}^{(2)}, P^{(2)}(t)) - L^{(2)}(t)L^{(2)}(t)^t + Q. \tag{10.50}$$

where $L^{(2)}(t) = P^{(2)}(t)H^t$ and $N(t,x) = N(\widehat{x(t)}^{(2)}, P^{(2)}(t))(x)$ and

$$L^{(2)}(t) = \int (x - \widehat{x(t)}^{(2)})(h(x) - h(\widehat{x(t)})^{(2)}))^t N(t,x)\,dx.$$

$$F^{(2)}(\widehat{x(t)}^{(2)}, P^{(2)}(t)) = \int (f(x) - f(\widehat{x(t)}^{(2)})(x - \widehat{x(t)}^{(2)})^t N(t,x)\,dx$$

$$+\int (x - \widehat{x(t)}^{(2)})(f(x) - f(\widehat{x(t))}^{(2)})^t N(t,x)\,dx.$$

Example (parameter dependent linear dynamics) Let $f(x,a) = A(a)x$ where a represents coefficients of matrix $A(a) \in R^{n\times n}$ for $a \in R^m$, i.e., $\frac{\partial A}{\partial a} = \dot{A} \in R^{n\times m}$. Assume that $N(t,m,\Sigma)$ is the normal distribution with mean $m = (\hat{x}, \hat{a})$ and covariance Σ

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Then,

$$\widehat{A(a)x} = A(\hat{a})\hat{x} + \int (A(a)(x - \hat{x}) + \dot{A}\hat{x}(a - \hat{a}) + (A(a) - A(\hat{a})(x - \hat{x})\,dN(t,m,\Sigma)$$

$$= A(\hat{a})\hat{x} + (\dot{A}\hat{x})\Sigma_{21}.$$

and

$$\int (A(a)x - \widehat{A(a)x})(x - \hat{x})^t\,dN(t,m,\Sigma) = A(\hat{a})\Sigma_{11} + \dot{A}\Sigma_{21}$$

Thus,we obtain the system for $(\hat{x}, \hat{a}, \Sigma)$:

$$\begin{cases} d\hat{x} = (A(\hat{a})\hat{x} + (\dot{A}\hat{x})\Sigma_{2,1})\,dt + \Sigma_{11}H^tR^{-1}(dy_t - H\hat{x}\,dt) \\ \\ d\hat{a} = \Sigma_{2,1}H^tR^{-1}(dy_t - H\hat{x}\,dt) \\ \\ \dfrac{d}{dt}\Sigma = J(\hat{x},\hat{a})\Sigma + \Sigma J(\hat{x},\hat{a})^t - \Sigma H^tR^{-1}H\Sigma + Q = 0 \end{cases}$$

where

$$J(\hat{x}, \hat{a}) = \begin{pmatrix} A(\hat{a}) & \dot{A}\hat{x} \\ \\ 0 & 0 \end{pmatrix}.$$

If we let $B = \dot{A}\hat{x}$, then the stationary equation is given by

$$\begin{cases} A\Sigma_{11} + \Sigma_{11}A^* + B\Sigma_{21} + \Sigma_{12}B^* - \Sigma_{11}H^*R^{-1}H\Sigma_{11} + Q_{11} = 0. \\ \\ A\Sigma_{12} + B\Sigma_{22} - \Sigma_{11}H^*R^{-1}H\Sigma_{12} + Q_{12} = 0. \\ \\ -\Sigma_{12}H^*R^{-1}H\Sigma_{12} + Q_{22} = 0. \end{cases}$$

Thus, if $Q_{12} = 0$, then

$$\Sigma_{12} = -(A - \Sigma_{11}H^*R^{-1}H)^{-1}B\Sigma_{22}$$

11 Optimal Stopping Problem and Variational inequality

The theory of optimal stopping is concerned with the problem of determining an optimal stopping time of the Markov process X_t to maximize an expected reward or minimize an expected cost. Optimal stopping problems can be found in areas of statistics, economics, and mathematical finance (related to the pricing of American options). Optimal stopping problems can often be written in the variational inequalty for a Bellman equation, and are therefore often solved using dynamic programming.

Consider the Ito diffusion $X_t,\ t \geq 0$. Problem is to find an optimal stopping time τ^* such that

$$E^x[g(X_{\tau^*})] = V(x) = \sup_\tau E^x[g(X_\tau)],$$

where $V(x)$ is the value function. A (continuous) function $g \geq 0$ is the reward function. We will show under appropriate conditions, the value function V satisfies the variational inequality:

$$\max(\mathcal{A}V, g - V) = 0, \quad a.e.$$

and for the domain D defined by

$$D = \{x \in \Omega : V(x) > g(x)\}$$

the optimal stopping time τ^* is given by

$$\tau^* = \inf\{t : X_t(\omega) \notin D\}.$$

Definition A measurable function f is called super-maenvalued function if $f(x) \geq E^x[f(X_\tau)]$ for all stopping time τ and $x \in R^n$. If in addition f is lower-semicontinuos, then f is called superharmonic.

Remark (1) By the Fatou's lemma

$$f(x) \leq E^x[\liminf_{n\to\infty} f(X_{\tau_n})] \leq \liminf_{n\to\infty} E^x[f(X_{\tau_n})]$$

for any sequence τ_n of stopping times satisfying $\tau_n \to 0^+$, if f is lower semicontinous. Thus if f is super harmonic,

$$f(x) = \lim_{n\to\infty} E^x[f(X_{\tau_n})] \quad \text{for all } x \in R^d$$

(2) If $f \in C^2(R^d)$ it follows from the Dynkin's formula that f is super harmonic if and only if

$$\mathcal{A}f \leq 0.$$

where $\mathcal{A}$ is the characteristic orator of the diffusion process X_t.
(3) If $X_t = x + B_t$ is the Brownian motion, then the super harmonic function coincides with the super harmonic function in the classical potential theory.

Definition f is the super harmonic majorant of g if $f \geq g$ and f is superharmonic. Define the least superharmonic majorant $\hat{g}$ if $\hat{h} \leq f$ for any super harmonic majorant of g.

Lemma (Construction of the least superharmonic majorant) Let $g = g_0$ be a lower nonnegative semicomtinuous function R^d and define the sequence $\{g_n\}$ by

$$g_n(x) = \sup_{t\in S_n} E^x[g_{n-1}(X_t)],$$

where $S_n = \{k2^{-n},\ 0 \leq k \leq 4^n\}$. Then $g_n \uparrow \hat{g}$ and $\hat{g}$ is the least superharmonic majorant of g.

Proof: By the definition $\{g_n\}$ is nondecreasing. Define $\tilde{g}(x) = \lim_{n\to\infty} g_n(x)$. Then,

$$\tilde{g}(x) \geq g_n(x) = E^x[g_{n-1}(X_t)] \text{ for all n and } t \in S_n.$$

Thus,

$$\tilde{g}(x) \geq \lim_{n\to\infty} E^x[g_{n-1}(X_t)] = E^x[\tilde{g}(X_t)]$$

for all $t \in S = \bigcup_{n=1}^\infty S_n$. Since $\tilde{g}$ is a monotone limit of lower semicontinuous functions, then $\tilde{g}$ is lower semicontinuous. For a fixed $t \in R$ choose $t_n \to t$ and we have

$$\tilde{g}(x) \geq \liminf_{n\to\infty} E^x[\tilde{g}(X_{t_n})] \geq E^x[\liminf_{n\to\infty} \tilde{g}(X_{t_n})] \geq E^x[\tilde{g}(X_t)]$$

by the Fatou lemma. Hence $\tilde{g}$ is an excessive function. It is shown in Dynkin [] that $\tilde{g}$ is a superharmonic majorant of g. If f is any superharmonic majorant of g, the by induction one can show that

$$f(x) \geq g_n(x) \text{ for all } n$$

and thus $f(x) \geq \tilde{g}(x)$. Hence $\tilde{g}$ is the least superharmonic majorant. □

Theorem (Optimal Stopping) Let $V(x)$ is the optimal reward function and $\hat{g}$ be the least superharmonic majorant of a continuous reward function $g \geq 0$. Then, $V(x) = \hat{g}$. Define the continuation region

$$D = \{x : V(x) > g(x)\}.$$

Suppose there exists an optimal stopping time τ^*, then $\tau^* \geq \tau_D$ for all $x \in D$ and $\tau^* = \tau_D$, i.e., $V(x) = E^x[g(X_{\tau_D})]$.

Proof: Since $\hat{g}$ is superharmonic majorant of g

$$\hat{g}(x) \geq E^x[\hat{g}(X_\tau)] \geq E^x[g(X_\tau)].$$

Thus,

$$V(x) = \sup_\tau E^x[g(X_\tau)] \leq \hat{g}(x)$$

Assume g is bounded. Define

$$D_\epsilon = \{x : g(x) < \hat{g}(x) - \epsilon\}$$

and

$$\hat{g}_\epsilon(x) = E^x[\hat{g}(X_{\tau_\epsilon})]$$

Then, $\hat{g}_\epsilon(x)$ is superharmonic . One can prove that

$$g(x) \le \hat{g}_\epsilon(x) + \epsilon$$

Thus, $\hat{g}_\epsilon + \epsilon$ is a superharmonic majorant of g and

$$\hat{g} \le \hat{g}_\epsilon + \epsilon = E^x[\hat{g}(X_{\tau_\epsilon})] + \epsilon \le E^x[(g + \epsilon(X_{\tau_\epsilon})] + \epsilon \le V(x) + 2\epsilon$$

which implies $\hat{g} = V(x)$ since $\epsilon > 0$ is arbitrary. Let $x \in D$ and τ be a stopping time with $P^x(\tau < \tau_D) > 0$. Since $g(X_\tau) < V(X_\tau)$ if $\tau < \tau_D$ and $g \le V$, we have

$$E^x[g(X_\tau)] = \int_{\tau<\tau_D} g(X_\tau)dP^x \int_{\tau\ge\tau_D} g(X_\tau)dP^x$$

$$< \int_{\tau<\tau_D} V(X_\tau)dP^x + \int_{\tau\ge\tau_D} g(X_\tau)dP^x = E^x[V(X_\tau)] \le V(x)$$

since $V = \hat{g}$ is superharmonic.

For $x \in D$ we have

$$V(x) = E^x[g(X_{\tau^*})] \le E^x[\hat{g}(X_{\tau^*})] \le E^x[\hat{g}(X_{\tau_D})]$$

$$= E^x[g(X_{\tau_D})] \le V(x)$$

For $x \in \partial D$ and assume $\tau_D > 0$. Let τ_k be a sequence of stopping times such that $0 < \tau_k < \tau_D$ and $\tau_k \to 0$ a.s. as $k \to \infty$. Since $X_{\tau_k} \in D$ and X_t is a strong Markov process we have

$$E^x[g(X_{\tau_D})] = E^x[\theta_{\tau_k} g(X_{|tau_D})] = E^x[E^{X_{\tau_k}}[g(X_{\tau_D})]] = E^x[V(X_{\tau_D})].$$

Hence since V is lower semi-continuous and by the Fatou's lemma

$$V(x) \le E^x[\liminf_{k\to\infty} V(X_{\tau_k})] \le \liminf_{k\to\infty} E^x[V(X_{\tau_k})] = E^x[g(X_{\tau_D})]$$

Hence $V(x) = E^x[g(X_{\tau_D})]$. □

Remark (1) Let X_t be a deterministic process and let

$$g(x) = \frac{x^2}{1+x^2}.$$

Then, $V(x) = 1$. But the optimal stooping time does not exists.
(2) If $g \in C^2(\Omega)$ and let

$$U = \{\mathcal{A}g > 0\}.$$

Then, $U \subset D$. In fact, for $x \in U$ by Dynkin's formula

$$V(x) \ge E^x[g(X_\tau)] = g(x) + E^x[\int_0^\tau Ag(X_s)\,ds] > g(x)$$

and thus $U \subset D$.

Example 1 (Running cost) Let τ_Ω be the stopping time of an open set Ω in R^d. Consider the optimal stopping time problem with a running cost of the form:

$$V(x) = \sup_\tau E^x[g(X_\tau) + \int_0^\tau f(X_s)\,ds]$$

over all stopping time $0 \le \tau \le \tau_\Omega$. Consider the augmented system:

$$\begin{cases} dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t, \quad X_0 = x \\ \\ dW_t = f(X_t)\,dt, \quad W_0 = 0. \end{cases}$$

The, we have

$$V(x) = \sup_\tau E^x[g(X_\tau) + W_\tau].$$

and the corresponding generator is given by

$$A\tilde{\phi}(x,w) = \mathcal{A}\tilde{\phi} + f\frac{\partial\tilde{\phi}}{\partial w}(x,w).$$

For $\tilde{\phi}(x,w) = \phi(x) + w$ we have

$$A\tilde{\phi}(x,w) = \mathcal{A}\phi + f(x).$$

<u>**Theorem (Variational Inequality I)**</u> Let $\phi \in H^2(\Omega)$ satisfies

$$\max(\mathcal{A}\phi + f, g - \phi) = 0.$$

Let

$$D = \{x \in \Omega : \phi(x) > g(x)\}$$

Then τ_D is the an optimal stooping time and $\phi = V$ is the optimal value function, i.e.,

$$\phi(x) = \sup_{0\le\tau\le\tau_\Omega} E^x[g(X_\tau) + \int_0^\tau f(X_s)\,ds].$$

Proof: For $\epsilon > 0$ let $\phi^\epsilon = \int_\Omega \epsilon^{-n}\zeta(\frac{x-y}{\epsilon})\phi(y)\,dy$ be a molifier of ϕ. Then, by the Dynkin's formula

$$E^x[\phi^\epsilon(X_\tau)] = \phi^\epsilon(x) + E^x[\int_0^\tau (\mathcal{A}\phi^\epsilon + f)(X_s)]$$

Letting $\epsilon \to 0^+$

$$E^x[\phi(X_\tau)] = \phi(x) + E^x[\int_0^\tau (\mathcal{A}\phi + f)(X_s)]$$

since $\mathcal{A}\phi^\epsilon \to \mathcal{A}\phi$ a.e. $x \in \Omega$. For $x \in D$ and $\tau < \tau_D$

$$E^x[g(X_\tau)] < E^x[\phi(X_\tau)] = \phi(x) + E^x[\int_0^\tau (\mathcal{A}\phi + f)(X_s)] = \phi(x)$$

and thus $V(x) = \phi(x)$. For $x \notin D$ since

$$E^x[\phi(X_\tau)] = E^x[g(X_\tau)] = \phi(x) + E^x[\int_0^\tau (\mathcal{A}\phi + f)(X_s)\,ds] \le \phi(x).$$

$V(x) = \phi(x)$. $\square$

<u>Example 2 (Time dependent case)</u>

$$V(x) = \sup_{0\le\tau\le T} E^x[e^{-\int_0^\tau q(X_s)\,ds} g(X_\tau) + \int_0^\tau e^{-\int_0^s q(X_r)\,dr} f(X_s)\,ds] = \sup_{0\le\tau\le T} E^x[Z_\tau g(X_\tau) + W_\tau],$$

where for the augmented dynamics

$$\begin{cases} dt = dt \\ \\ dX_t = b(t,,X_t)\,dt + \sigma(X_t)\,dB_t, \quad X_0 = x \\ \\ dW_t = Z_t f(t,X_t)\,dt, \quad W_0 = 0. \\ \\ dZ_t = -q(X_t)Z_t\,dt, \quad Z_0 = 1, \end{cases}$$

the generator is given by

$$A\tilde{\phi}(t,x,w,z) = \frac{\partial}{\partial t}\tilde{\phi} + \mathcal{A}\phi + zf(t,x)\frac{\partial}{\partial w}\tilde{\phi} - q(x)z\frac{\partial}{\partial z}\tilde{\phi}.$$

For $\tilde{\phi}(t,x,w,z) = z\phi(t,x) + w,\ z > 0$, we have

$$A\tilde{\phi}(x,w) = z(\frac{\partial}{\partial t}\phi + \mathcal{A}\phi + f(t,x) - q(x)\phi).$$

Theorem (Variational Inequality II) Let $\phi \in H^{1,2}(0,T,\Omega)$ satisfies

$$\max\left(\frac{\partial}{\partial t}v + \mathcal{A}v - q(x)v + f(t,x), g - v\right) = 0, \quad v(T,x) = g(x).$$

Let

$$D = \{x \in \Omega : v(t,x) > g(x)\}$$

Then τ_D is the an optimal stooping time and v is the optimal value function, i.e.,

$$v(t,x) = \sup_{t \le \tau \le T} E^{t,x}[e^{-\int_t^\tau q(X_s)\,ds} g(X_\tau) + \int_0^\tau e^{-\int_t^s q(X_r)\,dr} f(s,X_s)\,ds].$$

American Option

$$v(t,x) = \max_{t \le \tau \le T} E^{t,x}[e^{-r(\tau - t)}\psi(X_\tau)]$$

for the stock process

$$dX_t = rX_t\,dt + \sigma X_t\,dB_t.$$

Let $\psi = (K-x)^+$ (put option) or $\psi = (x-K)^+$ (call option), where K is the strike price and $T > 0$ is the maturity time. Then the Black Scholes American option equation is given by

$$\max\left(\frac{\partial}{\partial t}v + rx\,v_x + \frac{\sigma^2 x^2}{2}v_{xx} - r\,v, \psi - v\right) = 0 \text{ a.e. in } (0,T) \times R, \quad v(T,x) = \psi(x).$$

In general, consider a stochastic X_t defined on the probability space $(\Omega, \mathcal{F}, P)$ with filtration $\mathcal{F}_t$ and assume that X_t is adapted to the filtration. The optimal stopping problem is to find the stopping time τ^* which maximizes the expected reward

$$V(t,x) = E^{t,x}[g(X_{\tau^*})] = E^{t,x}_{t \le \tau \le T}[g(X_\tau)]$$

where $V(t,x)$ is called the value function. Here T can take value ∞. A more specific formulation is as follows. We consider an $\mathcal{F}_t$-adapted strong Markov process $X = \{X_t,\ t \ge 0\}$. Given continuous functions $g,\ f$, the optimal stopping problem is

$$V(x) = \sup_{0 \le \tau \le T} E^{0,x}[g(X_\tau) + \int_0^\tau f(X_t)\,dt].$$

Let X_t be a Levy process in R^d given by the SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t + \int_{R^n} \gamma(X_{t_-}, z)\bar{N}(dt,dz), \quad X_0 = x$$

where B_t is an m-dimensional Brownian motion, $\bar{N}$ is an l-dimensional compensated Poisson random measure, $b : R^d \to R^d$, $\sigma : R^d \to R^{d\times m}$, and $\gamma : R^n \times R^n \to R^{k \times l}$ are given functions such that a unique solution X_t exists. Let $\Omega \subset R^n$ be an open set and τ_Ω is the exit time from Ω of X_t:

$$\tau_\Omega = \inf\{t > 0 : X_t \notin \Omega\}.$$

$$V(x) = \sup_{\tau \le \tau_\Omega} J^\tau(y) = \sup_{\tau \le \tau_S} E^{0,x}[g(X_\tau) + \int_0^\tau f(X_t)\,dt].$$

It turns out that under some regularity conditions, the following verification theorem holds:

If a function $\phi : \bar{\Omega} \to R$ satisfies

- $\phi \in C(\bar{\Omega}) \cap C^1(\Omega) \cap C^2(\Omega \setminus \partial D)$, where the continuation region is $D = \{x \in \Omega : \phi(x) > g(x)\}$,
- $\phi \geq g$ on Ω, and
- $\mathcal{A}\phi + f \leq 0$ on $\Omega \setminus \partial D$, where $\mathcal{A}$ is the infinitesimal generator of X_t,

then $\phi(x) \geq V(x)$ for all $x \in \bar{\Omega}$. Moreover, if

$$\mathcal{A}\phi + f = 0 \text{ on } D$$

Then $\phi(x) = V(x)$ for all $x \in \bar{\Omega}$ and $\tau^* = \inf\{t > 0 : X_t \notin D\}$ is an optimal stopping time.

These conditions can also be written is a more compact form (the integro-differential variational inequality):

$$\max\{\mathcal{A}\phi + f, g - \phi\} = 0 \text{ on } \Omega \setminus \partial D.$$

12 Stochastic Optimal Control and Hamilton-Jacobi-Bellman equation

In this section we consider the stochastic optimal control problem.

12.1 Discrete time

Consider the discrete time control problem; minimize

$$J^u = E_{\{w_k\}_{k=i}^{N-1}}[\sum_{k=0}^{N-1} Q(X_k) + h(u_k) + G(X_N)|X_0 = x]$$

over controls $\{u_k\}_{k=0}^{N-1} \in U$, subject to

$$X_{n+1} = f(X_n, u_n, w_n), \quad X_0 = x,$$

where X_n is an S-valued process, u_k is $\mathcal{F}_k$-adapted, and $\{w_k\}$ is the independent, identically distributed random variables.
Let $\{V_n\}$ be the value function

$$V_n(x) = \min_{\{u_k\}_{k=n}^{N-1} \in U} E[\sum_{k=n}^{N-1}(Q(X_k) + h(u_k) + g(X_N)|X_n = x].$$

The optimality principle is given by the so-called Bellman's dynamic programming:

$$V_n(x) = \min_{u \in U} E_w[Q(x) + h(u) + V_{n+1}(f(x,u,w))], \quad V_N(x) = g(x) \tag{12.1}$$

for $n \leq N-1$ and the optimal feedback law is

$$u_n(x) = \text{argmin}_{u \in U} E_w[Q(x) + h(u) + V_{n+1}(f(x,u,w))].$$

In fact, since $\{w_k\}$ are independent

$$V_n(x) = \min_{\{u_k\}_{k=n}^{N-1} \in U} E_{\{w_k\}_{k=n}^{N-1}}[\textstyle\sum_{k=n}^{N-1}(Q(x_k) + h(u_k) + g(X_N))$$

$$= \min_{u_n \in U) \in U} E_{w_n}[Q(x) + h(u_n) + \min_{\{u_k\}_{k=n+1}^{N-1} \in U} \sum_{k=n+1}^{N-1} E_{\{w_k\}_{k=n+1}^{N-1}}[(Q(x_k) + h(u_k) + g(X_N))]$$

$$= E_w[(Q(x) + h(u)) + V_{n+1}(y)|y = f(x,u,w)]$$

Example (Markov chain case) In the case finite state control Markov chain with Let $S = \{1, 2, \cdots n\}$ and P^u_{ij} be the controlled transition probability. Then, the DP principle (**??**) is written as

$$V_n(i) = \min_{u\in U} \{Q(i) + h(u) + \sum_j P^u_{ij} V_{n+1}(j)\}.$$

Example (Deterministic case) In the deterministic the DP principle (**??**) reduces to

$$V_n(x) = \min_{u\in U} \{Q(x) + h(u) + V_{n+1}(f(x,u))\}.$$

Consider the linear quadratic case; minimize

$$\sum_{k=0}^{N-1} (x_k, Qx_k) + (Ru_k, u_k) + (x_N, Gx_N)$$

subject to

$$x_{n+1} = Ax_n + Bu_n$$

Then,

$$V_n(x) = (x, P_n x)$$

where the self-adjoint operator on X is defined by

$$P_n = A^* P_{n+1} A - A^* PB(R + B^* P_{n+1} B)^{-1} B^* PA + Q$$

with $P_N = G$. In fact

$$(x, Qx) + (u, Ru) + (Ax + Bu, P_{n+1}(Ax + Bu))$$

$$= ((R + B^* PB)u, u) + (B^* P_{n+1} Ax, u) + (x, Qx) + (P_{n+1} Ax, Ax)$$

is minimized at

$$u^* = -(R + B^* PB)^{-1} B^* P_{n+1} Ax$$

and thus

$$V_n(x) = (x, Qx) + (P_{n+1} Ax, Ax) - ((R + B^* PB)^{-1} B^* P_{n+1} Ax, B^* P_{n+1} Ax) = (x, P_n x).$$

12.2 Continuous time stochastic optimal control problem

Let Ω be an open set in R^d and $\tau = \tau_\Omega$ is the exit time from Ω. We consider the stochastic optimal control problem:

$$\min \quad J^u(s,x) = E^{s,x}[\int_s^{T\wedge\tau} f^0(t, X_t, u_t)\,dt + g(X_{T\wedge\tau})] \tag{12.2}$$

subject to

$$dX_t = b(X_t, u_t)\,dt + \sigma(X_t, u_t)\,dB_t \tag{12.3}$$

over $u_t \in U_{ad}$, i.e., a progressively measurable control such that $\{u \in U \text{ a.s.}\}$ where U is a closed convex set in R^m. Here we assume the running cost f^0 and the terminal cost g is Lipschitz. Define the optimal value function V by

$$\min \quad V(s,x) = \inf_{u\in U_{ad}} J^u(s,x).$$

Theorem (Hamilton-Jacobi-Bellman) Suppose $V \in C^{1,2}(0,T)$ satisfies

$$\frac{\partial}{\partial t} V(t,x) + \min_{u\in U} \{\mathcal{A}^u V + f^0(t,x,\nabla V)\} = 0, \quad V(T,x) = g(x), \quad x \in \Omega \text{ and } V(t,x) = g(x), \quad x \in \partial\Omega, \tag{12.4}$$

where the generator $\mathcal{A}^u,\ u \in U$ is given by

$$\mathcal{A}^u V = b(t,x,u)\cdot\nabla V + \frac{1}{2}\,\mathrm{tr}\,((\sigma(t,x,u)\sigma(t,x,u)^t)\nabla^2 V),$$

Then, V is the value function and the optimal control is a Markov control

$$u_t^* = \mathrm{argmin}_{u\in U}\{\mathcal{A}^u V(t,X_t) + f^0(t,X_t,u)\}. \tag{12.5}$$

Proof: By the Dynkin's formula

$$J^u(s,x) = V(s,x) + E^{s,x}[\int_s^{T\wedge\tau}(\frac{\partial V}{\partial t} + \mathcal{A}^u + f^0(t,X_t,u_t))\,dt] \geq V(s,x)$$

for all $u_t \in U_{ad}$. Moreover the equality hold if and only if u_t satisfies (**??**). □

Note that

$$u_t^* = \mathrm{argmin}_{u\in U}\{b(t,x,u)\cdot\nabla V(t,x) + \tfrac{1}{2}tr\sigma\sigma(t,x,u)^t\nabla^2 V + f^0(t,x,u)\}$$

$$= \Psi(t,x,\nabla V(t,x),\nabla^2 V(t,x))$$

(**??**) is a Markov control.

For example consider the case when $b(t,x,u) = f(x) + B(x)u$, $\sigma(t,x,u) = \sigma(t,x)$ and $f^0(t,x,u) = \ell(x) + h(u)$. Then, we have

$$u_t^* = \mathrm{argmin}_{u\in U}(h(u) + (\nabla V(t,x), B(x)u))$$

Let $h^*(\lambda)$ is the convex conjugate of h defined by

$$h^*(\lambda) = \sup_{u\in U}\{(\lambda,u) - h(u)\}.$$

Then,

$$\min_{u\in U}\{\mathcal{A}^u V + f^0(t,x,u)\} = -h^*(-B(x)^*\nabla_x V(t,x)) + \ell(x) + \mathcal{A}_0 V(t,x).$$

Note that $h*(\lambda)$ is convex and if ∂h^* is the subdifferential of h^* then

$$\mathrm{argmax}_{u\in U}\{(\lambda,u) - h(u)\} \in \partial\, h^*(p).$$

In fact, for $t \in [0,1]$

$$h^*(t\lambda_1 + (1-t)\lambda_2) = \sup_{u\in U}\{t((\lambda_1,u) - h(u)) + (1-t)((\lambda_1,u) - h(u)\} = \leq t\,h^*(\lambda_1) + (1-t)\,h^*(\lambda_2).$$

If u^* is maximizer at $\bar\lambda$ then

$$h^*(\lambda) \geq (\lambda - \bar\lambda, \bar{u^*}) + h^*(\bar\lambda)$$

for all λ, i.e., $u^* \in \partial h^*(\bar\lambda)$.

In general we have

$$u^* = \mathrm{argmin}_{u\in U}(h(u) + (b,p) + \frac{1}{2}\mathrm{Tr}(\sigma^* Q\sigma\}$$

for $p = \nabla_x V$ and $Q = V_{xx}$ (the Heasean of V).

Corollary (Discrete State S) Consider the control Markov process on $S = \{1,\cdots n\}$ with the transition probability P_{ij}^u. The value function V satisfies

$$\frac{d}{dt}V(t,i) + \min_{u\in U}\{f^0(t,i,u) + \sum_j P_{ij}^u V(t,j)\} = 0$$

and the optimal control is given by

$$u_t(i) = \text{argmin}_{u\in U}\{f^0(t,i,u) + \sum_j P^u_{ij} V(t,j)\}.$$

In general let $\{X_t\}$ be a Markov process with the controlled generator $\mathcal{A}^u$. Consider the optimal control of the form

$$\min \quad E^{s,x}[J^u(s,x)] \text{ over } u \in \mathcal{U}_{ad}.$$

Then, by the Dynkin's formula for the Markov process, the HJB theorem holds.

Example (Linear Quadratic Gaussian regulator problem) Consider the linear control system

$$dX_t = (AX_t + Bu_t)\,dt + dW_t, \quad X_0 \in N(m,\Sigma)$$

and the quadratic costfunctional

$$E^{s,x}[\int_0^T (Qx_t,x_t) + |u_t|^2)\,dt + (x_T, Gx_t)].$$

Then, let $V(x) = (x,P,x)$.

Example (Optimal Portfolio selection)

$$\min \quad E^{t,x}[N(Z_{\tau_G\wedge T})]$$

subject to $0 \le u_t \le 1$ and

$$dZ^u_t = (au_t + b(1-u_t))Z_t\,ds + \alpha u_t Z_t\,dB_t$$

$$V_t + \max_{u\in[0,1]}\{(au+b(1-u))xV_x + \frac{1}{2}\alpha^2u^2x^2\,V_{xx}\} = 0$$

The minimum of

$$(au+b(1-u)))xV_x + \frac{1}{2}\alpha^2x^2u^2\,V_{xx}\}$$

over $[0,1]$ is attained at

$$u^* = \min(1, \max(0, -\frac{(a-b)V_x}{\alpha^2xV_{xx}}))$$

assuming $V_{xx} > 0$. Thus,

$$V_t + bV_x - \frac{1}{2}(a-b)^2|V_x|^2)/(\alpha^2V_{xx}), \quad V(T,x) = N(x). \tag{12.6}$$

For the utility function $U(x) = x^\gamma$ we assume

$$V(t,x) = f(t)x^\gamma, \quad \text{for } N(x) = x^\gamma.$$

Then, from (**??**)

$$f'(t) + b\gamma f + \frac{1}{2}\frac{\gamma(a-b)^2}{\alpha^2(1-\gamma)}\,f$$

and

$$f(t) = e^{\lambda(T-t)}, \quad \lambda = b\gamma + \frac{1}{2}\frac{\gamma(a-b)^2}{\alpha^2(1-\gamma)}.$$

Thus,

$$u^*_t = \frac{b-a}{\alpha^2(1-\gamma)}$$

In general by Dynkin's formula

$$E^{t,x}[N(Z_{\tau_G})] = N(x) + E^{t,x}[\int_t^{\tau_G} (\mathcal{A}^u N)(X_s)\, ds]$$

and thus

$$u_t^* = \text{argmax}_{u\in[0,1]}(\mathcal{A}^u N)(X_t)$$

is the optimal policy. For example if $N(x) = \log(x)$ we have

$$E^{t,x}[\text{log}Z_t] = \text{log}x + E^{t,x}\int_0^\tau (au + b(1-u) - \frac{1}{2}\alpha^2)\, ds$$

Thus,

$$u^* = \frac{a-b}{\alpha^2}$$

is the optimal control for the Kelly condition $N(x) = \text{log}x$.

12.3 Viscosity solution and Verification Theorem

Definition (Viscosity solution

$$-\phi_t - \min_{u\in U}\{A^u\phi(t,x) + f^0(t,x,u)\} \le 0$$

$$-\psi_t - \min_{u\in U}\{A^u\psi(t,x) + f^0(t,x,u)\} \ge 0$$

Definition It is said that $(q,p,Q) \in (D^{1,2}_{t,x})^+ v(t,x)$ is the second order one-sided parabolic supper-differential of v at (t,x) if for $s \ge t$ and $y \in R^n$

$$v(s,y) \le v(t,x) + q(s-t) + (p, y-x) + \frac{1}{2}(y-x, Q(y-x)) + o(s-t+|y-x)|^2).$$

Theorem (Verification) Let $v \in C(0,T;R^n)$ be a viscosity sub-solution of the HJB equation satisfying $|v(t,x)| \le C(1+|x|^k)$ for some k such that $v(T,x) = g(x)$. Then,
(i) $v(s,y) \le J^u(s,y)$ for all (s,y) and admissible control u.
(ii) Given, (s,y), let (x^*,u^*) is an admissible pair of (**??**) . Suppose there exists a square integral $\mathcal{F}_t$ adapted process (q^*,p^*,Q^*) such that

$$(q^*(t),p^*(t),Q^*(t)) \in (D^{1,2}_{t,x})^+ v(t,x) \quad dt\times dP \text{ a.s.},$$

and

$$E[\int_s^T q^*(t) + f^0(t,x^*(t),u^*(t)) + (b(t,x^*(t),u^*(t)),p^*(t))$$

$$+\frac{1}{2}Tr(\sigma(t,x^*(t),u^*(t))^* Q\sigma(t,x^*(t),u^*(t)))\, dt] \le 0$$

Then, (x^*,u^*) is an optimal pair for the optimal control problem (**??**)–(**??**).

13 Stochastic Calculus with Jump Process

14 Backward SDE